FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics
TU WIEN

W|W|T|F

RiSE
Rigorous Systems Engineering

The QBF Gallery 2013: A Non-Competitive Evaluation of QBF Tools

Florian Lonsing, Martina Seidl, Allen Van Gelder

JKU
JOHANNES KEPLER UNIVERSITY LINZ

Overview

Recent Progress in QBF Solving:
- Powerful preprocessing techniques, new solvers.
- Certificate generation, new applications.

Purpose of the QBF Gallery 2013:
- Evaluate the state of the art in QBF solving and related aspects.
- No competition, no winners, no prizes: collect and analyze data.
- Community-driven organization, much interactions during runs.
- Ultimate goal: Identify promising research directions.

Solvers:
- Narizzano: + sq-qube3.0, Van Gelder: ○ hiqqer3, Janota: □ rareqs1.1
- Goultiaeva: ▼ ooq , ✳ dual_ooq , ▽ sDual_ooq
- Klieber: ● ghost , ■ ghost-CEGAR, × bGhost-CEGAR
- Lonsing: ▲ nenofex , △ DepQBF, ◇ DepQBF-lqpup

Considered Benchmarks and Tools:
- Four preprocessors, 14 CNF solvers.
- Three 2QBF-solvers (Bayless), one NNF-solver (Janota).
- Two tool suites for certificate generation and checking + 4 new sets.
- Benchmarks randomly selected from QBFLIB.
- New benchmarks from various application domains submitted by the participants.

Showcases:
- Preprocessing, Solving, Applications, Certificates.

Details:

`http://www.kr.tuwien.ac.at/events/qbfgallery2013/`

Some Excerpts of the Different Showcases

Preprocessing

	A Hiqqer3e			B Bloqqer			C Hiqqer3p			D SqueezeBF		
	solved	sat	unsat	solved	sat	unsat	solved	sat	unsat	solved	sat	unsat
eval2012r2	19	0	19	69	33	36	77	35	42	11	3	8
qbf-hardness	0	0	0	49	12	37	51	12	39	12	0	12
sauer-reimer	81	0	81	137	24	113	153	29	124	78	9	69
planning-CTE	0	0	0	3	2	1	7	6	1	0	0	0
conf.-planning	646	0	646	489	11	478	486	12	474	48	0	48
red.-finding	176	0	176	1496	837	659	1650	924	726	674	326	348

	eval2012r2												
	1	2	3	4	5	6	7	8	9	10	11	12	Σ
$(ABCD)^{12}$ fixpoints	3	25	62	64	38	5	4	2	0	2	0	1	206
$(ABCD)^{12}$ solved	93	9	3	1	1	1	0	0	0	0	0	0	108

- Hiqqer3e and Hiqqer3p by Allen Van Gelder.
- Bloqqer by Martina Seidl and Armin Biere.
- SqueezeBF by Massimo Narizzano.
- Evaluate individual preprocessors.
- Evaluate combinations of preprocessors.
- Time-limited incremental runs in multiple rounds, 120 seconds timeout per individual call of a preprocessor.
- Fixpoint detection.

	eval2012r2		
	solved	sat	unsat
$(A^2B^2C^2D^2)^6$	111	46	65
$(A^2B^2D^2C^2)^6$	111	45	66
$(A^2D^2B^2C^2)^6$	104	42	62
$(B^2C^2D^2A^2)^6$	103	42	61
$(D^2A^2B^2C^2)^6$	102	38	64
A	19	0	19
B	69	33	36
C	77	35	42
D	11	3	8
$(ABCD)^6$	107	44	63
$A+B+C+D$	87	36	51

Solving: Sampling Experiment

- 7 trials, 8 solvers, 1900 different formulas.
- Timeout per formula: 200s.
- Anonymized solver names.
- Stratification: 6 samples per family.
- Future work:
  - Longer timeouts, different sampling.

Seven trials ranked by number solved (out of 455)

BB	312	BB	317	BB	316	BB	315	BB	315	BB	310	BB	319
HH	293	HH	298	HH	300	HH	298	HH	291	HH	290	HH	299
AA	284	AA	285	AA	274	AA	272	AA	279	AA	275	AA	284
CC	274	CC	270	CC	265	CC	270	CC	270	CC	273	CC	274
DD	241	DD	237	DD	236	DD	242	DD	232	DD	241	DD	239
EE	211	EE	211	EE	205	EE	216	EE	208	EE	215	EE	211
GG	204	GG	201	GG	200	GG	202	GG	194	GG	194	GG	207
FF	159	FF	162	FF	161	FF	162	FF	159	FF	156	FF	171

Seven trials ranked by PAR10 score on 455 instances

BB	289789	BB	279394	BB	281633	BB	284161	BB	283213	BB	293765	BB	275484
HH	329163	HH	317815	HH	315260	HH	319013	HH	331889	HH	333963	HH	316156
AA	345379	AA	342510	AA	365017	AA	368546	AA	355440	AA	362225	AA	344501
CC	363203	CC	370937	CC	381240	CC	371138	CC	370969	CC	365632	CC	363675
DD	430920	DD	438119	DD	440782	DD	428847	DD	448715	DD	430942	DD	434698
EE	490114	EE	489906	EE	502105	EE	480172	EE	495984	EE	482456	EE	490329
GG	504751	GG	510314	GG	513311	GG	509006	GG	524914	GG	524489	GG	498584
FF	593122	FF	587675	FF	589655	FF	587346	FF	593564	FF	599697	FF	569816

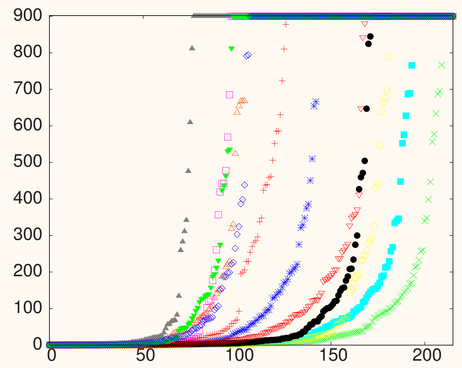
Certificates

- ResQu by Valeriy Balabanov and Jie-Hong Roland Jiang.
- QBFcert by Aina Niemetz and Mathias Preiner.
- Resource limits: 600 seconds, 3 GB memory.
- Upper part: officially submitted tools.
  - Certification failures due to resource limits only.
- Lower part: selected publicly available tools.

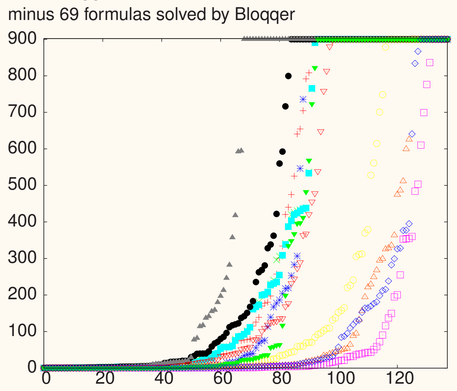
	eval2012r2	
Certification Workflow	Solved	Certified
DepQBF and QBFcert	91 (34 s, 57 u)	67 (20 s, 47 u)
DepQBF and ResQu	91 (34 s, 57 u)	63 (22 s, 41 u)
sKizzo and ozziKs	88 (36 s, 52 u)	35 (35 s, 0 u)
squolem and qbv	38 (19 s, 19 u)	38 (19 s, 19 u)
squolem and ResQu	38 (19 s, 19 u)	19 (0 s, 19 u)
qube-cert and checker	80 (25 s, 55 u)	32 (11 s, 21 u)
qube-cert and ResQu	80 (25 s, 55 u)	52 (17 s, 35 u)

Solving (Benchmarks of Eval 2012 Round 2)

Without Bloqqer:
345 formulas



With Bloqqer
minus 69 formulas solved by Bloqqer



- Observation: Preprocessing strongly influences solvers, for some it’s good, for others not.
- Idea: let solver choose if it wants preprocessing.
- Preprocessor: Bloqqer

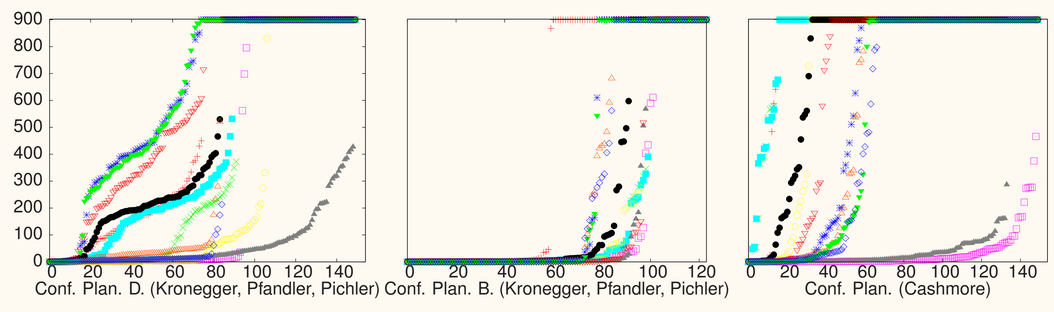
Virtual Experiment:

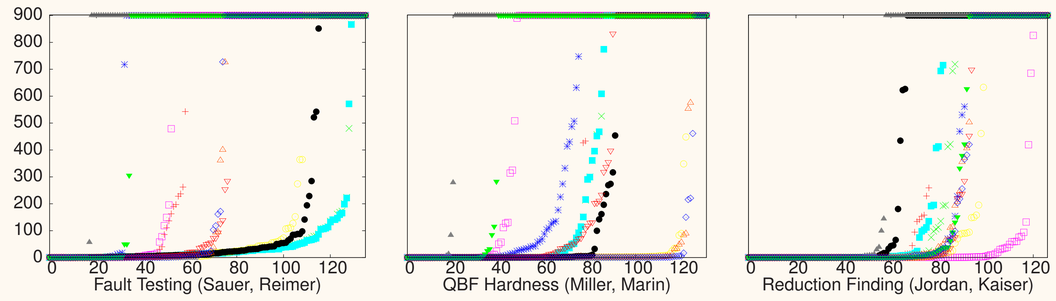
WANT Bloqqer		
	Number Solved	
Catgy./ Solver	Best Foot	Worst Foot
DepQBF	125	86
DepQBF-lqpup	127	88
nenofex	73	50
ooq	111	65
rareqs1.1	132	79

NO Bloqqer		
	Number Solved	
Catgy./ Solver	Best Foot	Worst Foot
dual_ooq_ooq	105	106
bGhost-CEGAR	142	92
ghost-CEGAR	142	92
ghost	122	85
hiqqer3	113	118
sDual_ooq	118	97
sq-qube3.0	90	91

Applications

- New benchmarks (not in QBFLIB).
- Sample size: 150.
- Timeout: 900 sec, Memory limit: 7 GB.
- Observation: behavior of solvers depends on set.





Conclusions

Lessons Learned:
- If this had been a competition, there would not be a clear winner.
- Preprocessing strongly influences solving.
- QBF solvers are not blackboxes, some use built-in preprocessing.
- Benchmark selection and scoring methods strongly influence rankings.
- Community-driven organization is challenging, but fruitful.

What’s next?
- More analysis of the available data.
- Establish fair benchmark sets for competitions and evaluations.
- Tighter integration of certificate generation.
- Common standards for input formats and testing workflows.
- Challenging long-term goal: bringing QBF to practice.

This work was supported by the Vienna Science and Technology Fund (WWTF) through project ICT10-018 and by the Austrian Science Foundation (FWF) under grants S11408-N23 and S11409-N23.