The QBF Gallery 2013
A Non-Competitive Evaluation of QBF Tools

Florian Lonsing\(^1\)  Martina Seidl\(^2\)  Allen Van Gelder\(^3\)

\(^1\)Vienna University of Technology  
http://www.kr.tuwien.ac.at/staff/lonsing

\(^2\)Johannes Kepler University, Linz, Austria  
http://fmv.jku.at/seidl

\(^3\)University of California at Santa Cruz, USA  
http://www.cse.ucsc.edu/~avg

This work is supported by the Austrian Science Fund (FWF) under grants S11409-N23 and S11408-N23 as well as by the Vienna Science and Technology Fund (WWTF) under grant ICT10-018.
No competition, no winners (nor losers!), no prizes

Goal:

- Evaluate the state-of-the-art in practical QBF research...
- ...by running QBF tools (any kind!) in an organized and centralized manner...
- ...and by collecting and evaluating data...
- ...in a community-driven manner with interaction / intervention opportunities during the runs.
Some Background Information

Organizational Details

- 1st QBF Gallery ever (complementary to biannual QBFEval competition).
- 4 (strongly related) showcases: Preprocessing, Solving, Applications, Certificates.
- Experiments on
  - FMV Cluster @ JKU Linz.
  - Infosys Cluster @ TU Vienna.
- > 7000 considered formulas (from QBFLIB and new benchmarks).
- > 114,000 runs in 3.92297e+07 seconds (11,000 hours).

Submissions:

- 23 contributors from 8 countries.
- 14 CNF-solvers, 1 Non-CNF-solver, 3 2QBF-solvers.
- 4 preprocessors.
- 2 certification tools.
- 5 new benchmark sets.

Details: http://www.kr.tuwien.ac.at/events/qbfgallery2013/
Excerpt of Showcase Preprocessing

- Comparison of individual preprocessors and combinations.
- Evaluation of solving power of preprocessors.
- Time-limited preprocessing
  - ... in multiple rounds
  - ... with different execution sequences
  - ... and fixpoint detection.
- Effects of preprocessing on solver performance.

<table>
<thead>
<tr>
<th></th>
<th>hiqqer3e</th>
<th>Bloqger</th>
<th>hiqqer3p</th>
<th>squeezebf</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t  s  u</td>
<td>t  s  u</td>
<td>t  s  u</td>
<td>t  s  u</td>
</tr>
<tr>
<td>eval2012r2</td>
<td>19  0  19</td>
<td>69  33  36</td>
<td>77  35  42</td>
<td>11  3  8</td>
</tr>
<tr>
<td>qbf-hardness</td>
<td>0   0   0</td>
<td>49  12  37</td>
<td>51  12  39</td>
<td>12  0  12</td>
</tr>
<tr>
<td>sauer-reimer</td>
<td>81  0  81</td>
<td>137 24  113</td>
<td>153 29  124</td>
<td>78  9  69</td>
</tr>
<tr>
<td>planning-CTE</td>
<td>0   0   0</td>
<td>3   2  1</td>
<td>7   6  1</td>
<td>0   0  0</td>
</tr>
<tr>
<td>conf.-planning</td>
<td>646 0  646</td>
<td>489 11  478</td>
<td>486 12  474</td>
<td>48  0  48</td>
</tr>
<tr>
<td>red.-finding</td>
<td>176 0  176</td>
<td>1496 837 659</td>
<td>1650 924 726</td>
<td>674 326 348</td>
</tr>
</tbody>
</table>

Individual preprocessors: solved instances \((t)\), solved satisfiable \((s)\) and solved unsatisfiable instances \((u)\).
■ 345 formulas.
■ 69 solved by preprocessor Bloqqer.
■ Solvers run on the remaining 276 formulas.
■ Question: is preprocessing always beneficial?
■ Best foot evaluation (virtual experiment): let solvers choose whether to use Bloqqer.
Excerpt of Showcase Applications (+Solving)

- 6 formula sets, 150 formulas each.
- Not in QBFLIB.
- 900s timeout, 7GB memory limit.

**Observation:**
At least one solver is good for one set (but it is not always the same!).
Excerpt of Showcase on Certificates

- Small, but very important showcase:
  - Only one solver and two tool suites submitted.
  - Urgently needed for practical applications.

- Additional experiments with publicly available tools not submitted by their authors.

- Requirements:
  - Need for standard proof formats and checkers.
  - More proof generating solvers.
  - Proof compression techniques.
  - Support from preprocessors.

<table>
<thead>
<tr>
<th>Workflow</th>
<th>eval2012r2 Solved</th>
<th>eval2012r2 Certified</th>
</tr>
</thead>
<tbody>
<tr>
<td>DepQBF and QBFcert</td>
<td>91 (34 s, 57 u)</td>
<td>67 (20 s, 47 u)</td>
</tr>
<tr>
<td>DepQBF and ResQu¹</td>
<td>91 (34 s, 57 u)</td>
<td>63 (22 s, 41 u)</td>
</tr>
</tbody>
</table>

¹Workflow involves proof format conversion.
Summary and Outlook

Lessons Learned

- If this had been a competition, there would not be a clear winner.
- Preprocessing strongly influences solving.
- Preprocessors are powerful (but incomplete) solvers.
- QBF solvers are not blackboxes, some use built-in preprocessing.
- Benchmark selection and scoring methods strongly influence rankings.
- Community-driven organization is challenging, but fruitful.

What’s next?

- More analysis of the available data.
- Establish fair benchmark sets for competitions and evaluations.
- More emphasis on special tracks (formulas needed!)
- Tighter integration of certificate generation.
- Common standards for input formats and testing workflows.

More details in the Poster Session
Thanks to the Contributors !!!!!!!!

- **Solvers**: S. Bayless, A. Goultiaeva, M. Janota, W. Klieber, F. Lonsing, M. Narizzano, A. Van Gelder

- **Preprocessors**: A. Biere, M. Narizzano, M. Seidl, A. Van Gelder

- **Certificates**: V. Balabanov, J.R. Jiang, A. Niemetz, M. Preiner


http://www.kr.tuwien.ac.at/events/qbfgallery2013/