11th International
Satisfiability Modulo Theories
Competition
SMT-COMP 2016

Sylvain Conchon    David Déharbe    Matthias Heizmann
Tjark Weber
The Numbers

- 17 teams participated

**Solvers:**

- Main track: 25 (2 non-competitive)
- Application track: 8 (3 non-competitive)
- Unsat-core track: 1 (4 non-competitive)

**Logics:**

- Main track: 40
- Application track: 14
- Unsat-core track: 40
- Unknown track: 26

**Benchmarks:**

- Main track: 154424
- Application track: 9856
- Unsat-core track: 93241
- Unknown track: 29724

Record numbers of solvers and benchmarks!
1,562,544 job pairs executed (+ some repeats)
Job Pairs by Track

- Main track: 64.2%
- Unknown track: 11.7%
- Unsat-core track: 2.1%
- Application track: 22.0%
StarExec

- All job pairs executed on StarExec
- Timeout: 40 minutes (unknown track: 10 minutes)
- \( \sim 12 \text{ days} \times 100 \text{ nodes} \times 2 \text{ processors/node} \) of compute time

**StarExec worked even better than last year**

- Thanks to Aaron Stump for prompt help when problems or questions arose
- Only very few (and minor) bug reports submitted to the StarExec developers
Machine Specifications

Hardware:
- Intel Xeon CPU E5-2609 @ 2.4 GHz, 10 MB cache
- 2 processors per node, 4 cores per processor
- Main memory capped at 60 GB per job pair

Software (upgraded in 2016):
- Red Hat Enterprise Linux Server release 7.2
- Kernel 3.10.0-327, gcc 4.8.5, glibc 2.17
- Virtual machine image available before the competition
Benchmarks and Logics

- Number of benchmarks in SMT-LIB almost unchanged since 2015
  - Very few new benchmarks
  - Some non-conforming benchmarks were removed

- No new logics

- Thanks to Clark Barrett for curation and uploading
Eligible Benchmarks

All eligible benchmarks were used for the competition. There was no further selection.
Important Rule Changes

- SMT-LIB 2.5 instead of 2.0
  - SMT-LIB not fully migrated yet
  - Fortunately, largely backwards-compatible

- Size-based weighting of benchmark families within divisions:
  
  \[ 1 + \log_e |F| \]

  Small benchmark families are more important than before.

- Unsat-core track reinstated
Competition Tools Improved

- New unsat-core track tools (scrambler and post-processor)

- New scrambling algorithm that makes it harder to identify the original benchmark (cf. yesterday’s talk)
Solvers
... primarily a (non-)termination and complexity bounds prover, but also ...

- SMT-LIB 2 front-end for QF_NIA
- use bit-blasting for binary arithmetic, back-end: MiniSat
- fixed bit-length for unknowns
- bit-length for constants, sums, products etc. as needed
- details on SAT encoding:
  [Fuhs, Giesl, Middeldorp, Schneider-Kamp, Thiemann, Zankl, SAT ’07]
- back-end for proof techniques for termination and complexity bounds, search space & time-out fixed in “tactics”
- approach for SMT-COMP
  - start with small search space
  - if MiniSat says satisfiable: return with model
  - else: retry with larger search space until satisfiable (or out of resources)
OpenSMT2 is an MIT-licensed SMT solver written in C++, Developed at Università della Svizzera Italiana, Switzerland

- By Antti, Leo & Matteo
- Check it out from http://verify.inf.usi.ch/opensmt

Version 2 has been under development since 2012

- Currently supports QF_UF and QF_LRA
- Labeled interpolation on Boolean, QF_UF and QF_LRA with proof compression
- Multicore and cluster/cloud based parallelization
- Provides C and Python API through a library
- Support for incrementality
- Compact size (55 000 LoC)
- Compact representation and efficient memory management for the data types
- An object-oriented design which (hopefully) makes the development of theory support easier
raSAT – an SMT Solver for Polynomial Constraints

Vu Xuan Tung, Mizuhito Ogawa @ JAIST, To Van Khanh @ VNU-UET

- raSAT: ICP + Testing + Intermediate Value Theorem (IVT).
  - Inequality
  - Equality
  - ICP: Interval Constraint Propagation = Interval Arithmetic + Constraint Propagation + Box Decomposition.
  - Testing to boost SAT detection of inequality.
  - Generalized IVT for (non-constructive) SAT detection of equality.

- Sound, but incomplete.
  - Outward rounding (ICP), confirmation by iRRAM (testing)

Download: http://www.jaist.ac.jp/~s1310007/raSAT/, or google “raSAT SMT”
Haniel Barbosa, David Déharbe and Pascal Fontaine
Loria, INRIA, Université de Lorraine (France), ClearSy and UFRN (Brazil)

What is new:

- cleaning, efficiency improvements, e.g. UF (space for improvement)
- (much) improved quantifier handling
- Other w.i.p.: (N|L)RA (Redlog), quantifier handling, proofs

Goals:

- clean, small SMT for UF(N|L)IRA with quantifiers and proofs
- for verification platforms B, TLA+
Selected Results
Results: QF_BV (Main Track)

<table>
<thead>
<tr>
<th>Solver</th>
<th>Error Score</th>
<th>Solved Score (Parallel)</th>
<th>Unsolved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boolector (pre)</td>
<td>0.000</td>
<td>24473.995</td>
<td>149</td>
</tr>
<tr>
<td>Boolector</td>
<td>0.000</td>
<td>24468.395</td>
<td>150</td>
</tr>
<tr>
<td>Minkeyrink</td>
<td>0.000</td>
<td>24434.194</td>
<td>193</td>
</tr>
<tr>
<td>smt-cms-mt</td>
<td>0.000</td>
<td>24244.599</td>
<td>216</td>
</tr>
<tr>
<td>smt-cms-st</td>
<td>0.000</td>
<td>24165.007</td>
<td>214</td>
</tr>
<tr>
<td>CVC4</td>
<td>0.000</td>
<td>23820.707</td>
<td>231</td>
</tr>
<tr>
<td>Z3</td>
<td>0.000</td>
<td>23732.215</td>
<td>304</td>
</tr>
<tr>
<td>smt-cms-exp</td>
<td>0.000</td>
<td>23640.669</td>
<td>270</td>
</tr>
<tr>
<td>ABC_glucose</td>
<td>0.000</td>
<td>23078.931</td>
<td>477</td>
</tr>
<tr>
<td>Yices2</td>
<td>0.000</td>
<td>22687.777</td>
<td>638</td>
</tr>
<tr>
<td>MathSat5</td>
<td>0.000</td>
<td>22496.779</td>
<td>544</td>
</tr>
<tr>
<td>MapleSTP-mt</td>
<td>0.000</td>
<td>22487.264</td>
<td>395</td>
</tr>
<tr>
<td>MapleSTP</td>
<td>0.000</td>
<td>21764.885</td>
<td>450</td>
</tr>
<tr>
<td>smt-minisat-st</td>
<td>0.000</td>
<td>20582.614</td>
<td>1058</td>
</tr>
<tr>
<td>ABC_default</td>
<td>0.000</td>
<td>18528.788</td>
<td>1354</td>
</tr>
<tr>
<td>Q3B</td>
<td>719.723</td>
<td>10397.757</td>
<td>4430</td>
</tr>
</tbody>
</table>
# Results: Competition-Wide Scoring (Main Track)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Solver</th>
<th>Score (sequential)</th>
<th>Score (parallel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CVC4</td>
<td>180.95</td>
<td>181.19</td>
</tr>
<tr>
<td>2</td>
<td>Yices</td>
<td>119.29</td>
<td>119.29</td>
</tr>
<tr>
<td>3</td>
<td>veriT</td>
<td>75.11</td>
<td>75.11</td>
</tr>
<tr>
<td>5</td>
<td>Vampire_parallel</td>
<td>65.36</td>
<td>65.62</td>
</tr>
</tbody>
</table>

**Best newcomer:**

- Vampire_parallel
<table>
<thead>
<tr>
<th>Logic</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANIA</td>
<td>Z3; CVC4</td>
</tr>
<tr>
<td>QF_ANIA</td>
<td>Z3; CVC4</td>
</tr>
<tr>
<td>QF_ALIA</td>
<td>Z3; SMTInterpol; Yices2; MathSat5; CVC4</td>
</tr>
<tr>
<td>QF_UFNIA</td>
<td>Z3; CVC4</td>
</tr>
<tr>
<td>LIA</td>
<td>Z3; CVC4</td>
</tr>
<tr>
<td>ALIA</td>
<td>Z3; CVC4</td>
</tr>
<tr>
<td>QF_UFLRA</td>
<td>Z3; Yices2; SMTInterpol; CVC4; MathSat5</td>
</tr>
<tr>
<td>UFLRA</td>
<td>Z3; CVC4</td>
</tr>
<tr>
<td>QF_UFLIA</td>
<td>Z3; CVC4; Yices2; SMTInterpol; MathSat5</td>
</tr>
<tr>
<td>QF_NIA</td>
<td>CVC4; Z3</td>
</tr>
<tr>
<td>QF_BV</td>
<td>MathSat5; Yices2; smt-cms-st; smt-cms-mt;</td>
</tr>
<tr>
<td></td>
<td>smt-cms-exp; CVC4; MapleSTP; MapleSTP-mt;</td>
</tr>
<tr>
<td></td>
<td>smt-minisat-st; Z3</td>
</tr>
<tr>
<td>QF_LRA</td>
<td>MathSat5; SMTInterpol; Z3; Yices2; CVC4</td>
</tr>
<tr>
<td>QF_LIA</td>
<td>Yices2; Z3; SMTInterpol; MathSat5; CVC4</td>
</tr>
<tr>
<td>QF_AUFLIA</td>
<td>Yices2; Z3; SMTInterpol; MathSat5; CVC4</td>
</tr>
</tbody>
</table>
## Selected Results: Unsat-Core Track

<table>
<thead>
<tr>
<th>Solver</th>
<th>Errors</th>
<th>Reductions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMTInterpol</td>
<td>0</td>
<td>1166535</td>
</tr>
<tr>
<td>toysmt</td>
<td>0</td>
<td>35886</td>
</tr>
<tr>
<td>veriT</td>
<td>26</td>
<td>68811</td>
</tr>
<tr>
<td>MathSat5</td>
<td>190</td>
<td>1527159</td>
</tr>
<tr>
<td>Z3</td>
<td>17079</td>
<td>4597883</td>
</tr>
</tbody>
</table>

- 182,367 job pairs
- In total, 83,450 (45.8%) unsat cores generated
- ... but also 17,097 (9.4%) wrong sat answers
- Each unsat core was checked with three solvers (CVC4, MathSat5 and Z3). 198 cores (2.4%) were found satisfiable by at least one solver.
Selected Results: Unknown Track

Most benchmarks solved:

<table>
<thead>
<tr>
<th>Solver</th>
<th>Benchmarks solved</th>
<th>Benchmarks attempted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yices2</td>
<td>18593</td>
<td>20473</td>
</tr>
<tr>
<td>Minkeyring</td>
<td>16724</td>
<td>17504</td>
</tr>
<tr>
<td>CVC4</td>
<td>16646</td>
<td>29509</td>
</tr>
</tbody>
</table>

In total, 21,542 benchmarks (72.5%) were solved.

However, disagreements on 79 benchmarks!
Further Thoughts

Benchmarks:
▶ Still more benchmarks needed, especially for small divisions
▶ Resolve semantics of partial operations, e.g., bvdiv, fp.min
▶ Benchmark curation deserves better tool support

Competition:
▶ Benchmark weights—good or bad?
▶ Integration of benchmarks with unknown status?
▶ Trophies? (T-shirts? Dinner? Funding?!) 

Teams:
▶ Congratulations on your accomplishments!
▶ Thanks for your participation!