Model Checking Contest

Report for 2011

Fabrice Kordon - LIP6/MoVe, UPMC, France
Alban Linard - CUI/SMV, Univ. Genève, Switzerland
Contents

Objectives

Evaluation procedure

The models

Participating tools

Analysis of the results

Concluding remarks
Contents

- Objectives
- Evaluation procedure
- The models
- Participating tools
- Analysis of the results
- Concluding remarks

Special thanks for those who helped to organize this MCC, in particular Clément Démoulins (infrastructure), Nicolas Gibelin (infrastructure and cluster), Lom Hillah (PNML), Emmanuel Paviot-Adet (models), Steve Hostettler (Properties) and Alexis Marechal (models)
Objectives
Lots of questions are raised...
  - To verify highly concurrent systems, should we use a symmetry-based or a partial order-based model checker?
  - For models with large variable domains, should we use decision diagram-based, or a symmetry-based model checker?
  - Can we combine structural reductions techniques with partial-order ones or symmetry-based ones?
  ...
**The Objectives...**

**MCC** is intended to:
- Exchange experience between tool programmers,
- Imagine some association of techniques, and thus better tools
- Stimulate development of tools
- Provide visibility to these tools

**MCC** can also be of great help for the PN community (and users):
- Define a common set of models for benchmarks
- Identify experimentally classes of problems (in models)
  - identify the techniques able to cope with a given class of problems...
- Improve communication between tools (and PNML ;-))
- Provides raw data for comparison

**This is a first edition**
- We hope more editions for an enhanced analysis and evaluation of tools
Evaluation Procedure
What to be measured?

The «enemies» of model checking
- Memory consumption
- CPU consumption

«Examinations» to be processed
- State space generation
- Deadlock detection
- Formula evaluation
  - Reachability
  - Verified and unverified

Etc... (for next editions)
- Temporal formulae (LTL, CTL)?
- Stochastic analysis?
- Time analysis?
**Evaluation procedure**

- **Execution on a dedicated cluster**
  - 22 2.4GHz bi-Pentium Xeon with 2Gbyte
  - One machine fully dedicated to tools (minimal preemption)
  - Tools must process «examination» for each models
    - Scale parameter to evaluate how far tools can run

- **Run = execution of a tool for one examination on one model/scale**

- **A benchmark script launching all runs**
  - with time confinement
    - 1800 sec per run
  - with memory confinement
    - 1.75 GByte per run
  - with both time and memory measures
  - deployment + execution on the cluster via OAR
    - OAR = resource manager for cluster developed at INRIA
Submission

- **Step 0**: read the rules to check for submission condition
- **Step 1**: download the submission kit (benchmark script + PNML files)
  - **Step 2**: integration of the tool in the execution system
    - translate the models (or parse PNML ;-) ) + parse formulas
    - wrap the tool for Unix and provide appropriate outputs
    - adapt setup.sh file
- **Step 3**: use the appropriate Makefile entry to build the submission archive
- **Step 4**: upload your file via a dedicated web page

Evaluation

- **Step 0**: check execution of tools on the cluster + adaptations
- **Step 1**: finalize the execution environment
  - Check for confinement
  - Check for the metrics capture
  - Extract results
- **Step 2**: run all submissions (with final values for confinement)
- **Step 3**: analyze results
Submission

Step 0: read the rules to check for submission condition

Step 1: download the submission kit (benchmark script + PNML files)

Step 2: integration of the tool in the execution system
  - translate the models (or parse PNML ;-) ) + parse formulas
  - wrap the tool for Unix and provide appropriate outputs
  - adapt setup.sh file

Step 3: use the appropriate Makefile entry to build the submission archive

Step 4: upload your file via a dedicated web page

Evaluation

Step 0: check execution of tools on the cluster + adaptations

Step 1: finalize the execution environment
  - Check for confinement
  - Check for the metrics capture

Step 2: run all submissions (with final values for confinement)

Step 3: analyze results

Memory measure was sort of a nightmare. It is not perfect yet.

However, all tools were evaluated the same way
Submission and evaluation procedures

Submission

Step 0: read the rules to check for submission condition
Step 1: download the submission kit (benchmark script + PNML files)
Step 2: integration of the tool in the execution system
   - translate the models (or parse PNML :-) ) + parse formulas
   - wrap the tool for Unix and provide appropriate outputs
   - adapt setup.sh file
Step 3: use the appropriate Makefile entry to build the submission archive
Step 4: upload your file via a dedicated web page

Evaluation

Step 0: check execution of tools on the cluster + adaptations
Step 1: finalize the execution environment
   - Check for confinement
   - Check for the metrics capture
   - Extract results
Step 2: run all submissions (with final values for confinement)
Step 3: analyze results

Memory measure was sort of a nightmare. It is not perfect yet.
However, all tools were evaluated the same way

Many thanks to the tool developers and their nice reactivity for solving troubles as well as for answering questions
The Models
P/T Nets

- **FMS (Flexible Manufacturing System)**
  - Bench from SMART

- **Kanban**
  - Bench from SMART

- **MAPK (Mitogen-activated protein kinase kaskade)**
  - Bench from Cottbus

Colored Nets

- **Peterson**
  - G.L. Peterson’s mutual exclusion algorithm

- **Philosophers**
  - Classical illustration for deadlocks

- **SharedMemory**
  - Bench from GreatSPN (colored version) presented in PNPM’1989

- **TokenRing**
  - Proposed by E.W. Dijkstra in CACM (1974)
<table>
<thead>
<tr>
<th>Models Characteristics</th>
<th>FMS</th>
<th>Kanban</th>
<th>MAPK</th>
<th>Peterson</th>
<th>Philosophers</th>
<th>SharedMemory</th>
<th>TokenRing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P/T</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Safe</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Colored</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cartesian product of color types</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Non equal guards</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Broadcast function</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Succ &amp; pred functions</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
# Models Characteristics

<table>
<thead>
<tr>
<th></th>
<th>FMS</th>
<th>Kanban</th>
<th>MAPK</th>
<th>Peterson</th>
<th>Philosophers</th>
<th>SharedMemory</th>
<th>TokenRing</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/T</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safe</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td></td>
</tr>
<tr>
<td>Colored</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cartesian product of color types</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non equal guards</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broadcast function</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Succ &amp; pred functions</td>
<td>✅</td>
<td>✅</td>
<td>✅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next MCC

- Better coverage of model characteristics
- More models
Participating Tools
<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Team</th>
<th>Institution</th>
<th>Country</th>
<th>Contact Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTIVITY-LOCAL</td>
<td>TIK</td>
<td>ETHZ</td>
<td>Switzerland</td>
<td>K. Lampka</td>
</tr>
<tr>
<td>AlPiNA</td>
<td>CUI/SMV</td>
<td>Univ. Geneva</td>
<td>Switzerland</td>
<td>D. Buchs</td>
</tr>
<tr>
<td>Crocodile</td>
<td>LIP6/MoVe</td>
<td>UPMC</td>
<td>France</td>
<td>M. Colange</td>
</tr>
<tr>
<td>ITS-Tools</td>
<td>LIP6/MoVe</td>
<td>UPMC</td>
<td>France</td>
<td>Y. Thierry-Mieg</td>
</tr>
<tr>
<td>LoLA</td>
<td>Team Rostock</td>
<td>Univ. Rostock</td>
<td>Germany</td>
<td>N. Lohmann and K. Wolf</td>
</tr>
<tr>
<td>PNXDD</td>
<td>LIP6/MoVe</td>
<td>UPMC</td>
<td>France</td>
<td>E. Paviot-Adet</td>
</tr>
<tr>
<td>PeTe</td>
<td>Stud. Group d402b</td>
<td>Univ. Aalborg</td>
<td>Denmark</td>
<td>J. Finnemann Jensen</td>
</tr>
<tr>
<td>Sara</td>
<td>Team Rostock</td>
<td>Univ. Rostock</td>
<td>Germany</td>
<td>H. Wimmel and K. Wolf</td>
</tr>
<tr>
<td>YASPA</td>
<td>TIK</td>
<td>ETHZ</td>
<td>Switzerland</td>
<td>K. Lampka</td>
</tr>
<tr>
<td>helena</td>
<td>LIPN/LCR</td>
<td>Univ. Paris 13</td>
<td>France</td>
<td>S. Evangelista</td>
</tr>
<tr>
<td>Tool</td>
<td>RG (Reachability Graph)</td>
<td>DL (Deadlock Detection)</td>
<td>F (Formula Evaluation)</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td>ACTIVITY-LOCAL</td>
<td>EXPPLICIT DECISION_DIAGRAMS OTHERS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AlPiNA</td>
<td>DECISION_DIAGRAMS</td>
<td>DECISION_DIAGRAMS</td>
<td>SYMMETRIES DECISION_DIAGRAMS</td>
<td></td>
</tr>
<tr>
<td>Crocodile</td>
<td>SYMMETRIES DECISION_DIAGRAMS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITS-Tools</td>
<td>DECISION_DIAGRAMS SYMMETRIES (opt)</td>
<td>DECISION_DIAGRAMS SYMMETRIES (opt)</td>
<td>DECISION_DIAGRAMS SYMMETRIES (opt)</td>
<td></td>
</tr>
<tr>
<td>LoLA</td>
<td>EXPPLICIT PARTIAL_ORDERS STATE_COMPRESSION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNXDD</td>
<td>DECISION_DIAGRAMS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PeTe</td>
<td>EXPPLICIT ABSTRACTIONS OTHERS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sara</td>
<td>ABSTRACTIONS PARTIAL_ORDERS OTHERS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YASPA</td>
<td>DECISION_DIAGRAMS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>helena</td>
<td>EXPPLICIT ABSTRACTIONS PARTIAL_ORDERS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### No formula evaluation for Peterson

- **SUMo 2011 - Model Checking Contest report - June 21, 2011**

### Less than 50% of submitted tools handling colored nets

<table>
<thead>
<tr>
<th></th>
<th>FMS</th>
<th>Kanban</th>
<th>MAPK</th>
<th>Peterson</th>
<th>Philosophers</th>
<th>Shared Memory</th>
<th>TokenRing</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTIVITY-LOCAL</td>
<td>RG</td>
<td>RG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIPiNA</td>
<td>DL RG</td>
<td>DL RG</td>
<td>DL RG</td>
<td>DL RG</td>
<td>DL RG</td>
<td>DL RG</td>
<td>DL RG</td>
</tr>
<tr>
<td>Crocodile</td>
<td>F RG</td>
<td>F RG</td>
<td>F RG</td>
<td></td>
<td>F RG</td>
<td>F RG</td>
<td>F RG</td>
</tr>
<tr>
<td>ITS-Tools</td>
<td>DL F RG</td>
<td>DL F RG</td>
<td>DL F RG</td>
<td></td>
<td>DL F RG</td>
<td>DL F RG</td>
<td>DL F RG</td>
</tr>
<tr>
<td>LoLA</td>
<td>F</td>
<td>F</td>
<td>FOK</td>
<td></td>
<td></td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>PNXDDD</td>
<td>RG</td>
<td>RG</td>
<td>RG</td>
<td>RG</td>
<td>RG</td>
<td>RG</td>
<td>RG</td>
</tr>
<tr>
<td>PeTe</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sara</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YASPA</td>
<td>RG</td>
<td>RG</td>
<td>RG (8)</td>
<td></td>
<td>DL RG</td>
<td>DL RG</td>
<td>DL RG</td>
</tr>
<tr>
<td>helena</td>
<td>DL RG</td>
<td>DL RG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No formula evaluation for Peterson
Less than 50% of submitted tools handling colored nets
Analysis of the Results
No interest in a «race»

More than 175 charts generated

- Synthesis to be published (TOPNOC report or tech report or web)
- Raw data will be available on-line or as an annex to the report
- It is annoying to show them all...

Identification (partial) of some «surprises» discovered when test were processed

- How tools scale up
  - P/T and colored
- Some observations on time and memory consumption
- Feed back with tools' characteristics
## How Did Tools Scale Up (P/T Models)

<table>
<thead>
<tr>
<th></th>
<th>A-LOCAL</th>
<th>AIPiNA</th>
<th>Crocodile</th>
<th>ITS-Tools</th>
<th>LoLA</th>
<th>PNXDD</th>
<th>PeTe</th>
<th>Sara</th>
<th>YASPA</th>
<th>helena</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FMS</strong></td>
<td>RG</td>
<td>20</td>
<td>10</td>
<td>2</td>
<td>100</td>
<td>500</td>
<td>500</td>
<td>100</td>
<td>500</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>10</td>
<td>50</td>
<td>200</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td>2</td>
<td>200</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td>2</td>
<td>50</td>
<td>20</td>
<td></td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td><strong>Kanban</strong></td>
<td>RG</td>
<td>20</td>
<td>10</td>
<td>200</td>
<td>500</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>10</td>
<td>200</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td>200</td>
<td>200</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td>200</td>
<td>10</td>
<td>1000</td>
<td></td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td><strong>MAPK</strong></td>
<td>RG</td>
<td>8</td>
<td>160</td>
<td>160</td>
<td>160</td>
<td>160</td>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>8</td>
<td>160</td>
<td>320</td>
<td>320</td>
<td>320</td>
<td>320</td>
<td>320</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td>160</td>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td>160</td>
<td>20</td>
<td></td>
<td></td>
<td>320</td>
<td>320</td>
<td>320</td>
<td>320</td>
<td></td>
</tr>
</tbody>
</table>
## Why Did Tools Failed

(P/T Models)

<table>
<thead>
<tr>
<th></th>
<th>A-LOCAL</th>
<th>AlPiNA</th>
<th>Crocodile</th>
<th>ITS-Tools</th>
<th>LoLA</th>
<th>PNXDD</th>
<th>PeTe</th>
<th>Sara</th>
<th>YASPA</th>
<th>helena</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FMS</strong></td>
<td>RG</td>
<td>EDNF</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td>MOVF</td>
<td>EDNF</td>
<td>MOVF</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td>MOVF</td>
<td>SOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td><strong>Kanban</strong></td>
<td>RG</td>
<td>EDNF</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>???</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVEF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td>MOVF</td>
<td>SOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td><strong>MAPK</strong></td>
<td>RG</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>EDNF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td>MOVF</td>
<td>SOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
</tr>
</tbody>
</table>

- **EDNF** = execution did not finished
- **MOVF** = memory overflow
- **SOVF** = stack overflow
- **???** = cannot determine failure
### How Did Tools Scale Up (Colored Models)

<table>
<thead>
<tr>
<th></th>
<th>A-LOCAL</th>
<th>AIPiNA</th>
<th>Crocodile</th>
<th>ITS-Tools</th>
<th>LoLA</th>
<th>PNXDD</th>
<th>PeTe</th>
<th>Sara</th>
<th>YASPA</th>
<th>helena</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peterson</strong></td>
<td>RG</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Philosophers</strong></td>
<td>RG</td>
<td>500</td>
<td>10</td>
<td>100000</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>100</td>
<td></td>
<td>100000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td>5</td>
<td>5000</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td>5</td>
<td>5000</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td><strong>SharedMemory</strong></td>
<td>RG</td>
<td>20</td>
<td>20</td>
<td>50000</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>20</td>
<td></td>
<td>50000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td>10</td>
<td>20000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td>10</td>
<td>20000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td><strong>TokenRing</strong></td>
<td>RG</td>
<td>10</td>
<td></td>
<td>50</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>10</td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>FOK</td>
<td></td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FNOK</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Why Did Tools Failed (Colored Models)

<table>
<thead>
<tr>
<th></th>
<th>A-LOCAL</th>
<th>APINa</th>
<th>Crocodile</th>
<th>ITIS-Tools</th>
<th>LoLA</th>
<th>PNXDD</th>
<th>PeTe</th>
<th>Sara</th>
<th>YASPA</th>
<th>helena</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peterson</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>RG</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
<td>???</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>DL</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOK</td>
<td>FOK</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNOK</td>
<td>FNOK</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Philosophers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>RG</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
<td>SOVF</td>
<td>???</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>DL</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOK</td>
<td>FOK</td>
<td>EDNF</td>
<td>MOVF</td>
<td>SOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNOK</td>
<td>FNOK</td>
<td>EDNF</td>
<td>MOVF</td>
<td>SOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SharedMemory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>RG</td>
<td>MOVF</td>
<td>MOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>DL</td>
<td>MOVF</td>
<td>MOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOK</td>
<td>FOK</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNOK</td>
<td>FNOK</td>
<td>MOVF</td>
<td>MOVF</td>
<td>MOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TokenRing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>RG</td>
<td>MOVF</td>
<td>MOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>DL</td>
<td>MOVF</td>
<td>MOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOK</td>
<td>FOK</td>
<td>MOVF</td>
<td>MOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNOK</td>
<td>FNOK</td>
<td>MOVF</td>
<td>MOVF</td>
<td></td>
<td></td>
<td>EDNF</td>
<td>EDNF</td>
<td>EDNF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **EDNF** = execution did not finished
- **SOVF** = stack overflow
- **MOVF** = memory overflow
- **???** = cannot determine failure
Memory measure for state space generation (FMS)

Tools
- ACTIVITY–LOCAL
- AlPiNA
- Crocodile
- ITS–Tools
- PNXDD
- YASPA
- helena
Memory measure for state space generation (FMS)

Efficiency of Decision diagrams
Memory measure for state space generation (FMS)

Crocodile in bad shape (P/T nets)
Memory measure for state space generation (FMS)

![Diagram showing memory measure for state space generation with different tools and their `Initial cost` for PNXDD.]
Memory measure for state space generation (FMS)

![Graph showing memory measure vs scaling parameter for different tools.]

Tools:
- ACTIVITY–LOCAL
- AlPiNA
- Crocodile
- ITS–Tools
- PNXDD
- YASPA
- helena

**JVM effect for AlPiNA**
helena is in bad condition because reductions are suppressed (unfair)
Memory measure for state space generation (FMS)

- **ACTIVITY-LOCAL**: $2.50 \times 10^9$
- **AlPîNA**: $1.90 \times 10^{25}$
- **Crocodile**: $6.03 \times 10^{12}$
- **ITS-Tools**: $2.70 \times 10^{21}$
- **PNXDD**: $2.50 \times 10^9$
- **YASPA**: $1.90 \times 10^{25}$
- **helena**: $6.03 \times 10^{12}$

Scaling parameter vs. Memory (Kbyte)
FMS, Deadlock Detection

CPU measure for deadlock detection (FMS)

- Tools:
  - AlPiNA
  - ITS-Tools
  - helena

Scaling parameter vs. Time (seconds)
CPU measure for deadlock detection (FMS)

- Impressive efficiency of reduction techniques (beats DD)

- Tools:
  - AlPiNA
  - ITS-Tools
  - helena

- 1.90x10^25

- Not computed
CPU measure for deadlock detection (FMS)

For helena, memory chart shows negligible consumption compared to allocation for execution.

1.90x10^25

not computed
CPU measure for the evaluation of verified formula (FMS)

CPU measure for the evaluation of unverified formula (FMS)
CPU measure for the evaluation of verified formula (FMS)

DD suffer but scale up well

CPU measure for the evaluation of unverified formula (FMS)
Class 1: tools that are excellent for OK formulæ, no need to go over the full state space.
Class 2: tools that are excellent for NOK formulæ, sort of structural analysis approach.
FMS, Reachability Analysis

CPU measure for the evaluation of verified formula (FMS)

CPU measure for the evaluation of unverified formula (FMS)

Class 3: Sara (but wait ;-)
Memory measure for the evaluation of verified formula (Kanban)

Memory measure for the evaluation of unverified formula (Kanban)
Similar observation than in the previous slide for memory... even Sara need some (OK formulæ)
Memory measure for the evaluation of verified formula (Kanban)

Similar observation than in the previous slide for memory... even Sara need some (OK formulæ)
MAPK, Reachability Analysis

Memory measure for the evaluation of verified formula (MAPK)

Memory measure for the evaluation of unverified formula (MAPK)
Once again, we observe the three classes of tools:
Interesting memory behavior for LoLA, partial order techniques need parallelism...
MAPK, Reachability Analysis

Memory measure for the evaluation of verified formula (MAPK)

Strange answers with OK formulæ for Sara (disagree with other tools)

Memory measure for the evaluation of unverified formula (MAPK)

Tools
- ITS
- LoLA
- PeTe
- Sara

Strange answers with OK formulæ for Sara (disagree with other tools)
MAPK, Reachability Analysis

Memory measure for the evaluation of verified formula (MAPK)

- Memory (Kbyte) vs. Scaling parameter

Memory measure for the evaluation of unverified formula (MAPK)

- Memory (Kbyte) vs. Scaling parameter

Tools:
- ITS
- LoLA
- PeTe
- Sara

- Memory measures:
  - ITS: $8.81 \times 10^{10}$
  - LoLA: $4.78 \times 10^{14}$
  - PeTe: $1.06 \times 10^{23}$
Peterson, State space generation

Memory measure for state space generation (Peterson)

CPU measure for state space generation (Peterson)
«bad» model for tools... strong synchronization that disable all suggested optimization techniques
«bad» model for tools... strong synchronization that disable all suggested optimization techniques
CPU measure for deadlock detection (Peterson)

Time (seconds)

Scaling parameter

Tools

AIPiNA

helena
Good scalability of reduction techniques compared to DD.
PETERSON, DEADLOCK DETECTION

CPU measure for deadlock detection (Peterson)

AlPiNA, computation time negligible compared to model loading and conversion.
CPU measure for state space generation (Philosophers)

Memory measure for state space generation (Philosophers)
Hierarchical DD + symmetries, specific «recursive folding» that provides huge scalability when enabled.
Hierarchical DD + symmetries, specific «recursive folding» that provides huge scalability when enabled.
Philosopher, State Space Generation

CPU measure for state space generation (Philosophers)

Memory measure for state space generation (Philosophers)

Scaling parameter

Tools
- AlPiNA
- Crocodile
- ITS
- PNXDD
- helena

Memory (Kbyte)

Time (seconds)
For regular models, we still see superiority of DD.
For regular models, we still see superiority of DD.
For formulae evaluation, ITS-Tools fails at 100000 ;-)
Memory measure for state space generation (SharedMemory)

CPU measure for state space generation (SharedMemory)
Once again, impressive application of the «recursive folding» technique.
Once again, impressive application of the «recursive folding» technique.

<table>
<thead>
<tr>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlPiNA</td>
</tr>
<tr>
<td>Crocodile</td>
</tr>
<tr>
<td>ITS</td>
</tr>
<tr>
<td>PNXDD</td>
</tr>
<tr>
<td>helena</td>
</tr>
</tbody>
</table>

Memory measure for state space generation (SharedMemory)

CPU measure for state space generation (SharedMemory)
Once again, impressive application of the «recursive folding» technique.

The presence of a non-equal guard seems to annoy DD-only-based tools.
Once again, impressive application of the «recursive folding» technique.

The presence of a non-equal guard seems to annoy DD-only-based tools.
TokenRing, State Space Generation

Memory measure for state space generation (TokenRing)

CPU measure for state space generation (TokenRing)

Tools
- AlPiNA
- ITS-Tools
- PNXDD
- helena
TokenRing, State Space Generation

Memory measure for state space generation (TokenRing)

CPU measure for state space generation (TokenRing)

No «recursive folding» for ITS-Tool, only hierarchy in DD
TokenRing, State Space Generation

---

**Memory measure for state space generation (TokenRing)**

![Memory Graph](image)

**CPU measure for state space generation (TokenRing)**

![CPU Graph](image)

No «recursive folding» for ITS-Tool, only hierarchy in DD

**Tools**
- AIPiNA
- ITS-Tools
- PNXDD
- helena

---

Scaling parameter

Memory (Kbyte)

- $10^3$
- $10^4$
- $10^5$
- $10^6$

Time (seconds)

- $10^{-3}$
- $10^{-2}$
- $10^{-1}$
- $10^0$
- $10^1$
- $10^2$

Tools

- AIPiNA
- ITS-Tools
- PNXDD
- helena

---

- AIPiNA: $9.70 \times 10^6$
- ITS-Tools: $1.69 \times 10^4$
- PNXDD: $6.56 \times 10^9$
- helena: $3.81 \times 10^{15}$
State Space Generation, Best Tools
State Space Generation, Best Tools

ITS-Tools

PNXDD

AIpINA

SharedMemory

TokenRing

Philosophers

FMS

MAPK

Kanban

Peterson

FMS

SharedMemory

TokenRing

Philosophers

FMS

MapK

Kanban

Peterson
State Space Generation, Best Tools

how far the tool goes for the model
Surface increased by excellent support of Philosopher & SharedMemory (thanks to «recursive folding»)
State Space Generation, Best Tools

Surface increased by excellent support of Philosopher & SharedMemory (thanks to «recursive folding»)

Use of hierarchical DD as for ITS-Tools (but in a different way)
State Space Generation, Best Tools

Use of hierarchical DD as for ITS-Tools (but in a different way)

PNXDD

Use of DD too

Surface increased by excellent support of Philosopher & SharedMemory (thanks to «recursive folding»)

ITS-Tools
State Space Generation, Best Tools
(P/T Nets)
State Space Generation, Best Tools (P/T Nets)

- ITS-Tools
- PNXDD
- MAPK
- FMS
- Kanban
- Peterson
- TokenRing
- SharedMemory
- Philosophers

Use of DD too
Hierarchical DD are also good for P/T nets

Use of DD too
Deadlock Detection, Best Tools
Deadlock Detection, Best Tools

ITS-Tools

AlPiNA

helena

FMS

Kanban

MAPK

Peterson

TokenRing

SharedMemory

Philosophers

1

2

3
Deadlock Detection, Best Tools

Good scalability for the reduction techniques
Deadlock Detection, Best Tools (P/T Nets)
For FMS, reductions are more efficient than DD
Reachability Analysis (Verified),
Best Tools
Reachability Analysis (Verified), Best Tools

LoLA

ITS-Tools

Sara

MAPK

FMS

Kanban

TokenRing

Peterson

SharedMemory

Philosophers
Partial order and state compression are doing good too
Reachability Analysis (Verified), Best Tools

LoLA

Partial order and state compression are doing good too

Abstractions and partial order are doing good too

Sara
Reachability Analysis (Verified),
Best Tools (P/T Nets)
Reachability Analysis (Verified), Best Tools (P/T Nets)
Reachability Analysis (Verified), Best Tools (P/T Nets)

For P/T nets, abstraction + partial order / state compression seem better than DD
Reachability Analysis (Unverified), Best Tools
Reachability Analysis (Unverified), Best Tools

**ITS-Tools**

- FMS
- Kanban
- MAPK
- Peterson
- TokenRing
- SharedMemory
- Philosophers

**PeTe**

- Kanban
- MAPK
- Peterson
- TokenRing
- SharedMemory
- Philosophers

**Sara**

- FMS
- Kanban
- MAPK
- Peterson
- TokenRing
- SharedMemory
- Philosophers
Reachability Analysis (Unverified), Best Tools (P/T Nets)
Reachability Analysis (Unverified), Best Tools (P/T Nets)

For P/T nets, abstraction + partial order/others seem better than DD.
**About Decision Diagrams and Symmetries**

- **Decision Diagrams**
  - thick = worst
  - thin = best

- **Symmetries**

**`technique by technique` `radar`**
- Subtract maximum of worse tools with maximum of best tools
- Shows progression for this technique
- Here, for state space generation

**Stacking good for colored models**
- Especially regular ones
- Good for: RG, DL, FOK, FNOK

**Makes less sense for P/T models**
- What about the model structure?
Both techniques are efficient
- No deadlock detection on MAPK
- For Peterson, see remark to come
- Seems to work on both P/T and Colored Models

Good scalability when stacked
- Sara are examples
Abstraction work on both FOK and FNOK

But only experimented on P/T models this year
Abstraction work on both FOK and FNOK

But only experimented on P/T models this year

State compression works well for FOK
Abstraction work on both FOK and FNOK
  But only experimented on P/T models this year

State compression works well for FOK

Partial orders work for formula evaluation
  Slightly better for FNOK
Abstraction work on both FOK and FNOK
But only experimented on P/T models this year

State compression works well for FOK

Partial orders work for formula evaluation
Slightly better for FNOK

«Other Techniques» works too
Need to be refined
«other»?
Excellent for FNOK
structural evaluation of formula
The «technique by model» radar

- Shows the appropriate technique for a given model

The Peterson model seems very resistant

- Techniques have difficulties to scale up

Special Note on the Peterson Model
Some tools are dedicated to specific classes
- AlPiNA for Algebraic Nets
- Crocodile for SN with Bags

Some tool are more general
- ACTIVITY-LOCAL supports multiple formalisms

Memory measure to be refined
- Strange behavior for PeTe?

These tools may not have been used in their best conditions...
Some tools are dedicated to specific classes
- AlPiNA for Algebraic Nets
- Crocodile for SN with Bags

Some tools are more general
- ACTIVITY-LOCAL supports multiple formalisms

Memory measure to be refined
- Strange behavior for PeTe?

Questions raised from this (partial) study
- Could we stack reduction techniques with DD as for Symmetries?
- Can abstractions be stacked with other techniques?
  - True for abstraction + DD (but not a competing tool)
- Can partial order be stacked with other techniques?

These tools may not have been used in their best conditions...
Some tools are dedicated to specific classes
- AlPiNA for Algebraic Nets
- Crocodile for SN with Bags

Some tool are more general
- ACTIVITY-LOCAL supports multiple formalisms

Memory measure to be refined
- Strange behavior for PeTe?

Questions raised from this (partial) study
- Could we stack reduction techniques with DD as for Symmetries?
- Can abstractions be stacked with other techniques?
  - True for abstraction + DD (but not a competing tool)
- Can partial order be stacked with other techniques?

Lots of new stuff to experiment for the next MCC ;-)

These tools may not have been used in their best conditions...
Concluding Remarks
What Could Come Out?

- A first attempt to evaluate model checkers with common criteria
- Some benchmarks (to be updated of course)
- A common language for queries? (to be updated of course)

Some inputs for tool developers
- Potential association of some techniques
- Identification of some problems

Toward adaptive model checking?
- What technique is appropriate for a given type of model

Example: in the first modeling (debugging) phase, select a tool performant with unverified properties, then chose a tool performant with verified properties
- Tool compatibilities issues then...
What Could Come Out?

- A first attempt to evaluate model checkers with common criteria
- Some benchmarks (to be updated of course)
- A common language for queries? (to be updated of course)

Some inputs for tool developers
- Potential association of some techniques
- Identification of some problems

Toward adaptive model checking?
- What technique is appropriate for a given type of model

Example: in the first modeling (debugging) phase, select a tool performant with unverified properties, then chose a tool performant with verified properties
- Tool compatibilities issues then...
Next Edition in 2012?

As a standalone event?
Summary + presentation of some interesting techniques and approaches

New stuff to be proposed in the next editions
More models: safe P/T, unsafe Colored, «industrial»
More Petri Net classes: timed PN? stochastic PN?
More properties: LTL? CTL?
Better identification of techniques
Enhanced confinement and measure environment

Do we suggest PNML as the only input for tools?
ACTIVITY-LOCAL
http://dblp.uni-trier.de/rec/bibtex/conf/mmb/LampkaS06 (reference but not the tool itself 😞)

AlPiNA
http://alpina.unige.ch

Crocodile
http://move.lip6.fr/software/Crocodile

ITS-Tools

LoLA
http://www.informatik.uni-rostock.de/tpp/lola/

PNXDD
https://srcdev.lip6.fr/trac/research/NEOPPOD/wiki/pnxdd

PeTe
https://github.com/jopsen/PeTe

Sara
http://www.service-technology.org/tools/download

YASPA
http://www.tik.ee.ethz.ch/~klampka

helena
http://helena.mc.sourceforge.net