Scalability Issues In Trace Based Analysis Tools

Jesus Labarta, Judit Gimenez
jesus@cepba.upc.edu / jesus@bsc.es
Traceland …

… aiming at detailed analyses …

… and flexibility in the tools.
Importance of details

- Variance is important
  - Along time
  - Across processors

- Highly non linear systems
  - Microscopic effects are important
    - May have large macroscopic impact
IBM Petascale Tools Strategy Workshop, TJWatson. May 3-4th

CEPBA-Tools

Data display tools

TraceDriver

MPIDtrace

Dimemas

Java, WAS GT4

AIXtrace

aixtrace2prv

LTTtrace

LTT2prv

Nanos Compiler

GPFS2prv

GPFStrace

Paraver

JIS

Paramedir

Aaa miss ratio 0.8
Bbb IPC 0.5
Ccc Efficiency 0.4
Ddd bandwidth 520

aixtrace2prv

LTT2prv

GPFS2prv
Index

- Scalability
- Dimensions
  - Acquisition
  - Presentation
  - Processing
- CEPBA-tools evolution
- Conclusions

Work partially supported by IBM, EC (HPC_Europa), MEC
The typical comment
- … it is not scalable!

Our questions
- What is scalability?
- How scalable is an analyst?

How far can a trace based approach go?
Scalability

What is scalability?

- Ability to handle large processors, time, event frequencies, …
- Ability to detect and focus on the relevant information at the appropriate level of detail
- Ability to perform a deep analysis of large applications without looking at the source code
- Applicable in many areas

Key issue

- size vs. dynamic range !!!
Tools: data/computation tradeoffs

Acquisition  Processing  Presentation

IBM Petascale Tools Strategy Workshop, TJWatson. May 3-4th
Scalability: Dimensions

- Acquisition
- Processing
- Presentation
Scalability of tracing

The issue:
- sufficient information / sufficiently detailed
- Usable by presentation tool

Techniques
- User specified on/off
  - Static in source code
    - Time
    - Space
      - `ompitrace_set_tracing_tasks(task_ini, task_end)`
  - Dynamic at run time
    - Environment variables
      - `setenv MPTRACE_FROM_TO routine_on, instance, routine_off, instance`
Scalability of tracing

- Techniques (cont.)
  - Automatic on/off
    - Periodicity detection
    - Application structure identification
Scalability of tracing

- **Techniques (cont.)**
  - Limit file size
    - Stop tracing when reached
    - Circular buffer. Dump on request (signal)
      - Support analysis of large runs.
      - Issue: correlation of local communication events
  - Drop some information
    - i.e. communications, state records, calls

Linpack 128 procs @ BGL, whole run 157s
Scalability of tracing

Techniques (cont.)

- Summarization
  - What/how
    - Events
      - Software counters
        MPI_Probe
  - When
    - Periodic samples
    - Correlated to other events (ie. function entry/exit)
    - Threshold frequency
Scalability of tracing

- Techniques (cont.)
  - Data structuring
    - ✓ Always good. Not THE solution.
  - Sampling
    - ✓ Not ALL the information.
Scalability: Dimensions

- Acquisition
- Processing
- Presentation
The problem
● Can we squeeze large amounts of data to obtain information and convey it to the analyst?
● Both qualitative presentation and precise/accurate statistics
● Support large dynamic range
● Actual information often
  ✓ In very small data sets.
  ✓ Scattered
  ✓ In the variance
  ✓ Within perturbed/noisy data

Finding needles in haystacks (select the right KB out of 200MB)
Scalability of visualization

- **Issues**
  - Non Linear Rendering
    - #pixels << Space x Time
    - Displayed value:
      - Minimum, maximum, average, last, random selection, …

AMBER. 512 procs @ Marenosrum
IBM Petascale Tools Strategy Workshop, TJWatson. May 3-4th
Scalability of visualization

- Issues (cont.)
  - Non Linear Rendering (cont.)
    - Gradient colors
      - Small # levels

Outlyers? Reason? Impact on averages?
Scalability of visualization

**Issues (cont.)**

- Display a subset of processes
  - Scroll bars, 2D zooms
    - Limited to contiguous processes.
    - Combined with rendering and **simultaneous** display of several views → support a large dynamic range

Sweep3d. 128 procs @ SP

Human eye: a great correlator
Scalability of visualization

**Issues (cont.)**

- Display a subset of processes (cont.)
  - Any subset

Linpack 128 procs @ BGL
IBM Petascale Tools Strategy Workshop, TJWatson. May 3-4th
Issues (cont.)

- Display a subset of processes (cont.)
  - “relevant” subsets?
    - Integrate analysis and visualization (selection by property)

Who sends how much

Some proc. delayed

MPI impl. does not overlap comp/comm

AMBER. 512 procs @ Marenostrum

Who sends to 140

Selection of senders to 140

Histogram
Issues (cont.)

- Aggregation
  - Computing an aggregated metric for groups of threads/processes
    - Following programming model structure of application
    - Following resource model

Functional rather than scalability motivation

NAS LU-MZ. MPI+OpenMP. 121 procs @ SP Nighthawk
Issues (cont.)

- Higher processing functionalities (semantic module)
  - Count
  - Latch
  - Mathematical morphology, ...
  - ...

Linpack 128 procs @ BGL
IBM Petascale Tools Strategy Workshop, TJWatson. May 3-4th
Scalability: Dimensions

- Acquisition
- Processing
- Presentation
Trace post processing

The issue

- Match tracing and visualization tool capabilities
- Minimum loss, maximum compatibility

User driven

Automatic analysis

- Selection of a proper subset of a trace
- Point to potential performance problems
Techniques

- Same as for tracing component.
  - Batch: more time for computation
  - Available information: only what was emitted to original trace.

- Trace size control
  - On/off: space / time
  - Limit file size
  - Automatic application structure identification

AMBER. 1024 procs (500-530) @ Marenostrum
IBM Petascale Tools Strategy Workshop, TJWatson. May 3-4th
Trace post processing

- Techniques (cont.)
  - Summarization (software counters)
    - Periodic samples: used to identify regions of activity, periodicities
    - Threshold frequency
    - Correlated to other events (i.e., function entry/exit)

*AMBER. 512 procs @ Marenostrum*

*Linpack 1024 procs @ Marenostrum*

IBM Petascale Tools Strategy Workshop, TJWatson. May 3-4th
Scalability

Dimensions
  • Acquisition
  • Presentation
  • Processing

CEPBA-tools evolution

Conclusions

Work partially supported by IBM, EC (HPC_Europa), MEC
CEPBA-tools towards scalability

- Large flat files
- Scalable merge
- Simple batch filters
- Combinations

- Speed up trace load, metric computation
- Increase semantic functionality
- Still memory eager
 Scalable merge

Partially resolved communications
CEPBA-tools towards scalability

- Paraver performance
  - Trace loading
  - Interval computation
  - Replication of records

... still room for improvement
IBM Petascale Tools Strategy Workshop, TJWatson. May 3-4th
Conclusion

- We do believe trace based approaches can be used at large scale systems
  - Consider detail of time and space variability.

- We need
  - Tracing and preprocessing tools
    - Intelligence/flexibility to select what to trace
  - Analysis tools
    - Flexible statistics and structure identification algorithms
    - Flexible selection / focusing mechanisms to expose effects
    - Loop between stats and selection mechanisms
  - Plus a lot of memory
  - Plus patience, curiosity,....