LinuxCon North America 2012

LTTng 2.0 : Tracing, Analysis and Views for Performance and Debugging.

E-mail: mathieu.desnoyers@efficios.com



> Presenter

- Mathieu Desnoyers
- EfficiOS Inc.
 - http://www.efficios.com
- Author/Maintainer of
 - LTTng, LTTng-UST, Babeltrace, Userspace RCU



> Content

- Tracing benefits,
- LTTng 2.0 Linux kernel and user-space tracers,
- LTTng 2.0 usage scenarios & viewers,
- New features ready for LTTng 2.1,
- Conclusion



> Benefits of low-impact tracing in a multi-core world

- Understanding interaction between
 - Kernel
 - Libraries
 - Applications
 - Virtual Machines
- Debugging
- Performance tuning
- Monitoring



> Tracing use-cases

- Telecom
 - Operator, engineer tracing systems concurrently with different instrumentation sets.
 - In development and production phases.
- High-availability, high-throughput servers
 - Development and production: ensure high performance, low-latency in production.
- Embedded
 - System development and production stages.



> LTTng 2.0

- Rich ecosystem of projects,
- Key characteristics of LTTng 2.0:
 - Small impact on the traced system, fast, useroriented features.
- Interfacing with: Common Trace Format (CTF)

Interoperability Between Tracing Tools with the Common Trace Format (CTF),
Mathieu Desnoyers, EfficiOS,
Tracing Summit,
Aug. 30, Room Nautilus 3, 14h45.

Tracing Well With Others: Integration of GDB Tracepoints Into Trace Tools,
Stan Shebs, Mentor Graphics,
Tracing Summit,
Aug. 30, Room Nautilus 3, 11h15.

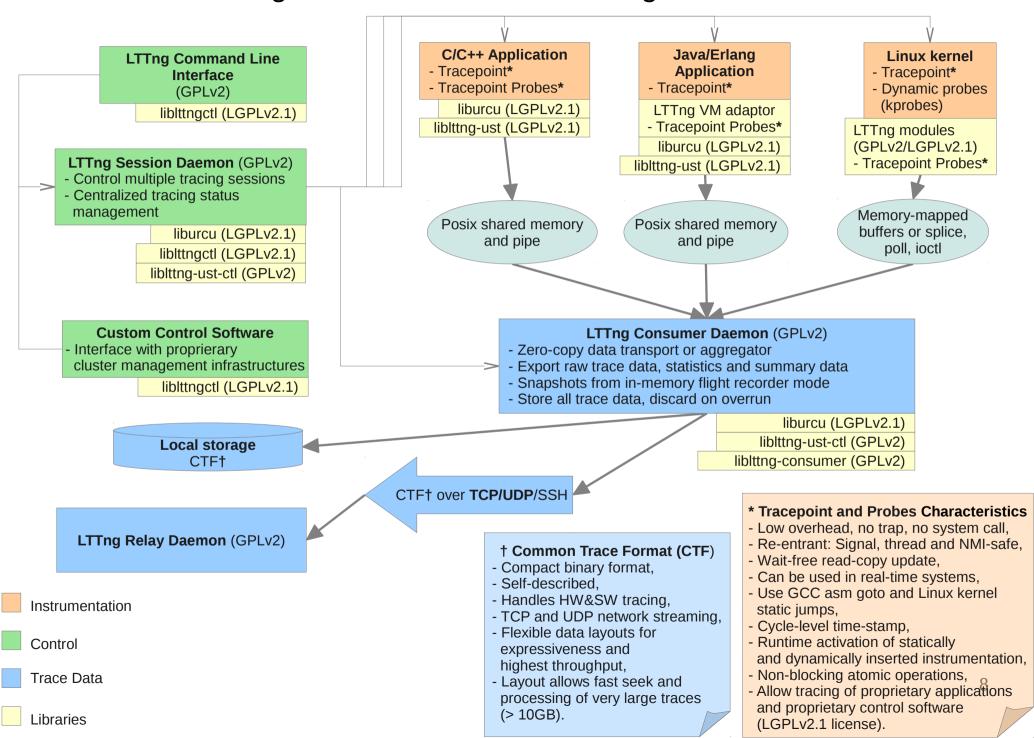


> LTTng 2.0 Tracers

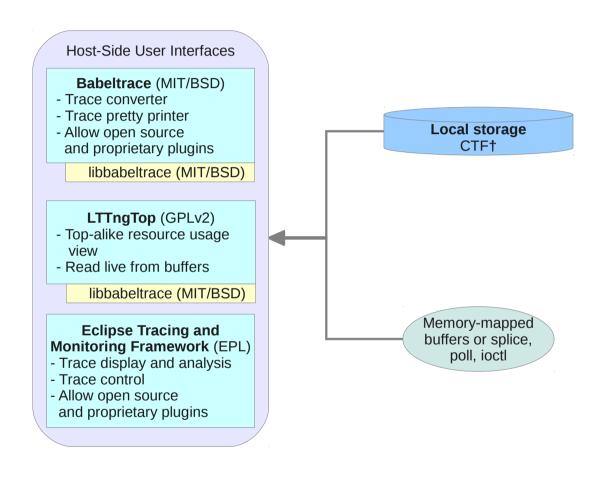
- Controlled by Ittng-sessiond(8),
- User interact with single command line UI, "lttng(1)",
- LTTng modules: Linux kernel tracing,
- LTTng-UST library: user-space tracing,



LTTng 2.0 Low-Overhead Tracing Architecture



LTTng 2.0 Common Trace Format Viewers



Trace Data

Libraries

9

> Trace Control Libraries and Bindings

- liblttng-ctl and lttng/lttng.h expose a C/C++ LGPL v2.1 API to control tracing.
- LTTng tracing control Python bindings planned to be merged into lttng-tools 2.2
 - http://git.lttng.org/?p=lttng-tools.git master branch, in extras/.
- Noteworthy: GDB is also using Python.



> LTTng 2.0 Linux Kernel Tracer

- Build against a vanilla or distribution kernel
 2.6.38+, without need for additional patches,
 and back to 2.6.32 with backport of 3 upstream kernel patches,
- Instrumentation sources: Tracepoints, System calls, Function tracer, kprobes, and kretprobes,
- Supports multiple concurrent tracing sessions,
- Flight recorder mode, snapshots, supported at the tracer level, not supported by lttng-tools 2.0 yet (planned for 2.2).



> LTTng 2.0 Kernel Tracer

- Supports dynamically selectable "context" information to augment event payload
 - Any Perf Performance Monitoring Unit counter
 - PID, PPID, TID, process name, VPID, VTID, ...
 - Dynamic Priority, nice value



> LTTng 2.0 User-Space Tracer

- Supports: Linux, FreeBSD, OpenBSD, NetBSD,
- Tracing performed directly in user-space through shared memory map, without calling the kernel (for speed),
- Support TRACEPOINT_EVENT instrumentation: dynamically enabled, statically defined, user-space instrumentation.
- Supports multiple concurrent tracing sessions.



> LTTng-UST 2.0 User-space Tracer Features

- TRACEPOINT_EVENT() API for application/library static instrumentation with sdt.h gdb/systemtap integration.
- Per-user tracing.
- System-wide tracing.
 - "tracing" group: no need to be root to perform system-wide tracing.



> TRACEPOINT_EVENT

In header:

```
TRACEPOINT_EVENT(ust_tests_hello, tptest,
     TP ARGS(int, anint, long *, values,
           char *, text, size t, textlen,
           double, doublearg, float, floatarg),
     TP FIELDS(
          ctf integer(int, intfield, anint)
          ctf_integer_hex(int, intfield2, anint)
          ctf array(long, arrfield1, values, 3)
          ctf_sequence(char, seqfield1, text,
                   size t, textlen)
          ctf string(stringfield, text)
          ctf float(float, floatfield, floatarg)
          ctf float(double, doublefield, doublearg)
```

Tracepoint name convention



> User-level Tracepoint

Name convention

```
< [com_company_]project[_component] >, < event >
Where "company" is the name of the company,
   "project" is the name of the project,
   "component" is the name of the project component (which may
   include several levels of sub-components, e.g.
   ...component_subcomponent_...) where the tracepoint is located
   (optional),
   "event" is the name of the tracepoint event.
```

Tracepoint invocation within the code:



> Usage Scenarios

- Tracing,
- Analyzing trace data,
- Tracing across kernel and user-space,
- Tracing across multiple nodes.



> Tracing: record kernel trace

```
File Edit View Search Terminal Help

test4@thinkos:~$ lttng create

Session auto-20120827-120832 created.

Traces will be written in /home/test4/lttng-traces/auto-20120827-120832

test4@thinkos:~$ lttng enable-event -k -a

All kernel events are enabled in channel channel0

test4@thinkos:~$ lttng start

Tracing started for session auto-20120827-120832

test4@thinkos:~$ lttng stop

Tracing stopped for session auto-20120827-120832

test4@thinkos:~$
```



> Tracing: view trace

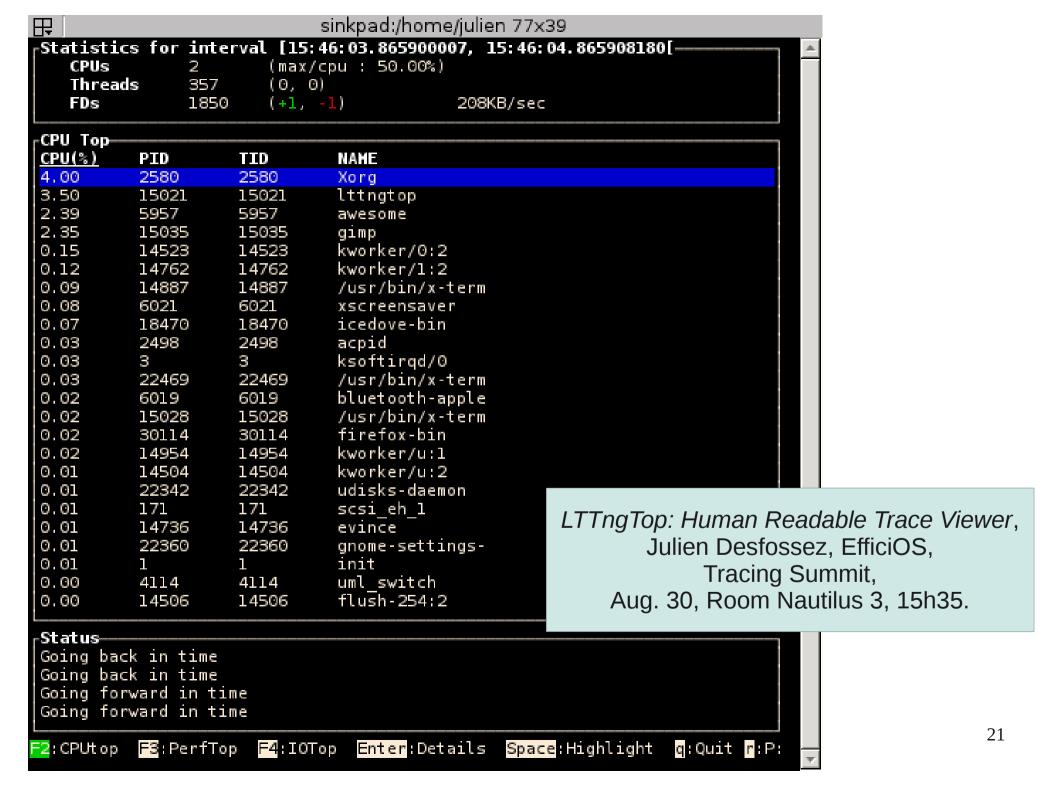
```
File Edit View Search Terminal Help
test4@thinkos:~$ lttng view |head
[12:09:27.319016156] (+?.????????) thinkos exit syscall: { cpu id = 0 }, { ret = 4400 }
[12:09:27.319019724] (+0.000003568) thinkos sys poll: { cpu id = 0 }, { ufds = 0x7F799E791CB0, nfds = 2,
 timeout msecs = -1 }
[12:09:27.319021368] (+0.000001644) thinkos exit syscall: { cpu id = 0 }, { ret = 1 }
[12:09:27.319022644] (+0.000001276) thinkos sys recvmsg: { cpu id = 0 }, { fd = 18, msg = 0x7F799E7909C0
 flags = 0 }
[12:09:27.319027091] (+0.000004447) thinkos exit syscall: { cpu id = 0 }, { ret = 1 }
[12:09:27.319037949] (+0.000010858) thinkos sys geteuid: { cpu id = 0 }, { }
[12:09:27.319038734] (+0.000000785) thinkos exit syscall: { cpu id = 0 }, { ret = 0 }
[12:09:27.319040053] (+0.000001319) thinkos sys pipe: { cpu id = 0 }, { fildes = 0x7F799E790920 }
[12:09:27.319046308] (+0.000006255) thinkos exit syscall: { cpu id = 0 }, { ret = 0 }
[12:09:27.319047858] (+0.000001550) thinkos sys mmap: { cpu id = 0 }, { addr = 0x0, len = 10485760, prot
 = 3, flags = 131362, fd = -1, offset = 0 }
test4@thinkos:~$
```



> LTTng 2.0 high-speed "strace"

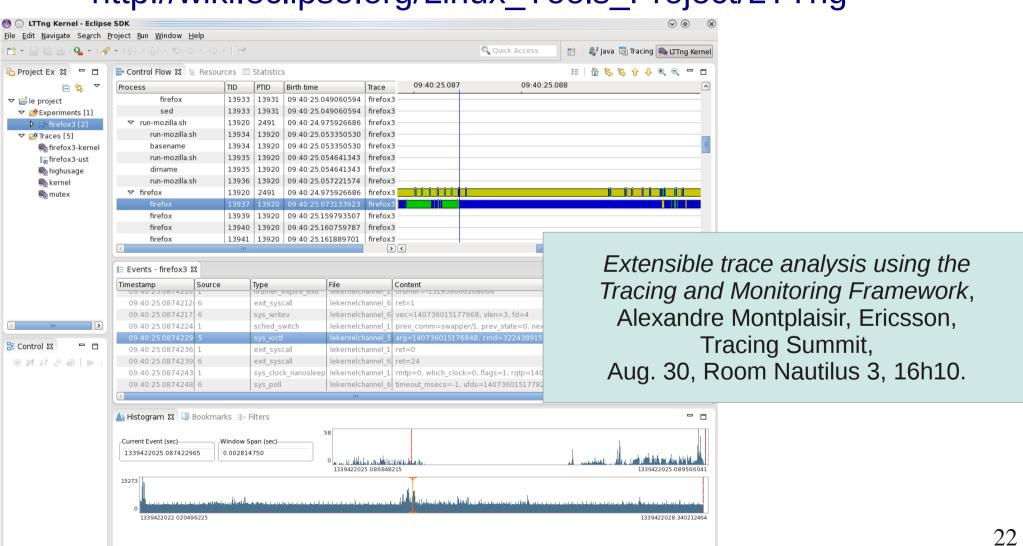
Ittng enable-event --syscall -a

```
compudi@squeeze-amd64: ~
name = sys brk, stream.packet.context = { cpu id = 1 }, event.fields = { brk = 28622848 }
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 28622848 }
name = sys read, stream.packet.context = { cpu id = 1 }, event.fields = { fd = 3, buf = 0x1B48008, count = 9645 }
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 9645 }
name = sys close, stream.packet.context = { cpu id = 1 }, event.fields = { fd = 3 }
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 0 }
name = sys open, stream.packet.context = { cpu id = 1 }, event.fields = { filename = "/root/.bash history", flags = 513, mode
name = exit_syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 3 }
name = sys write, stream.packet.context = \{ cpu id = 1 \}, event.fields = \{ fd = 3, buf = 0x1B48081, count = 9524 \}
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 9524 }
name = sys close, stream.packet.context = { cpu id = 1 }, event.fields = { fd = 3 }
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 0 }
name = sys rt sigprocmask, stream.packet.context = { cpu id = 1 }, event.fields = { how = 0, nset = 0x7FFF28A2A040, oset = 0
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 0 }
name = sys ioctl, stream.packet.context = \{ cpu id = 1 \}, event.fields = \{ fd = 255, cmd = 21520, arg = 140733875134380 \}
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 0 }
name = sys rt sigprocmask, stream.packet.context = { cpu id = 1 }, event.fields = { how = 2, nset = 0x7FFF28A29FC0, oset = 0x7FF728A29FC0, oset = 0x7FF728A29FC0
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 0 }
name = sys setpgid, stream.packet.context = { cpu id = 1 }, event.fields = { pid = 0, pgid = 4235 }
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 0 }
name = sys exit group, stream.packet.context = { cpu id = 1 }, event.fields = { error code = 0 }
name = exit_syscall, stream.packet.context = { cpu_id = 1 }, event.fields = { ret = 1 }
name = sys gettimeofday, stream.packet.context = { cpu id = 1 }, event.fields = { tv = 0x7FFF0E61DC10, tz = 0x0 }
name = exit syscall, stream.packet.context = { cpu id = 1 }, event.fields = { ret = 0 }
```

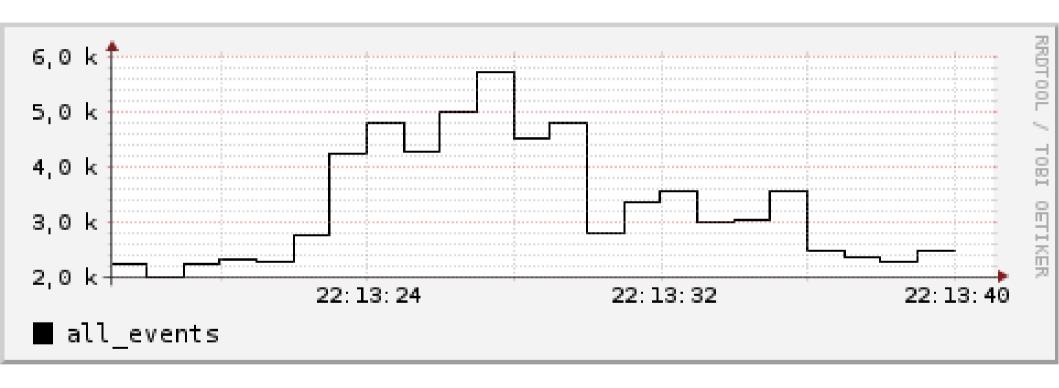


> Eclipse Linux Tools Project: LTTng support

http://wiki.eclipse.org/Linux_Tools_Project/LTTng



> LTTng-Graph



http://git.dorsal.polymtl.ca/~jdesfossez?p=lttng-graph



> Viewer Libraries and Bindings

- Lib Babeltrace: a C/C++ library for reading CTF traces (MIT BSD-style license)
- Python bindings over Lib Babeltrace planned to be merged into Babeltrace 1.1
 - http://git.efficios.com/?p=babeltrace.git bindings/python branch.



> Tracing across kernel and user-space

```
File Edit View Search Terminal Help

test4@thinkos:~$ lttng create

Session auto-20120827-133620 created.

Traces will be written in /home/test4/lttng-traces/auto-20120827-133620

test4@thinkos:~$ lttng enable-event -u -a

All UST events are enabled in channel channel0

test4@thinkos:~$ lttng enable-event -k -a

All kernel events are enabled in channel channel0

test4@thinkos:~$ lttng start

Tracing started for session auto-20120827-133620

test4@thinkos:~$ lttng destroy

Session auto-20120827-133620 destroyed

test4@thinkos:~$
```



> Tracing across kernel and user-space (2)

babeltrace -f trace:domain /home/test4/lttng-traces/auto-20120827-133620 \ | grep "{ cpu_id = 0 }" | less

```
File Edit View Search Terminal Help
[13:36:30.119503690] (+0.0000\4293) kernel exit_syscall: { cpu_id = 0 }, { ret = 139998849839104 }
[13:36:30.119513141] (+0.00000027) kernel sys rt sigprocmask: { cpu id = 0 }, { how = 2, nset = 0x
7FFF504E8640, oset = 0x0, sigsetsize - 0 ;
[13:36:30.119513896] (+0.000000755) kernal exit syscall: { cpu id = 0 }, { ret = 0 }
[13:36:30.119523258] (+0.000001754) ust ust_tests_demo:starting: { cpu_id = 0 }, { value = 123 }
[13:36:30.119531493] (+0.00000100) ust ust_tests_demo2:loop: { cpu_id = 0 }, { intfield = 0, intfi
eld2 = 0x0, longfield = 0, netintfield = 0, netintfieldhex = 0x0, arrfield1 = [0] = 1, [1] = 2,
2] = 3 ], arrfield2 = "test", seqfield1 length = 4, seqfield1 = [ [0] = 116, [1] = 101, [2] = 115,
[3] = 116], segfield2 length = 4, segfield2 = "test", stringfield = "test", floatfield = 2222, d
oublefield = 2 }
[13:36:30.119535331] (+0.000003838) ust ust tests demo2:loop: { cpu id = 0 }, { intfield = 1, intfi
eld2 = 0x1, longfield = 1, netintfield = 1, netintfieldhex = 0x1, arrfield1 = [ [0] = 1, [1] = 2, [
2] = 3 ], arrfield2 = "test", seqfield1 length = 4, seqfield1 = [ [0] = 116, [1] = 101, [2] = 115,
[3] = 116], seqfield2 length = 4, seqfield2 = "test", stringfield = "test", floatfield = 2222, d
oublefield = 2 }
[13:36:30.119538004] (+0.000002673) ust ust tests demo2:loop: { cpu id = 0 }, { intfield = 2, intfi
eld2 = 0x2, longfield = 2, netintfield = 2, netintfieldhex = 0x2, arrfield1 = [[0] = 1, [1] = 2,
2] = 3 ], arrfield2 = "test", segfield1 length = 4, segfield1 = [ [0] = 116, [1] = 101, [2] = 115,
 [3] = 116], seqfield2 length = 4, seqfield2 = "test", stringfield = "test", floatfield = 2222, d
oublefield = 2 }
[13:36:30.119540487] (+0.000001751) ust ust tests demo2:loop: { cpu id = 0 }, { intfield = 3, intfi
```



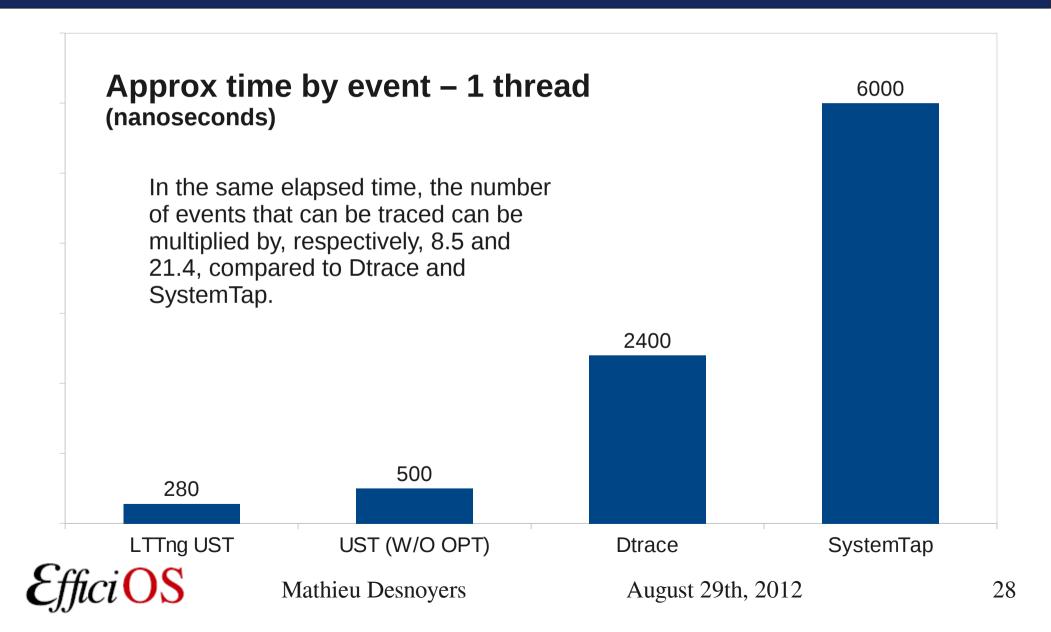
> Tracing across nodes

babeltrace --clock-force-correlate trace1 trace2 |grep -e sys_read -e sys_write |less

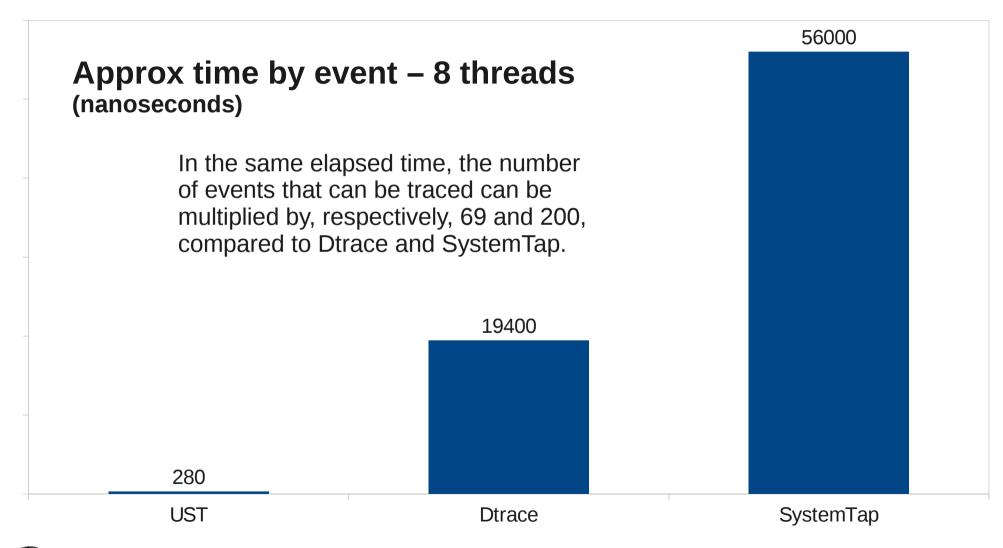
```
[21:43:16.653417520] (+0.000001302) thinkos by read: { cpu id = 0 }, { fd = 17, buf = 0x7FFFEE5456}
60. count = 16 
[21:43:16.653419555] (+0.000001022) thinkos sys read: { cpu id = 0 }, { fd = 8, buf = 0x7FFFEE54546
0, count = 512 }
[21:43:16.654177499] (+0.000001982 squeeze-i386 sys writev: { cpu id = 1 }, { fd = 3, vec = 0xBFC9}
1B08. vlen = 3 
[21:43:16.654284531] (+0.000007210) squeeze-i386 sys read: { cpu id = 1 }, { fd = 23, buf = 0x91639}
78, count = 4096 }
[21:43:16.654300013] (+0.000007367) squeeze-i386 sys read: { cpu_id = 1 }, { fd = 23, buf = 0x91639}
78. count = 4096 
[21:43:16.654304299] (+0.000002903) squeeze-i386 sys writev: { cpu id = 1 }, { fd = 23, vec = 0xBFC
CBF08. vlen = 1 }
[21:43:16.654351523] (+0.000000309) squeeze-i386 sys read: { cpu id = 1 }, { fd = 3, buf = 0x81477C
0, count = 4096 }
[21:43:16.654362090] (+0.000005374) squeeze-i386 sys read: { cpu id = 1 }, { fd = 3, buf = 0x81477C
0, count = 4096 }
[21:43:16.654370802] (+0.000005520) thinkos sys read: { cpu id = 0 }, { fd = 17, buf = 0x7FFFEE5456
60, count = 16 }
[21:43:16.654374957] (+0.000002107) thinkos sys read: { cpu id = 0 }, { fd = 7, buf = 0x7FFFEE54554
0, count = 128 }
[21:43:16.654375457] (+0.000000500) squeeze-i386 sys read: { cpu id = 1 }, { fd = 3, buf = 0x81477C
```



> Userspace Tracing Benchmark

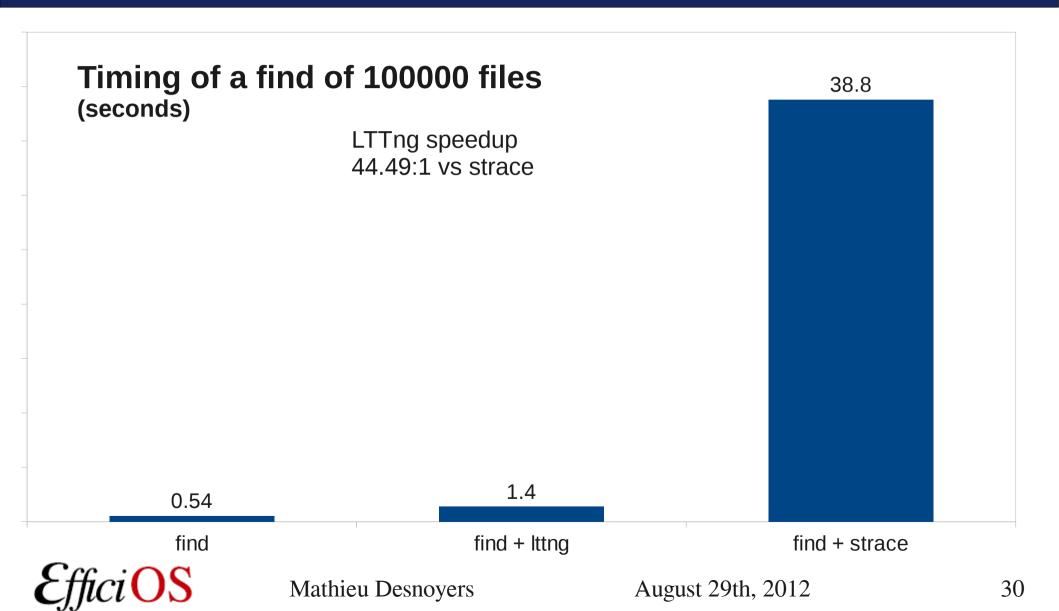


> Userspace Tracing Benchmark





> Strace vs LTTng Tracing

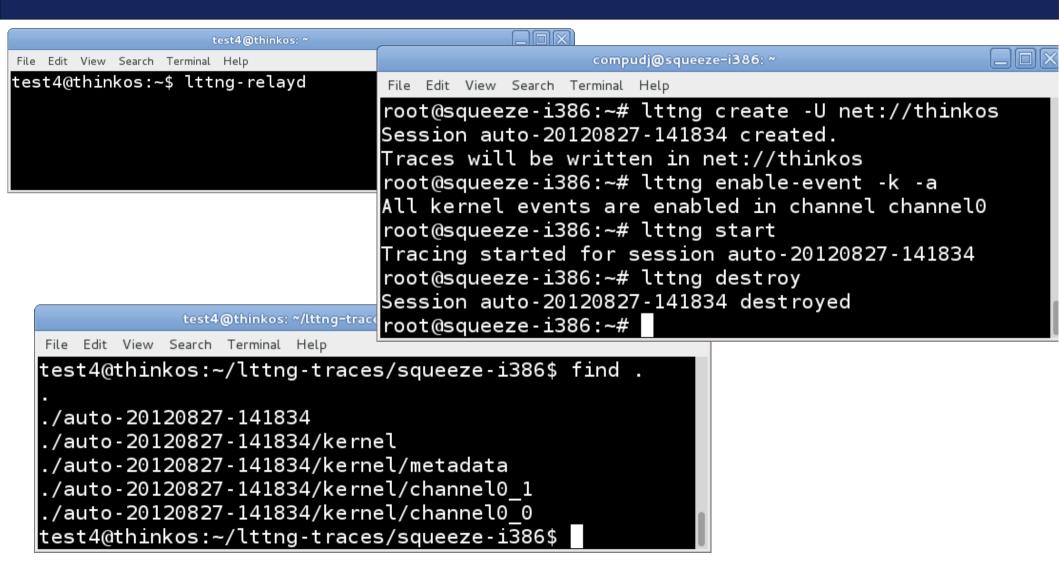


> New Features (LTTng 2.1)

- Network Streaming over TCP
- LTTng-UST filtering before data collection
- Ittng-sessiond(8) health monitoring API.



> Network Streaming over TCP





> LTTng-UST Filtering

```
File Edit View Search Terminal Help

test4@thinkos:~$ lttng create

Session auto-20120827-142450 created.

Traces will be written in /home/test4/lttng-traces/auto-20120827-142450

test4@thinkos:~$ lttng enable-event -u -a --filter "(intfield > 42 && intfield <= 44)

) || longfield == 1"

All UST events are enabled in channel channel0

test4@thinkos:~$ lttng start

Tracing started for session auto-20120827-142450

test4@thinkos:~$ lttng destroy

Session auto-20120827-142450 destroyed

test4@thinkos:~$
```



> LTTng-UST Filtering (2)

```
test4@thinkos: ~/lttng-traces
File Edit View Search Terminal Help
test4@thinkos:~/lttng-traces$ babeltrace auto-20120827-142450/
[14:25:09.348326990] (+?.????????) thinkos:lt-hello:22200 ust_tests_hello:tp
test: { cpu_id = 0 }, { intfield = 1, intfield2 = 0x(, longfield = 1, netintf
ield = 1, netintfieldhex = 0x1, arrfield1 = [0] = 1, <math>[1] - 2, [2] = 3, [2] = 3
rfield2 = "test", seqfield1 length = 4, seqfield1 = [ [0] = 116, [1] = 101,
[2] = 115, [3] = 116 ], segfield2 length = 4, segfield2 = "test", stringfiel
d = "test", floatfield = 2222, doublefield = 2, boolfield = 1 }
[14:25:09.348343662] (+0.000010072) thinkos:lt-hello:22206 ust_tests_hello:tp
ntfield = 4\overline{3}, netintfieldhex = 0\times2B, arrfield1 = [0] = 1, [1] = 2, [2] = 3
], arrfield2 = "test", _seqfield1_length = 4, seqfield1 = [ [0] = 116, [1] =
101, [2] = 115, [3] = \overline{116} ], _seqfield2_length = 4, seqfield2 = "test", strin
gfield = "test", floatfield = 2222, doublefield = 2, boolfield = 1 }
[14:25:09.348347126] (+0.000003464) thinkos:lt-hello:22206 ust_tests_hello:tp
ntfield = 4\overline{4}, netintfierdiex 0\times2C, arrfield1 = [0] = 1, [1] = 2, [2] = 3
], arrfield2 = "test", _seqfield1_length = 4, seqfield1 = [ [0] = 116, [1] =
101, [2] = 115, [3] = 116 ], _seqfield2_length = 4, seqfield2 = "test", strin
gfield = "test", floatfield = 2222, doublefield = 2, boolfield = 1 }
test4@thinkos:~/lttng-traces$
```



> Distributions / Integration

- LTTng 0.x
 - Wind River Linux, Montavista, STlinux, Linaro, Yocto, Mentor Embedded Linux, ELinOS.
- LTTng 2.0
 - Ubuntu 12.04 LTS
 - Debian
 - SuSE Enterprise RT Linux
 - Linux Foundation LTSI
 - Fedora: process ongoing



> Collaborations

- Interfacing: GDB tracepoints can interoperate with LTTng UST tracepoints, the Eclipse Tracing Monitoring Framework supports LTTng CTF traces, Perf PMU counters are used by LTTng, The Multi Core Association is defining a Common Trace Format (CTF), for which LTTng 2.0 is a reference implementation.
- Collaborators: Autodesk, CAE, C2 microsystems, Defence R&D Canada, Ecole Polytechnique de Montréal, EfficiOS, Ericsson, Fujitsu, Freescale, Google, Harvard University, IBM, Linux Foundation CE, Mentor, Multicore Association, Nokia, Opal-RT, Portland State University, RedHat, Revolution Linux, Sony, ST Microelectronics, Siemens, SUSE, WindRiver, CRIAQ, NSERC, Prompt, etc.



Tomorrow 8h30 in Nautilus 3

The Tracing Summit 2012 will be held in San Diego, on August 30th, 2012, as part of the Linux Plumbers Conference 2012 .

- 8h30 Can mainstream tracing meet embedded needs?, Frank Rowand, Sony
- 8h55 What We Want in Our Toolkit: Thoughts From a Mission Critical Low Latency Environment, Vinod Kutty, CME

Group

- 9h20 Troubleshooting complex problems with built-in diagnostic, Dominique Toupin, Ericsson
- 9h45 The Linux Perf Tools: Overview and Current Developments, Arnaldo Carvahlo de Melo, Red Hat
- 10h10 10h25 break
- 10h25 Tracing the Guest OS from Host via Shared Memory, Masami Hiramatsu, Hitachi
- 10h50 Shrinking core dump on the fly, Thomas Gleixner, Linutronix
- 11h15 Tracing Well With Others: Integration of GDB Tracepoints Into Trace Tools, Stan Shebs, Mentor Graphics
- 11h40 Ftrace and Multiple buffers, Steven Rostedt, Red Hat
- 12h05 13h30 lunch
- 13h30 The Road ahead of Uprobes: Plans and features in pipeline, Srikar Dronamraju, IBM
- 13h45 Systemtap and new connections: dyninst, pcp, uprobes, David Smith and Josh Stone, Redhat
- 14h20 LTTng and Nexus Trace for Freescale QorlQ Devices, Ed Martinez, Freescale
- 14h35 14h45 break
- 14h45 Interoperability Between Tracing Tools with the Common Trace Format (CTF), Mathieu Desnoyers, EfficiOS
- 15h10 Making linsched useful, Dhaval Giani, University of Toronto
- 15h35 LTTngTop: Human Readable Trace Viewer, Julien Desfossez, EfficiOS
- 16h00 16h10 break
- 16h10 Extensible trace analysis using the Tracing and Monitoring Framework, Alexandre Montplaisir, Ericsson
- 16h45 18h00 Open Discussion, Where do we go from here?

> Questions?

LTTng 2.0 available at http://lttng.org





- http://www.efficios.com
- LTTng Information
 - http://lttng.org
 - Ittng-dev@lists.Ittng.org

