Reactive Synthesis Competition SYNTCOMP 2015

Swen Jacobs Saarland University

Roderick Bloem TU Graz

18 July 2015 – SYNT Workshop

SYNTCOMP: Goals

- Establish **benchmark format**
- Collect **benchmark library**
- Make synthesis tools **comparable**
- Encourage implementation of mature, **push-button tools**
- Improve state of the art through challenging benchmarks

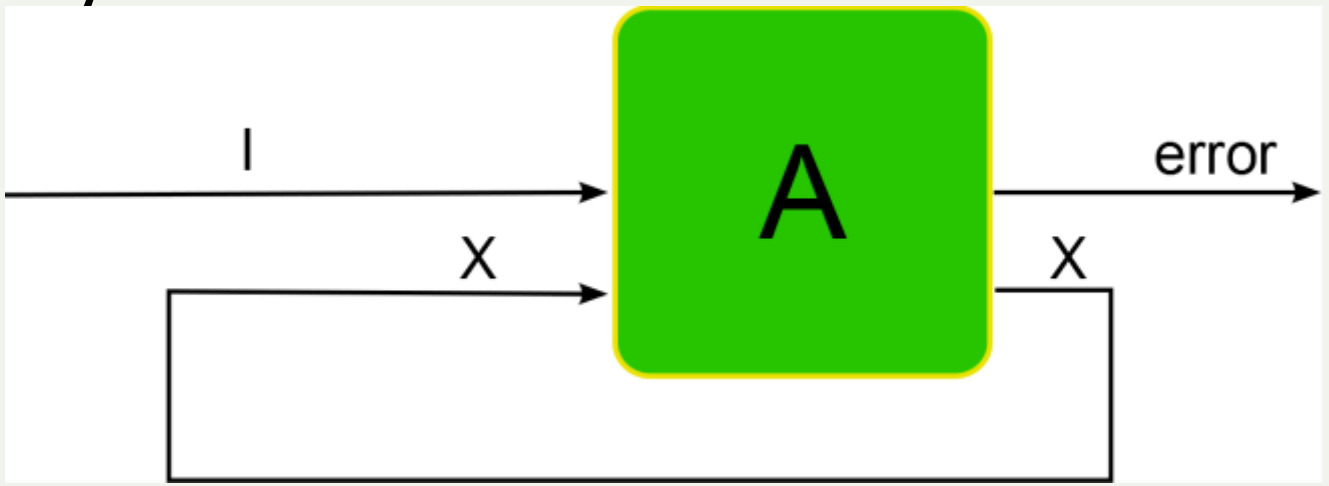
SYNTCOMP: Design Choices

- Low entry-barrier: restrict to safety properties, low-level format
- Re-use existing standards: extend AIGER format
- Synthesis Artifacts are non-trivial:
  - Correctness needs to be checked: use model checkers for verification
  - Output quality is a major issue: needs to be reflected in tool ranking

AIGER Format (for model checking)

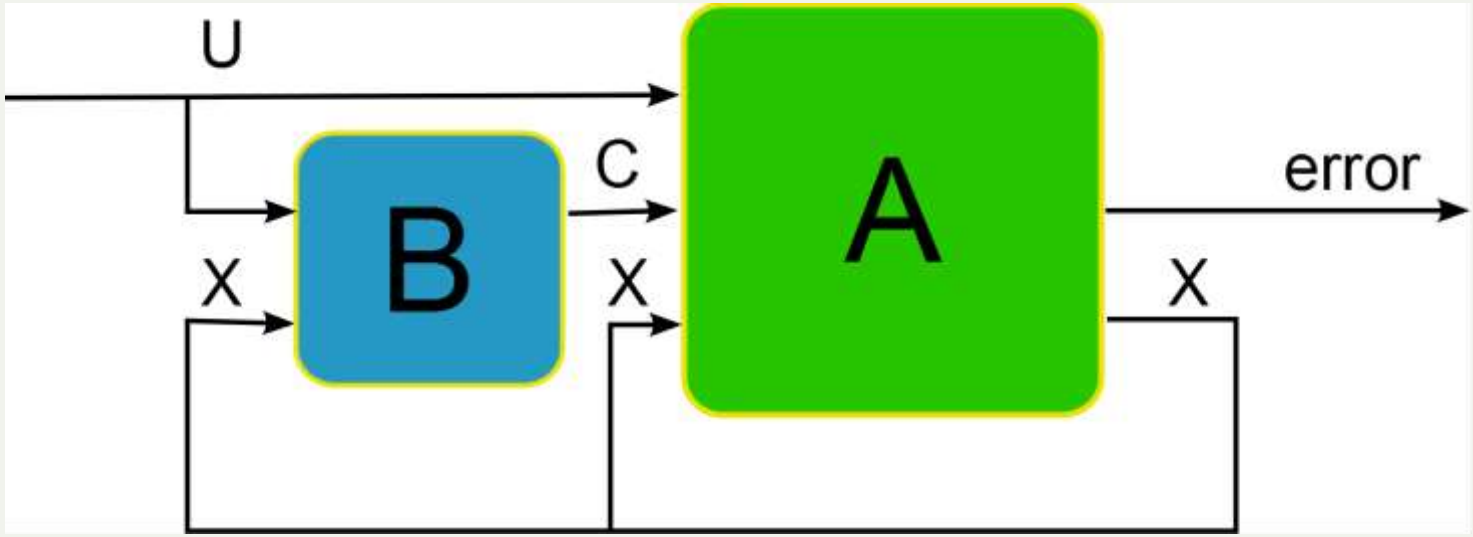
- AIGER format defines system and spec as a circuit A, composed of And-Gates, Inverters, and Latches

- For safety specs, single output is *error*;
  system is correct iff *error* is always false



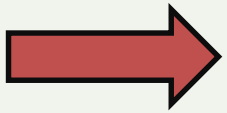
Extended AIGER Format for Synthesis

- For synthesis problems, partition inputs I of system into **controllable** inputs C and **uncontrollable** inputs U
- A solution of synthesis problem is an AIG that includes original AIG A, and adds control structure B for inputs C such that resulting system is correct



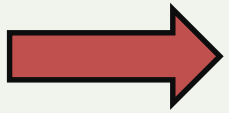
SYNTCOMP 2014: Lessons learned

- 569 benchmarks in 6 benchmark classes
- 5 tools competed in (effectively) 12 configurations
- Separated into Realizability and Synthesis Track, sequential and parallel execution mode
- Realizability Track: fastest tool gets most points (per benchmark)



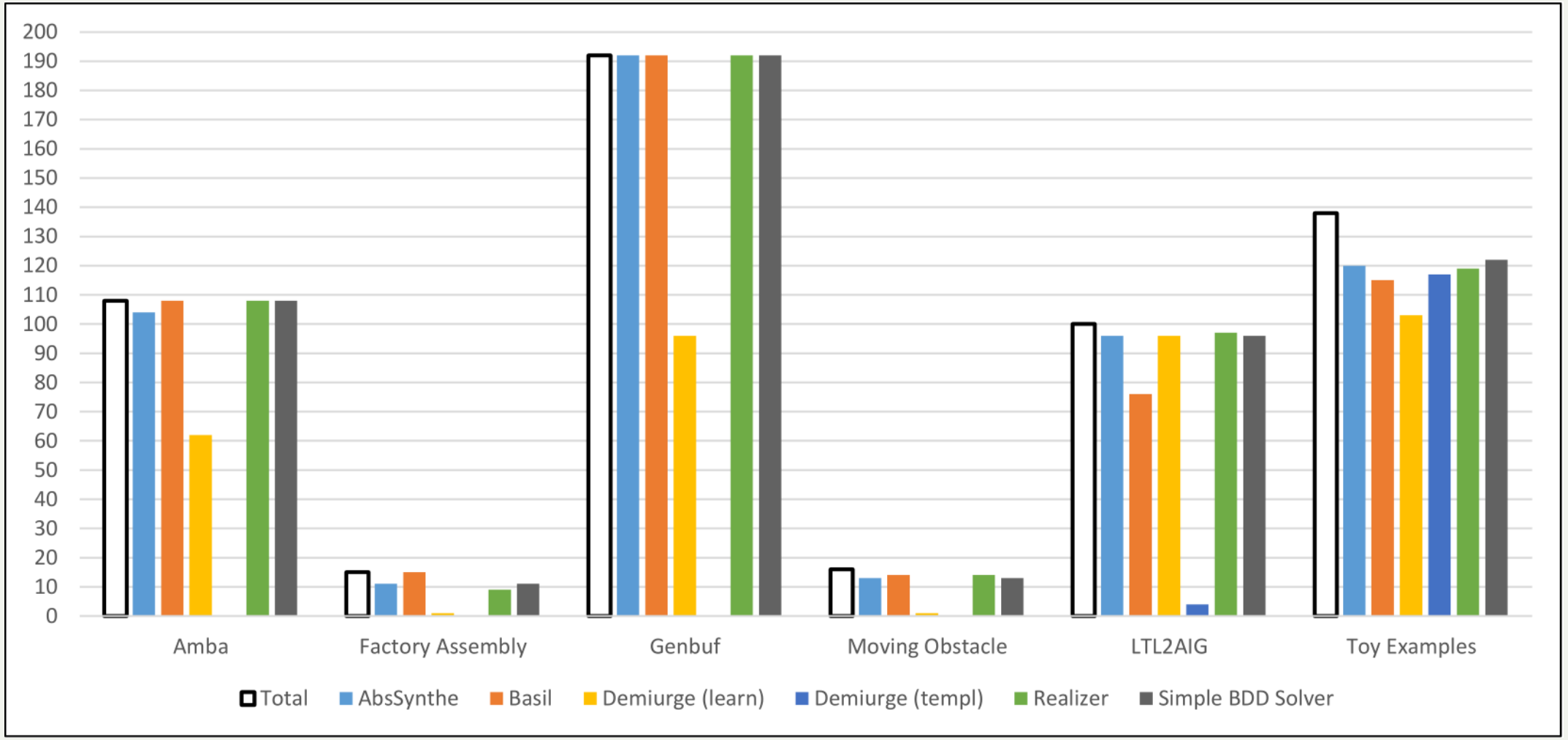
much weight on fast start-up time of tools

- Synthesis Track: tool with smallest solution gets most points



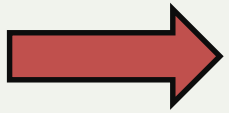
only realizable benchmarks;
no track with “complete” evaluation of synthesis tool

SYNTCOMP 2014: Results by Category (Realizabililty, sequential)



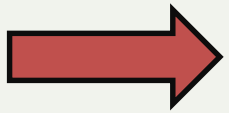
SYNTCOMP 2014: Lessons learned

- Amba and Genbuf benchmarks: most tools solve all benchmarks
- No selection or weighting of instances



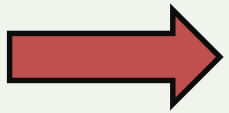
much weight on simple benchmarks and classes with many instances

- Overall, the best approach solves 542 out of 569 instances ($> 95\%$)



overall not very challenging

- Technical issues and time constraints led to a number of problems, incl. additional configurations of tools that did not run in the competition



could have been prevented with better planning,
or solved with more time

SYNTCOMP 2015: Benchmark Collection

New Benchmarks:

- Challenging instances of some classes from 2014 (AMBA, Genbuf, a number of toy examples)
- More LTL2AIG translations of Acacia benchmarks
- Matrix multiplication benchmarks
- Cycle scheduler benchmarks
- Driver synthesis benchmarks
- Controller synthesis for unsafe HWMCC benchmarks
- Huffman encoder
- HyperLTL properties

More than 2000 new benchmark instances

Challenging!

SYNTCOMP 2015: Benchmark Classification

- 2 benchmark classes from 2014 stayed as before:
  Factory Assembly Line, Moving Obstacle
- 4 benchmark classes from 2014 received **new instances**:
  AMBA, Genbuf, Toy Examples, LTL2AIG
- 2 benchmark classes from 2014 were **split into several classes** for 2015:
  Toy Examples, LTL2AIG
- **6 new benchmark classes**:
  Matrix multiplication, Cycle scheduler, Driver synthesis, HWMCC,
  Huffman encoder, HyperLTL properties

Selection of benchmarks: same number of instances per (small or large) class

SYNTCOMP 2015: Weighted benchmark classes

Class	# Benchmarks	Class	# Benchmarks
Amba	16	Moving Obstacle	16
Cycle Scheduler	15	Matrix Multiplication	16
Demo (LTL2AIG)	16	Add (Toy Examples)	8
Driver Synthesis	16	Bitshift (Toy Examples)	8
Factory Assembly Line	15	Count (Toy Examples)	8
Genbuf	16	Genbuf (LTL2AIG)	8
HWMCC	16	Huffman Encoder	5
HyperLTL	15	Mult (Toy Examples)	8
Load Balancer (LTL2AIG)	16	Mv/Mvs (Toy Examples)	8
LTL2DBA/LTL2DPA (LTL2AIG)	16	Stay (Toy Examples)	8

Total: 250 instances

SYNTCOMP 2015: Difficulty Rating

To balance weight on different difficulties, rating takes into account

- Ratio of tools that solved existing benchmark instance in 2014, or
- Ratio of tools (out of 3 best from 2014) that solved new instances in a special *classification run*

Out of every class, select benchmark instances for 2015 with **even distribution over all difficulties**

incl. ~ 20% unsolved instances

Format Extension: SYNTCOMP Tags

Include Meta-Information into benchmark instances (similar to CASC/SMT-COMP):

```
#!SYNTCOMP
STATUS : realizable
SOLVED_BY : 8/8 [SYNTCOMP2014-RealSeq]
SOLVED_IN : 0.008 [SYNTCOMP2014-RealSeq]
REF_SIZE : 203
#.
```

Used for:

- **checking correctness of results**
- **rating of benchmarks and solutions**

SYNTCOMP 2015: Entrants

- **AbsSynthe**: Realizability and Synthesis, 10 configurations
- **Demiurge**: Realizability and Synthesis, 4 configurations
- **Realizer**: Realizability, 2 configurations
- **Simple BDD Solver**: Realizability, 2 configurations
- Hors concours:
  - 2014 versions of AbsSynthe, Demiurge and Simple BDD Solver
  - reference implementation Aisy

Swiss AbsSynthe v1.0

- **Authors**: Romain Brenguier, Ocan Sankur, Guillermo A. Pérez, Jean-François Raskin (*ULB*)
- **Approach**: BDD-based fixpoint computation
- **Implemented in**: C++
- **Uses**: CUDD, AIGER tools
- **New**: compositional approach (and parallel versions)

Demiurge v1.2.0

- **Authors**: Robert Könighofer (*TU Graz*), Martina Seidl (*JKU Linz*)
- **Approach**: different SAT-based game solving approaches
- **Implemented in**: C++
- **Uses**: MiniSAT, Lingeling, DepQBF, Bloqqer, QBFcert
- **Improved**: learning approach (partial quantifier expansion), template-based approach (additional strategy based on SAT and CEGIS)
- **New**: parallel mode with **3 cooperating approaches** (learning, template, incremental induction) that **share information** about winning region

Realizer 2015

- **Author**: Leander Tentrup (*Saarland University*)
- **Approach**: BDD-based fixpoint computation
- **Implemented in**: Python
- **Uses**: CUDD, PyCUDD
- **Improved**: Bug fixes, memory management, parallel version with 2 different strategies

Simple BDD Solver 2015

- **Authors**: Leonid Ryzhyk (*NICTA, CMU*), Adam Walker (*NICTA*)
- **Approach**: BDD-based fixpoint computation
- **Implemented in**: Haskell
- **Uses**: CUDD, Attoparsec
- **Improved**: memory management
- **New**: abstraction-based approach

SYNTCOMP 2015: Rules

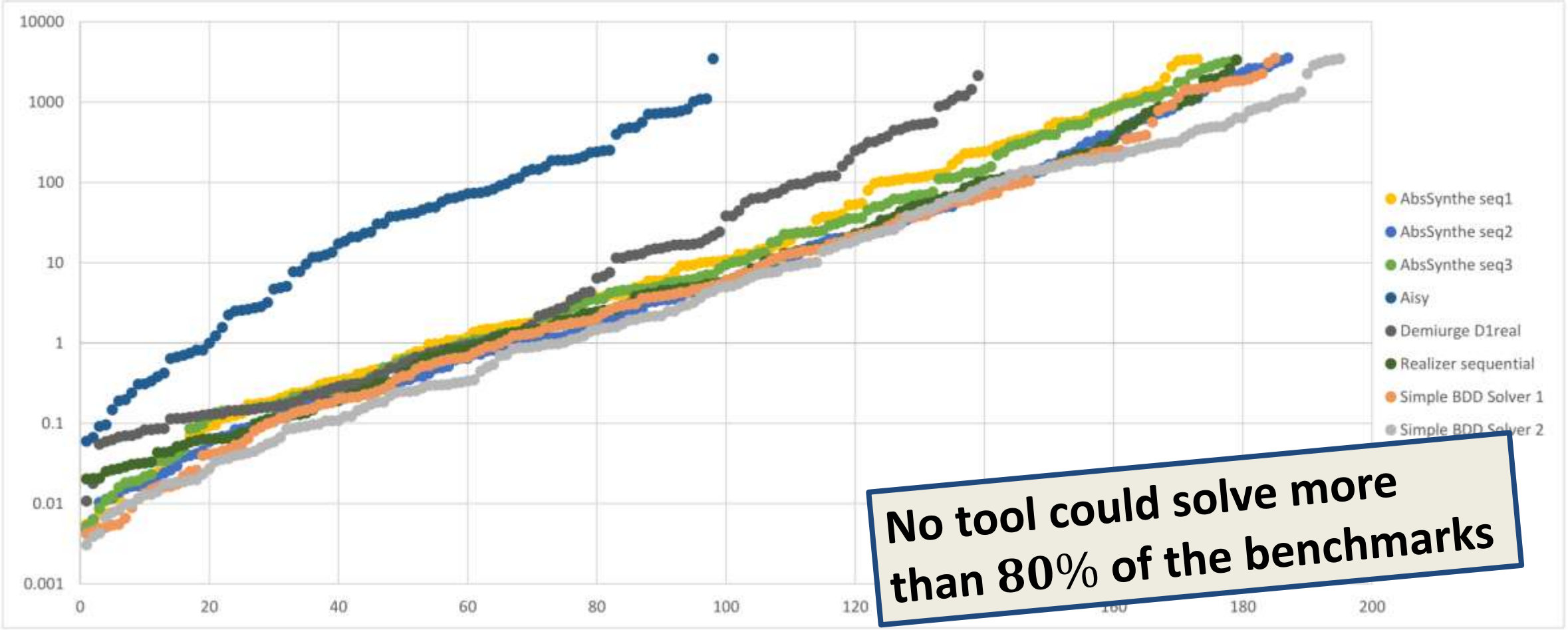
- **Realizability Track**:
  - Determine realizability within time bound
  - Tool with highest number of correct answers wins (incorrect answers are punished, in theory)
- **Synthesis Track**:
  - Return solution or “unrealizable” within time bound
  - Solutions need to be verifiable within separate time bound
  - Tool with highest number of correct answers wins
  - Additional quality ranking: bonus points based on relative size of solution

SYNTCOMP 2015: Execution

- run at Saarland University
- **EDACC** execution & evaluation system
- compute nodes: Quad-Core Intel processors (quad-core, 3.6GHz), 32 GB RAM, 480 GB SSD
- each job runs isolated on one node
- sequential mode: 3600s CPU Time
- parallel mode: 3600s Wall Time
- model checker: iimc (with v3 and ABC as backup)

SYNTCOMP 2015: Results (Realizability)

Sequential mode:



No tool could solve more than 80% of the benchmarks

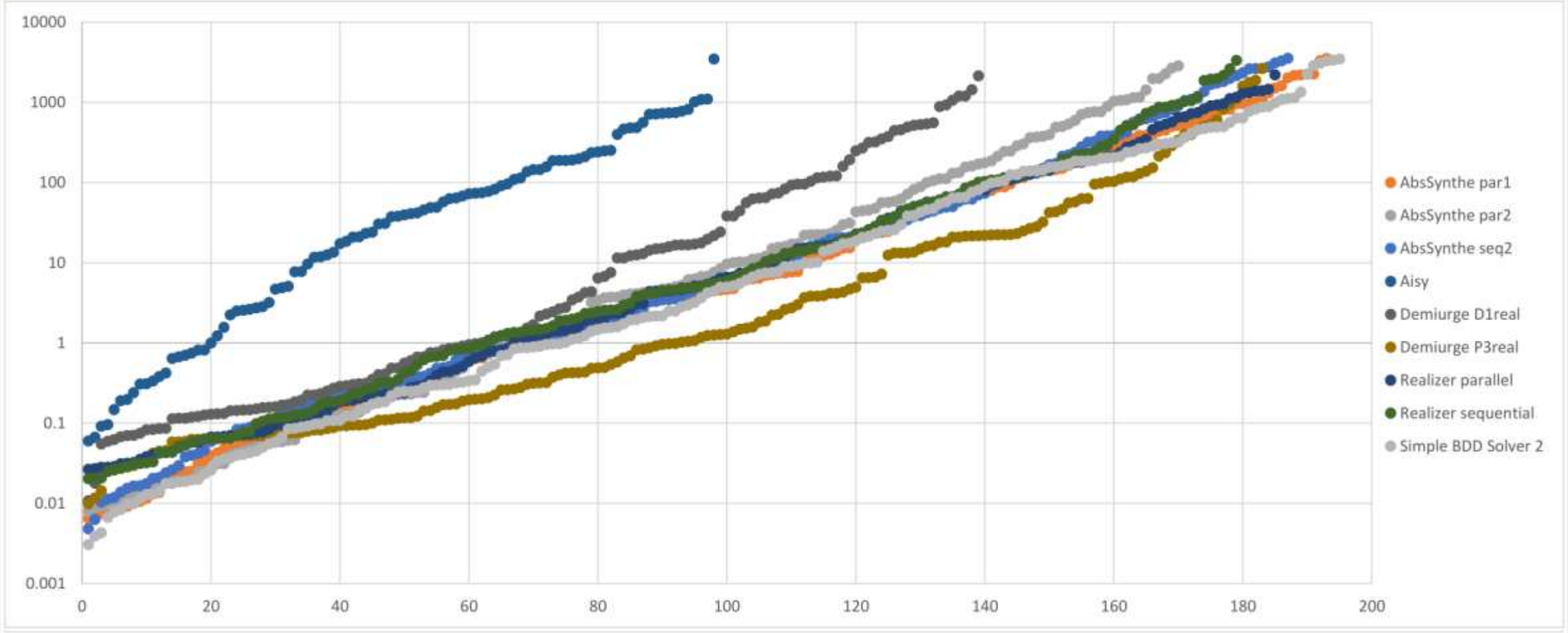
SYNTCOMP 2015: Results (Realizability)

Sequential mode:

Rank	Tool (conf)	Solved	Unique
1	Simple BDD Solver (2)	195	10
2	AbsSynthe (seq2)	187	2
3	Simple BDD Solver (1)	185	
4	AbsSynthe (seq3)	179	
	Realizer (sequential)	179	
6	AbsSynthe (seq1)	173	1
7	Demiurge (D1real)	139	5
	Aisy	98	

SYNTCOMP 2015: Results (Realizability)

Parallel mode (best sequential conf.s for comparison):

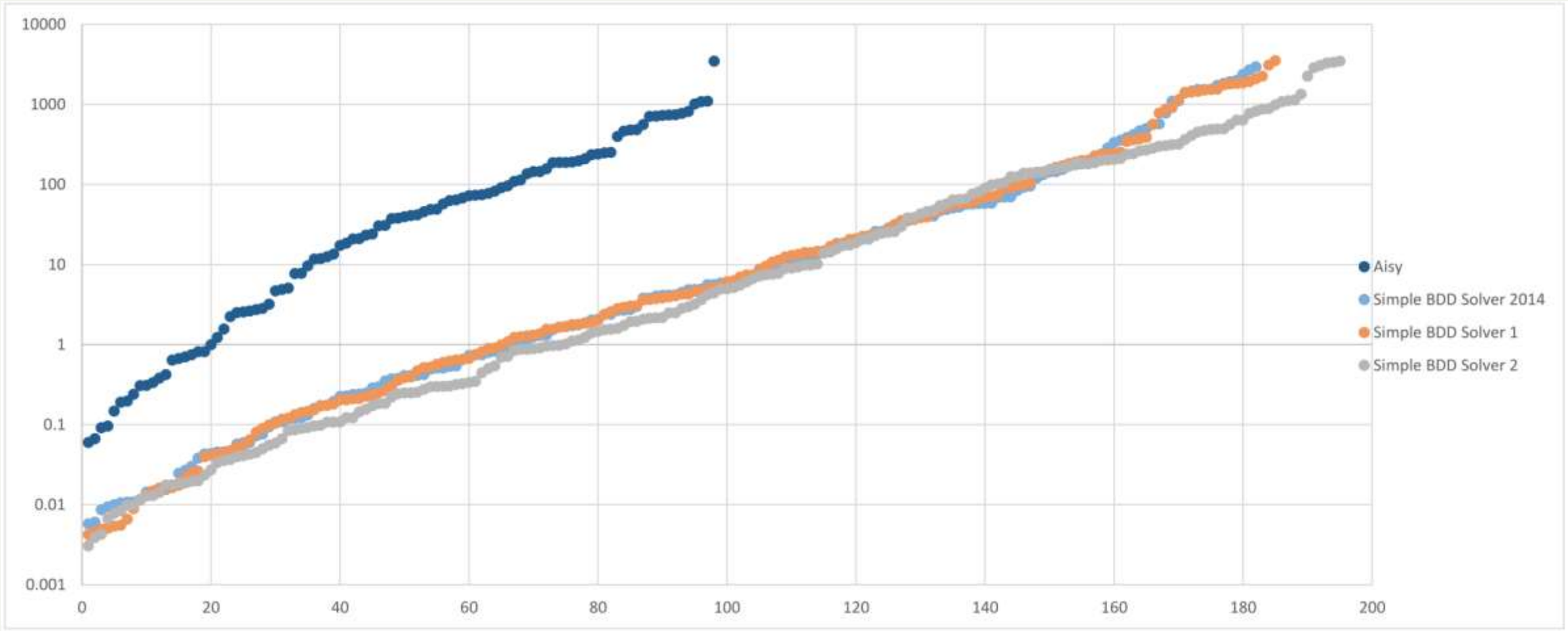


SYNTCOMP 2015: Results (Realizability)

Parallel & sequential modes:

Rank	Tool (conf)	Solved	Unique
1	Simple BDD Solver (2)	195	2
2	**AbsSynthe (par1)**	**193**	
3	AbsSynthe (seq2)	187	
4	Simple BDD Solver (1)	185	
	Realizer (parallel)	**185**	**3**
6	**Demiurge (P3real)**	**183**	**17**
7	AbsSynthe (seq3)	179	
	Realizer (sequential)	179	
9	AbsSynthe (seq1)	173	
10	**AbsSynthe (par2)**	**170**	
11	Demiurge (D1real)	139	
	Aisy	98	

SYNTCOMP 2015: Improvement over 2014 (Realizability)



SYNTCOMP 2015: Synthesis Track

Selection of instances: only those solved in realizability track

Standard ranking:

Which tool can solve most problems?

(in case of realizability, solution must be verifiably correct)

Quality ranking:

- 1 point for detecting unrealizability
- $2 - \log_{10}(\frac{solutionsize}{referencesize})$ points for a (verifiably correct) solution
- Reference size is smallest known implementation from synthesis tool

Entrants: AbsSynthe, Demiurge

SYNTCOMP 2015: Results (Synthesis)

Sequential mode:

Rank	Tool (conf)	Solved	Unique	MC timeout
1	**AbsSynthe (seq_synth2)**	**161**	**4**	**16**
2	**AbsSynthe (seq_synth3)**	**152**	**1**	**16**
3	**AbsSynthe (seq_synth1)**	**148**	**6**	**18**
	AbsSynthe (2014)	145	%	16
4	**Demiurge (D1synt)**	**127**	**8**	**4**
	Demiurge (2014,learn)	83	%	1
	Aisy	75	%	3

No tool could solve more than 80% of the benchmarks

SYNTCOMP 2015: Results (Synthesis)

Sequential mode, Quality ranking:

Rank	Tool (conf)	Solved	Unique	MC timeout	Quality
1	**AbsSynthe (seq_synth2)**	**161**	**4**	**16**	**254**
2	**AbsSynthe (seq_synth3)**	**152**	**1**	**16**	**241**
3	**AbsSynthe (seq_synth1)**	**148**	**6**	**18**	**234**
	AbsSynthe (2014)	145	%	16	231
4	**Demiurge (D1synt)**	**127**	**8**	**4**	**214**
	Demiurge (2014,learn)	83	%	1	138
	Aisy	75	%	3	105

SYNTCOMP 2015: Results (Synthesis)

Parallel mode:

Rank	Tool (conf)	Solved	Unique	MC timeout
1	**Demiurge (P3Synt)**	**180**	**28**	**1**
2	**AbsSynthe (par_synth1)**	**167**	**2**	**20**
3	AbsSynthe (seq_synth2)	161	4	16
4	AbsSynthe (seq_synth3)	152	1	16
5	AbsSynthe (seq_synth1)	148	6	18
	AbsSynthe (par_synth2)	**148**	**0**	**17**
	AbsSynthe (2014)	145	%	16
7	**Demiurge (D1synt)**	**127**	**8**	**4**
	Demiurge (2014,parallel)	88	0	1
	Demiurge (2014,learn)	83	%	1
	Aisy	75	%	3

SYNTCOMP 2015: Results (Synthesis)

Parallel mode,

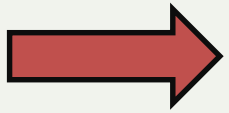
Quality Ranking:

Rank	Tool (conf)	Quality	Solved
1	**Demiurge (P3Synt)**	**317**	**180**
2	**AbsSynthe (par_synth1)**	**263**	**167**
3	AbsSynthe (seq_synth2)	254	161
4	AbsSynthe (seq_synth3)	241	152
5	**AbsSynthe (par_synth2)**	**236**	**148**
6	AbsSynthe (seq_synth1)	235	148
	AbsSynthe (2014)	231	145
7	**Demiurge (D1synt)**	**215**	**127**
	Demiurge (2014,parallel)	144	88
	Demiurge (2014,learn)	138	83
	Aisy	105	75

SYNTCOMP 2015: Results (Synthesis)

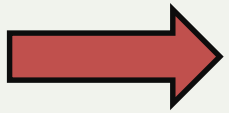
Model Checking Problem:

With more difficult problem instances, also solutions become more difficult to model check



up to 20 solutions per solver that could not be checked

Easy fix (using another model checker) did not work even for the smallest of these solutions



Additional information for model checker?
(winning region as invariant?)

SYNTCOMP 2015: Results

A web frontend of our EDACC system is available online, with detailed data on all experiments for SYNTCOMP 2015:

http://syntcomp.cs.uni-saarland.de/syntcomp2015/experiments/

News and announcements for SYNTCOMP are available on

http://www.syntcomp.org

Conclusions

- Many **new and challenging benchmarks**
- Better selection of benchmarks, better rating system, better execution than last year
- **Basil did not compete, no new tools**
- All other tools competed with **interesting improvements**

SYNTCOMP 2016: New Challenges?

- Encourage real progress, not implementation details:
  Special **challenges**? Specific **classes of benchmarks**?
- Extension of **specification format**: liveness properties, full LTL?
- Extension of **system class**: timed systems?