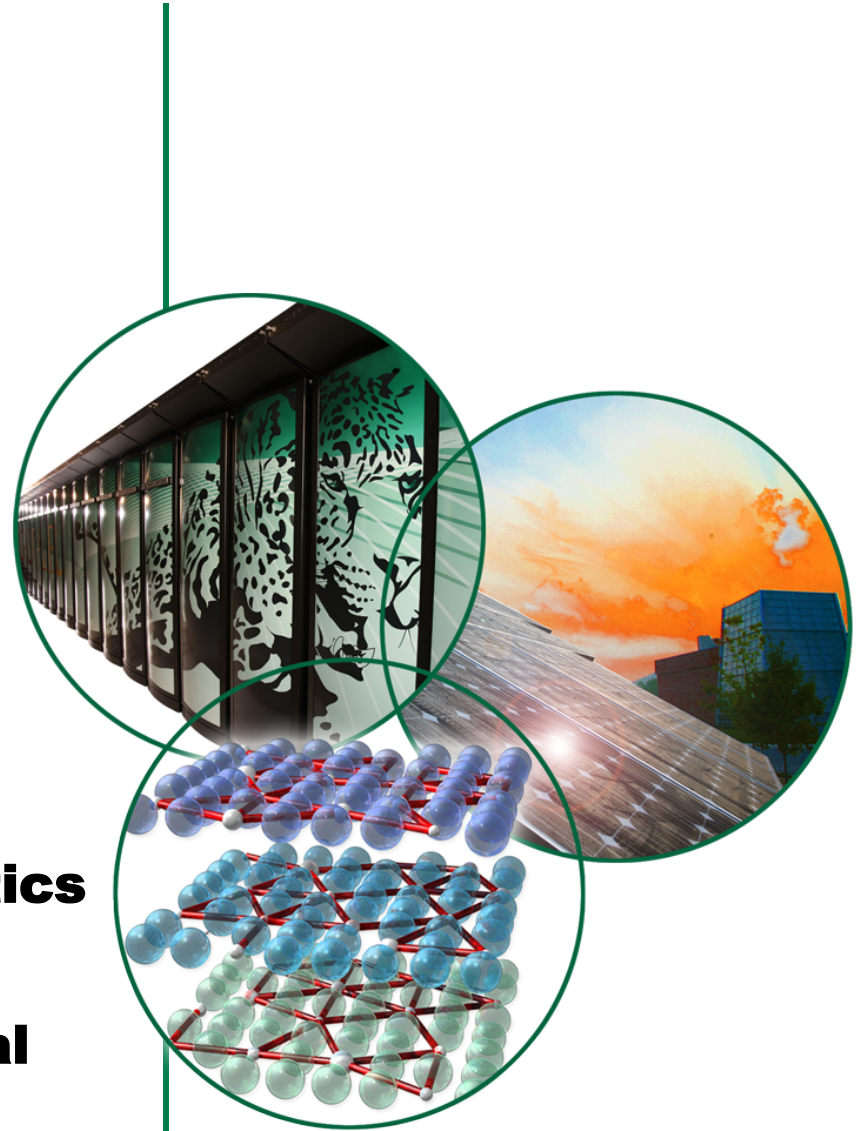


RAVEN: RAS data Analysis through Visually Enhanced Navigation

**Byung-Hoon Park, Junseong Heo,
Guruprsad Kora, and Al Geist**

**Computer Science and Mathematics
Division, ORNL**

**National Center for Computational
Science (NICS)**



**U.S. DEPARTMENT OF
ENERGY**

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

Motivation

- RAS Logs are often the only resource for obtaining clues for system failure (or system abnormality).
- Ever-increasing volume of Logs is beyond the capacity of manual analysis by humans.
- Console logs are mixture of free outputs from many open source packages.
- Rendering unstructured RAS logs in intuitive and human friendly format is greatly desired.



```
[2009-12-21 04:0...
[2009-12-21 04:0...
[2009-12-21 04:0...
[2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:1303:lnet_send()) No route to 12345-10.36.227.200@o2ib (all routers down)
[2009-12-21 04:06:25][c18-3c2s4n1]beer: cpu_id 9: unable to initiate communication with nid 5680, cpu 0 after 1200 seconds
[2009-12-21 04:06:30][c18-3c2s4n1]LustreError: 761:0:(ptlInD_tx.c:469:kptlInD_tx_callback()) Skipped 14 previous similar messages
[2009-12-21 04:06:30][c18-3c2s4n1]beer: tx=ffff8103fc985a40 fail=PTL_NAL_FAILED(4) unlinked=1
[2009-12-21 04:06:30][c18-3c2s4n1]LustreError: 25768:0:(ptlInD_ptltrace.c:38:kptlInD_ptltrace_to_file()) dumping ptltrace to /tmp/lnet-ptltrace.1261386390.25249
[2009-12-21 04:06:30][c18-3c2s4n1]beer: cpu_id 9: unable to initiate communication with nid 6371, cpu 0 after 1200 seconds
```

Properties of RAS Logs

Irregular

- Event occurrences are irregular, often come in bulk.

Redundancy

- The same root cause triggers different observations of many components.

Implicit Correlation

- Some event types are inherently correlated.
- Can be used to confirm the situation.

Implicit Context

- Many Messages embed pair wise relations.
- Messages are clustered based on the user application.

Better Use of RAS Data

[2009-12-21 04:01:31][c16-0c2s0n2]cpu 0 apic_timer_irqs=0x3610de5
[2009-12-21 04:01:31][c16-0c2s0n3]cpu 0 apic_timer_irqs=0x3610ed6
[2009-12-21 04:01:31][c16-0c2s5n3]cpu 0 apic_timer_irqs=0x3610e8a
[2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:1303:inet_send()) No route to 12345-10.36.227.200@o2ib (all routers down)
[2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:2285:LNetPut()) Skipped 2693 previous similar messages
[2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:1303:inet_send()) Error sending PUT to 12345-10.36.227.200@o2ib: -113
[2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:2285:LNetPut()) Skipped 2693 previous similar messages
[2009-12-21 04:06:30][c18-3c2s4n1]beer: cpu_id 9: unable to initiate communication with nid 5680, cpu 0 after 1200 seconds
[2009-12-21 04:06:30][c18-3c2s4n1]LustreError: 761:0:(ptlnd_tx.c:469:kptlnd_tx_callback()) Portals error to 12345-5680@ptl1:
PTL_EVENT_SEND_END(9) tx=ffff8103fc985a40 fail=PTL_NAL_FAILED(4) unlinked-1
[2009-12-21 04:06:30][c18-3c2s4n1]LustreError: 761:0:(ptlnd_tx.c:469:kptlnd_tx_callback()) Skipped 14 previous similar messages
[2009-12-21 04:06:30][c18-3c2s4n1]beer: cpu_id 9: unable to initiate communication with nid 6004, cpu 0 after 1200 seconds
[2009-12-21 04:06:30][c18-3c2s4n1]Lustre: 25768:0:(ptlnd_ptltrace.c:120) Writing ptltrace to /tmp/inet-ptltrace.
1261386390.25249
[2009-12-21 04:06:30][c18-3c2s4n1]beer: cpu_id 9: unable to initiate communication with nid 6004, cpu 0 after 1200 seconds

Multi-Resolution View

Views of Different Contexts

Event Synopsis

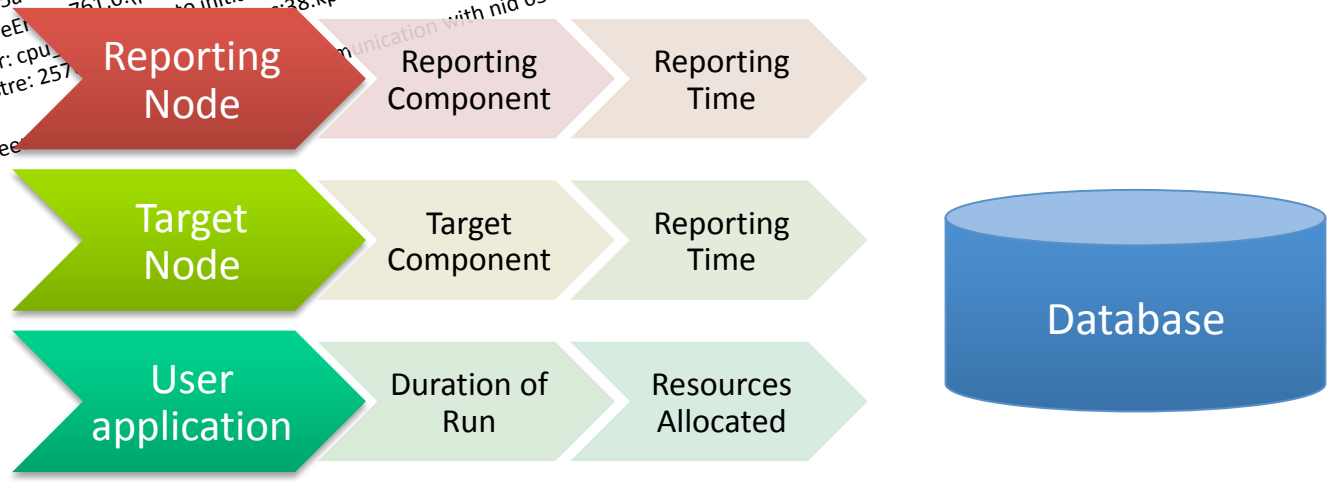


Context Driven Temporal and Spatial Analysis

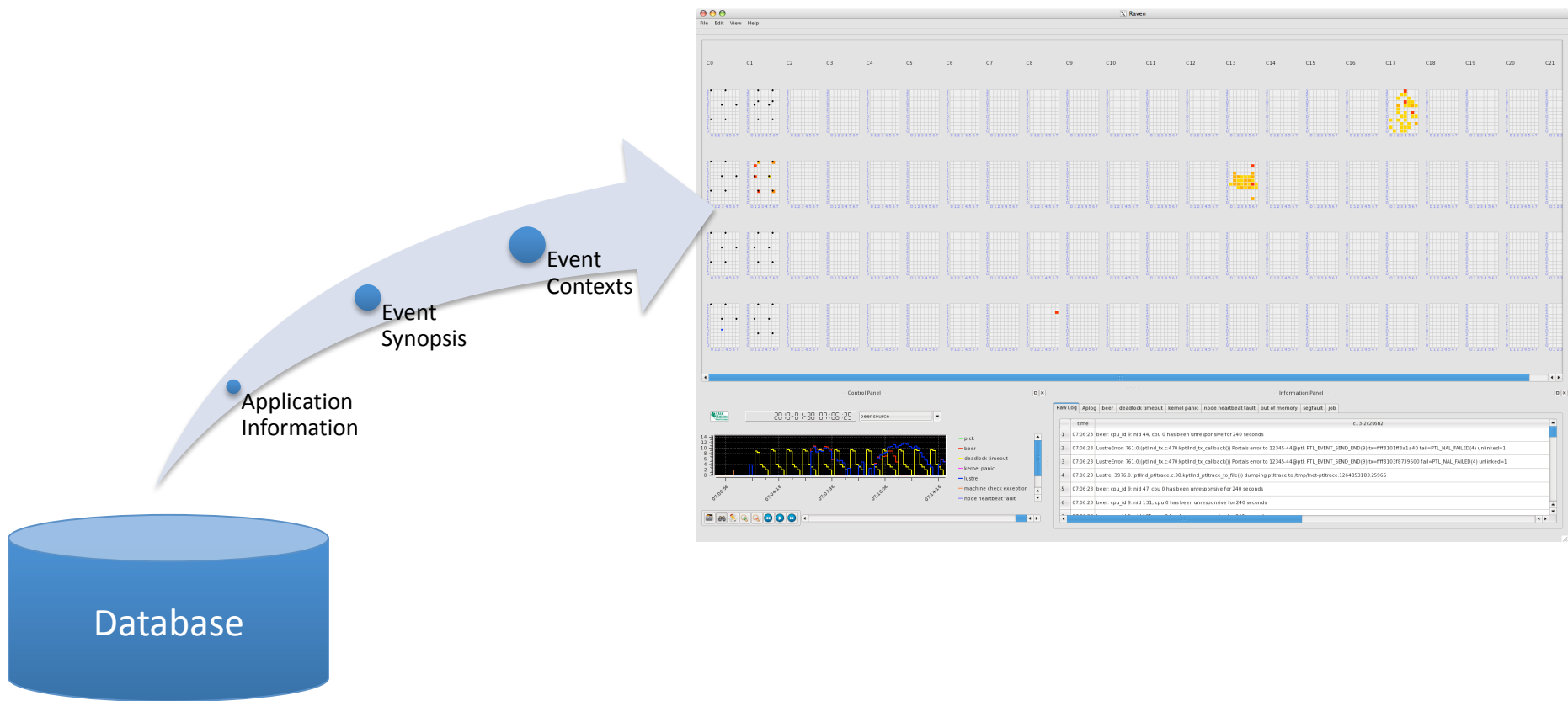


Abstraction of RAS Data

[2009-12-21 04:01:31][c16-0c2s0n2]cpu 0 apic_timer_irqs=0x3610de5
 [2009-12-21 04:01:31][c16-0c2s0n3]cpu 0 apic_timer_irqs=0x3610e8a
 [2009-12-21 04:01:31][c16-0c2s5n3]cpu 0 apic_timer_irqs=0x3610e8a
 [2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:1303:lnet_send()) No route to 12345-10.36.227.200@o2ib (all routers down)
 [2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:1303:lnet_send()) Skipped 2693 previous similar messages
 [2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:2285:LNetPut()) Error sending PUT to 12345-10.36.227.200@o2ib: -113
 [2009-12-21 04:06:25][c18-3c2s4n1]LustreError: 25496:0:(lib-move.c:2285:LNetPut()) Skipped 2693 previous similar messages
 [2009-12-21 04:06:30][c18-3c2s4n1]beer: cpu_id 9: unable to initiate communication with nid 5680, cpu 0 after 1200 seconds
 [2009-12-21 04:06:30][c18-3c2s4n1]LustreError: 761:0:(ptlnd_tx.c:469:kptlnd_tx_callback()) Portals error to 12345-5680@ptl1:
 PTL_EVENT_SEND_END(9) tx=ffff8103fc985a40 fail=PTL_NAL_FAILED(4) unlinked=1
 [2009-12-21 04:06:30][c18-3c2s4n1]LustreError: 761:0:(ptlnd_tx.c:469:kptlnd_tx_callback()) Skipped 14 previous similar messages
 [2009-12-21 04:06:30][c18-3c2s4n1]beer: cpu_id 9: unable to initiate communication with nid 6004, cpu 0 after 1200 seconds
 [2009-12-21 04:06:30][c18-3c2s4n1]Lustre: 2571:0:(ptlnd_tx.c:469:kptlnd_tx_callback()) dumping pttltrace to /tmp/lnet-ptttrace.
 1261386390.25249
 [2009-12-21 04:06:30][c18-3c2s4n1]beer: cpu_id 9: unable to initiate communication with nid 6371, cpu 0 after 1200 seconds



RAVEN: RAS data Analysis through Visually Enhanced Navigation



RAVEN At a Glance



Raven is a system independent tool

- Database Driven (MySQL)
- Configurable through XML files.



RAVEN is GUI tool.

- Frontend Interface is written with Qt4.6
- Graph Plotting is written with Qwt5.2



RAVEN has been ported to

- Cray XT5 (Jaguar, Kraken)
- IBM BG/P at ANL (Underway)

RAVEN RAS Events for Cray XT5

Console

- Lustre, BEER, Segfault, OOM, MCE, Kernel Panic, etc.

Netwatch

- Link Inactive, Deadlock Timeout, uPacket Squash (off), etc.

Consumer

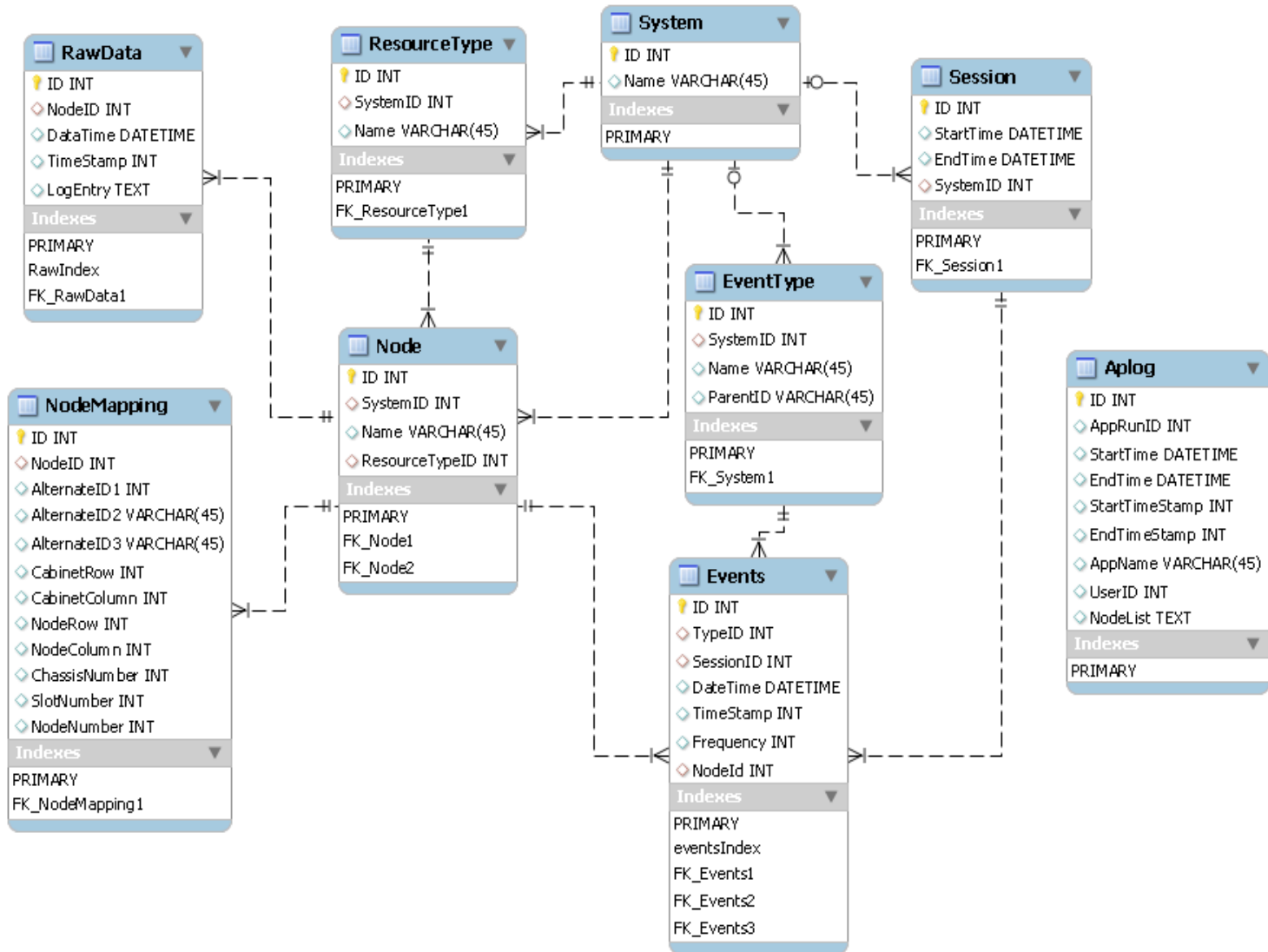
- Node Heartbeat Failure, Seastar heartbeat failure, Node voltage fault, Verty health check fault, etc.

Apsched

- Application Name, User ID, Duration of Run, Allocated Nodes, ResID, etc.

RAVEN Backend Database Design

RAVEN RAS Log Backend Design



RAVEN Frontend

The screenshot displays the RAVEN Frontend interface, which is used for monitoring and controlling a system. The interface is divided into several key sections:

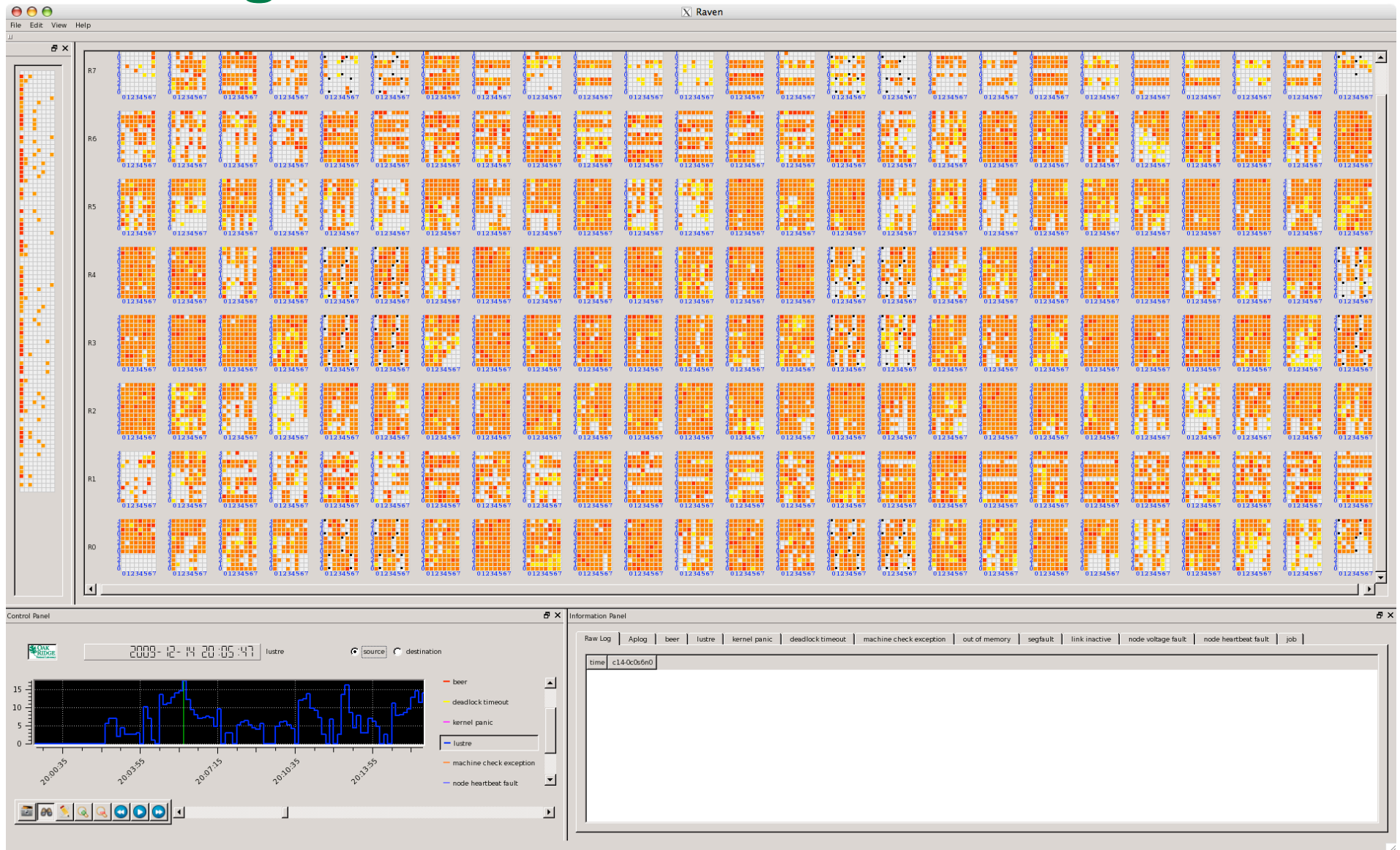
- System Layout Map:** A grid of 22 columns (C0 to C21) and 4 rows of small plots. Each plot shows a grid with various colored dots and lines, representing the state of different components in the system. A blue callout box labeled "System Layout Map" points to this area.
- Control Panel:** A horizontal bar at the bottom left containing a "Control Panel" label, a "Control Panel" button, and a "Control Panel" dropdown menu. Below this is a "Control Panel" window showing a time-series plot of various metrics (pick, beer, deadlock timeout, kernel panic, lustre, machine check exception, node heartbeat fault) over time. A blue callout box labeled "Control Panel" points to this area.
- Information Panel:** A horizontal bar at the bottom right containing an "Information Panel" label and an "Information Panel" button. Below this is an "Information Panel" window showing a log of system events, including error messages and status updates. A blue callout box labeled "Information Panel" points to this area.

The main window also features a menu bar (File, Edit, View, Help) and a title bar (Raven). The System Layout Map plots are arranged in a grid, with columns C0-C8 and C13-C21 visible. The Control Panel window shows a time-series plot with a legend and a play button. The Information Panel window shows a log of system events with a scroll bar.

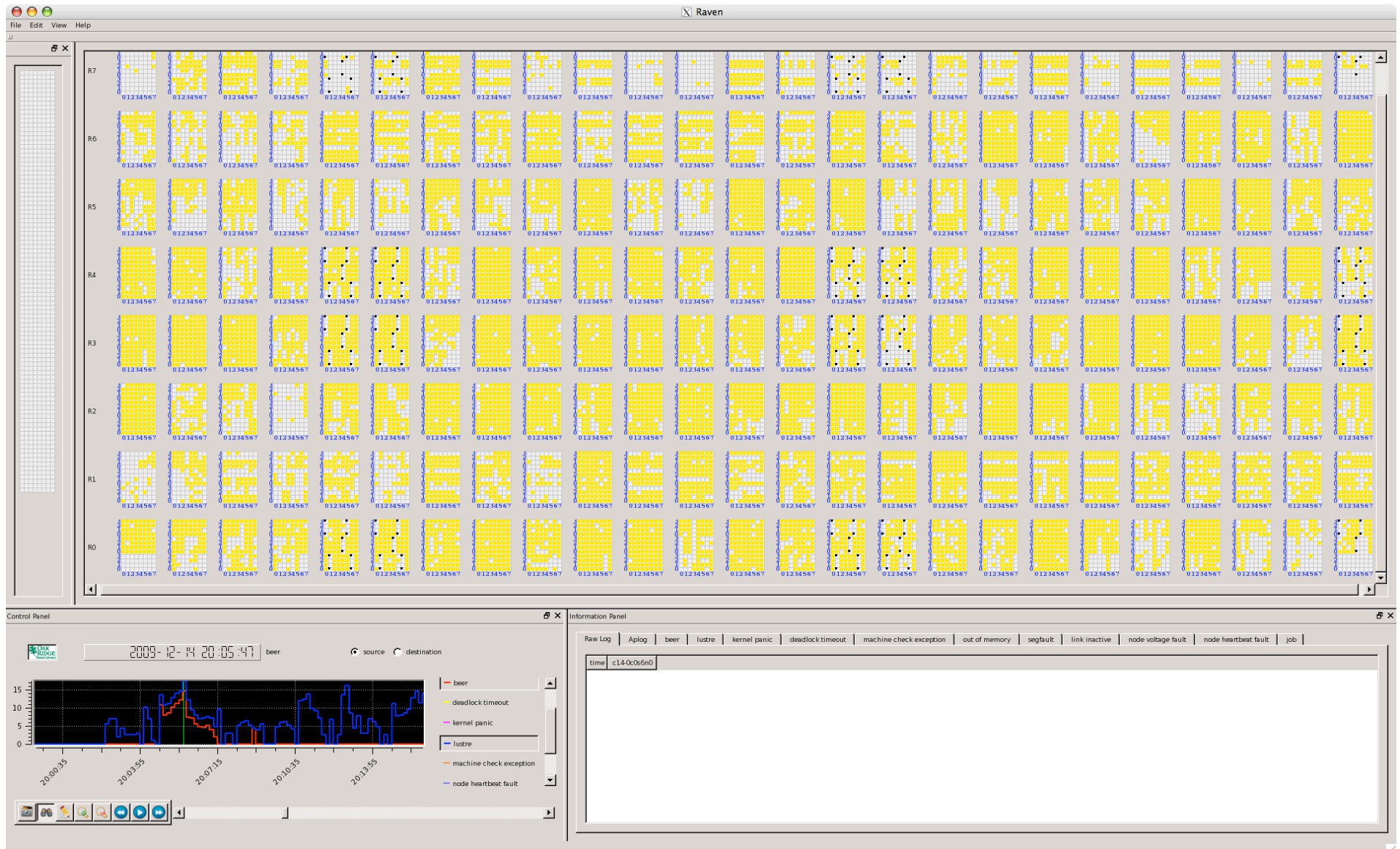
Application Displacement

The screenshot displays the Raven application monitoring interface. The main area is a grid of heatmaps for 22 nodes (C0-C21) across 4 rows. Each heatmap shows application displacement over time, with axes labeled 0-1234567. The bottom section features a Control Panel with a timeline graph showing various events (pick, beer, deadlock timeout, kernel panic, lustre, machine check exception, node heartbeat fault) from 07:00:56 to 07:14:16. The Information Panel shows a log of events for node c13-2c256n2, including messages like 'beer: cpu_id 9: nid 44, cpu 0 has been unresponsive for 240 seconds' and 'LustreError: 761:0:(ptlnd_tx.c:470:kptlnd_tx_callback()) Portals error to 12345-44@ptl: PTL_EVENT_SEND_END(9) tx=fff8101f3a1a40 fail=PTL_NAL_FAILED(4) unlinked=1'.

Case Study: A Flood of Lustre Messages



And a Flood of BEER Messages Together



Pointing to a Single Router Node

The screenshot displays the Raven network monitoring application. The main window shows a grid of 180 small network graphs arranged in 6 rows (R0-R5) and 30 columns. A red circle highlights a specific graph in the R0 row, with a red arrow pointing to it. Below the grid is a 'Control Panel' with a line graph showing various metrics over time, and an 'Information Panel' displaying a log of system events.

Control Panel

2009-12-14 20:05:47 beer source destination

15
10
5
0

20:00:55 20:03:55 20:07:55 20:10:55 20:13:55

- beer
- deadlock timeout
- kernel panic
- lustre
- machine check exception
- node heartbeat fault

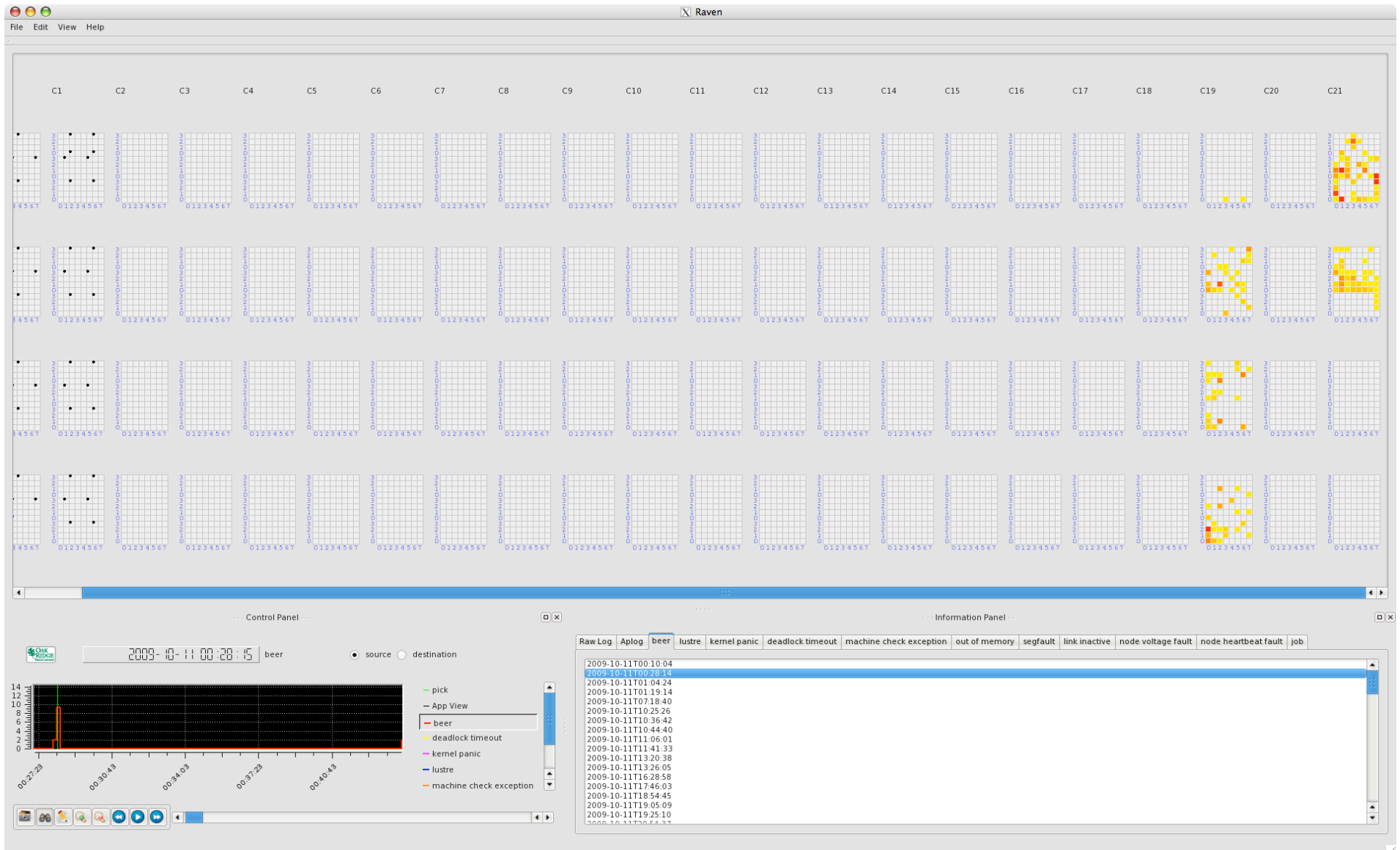
Information Panel

Raw Log | Alog | beer | lustre | kernel panic | deadlock timeout | machine check exception | out of memory | segfault | link inactive | node voltage fault | node heartbeat fault | job

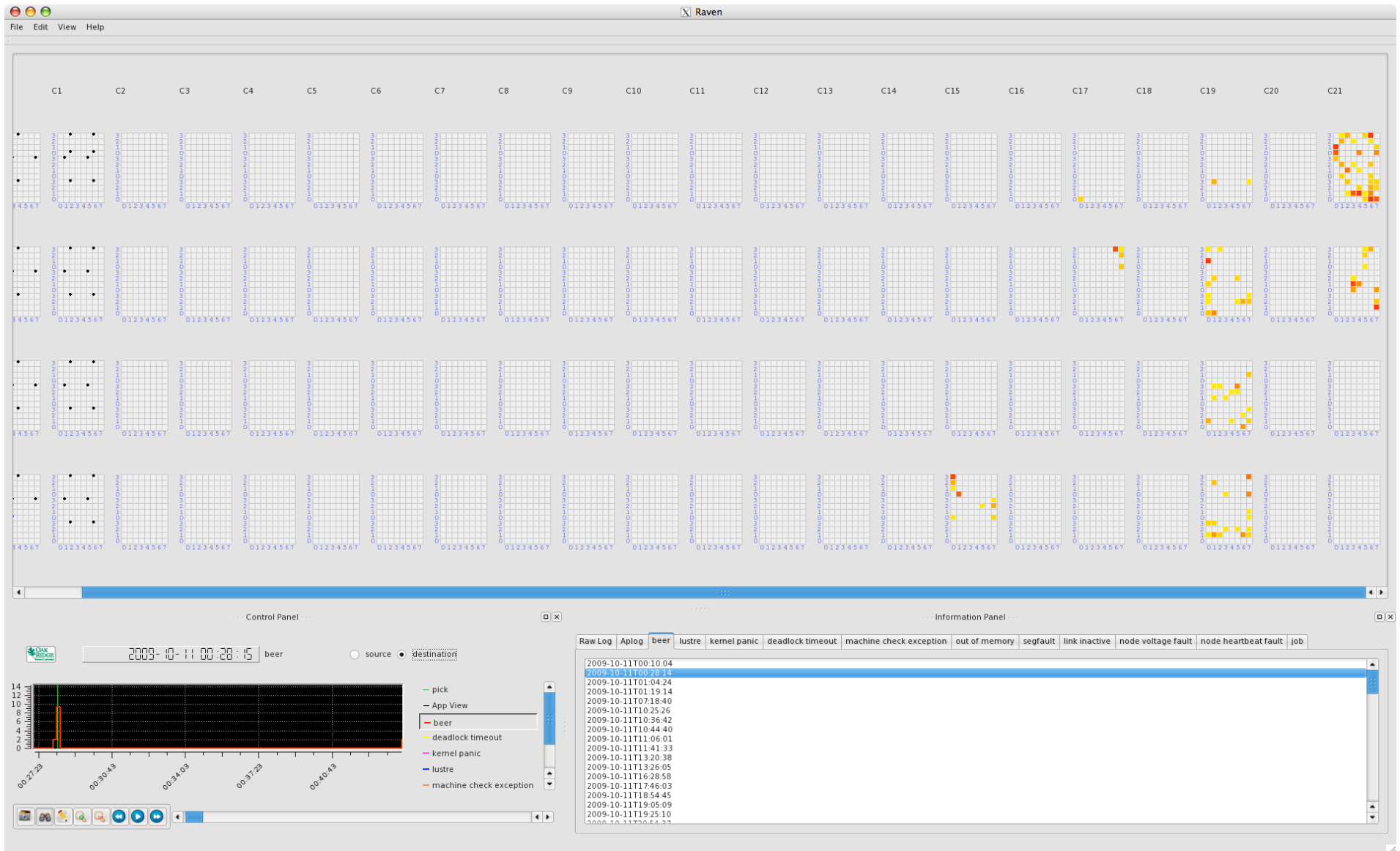
c9-0c05n3

time	message
2005.45	beer: cpu_id 9: nid 9752, cpu 0 has been unresponsive for 240 seconds
2005.45	LustreError: 761.0:(ptlInd_tx.c:469:kptlInd_tx_callback()) Portals error to 12345-9752@pt1: PTL_EVENT_SEND_END(9) tx=ffff8103f09b7900 fail=PTL_NAL_FAILED(4) unlinked=1
2005.45	LustreError: 25234.0:(events.c:55:request_out_callback()) @@ type 4, status -5 req@ffff810038c03400 x549377f0 o400->widowO-OST0076_UIID@10.36.227.23@o2ib:28/4 lens 128/256 e 0 to 1 dl 1260839503 ref 2 fi Rpc
2005.45	Lustre: 26016.0:(ptlInd_ptltrace.c:38:kptlInd_ptltrace_to_file()) dumping ptltrace to /tmp/ine-ptltrace.1260839144.25237
2005.45	Lustre: Request x549377 sent from widowO-OST0076-osc:ffff8103f050bbc00 to NID 10.36.227.23@o2ib 241s ago has timed out (limit: 600s).
2005.45	Lustre: widowO-OST0076-osc:ffff8103f050bbc00: Connection to service widowO-OST0076 via nid 10.36.227.23@o2ib was lost; in progress operations using this service will wait for recovery to complete.
2005.45	Lustre: widowO-OST0076-osc:ffff8103f050bbc00: Connection to service widowO-OST0076 via nid 10.36.227.23@o2ib was lost; in progress operations using this service will wait for recovery to complete.

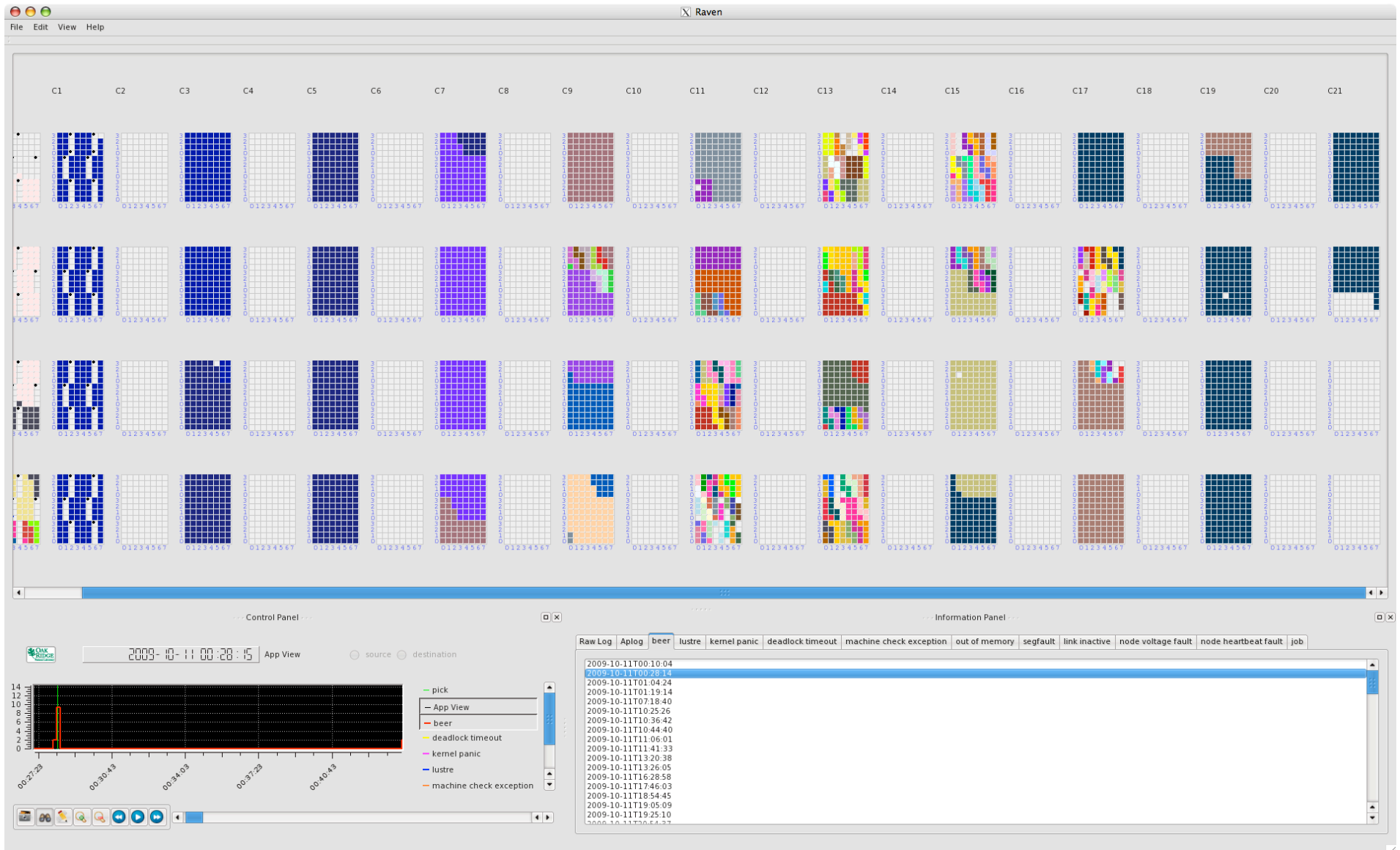
Case Study: BEER Messages Between Nodes of the Same Application



BEER Messages Between Nodes of the Same Application



BEER Messages Between Nodes of the Same Application



Next Step

- **Provide live feed of RAS Events into RAVEN**
- **Incorporate a richer set of RAS event types**
- **Embed a intelligent module that sifts abnormal signatures from live feed of logs**
 - **Cluster of nodes showing abnormal behavior.**
 - **Time interval showing abnormal behavior.**
 - **Job showing abnormal behavior.**
 - **Etc.**
- **A Web-based RAVEN is under development.**

Acknowledgement

- Raghul Gunasekaran (ORNL)
- David Dillow (ORNL)
- Galen Shipman (ORNL)
- Don Maxwell (NCCS, ORNL)
- Jeff Beckleheimer (Cray Inc.)
- Rick Mohr (NICS, UTK)



OAK RIDGE NATIONAL LABORATORY

Managed by UT-Battelle for the Department of Energy