

MODEL CHECKING CONTEST

RESULTS FOR 2014

Fabrice Kordon - LIP6, Univ. P. & M. Curie, France

Hubert Garavel - Inria/Laboratoire d'Informatique de Grenoble, France

Lom-Messan Hillah, LIP6, Univ. P. & M. Curie & Univ. Paris Ouest, France

Alban Linard - LSV, Inria/École Normale Supérieure de Cachan, France

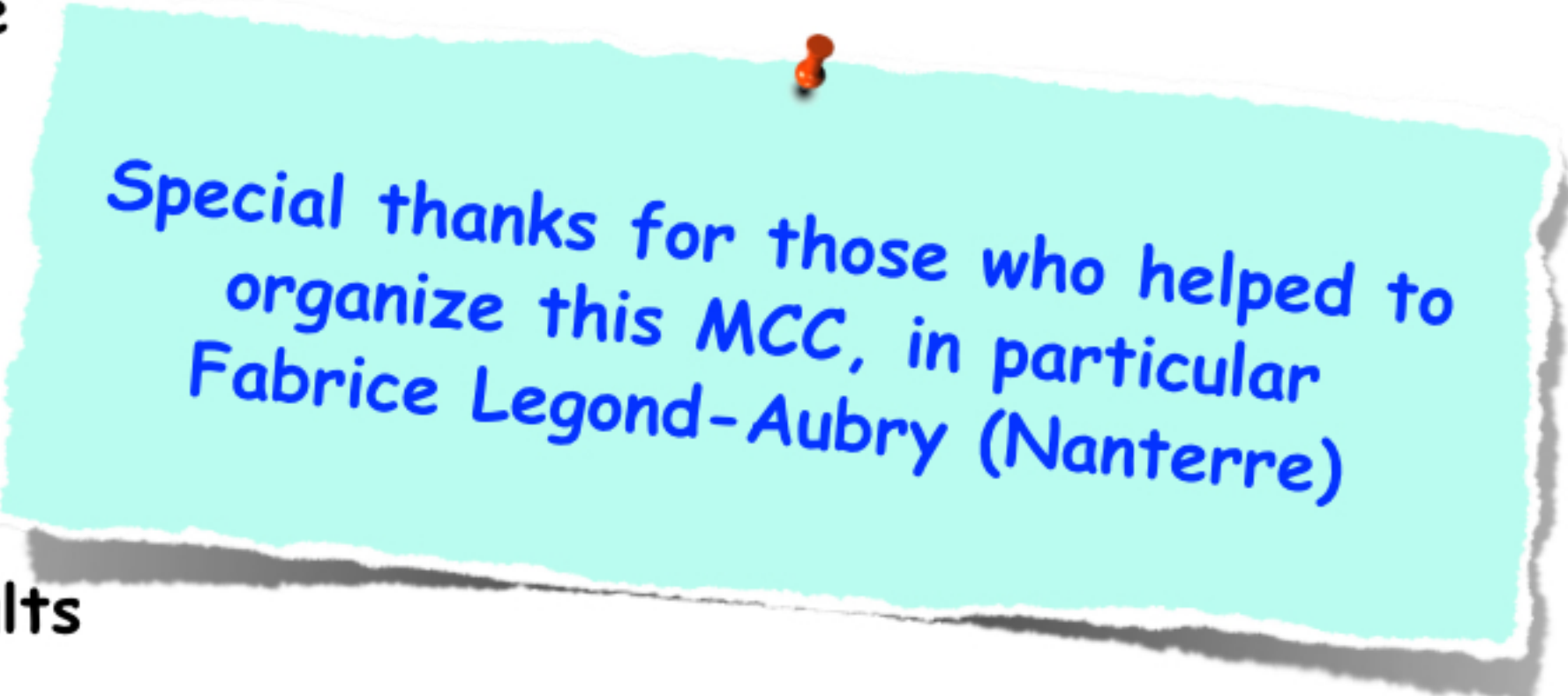
Francis Hulin-Hubard - LSV, CNRS/École Normale Supérieure de Cachan, France

MCC
2014

Model Checking Contest @



- 🏆 Objectives
- 🏆 Evaluation procedure
- 🏆 The models
- 🏆 Participating tools
- 🏆 Analysis of the results
- 🏆 Concluding remarks





Special thanks for those who helped to organize this MCC, in particular Fabrice Legond-Aubry (Nanterre)

**MCC
2014**

OBJECTIVES

 Lots of questions are raised...

 To verify highly concurrent systems, should we use a symmetry-based or a partial order-based model checker?

 For models with large variable domains, should we use decision diagram-based, or a symmetry-based model checker?

 Can we combine structural reductions techniques with partial-order ones or symmetry-based ones?

 How do tools evolve in the community?

 ...

 A large variety of model checking techniques

 and their potential combination

 A large variety of model categories

 A challenge with large scale specifications **and variety of models**

 **A need to evaluate in the fairest way current MC implementations**

MCC is intended to:

- Exchange experience between tool programmers,
- Imagine some association of techniques, and thus better tools
- Stimulate development of tools
- Provide visibility to these tools

MCC can also be of great help for the PN community (and users):

- Define a common set of models for benchmarks
- Identify experimentally classes of problems (in models)
 - identify the techniques able to cope with a given class of problems...
- Improve communication between tools (and PNML ;-)
- Provides raw data for comparison

This is the fourth edition

- Stabilized evaluation procedure (BenchKit) + better reproducibility
- Enriched Benchmark...
- ...still elements to be improved



**MCC
2014**

**EVALUATION
PROCEDURE**

The «enemies» of model checking

- Memory consumption
- CPU consumption

13 classes of «Examinations» to be processed

- State space generation
- Formula evaluation
 - Reachability Formulas (6 subclasses)
 - CTL formulas (3 subclasses)
 - LTL formulas (3 subclasses)
- Sub-categories in formulas



Hard time in some cases

	Bounds	Compute bounds	Deadlock	Cardinality	Fireability simple	Fireability
Reachability	✓	✓	✓	✓	✓	✓
CTL	✗	✗	✗	✓	✓	✓
LTL	✗	✗	✗	✓	✓	✓

	Quadhexa-2	Rented machine	Total
Cores	24 @ 2.4GHz	40 @ 2.5GHz	
Memory	128 GB	256 GB	
Used Cores	21	39	
number of runs	48 395	34 913	83 308
total computation time	157 days 0 hours 10 minutes 51 seconds		
time spend	8 days 21 hours 50 minutes 29 seconds		
Boot time cumulated	Approx 19 days 4 hours of CPU (for 20s boot-time)		

BenchKit-2

- Powerful tool to operate a large
- See &ACSD net friday

A Big Benchmark



	Quadhexa-2	Rented machine	Total
Cores	24 @ 2.4GHz	40 @ 2.5GHz	
Memory	128 GB	256 GB	
Used Cores	21	39	
number of runs	48 395	34 913	83 308
total computation time	157 days 0 hours 10 minutes 51 seconds		
time spend	8 days 21 hours 50 minutes 29 seconds		
Boot time	Approx 19 days 4 hours of CPU (for 20s boot-time)		

Thanks!

BenchKit

Powerful generate a large
See & ACS Friday

quadhexa-2 lent by
Univ. Paris Ouest Nanterre

A Big Benchmark



	Quadhexa-2	Rented machine	Total
Cores	24 @ 2.4GHz	40 @ 2.5GHz	
Memory	128 GB	256 GB	
Used Cores	21	39	
number of runs	48 395	34 913	83 308
total computation time	157 days 0 hours 10 minutes 51 seconds		
time spend	8 days 21 hours 50 minutes 29 seconds		
Boot time	Approx 19 days 4 hours of CPU (for 20s boot-time)		

Thanks!

BenchKit

Powerful generate a large
See & ACS Friday

server rented

thanks to the ImPro ANR project

A Big Benchmark



- 🏆 **First of all correct results !!!**
 - 🎤 How to be sure?
 - 🎤 Compare results (we had time this year)
- 🏆 **Basic algorithm «per line»**
 - 🎤 One line = all results for a model/scale value
 - 🎤 The good one has the majority
 - 🎤 Undecided when no majority
 - 🎤 Comparison between «known» and «scrambled» models
- 🏆 **Points are counted only if the result is true!**
- 🏆 **Interesting side effect**
 - 🎤 Developers have a detailed feed-back on their tool capabilities and



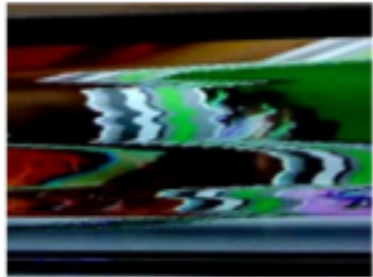
**MCC
2014**

THE MODELS

28 «known» models (7 from 2011 + 12 from 2012 + 9 from 2013)

28 presented in the original disk image for submission

28 «scrambled» (exactly the same but without tool specific information)



15 new «Surprise models»

ENS de Cachan

Inria/LIG

ST microelectronics & Inria

Univ. Cottbus

Univ P. & M. Curie

Univ. Saarland

Univ. Utha

Each tool confronted to
670 execution
for each examination

**MCC
2014**

**PARTICIPATING
TOOLS**

 9 tools as for last year Only one version per tool

	Origin	VM published
Cunf	ENS de Cachan/ Univ. Oxford (France/UK)	✓
GreatSPN	Univ. Torino (Italy)	✓
Helena	Univ. Paris 13 (France)	✓
IoLA	Univ. Rostock (Germany)	✓
Marcie	Univ. Cottbus (Germany)	✓
PNMC	ISAE (Toulouse, France)	✓
PNXDD	Univ P. & M. Curie (France)	✓
Stratagem	Univ. Geneva (Switzerland)	✓
Tapaal	Univ. Aalborg (Denmark)	✓



May not be activated simultaneously



	Reported Techniques
Cunf	NET_UNFOLDING SAT_SMT
GreatSPN	DECISION_DIAGRAMS SYMMETRIES
Helena	EXPLICIT STUBBORN_SETS
IoLA	EXPLICIT STATE_COMPRESSION STUBBORN_SETS SYMMETRIES TOPOLOGICAL UNFOLDING_TO_PT
Marcie	DECISION_DIAGRAMS
PNMC	DECISION_DIAGRAMS
PNXDD	DECISION_DIAGRAMS TOPOLOGICAL
Stratagem	DECISION_DIAGRAMS TOPOLOGICAL
Tapaal	EXPLICIT STRUCTURAL_REDUCTION

**MCC
2014**

**ANALYSIS OF THE
RESULTS**

The HTML report is being mostly generated automatically!

- Outputs are HUGE
 - 2.7 Gbyte of csv + ASCII (306+ Kfiles)
 - Need for automated analysis

- Rating based on correctness of results
 - Computation of a «result mask»

- ✓ All points (may be honor)
- ✓ Only a part of the points
- ✓ No point

model	class	parameter	Tools' results									Computed data from the results		
			T1	T2	T3	T4	T5	T6	T7	T8	T9	Result	majority	
Model1	PT	2	TF?F?FF???	XF?F?FF???	-----	TX?X?XF???	-----	-----	-----	-----	-----	TF?F?FF???	TF?F?FF???	3303034000
Model1	PT	3	FTFTTFTFT?	FXFXFXFX?	-----	--X-TX-X-?	-----	-----	-----	-----	-----	FTFTTFTFT?	F?T?TFTT?	3232332320
Model1	PT	4	DNF	MOVF	-----	-?-?-T?-	-----	-----	-----	-----	-----	F?T?TFTT?	F?T?TFTT?	1010111201
Model1	PT	5	CC	MOVF	-----	-----?-	-----	-----	-----	-----	-----	T-F-F-TT?F	T?F?F?TT?F	1010101101
Model1	PT	6	CC	MOVF	-----	T--T-T----	-----	-----	-----	-----	-----	DNF	T??T?T???	1001010000
Model1	PT	7	CC	DNF	-----	-----	-----	-----	-----	-----	-----	DNF	??????????	0000000000
Model1	PT	8	CC	DNF	-----	-----	-----	-----	-----	-----	-----	DNF	??????????	0000000000
model2	PT	10	TTTTTTTTTT	TXXXTTXTXT	-----	TTTTTTTTTT	-----	-----	-----	-----	-----	TTTTTTTTTT	TTTTTTTTTT	4333443434
model2	PT	15	TTTTTTTTTT	TTTTXTTTT	-----	TTTTTTTTTT	-----	-----	-----	-----	-----	TTTTTTTTTT	TTTTTTTTTT	4444434444
model2	PT	20	TTTTTTTTTT	TXTTXXTTX	-----	TTTTTTTTTT	-----	-----	-----	-----	-----	TTTTTTTTTT	TTTTTTTTTT	4344434443
model2	PT	50	DNF	DNF	-----	TTTTTTTTTT	-----	-----	-----	-----	-----	TTTTTTTTTT	TTTTTTTTTT	2222222222
model2	PT	##	CC	MOVF	-----	TTTTTTTTTT	-----	-----	-----	-----	-----	TTTTTTTTTT	TTTTTTTTTT	2222222222

- A tool may be wrong between «known» and «scrambled»

Point gained by a tool

- Those for the examinations where
 - You get the good result (even partially)
 - Partially = you get a part of the points :-)
- The result can be confirmed
 - Your tool is not the only one to compete on this specific one
- No point provided when we cannot state on a given result
 - The algorithm could be improved
 - Consider a «memory» of «right tools»

Honor points

- Those gained when a tool provided a result...
- The result was not confirmed by another tool

Ranking does not include honor points

- These are indicative

Decided before computation was done

To avoid any influence

Applies when results are OK

An full successful emanation = 10 pts

- State space → size = 4pts, others = opts
- Deadlock Formulas → 5pts per result
- Other formulas → 10pts per result

Best tool (goes rarest when model have scaling parameter) — model basis

+2

Last value of scaling parameter for a model — model basis

+3

Fastest on a «line»

+1

Smallest memory foot print on a «line»

+1

Multiplicator coefficient

«known» = x1 — «scrambled» = x2 — «surprise» = x3

- 🏆 Las year, too many examination classes
 - 🎤 Not readable in general
- 🏆 2014, covering some use cases
 - 🎤 State Space
 - 🎤 Gathers StatsSpace
 - 🎤 Reachability
 - 🎤 Gathers ReachabilityComputeBounds ReachabilityBounds ReachabilityDeadlock ReachabilityCardinality ReachabilityFireabilitySimple ReachabilityFireability
 - 🎤 CTL
 - 🎤 Gathers CTLCardinality CTLFireabilitySimple CTLFireability
 - 🎤 LTL
 - 🎤 Gathers LTLCardinality LTLFireabilitySimple LTLFireability
- 🏆 So, 4 trophies
 - 🎤 We will see in fact 😇



Some information

Various information on Colored nets

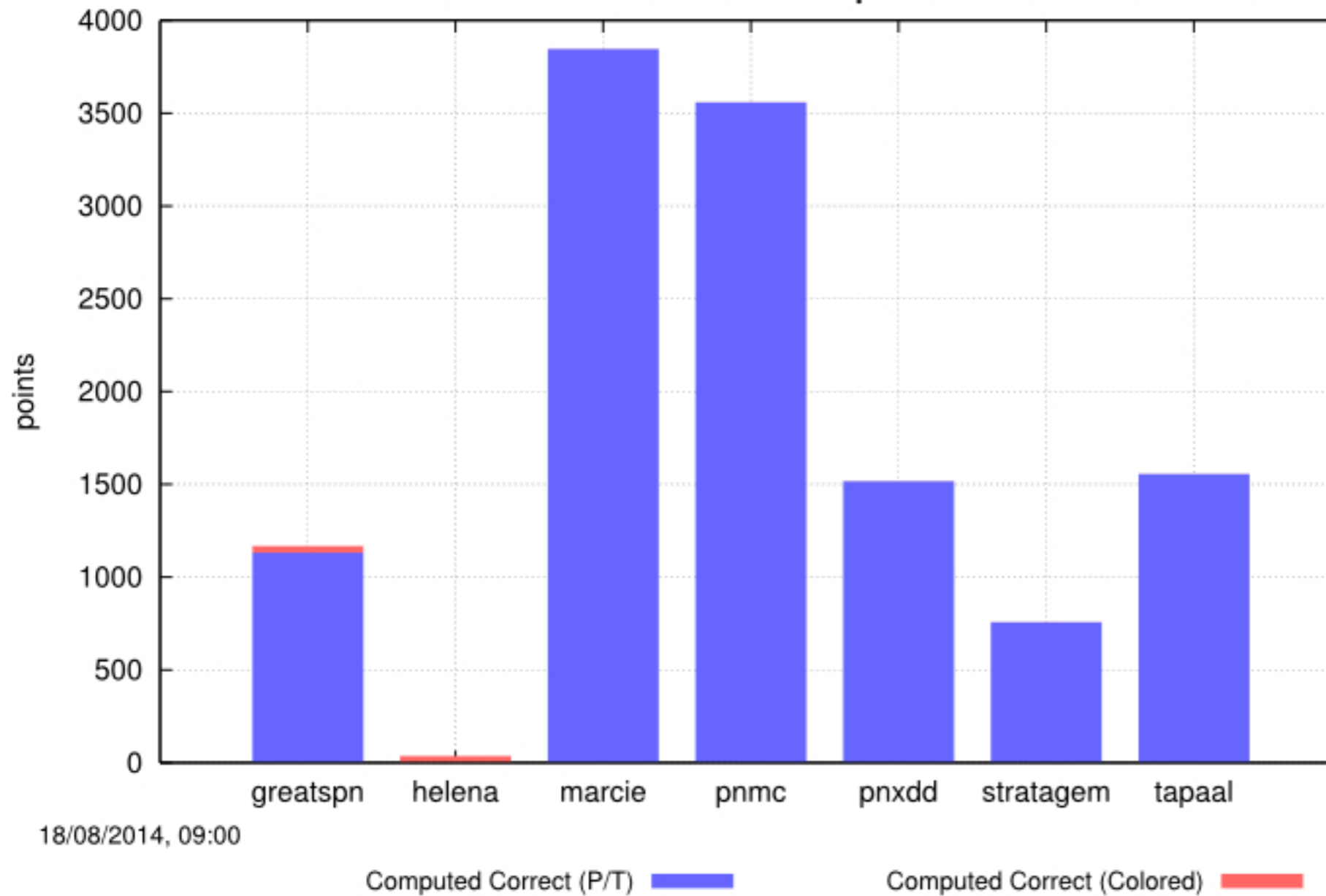
	cunf	greatspn	helena	lola	marcie	pnmc	pnxdd	stratagem	tapaal
number of errors	0	0	0	0	0	0	0	0	0
Alone to compute	0	7	48	0	0	0	0	0	0
Computed correctly	0	5	5	0	0	0	0	0	0
computed / models (Colored nets)	0.00%	4.98%	21.99%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
computed correctly / models (Colored nets)	0.00%	2.08%	2.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
computed alone / models (Colored nets)	0.00%	2.91%	19.91%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
computed erroneously / models (Colored nets)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Various information on P/T nets

	cunf	greatspn	helena	lola	marcie	pnmc	pnxdd	stratagem	tapaal
number of errors	0	54	0	0	0	4	0	0	0
Alone to compute	0	44	0	0	4	22	4	0	0
Computed correctly	0	83	0	0	211	218	191	119	103
computed / models (P/T nets)	0.00%	43.83%	0.00%	0.00%	52.05%	59.08%	47.22%	28.81%	24.94%
computed correctly / models (P/T nets)	0.00%	20.09%	0.00%	0.00%	51.09%	52.78%	46.25%	28.81%	24.94%
computed alone / models (P/T nets)	0.00%	10.66%	0.00%	0.00%	0.97%	5.33%	0.97%	0.00%	0.00%
computed erroneously / models (P/T nets)	0.00%	13.08%	0.00%	0.00%	0.00%	0.97%	0.00%	0.00%	0.00%



Tool classification for StateSpace examinations



Marcie



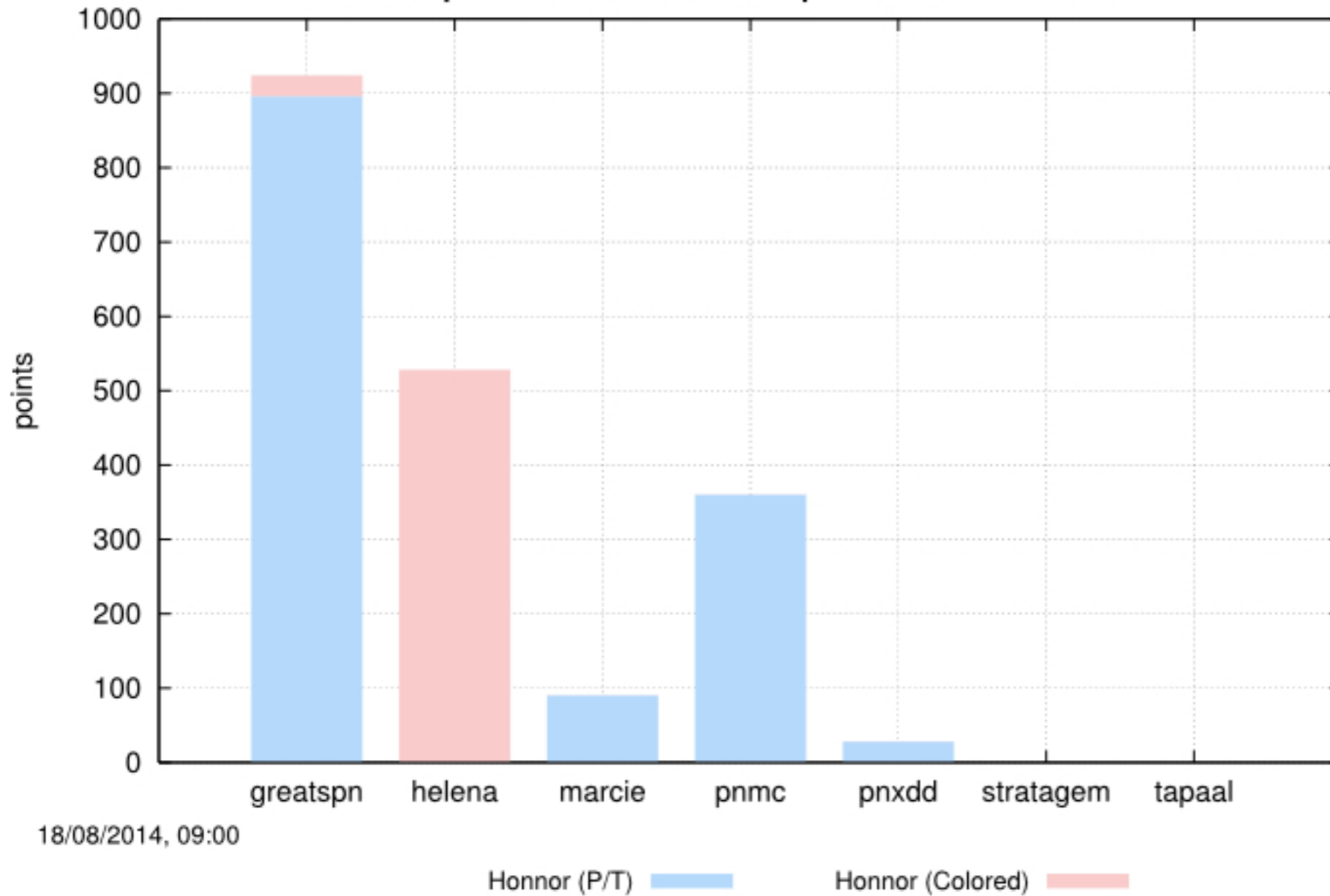
PNMC



Tapaal



Honor points for StateSpace examinations



18/08/2014, 09:00



Marcie



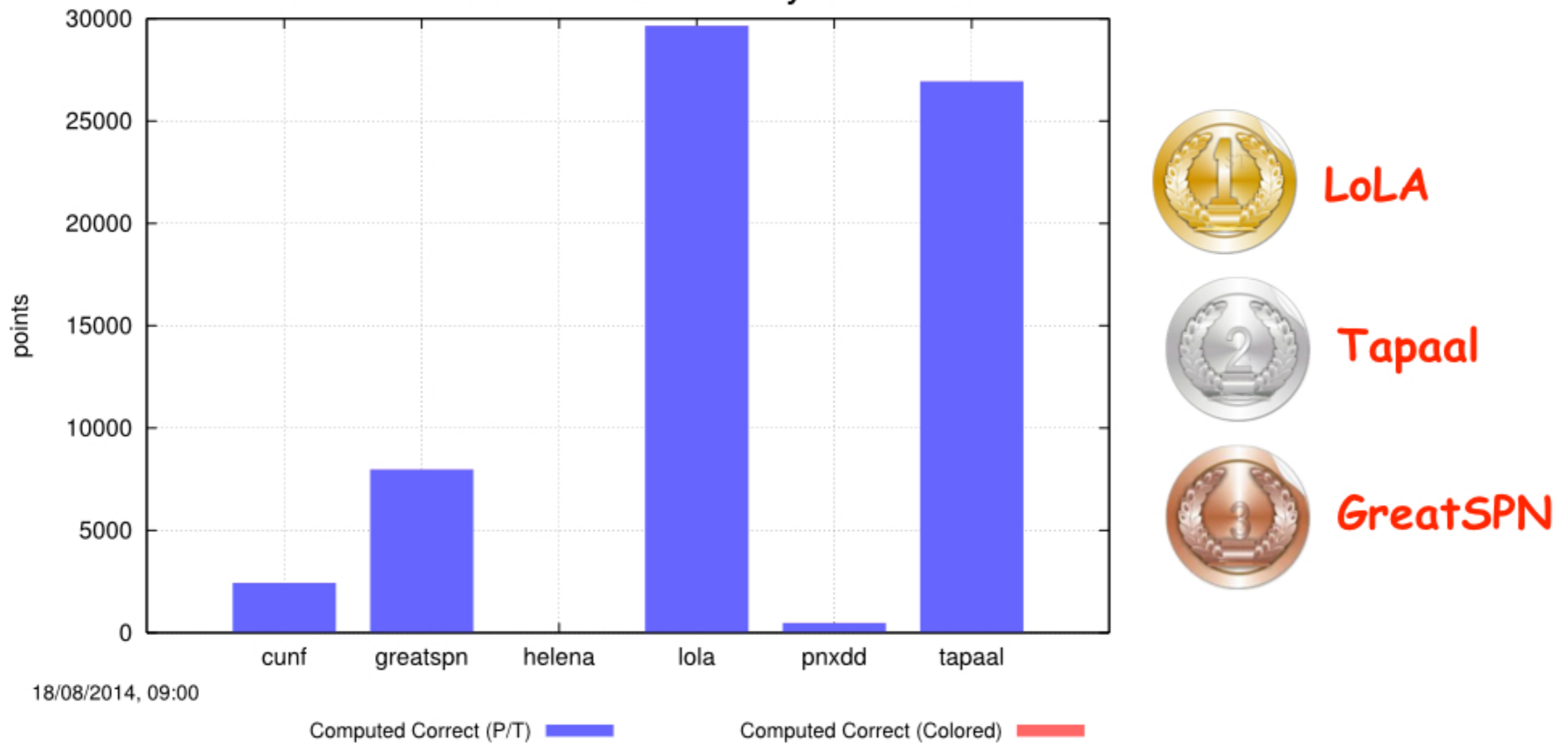
PNMC



Tapaal

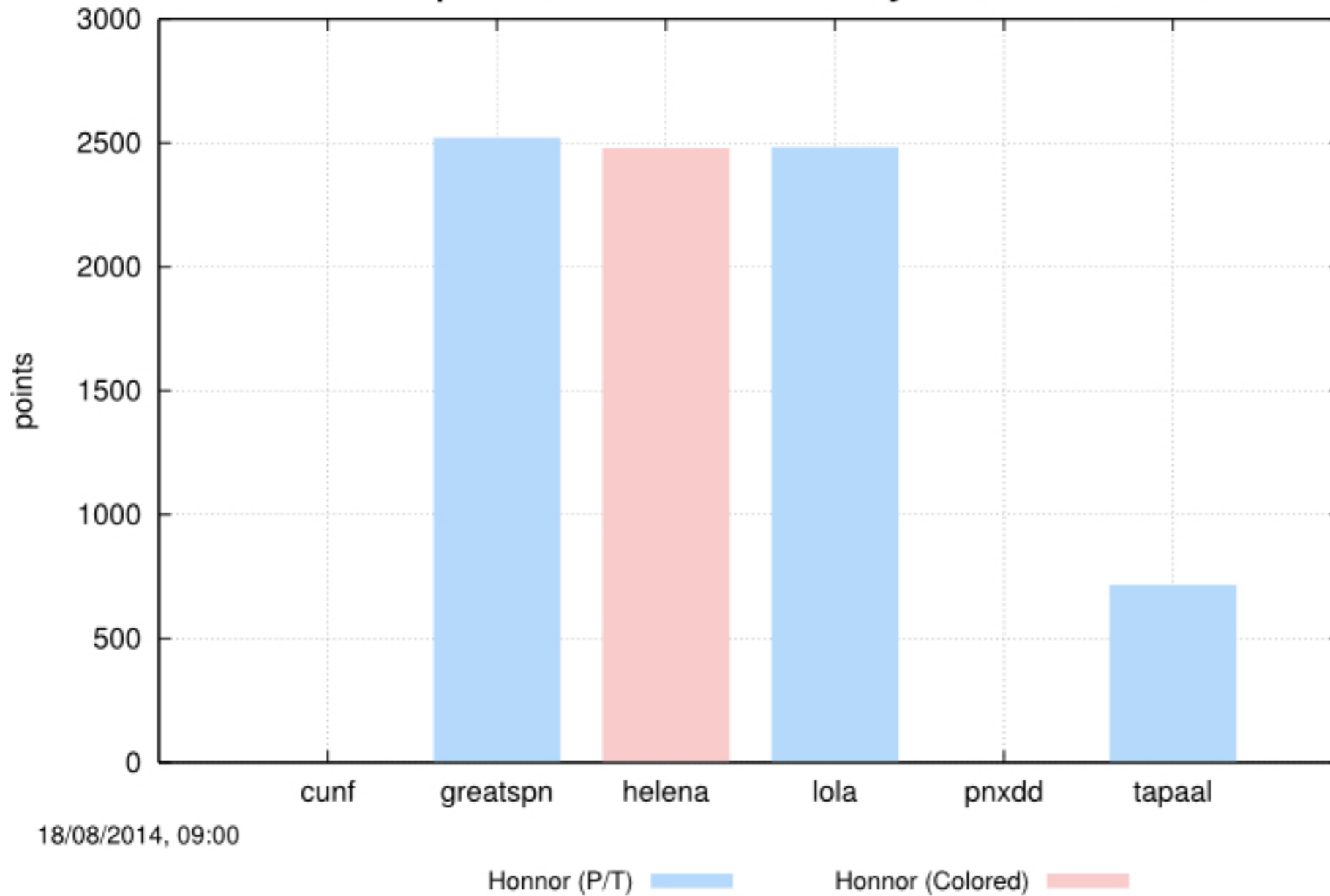


Tool classification for Reachability examinations





Honor points for Reachability examinations



LoLA



Tapaal

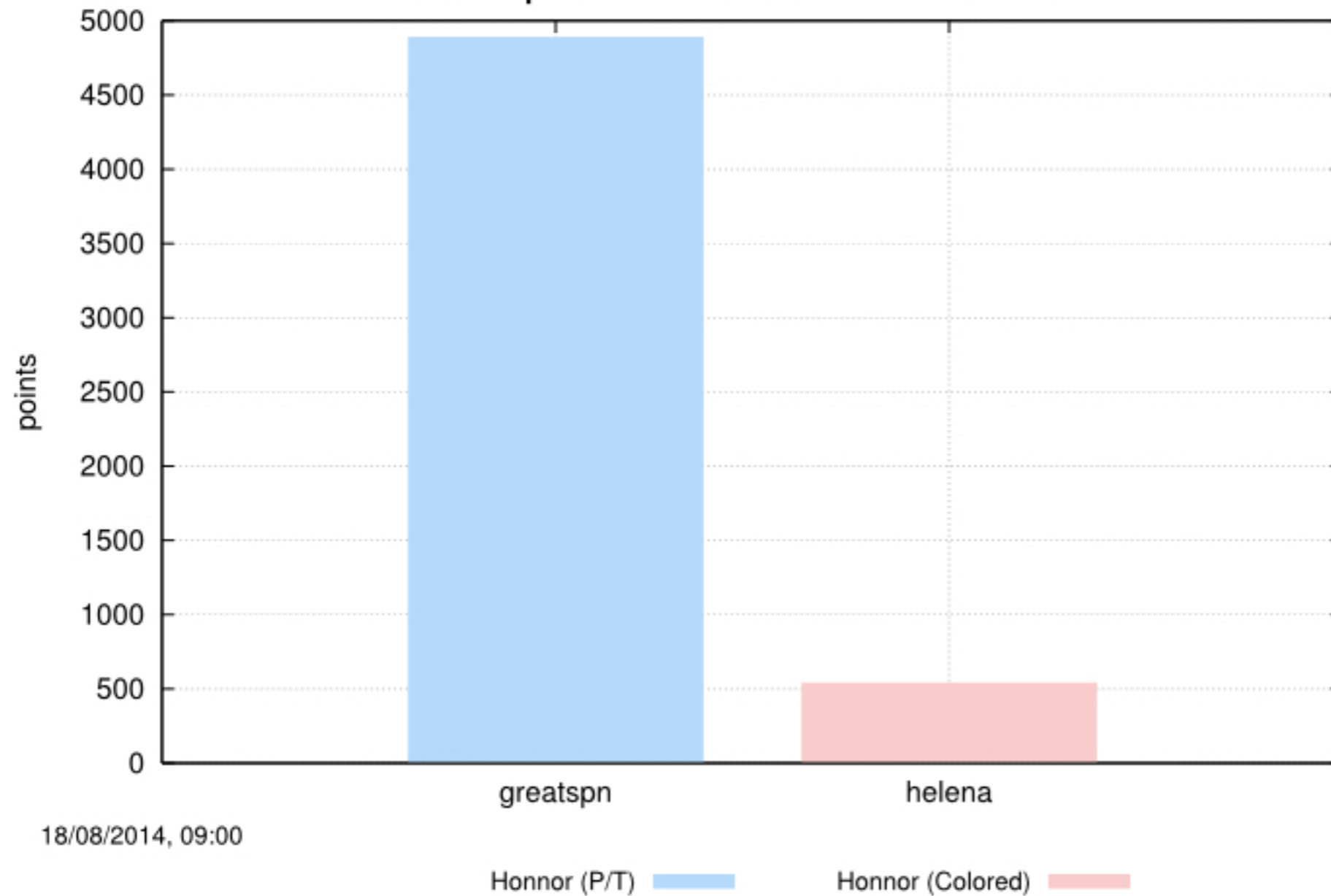


GreatSPN



Only honor points!!!

Honor points for CTL examinations

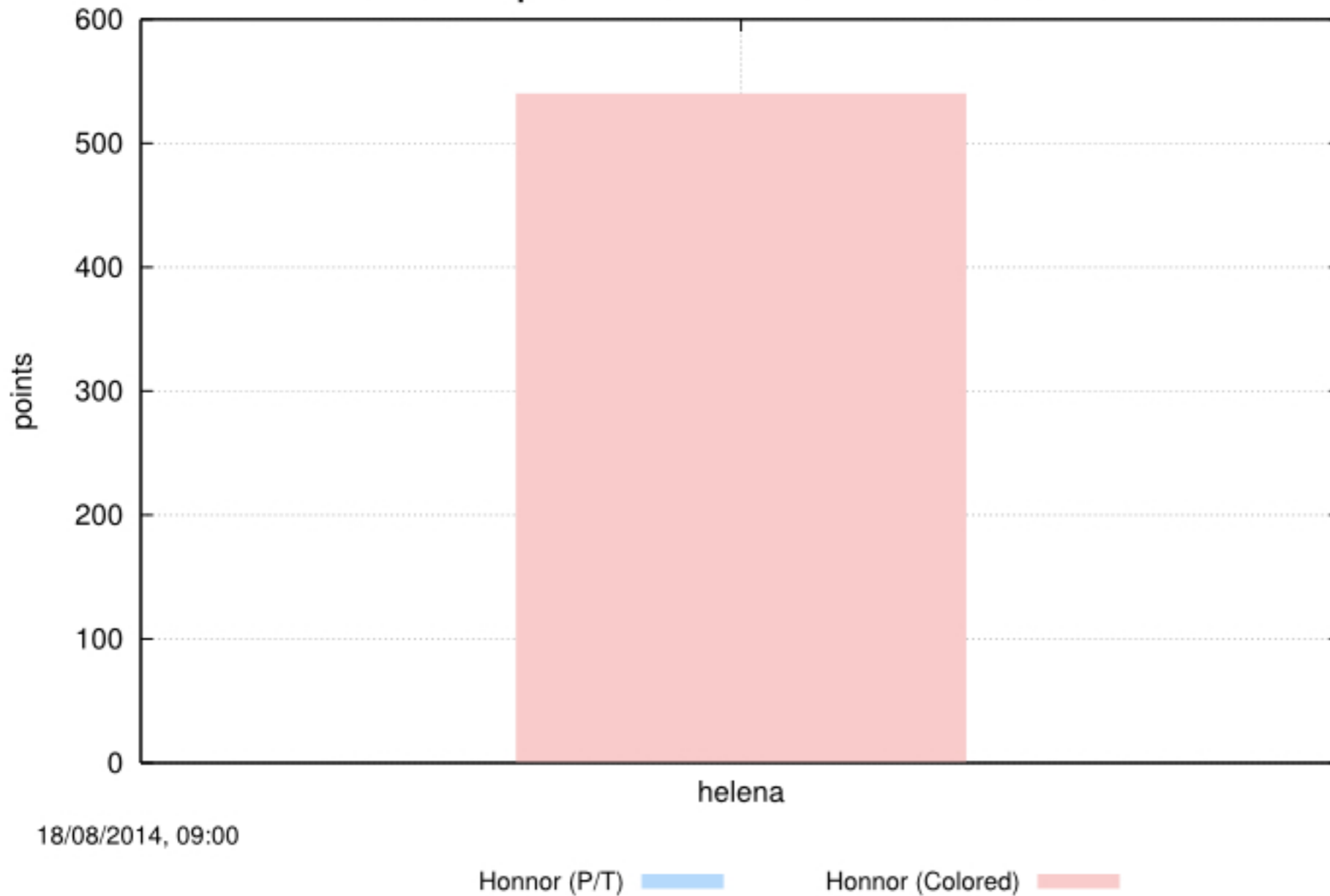


18/08/2014, 09:00



Only one tool participating

Honor points for LTL examinations



18/08/2014, 09:00

- These podiums do not show all aspects of tools
 - Expressivity
 - Formula translation problems

- There is probably a lot of useful information in the detailed report
 - To appear soon (generator to be enhanced)
 - HTML (online) and pdf (CoRR — Cornell Research Repository)

- So this should not be considered as an absolute classification!
 - See it as outputs for tool developers
 - They can focus their effort and evaluate some strategies

**MCC
2014**

**CONCLUDING
REMARKS**



On the models

- Huge effort to increase consistency
 - Hubert Garavel + Lom Hillah

On the formulas

- Huge effort to generate better formulas
 - Alban Linard

On the execution environment

- Huge effort to increase «ease to use» and efficiency
 - Fabrice Kordon + Francis Hulin-Hubard

On the results

- Much more efficient analysis
 - 12h down to 35mn from raw data
 - Evaluation of correctness

These may be reused for better state of the art comparisons

To come

- Diffusion of results + VM and environment
 - Reproducibility of results if needed (training for next year? 🤔)

- 🏆 The team is ready
- 🏆 Challenge, the machines
 - 🔊 Quadhexa from Nanterre
 - 🔊 Rented OVH server? (contract used in 2014 finished)
 - 🔊 BenchKit 2 shows efficiency = gain of time
- 🏆 Challenge, increase detection of «good results»
 - 🔊 Weight tools according to their detected accuracy
- 🏆 Challenge, increase the generation of formulas
 - 🔊 50 000+ formula needed
- 🏆 And also...
 - 🔊 «live event» (quite complex)



We really need help!

**MCC
2014**

**THANK YOU FOR YOUR
ATTENTION**

**READY FOR
DISCUSSION?**