Hardware Model Checking Competition 2017

Armin Biere, Tom van Dijk, Keijo Heljanko

FMCAD'17

17th International Conference on
Formal Methods in Computer-Aided Design

Vienna, Austria
October 5, 2017

http://fmv.jku.at/hwmcc17

2006	2007	2008	2010	2011	2012	2013	2014	2015	**2017**
AIGER format AVM'06 Ascona	1st HWMCC CAV'07 Berlin	2nd HWMCC CAV'08 Princeton	3rd HWMCC CAV'10 FLOC'10 Edinburgh	4th HWMCC FMCAD'11 Austin	5th HWMCC FMCAD'12 Cambridge	6th HWMCC FMCAD'13 Portland	7th HWMCC CAV'14 FLOC'14 Vienna	8th HWMCC FMCAD'15 Austin	**9th** **HWMCC** **FMCAD'17** **Vienna**
Founding Lunch CAV'06 FLOC'06 Seattle		HWMCC Lunch FMCAD'08 Portland							

- HWMCC'07 – HWMCC'14:

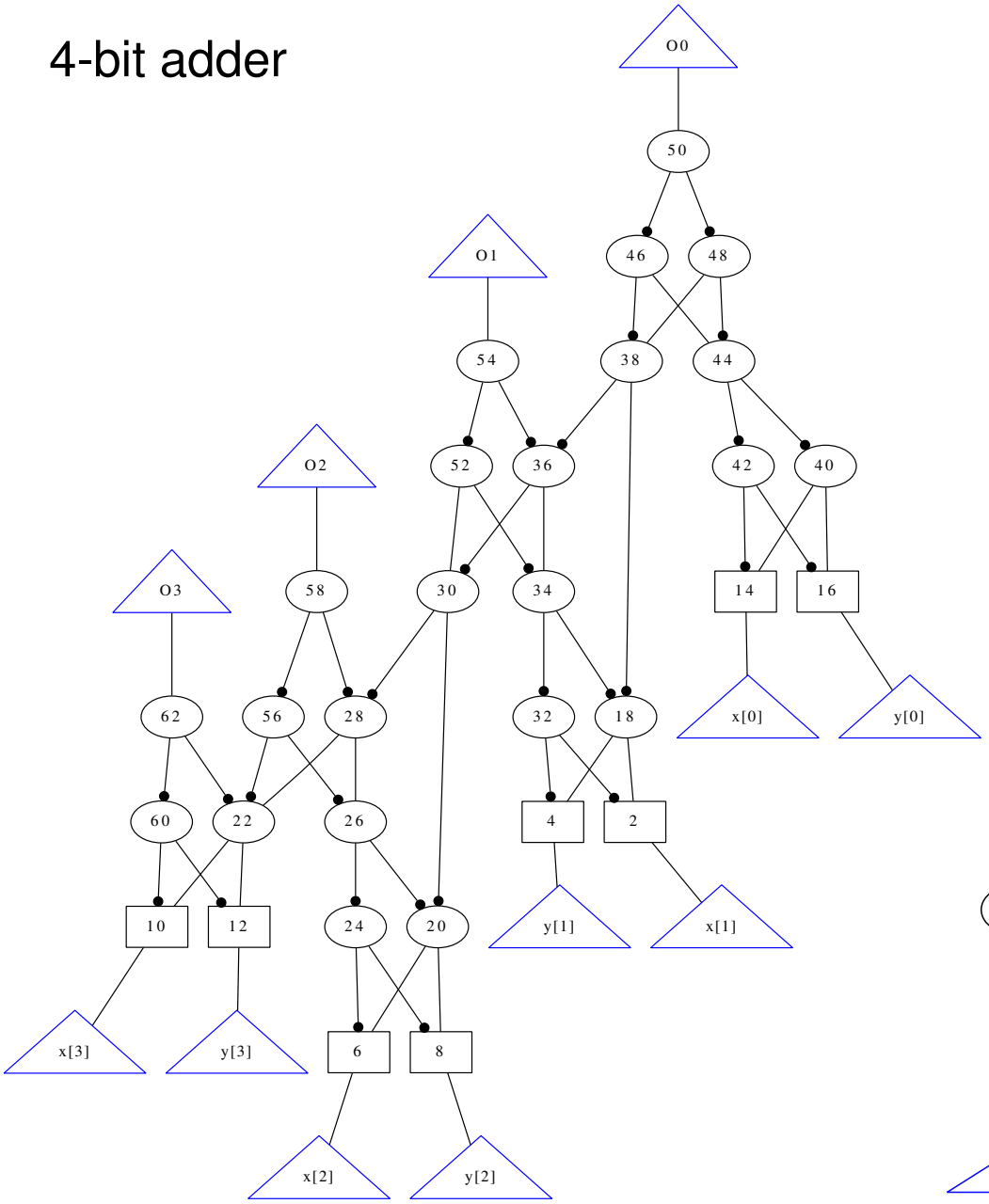
  G. Cabodi, C. Loiacono, M. Palena, P. Pasini, D. Patti, S. Quer, D. Vendraminetto, A. Biere, K. Heljanko. Hardware Model Checking Competition **2014**: An Analysis and Comparison of Model Checkers and Benchmarks. *Journal on Satisfiability, Boolean Modeling and Computation (JSAT)*, **9**, pages 135-172, 2015 (published 2016).

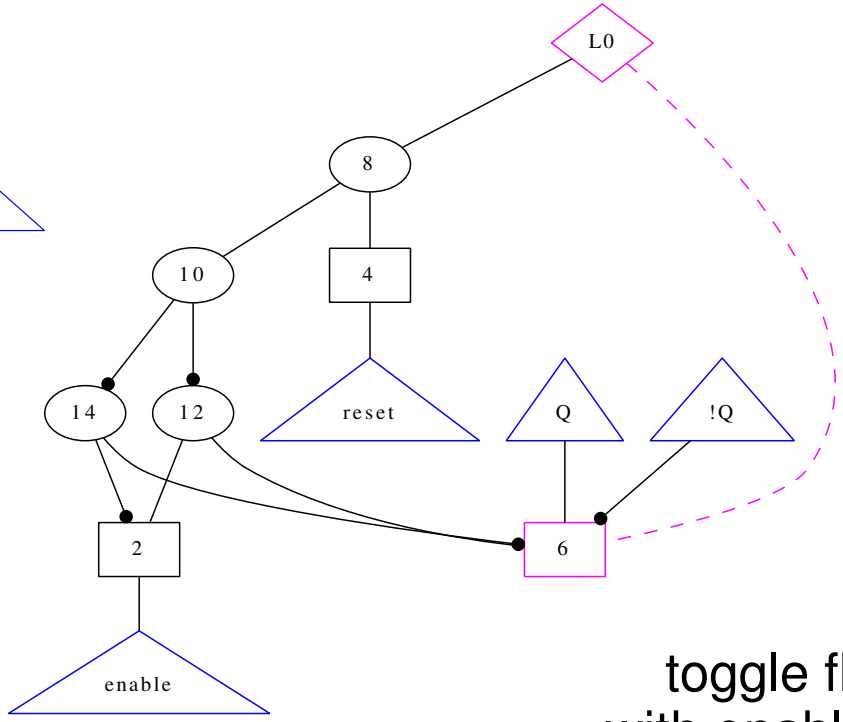
- HWMCC'17 rules did not change (in essence since HWMCC'12)
  - **SINGLE** safety (bad state) property track requires witnesses
  - how **DEEP** model checkers go on unsolved SINGLE instances (Oski award $500)
  - **LIVENESS** track (single "justice" property)
- Intel Xeon E5-2620 v4 2.10GHz, 16 cores, 128 GB main, 1h time limit memory

AIGER

http://fmv.jku.at/aiger

4-bit adder





toggle flip-flop
with enable & reset

Benchmark Selection

- same LIVE benchmark set as in HWMCC'15
  - since only one new model checker (version): abclive
  - single justice benchmarks (negation of liveness properties)
- new SINGLE benchmark set for HWMCC'17
  - singe bad state properties in pre-AIGER-1.9 format
    - encoded as single output, enforced zero initialized latches
  - known benchmarks (from previous competitions HWMCC'07 - HWMCC'15)
    - 100 randomly sampled unsolved benchmarks (out of 169 filtered from 387)
    - 100 randomly sampled solved benchmarks (out of 517 filtered from 1226)
  - new benchmarks (not used in competitions before)
    - 50 newly generated and sampled from 6s suite (Jason Baumgartner, IBM)
    - 50 new benchmarks sampled from:

      P. Subramanyan, Y. Vizel, S. Ray, S. Malik: Template-based Synthesis of Instruction-Level Abstractions for SoC Verification. FMCAD 2015: 160-167
- thus **300** SINGLE and **223** LIVE benchmarks

ABC

- Robert K. Brayton, Baruch Sterin, Alan Mishchenko (Univ. Berkeley)
- abcsuperprove (super_prove)
  - Fine tuning parameters to control timers and engines run in parallel
  - Integrating improved SAT solvers in BMC versions.
  - Improved versions of PDR with internal abstractions.
- abcdeep (super_deep)
  - A new engine running for the DEEP track.
  - Runs several BMC engines in parallel, reporting bounds as they arrive.
  - In addition a PDR engine is used to terminate early on UNSAT examples.
  - The engines used are ABC-ZZ's , `bmc`, ABC's `bmc3` and the new `&bmcs -g`.
  - simplification of model before launching additional copy of `&bmcs -g`.
- abclive (super_live)
  - Fixed several bugs in last year's submission that limited performance.

AVY

- Yakir Vizel (Univ. Princeton), Arie Gurfinkel (Univ. Waterloo)
- 4 configurations
  - one configuration aims at counterexamples – uses a BMC engine
  - 3 configurations are for proofs: including AVY and PDR
- additions in this version:
  - finding inductive clauses during proof construction (push to infinity)
  - applied both to AVY and PDR

Hors Concours

- blimc, aigbmc (Armin Biere, JKU Linz)
- nuxmv (Alberto Griggio et.al., FBK, Trento)
- iimc (Bradley, Somenzi, ... Boulder, Colorado, USA)

iProver

- Konstantin Korovin, Dmitry Tsarkov (University of Manchester)
- general purpose theorem prover for first-order logic
- combines SAT reasoning with first-order instantiations in a model directed way
- AIG problems are translated into the EPR-fragment of first-order logic
- incldues EPR-based BMC, k-induction and CEGAR

pdtrav

- Gianpiero Cabodi et.al. (Politecnico di Torino)
- updates from HWMCC'15 version as follows
- news in ptravthread
  - improved uncovering (reverse engineering) of hidden constraints
  - improved IGR (based on interpolation) engine
- news in ptravdeep
  - multithreaded BMC Tarmo-style
  - distributing different BMC bounds to available threads
  - with speculative assumption of bounds under check

ShiftBMC, Co-NPCheck

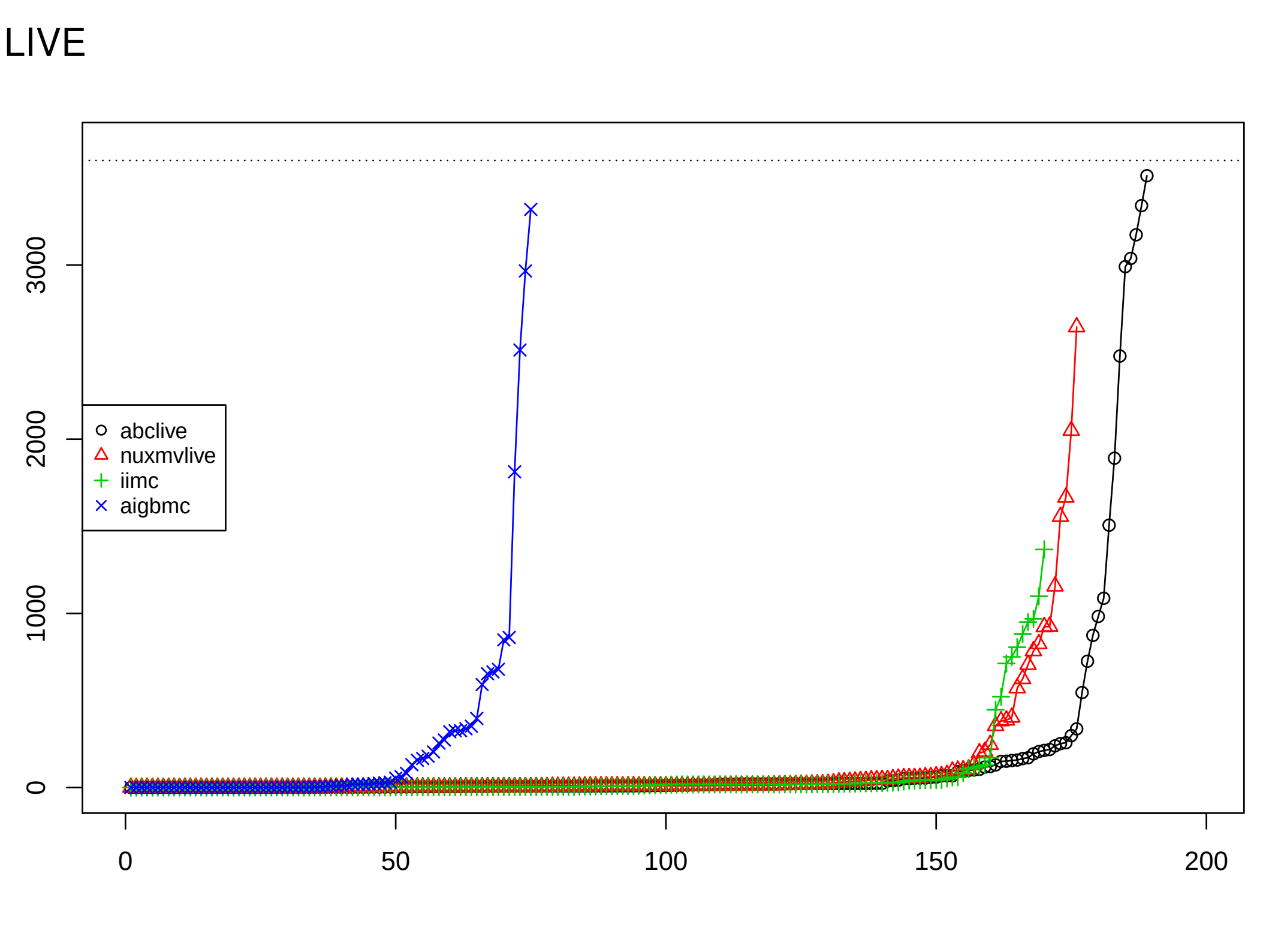
- Norbert Manthey (hobbyist, former Post-Doc @ TU Dresden)
- ShiftBMC
  - idea: aims at fast unrolling, simplifies AIG and CNF to ”shift” before actual BMC
  - same ShiftBMC configuration as in 2015
  - recent riss (https://github.com/nmanthey/riss-solver)
  - with dropped simplification during incremental solving
  - recent ABC
- Co-NPCheck
  - idea: combine incomplete BMC with PDR for unsatisfiable benchmarks
  - python portfolio of ShiftBMC+BIP+BIP.
  - BIPs get simplified aiger from ABC (same as ShiftBMC);
  - BIP(1) uses CLI option ”-abs”
  - BIP(2) uses hacked-in reverse-core-refinement [1] and no extra CLI options
  - many thanks to Baruch Sterin for discussing usage of ABC and BIP

truss

- Ryan Berryhill[1], Alexander Ivrii[2], Neil Veira[1], Andreas Veneris[1].

  [1] University of Toronto, [2] IBM Research Haifa

- Truss: Testing Reachability Using Support Sets
- IC3-based algorithm
- incorporating features of
  - Quip and
  - new recursive blocking procedure
- simple portfolio approach is used consisting
  - Truss, Quip, IC3, and BMC

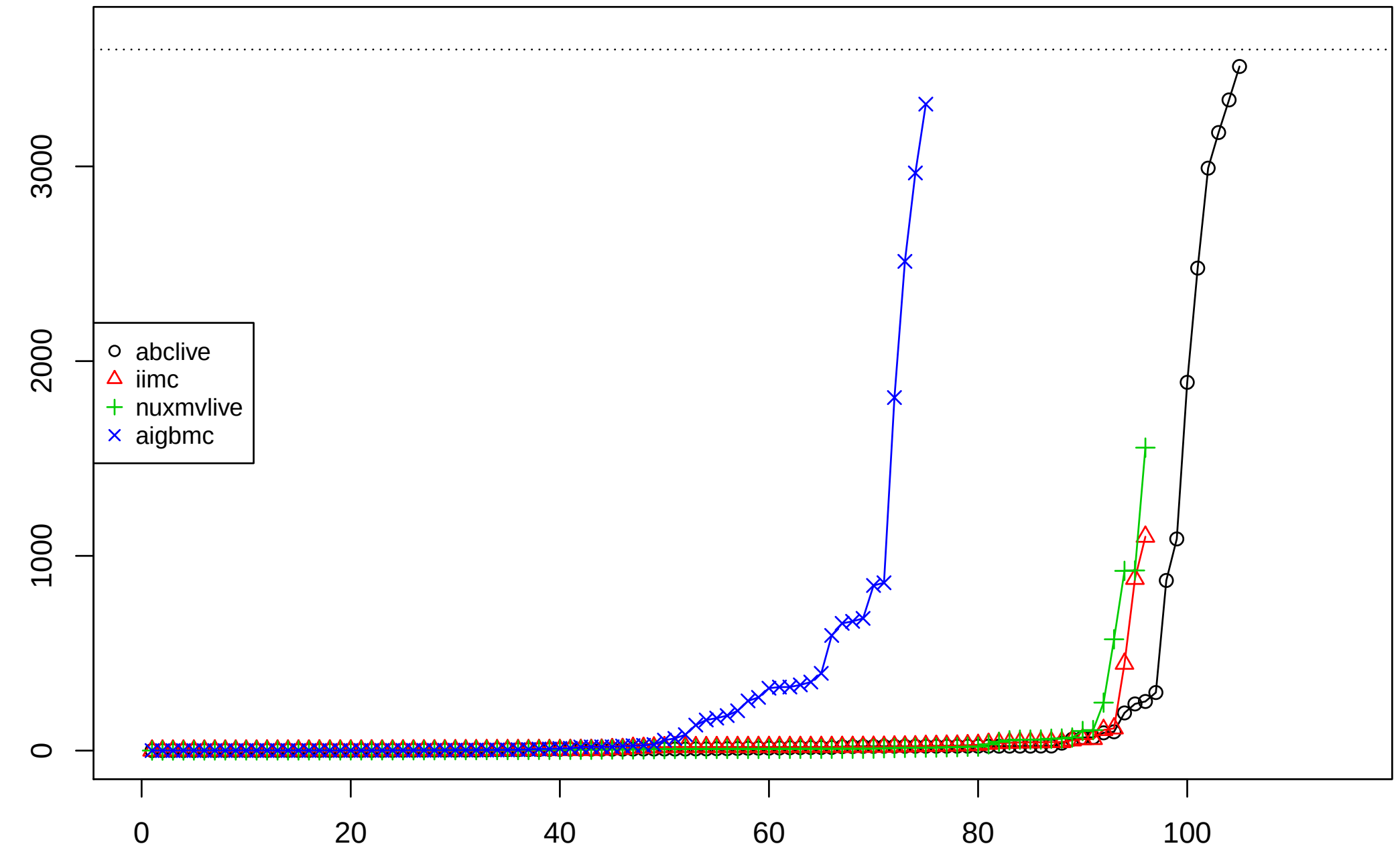

LIVE
abclive
nuxmvlive
iimc
aigbmc
0
1000
2000
3000
0
50
100
150
200

LIVE

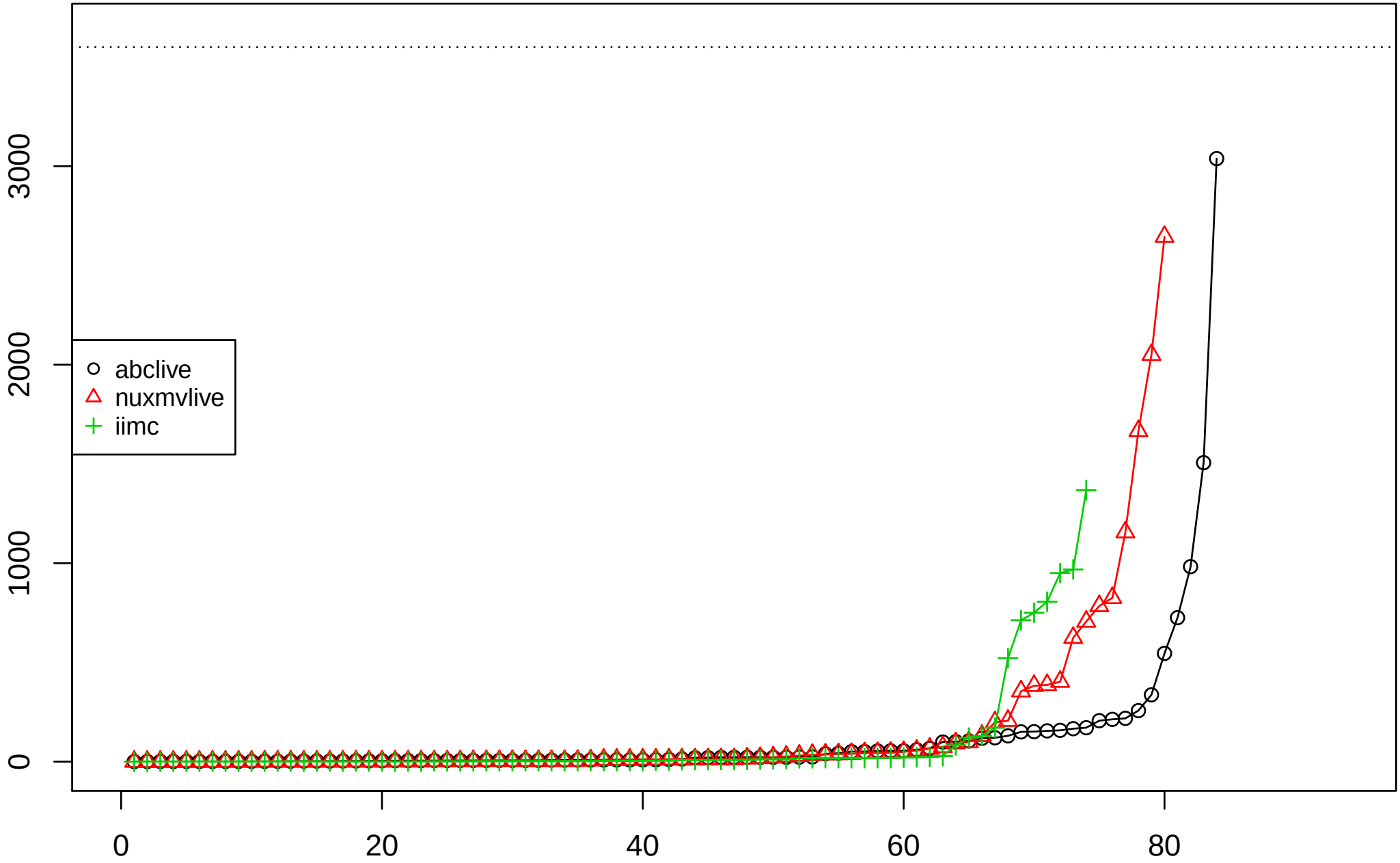
	solved	sat	uns	fld	to	unk	real	time	space	max	best	uniq
1	abclive	189	105	84	34	34	0	32149	470878	433817	14402	62	9
	nuxmvlive	176	96	80	47	47	0	18943	75038	110768	5273	58	0
	iimc	170	96	74	53	53	0	10902	33522	135507	13472	52	1
	aigbmc	75	75	0	148	66	82	18803	18759	13023	4758	20	2

solved = sat + unsat
sat = (justice) property reached (trace exists)
unsat = (justice) property unsatisfiable (trace does not exist)
fld = failed runs
to = number of runs with time out
unk = unknown failure (bound 1000 reached)
real = sum of wall clock time of solved runs
time = sum process time of solved runs
space = sum of used memory in MB
max = maximum memory used in MB
best = number of best runs
uniq = number of uniquely solved runs

LIVE SAT


abclive
iimc
nuxmvlive
aigbmc
0
1000
2000
3000
0
20
40
60
80
100

LIVE UNSAT



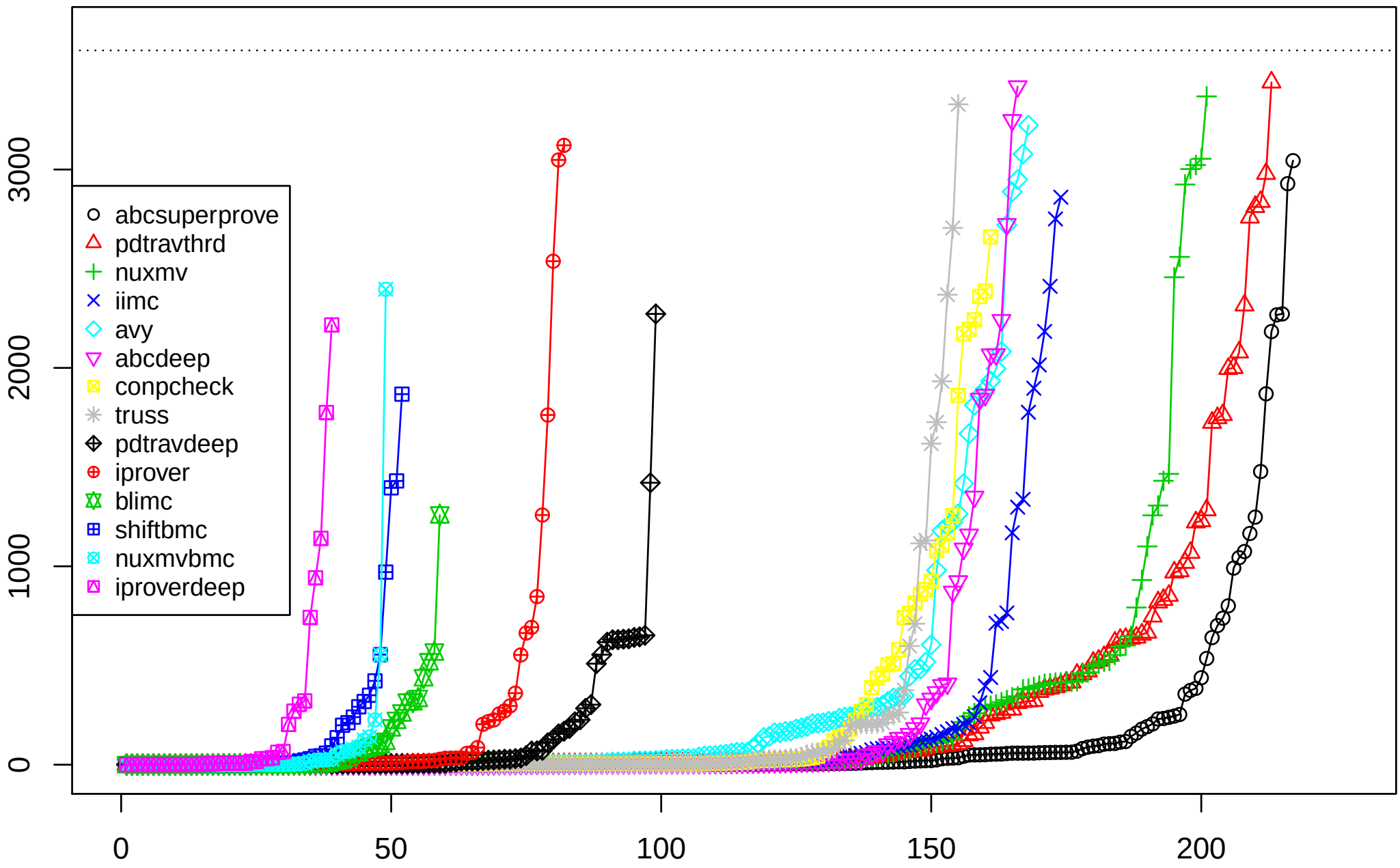
LIVE SAT

		sat	fld	to	unk	real	time	space	max	best	uniq
1	abclive	105	2	2	0	21573	325508	254811	13290	42	7
	iimc	96	11	11	0	3917	12994	42583	3226	19	0
	nuxmvlive	96	11	11	0	5386	21149	51949	5273	26	0
	aigbmc	75	32	6	26	18803	18759	13023	4758	20	2

LIVE UNSAT

		uns	fld	to	unk	real	time	space	max	best	uniq
1	abclive	84	1	1	0	10575	145369	179006	14402	20	2
	nuxmvlive	80	5	5	0	13557	53889	58820	4859	32	0
	iimc	74	11	11	0	6986	20527	92924	13472	33	1

SINGLE

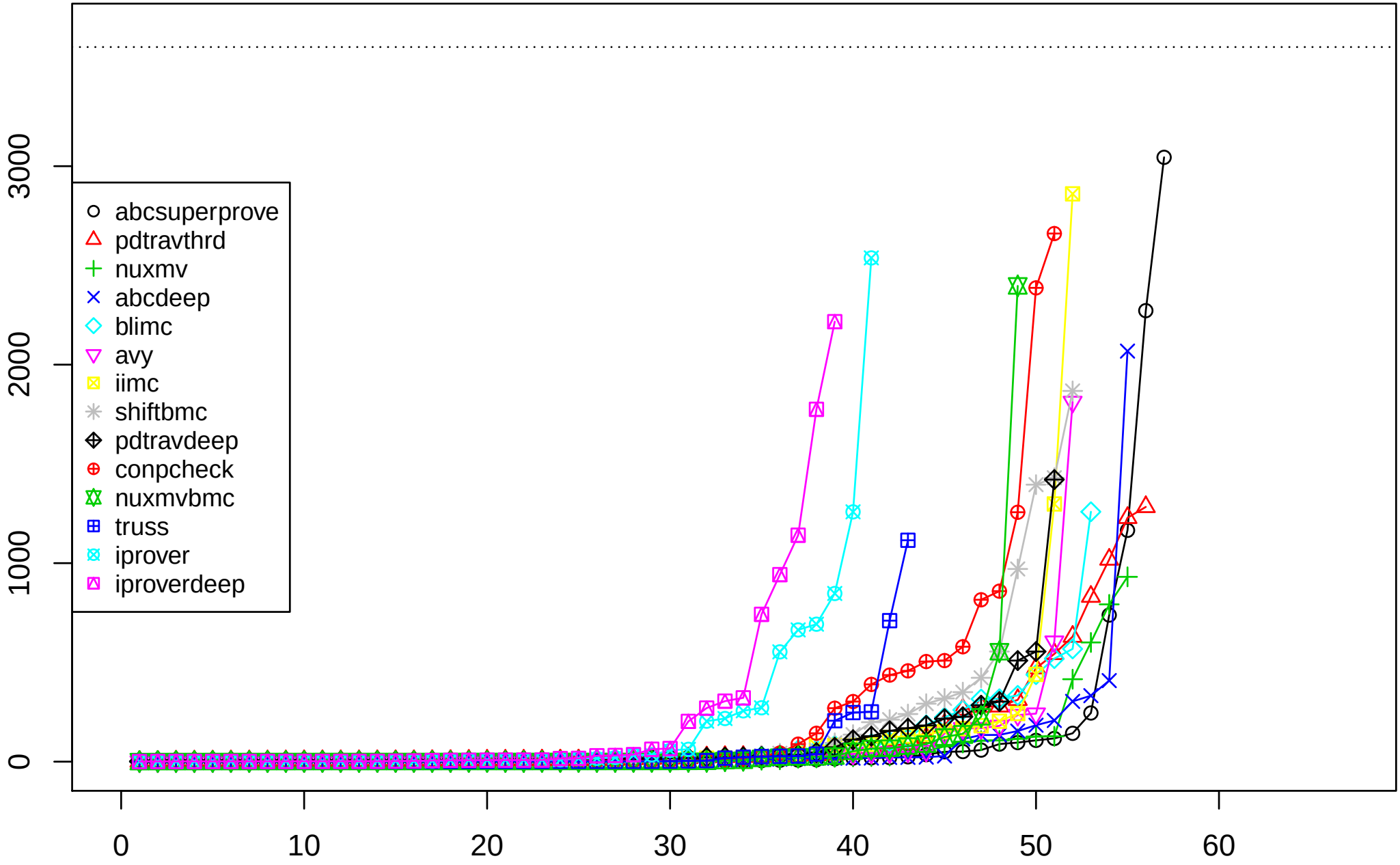

abcsuperprove
pdtravthrd
nuxmv
iimc
avy
abcdeep
conpcheck
truss
pdtravdeep
iprover
blimc
shiftbmc
nuxmvbmc
iproverdeep
0
1000
2000
3000
0
50
100
150
200

SINGLE

		solved	sat	uns	fld	to	mo	s11	s6	unk	real	time	space	max	best	uniq
1	abcsuperprove	217	57	160	83	82	0	0	0	1	31328	272099	337693	66991	97	4
2	pdtravthrd	213	56	157	87	82	0	0	0	5	55645	282927	444468	52858	19	8
	nuxmv	201	55	146	99	99	0	0	0	0	43770	174436	106993	9760	30	1
	iimc	174	52	122	126	125	0	0	1	0	26498	189886	100458	10838	43	3
3	avy	168	52	116	132	131	0	0	0	1	45426	177884	119035	17774	13	3
	abcdeep	166	55	111	134	132	2	0	0	0	28730	217443	228342	45519	39	0
	conpcheck	161	51	110	139	138	0	0	0	1	30633	85855	29351	4104	4	0
	truss	155	43	112	145	144	0	0	1	0	20972	62167	75393	6673	20	2
	pdtravdeep	99	51	48	201	200	0	1	0	0	12365	40523	154541	35092	4	0
	iprover	82	41	41	218	209	0	0	0	9	16932	112234	918713	124463	6	1
	blimc	59	53	6	241	184	0	0	0	57	5183	5150	7424	1904	21	0
	shiftbmc	52	52	0	248	151	0	0	0	97	8764	8730	9776	2417	11	0
	nuxmvbmc	49	49	0	251	161	0	0	0	90	3946	3913	10645	1963	14	0
	iproverdeep	39	39	0	261	238	0	0	0	23	8252	33001	363106	72672	1	0

mo = runs with memory out, s11 = ... segmentation fault, s6 = ... assertion failure

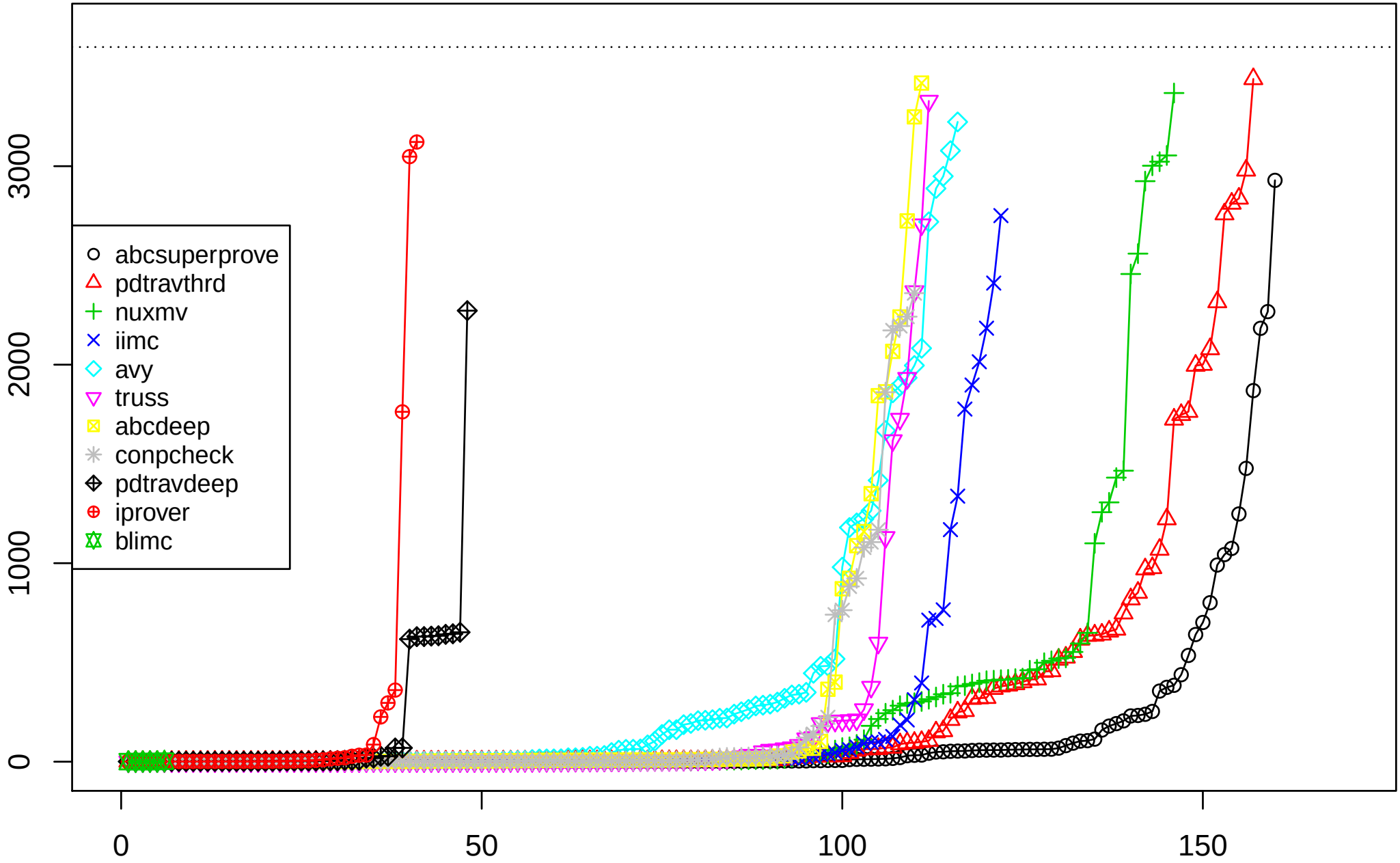
SINGLE SAT


abcsuperprove
pdtravthrd
nuxmv
abcdeep
blimc
avy
iimc
shiftbmc
pdtravdeep
conpcheck
nuxmvbmc
truss
iprover
iproverdeep
0
1000
2000
3000
0
10
20
30
40
50
60

SINGLE SAT

		sat	fld	to	s6	unk	real	time	space	max	best	uniq
1	abcsuperprove	57	4	3	0	1	8378	58143	133439	66991	13	1
2	pdtravthrd	56	5	3	0	2	8101	39659	119279	11071	4	1
	nuxmv	55	6	6	0	0	3788	14960	30122	9760	5	0
-	abcdeep	55	6	6	0	0	4324	27600	56432	16224	6	0
	blimc	53	8	8	0	0	5178	5147	7407	1904	3	0
3	avy	52	9	8	0	1	3996	14229	31882	9683	2	0
	iimc	52	9	8	1	0	6496	44677	39874	8419	5	1
	shiftbmc	52	9	9	0	0	8764	8730	9776	2417	2	0
	pdtravdeep	51	10	10	0	0	4699	12548	102883	22336	3	0
	conpcheck	51	10	9	0	1	11857	35229	12433	3042	0	0
	nuxmvbmc	49	12	12	0	0	3946	3913	10645	1963	11	0
	truss	43	18	18	0	0	2727	10098	21770	4477	6	0
	iprover	41	20	20	0	0	7821	52184	572521	124463	1	0
	iproverdeep	39	22	21	0	1	8252	33001	363106	72672	0	0

SINGLE UNSAT



SINGLE UNSAT

		uns	fld	to	mo	s11	unk	real	time	space	max	best	uniq
1	abcsuperprove	160	21	21	0	0	0	22950	213956	204254	37042	84	3
2	pdtravthrd	157	24	23	0	0	1	47544	243269	325189	52858	15	7
	nuxmv	146	35	35	0	0	0	39981	159476	76871	6923	23	1
	iimc	122	59	59	0	0	0	20002	145209	60584	10838	33	2
3	avy	116	65	65	0	0	0	41431	163655	87153	17774	11	3
	truss	112	69	69	0	0	0	18246	52069	53624	6673	9	2
	abcdeep	111	70	68	2	0	0	24405	189843	171909	45519	1	0
	conpcheck	110	71	71	0	0	0	18776	50626	16918	4104	2	0
	pdtravdeep	48	133	132	0	1	0	7666	27975	51658	35092	1	0
	iprover	41	140	131	0	0	9	9110	60049	346193	113919	2	1
	blimc	6	175	120	0	0	55	6	2	16	9	0	0

DEEP Bound Track

award of $500 sponsored by Oski Technology

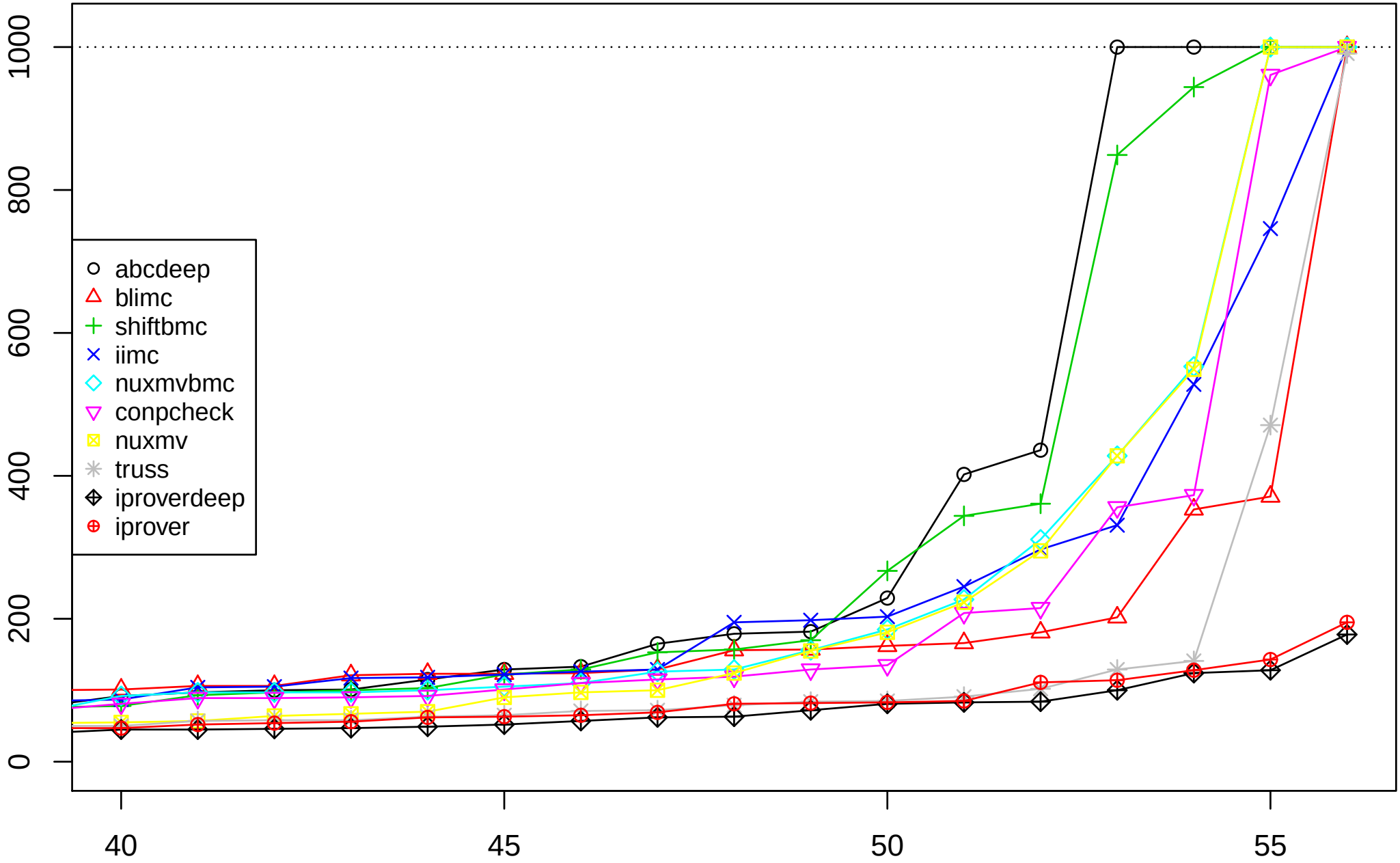
58 unsolved benchmarks in SINGLE (out of 300)

reached bounds capped at 1000 $\Rightarrow$ bound$_i$

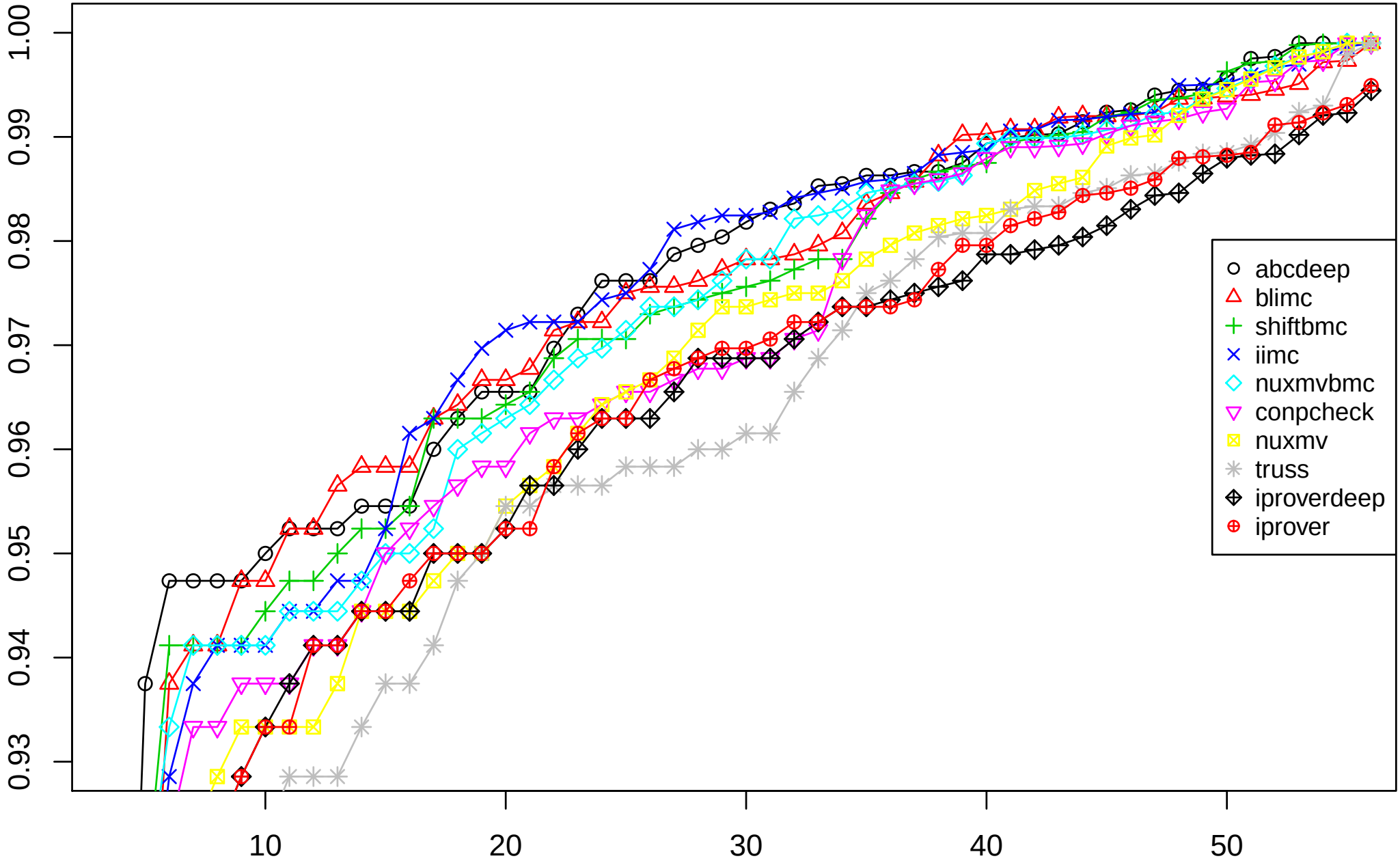
$$\texttt{deep} \quad = \quad \frac{1}{58} \cdot \sum_{i=1}^{58} (1 - 1/(2 + bound_i))$$

$bound_i$	=	-1	contributes	0%
$bound_i$	=	0	contributes	50%
$bound_i$	=	1	contributes	75%
			$\vdots$	
$bound_i$	=	8	contributes	90%
			$\vdots$	
$bound_i$	=	98	contributes	99%
			$\vdots$	

DEEP BOUNDS


1000
800
600
400
200
0
40
45
50
55
abcdeep
blimc
shiftbmc
iimc
nuxmvbmc
conpcheck
nuxmv
truss
iproverdeep
iprover

DEEP SCORES


1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.93
10
20
30
40
50
abcdeep
blimc
shiftbmc
iimc
nuxmvbmc
conpcheck
nuxmv
truss
iproverdeep
iprover

DEEP BOUND TRACK

sponsored by Oski TECHNOLOGY

		to	s6	unk	best	deep
1	abcdeep	58	0	0	32	0.92704
	blimc	56	0	2	18	0.92544
2	shiftbmc	56	0	2	9	0.92490
	iimc	58	0	0	5	0.92388
	nuxmvbmc	56	0	2	3	0.92227
-	conpcheck	58	0	0	2	0.91861
	nuxmv	58	0	0	2	0.91753
3	truss	57	1	0	5	0.91067
	iproverdeep	58	0	0	1	0.90927
	iprover	58	0	0	3	0.90633

used the following 58 unsolved instances in SINGLE track:

6s105 6s158 6s160 6s163 6s185 6s186 6s188 6s191 6s195 6s22 6s267rb3 6s268r 6s274r
6s279r 6s280r 6s29 6s316b421 6s316b460 6s322rb646 6s329rb19 6s341r 6s342rb122 6s365r
6s367r 6s37 6s376r 6s377r 6s399b02 6s42 6s44 6s514r 6s516r 6s517rb0 6s128 6s398b09
beemandrsn6b1 beemlifts3b1 bobpcihm nusmvdme216 oski15a01b00s oski15a07b0s oski15a10b02s
intel012 intel013 intel014 intel016 intel027 intel028 intel032
oc8051gm06iram oc8051gm3bacc oc8051gm43acc oc8051gm49acc oc8051gm63iram
oc8051gm88iram oc8051gma4pc oc8051gmbfpc oc8051gmd7acc

unsat-bound 3620 in 'pdtravdeep/beemadd4b1' >= witness length 15 in 'abcdeep/beemadd4b1'

Conclusion

- negative
  - only 4 new model checkers
  - only one new set of benchmarks (from 2015)
- positive
  - ABC clearly improved (in all tracks)
  - there are some new papers on HWMCC
- future
  - HWMCC'18 @ FLOC
  - regression and testing support!
  - maybe go towards single core track?
  - make word-level track finally happen (BTOR)!
  - new scoring for DEEP track or alternative track?

Thanks to all submitters!