

# The SAT 2011 competition

## Results part 2

Matti Järvisalo   Daniel Le Berre   Olivier Roussel

with the support of CRIL and Intel

SAT 2011 conference, Ann Arbor, 18 June 2011

# Heard about the SAT competition

- ▶ Focus on speed, not on ideas
- ▶ Could lead to monoculture for CDCL, preprocessing, etc.
- ▶ Competition not driven in a “scientific way” (rankings, multiple runs, etc.)

# Purpose of the competition

- ▶ Gather empirical evidences of the behavior of the solvers on benchmarks
- ▶ Promote and stimulate the design of running systems
- ▶ Gather new benchmarks/applications
- ▶ Have fun

# Purpose of the competition

- ▶ Gather empirical evidences of the behavior of the solvers on benchmarks
- ▶ Promote and stimulate the design of running systems
- ▶ Gather new benchmarks/applications
- ▶ Have fun

Limits :

- ▶ Snapshot of the solvers under specific conditions
- ▶ Ranking heavily depends on benchmarks, chance, hardware

# Purpose of the competition

- ▶ Gather empirical evidences of the behavior of the solvers on benchmarks
- ▶ Promote and stimulate the design of running systems
- ▶ Gather new benchmarks/applications
- ▶ Have fun

Limits :

- ▶ Snapshot of the solvers under specific conditions
- ▶ Ranking heavily depends on benchmarks, chance, hardware

Our motivation :

Allow scientists to start the **observe**/analyze/make hypothesis/check hypothesis/conclude cycle on SAT solvers

# Key benefits of the SAT competition

- ▶ Provide **fast, reliable and reusable** solvers to the community
- ▶ Showcase for SAT solvers (*SAT store, SAT Market*)
- ▶ Provide fast and reliable solvers for other communities
- ▶ Provide independent assessment of solvers
- ▶ Alternation of competitions and races to please everybody

# Preview of MUS results (selection of 300 benchmarks)

Caution : answers not checked, results not reviewed by submitters.

Solver	#Solved
Haifa-MUC 1	248
Haifa-MUC 2	250
Haifa-MUC-A	210
MiniUnsat2	185
MoUsSaKa	243
MUSer	241
OptimuS a100	114
OptimuS p100	117
OptimuS x100	126
PicoMUS	225
Sat4j (Qxplain)	194

# Preview of Group/High Level MUS (197 benchmarks)

Caution : answers not checked, results not reviewed by submitters.

Solver	#Solved
Haifa-HLMUC 1	197
Haifa-HLMUC 2	187
Haifa-HLMUC-A 1	196
Haifa-HLMUC-A 2	196
MoUsSaKa	194
MUSer	184
PicoMUS	181
Sat4j (Deletion)	140



# The second stage results

- ▶ 55 medals to distribute
- ▶ Moshe Vardi's price will be decided after analyzing the results and looking at the solvers descriptions.
- ▶ Medals to be given to all winners after presentation of all the results



CPU time						
	SAT+UNSAT		SAT		UNSAT	
Rank	Solver	#	Solver	#	Solver	#
Ref	SATzilla_R	384	TNM	324	march_hi	123
1	3S (coach)	408	sparrow2011	362	march_rw	123
2	ppfolio //	377	sattime2011	334	mphasesat_m	104
3	ppfolio seq	375	eagleup	328	ppfolio //	101

CPU time						
	SAT+UNSAT		SAT		UNSAT	
Rank	Solver	#	Solver	#	Solver	#
Ref	SATzilla_R	384	TNM	324	march_hi	123
1	3S (coach)	408	sparrow2011	362	march_rw	123
2	ppfolio //	377	sattime2011	334	mphasesat_m	104
3	ppfolio seq	375	eagleup	328	ppfolio //	101

  

WC time						
	SAT+UNSAT		SAT		UNSAT	
Rank	Solver	#	Solver	#	Solver	#
1	ppfolio //	444	sparrow2011	362	march_rw	123
2	3S (coach)	408	csls	334	ppfolio //	122
3	ppfolio seq	375	sattime2011	334	mphasesat_m	104



CPU time						
	SAT+UNSAT		SAT		UNSAT	
Rank	Solver	#	Solver	#	Solver	#
Ref	clasp	149	clasp	85	clasp	64
1	3S (coach)	163	ppfolio //	120	clasp	62
2	ppfolio //	155	ppfolio seq	114	3S (coach)	51
3	ppfolio seq	152	3S (coach)	112	glucose	42

CPU time						
	SAT+UNSAT		SAT		UNSAT	
Rank	Solver	#	Solver	#	Solver	#
Ref	clasp	149	clasp	85	clasp	64
1	3S (coach)	163	ppfolio //	120	clasp	62
2	ppfolio //	155	ppfolio seq	114	3S (coach)	51
3	ppfolio seq	152	3S (coach)	112	glucose	42

WC time						
	SAT+UNSAT		SAT		UNSAT	
Rank	Solver	#	Solver	#	Solver	#
1	ppfolio //	187	ppfolio //	127	claspmt	70
2	claspmt	170	ppfolio seq	114	clasp	62
3	3S (coach)	163	3S (coach)	112	ppfolio //	60





CPU time						
	SAT+UNSAT		SAT		UNSAT	
Rank	Solver	#	Solver	#	Solver	#
Ref	precosat	206	precosat	96	precosat	110
1	glucose	215	contrasat	99	glueminisat	126
2	glueminisat	211	cirminisat	99	glucose	121
3	lingeling	208	mphasesat64	99	qutersat	120

CPU time						
	SAT+UNSAT		SAT		UNSAT	
Rank	Solver	#	Solver	#	Solver	#
Ref	precosat	206	precosat	96	precosat	110
1	glucose	215	contrasat	99	glueminisat	126
2	glueminisat	211	cirminisat	99	glucose	121
3	lingeling	208	mphasesat64	99	qutersat	120

WC time						
	SAT+UNSAT		SAT		UNSAT	
	Solver	#	Solver	#	Solver	#
1	plingeling	230	ppfolio //	107	cryptoms	133
2	cryptoms	230	plingeling	106	glueminisat	126
3	ppfolio //	229	contrasat	99	plingeling	124

Solver	Gold	Silver	Bronze	Total
ppfolio //	5	3	3	11
3S (coach)	2	2	3	7
cryptoms //	1	1		2
sparrow2011	2			2
march_rw	2			2
glueminisat	1	2		3
plingeling	1	1	1	3
glucose	1	1	1	3
claspmt	1	1		2
clasp seq	1	1		2
contrasat	1		1	2
ppfolio seq		2	3	5
sattime2011		1	1	2
mphasesatm		1	1	2
csls		1		1
cirminisat		1		1
lingeling			1	1
mphasesat64			1	1
qutersat			1	1
eagleup			1	1

## Two minute observation of the results

- ▶ Competition favorable to portfolio (e.g. random SAT+UNSAT, crafted SAT and SAT+UNSAT)
- ▶ Many good minisat hack solvers for application SAT
- ▶ Glue-based solvers good on application UNSAT benchmarks
- ▶ Improvement in random SAT
- ▶ Wall clock ranking favors “brute force” parallel approaches

## ppfolio : first solver to answer per category

*reminder : ppfolio is essentially equivalent to typing*  
solver1 & solver2 & ... solvern &

name	answering solver	count
APPLICATION	plingeling	121
APPLICATION	cryptominisat	51
APPLICATION	clasp	49
APPLICATION	TNM	4
APPLICATION	march_hi	4
CRAFTED	TNM	75
CRAFTED	cryptominisat	35
CRAFTED	clasp	31
CRAFTED	march_hi	24
CRAFTED	plingeling	22
RANDOM	TNM	320
RANDOM	march_hi	124

# ppfolio : first solver to answer per category, SAT/UNSAT

name	answer	answering solver	count
APPLICATION	SAT	plingeling	68
APPLICATION	SAT	clasp	26
APPLICATION	SAT	cryptominisat	9
APPLICATION	SAT	TNM	4
APPLICATION	UNSAT	plingeling	53
APPLICATION	UNSAT	cryptominisat	42
APPLICATION	UNSAT	clasp	23
APPLICATION	UNSAT	march_hi	4
CRAFTED	SAT	TNM	75
CRAFTED	SAT	cryptominisat	24
CRAFTED	SAT	plingeling	18
CRAFTED	SAT	clasp	7
CRAFTED	SAT	march_hi	3
CRAFTED	UNSAT	clasp	24
CRAFTED	UNSAT	march_hi	21
CRAFTED	UNSAT	cryptominisat	11
CRAFTED	UNSAT	plingeling	4
RANDOM	SAT	TNM	320
RANDOM	SAT	march_hi	2
RANDOM	UNSAT	march_hi	122

## 32 cores results, application benchmarks

Rank	Solver	#solved	Detail	%inst.	%VBS
<i>Total number of instances : 300</i>					
<i>Virtual Best Solver (VBS)</i>		258	116 SAT, 142 UNSAT	86%	100%
1	CryptoMS	236	103 SAT, 133 UNSAT	79%	91%
2	ppfolio	235	113 SAT, 122 UNSAT	78%	91%
3	Plingeling	228	105 SAT, 123 UNSAT	76%	88%
4	claspmt	219	97 SAT, 122 UNSAT	73%	85%

Number of instances solved :

- ▶ 215 sequential (131771 CPU s)
- ▶ 230 with 8 cores (905080 CPU s) : seq  $\times$  6.8 CPU effort gains 15
- ▶ 236 with 32 cores (2876697 CPU s) : seq  $\times$  21.8 CPU effort gains 21

- ▶ Before 2011 : Satzilla
  - ▶ 2009 gold in all categories
  - ▶ 2007 gold in crafted and random
  - ▶ 2003 first submission to the competition
- ▶ This year, many portfolios : pportfolio, 3S (coach), borg

How should we handle portfolio in the future ?