The SAT 2009 competition results

does theory meet practice?

Daniel Le Berre Olivier Roussel Laurent Simon
Andreas Goerdt Ines Lynce Aaron Stump

Supported by CRIL, LRI and French ANR UNLOC

SAT 2009 conference, Swansea, 3 July 2009

For those having a computer and Wifi access

- See the rules and benchmarks details at : `http://www.satcompetition.org/2009/`
- See the results live at : `http://www.cril.univ-artois.fr/SAT09/`

The team

Organizers
- Daniel Le Berre
- Olivier Roussel
- Laurent Simon (apart main track)

Judges
- Andreas Goerdt
- Ines Lynce
- Aaron Stump

Computer infrastructure provided by CRIL (96 bi-processor cluster) and LRI (48 quad-core cluster + one 16 core machine (68GB) for the parallel track).

The tracks

Main track sequential solvers

- **competition** Source code of the solver should be available after the competition
- **demonstration** Binary code should be available after the competition (for research purpose)

Parallel Solvers tailored to run on multicore computers (up to 16 cores)

Minisat Hack Submission of (small) patches against latest public release of Minisat2

Preprocessing track competition of preprocessors in front of Minisat2.

Integration of the competition in the conference

Tuesday

- Efficiently Calculating Tree Measures Using SAT : bio_2 benchmarks
- Finding Efficient Circuits Using SAT solvers : mod_circuits benchmarks

Wednesday

- On the fly clause improvement : Circus, main track
- Problem sensitive restarts heuristics for the DPLL procedure : Minisat09z, minisat hack
- Improved Conflict-Clause Minimization Leads to Improved Propositional Proof Traces : Minisat2Hack, minisat hack
- A novel approach to combine SLS and a DPLL solver for the satisfiability problem : hybridGM, main track
- Building a Hybrid SAT solver via Conflict Driven, Look-Ahead and Xor reasoning techniques : MoRsat, main track
- Improving Variable Selection Process in Stochastic Local Search for Propositional Satisfiability : slstc, main track
- VARSAT : Integrating Novel Probabilistic Inference Techniques with DPLL Search :VARSAT, main track

Thursday

- Width-Based Restart Policies for Clause Learning : Rsat, main track

Common rules to all tracks

- No more than 3 solvers per submitter
- Compared using a simple static ranking scheme
- Results available for SAT, UNSAT and SAT+UNSAT benchmarks.
- Results available to the submitters for checking : It is the responsibility of the competitor to check that his system performed as expected !

New scoring scheme

- Purse based scoring since 2005 (designed by Allen van Gelder).
  - pros
    - Take into account various aspects of the solver (power, robustness, speed).
    - Focus on singular solvers
  - cons
    - Difficult to check (and understand)
    - Too much weight on singularity ?
    - Depends on the set of competitors
- "Spec 2009" static scoring scheme desirable
  - To compare easily other solvers (e.g. reference solvers) without disturbing the ranking of the competitors.
  - To allow anybody to compare his solver to the SAT 2009 competitors on similar settings.

Available metrics

NBTOTAL	Total number of benchmarks to solve
NBSOLVED	Total number of benchmarks solved within a given timeout
NBUNSOLVEDSERIES	Total number of set of benchmarks for which the solver was unable to solve any element.
TIMEOUT	Time allowed to solve a given benchmark
ti	Time needed to solve a given benchmark, within the time limit
PENALTY	Constant to use as a penalty for benchmarks not solved within the timeout
SERIESPENALTY	Constant to use as a penalty for a set of benchmarks in which all members cannot be solved by the solver.

Spec 2009 proposals

- Lexicographical $NBSOLVED, \sum ti$
- Cumulative time based, with timeout penalty

$$\sum t_i + (NBTOTAL - NBSOLVED) * TIMEOUT * PENALTY$$

- Cumulative time based, with timeout penalty, log based

$$\sum log_{10}(1+ti)+(NBTOTAL-NBSOLVED)*log_{10}((1+TIMEOUT)*PENALTY)$$

- Cumulative time based, with timeout and robustness penalties (Proposed by Marijn Heule)
  $\sum t_i + (NBTOTAL - NBSOLVED) * TIMEOUT * PENALTY +$
  $NBUNSOLVEDSERIES * SERIESPENALTY$
- SAT 2005 and 2007 purse based scoring

Spec 2009 proposals and results of the votes

- Lexicographical $NBSOLVED, \sum ti$ 9 votes
- Cumulative time based, with timeout penalty 3 votes

$$\sum t_i + (NBTOTAL - NBSOLVED) * TIMEOUT * PENALTY$$

- Cumulative time based, with timeout penalty, log based

$$\sum log_{10}(1+ti)+(NBTOTAL-NBSOLVED)*log_{10}((1+TIMEOUT)*PENALTY)$$

- Cumulative time based, with timeout and robustness penalties (Proposed by Marijn Heule) 4 votes
  $\sum t_i + (NBTOTAL - NBSOLVED) * TIMEOUT * PENALTY +$
  $NBUNSOLVEDSERIES * SERIESPENALTY$
- SAT 2005 and 2007 purse based scoring

Industrial vs Application

- Many instances in the industrial category do not come from industry
- Application better reflects the wide use of SAT technology

Benchmarks selection : Random category

Based on O. Kullmann recommendations in 2005 (see [OK JSAT06] for details)

	3-SAT			5-SAT			7-SAT		
Generated benchmarks parameters									
	ratio	start-stop	step	ratio	start-stop	step	ratio	start-stop	step
Medium	4.26	360-560	20	21.3	90-120	10	89	60-75	5
Large	4.2	2000-18000	2000	20	700-1100	100	81	140-220	20
Number of generated benchmarks									
	SAT	UNKNOWN		SAT	UNKNOWN		SAT	UNKNOWN	
Medium	110	110		40	40		40	40	
Large	90	-		50	-		50	-	

- Balanced number of SAT/UNKNOWN benchmarks for complete solvers : 190/190
- Specific benchmarks for complete SAT solvers : 190
- Specific benchmarks for incomplete SAT solvers 190
- Satisfiability of medium benchmarks checked using gNovelty+.
- Satisfiability of large benchmarks per construction (ratio $<$ threshold).
- 100 benchmarks generated for each setting.
- Randomly selected benchmarks 10 using judges random seed
- 40 large 3-SAT benchmarks (20K-26K variables) added for the second stage

How to predict benchmark hardness for non-random benchmarks ?

- Problem : we need benchmarks to discriminate solvers (i.e. not too easy, not too hard).
- Challenging benchmarks necessary to see the limit of current approaches
- Idea : use a small set of last SAT winners in each categories
- Rsat, Minisat and picosat to rank application benchmarks
- March-KS, Satzilla-Crafted and Minisat for crafted benchmarks

easy	Solved within 30s by all the solvers
hard	Not solved by any of the solvers (timeout used)
medium	Remaining instances

Judges decisions regarding the selection of submitted vs existing benchmarks

- No more than 10% of the benchmarks should come from the same source.
- The final selection of benchmarks should contain 45% existing benchmarks and 55% submitted benchmarks.
- The final selection should contain 10% easy, 40% medium and 50% hard benchmarks.
- Duplicate benchmarks found after the selection was done will simply be removed from the selection. No other benchmarks will be added to the selection.

Application benchmarks submitted to the competition

Aprove	(*Carsten Fuhs*) Term Rewriting systems benchmarks.
BioInfo I	(*Fabien Corblin*) Queries to find the maximal size of a biological behavior without cycles in discrete genetic networks.
BioInfo II	(*Maria Louisa Bonet*) Evolutionary trees (presented on Tuesday).
Bit Verif	(*Robert Brummayer*) Bit precise software verification generated by the SMT solver Boolector.
C32SAT	Submitted by Hendrik Post and Carsten Sinz. Software verification generated by the C32SAT satisfiability checker for C programs.
Crypto	(*Milan Sesum*) Encode attacks for both the DES and MD5 crypto systems.
Diagnosis	(*Anbulagan and Alban Grastien*) 4 different encodings of discrete event systems.

Application benchmarks : classification

Origin	EASY		MEDIUM		HARD			Total
	SAT	UNSAT	SAT	UNSAT	SAT	UNSAT	UNKOWN	
SAT RACES	6	18	43	50	3	21	-	141
SAT COMP 07	6	15	47	49	7	12	45	181
SUBMITTED 09	60	38	38	60	8	12	102	318
Total	72	71	128	159	18	45	147	640

Origin	EASY		MEDIUM		HARD			Total
	SAT	UNSAT	SAT	UNSAT	SAT	UNSAT	UNKOWN	
Aprove	21	-	4	-	-	-	-	25
BioInfo I	3	-	6	11	-	-	-	20
BioInfo II	9	-	4	3	-	-	24	40
Bit Verif	-	14	-	22	-	6	23	65
C32SAT	-	1	1	3	-	3	2	10
Crypto	5	-	7	6	4	-	40	62
Diagnosis	22	23	16	15	4	3	13	96
Total	60	38	38	60	8	12	102	318

Application benchmarks, final selection

	EASY			MEDIUM			HARD				Total
Origin	SAT	UNSAT	ALL	SAT	UNSAT	ALL	SAT	UNSAT	UNK	ALL	
old	1	9	10	21	33	54	6	23	34	63	127
new	18	1	19	25	40	65	8	10	63	81	165
Total	19	10	29	46	73	119	14	33	97	144	292

Crafted benchmarks submitted to the competition

Edge Matching Submitted by Marijn Heule. Four encodings of edge matching problems

Mod Circuits submitted by Grigory Yaroslavtsev. Presented on Tuesday.

Parity Games submitted by Oliver Friedmann. The generator encodes parity games of a fixed size n that forced the strategy improvement algorithm to require at least i iterations.

Ramsey Cube Submitted by Philipp Zumstein.

RB SAT Submitted by Nouredine Ould Mohamedou. Random CSP problems encoded into SAT.

Sgen submitted by Ivor Spence. Small but hard satisfibility benchmarks, either SAT or UNSAT.

SGI submitted by Calin Auton. Random SGI model -SRSGI. Sub Graph isomorphism problems.

Difficulty of crafted benchmarks

Origin	EASY		MEDIUM		HARD			Total
	SAT	UNSAT	SAT	UNSAT	SAT	UNSAT	UNKOWN	
Edge Matching	-	-	20	-	6	-	6	32
ModCircuits	-	1	4	1	-	-	13	19
Parity Games	6	8	7	2	-	-	1	24
Ramsey Cube	1	-	5	3	-	-	1	10
RBSAT	-	-	34	1	-	-	325	360
SGEN	5	1	4	2	-	-	9	21
SGI	106	-	1	-	-	-	-	107
Total	118	10	75	9	6	-	355	573

	EASY			MEDIUM			HARD				Total
Origin	SAT	UNSAT	ALL	SAT	UNSAT	ALL	SAT	UNSAT	UNK	ALL	
old	-	4	4	19	42	61	4	12	58	74	139
new	19	7	26	50	9	59	6	-	70	76	161
Total	19	11	30	69	65	120	11	10	129	150	300

Preprocessor track : aim

Back to the first competition aim :

- a lot of new methods exist, but hard to tell which one is the best
- Satelite is widely used, but getting old
- We want to encourage new methods
- Allow to easily enhance all solvers by just adding preprocessors in front of them

Preprocessor track : competitors

Solver name	Authors
Competition division	
IUT_BMB_SIM 1.0	Abdorrahim Bahrami, Seyed Rasoul Mousavi, Kiarash Bazargan
ReVivAl 0.23	Cédric Piette
ReVivAl 0.23 + SatElite	Cédric Piette
SatElite + ReVivAl 0.23	Cédric Piette
Demonstration division	
kw_pre	Johan Alfredsson
Reference solvers	
minisat2-core	Niklas Een and Niklas Sorensson
minisat2-simp	Niklas Een and Niklas Sorensson

Preprocessing track : experimental settings

Benchmarks the one from the main track in both application and crafted categories.

SAT engine Minisat2 070721 core solver (without preprocessing).

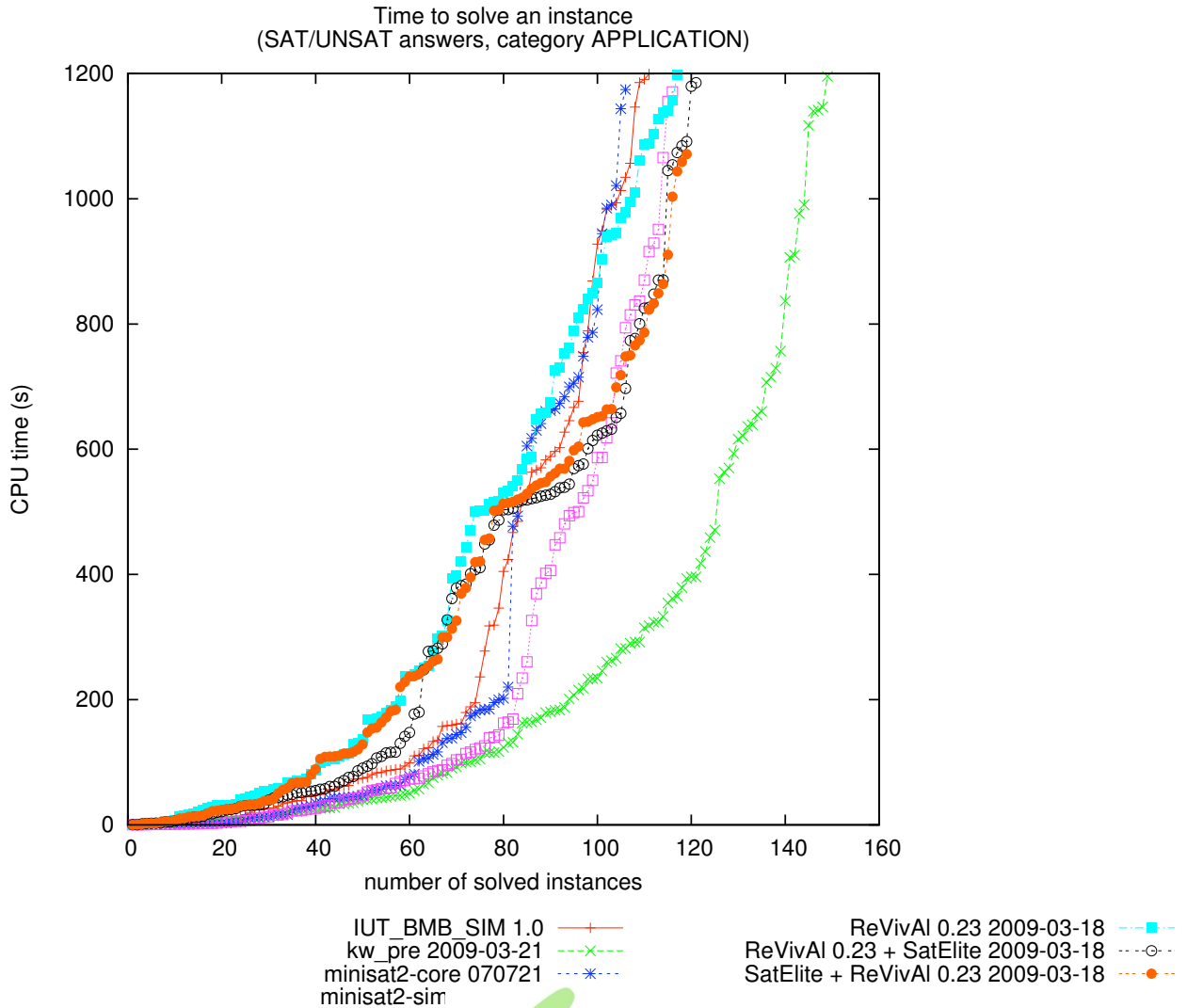
Comparison criteria the preprocessor and the engine seen as a black box.

Timeout 1200s.

Preprocessing track : the results in application category

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	163	67	96	33886.97
1	kw_pre	149	58	91	34591.65
2	ReVivAl 0.23 + SatElite	121	51	70	39093.24
3	SatElite + ReVivAl 0.23	119	48	71	38374.13
4	ReVivAl 0.23	117	53	64	44067.36
5	minisat2-simp	116	46	70	25111.90
6	IUT_BMB_SIM 1.0	111	46	65	30273.14
7	minisat2-core	106	47	59	23477.71

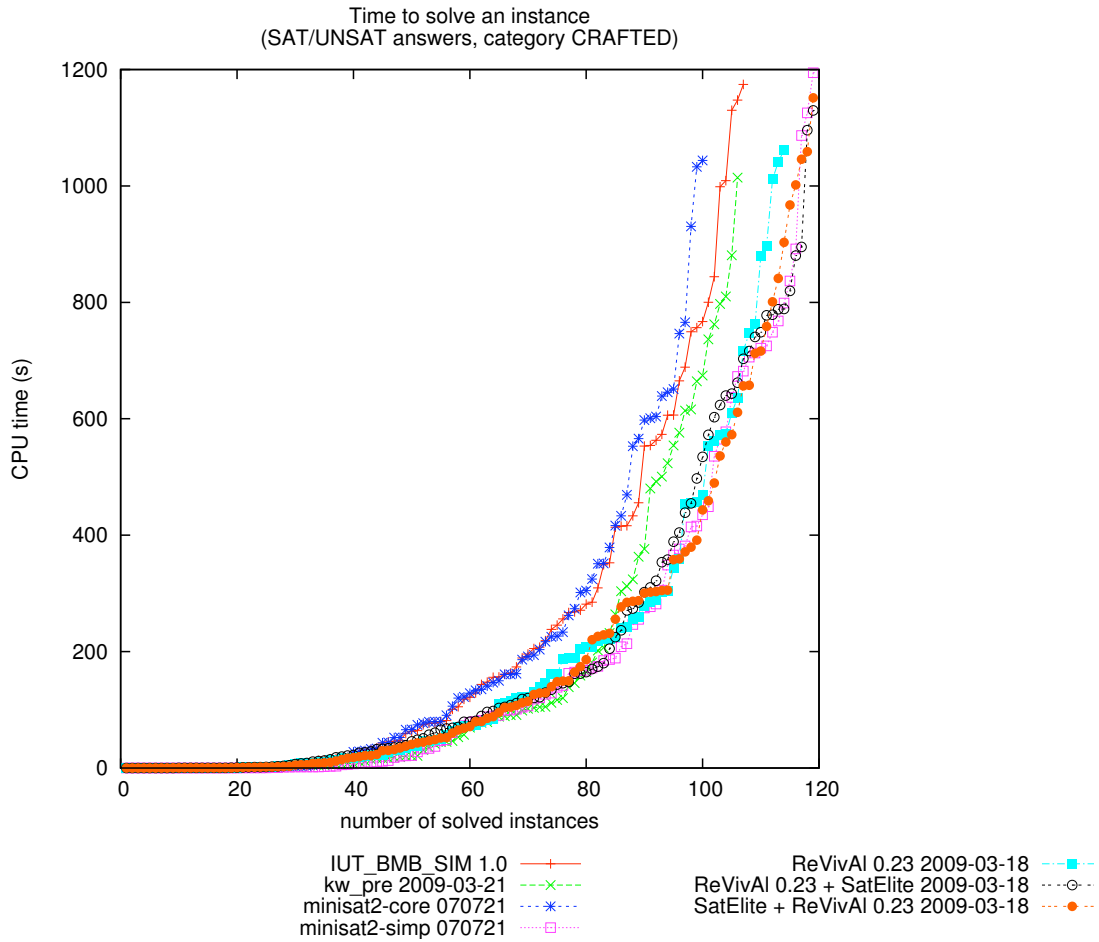
Preprocessing track running time : application (SAT+UNSAT)



Preprocessing track : the results in crafted category

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	137	92	45	20732.67
1	minisat2-simp	119	76	43	23212.54
2	SatElite + ReVivAl 0.23	119	75	44	24059.71
3	ReVivAl 0.23 + SatElite	119	75	44	24622.54
4	ReVivAl 0.23	114	72	42	20435.40
5	IUT_BMB_SIM 1.0	107	74	33	23163.33
6	kw_pre	106	72	34	16298.74
7	minisat2-core	100	69	31	17639.05

Preprocessing track running time : crafted (SAT+UNSAT)



Minisat Hack track : aim

- 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)
- Encourage easy entries to the competition (e.g. Master or first year PhD student)

Minisat Hack competitors

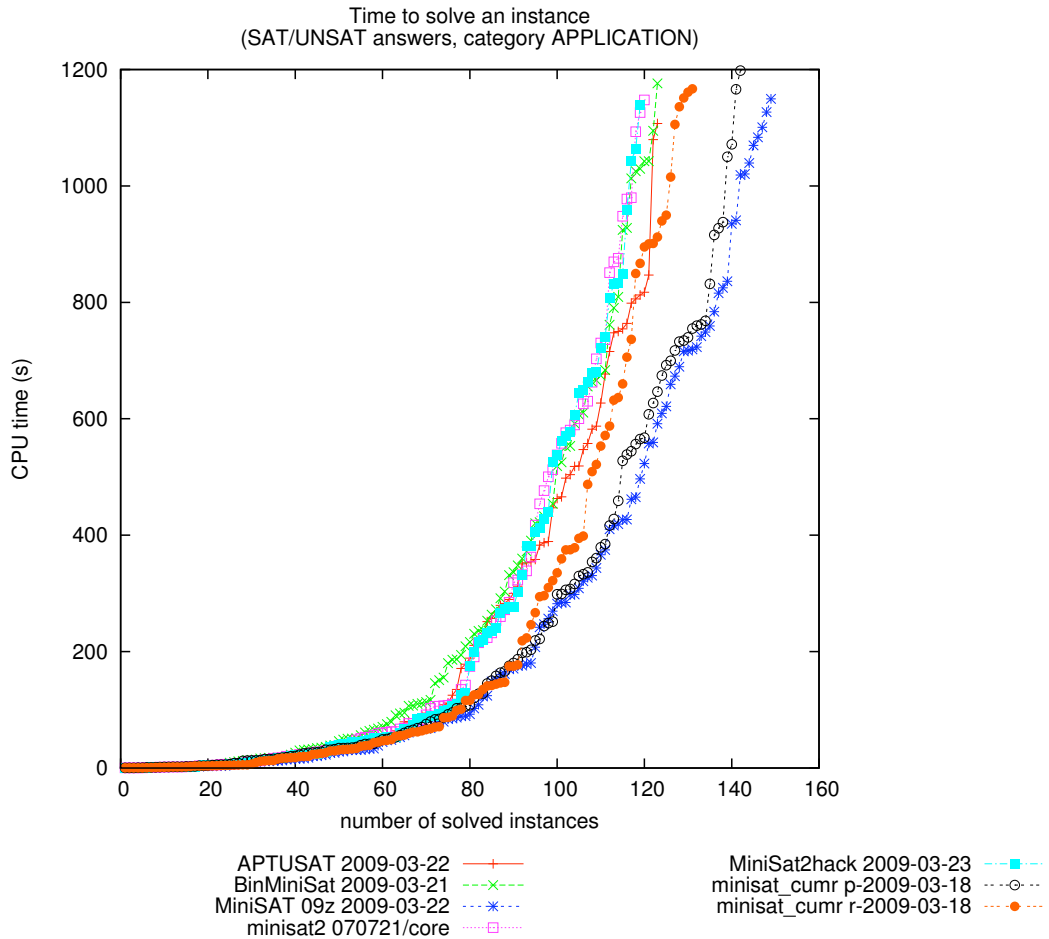
Solver name	Authors
Submissions	
APTUSAT	Alexander Mishunin and Grigory Yaroslavtsev
BinMiniSat	Kiyonori Taniguchi, Miyuki Koshimura, Hiroshi Fujita, and Ryuzo Hasegawa
MiniSAT 09z	Markus Iser
MiniSat2hack	Allen Van Gelder
minisat_cumr p/r	Kazuya Masuda and Tomio Kamada
Reference solvers	
minisat2 core	Niklas Een and Niklas Sorensson

Solvers presented during the SAT 2009 conference

Minisat hack results

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	169	71	98	40959.35
1	MiniSAT 09z	149	59	90	37228.91
2	minisat_cumr p	142	58	84	32636.31
3	minisat_cumr r	131	60	71	29316.97
4	APTUSAT	123	54	69	25418.27
5	BinMiniSat	123	48	75	29326.67
6	minisat2 core	120	53	67	25600.16
7	MiniSat2hack	119	52	67	24024.97

Minisat hack running time : application (SAT+UNSAT)



Parallel (multithreads) track : aim

We'll have to deal with multicores computers, let's start thinking about it.

- Naive parellelization should not work on many cores : memory access is a hard bottleneck for SAT solvers
- We would like to observe if multithreaded solvers scale well on a machine with 16 cores.

Parallel (multithreads) track : aim

We'll have to deal with multicores computers, let's start thinking about it.

- Naive parellelization should not work on many cores : memory access is a hard bottleneck for SAT solvers
- We would like to observe if multithreaded solvers scale well on a machine with 16 cores.
- Problem : not enough competitors !

Parallel track : the competitors

Solver name	Authors
No limit on threads	
gNovelty+-T	Duc-Nghia Pham and Charles Gretton
satake	Kota Tsuyuzaki
ttsth-5-0	Ivor Spence
Limited to 4 threads	
ManySAT 1.1 aimd 0/1/2	Youssef Hamadi, Saïd Jabbour, Lakhdar Saïs

Parallel track : the settings

- Parallel solvers ran on 3 different computers :
  - 2 processors with the main track, first stage, at CRIL.
  - 4 cores on a cluster of 4 core computers at LRI.
  - 16 cores on one specific 16 core computer at LRI.
- The solvers are given 10000s CPU time to be shared by the different threads : to be compared with the second stage of the main track.
- We ran only solvers able to use the 16 cores on the 16 core computer.

Parallel track : the results

Solver	Total	SAT	UNSAT	CPU Time
Application				
2 Threads (CRIL)				
ManySAT 1.1 aimd 1	193	71	122	173344.71
4 Threads (LRI)				
ManySAT 1.1 aimd 1	187	68	119	112384.15
ManySAT 1.1 aimd 0	185	69	116	103255.01
ManySAT 1.1 aimd 2	181	65	116	104021.63
satake	118	52	66	50543.61
ttsth-5-0	7	3	4	2274.38
16 Threads (LRI)				
satake	106	40	66	130477.38
ttsth-5-0	7	3	4	9007.53
Random				
gNovelty+-T (2 threads CRIL)	314	314	-	143439.69
gNovelty+-T (4 threads LRI)	296	296	-	95118.33
gNovelty+-T (16 threads LRI)	237	237	-	68173.49

The main track : competitors

Solver name	Authors
adaptg2wsat2009/++	chuMin Li, Wanxia Wei
CircUs	Hyojung Han
clasp 1.2.0-SAT09-32	Benjamin Kaufmann
CSat 2009-03-22	Guanfeng Lv, Qian Wang, Kaile Su
glucose 1.0	Gilles Audemard and Laurent Simon
gnovelty+/2/2-H	Duc-Nghia Pham and Charles Gretton
Hybrid2	Wanxia Wei, Chu Min Li, and Harry Zhang
hybridGM 1/3/7	Adrian Balint
HydraSAT base/flat/multi	Christoph Baldow, Friedrich Gräter, Steffen Hölldobler, Norbert Manthey, Max Seelemann, Peter Steinke, C
iPAWS	John Thornton and Duc Nghia Pham
IUT_BMB_SAT 1.0	Abdorrahim Bahrami, Seyed Rasoul Mousavi, Kiarash Bazargan
LySAT c/i	Youssef Hamadi, Saïd Jabbour, Lakhdar Saïs
march_hi/nn	Marijn Heule
MoRsat	Jingchao Chen
MXC	David Bregman
NCVWr	Wanxia Wei, Chu Min Li, and Harry Zhang
picosat 913	Armin Biere
precosat 236	Armin Biere
Rsat	Knot Pipatsrisawat and Adnan Darwiche
SApperloT base/hrp	Stephan Kottler
SAT4J CORE 2.1 RC1	Daniel Le Berre
SATzilla2009_C/I/R	Lin Xu, Frank Hutter, Holger H. Hoos and Kevin Leyton-Brown
slstc 1.0	Anton Belov, Zbigniew Stachniak
TNM	Wanxia Wei and Chu Min Li
tts-5-0	Ivor Spence
VARSAT-crafted/random/industrial	Eric Hsu
kw 2009-03-20	Johan Alfredsson
MiniSat 2.1 (Sat-race'08 Edition)	Niklas Sorensson, Niklas Een

The main track : reference solvers from 2007

Solver name	Authors
Random	
adaptg2wsat+	Wanxia Wei, Chu-Min Li and Harry Zhang
gnovelty+	Duc Nghia Pham and Charles Gretton
March KS	Marijn Heule and Hans van Maaren
SATzilla RANDOM	Lin Xu, Frank Hutter, Holger H. Hoos and Kevin Leyton-Brown
Application	
picosat 535	Armin Biere
Rsat 07	Knot Pipatsrisawat and Adnan Darwiche
Crafted	
SATzilla CRAFTED	Lin Xu, Frank Hutter, Holger H. Hoos and Kevin Leyton-Brown
minisat SAT 2007	Niklas Sorensson and Niklas Een

Main track : phase 1, application

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	196	79	117	33863.84
1	precosat 236	164	65	99	37379.67
2	MiniSat 2.1	155	65	90	27011.56
3	LySAT i	153	57	96	35271.11
4	glucose 1.0	152	54	98	34784.84
5	MiniSAT 09z	152	59	93	37872.87
6	kw	150	58	92	35080.23
7	ManySAT 1.1 aimd 1	149	54	95	34834.19
8	ManySAT 1.1 aimd 0	149	54	95	38639.59
9	MXC	147	62	85	27968.90
10	ManySAT 1.1 aimd 2	145	51	94	34242.50
11	CircUs	144	59	85	36680.28
12	Rsat	143	53	90	31000.89
13	SATzilla2009_I	142	60	82	33608.36
14	minisat_cumr p	141	58	83	29304.08
15	picosat 913	139	63	76	34013.47
16	clasp 1.2.0-SAT09-32	138	53	85	33317.37
17	Rsat 2007	133	56	77	28975.23
18	SApperloT base	129	55	74	31762.78
19	picosat 535	126	59	67	33871.13

Main track : phase 1, application continued

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	196	79	117	33863.84
20	LySAT c	123	51	72	26865.49
21	IUT_BMB_SAT 1.0	116	46	70	20974.40
22	HydraSAT Base	116	53	63	26856.33
23	HydraSAT-Flat Flat	115	51	64	26016.34
24	VARSAT-industrial	110	49	61	22753.77
25	SApperloT hrp	107	42	65	20954.19
26	HydraSAT-Multi	106	49	57	16308.48
27	SATzilla2009_C	106	45	61	25974.72
28	VARSAT-crafted	99	44	55	23553.01
29	SAT4J CORE 2.1 RC1	95	46	49	25380.84
30	satake	92	40	52	18309.62
31	CSat 2009-03-22	91	40	51	20461.14
32	SATzilla2009_R	59	36	23	6260.03
33	VARSAT-random	59	25	34	16836.65
34	march_hi	21	9	12	5170.80
35	march_nn	21	10	11	6189.51
36	Hybrid2	12	11	1	3851.84
37	adaptg2wsat2009	11	8	3	1746.45
38	adaptg2wsat2009++	11	8	3	1806.37

Main track : phase 1, application continued two

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	196	79	117	33863.84
39	slstc 1.0	10	9	1	2093.29
40	tts	10	6	4	2539.03
41	NCVWr	10	9	1	2973.72
42	iPAWS	8	8	3	1400.34
43	ttsth-5-0	8	4	4	2937.42
44	hybridGM7	7	7	-	468.76
45	gnovelty+	7	7	-	1586.83
46	gNovelty+-T	7	7	-	1826.46
47	TNM	6	5	1	1157.83
48	hybridGM 1	5	5	-	731.62
49	hybridGM3	5	5	-	1103.11
50	gnovelty+2	4	4	-	91.85

First stage : crafted

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	194	124	70	19204.67
1	clasp 1.2.0-SAT09-32	131	78	53	22257.76
2	SATzilla2009_I	128	86	42	21700.11
3	SATzilla2009_C	125	73	52	16701.85
4	MXC 2009-03-10	124	80	44	22256.57
5	precosat 236	122	81	41	22844.50
6	IUT_BMB_SAT 1.0	120	76	44	22395.97
7	minisat SAT 2007	119	76	43	22930.58
8	SATzilla CRAFTED	114	82	32	18066.80
9	MiniSat 2.1 (Sat-race'08 Edition)	114	74	40	18107.02
10	glucose 1.0	114	75	39	20823.96
11	VARSAT-industrial	113	73	40	22306.77
12	SApperloT base	113	73	40	22826.65
13	picosat 913	112	80	32	17111.73
14	LySAT c	112	70	42	21080.61
15	CircUs	107	70	37	16148.01
16	kw	106	72	34	16460.37
17	Rsat	105	71	34	14010.73
18	SATzilla2009_R	104	78	26	14460.38
19	ManySAT 1.1 aimd 1	103	72	31	14991.64

First stage : crafted continued

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	194	124	70	19204.67
20	HydraSAT-Multi	103	70	33	20825.53
21	HydraSAT-Flat	102	70	32	17796.15
22	SApperloT hrp	102	69	33	20647.84
23	minisat_cumr p	102	75	27	23176.38
24	VARSAT-crafted	102	61	41	23304.40
25	LySAT i	100	69	31	14874.18
26	ManySAT 1.1 aimd 2	99	70	29	14211.48
27	ManySAT 1.1 aimd 0	99	71	28	15251.61
28	HydraSAT base	99	66	33	16718.94
29	MiniSAT 09z	99	72	27	17027.31
30	VARSAT-random	84	47	37	14023.19
31	satake	75	55	20	16261.12
32	iPAWS	71	71	-	7352.89
33	SAT4J CORE 2.1 RC1	71	50	21	15136.95
34	adaptg2wsat2009	70	68	2	9425.51
35	adaptg2wsat2009++	66	64	2	5796.69
36	Hybrid2	66	66	-	10425.56
37	CSat	65	50	15	10319.33

First stage : crafted continued two

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	194	124	70	19204.67
38	march_hi	63	45	18	10622.02
39	TNM	62	62	-	8181.19
40	March KS	61	42	19	9021.93
41	march_nn	58	43	15	6232.17
42	gnovelty+	54	54	-	5853.95
43	gNovelty+-T	53	53	-	5073.82
44	hybridGM	51	51	-	5298.30
45	hybridGM3	51	51	-	6737.29
46	NCVWr	48	48	-	12116.63
47	tts 5-0	46	25	21	2507.80
48	ttsth-5-0	46	24	22	4020.68
49	gnovelty+2	46	44	2	4840.28
50	hybridGM7	38	38	-	4385.10
51	slstc 1.0	33	33	-	4228.67

Random results

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	459	359	100	62339.75
1	SATzilla2009_R	365	299	66	51997.72
2	TNM	317	317	-	35346.17
3	gnovelty+2	305	305	-	26616.48
4	hybridGM3	299	299	-	23272.79
5	hybridGM7	298	298	-	25567.23
6	adaptg2wsat2009++	297	297	-	26432.65
7	hybridGM 1	294	294	-	23732.78
8	adaptg2wsat2009	294	294	-	26658.47
9	Hybrid2	290	290	-	30134.40
10	gnovelty+	281	281	-	25523.72
11	NCVWr	278	278	-	31132.10
12	gnovelty+	272	272	-	21956.28
13	SATzilla RANDOM	268	177	91	42919.16
14	gNovelty+-T	266	266	-	22823.37
15	adaptg2wsat+	265	265	-	22333.18
16	iPAWS	258	258	-	19296.93
17	march_hi	247	147	100	65568.89
18	march_nn	243	145	98	66494.85
19	March KS	239	149	90	57869.03
20	SATzilla2009_I	145	90	55	37645.86

Random results : weak solvers

Rank	Solver	Total	SAT	UNSAT	CPU Time
21	slstc 1.0	118	118	-	13250.77
22	clasp 1.2.0-SAT09-32	84	66	18	32979.32
23	VARSAT-random	83	72	11	30273.41
24	picosat 913	79	57	22	29440.52
25	SATzilla2009_C	73	61	12	22395.73
26	VARSAT-industrial	71	61	10	27295.84
27	VARSAT-crafted	70	60	10	27367.38
28	SApperloT base	70	53	17	28249.79
29	IUT_BMB_SAT 1.0	63	50	13	25630.38
30	MXC	61	50	11	28069.37
31	LySAT c	60	48	12	24329.68
32	MiniSat 2.1 (Sat-race'08 Edition)	41	37	4	16957.09
33	minisat_cumr p	29	29	-	14078.15
34	precosat 236	27	25	2	9522.84
35	satake	24	24	-	11034.05
36	SApperloT hrp	17	13	4	7724.34
37	glucose 1.0	17	17	-	7772.56

Random problems : very bad solvers

Rank	Solver	Total	SAT	UNSAT	CPU Time
38	HydraSAT-Flat	16	15	1	5738.82
39	HydraSAT-Multi	16	16	-	7836.19
40	HydraSAT Base	13	13	-	4930.65
41	CircUs	8	8	-	2553.24
42	ManySAT 1.1 aimd 0	7	7	-	1783.47
43	ManySAT 1.1 aimd 2	6	6	-	957.09
44	LySAT i	6	6	-	2124.49
45	CSat	6	6	-	2263.92
46	Rsat	5	5	-	1801.20
47	ManySAT 1.1 aimd 1	5	5	-	4144.36
48	kw	4	4	-	635.52
49	SAT4J CORE 2.1 RC1	4	4	-	1440.19
50	MiniSAT 09z	3	3	-	1096.04
51	tts 5.0	0	0	0	0.00
52	ttsth-5-0	0	0	0	0.00

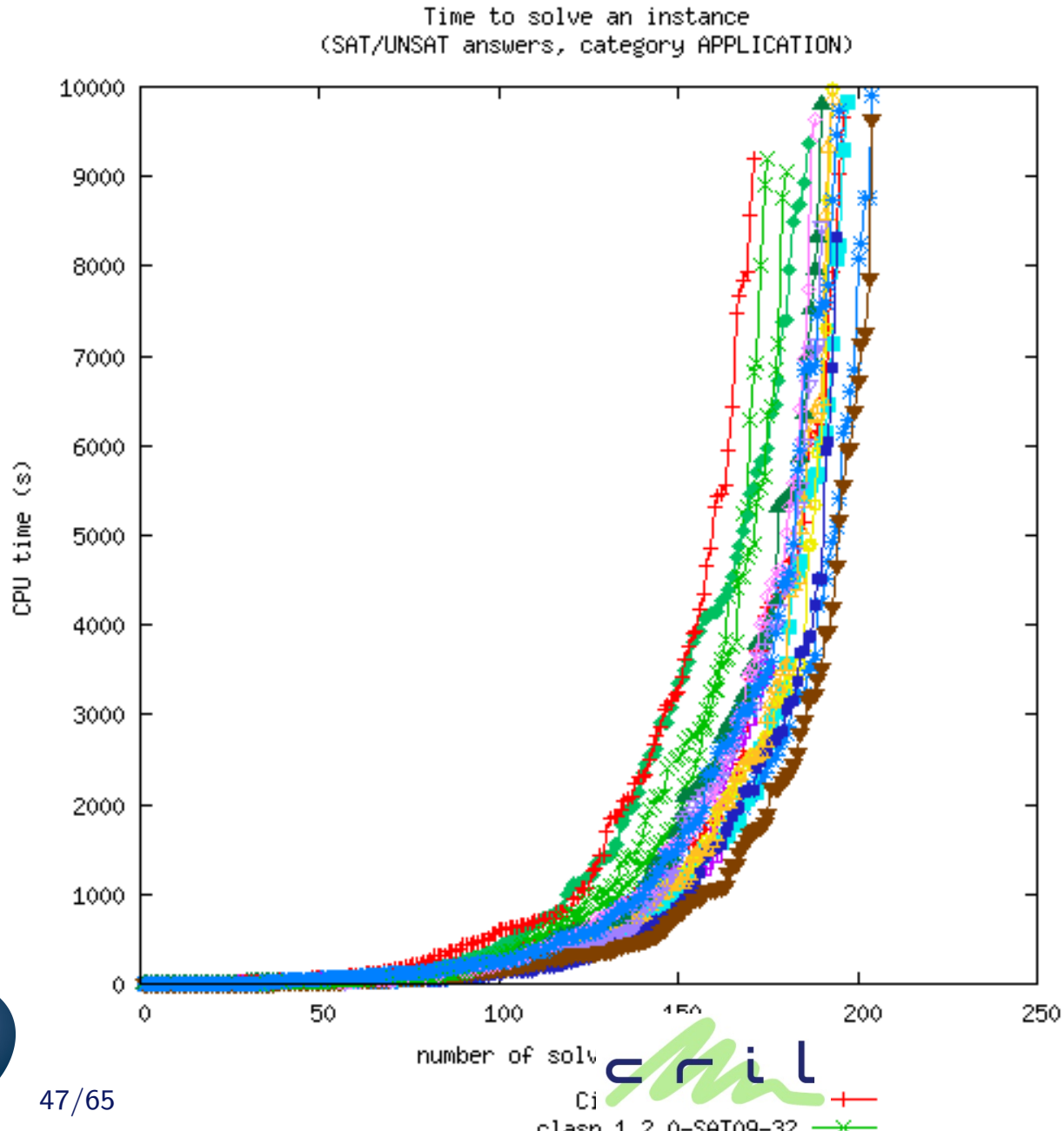
Finally

The results of the second stage !

Final results, Application, SAT+UNSAT

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	229	91	138	153127.06
1	precosat 236	204	79	125	180345.80
2	glucose 1.0	204	77	127	218826.10
3	LySAT i	197	73	124	198491.53
4	CircUs	196	77	119	229285.44
5	SATzilla2009_I 2009-03-22	195	81	114	234743.41
6	MiniSat 2.1 (Sat-race'08 Edition)	194	78	116	144548.45
7	ManySAT 1.1 aimd 1	193	71	122	173344.71
8	MiniSAT 09z	193	78	115	184696.75
9	MXC	190	79	111	180409.82
10	minisat_cumr p	190	75	115	206371.06
11	Rsat	188	74	114	187726.95
12	SApperloT base	186	78	108	282488.39
13	Rsat 2007-02-08	180	69	111	195748.38
14	kw	175	67	108	90213.34
15	clasp 1.2.0-SAT09-32	175	60	115	163460.74
16	picosat 535	171	76	95	209004.97

Cactus plot : application SAT+UNSAT



UNIVERSITÉ D'ARTOIS

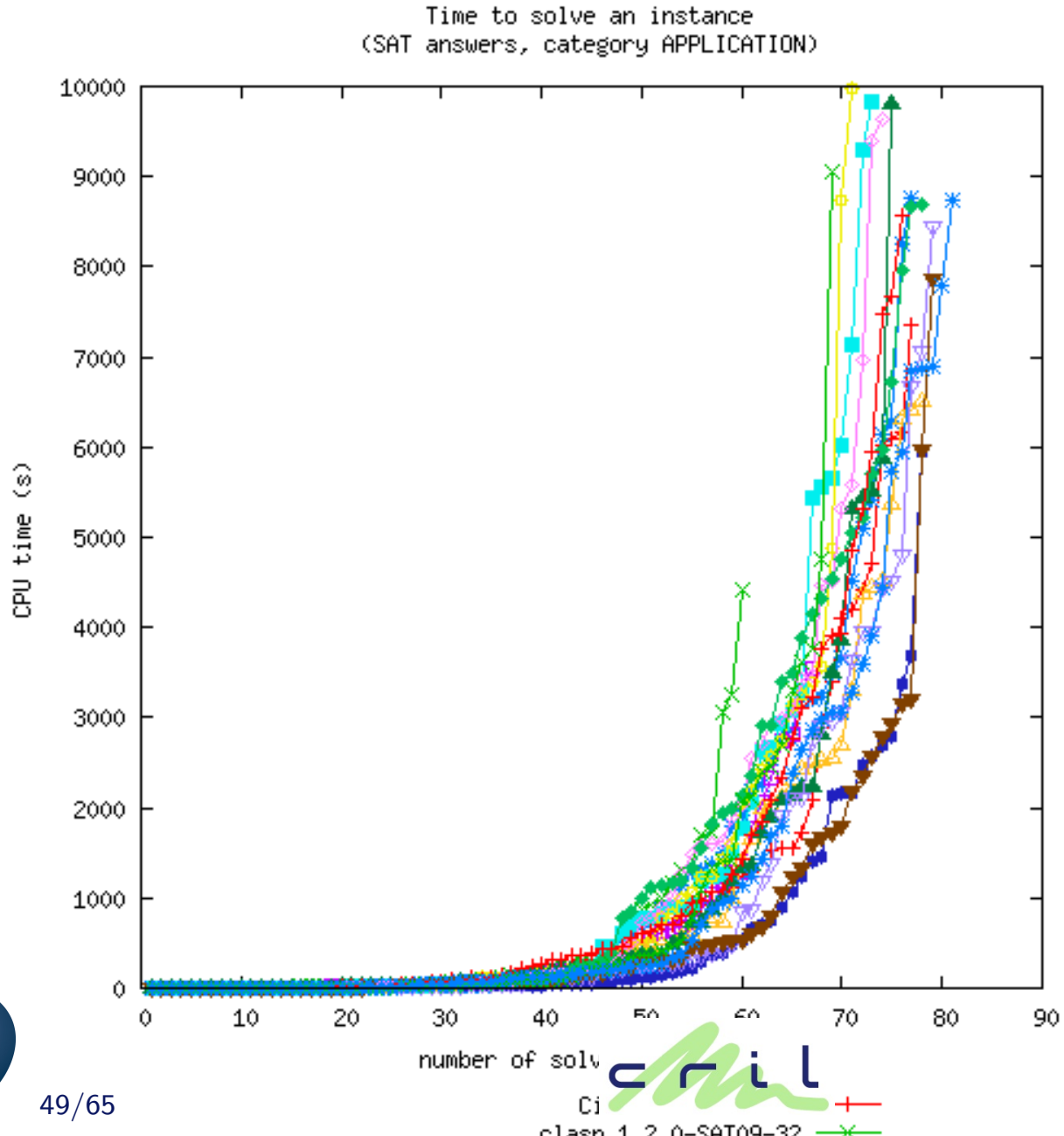
Final results, Application, SAT only

Rank	Solver	SAT	CPU Time
	Virtual Best Solver (VBS)	91	52336.24
1	SATzilla2009_I	81	96609.87
2	precosat 236	79	52903.18
3	MXC	79	75203.55
4	MiniSat 2.1 (Sat-race'08 Edition)	78	42218.37
5	MiniSAT 09z	78	75075.48
6	SApperloT base	78	111286.45
7	CircUs	77	74720.59
8	glucose 1.0	77	90532.72
9	picosat 535	76	84382.33
10	minisat_cumr p	75	67373.20
11	Rsat 2009-03-22	74	85363.26
12	LySAT i	73	81793.98
13	ManySAT 1.1 aimd 1	71	62994.30
14	Rsat 2007-02-08	69	47294.67
15	kw	67	31254.87
16	clasp 1.2.0-SAT09-32	60	25529.94

Final results, Application, SAT only

Rank	Solver	SAT	CPU Time
	Virtual Best Solver (VBS)	91	52336.24
1	SATzilla2009_I	81	96609.87
2	precosat 236	79	52903.18
3	MXC	79	75203.55
4	MiniSat 2.1 (Sat-race'08 Edition)	78	42218.37
5	MiniSAT 09z	78	75075.48
6	SApperloT base	78	111286.45
7	CircUs	77	74720.59
8	glucose 1.0	77	90532.72
9	picosat 535	76	84382.33
10	minisat_cumr p	75	67373.20
11	Rsat 2009-03-22	74	85363.26
12	LySAT i	73	81793.98
13	ManySAT 1.1 aimd 1	71	62994.30
14	Rsat 2007-02-08	69	47294.67
15	kw	67	31254.87
16	clasp 1.2.0-SAT09-32	60	25529.94

Cactus plot : application SAT (timeout matters !)

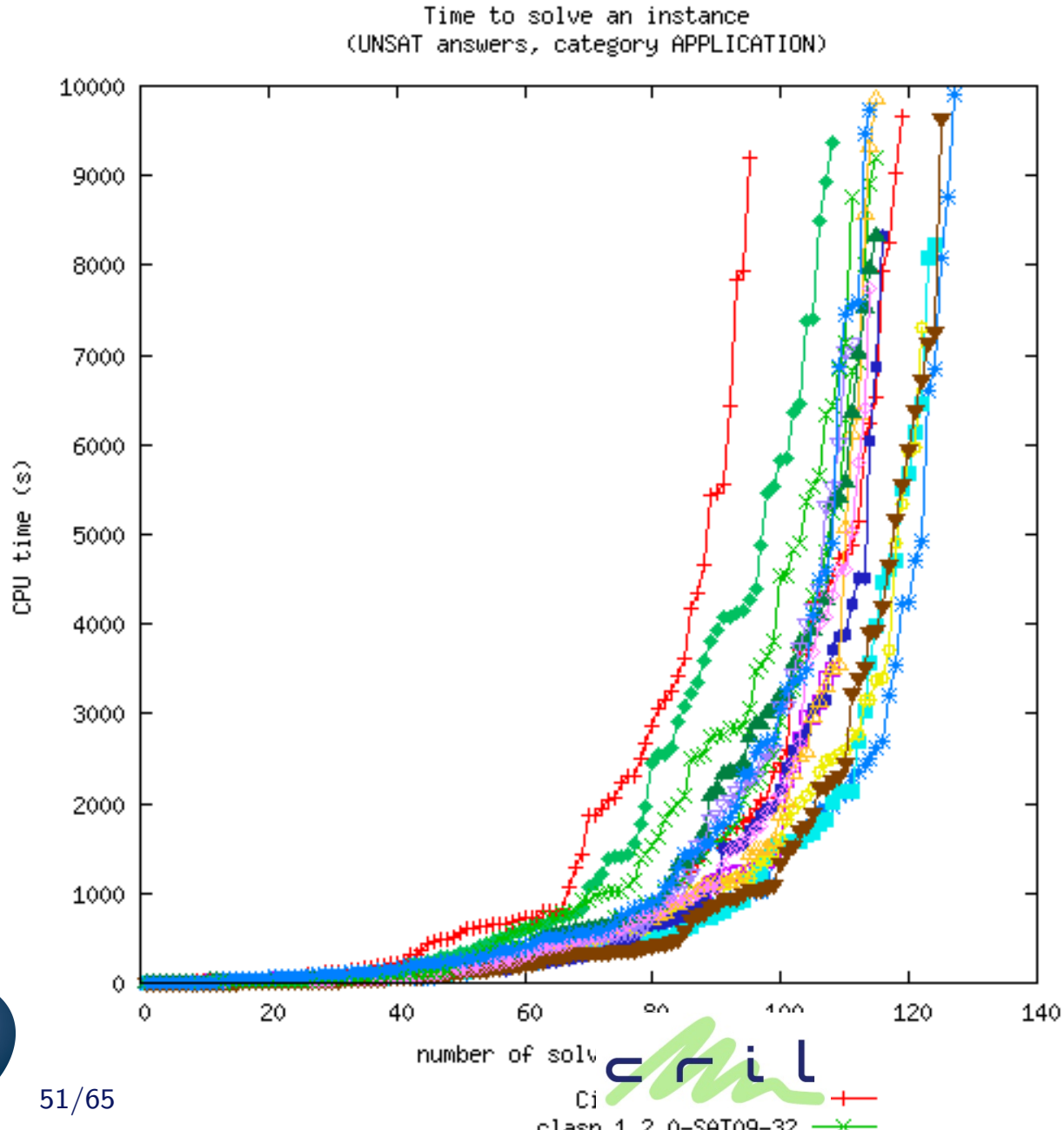


UNIVERSITÉ D'ARTOIS

Final results, Application, UNSAT only

Rank	Solver	UNSAT	CPU Time
	Virtual Best Solver (VBS)	138	100790.82
1	glucose 1.0	127	128293.39
2	precosat 236	125	127442.62
3	LySAT i	124	116697.55
4	ManySAT 1.1 aimd 1	122	110350.41
5	CircUs	119	154564.85
6	MiniSat 2.1 (Sat-race'08 Edition)	116	102330.08
7	MiniSAT 09z	115	109621.27
8	clasp 1.2.0-SAT09-32	115	137930.80
9	minisat_cumr p	115	138997.86
10	Rsat	114	102363.69
11	SATzilla2009_I	114	138133.54
12	MXC	111	105206.27
13	Rsat 2007-02-08	111	148453.71
14	kw 2009-03-20	108	58958.47
15	SApperloT base	108	171201.93
16	picosat 535	95	124622.64

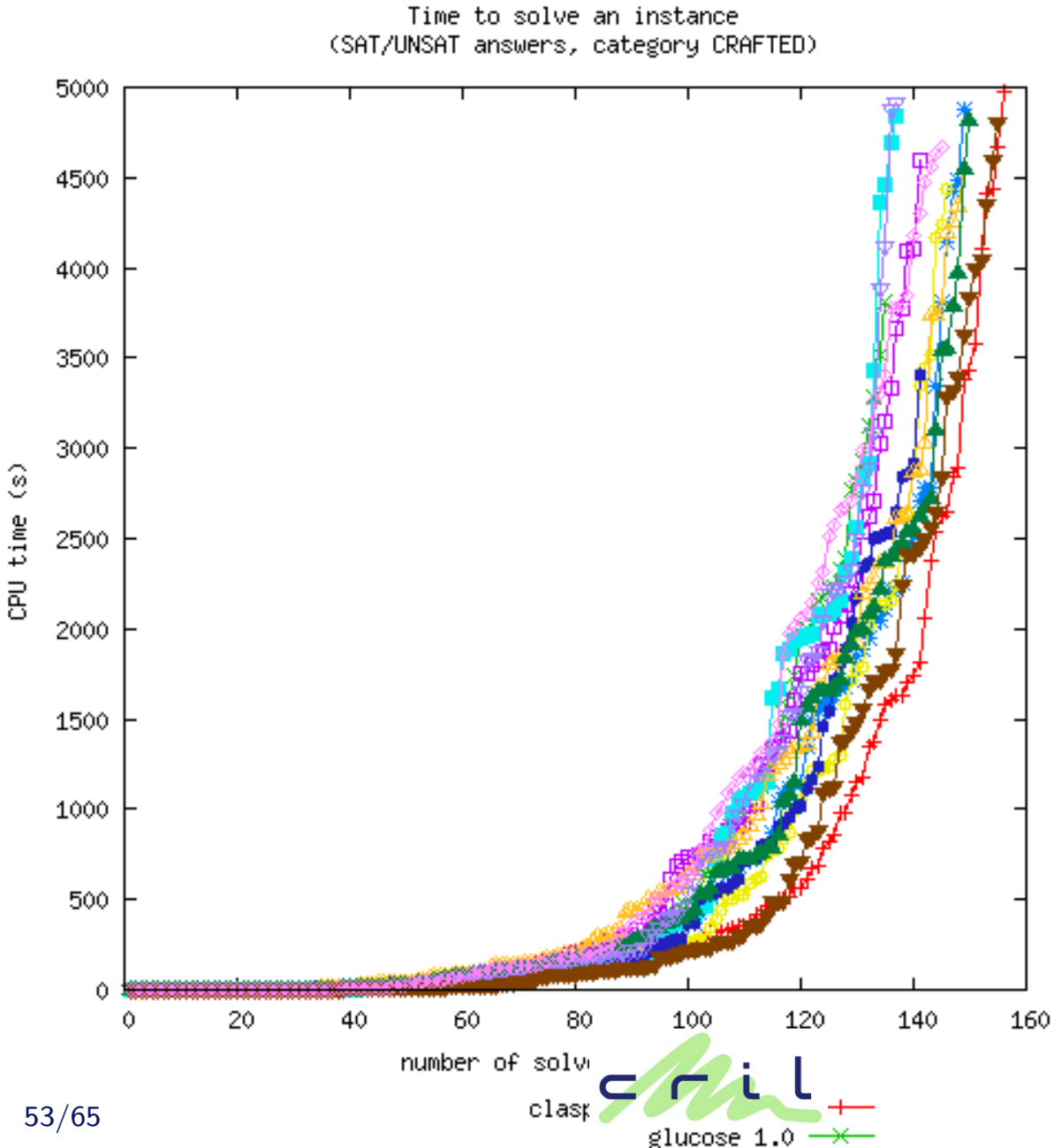
Cactus plot : application UNSAT (timeout matters !)



Final results, Crafted, SAT+UNSAT

Rank	Solver	Total	SAT	UNSAT	CPU Time
	Virtual Best Solver (VBS)	187	108	79	62264.60
1	clasp 1.2.0-SAT09-32	156	92	64	89194.49
2	SATzilla2009_C	155	83	72	94762.27
3	minisat SAT 2007	150	90	60	99960.89
4	IUT_BMB_SAT 1.0	149	89	60	93502.16
5	SApperloT base	149	92	57	108298.52
6	MXC	146	91	55	76965.59
7	VARSAT-industrial	145	85	60	119365.13
8	precosat 236	141	90	51	66318.44
9	LySAT c	141	83	58	89925.84
10	SATzilla CRAFTED	137	84	53	76856.90
11	MiniSat 2.1 (Sat-race'08 Edition)	137	87	50	78381.80
12	glucose 1.0	135	86	49	70385.63

Cactus plot : crafted SAT+UNSAT

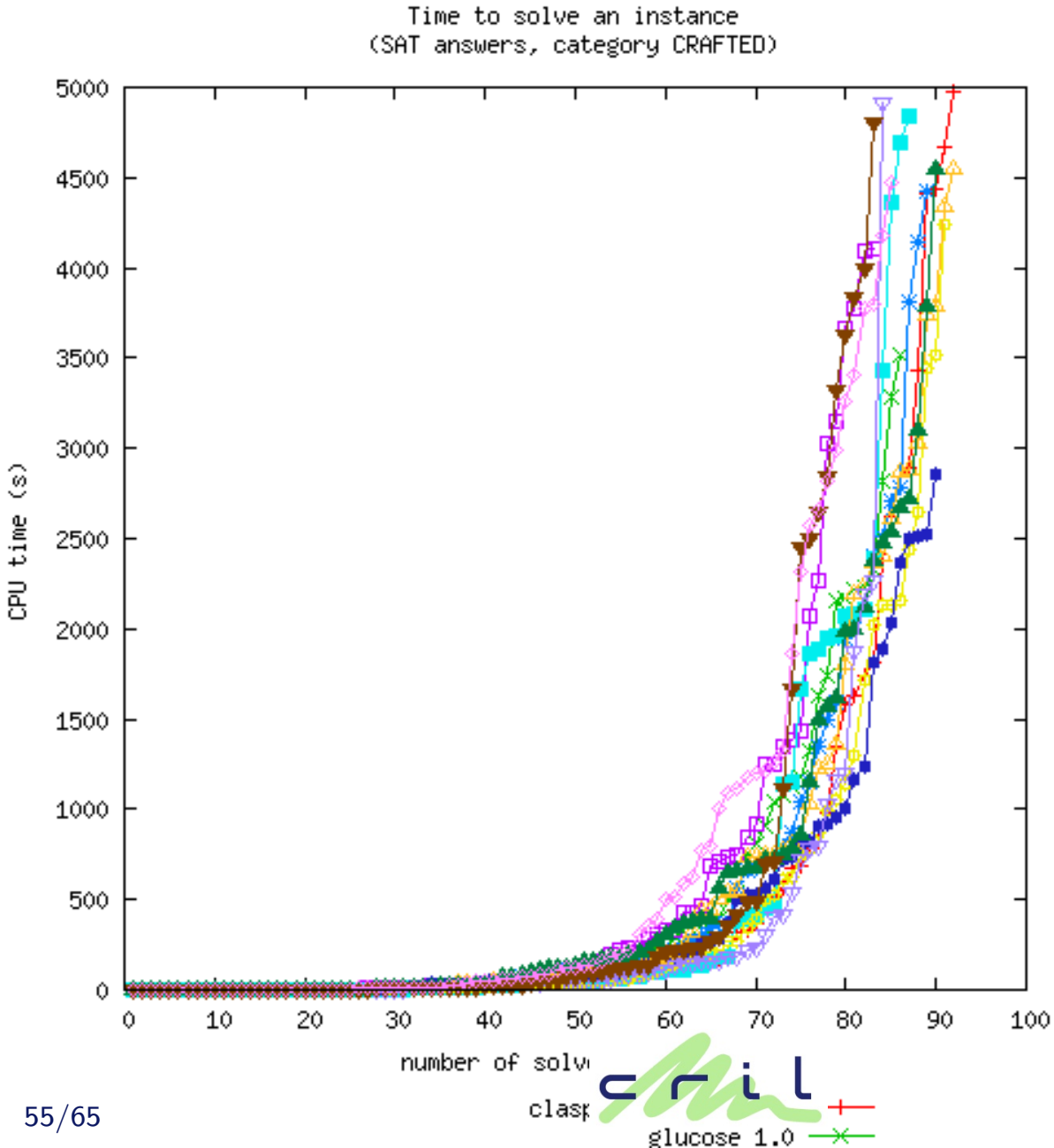


UNIVERSITÉ D'ARTOIS

Final results : crafted SAT only

Rank	Solver	SAT	CPU Time
	Virtual Best Solver (VBS)	108	21224.84
1	clasp 1.2.0-SAT09-32	92	49775.04
2	SApperloT base	92	54682.14
3	MXC 2009-03-10	91	39227.16
4	precosat 236	90	34447.16
5	minisat SAT 2007	90	48346.20
6	IUT_BMB_SAT 1.0	89	45287.01
7	MiniSat 2.1 (Sat-race'08 Edition)	87	41994.77
8	glucose 1.0	86	37779.61
9	VARSAT-industrial	85	54521.77
10	SATzilla CRAFTED	84	21726.48
11	SATzilla2009_C	83	39383.44
12	LySAT c	83	42073.80

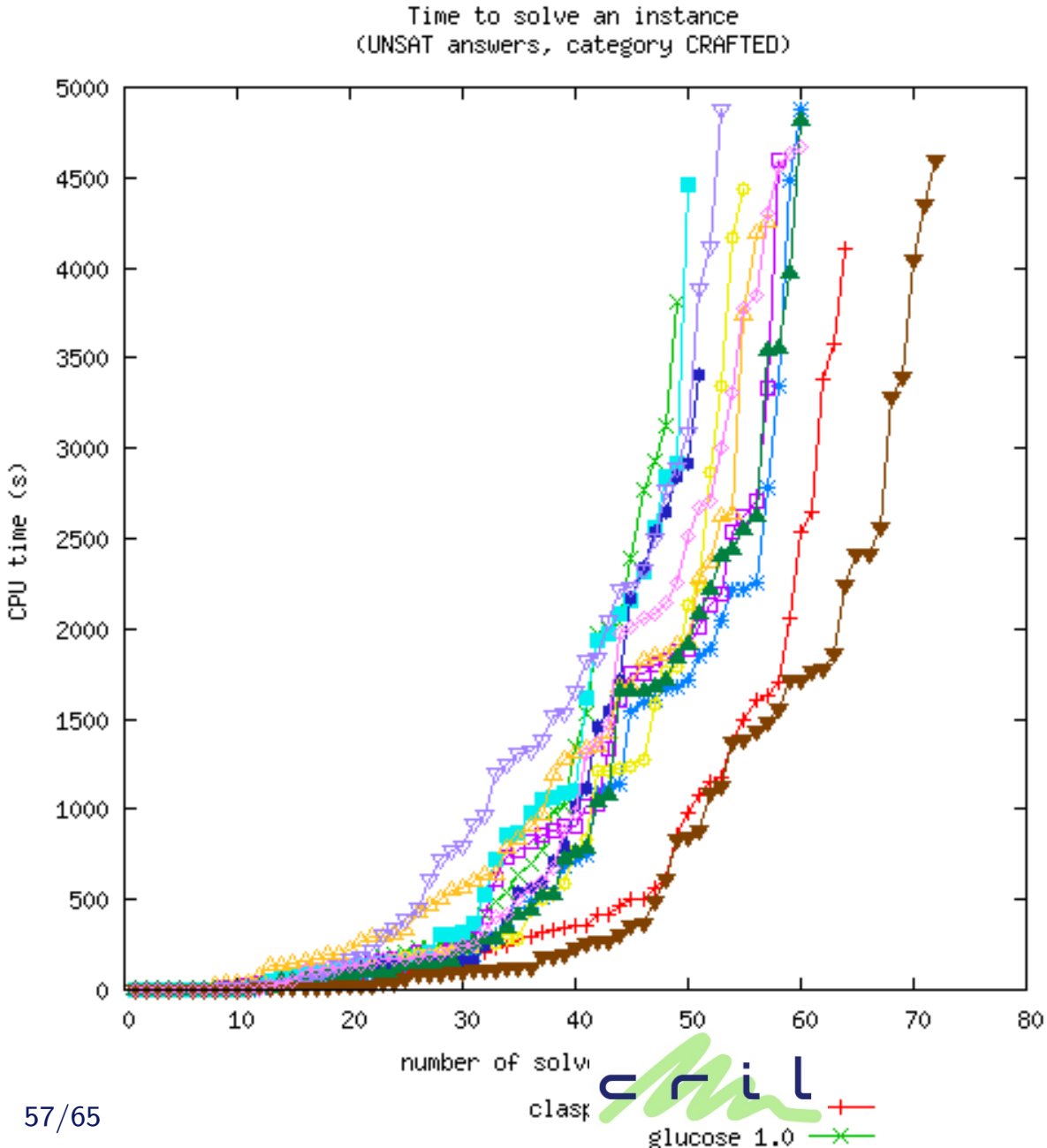
Cactus plot : crafted SAT only



Final results : crafted UNSAT only

Rank	Solver	UNSAT	CPU Time
	Virtual Best Solver (VBS)	79	41039.76
1	SATzilla2009_C	72	55378.83
2	clasp 1.2.0-SAT09-32	64	39419.45
3	IUT_BMB_SAT 1.0	60	48215.14
4	minisat SAT 2007	60	51614.69
5	VARSAT-industrial	60	64843.36
6	LySAT c	58	47852.03
7	SApperloT base	57	53616.38
8	MXC	55	37738.43
9	SATzilla CRAFTED	53	55130.42
10	precosat 236	51	31871.28
11	MiniSat 2.1 (Sat-race'08 Edition)	50	36387.03
12	glucose 1.0	49	32606.02

Cactus plot : crafted UNSAT only

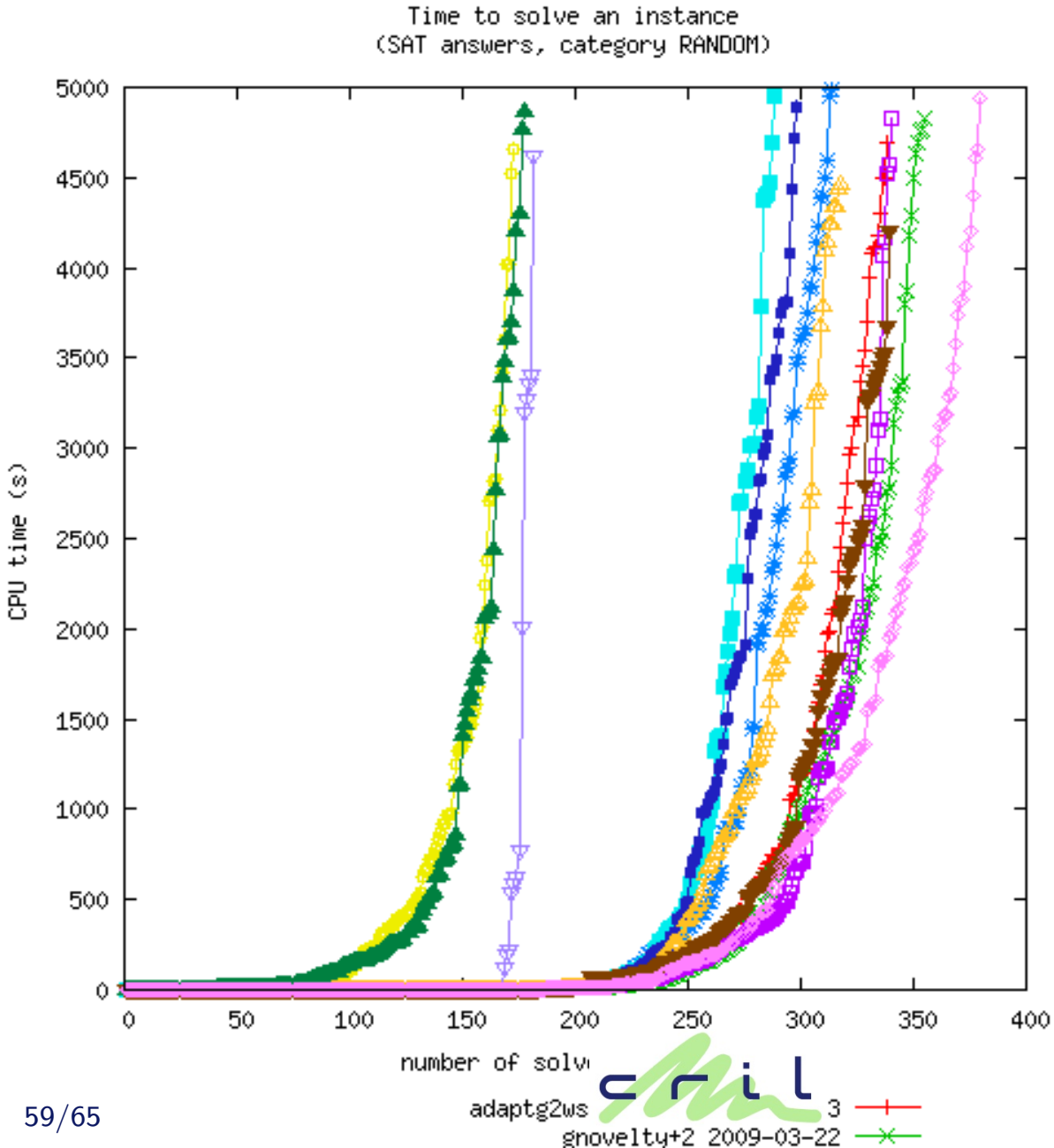


UNIVERSITÉ D'ARTOIS

Random, SAT (420/380 benchmarks)

Rank	Solver	SAT	CPU Time
	Virtual Best Solver (VBS)	404 /371	97656.83
1	TNM	379/353	194780.22
2	gnovelty+2	355/352	154503.93
3	hybridGM3	340/309	101986.32
4	SATzilla2009_R	339/335	122158.36
5	adaptg2wsat2009++	338/337	133641.90
6	gnovelty+ 2007-02-08	318/311	130357.30
7	gNovelty+-T	314/309	143439.69
8	adaptg2wsat+ 2007-02-08	298	117302.89
9	iPAWS	288	93855.93
10	SATzilla RANDOM	181	23793.38
11	March KS 2007-02-08	177	98629.25
12	march_hi	173	90433.09

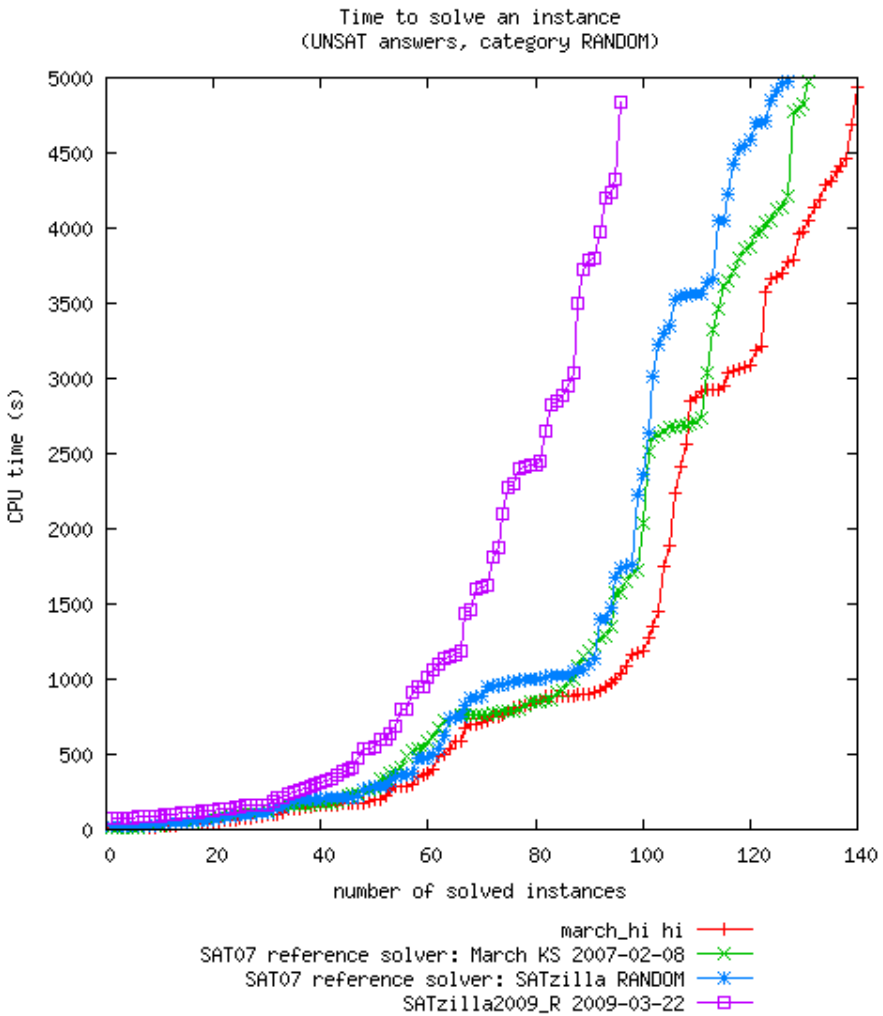
Cactus plot : random SAT only



Random, UNSAT/SAT+UNSAT

Rank	Solver	Total	SAT	UNSAT	CPU Time
SAT+UNSAT (610 benchmarks)					
1	SATzilla2009_R	435/431	339/335	96	231051.45
2	march_hi	313	173	140	261826.59
3	SATzilla RANDOM	308	181	127	186335.14
4	March KS 2007-02-08	308	177	131	258763.45
UNSAT					(190 benchmarks)
1	march_hi	-	-	140	171393.50
2	March KS 2007-02-08	-	-	131	160134.20
3	SATzilla RANDOM	-	-	127	162541.76
4	SATzilla2009_R	-	-	96	108893.09

Cactus plot : random UNSAT only



Awards Summary

Category	Gold	Silver	Bronze
Application			
SAT	SATzilla2009_I	precosat	MXC
UNSAT	glucose	precosat	Lysat
SAT+UNSAT	precosat	glucose	Lysat
Crafted			
SAT	clasp	SApperloT	MXC
UNSAT	SATzilla2009_C	clasp	IUT_BMB_SAT
SAT+UNSAT	clasp	SATzilla2009_C	IUT_BMB_SAT
Random			
SAT	TNM	gNovelty2+	hybridGM3 / adaptg2wsat2009++
UNSAT	March_hi	SATzilla2009_R	-
SAT+UNSAT	SATzilla2009_R	March_hi	-

Special prizes

ManySAT Best parallel SAT solver for application benchmarks

gNovelty+-T Best parallel solver for random benchmarks

Minisat 09z Best Minisat Hack solver

Summary of the competition

- Steady improvement since SAT competition 2007
- Huge improvements in SLS solvers (Random SAT)
- Many good solvers !
- Many new solvers awarded !
- The portfolio approach is quite accurate despite 55% of new benchmarks in the competition !
- Parallel SAT solving evaluation needs to be discussed !

Discussion

Is the competition good or bad for the community ?
It depends of the use of the results !

- independent public results available
- new benchmarks appear each year
- high visibility outside the community
- reward for solver designers

Main problem : how to spot good new idea ?

- purse based scoring ?
- time independent metrics ?
- submit also less mature solvers !

Provocative idea

Should we remove the random category to force incomplete solver designers to tune their approach to application benchmarks ?