The SAT 2011 competition

Results of phase 1

Matti Järvisalo Daniel Le Berre Olivier Roussel

with the support of CRIL and Intel

SAT 2011 conference, Ann Arbor, 18 June 2011

The team

Organizers

- Matti Jarvisalo
- Daniel Le Berre
- Joao Marques-Silva (MUS/HLMUS)
- Olivier Roussel
- Allen Van Gelder (Certified Unsat)

Judges

- Uwe Egly
- Alexander Nadel
- Ashish Sabharwal
- Moshe Vardi

Computer infrastructure (provided by CRIL)

- 49 2x4 cores Xeon @ 2.66 GHz node cluster with 32GB RAM
  - 3.3 years CPU time used for stage 1 (103 "solvers", 44 submitters)
  - 4.7 years CPU time used for stage 2 (60 "solvers", 36 submitters)
- 1 4x8 cores Xeon @ 2GHz computer with 256GB RAM.
  - 1.7 years CPU time used (4 solvers, 4 submitters)

Policy regarding organizers/judges submissions

Organizers and judges can submit (and win)

- All decisions taken by judges are based on anonymous results
- Benchmarks selection done without knowledge of the competitors
- Olivier Roussel is the only one forced to know the details (to compile and run the solvers)
- md5s of the solvers or their description were available from the competition website on February 6.

Judges and organizers submissions

```
f72932eb16684ee853cfb1f2afe2ba7f sat4j-mus-v20110206.jar (D. Le Berre)
c497f12e2c0bfdd55cb0e90753f11d75 org.sat4j.core.jar (D. Le Berre)
2a44b64d102fd127cb22562d07f4274d muser.bz2 (J. Marques Silva)
a12e32960624463bb1934039fff35cbf solver1.tex (O. Roussel)
ccf3680ef9f3e2909f0a9b87d026e6b3 solver2.tex (O. Roussel)
645f4c8dda7608af929225b163078de1 solver1-minisathack.pdf (A. Sabharwal)
7def65d9245316d8f25505970498d88d solver2-maintrack.pdf (A. Sabharwal et al)
```

Main track

- Two stage process :
  - Run all the solvers on registered categories with medium timeout (1200s)
  - Run promising solvers with extended timeout (5000s) for the award
- Test phase just for sanity check (I/O conformance)
- Source code required for the competition
- Binary only solvers accepted for demonstration

CPU and wall clock rankings

CPU based ranking Measure computational effort, reward solvers using efficiently the resources.

- Solvers are given TIMEOUT seconds of CPU time to solve the instance.
- Any answer given after more than TIMEOUT seconds of CPU time is ignored.

Wall Clock based ranking Measure user's perception of the solver efficiency, rewards fast solvers regardless of resources consumption.

- Solvers are given TIMEOUT seconds of WC time to solve the instance.
- Any answer given after more than TIMEOUT seconds of WC time is ignored.

WC vs CPU time remarks

Expected results for WC ranking

Rank	Solver	Number of solved instances	Detail	% of all instances	% of VBS	Cumulated CPU time on solved instances	Average CPU time per solved instance	Cumulated WC time on solved instances	Average WC time per solved instance	CPU/WC
Total number of instances in the category: 300										
Virtual Best Solver (VBS)		241	129 UNSAT, 112 SAT	80%	100%	65021.06	269.80	-	-	-
1	solver-1848 (complete, team 222)	192	96 SAT, 96 UNSAT	64%	80%	109844.85	572.11	27875.01	145.18	3.94
2	solver-1844 (complete, team 260)	192	87 SAT, 105 UNSAT	64%	80%	131387.77	684.31	33587.71	174.94	3.91
3	solver-1586 (complete, team 260)	191	88 SAT, 103 UNSAT	64%	79%	132983.21	696.25	33820.23	177.07	3.93
4	solver-1605 (complete, team 222)	190	95 SAT, 95 UNSAT	63%	79%	109936.90	578.62	27912.74	146.91	3.94
5	solver-1612 (complete, team 211)	183	86 SAT, 97 UNSAT	61%	76%	122546.35	669.65	31289.47	170.98	3.92
6	solver-1588 (complete, team 260)	181	77 SAT, 104 UNSAT	60%	75%	117630.79	649.89	30300.60	167.41	3.88
7	solver-1853 (complete, team 237)	181	89 SAT, 92 UNSAT	60%	75%	119900.30	662.43	30520.98	168.62	3.93
8	solver-1846 (complete, team 260)	180	78 SAT, 102 UNSAT	60%	75%	111419.81	619.00	28922.57	160.68	3.85
9	solver-1860 (complete, team 220)	179	85 SAT, 94 UNSAT	60%	74%	120611.14	673.81	42756.99	238.87	2.82
10	solver-1683 (complete, team 237)	178	89 SAT, 89 UNSAT	59%	74%	96265.08	540.82	24546.73	137.90	3.92
11	solver-1706 (complete, team 220)	175	86 SAT, 89 UNSAT	58%	73%	118286.46	675.92	42570.37	243.26	2.78
12	solver-1651 (complete, team 254)	174	82 SAT, 92 UNSAT	58%	72%	32565.62	187.16	32569.63	187.18	1.00
13	solver-1845 (complete, team 260)	172	77 SAT, 95 UNSAT	57%	71%	32304.49	187.82	32319.98	187.91	1.00
14	solver-1587 (complete, team 260)	172	77 SAT, 95 UNSAT	57%	71%	33519.81	194.88	33521.71	194.89	1.00
15	solver-1606 (complete, team 222)	171	79 SAT, 92 UNSAT	57%	71%	28545.15	166.93	28535.25	166.87	1.00
16	solver-1849 (complete, team 222)	171	79 SAT, 92 UNSAT	57%	71%	28641.98	167.50	28655.39	167.58	1.00
17	solver-1856 (complete, team 218)	170	77 SAT, 93 UNSAT	57%	71%	28607.58	168.28	28728.50	168.99	1.00

WC vs CPU time remarks

Expected results for WC ranking

Solver	Number of solved instances	Detail	% of all instances	% of VBS	Cumulated CPU time on solved instances	Average CPU time per solved instance	Cumulat time on sol instan
Total number of instances in the category: 300							
t Solver (VBS)	241	129 UNSAT, 112 SAT	80%	100%	65021.06	269.80	
te, team 222)	192	96 SAT, 96 UNSAT	64%	80%	109844.85	572.11	2
te, team 260)	192	87 SAT, 105 UNSAT	64%	80%	131387.77	684.31	3
te, team 260)	191	88 SAT, 103 UNSAT	64%	79%	132983.21	696.25	3
te, team 222)	190	95 SAT, 95 UNSAT	63%	79%	109936.90	578.62	2
te, team 211)	183	86 SAT, 97 UNSAT	61%	76%	122546.35	669.65	3
te, team 260)	181	77 SAT, 104 UNSAT	60%	75%	117630.79	649.89	3
te, team 237)	181	89 SAT, 92 UNSAT	60%	75%	119900.30	662.43	3
te, team 260)	180	78 SAT, 102 UNSAT	60%	75%	111419.81	619.00	2
te, team 220)	179	85 SAT, 94 UNSAT	60%	74%	120611.14	673.81	4
te, team 237)	178	89 SAT, 89 UNSAT	59%	74%	96265.08	540.82	2
te, team 220)	175	86 SAT, 89 UNSAT	58%	73%	118286.46	675.92	4
te, team 254)	174	82 SAT, 92 UNSAT	58%	72%	32565.62	187.16	3
te, team 260)	172	77 SAT, 95 UNSAT	57%	71%	32304.49	187.82	3
te, team 260)	172	77 SAT, 95 UNSAT	57%	71%	33519.81	194.88	3
te, team 222)	171	79 SAT, 92 UNSAT	57%	71%	28545.15	166.93	2
te, team 222)	171	79 SAT, 92 UNSAT	57%	71%	28641.98	167.50	2

WC vs CPU time remarks

Expected results for WC ranking

Cumulated CPU time on solved instances	Average CPU time per solved instance	Cumulated WC time on solved instances	Average WC time per solved instance	CPU/WC
: 300				
65021.06	269.80	-	-	-
109844.85	572.11	27875.01	145.18	3.94
131387.77	684.31	33587.71	174.94	3.91
132983.21	696.25	33820.23	177.07	3.93
109936.90	578.62	27912.74	146.91	3.94
122546.35	669.65	31289.47	170.98	3.92
117630.79	649.89	30300.60	167.41	3.88
119900.30	662.43	30520.98	168.62	3.93
111419.81	619.00	28922.57	160.68	3.85
120611.14	673.81	42756.99	238.87	2.82
96265.08	540.82	24546.73	137.90	3.92
118286.46	675.92	42570.37	243.26	2.78
32565.62	187.16	32569.63	187.18	1.00
32304.49	187.82	32319.98	187.91	1.00
33519.81	194.88	33521.71	194.89	1.00
28545.15	166.93	28535.25	166.87	1.00
28641.98	167.50	28655.39	167.58	1.00

WC vs CPU time remarks

Expectable results for WC ranking

Rank	Solver	Number of solved instances	Detail	% of all instances	% of VBS	Cumulated CPU time on solved instances	Average CPU time per solved instance	Cumulated WC time on solved instances	Average WC time per solved instance	CPU/WC
Total number of instances in the category: 300										
Virtual Best Solver (VBS)		207	64 UNSAT, 143 SAT	69%	100%	42190.88	203.82	-	-	-
1	solver-1612 (complete, team 211)	151	113 SAT, 38 UNSAT	50%	73%	63866.59	422.96	16909.26	111.98	3.78
2	solver-1613 (complete, team 211)	138	106 SAT, 32 UNSAT	46%	67%	17991.42	130.37	17998.11	130.42	1.00
3	solver-1618 (complete, team 266)	136	107 SAT, 29 UNSAT	45%	66%	24195.31	177.91	24336.17	178.94	0.99
4	solver-1853 (complete, team 237)	133	90 SAT, 43 UNSAT	44%	64%	92773.67	697.55	23307.75	175.25	3.98
5	solver-1683 (complete, team 237)	130	88 SAT, 42 UNSAT	43%	63%	87215.59	670.89	21856.66	168.13	3.99
6	solver-1846 (complete, team 260)	113	76 SAT, 37 UNSAT	38%	55%	77402.62	684.98	19581.24	173.29	3.95
7	solver-1588 (complete, team 260)	111	73 SAT, 38 UNSAT	37%	54%	74220.58	668.65	18772.10	169.12	3.95

WC vs CPU time remarks

Expectable results for WC ranking

Solver	Number of solved instances	Detail	% of all instances	% of VBS	Cumulated CPU time on solved instances	Average CPU time per solved instance	Cumulat on solve
Total number of instances in the category: 300							
Solver (VBS)	207	64 UNSAT, 143 SAT	69%	100%	42190.88	203.82	
ete, team 211)	151	113 SAT, 38 UNSAT	50%	73%	63866.59	422.96	
ete, team 211)	138	106 SAT, 32 UNSAT	46%	67%	17991.42	130.37	
ete, team 266)	136	107 SAT, 29 UNSAT	45%	66%	24195.31	177.91	
ete, team 237)	133	90 SAT, 43 UNSAT	44%	64%	92773.67	697.55	
ete, team 237)	130	88 SAT, 42 UNSAT	43%	63%	87215.59	670.89	
ete, team 260)	113	76 SAT, 37 UNSAT	38%	55%	77402.62	684.98	
ete, team 260)	111	73 SAT, 38 UNSAT	37%	54%	74220.58	668.65	

WC vs CPU time remarks

Expectable results for WC ranking

Cumulated CPU time on solved instances	Average CPU time per solved instance	Cumulated WC time on solved instances	Average WC time per solved instance	CPU/WC
gory: 300				
42190.88	203.82	-	-	-
63866.59	422.96	16909.26	111.98	3.78
17991.42	130.37	17998.11	130.42	1.00
24195.31	177.91	24336.17	178.94	0.99
92773.67	697.55	23307.75	175.25	3.98
87215.59	670.89	21856.66	168.13	3.99
77402.62	684.98	19581.24	173.29	3.95
74220.58	668.65	18772.10	169.12	3.95

WC vs CPU time remarks

Unexpected results for CPU ranking

Rank	Solver	Number of solved instances	Detail	% of all instances	% of VBS	Cumulated CPU time on solved instances	Average CPU time per solved instance	Cumulated WC time on solved instances	Average WC time per solved instance	CPU/WC
Total number of instances in the category: 300										
Virtual Best Solver (VBS)		207	64 UNSAT, 143 SAT	69%	100%	42190.88	203.82	-	-	-
1	solver-1613 (complete, team 211)	138	106 SAT, 32 UNSAT	46%	67%	17991.42	130.37	17998.11	130.42	1.00
2	solver-1618 (complete, team 266)	136	107 SAT, 29 UNSAT	45%	66%	24195.31	177.91	24336.17	178.94	0.99
3	solver-1612 (complete, team 211)	134	103 SAT, 31 UNSAT	45%	65%	20773.94	155.03	5345.35	39.89	3.89
4	solver-1520 (complete, team 217)	107	81 SAT, 26 UNSAT	36%	52%	14248.05	133.16	14248.58	133.16	1.00
5	solver-1855 (complete, team 217)	107	80 SAT, 27 UNSAT	36%	52%	18057.38	168.76	18057.80	168.76	1.00
6	solver-1853 (complete, team 237)	106	74 SAT, 32 UNSAT	35%	51%	20331.92	191.81	5129.44	48.39	3.96
7	solver-1563 (incomplete, team 253)	104	104 SAT	35%	50%	6153.68	59.17	6154.63	59.18	1.00

WC vs CPU time remarks

Unexpected results for CPU ranking

Solver	Number of solved instances	Detail	% of all instances	% of VBS	Cumulated CPU time on solved instances	Average CPU time per solved instance	Cumulat on solve
Total number of instances in the category: 300							
Solver (VBS)	207	64 UNSAT, 143 SAT	69%	100%	42190.88	203.82	
ete, team 211)	138	106 SAT, 32 UNSAT	46%	67%	17991.42	130.37	
ete, team 266)	136	107 SAT, 29 UNSAT	45%	66%	24195.31	177.91	
ete, team 211)	134	103 SAT, 31 UNSAT	45%	65%	20773.94	155.03	
ete, team 217)	107	81 SAT, 26 UNSAT	36%	52%	14248.05	133.16	
ete, team 217)	107	80 SAT, 27 UNSAT	36%	52%	18057.38	168.76	
ete, team 237)	106	74 SAT, 32 UNSAT	35%	51%	20331.92	191.81	
plete, team 253)	104	104 SAT	35%	50%	6153.68	59.17	

WC vs CPU time remarks

Unexpected results for CPU ranking

Average CPU time per solved instance	Cumulated WC time on solved instances	Average WC time per solved instance	CPU/WC
203.82	-	-	-
130.37	17998.11	130.42	1.00
177.91	24336.17	178.94	0.99
155.03	5345.35	39.89	3.89
133.16	14248.58	133.16	1.00
168.76	18057.80	168.76	1.00
191.81	5129.44	48.39	3.96
59.17	6154.63	59.18	1.00

Time limits

Theory was easy :

- phase 1 : TIMEOUT= 1200 s
- phase 2 : TIMEOUT= 5000 s

Practice proved more difficult :

- We want to run one single experiment for both rankings.
- CPU limit more reliable than WC limit
- WC limit cannot be set to TIMEOUT because the system might pause the solver during a small fraction of time. Experiments were done with a WC limit set to timeout + 100 s (large safety margin).
- In phase 1, CPU limit was set to TIMEOUT * number of allocated cores. This was a mistake, because sequential solvers were allocated 2 cores but actually used 1 core.
- In phase 2, CPU limit was set to TIMEOUT for sequential solvers and TIMEOUT * allocated cores for parallel solvers

Post-processing of results

From the single experiment, we enforce the correct limits for sequential and parallel solver by post-processing (replace answers obtained after the TIMEOUT by UNKNOWN).

Actual limits of the rankings (seconds) :

Ranking	WC limit all solvers	CPU limit seq. solvers	CPU limit // solvers
phase 1, CPU	1210	1200	1200
phase 1, WC	1210	1200	4800
phase 2, CPU	5000	5000	5000
phase 2, WC	5000	5000	40000

Allocation of cores

- Our hosts are bi quad-core processors, 32 GB RAM
- We want to optimize the use of the cluster and run as many solvers as possible on one host
- But we also want to have times almost equivalent to the ones of a solver running alone on the host

After experimentations on Minisat, it was decided to

- In phase 1, run 4 sequential solvers concurrently on a node (2 cores/solver, implies approximately a 10% penalty for minisat), and 2 parallel solvers on a node (4 cores per solver)
- In phase 2, run 2 sequential solvers concurrently on a node (4 cores/solver, almost no penalty for minisat), and 1 parallel solver on a node (8 cores per solver)
- We do not run 2 different solvers concurrently on a node.

Allowed memory

- Simple policy : share equally the memory between the solvers. Hence, solvers were allowed to use 31GB divided by the number of concurrent solvers.
- In phase 1 : 7.7 GB for seq. solvers, 15.5 GB for // solvers
- In phase 2 : 15.5 GB for seq. solvers, 31 GB for // solvers

Hence, parallel solvers were allocated twice the memory of a sequential solver !

- Looks unfair ?
- But parallel solvers necessarily need more memory than sequential solvers. Hence, enforcing the same limit would not be fair either !
- No way to be fair.
- Indirect way to encourage the development of // solvers.

Minisat Hack track : aims

- Observe the effect of clearly identified "small changes" in a widely used solver
- Help understand what is really important in Minisat, what can be improved, ...
- Ensure that all solvers are comparable (small syntactic changes)
- Lower entry level for the competition (e.g. Master or first year PhD student)
- Track initiated for SAT 2009 competition with Minisat 2.0

Based on Minisat 2.2 this year

Minimally Unsatisfiable Subset Track

- Plain Mus Track
  Success of unsat core guided maxsat solvers
- Group/High Level MUS
  Use of MUS in real applications
- Benchmarks provided by Joao Marques Silva and Alexander Nadel groups
- Additional unsat benchmarks from the main track

Resources also needed to check the results !

Submission policy

- parallel, sequential and minisat hack submissions are separated categories
- 3 submissions per category max for stage 1
- 1 solver per category max for stage 2

Data analysis track

- The SAT competition is not just a ranking : competition results at least as important
  - Check results of solvers on individual benchmarks
  - Check individual results are repeatable with the traces (e.g. for scaling).
  - Results reviewed by both the submitters and organizers
- Idea : promote the use of competition results to the community
- How : accept data analyzer for the competition

Benchmarks selection

- Random category : 600 randomly generated $k \in 3, 5, 7$-CNF, 400 SAT/400 Medium size
- Crafted and application categories : 300 benchmarks, 150 new/150 existing
  - difficulty of benchmarks using SAT 2009 reference solvers
  - selection of existing benchmarks among all available benchmarks used in SAT competitions or Races since 2002

	EASY	MEDIUM	HARD	XHARD
Application	48	124	53	75
Crafted	38	70	70	122

Categories

- 3 categories of benchmarks (APPLICATION, CRAFTED, RANDOM)
- 3 subcategories (SAT+UNSAT, SAT, UNSAT)
- 2 rankings (based on CPU time/WC time)
- 3 medals (gold, silver, bronze)
- Best minisat hack solver (Application SAT+UNSAT)
- Moshe Vardi's award against CDCL monoculture

A total of 18 rankings and 54+2 medals (3 kg of medals !)

First stage results

Available now from `http://www.cril.univ-artois.fr/SAT11/`

About ppfolio

- ppfolio stands for both Pico-PortFOLIO and Parallel PortFOLIO
- was written by O. Roussel after a discussion between the organizers and H. Hoos about VBS and portfolios
- ppfolio was designed to be a very naive and straightforward implementation of both a portfolio and a parallel solver
- the goal was to serve as a reference for both portfolios and parallel solver

ppfolio "algorithm"

- The idea was to create an approximation of the VBS that would run on a single host, using each available core.
- The VBS (Virtual Best Solver) is the solver we obtain by running in parallel each available solver on its own computer (it's by definition the best solver on Earth, but requires a cluster to run !)
- The "algorithm" is straightforward :
  - select the solvers to run according to the number of cores available
  - run a solver (or more) on each available core
  - filter the solvers output and only display 's' an 'v' lines

Comments on ppfolio

The author shamelessly claims that :

- it's probably the laziest and most stupid solver ever written
- it's so lazy it doesn't even parse the CNF
- the main program knows nothing about clauses
- ppfolio is just a system tool to run solvers in parallel, without any communication without the solvers.

About ppfolio

ppfolio starts the following solvers :

1. cryptominisat 2.7.1 (M. Soos)
2. lingeling/plingeling 276-6264d55-100731 (A. Biere)
3. clasp 1.3.6 (M. Gebser, B. Kaufmann, and T. Schaub)
4. march_hi 2009 (M. Heule and H. Van Maaren)
5. TNM 2009 (W. Wei and Chu Min Li)

Two versions of ppfolio :

- sequential (seq) : run (1), (3) and (5) on 1 core
- parallel (par) :

4 cores	run (1), (2) and (3) on their own core, (4) and (5) on the last core
>5 cores	run (1), (3), (4) and (5) on their own core, run plingeling on the remaining cores

Preliminary comments on ppfolio

- The results of ppfolio are unexpectedly good
- Even the sequential version gets decent results
- Raises a number of interesting questions...

More information on ppfolio :

- See the description on the conference USB stick
- Mode information and code on `http://www.cril.univ-artois.fr/~roussel/ppfolio`