

Fifth QBF solvers competition (QBFEVAL'12)

Massimo Narizzano

Laboratory of Systems and Technologies for Automated Reasoning (STAR-Lab)
Department of Communication, Computer and System Sciences (DIST)
University of Genoa



QBFEVAL Evolution

Years	#cores	speed(GHZ)	RAM(MB)	#Solvers
2003	8	2.4	512	11(11,0,0)
2004	8	2.4	512	16(4,8,4)
2005	10	3.2	1024	13(5,4,4)
2006	10	3.2	1024	21(7,9,5)
2007	10	3.2	1024	18(12,6,0)
2008	18	2.13	4096	10(8,2,0)
2010	18	2.13	4096	13(9,4,0)

(new,updated, no change)



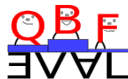
Outline

- 1 The Setup
- 2 Data Collection and Check
- 3 Running the Evaluation
- 4 Checking the Results
- 5 Results
- 6 Conclusion and Future Work



Outline

- 1 The Setup
- 2 Data Collection and Check
- 3 Running the Evaluation
- 4 Checking the Results
- 5 Results
- 6 Conclusion and Future Work



The Setup

Data Collection and Check

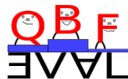
Running the Evaluation

Checking the Results

Results

Conclusion and Future Work

The old infrastructure



The old infrastructure

“cluster” used in 2003 and 2004



The old infrastructure

“cluster” used in 2003 and 2004



cluster used from 2005 till 2007



The old infrastructure

“cluster” used in 2003 and 2004



cluster used from 2005 till 2007



“cluster” used in 2008 and 2010



Infrastructure

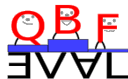


- Cineca is a non profit Consortium, made of 54 Italian universities, MIUR, CNR and others
- It is the largest Italian computing center, one of the most important worldwide.
- Star-lab won a grant for running the QBFEVAL'12



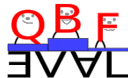
Infrastructure

- IBM pSeries 575 in an Infiniband Cluster



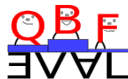
Infrastructure

- IBM pSeries 575 in an Infiniband Cluster
- 168 Computing Nodes, each one with 16 chip, each chip is dual core



Infrastructure

- IBM pSeries 575 in an Infiniband Cluster
- 168 Computing Nodes, each one with 16 chip, each chip is dual core
- 5376 cores 4.7 GHZ



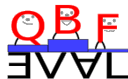
Infrastructure

- IBM pSeries 575 in an Infiniband Cluster
- 168 Computing Nodes, each one with 16 chip, each chip is dual core
- 5376 cores 4.7 GHZ
- RAM : 21Tera -> 128 GB per Node -> 4 GB per core



Infrastructure

- IBM pSeries 575 in an Infiniband Cluster
- 168 Computing Nodes, each one with 16 chip, each chip is dual core
- 5376 cores 4.7 GHZ
- RAM : 21Tera -> 128 GB per Node -> 4 GB per core



Infrastructure

- IBM pSeries 575 in an Infiniband Cluster
- 168 Computing Nodes, each one with 16 chip, each chip is dual core
- **5376** cores 4.7 GHZ
- RAM : 21Tera -> 128 GB per Node -> 4 GB per core



“Known” issues

- Porting from the “old” Intel x86/x86_64 architecture to the new IBM Power 6 architecture is not **easy**
- The claimed “Linux affinity” is far from been real
- No backward compatibility, only maintained solvers can participate



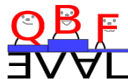
Main Rules

- Participants should provide the source code or a compiled *xcoff64* executable for AIX
- Solvers may assume that formulas are closed QBF
- Formulas must be compliant with QDIMACS
- Formula selection from QBFLIB will be done on the basis of several parameters, mostly based on the results of past QBFEVALs.
- Two tracks: Main and 2QBF
- Correctness? no certificates, majority voting as a temporary measure



Outline

- 1 The Setup
- 2 Data Collection and Check
- 3 Running the Evaluation
- 4 Checking the Results
- 5 Results
- 6 Conclusion and Future Work



Data Collected

- 11 new solvers submitted by 5 developers



Data Collected

- 11 new solvers submitted by 5 developers
- 3 of them accepting only 2QBFs as input



Data Collected

- 11 new solvers submitted by 5 developers
- 3 of them accepting only 2QBFs as input
- 3 new families submitted



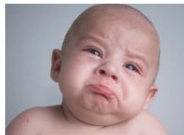
Data Collected

- 11 new solvers submitted by 5 developers
- 3 of them accepting only 2QBFs as input
- 3 new families submitted
- 1512 new formulas submitted



Data Collected

- 11 new solvers submitted by 5 developers
- 3 of them accepting only 2QBFs as input
- 3 new families submitted
- 1512 new formulas submitted



Solvers Collected

Solver	Submitter
Hiqqr3	Allen Van Gelder
bloqqr_rareqs_uubph3	mikolas janota
ghostq-bq-ceg	will klieber
ghostq-cegar	will klieber
ghostq-plain	will klieber
rareqs_uubph3	mikolas janota
free2qbf	Sam Bayless
mini2qbf	Sam Bayless
pico2qbf	Sam Bayless
depqbf	Florian Lonsing
bdepqbf	Florian Lonsing

2QBF track



Formulas Collected

New Benchmarks

Suites	Submitted by	#
BooM-NP	Jie-Hong R. Jiang	654
BooM-P	Jie-Hong R. Jiang	718
Lights3_035_qdgz	Allen Van Gelder	140



Formulas Collected

New Benchmarks

Suites	Submitted by	#
BooM-NP	Jie-Hong R. Jiang	654
BooM-P	Jie-Hong R. Jiang	718
Lights3_035_qdgz	Allen Van Gelder	140

Benchmarks from QBFLIB Repository

Suites	Submitted by	#
2010 Miscellanea	various authors	507
2010 2QBF	various authors	200



Checks

Formulas Compliance

- Compliance with QDIMACS
- Closed QBF
- No more than 10% of the total test set comes from a submitter also authoring a competitive solver.

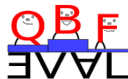
Solvers Compliance

- Compliance with input/output command line
- Single thread solver



Benchmark's checking

CATEGORY	QDIMACS	Closed	Number	%
2QBF	OK	OK	200	9.91
2010	OK	OK	507	25.11
BooM-NP	OK	OK	654	32.39
BooM-P	OK	OK	718	35.56
Lights3_035_qdgz	OK	OK	140	6.93



Solver's checking

The check needs to be done on the sp6 machine



Solver's checking

The check needs to be done on the sp6 machine

login as XYZ@sp6.cineca



Solver's checking

The check needs to be done on the sp6 machine

login as XYZ@sp6.cineca

Welcome back XYZ to the CINECA system, please make a choice.

bla bla bla

bla bla bla

bla bla bla

- * Please notice the system will be **shut-down** at any time in between
- * the 10th to the 15th of May, please consider to not run anything after the 10th.

bla bla bla

bla bla bla

bla bla bla

XYZ@sp6>



Solver's checking

The check needs to be done on the sp6 machine

```
login as XYZ@sp6.cineca ....
```

```
Welcome back XYZ to the CINECA system, please make a choice.
```

```
bla bla bla .....
```

```
bla bla bla .....
```

```
bla bla bla .....
```

```
*****
```

```
* Please notice the system will be shut-down at any time in between
```

```
* the 10th to the 15th of May, please consider to not run anything after the 10th.
```

```
*****
```

```
bla bla bla .....
```

```
bla bla bla .....
```

```
bla bla bla .....
```

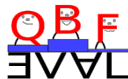
```
XYZ@sp6>
```

This at the end of April



Solver's checking

“Well from 45 days, down to 15 days ... It is still possible!”

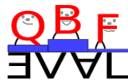


Solver's checking

“Well from 45 days, down to 15 days ... It is still possible!”

Murphy Law

“If anything can go wrong, it will”



Solver's checking

“Well from 45 days, down to 15 days ... It is still possible!”

Murphy Law

“If anything can go wrong, it will”

Many solvers present segmentation faults

This was mainly due to the different architecture.



Solver's checking

“Well from 45 days, down to 15 days ... It is still possible!”

Murphy Law

“If anything can go wrong, it will”

Many solvers present segmentation faults

This was mainly due to the different architecture.

Fixing the solvers required extra days.



Outline

- 1 The Setup
- 2 Data Collection and Check
- 3 Running the Evaluation**
- 4 Checking the Results
- 5 Results
- 6 Conclusion and Future Work



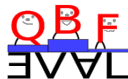
QBFEVAL'12 in numbers

- time limit fixed to 1200 seconds
- memory limit fixed to 2GB
- Two tracks with 2219 benchmarks:
 - Main Tracks: All but 2QBF solvers, all benchmarks
 - 2QBF Tracks: All the solvers, only 2QBF benchmarks
- The competition ran for more than 5000 h, in approximately 8 days.
- We used more than 30 cores in parallel at the time



Outline

- 1 The Setup
- 2 Data Collection and Check
- 3 Running the Evaluation
- 4 Checking the Results**
- 5 Results
- 6 Conclusion and Future Work



Discrepancies

BENCHMARK	SAT	UNSAT	QBFLIB
aim-100-1_6-yes1-3-90-shuffled	7	1	SAT
sortnetsort7.AE.stepl.003	6	1	SAT
sortnetsort5.AE.stepl.003	9	1	SAT



Discrepancies

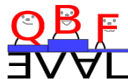
BENCHMARK	SAT	UNSAT	QBFLIB
aim-100-1_6-yes1-3-90-shuffled	7	1	SAT
sortnetsort7.AE.stepl.003	6	1	SAT
sortnetsort5.AE.stepl.003	9	1	SAT

Hiqqr3



Digging the Data

Few Solvers were reporting unusual exit code



Digging the Data

Few Solvers were reporting unusual exit code

Murphy (second) Law

“If everything seems to be going well, you have obviously overlooked something”



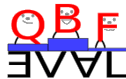
Digging the Data

Few Solvers were reporting unusual exit code

Murphy (second) Law

“If everything seems to be going well, you have obviously overlooked something”

BooM-P and BooM-NP, were not compliant with QDIMACS.



Digging the Data

Few Solvers were reporting unusual exit code

Murphy (second) Law

“If everything seems to be going well, you have obviously overlooked something”

BooM-P and BooM-NP, were not compliant with QDIMACS.
(yes ... there was a bug in the checking script)



Digging the Data

Few Solvers were reporting unusual exit code

Murphy (second) Law

“If everything seems to be going well, you have obviously overlooked something”

BooM-P and BooM-NP, were not compliant with QDIMACS.
(yes ... there was a bug in the checking script)

Benchmarks could be fixed ... unlikely the system shut-down



Digging the Data

Few Solvers were reporting unusual exit code

Murphy (second) Law

“If everything seems to be going well, you have obviously overlooked something”

BooM-P and BooM-NP, were not compliant with QDIMACS.
(yes ... there was a bug in the checking script)

Benchmarks could be fixed ... unlikely the system shut-down
BooM-P and BooM-NP families has been removed



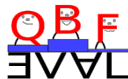
Outline

- 1 The Setup
- 2 Data Collection and Check
- 3 Running the Evaluation
- 4 Checking the Results
- 5 Results**
- 6 Conclusion and Future Work



Best 2012 solvers

<u>Place</u>	<u>Solver</u>	<u># Solved</u>
--------------	---------------	-----------------



Best 2012 solvers

Place	Solver	# Solved
-------	--------	----------





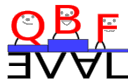
ghostq-cegar

582






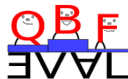
Best 2012 solvers

Place	Solver	# Solved
	ghostq-bq-ceg	664
	ghostq-cegar	582



Best 2012 solvers

Place	Solver	# Solved
	bdepqbf	711
	ghostq-bq-ceg	664
	ghostq-cegar	582



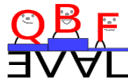
Results for Main track

Solver	#			TIME(h)		
	Tot	SAT	UNSAT	Tot	SAT	UNSAT
Hiqqr3	737	413	324	6.6	4.83	1.77
bdepqbf	711	414	297	5.38	4.3	1.07
ghostq-bq-ceg	664	367	297	2.94	1.2	1.73
ghostq-cegar	582	297	285	3.38	1.54	1.84
ghostq-plain	515	268	247	3.44	1.66	1.78
depqbf	400	181	219	6.21	4.84	1.37
rareqs_uubph3	283	119	164	1.56	0.89	0.67
bloqqr_rareqs_uubph3	283	119	164	1.64	0.89	0.75



Best 2012 solvers on 2QBF

<u>Place</u>	<u>Solver</u>	<u># Solved</u>
--------------	---------------	-----------------



Best 2012 solvers on 2QBF

Place	Solver	# Solved
--------------	---------------	-----------------





ghostq-cegar

111






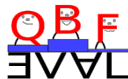
Best 2012 solvers on 2QBF

Place	Solver	# Solved
	bdepqbf	156
	ghostq-cegar	111



Best 2012 solvers on 2QBF

Place	Solver	# Solved
	ghostq-bq-ceg	179
	bdepqbf	156
	ghostq-cegar	111



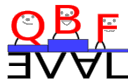
Results for 2QBF track

Solver	#			TIME(m)		
	Tot	SAT	UNSAT	Tot	SAT	UNSAT
Hiqqr3	179	121	58	84.31	82.93	1.38
ghostq-bq-ceg	179	126	53	935.78	920.44	15.34
bdepqbf	156	124	32	165.4	162.69	2.71
ghostq-cegar	113	60	53	909.87	894.95	14.92
mini2qbf	91	68	23	937.33	921.96	15.37
ghostq-plain	87	38	49	1226.11	1206.01	20.1
pico2qbf	86	66	20	1297.23	1275.96	21.27
free2qbf	60	50	10	93.38	91.85	1.53
bloqqr_rareqs_uubph3	32	24	8	716.77	705.01	11.75
rareqs_uubph3	32	24	8	731.47	719.48	11.99
depqbf	26	20	6	1132.78	1114.21	18.57



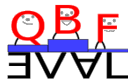
Outline

- 1 The Setup
- 2 Data Collection and Check
- 3 Running the Evaluation
- 4 Checking the Results
- 5 Results
- 6 Conclusion and Future Work



Conclusion

Bad



Conclusion

Bad

- The new powerful infrastructure



Conclusion

Bad

- The new powerful infrastructure
- Need for certificates



Conclusion

Bad

- The new powerful infrastructure
- Need for certificates

Good

- QBF competition is still attracting new people



Conclusion

Bad

- The new powerful infrastructure
- Need for certificates

Good

- QBF competition is still attracting new people
- New powerful infrastructure



Conclusion

Bad

- The new powerful infrastructure
- Need for certificates

Good

- QBF competition is still attracting new people
- New powerful infrastructure
- The competition will be re-run after the conference as soon the new Infrastructure will be ready



Thank you!

