HWMCC'07

Hardware Model Checking Competition 2007

Armin Biere, Toni Jussila

Institute for Formal Models and Verification
Johannes Kepler University Linz, Austria

CAV'07

Berlin

July 7, 2007

Organization

- Chairs
  - Armin Biere (JKU, Linz, Austria)
  - Toni Jussila (JKU, Linz, Austria)
- Committee
  - Alessandro Cimatti (IRST, Trento, Italy)
  - Koen Lindström Claessen (Chalmers, Gothenburg, Sweden)
  - Ken McMillan (Cadence Berkeley Labs, Berkeley, USA)
  - Fabio Somenzi (University of Colorado, Boulder, USA)

Motivation

- advance model checking technology and research:
  - generate a large set of public available benchmarks
  - encourage researchers to work on novel model checking engines
  - provide a platform for comparison
- repeat success story of SAT competition:
  - exponential improvement of SAT solvers
  - enhances visibility and generates more and more applications
- first things first: synchronous gate level models

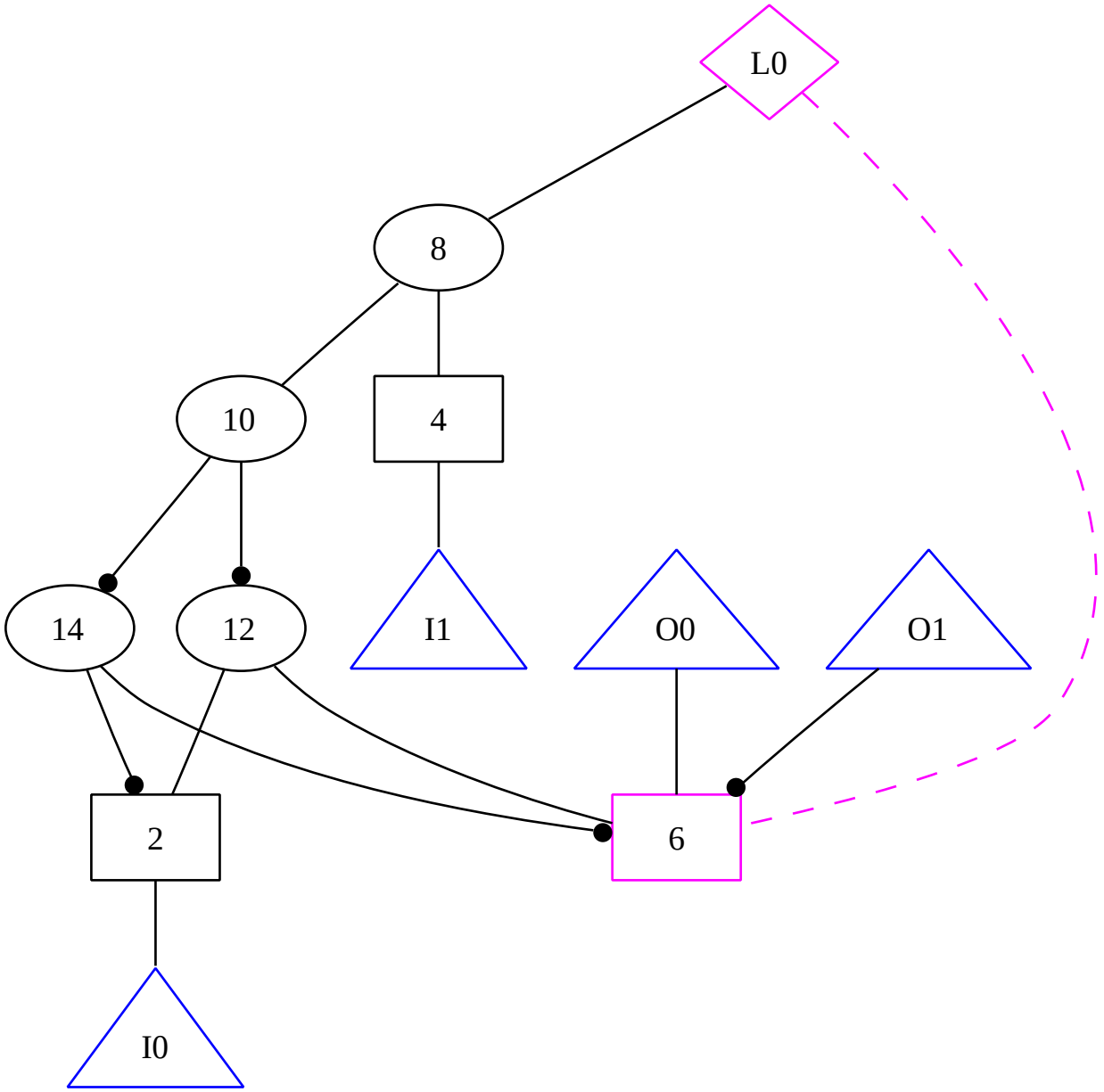
AIGER Input Format

[http://fmv.jku.at/aiger]

- Multi-Rooted Sequential And-Inverter Graphs
  - binary AND gate as single type of operator
  - roots denote outputs or next state functions
  - latches are treated as inputs, initialized to zero

- nodes / inputs / outputs represented by unsigned numbers (literals)
  - boolean constants 0, 1
  - least significant bit (LSB) used as **sign bit**
  - 2 = first variable, 3 = not first variable

D-Flip-Flop with Enable and Reset



D-Flip-Flop with Enable and Reset


```
    M I L O A       MaxVar Inputs Latches Outputs Ands

aag 7 2 1 2 4
2                   input 0         'enable'
4                   input 1         'reset'
6 8                 latch 0         Q next(Q)
6                   output 0        Q
7                   output 1        !Q
8 4 10              AND gate 0      reset & (enable ^ Q)
10 13 15            AND gate 1      enable ^ Q
12 2 6              AND gate 2      enable & Q
14 3 7              AND gate 3      !enable & !Q
```

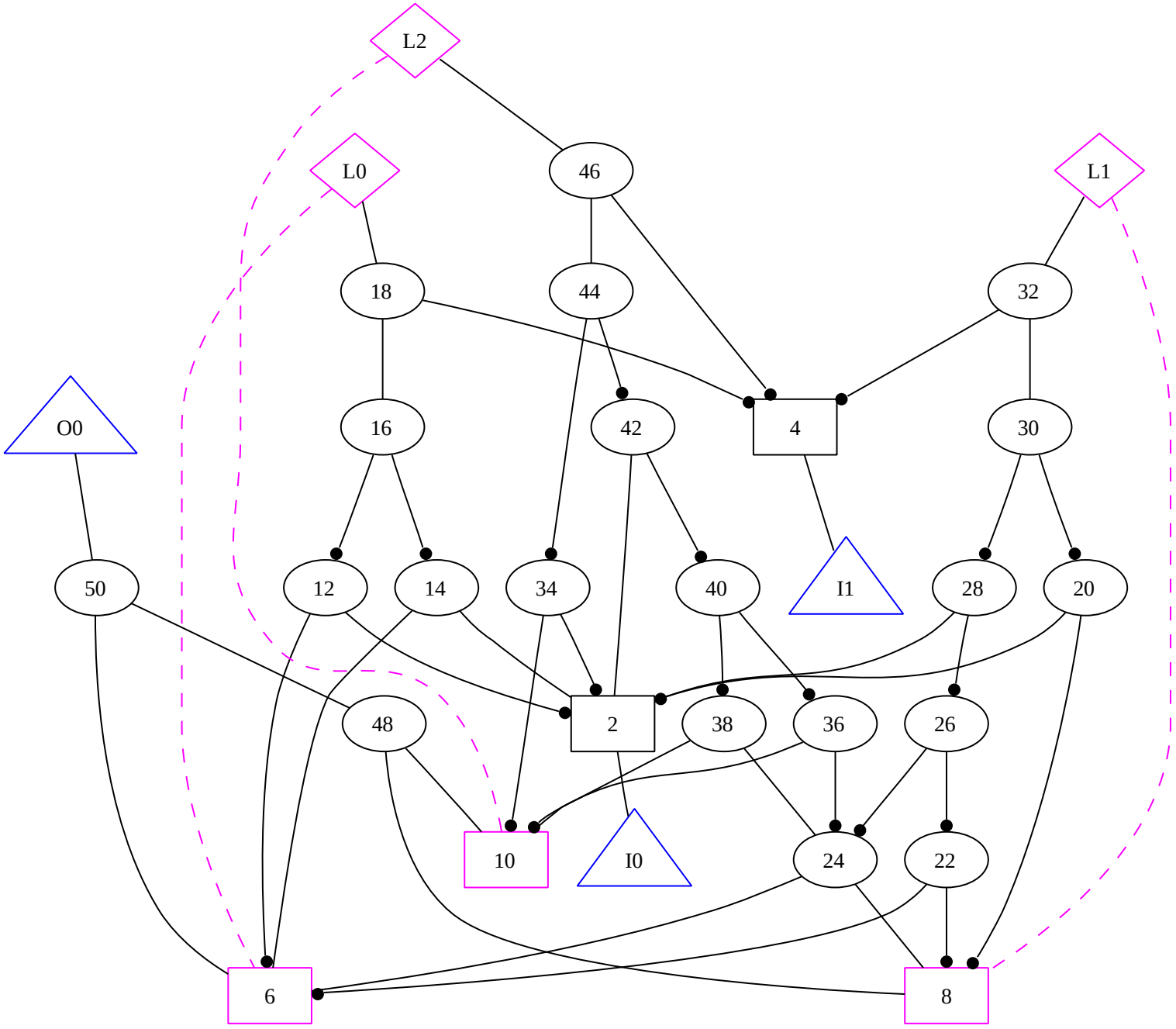
Same Example for Property Checking


```
    M I L O A        MaxVar Inputs Latches Outputs Ands

aag 8 2 1 1 5
2                    input 0          'enable'
4                    input 1          'reset'
6 8                  latch 0          Q next(Q)
16                   output 0         Q & !q
8 4 10               AND gate 0       reset & (enable ^ Q)
10 13 15             AND gate 1       enable ^ Q
12 2 6               AND gate 2       enable & Q
14 3 7               AND gate 3       !enable & !Q
16 6 7               AND gate 4       Q & !Q
```

SAT *iff* output can be set to 1 UNSAT *iff* output can *never* be set to 1

Modulo 8 (3 Bit) Counter with Enable and Reset



More Details on AIGER

- supports combinational and sequential circuits and property checking
- compact binary format (10x – 100x size reduction)
  - reasonably easy to parse
  - multiple parsers available
  - symbol tables
- AIGER is a format , a library and a tool suite
  - smvtoaig, aigtosmv, bliftoaig, aigtoblif, aigdd, aigtodot, aigtocnf, ...
  - MIT style license (with the exception of bliftoaig)
- third party software: ABC, ...

344 Benchmarks in 4 Suites

- L2S (submitted by Viktor Schuppan) — SMV
  - various sources, LTL properties, liveness to safety — 175 instances
- TIP (originally generated by Niklas Eén) — SMV, Verilog, BLIF, ISCAS
  - various sources, all safety — 118 instances
- Intel (submitted by Zurab Khasidashvili) — SMV
  - varying hardness, contain monitors — 42 instances
- AMBA (submitted by Barbara Jobstman, Roderick Bloem) — Verilog
  - automatically synthesized circuits, scalable — 9 instances

344 instances

19 Entrants

	aiger	JKU Linz	BMC with PicoSAT
+	aigtrav	JKU Linz	BDDs + localization
+	ebmc-interpolate	ETH Zürich	
+	ebmc-k-induction	ETH Zürich	
	nusmv-bdd	IRST Trento	BDDs version 2.4.3
	nusmv-bmc	IRST Trento	BMC with MiniSAT 2.0
+	pdtrav-bdd	Politecnico di Torino	
+	pdtrav-cbq	PoliTecnico di Torino	
+	pdtrav-inv	PoliTecnico di Torino	
+	pdtrav-itp	PoliTecnico di Torino	
	smv2qbf	JKU Linz	k-induction with PicoSAT
	smv-cmu	CMU Pittsburgh	BDDs
	smv-bwolen	CMU Pittssburgh	BDDs
	smv-cadence	Cadence Berkeley	BDDs
	tip	Chalmers Gothenburg	k-induction with Satzoo
+	vis-bmc-comp	CU-Boulder	
+	vis-bmc-incomp	CU-Boulder	
+	vis-grab	CU-Boulder	8 already existed
+	vis-puresat	CU-Boulder	11 actually submitted

- 15 node cluster running Ubuntu Linux 7.04
  - Fully Automatic Install (FAI)
  - Sun's Grid Engine (SGE)
- identical nodes with Intel Pentium IV, 3 GHz, 2 GB main memory
- limits enforced by resource sampling `run` utility
  - time limit: 900 seconds (only in final run)
  - space limit: 1.5 GB

Resources

- we started with **small runs**
  - some checkers when killed stalled cluster nodes
  - some checkers had bugs (partially fixed by authors)
  - multiple fixes to *run* and *analysis* scripts
- **final run** took slightly more than 48 hours wall clock time
  - finished Sunday June 24
  - CPU time of 720 hours = 30 days on a single computer
  - used up roughly 106 kWh

Issues

- root disk of our cluster master crashed at the start of the competition
  - Murphy's Law: back up also lost (which is a long story)
  - lost one full person week to get the cluster running again

    so we had a busy June ...

- surprisingly no discrepancies among model checkers (solvers)
  - discrepancy: two solvers return different results
  - very encouraging compared to SAT/QBF/SMT competitions

Solved Instances

	total	solved	SAT	UNSAT
all	344	307	194	113
L2S	175	175	166	9
TIP	118	108	28	80
Intel	42	9	0	9
AMBA	9	8	0	8

Smallest Unsolved Instance

- `cmu.dme1.B.aig`

M	I	L	O	A
379	54	61	1	264

- the Intel suite is the hardest one
  - first four instances are relatively easy
  - only aigtrav and pdtrav-itp occasionally can solve harder ones

- AMBA shows nice scalable behavior

spec	1	2	3	4	5	6	7	9	10
solved by	11	9	7	5	4	2	1	1	0

8 missing, since
$|\text{spec } 8| > |\text{spec } 10|$

1. interpretation: run all solvers in parallel
   - stop as soon one finds a result
   - *this* SOTA solver solves 307 benchmarks

2. interpretation: a SOTA solver solves one instance uniquely (CASC)

	uniquely solved
nusmv-bmc	7
pdtrav-bdd	3
pdtrav-{inv,itp}, aigtrav	2
vis-grab	1
others	0

Ranking

	solved	SAT	UNSAT	timeout	spaceout	signal
SOTA	307	194	113	0	0	0
pdtrav-inv	251	163	88	78	1	14
pdtrav-itp	248	170	78	80	3	13
vis-puresat	236	177	59	101	4	3
tip	222	183	39	97	25	0
vis-bmc-comp	210	173	37	134	0	0
smv2qbf	210	173	37	26	108	0
vis-bmc-incomp	199	183	16	131	14	0
ebmc-k-induction	199	170	29	52	93	0
nusmv-bmc	193	193	0	134	3	14
aiger	182	182	0	65	97	0
ebmc-interpolate	178	155	23	121	45	0
pdtrav-bdd	125	42	83	213	0	6
pdtrav-cbq	118	39	79	212	6	8
smv-cadence	106	33	73	49	171	18
vis-grab	94	19	75	71	0	179
aigtrav	77	22	55	16	251	0
smv-bwolen	51	18	33	42	1	250
nusmv-bdd	47	13	34	35	256	6
smv-cmu	30	12	18	296	0	18

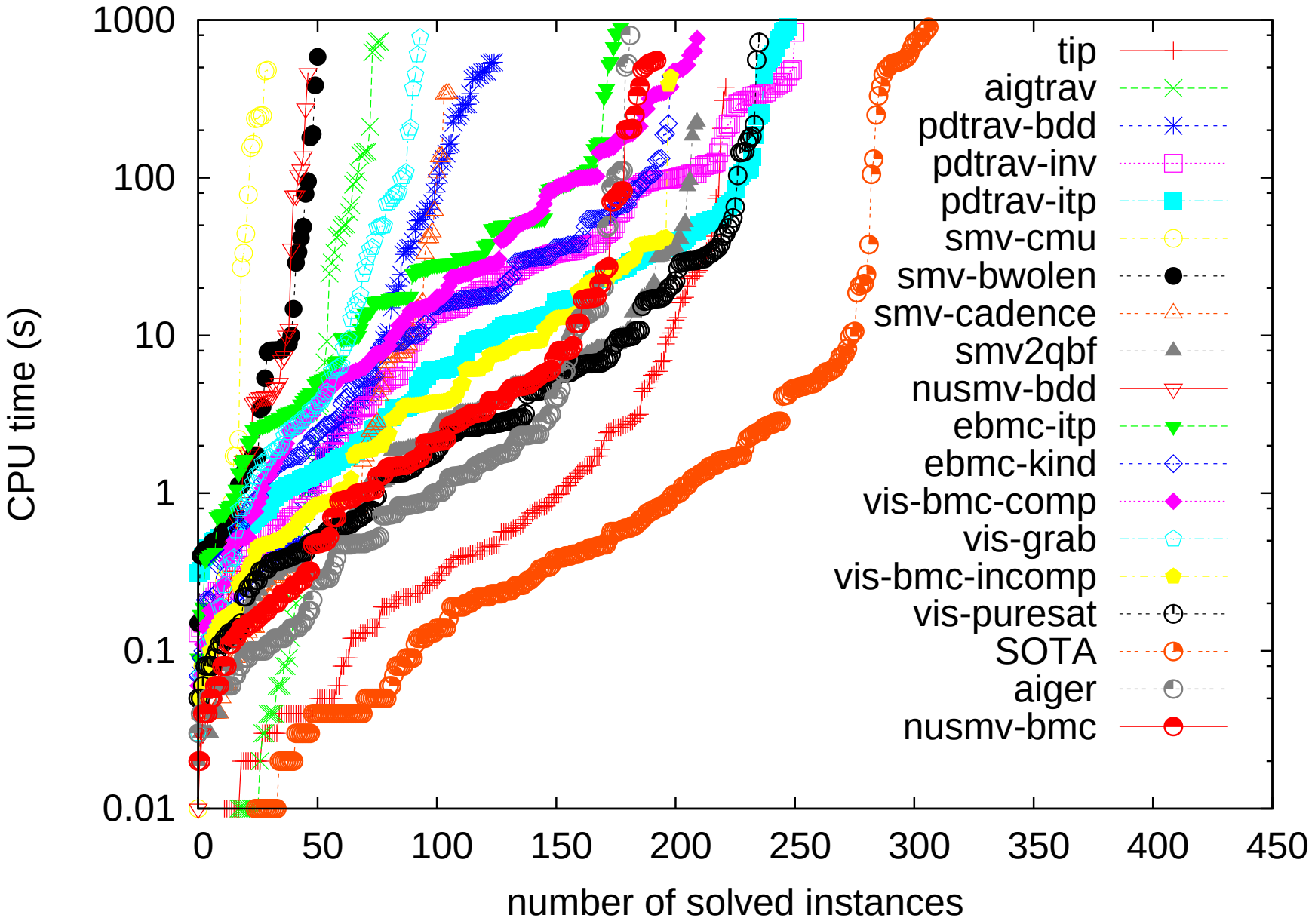
Ranking SAT

	solved	**SAT**	UNSAT	timeout	spaceout	signal
SOTA	307	194	113	0	0	0
nusmv-bmc	193	193	0	134	3	14
tip	222	183	39	97	25	0
vis-bmc-incomp	199	183	16	131	14	0
aiger	182	182	0	65	97	0
vis-puresat	236	177	59	101	4	3
vis-bmc-comp	210	173	37	134	0	0
smv2qbf	210	173	37	26	108	0
pdtrav-itp	248	170	78	80	3	13
ebmc-k-induction	199	170	29	52	93	0
pdtrav-inv	251	163	88	78	1	14
ebmc-interpolate	178	155	23	121	45	0
pdtrav-bdd	125	42	83	213	0	6
pdtrav-cbq	118	39	79	212	6	8
smv-cadence	106	33	73	49	171	18
aigtrav	77	22	55	16	251	0
vis-grab	94	19	75	71	0	179
smv-bwolen	51	18	33	42	1	250
nusmv-bdd	47	13	34	35	256	6
smv-cmu	30	12	18	296	0	18

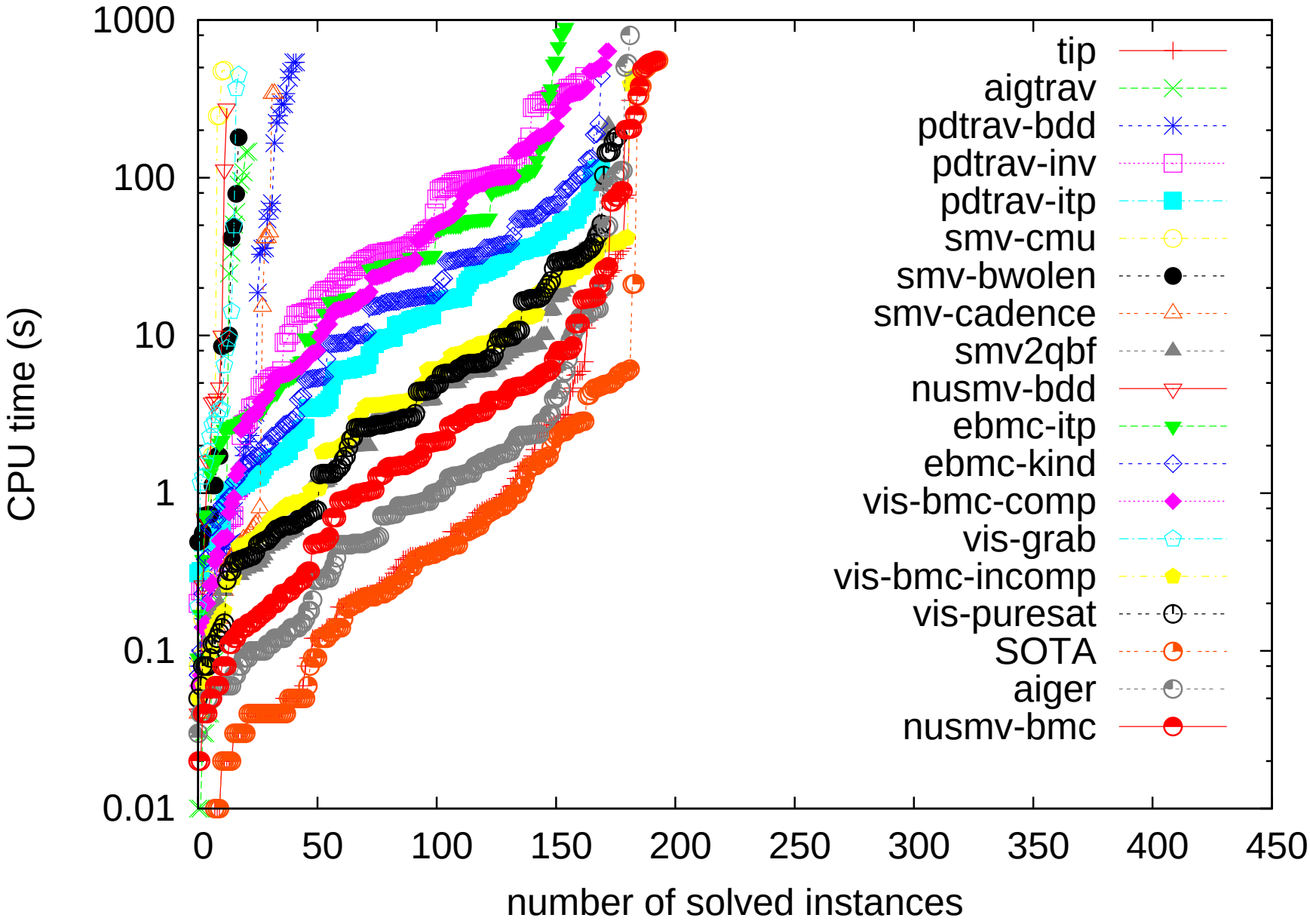
Ranking UNSAT

	solved	SAT	**UNSAT**	timeout	spaceout	signal
SOTA	307	194	113	0	0	0
pdtrav-inv	251	163	88	78	1	14
pdtrav-bdd	125	42	83	213	0	6
pdtrav-cbq	118	39	79	212	6	8
pdtrav-itp	248	170	78	80	3	13
vis-grab	94	19	75	71	0	179
smv-cadence	106	33	73	49	171	18
vis-puresat	236	177	59	101	4	3
aigtrav	77	22	55	16	251	0
tip	222	183	39	97	25	0
vis-bmc-comp	210	173	37	134	0	0
smv2qbf	210	173	37	26	108	0
nusmv-bdd	47	13	34	35	256	6
smv-bwolen	51	18	33	42	1	250
ebmc-k-induction	199	170	29	52	93	0
ebmc-interpolate	178	155	23	121	45	0
smv-cmu	30	12	18	296	0	18
vis-bmc-incomp	199	183	16	131	14	0
nusmv-bmc	193	193	0	134	3	14
aiger	182	182	0	65	97	0

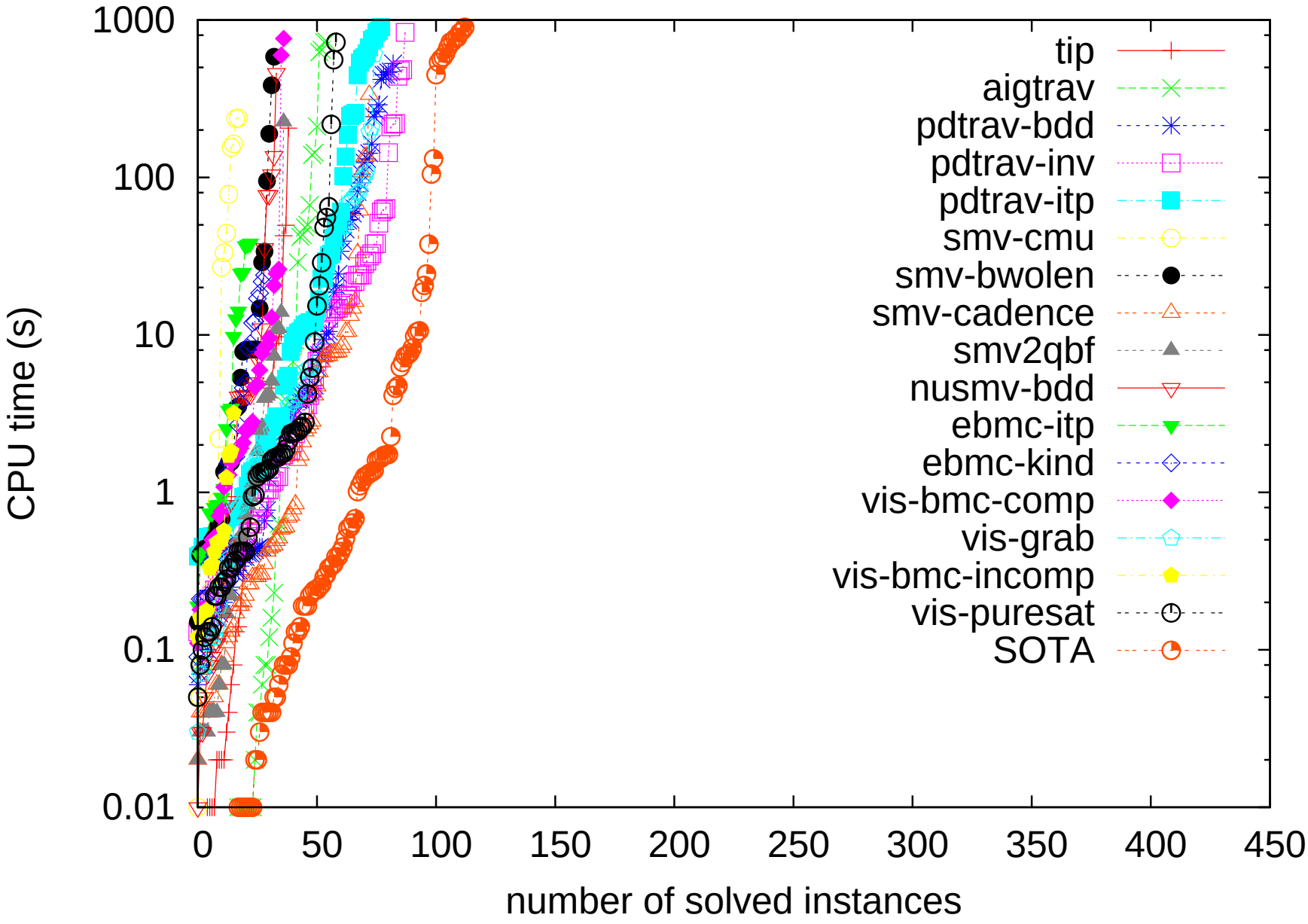
Cactus


1000
100
10
1
0.1
0.01
CPU time (s)
0
50
100
150
200
250
300
350
400
450
number of solved instances
tip
aigtrav
pdtrav-bdd
pdtrav-inv
pdtrav-itp
smv-cmu
smv-bwolen
smv-cadence
smv2qbf
nusmv-bdd
ebmc-itp
ebmc-kind
vis-bmc-comp
vis-grab
vis-bmc-incomp
vis-puresat
SOTA
aiger
nusmv-bmc

SAT Cactus


1000
100
10
1
0.1
0.01
CPU time (s)
0
50
100
150
200
250
300
350
400
450
number of solved instances
tip
aigtrav
pdtrav-bdd
pdtrav-inv
pdtrav-itp
smv-cmu
smv-bwolen
smv-cadence
smv2qbf
nusmv-bdd
ebmc-itp
ebmc-kind
vis-bmc-comp
vis-grab
vis-bmc-incomp
vis-puresat
SOTA
aiger
nusmv-bmc

UNSAT Cactus



- AIGER Format 2.0 (full binary, secondary outputs, QBF)
- OS and I/O conformance
  - ideally: single statically linked binary, temporary files in `/tmp`
  - forbidden: processes, environment assumptions
  - clearly: witnesses / counterexample traces
- two rounds
  - 1st round: weed out problems
  - 2nd round: only one entrant per "group"
- more model checkers and more benchmarks