A SPATIOTEMPORAL DATA AGGREGATION TECHNIQUE FOR PERFORMANCE ANALYSIS OF LARGE-SCALE EXECUTION TRACES

IEEE Cluster 2014 - 25th September 2014

Damien Dosimont 1 2, Robin Lamarche-Perrin 3, Lucas M. Schnorr 4, Guillaume Huard 2 1, Jean-Marc Vincent 2 1

1 Inria, 2 Univ. Grenoble Alpes, LIG, CNRS, F-38000 Grenoble, France
first.last@imag.fr
3 MPI for Mathematics in the Sciences, 04103 Leipzig, Germany
robin.lamarche-perrin@mis.mpg.de
4 Informatics Institute, UFRGS, Porto Alegre
schnorr@inf.ufrgs.br
INTRODUCTION
TRACE VISUALIZATION PROBLEMATIC

- Trace contents:
  - SPACE = application structure:
    - hardware components: clusters, machines, cores, etc.
    - software components: processes, threads, etc.
  - TIME = timestamped events:
    - function calls, communications, CPU load, malloc, etc.

- Traces can be HUGE
  → scalability issues of space-time representations
PROBLEMATIC VISUALIZATION
OBJECTIVE: SPATIOTEMPORAL OVERVIEW...

- Overcoming these issues thanks to data aggregation
- Showing meaningful information

Grid'5000 Nancy, 700 Cores, NASBP LU, class C
TABLE OF CONTENTS

1. Introduction
2. Previous works
3. Spatiotemporal data aggregation
4. Conclusion
PREVIOUS WORKS
DATA AGGREGATION METHODOLOGY

- A1. Choose a **model** and a **metric**
- A2. Choose on **which dimension(s) aggregate**
- A3. Define the **operands**
- A4. **Constrain** the aggregation: $\rightarrow$ partitions $\mathcal{P}$ allowed
- A5. Define the **operator**
- A6. Define the **trigger** - the aggregation condition
- A7. Build the **algorithm** satisfying A1-A6
INFORMATION LOSS: KL DIVERGENCE

\[
\text{loss}_E = \sum_{e \in E} \rho_e \log_2 \left( \frac{\rho_e}{\rho_E} \right)
\]
COMPLEXITY REDUCTION: SHANNON ENTROPY

\[ \text{gain}_E = \rho_E \log_2 \rho_E - \sum_{e \in E} \rho_e \log_2 \rho_e \]
TRADE-OFF: PIC

\[ pIC_E = p \text{gain}_E - (1-p) \text{loss}_E \]

\[ pIC_P = \sum_{E \in \mathcal{P}} pIC_E \]

- For a given \( p \): choose \( \mathcal{P} \) with the highest \( pIC \)
- Aggregate in priority most homogeneous values
VIVA: SPATIAL AGGREGATION

A Hierarchy: Cluster (3) - Machine (50) - Process (433)

B Ratio Gain/Loss with P = 10%

C Ratio Gain/Loss with P = 30%
OCELOT L: TEMPORAL AGGREGATION
SPATIOTEMPORAL CRITERIA

- M1. Spatiotemporal representation
- M2. Aggregation coherence
SPATIOTEMPORAL DATA AGGREGATION
A.1 MICROSCOPIC MODEL

\[ |X| = 2, \rho_x(s, t) = \frac{d_x(s, t)}{d(t)} \in [0, 1], \rho_1(s, t) = 1 - \rho_2(s, t) \]
A2-A5

- A2. We aggregate simultaneously on $T$ and $S$
- A3. Operands: $(s, t) \in S \times T$
- A4. Constraint: $\mathcal{A}(S \times T) = \mathcal{H}(S) \times \mathcal{I}(T)$
  Aggregation result is a partition $\mathcal{P}(S \times T) \in \mathcal{A}(S \times T)$
- A5. Operator: $+$
- A6. Trigger: maximize $pIC$ of the partition $\mathcal{P}(S \times T)$
BEST CUT ALGORITHM

Compute the partition with the highest $pIC$:
- Cut an area: time, space (or no cut)
- Best cut: the partition $\mathcal{P}$ where $\sum_{E \in \mathcal{P}} pIC_E$ is max
- Recursively cut and evaluate the partitions of $E_1, E_2 \in \mathcal{P}$
- Useless recomputation is avoided
INFLUENCE OF P

\[ p = 0 \]

\[ p = a, \ 0 < a < 1 \]

\[ p = b, \ 0 < a < b < 1 \]

\[ p = 1 \]
CG CLASS C, 64 PROCESSES ON G5K RENNES
LU CLASS C, 700 PROCESSES ON G5K NANCY

- **Introduction**
- **Previous works**
- **Spatiotemporal data aggregation**
- **Conclusion**

**Diagram:**
- MPI_Init
- MPI_Allreduce
- MPI_Recv

**Time:** 0s - 60s

- $S_A$
- $S_B$
- $S_C$

**Annotations:**
- Heterogeneous Spatiotemporal Behavior
- Temporal Perturbation
## PERFORMANCES

<table>
<thead>
<tr>
<th>Case</th>
<th>Application</th>
<th>Processes</th>
<th>Site</th>
<th>Clusters (nodes)</th>
<th>Event number</th>
<th>Trace size</th>
<th>Ocelotl computation times (30 time slices)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>CG, class C</td>
<td>64</td>
<td>Rennes</td>
<td>parapide(8)</td>
<td>3,838,144</td>
<td>136.9 MB</td>
<td>Trace reading + Microscopic description</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>adonis(9), del(24), genepi(31)</td>
<td>49,149,440</td>
<td>1.8 GB</td>
<td>5s + &lt;1s</td>
</tr>
<tr>
<td>B</td>
<td>CG, class C</td>
<td>512</td>
<td>Grenoble</td>
<td>adonis(9), genepi(31)</td>
<td>49,149,440</td>
<td>1.8 GB</td>
<td>31s + &lt;1s</td>
</tr>
<tr>
<td>C</td>
<td>LU, class C</td>
<td>700</td>
<td>Nancy</td>
<td>graphene(26), graphite(4), griffon(67)</td>
<td>218,457,456</td>
<td>8.3 GB</td>
<td>222s + 2s</td>
</tr>
<tr>
<td>D</td>
<td>LU, class B</td>
<td>900</td>
<td>Rennes</td>
<td>paradent(38), parapide(21), parapluie(18)</td>
<td>177,376,729</td>
<td>6.7 GB</td>
<td>174s + 2s</td>
</tr>
</tbody>
</table>

**Note:** The table provides a comparison of different cases (A to D) in terms of application, processes, site, clusters, event number, trace size, and Ocelotl computation times for the tasks of trace reading and aggregation.
CONCLUSION
CONCLUSION

- Visualization based on spatiotemporal data aggregation
  - Solves screen, computing and analyst capability limitations
  - Gives meaningful information about homogeneity (phases, perturbations)
  - Two use cases show its relevancy

- Future work:
  - Improve visualization and interaction to get more details
  - Extend methodology and design new algorithms
    \((\mathcal{H}(S) \times \mathcal{H}(S) \times \mathcal{I}(T), \text{surface, etc.})\)
**LINKS**

<table>
<thead>
<tr>
<th>Ocelotl:</th>
<th><a href="http://github.com/dosimont/ocelotl">http://github.com/dosimont/ocelotl</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Framesoc:</td>
<td><a href="http://github.com/generoso/framesoc">http://github.com/generoso/framesoc</a></td>
</tr>
<tr>
<td>Viva:</td>
<td><a href="http://github.com/schnorr/viva">http://github.com/schnorr/viva</a></td>
</tr>
</tbody>
</table>
THANK YOU FOR YOUR ATTENTION
OCELOT: TEMPORAL AGGREGATION (1)
OCELOTL: TEMPORAL AGGREGATION (2)
ELMQVIST-FEKETE CRITERIA

- **Shneiderman**: overview, zoom and filter, then get details on demand
- **Elmqvist & Fekete**: guidelines to design an overview visualization based on hierarchical aggregation
  - G1. Entity Budget
  - G2. Visual Summary
  - G3. Visual Simplicity
  - G4. Discriminability
  - G5. Fidelity
  - G6. Interpretability
Example of Gantt chart - space-time diagram
VISUALIZATIONS NOT FULFILLING THESE CRITERIA (2)

KPTre: $G_1$ (time), $G_2$, $G_4$, $G_5$
Pajé: G1 (space), G2