Model Checking Contest results for 2016

Fabrice Kordon — LIP6, Univ. P. & M. Curie, France
Hubert Garavel — Inria/LIG, France
Lom Messan Hillah — LIP6 & Univ. Paris Ouest Nanterre, France
Francis Hulin-Hubard — LSV, CNRS/ENS de Cachan, France

Emmanuel Paviot-Adet — LIP6 & Univ. Paris Descartes, France
Loïg Jézequel, IRCCyN, Univ. Nantes, France
César Rodrígez — LIPN, Univ. Paris 13, France

MCC 2016

Objectives

- **Promoting model checking tools**
  - Compare and debug
    - **Oracle handled by the developers themselves**
  - Enhance reproducibility of results
    - **BenchKit + dedicated environment using virtualization (easier replay)**
    - **Submissions available online**
  - Encourage tools and tool support
    - **Observatory for the community**
    - **Provide reusable and fair comparison charts and data**
- **Creating a common database of benchmark**
  - Models from various origins (more to tell later)
    - **PNML is a good format for this**
- **Competing tools not only dedicated to Petri nets**
  - Tools coming from other communities

MCC 2016 Model Checking Contest — who does what?

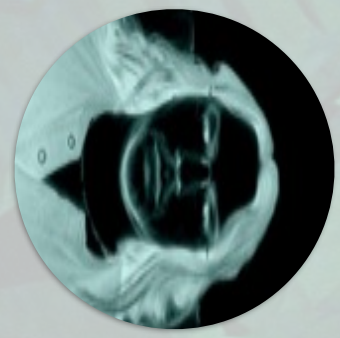
Hubert Garavel (Inria)

Lom Hillah (UPOND)

Managing Models

Managing Execution + analysis

Fabrice Kordon (UPMC)



Francis Hulin-Hubard (CNRS)

Loïg Jezequel (U. Nantes)

Emmanuel Paviot-Adet (UP5)

César Rodríguez (UP13)

Managing Formulas

MCC 2016 Tools Submitted this Year

- **ITS-Tools**
  - Univ. P. & M. Curie, F
- **LoLA**
  - Univ. Rostock, D
- **LTSMin**
  - Univ. Twente, NL
- **MARCIE**
  - Univ. Cottbus, D
- **PeCan (new)**
  - Univ. HoChiMinh, VN
- **pnmc**
  - Steery.io, F
- **PNXDD**
  - Univ. P. & M. Curie, F
- **Smart (new)**
  - Iowa State Univ, USA
- **tapaal**
  - Univ. Aalborg, DK
    - **3 variants (PAR, SEQ, EXP)**
- **ydd-pt (new)**
  - Univ. Geneva, CH

MCC 2016 Tools Submitted this Year

- **ITS-Tools**
  - Univ. P. & M. Curie, F
- **LoLA**
  - Univ. Rosto
- **LTSMin**
  - Univ. Twente, NL
- **MARCIE**
  - Univ. Cottb
- **PeCan (r**
  - Univ. HoC

Not present this year

Cunf, GreatSPN, StraTAGem

- **pnmc**
  - St
  - Curie, F
- **Smart (new)**
  - Iowa State Univ, USA
- **tapaal**
  - rg, DK

PAR, SEQ, EXP)

ew)

  - Univ. Geneva, CH

All VMs will be published

Reproducibility can be achieved

Techniques Reported by Tools

Tools	parallelism	Techniques
Marcie	/	SEQUENTIAL_PROCESSING DECISION_DIAGRAMS UNFOLDING_TO_PT
PeCan	/	EXPLICIT
pnmc	/	DECISION_DIAGRAMS USE_NUPN
PNXDD	/	DECISION_DIAGRAMS TOPOLOGICAL
Smart	/	DECISION_DIAGRAMS
tapaal(EXP)	/	EXPLICIT STRUCTURAL_REDUCTION STATE_COMPRESSION STATE_EQUATIONS
tapaal(SEQ)	/	EXPLICIT STRUCTURAL_REDUCTION STATE_EQUATIONS
ydd-pt	/	DECISION_DIAGRAMS
ITS-Tools	MC	DECISION_DIAGRAMS SAT_SMT INITIAL_STATE TOPOLOGICAL USE_NUPN
LoLA	MC	PARALLEL_PROCESSING EXPLICIT SAT_SMT STATE_COMPRESSION STUBBORN_SETS TOPOLOGICAL
LTSMin	PAR	DECISION_DIAGRAMS EXPLICIT STATIC_VARIABLE_REORDERING USE_NUPN
tapaal(PAR)	PAR	EXPLICIT COMPRESSION STRUCTURAL_REDUCTION STATE_EQUATIONS

Processing Capacity

	bluewhale03	Ebro	Quadhexa-2	Small (cluster)	Total
Cores	40 @ 2.8GHz	64 @ 2.7GHz	24 @ 2.66GHz	11x24 @ 2.4GHz	-
Memory (GB)	512	1024	128	11x64	-
Used Cores (1 per VM) for sequential tools	31 31 VM in //	63 63 VM in //	7 7 VM in //	11x3, 5x3 VM in //	-
Used Cores (4 per VM) for parallel tools	36, 9 VM in //	60, 15 VM in //	20, 5 VM in //	11x3, 5x3 VM in //	-
Number of runs	13 374	36 936	15 768	62 604	**128 682**
Total CPU required	156d, 17h, 44m, 59s	485d, 19h, 27m, 43s	203d, 0h, 25m, 47s	636d, 9h, 11m, 36s	**1481d, 22h, 50m, 5s**
Total CPU	**about 4 years and 20 days**				-
Time spent to complete benchmarks	**about 22 days and 1 hours**				-
VM boot time of VMs + management (overhead)	**22 d, 8h (Included in total CPU)**				-

Processing Capacity

	bluewhale03	Ebro	Quadhexa-2	Small (cluster)	Total
				@ Hz	-
Memor				4	-
Used Cores (1 pe for sequentia				in //	-
Used Cores (4 per VM) for parallel tools	36, 9 VM in //	60, 15 VM in //	20, 5 VM in //	11x3, 5x3 VM in //	-
Number of runs				604	**128 682**
Total CPU required				, 9h, 36s	**1481d, 22h, 50m, 5s**
Total CPU					-
Time spent to complete benchmarks					-
VM boot time of VMs + management (overhead)			**22 d, 8h (Included in total CPU)**		-

Less CPU than in 2015

128 682 runs instead of 169 078
but more completed runs

Thank you very much

Université de Genève
Rostock University
Université Paris Ouest
Université P. & M. Curie

Categories of Models

«known» models

- Those from past years
  - **Test the tool as used by its developers**

«Stripped» models

- «known» (original archive) and set as «surprise» ones
  - **Test the tool as used by «non experts» of the tool**

«Surprise» models

- New models proposed by the community this year
  - **Test the tool as used by «non experts» of the tool**
  - **new situations for the tool**

Categories of Models

- **«known» models**
  - Those from past years
    - **Test the tool as used by its develope**

Coefficients (after pool)

«known» = x1
«stripped» = x3
«surprise» = x5

- **«Stripped» models**
  - «known» (original archive) and set as «surprise» ones
    - **Test the tool as used by «non experts» of the tool**

- **«Surprise» models**
  - New models proposed by the co
    - **Test the tool as used by «non exper**
    - **new situations for the tool**

Execution consistency

On the same machine
«known» / «stripped»
colored + associated P/T

F. Kordon - Université P. & M. Curie - CC2016

11 New Models for 2016

- **B. Barbot**
  - PaceMaker
- **B. Barbot and M. Kwiatkowska**
  - DNAwalker
- **H. Evrard and F. Lang**
  - DLCshifumi
- **M. Heiner**
  - GPPP
- **F. Jebali and E. Jenn**
  - AutoFlight
- **F. Kordon**
  - AirplaneLD
- **G. Salaün**
  - CloudDeployment
- **W. Serwe and H. Garavel**
  - DES
- **T. Shmeleva**
  - TriangularGrid
- **D. Zaistev**
  - HypertorusGrid
  - TCPcondis

11 New Models for 2016

- **B. Barbot**
  - PaceMaker
- **B. Barbot and M. Kwiatkowska**
  - CloudDeployment
- **W. Serwe and H. Garavel**
  - DES
- **T. Shmeleva**
  - TriangularGrid
- **M. Heiner**
  - GPPP
- **F. Jebali and E. Jenn**
  - AutoFlight

With scaling parameters

139 models in fact

Thanks!!!

We really need various models

Already from past years

525 instances of models

Examinations

- **StateSpace**
- **UpperBound**
- **Reachability**
  - ReachabilityDeadlock
  - ReachabilityCardinality → atomic propositions refer to tokens
  - ReachabilityFireability → atomic propositions refer to firing
- **CTL**
  - CTLCardinality → atomic propositions refer to tokens
  - CTLFireability → atomic propositions refer to firing
- **LTL**
  - LTLCardinality → atomic propositions refer to tokens
  - LTLFireability → atomic propositions refer to firing

The Submission Protocol

- **May 1st, delivery of disk images**
  - Qualification phase
  - Completed by mid May
    - **~37 500 test runs**
- **May 17, starting to operate tools**
  - 128 682 runs distributed over 4 different machines over Europe
  - VM with 4 cores / 16GB
    - **ITS-Tools, LTSMin, TAPAAL(PAR), LoLa**
  - WM with 1 core / 16 GB
    - **Marcie, PeCan, pnmc, PNXDD Tapaal (SEQ, EXP), ydd-pt**
  - Time confinement, 1h

The Analysis Protocol

- **Mid June, consolidation + analysis of outcomes**
  - 31 GByte of logs and CSV files
    - **Post analysis = ~18KLOC Ada + ~800 LOC bash**
- **Analysis Protocol**
  - Pass 1, computing results for the majority in a «line»
    - **All tools for an examination for a model instance**
  - Pass 2, evaluating tool reliability
    - **Only considering values with a large majority**
  - Pass 3, reconstructing the results using tool reliability
    - **Help to decide when only 2 different answers**
    - **A result must be of confidentiality 0.93 or more (0.9 in 2015)**
    - **Some results are tagged «insecure»**
  - Pass 4 computing scores
    - **«insecure» results not considered when counting points**

The Analysis Protocol

- **Mid June, consolidation**
  - 31 GByte of logs and CSV file
    - **Post analysis = ~18KLOC Ada +**
- **Analysis Protocol**
  - Pass 1, computing results for the majority in a «line»
    - **All tools for an examination f**
  - Pass 2, evaluating tool relia
    - **Only considering values with**
  - Pass 3, reconstructing the re
    - **Help to decide when only 2 different answers**
    - **A result must be of confidenti**
    - **Some results are tagged «inse**
  - Pass 4 computing scores
    - **«insecure» results not conside**

Scoring

StateSpace, 10 / 2 / 2 / 2
Deadlock, 16
Other formulas, 1 per formula

Bonus for a «line»

+4 for the fastest tool
+4 for the smallest memory footprint

Penalty for mistakes

Twice the score for a good value
No bonus if at least one error

Checking the Results

- **Consistency checks**
  - Colored versus equivalent P/T nets
  - «known» models versus «stripped» models
- **Computing the «reliability rate»**
  - Section III.2 in http://mcc.lip6.fr/rules.php
  - Computing **V**, the set of values with a majority of 3 and more tools
  - For each tool t, selecting $\mathbf{V_t}$, the values computed $\in$ **V**
  - For each tool t, selecting $\mathbf{V_{tt}}$, the correct values computed $\in$ **V**
  - Reliability rate = $\frac{|\mathbf{V_{tt}}|}{|\mathbf{V_t}|}$

Tool Reliability in 2015

Tools	Reliability	success	selected	Examinations
Cunf	96,96 %	4728	4 876	3 (Reach)
GreatSPN-Meddly	62,30 %	11 966	19 206	10 (Sate, Reach, CTL)
ITS-Tools	64,05 %	10 890	17 003	4 (Sate, Reach)
LoLA 2.0	97,80 %	25 796	26 378	6 (Reach)
LTSMin	79,13 %	13 995	17 687	5 (State, Reach)
Marcie	92,52 %	18 443	19 934	10 (Sate, Reach, CTL)
pnmc	99,59 %	741	744	1 (State)
PNXDD	88,89 %	56	63	1 (State)
STrataGEM0.5.0	100,00 %	243	243	1 (State)
TAPAAL (SEQ)	99,88 %	22 880	22 907	7 (State, reach)
TAPAAL(MC)	99,75 %	23 247	23 306	7 (State, reach)
TAPAAL-OTF (SEQ)	96,19 %	19 001	19 733	7 (State, reach)
TAPAAL-OTF(PAR)	88,43 %	15 253	17 248	7 (State, reach)

Tool Reliability in ~~2015~~ 2016

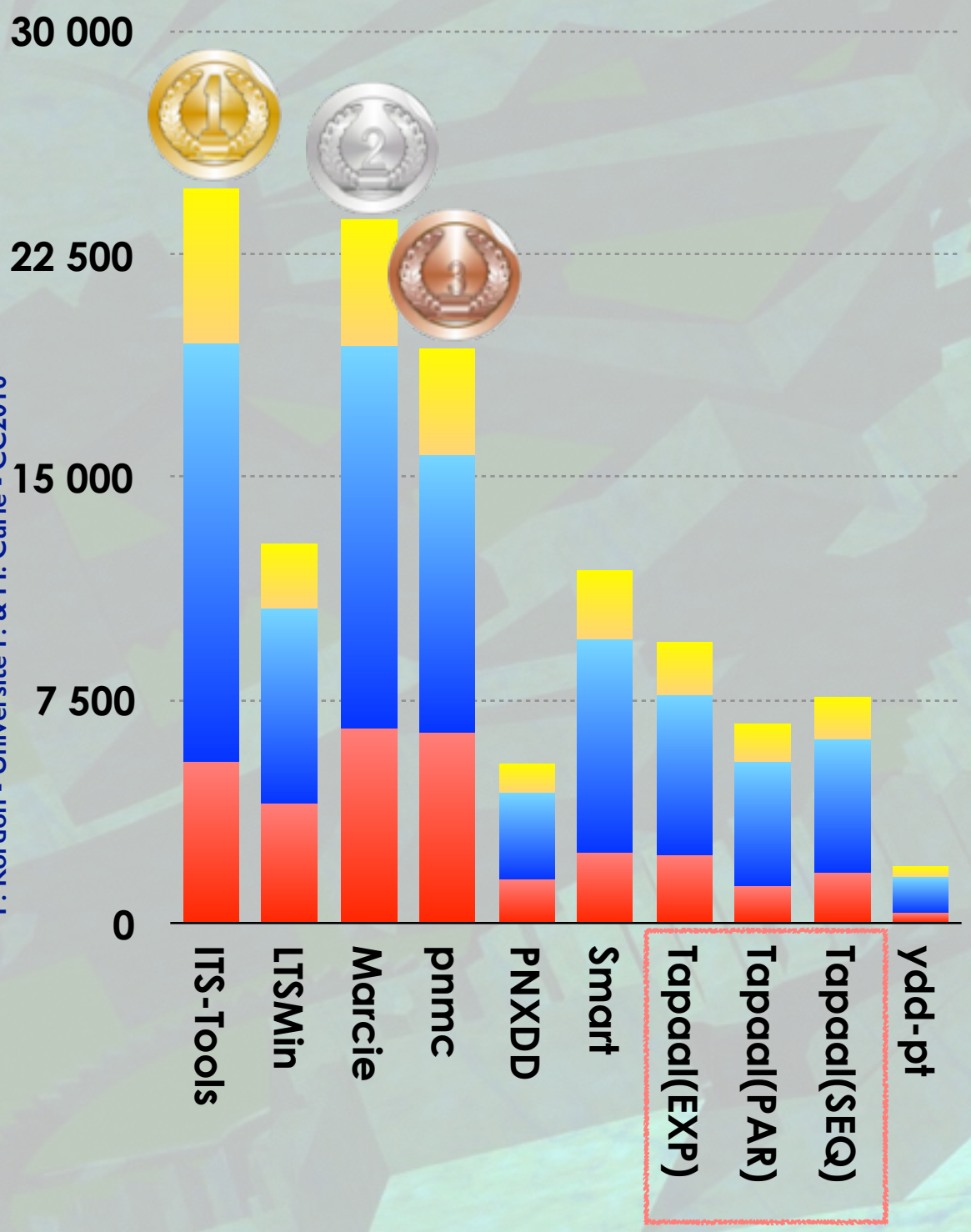
Tools	Reliability	success	selected	Examinations
ITS-Tools	98,38 %	33 634	34 189	9 (SS, UB, Reach, CTL, LTL)
LoLa	99,22 %	41 011	41 335	8 (UB, Reach, CTL, LTL)
LTSMin	99,98 %	34 902	34 910	8 (SS, Reach, CTL, LTL)
Marcie	99,99 %	27 361	27 364	7 (SS, UB, Reach, CTL)
PeCan	37,54 %	3 967	10 568	5 (Reach, LTL)
pnmc	99,84 %	1 219	1 221	1 (State Space)
PNXDD	99,11 %	222	224	1 (State Space)
Smart	98,72 %	926	938	1 (State Space)
ydd-pt	97,70 %	85	87	2 (SS, UB)
Tapaal(EXP)	99,95 %	22 421	22 434	5 (SS, UB, Reach)
Tapaal(PAR)	99,98 %	19 555	19 558	7 (SS, UB, Reach, CTL)
Tapaal(SEQ)	99,97 %	30 130	30 140	7 (SS, UB, Reach, CTL)

Tool Reliability in ~~2015~~ 2016

Answering protocol not respected

Tools	Reliability	success	selected	Examinations
ITS-Tools	98,38 %	33 634	34 189	9 (SS, UB, Reach, CTL, LTL)
LoLa	99,22 %	41 011	41 335	8 (UB, Reach, CTL, LTL)
LTSMin	99,98 %	34 902	34 910	8 (SS, Reach, CTL, LTL)
Marcie	99,99 %	27 361	27 364	7 (SS, UB, Reach, CTL)
PeCan	37,54 %	3 967	10 568	5 (Reach, LTL)
pnmc	99,84 %	1 219	1 221	1 (State Space)
PNXDD	99,11 %	222	224	1 (State Space)
Smart	98,72 %	926	938	1 (State Space)
ydd-pt	97,70 %	85	87	2 (SS, UB)
Tapaal(EXP)	99,95 %	22 421	22 434	5 (SS, UB, Reach)
Tapaal(PAR)	99,98 %	19 555	19 558	7 (SS, UB, Reach, CTL)
Tapaal(SEQ)	99,97 %	30 130	30 140	7 (SS, UB, Reach, CTL)

StateSpace Examination

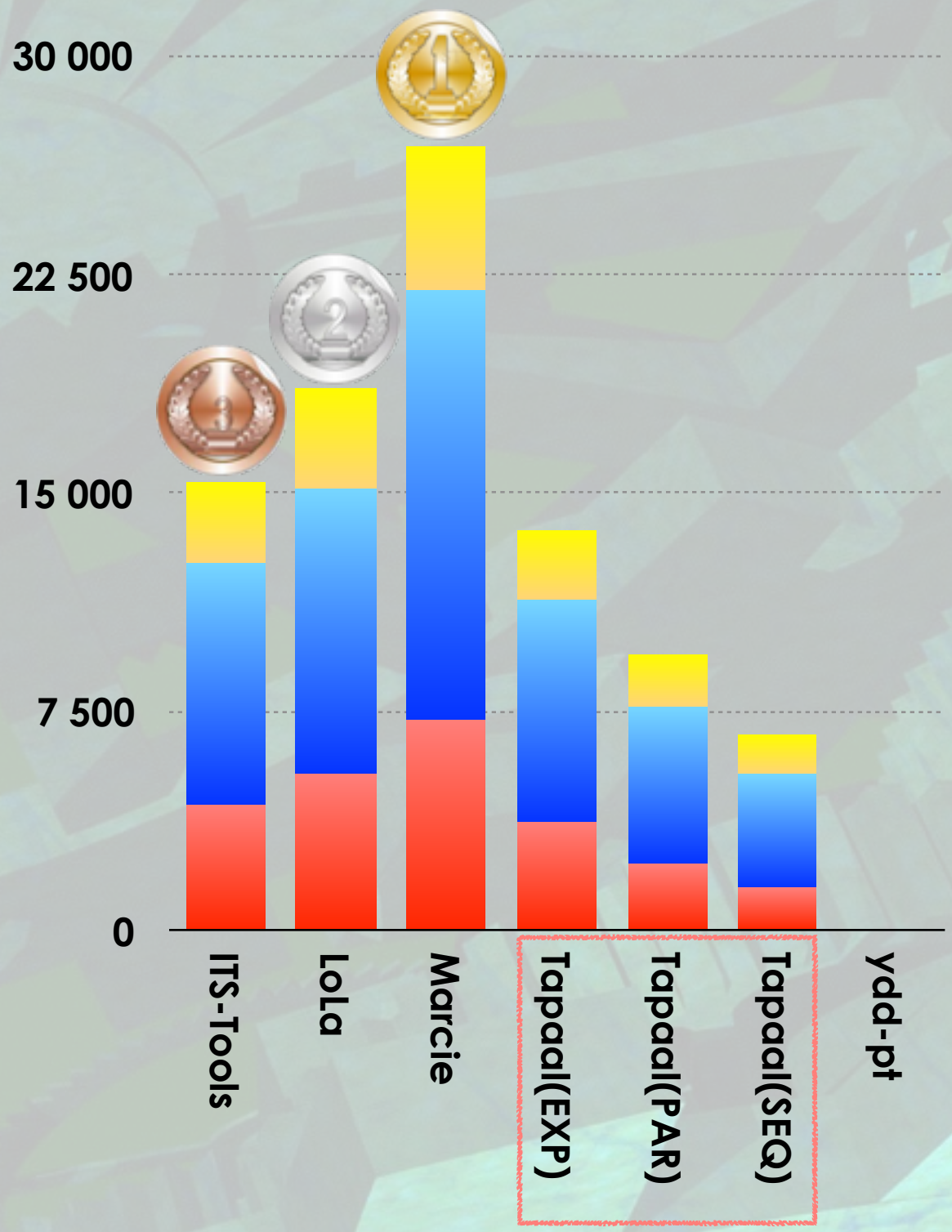


The most attended one

- 10 tools/variants participating
  - **Out of 12**

F. Kordon - Université P. & M. Curie - C2016

UpperBound Examination



A popular one

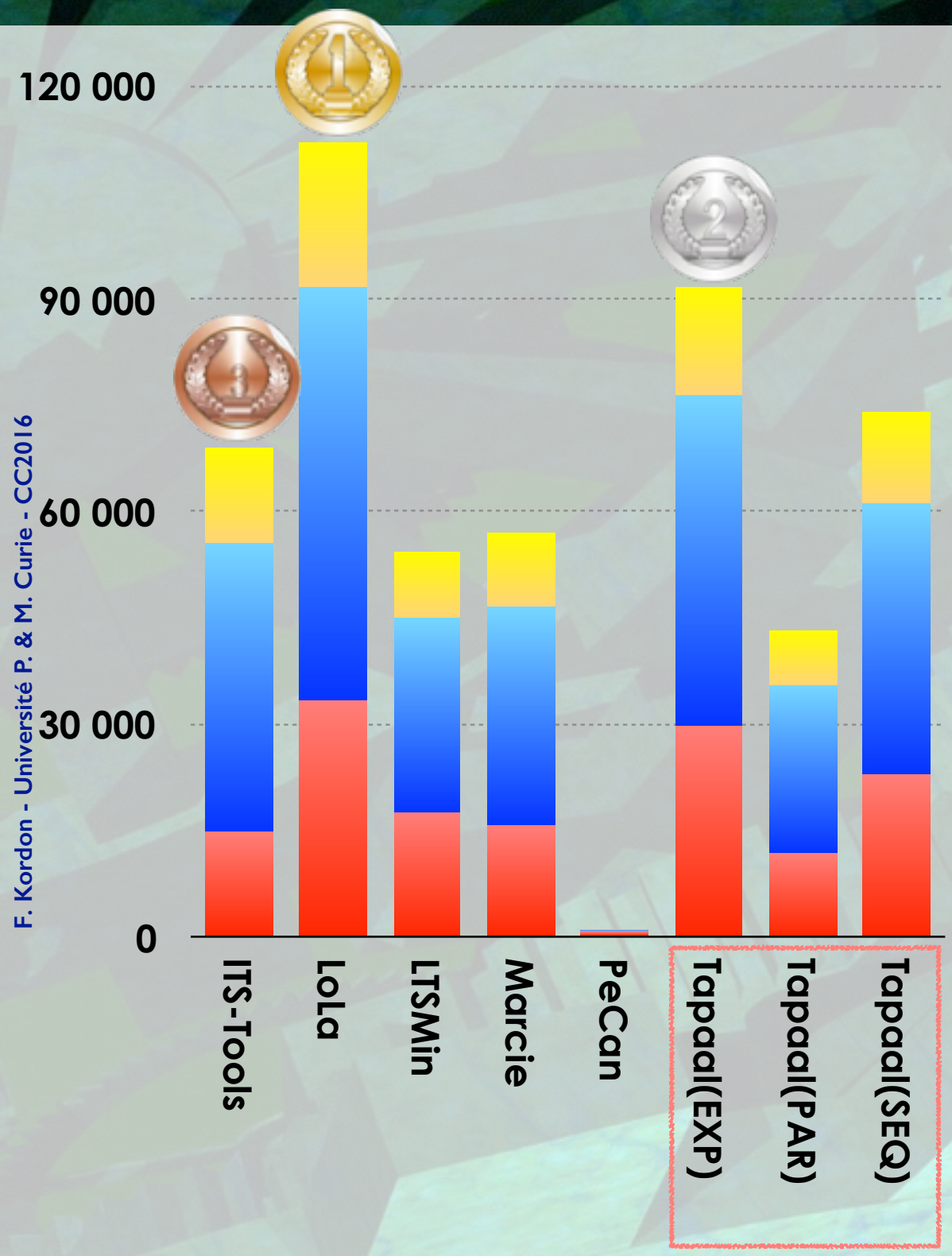
- 7 tools/variants participating
  - **Out of 12**

Ydd-pt

- Not really participating
- Answering problem
  - **Should always answers DNC**

F. Kordon - Université P. & M. Curie - CC2016

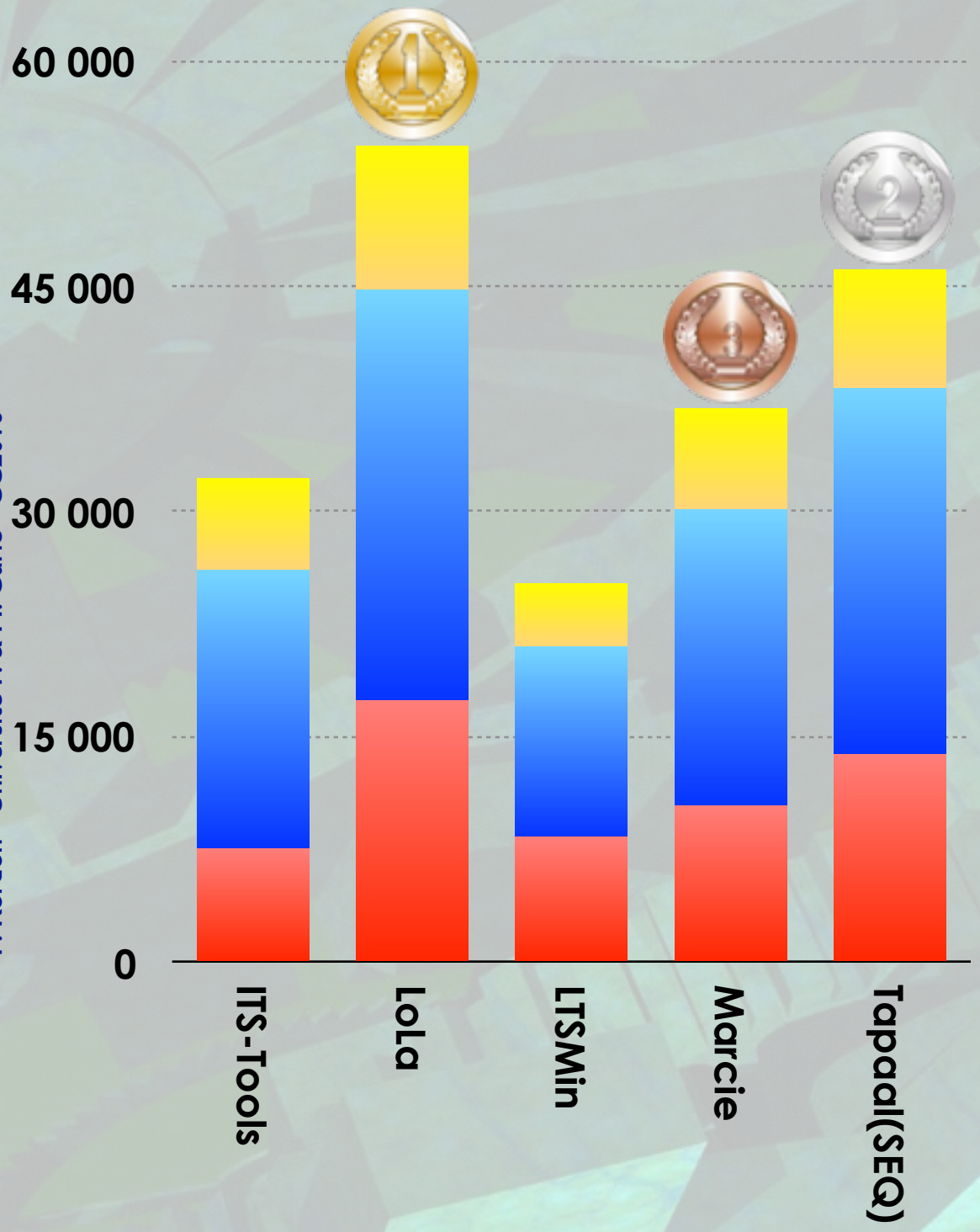
MCC 2016 All Reachability Examinations



- **A popular one**
  - 8 tools/variants participating
    - **Out of 12**
- **PeCan**
  - States erroneous values in case where it should state CC
    - **Negatives score in**
      - **ReachabilityFireability**
      - **ReachabilityCardinality**

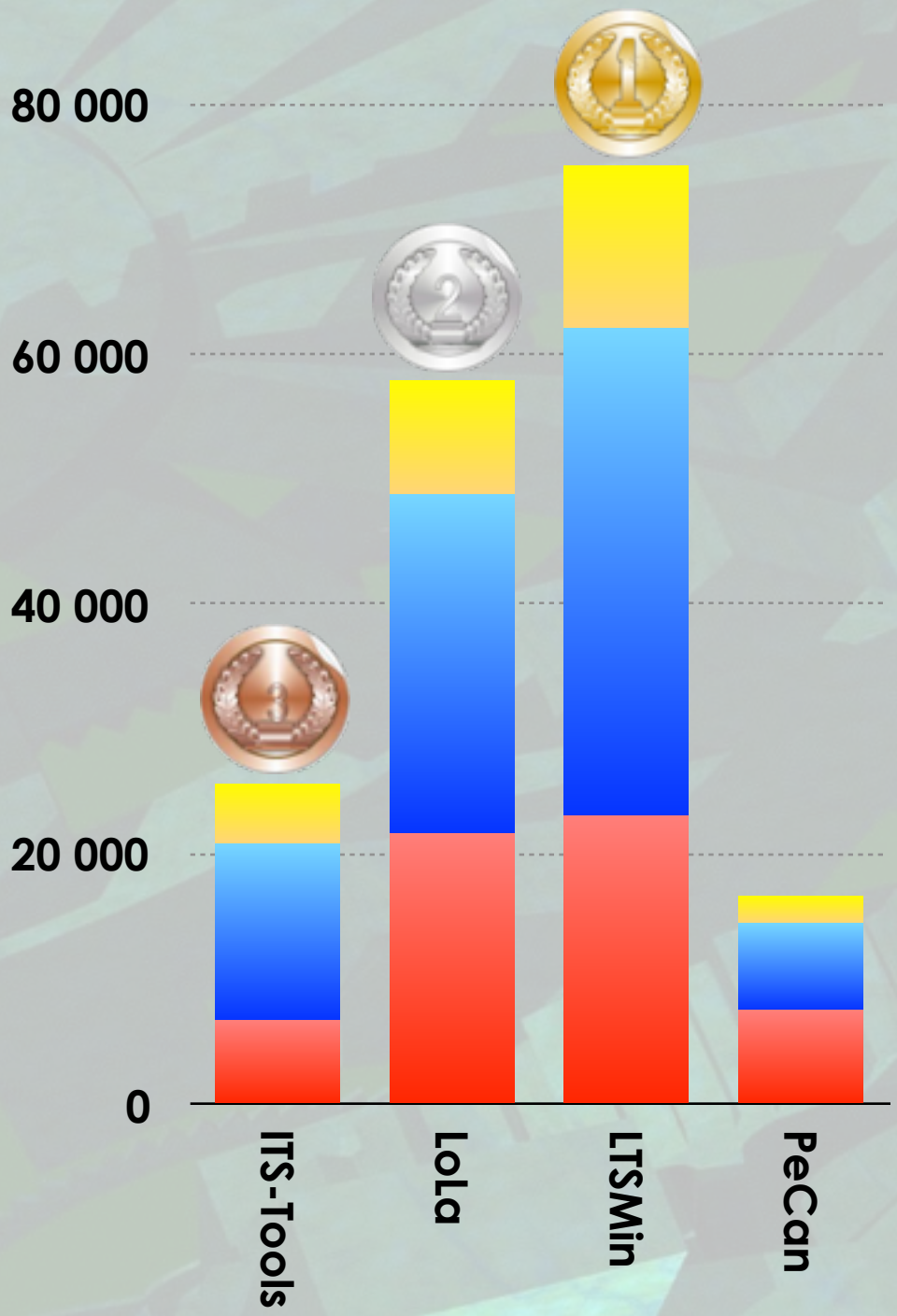
F. Kordon - Université P. & M. Curie - CC2016

All CTL Examinations



- **Less popular**
  - 6 (-1) tools/variants participating
    - **Out of 12**
- **Tapaal (par)**
  - Compilation optimization issue lately detected
    - **Crash for CTL in numerous situations**
    - **The parallel version was withdrawn**

All LTL Examinations

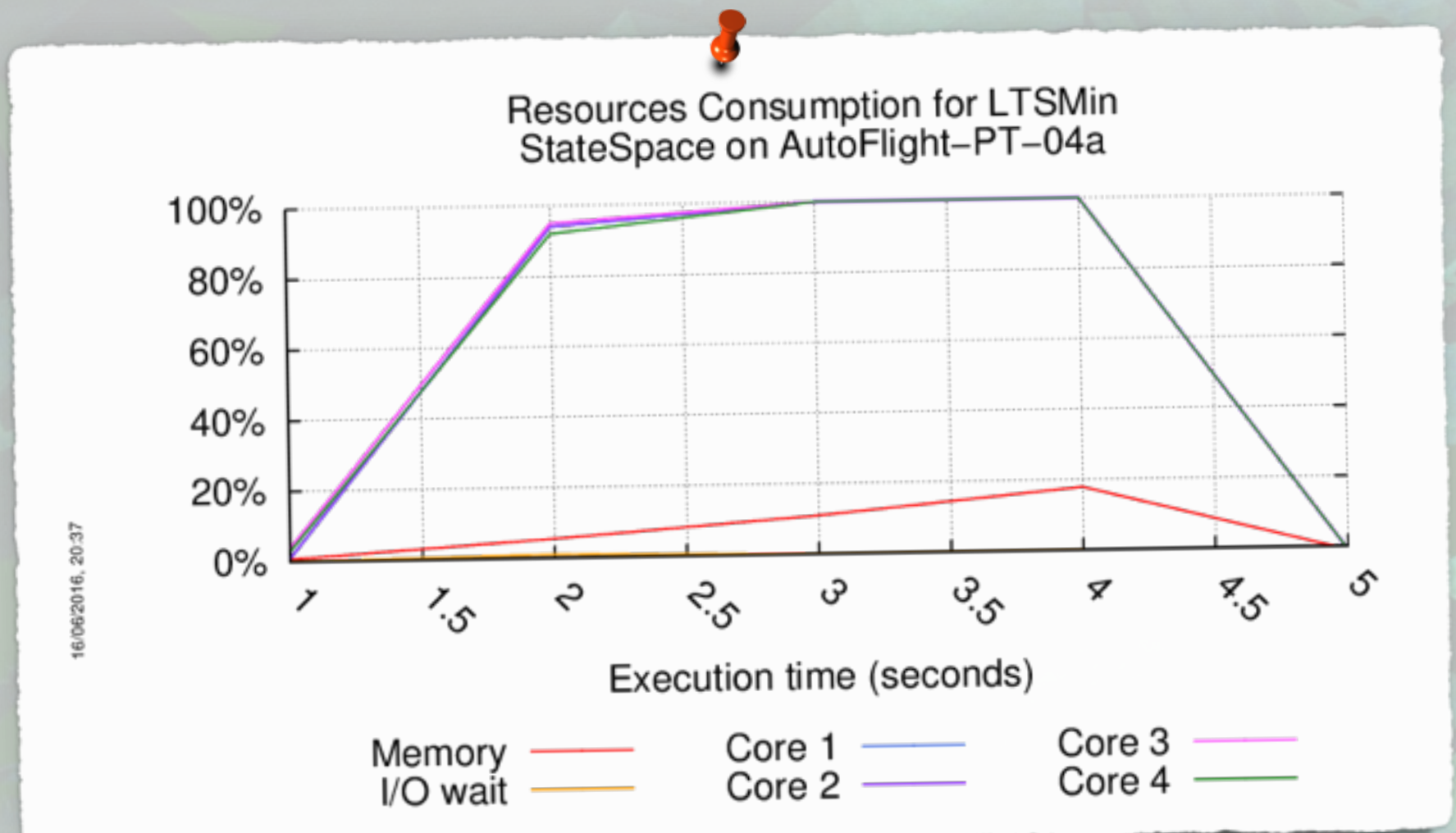


- No participating tool in 2015
  - 4 tools/variants participating
    - Out of 12

Generated Report

Full HTML report

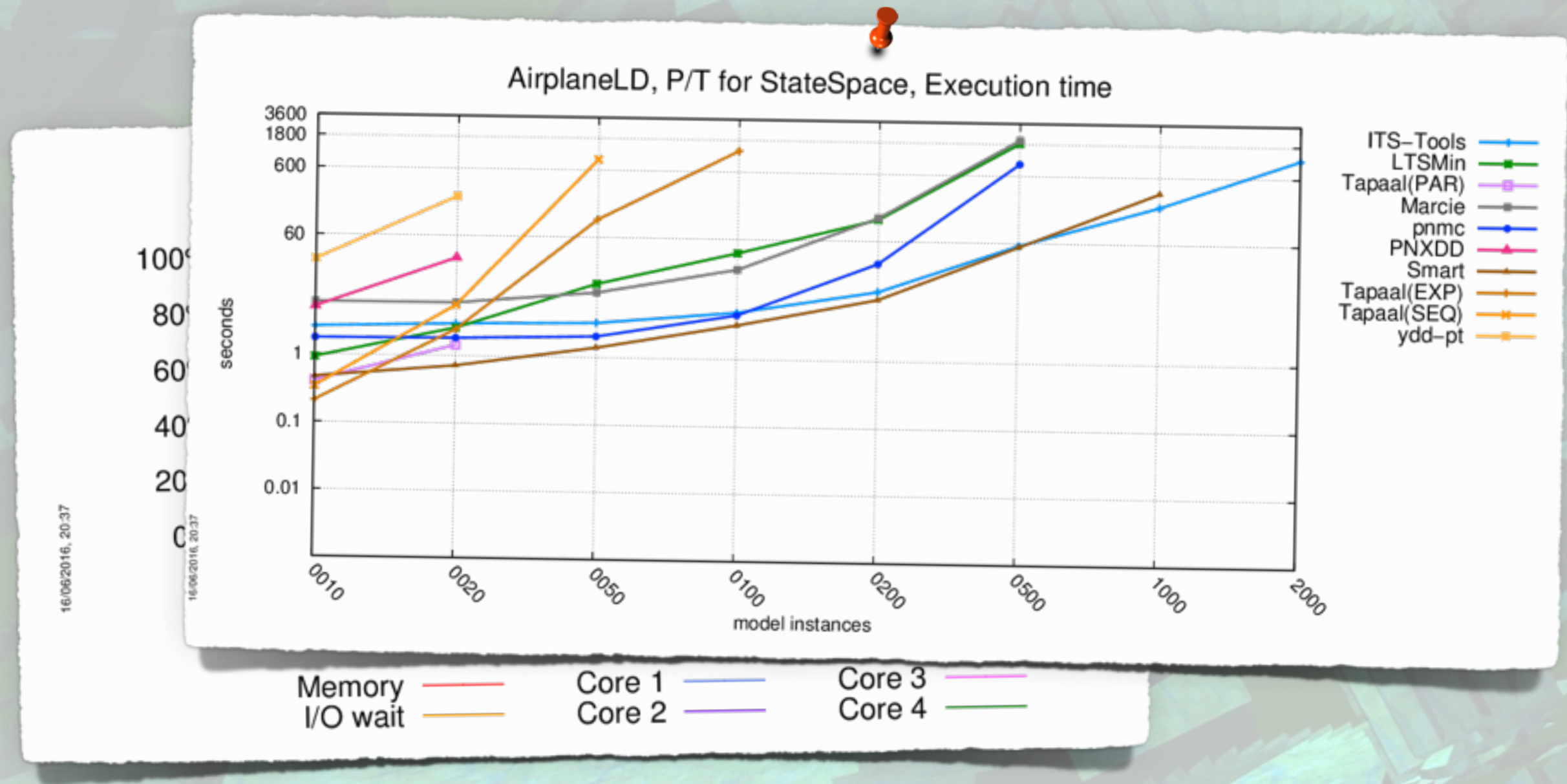
- 64 481 charts and 58 828 web pages



Generated Report

Full HTML report

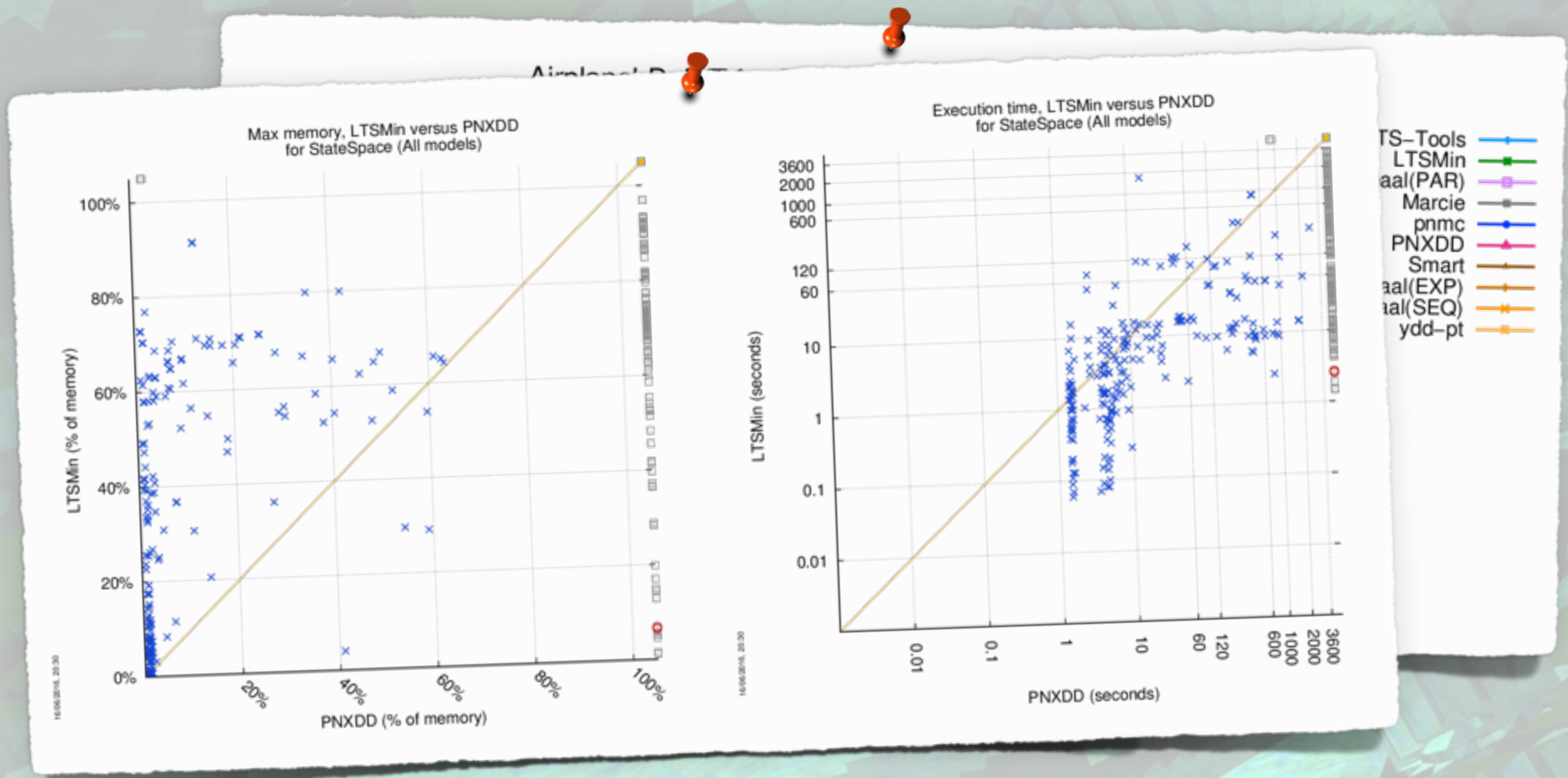
- 64 481 charts and 58 828 web pages



Generated Report

Full HTML report

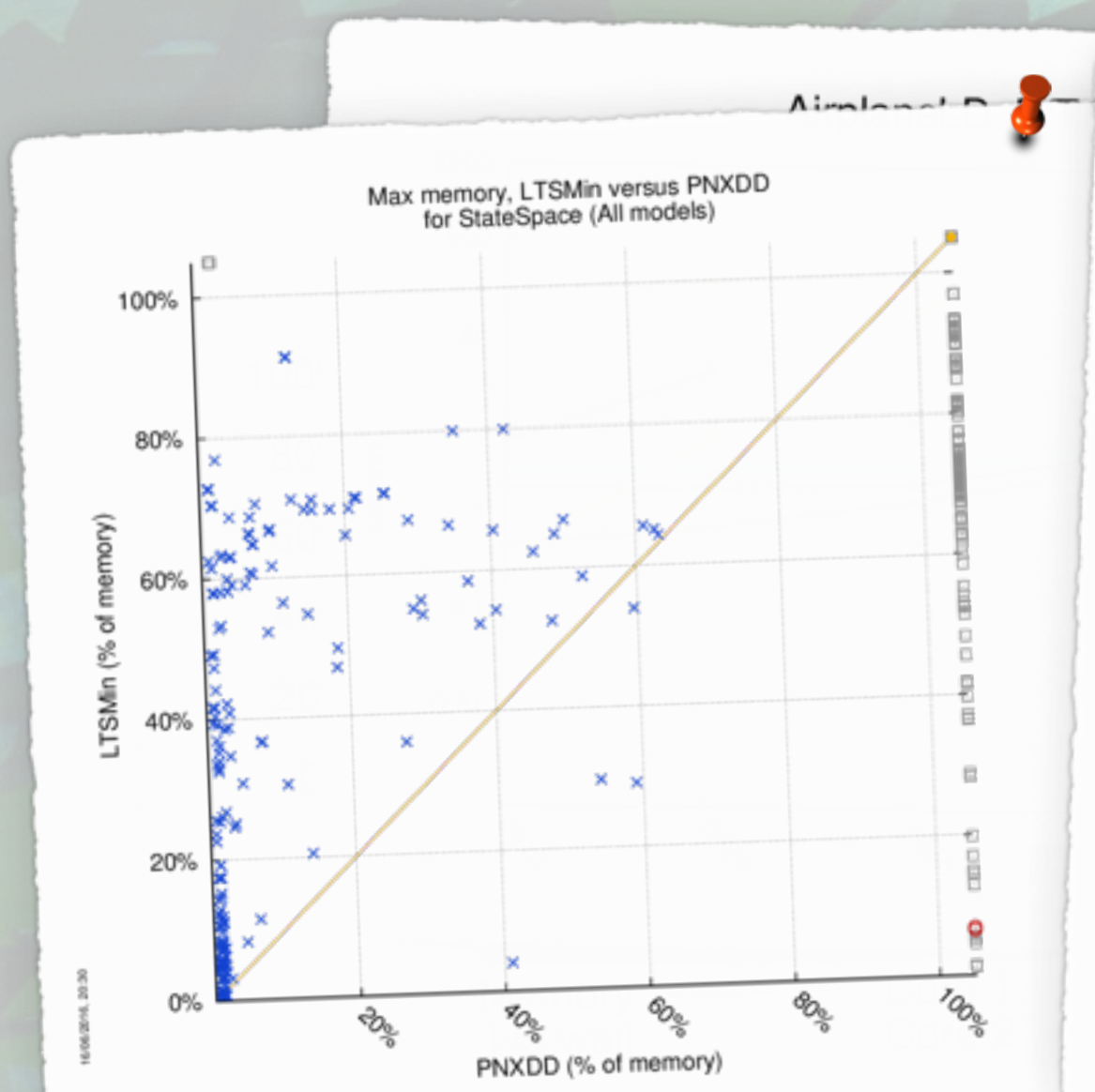
- 64 481 charts and 58 828 web pages

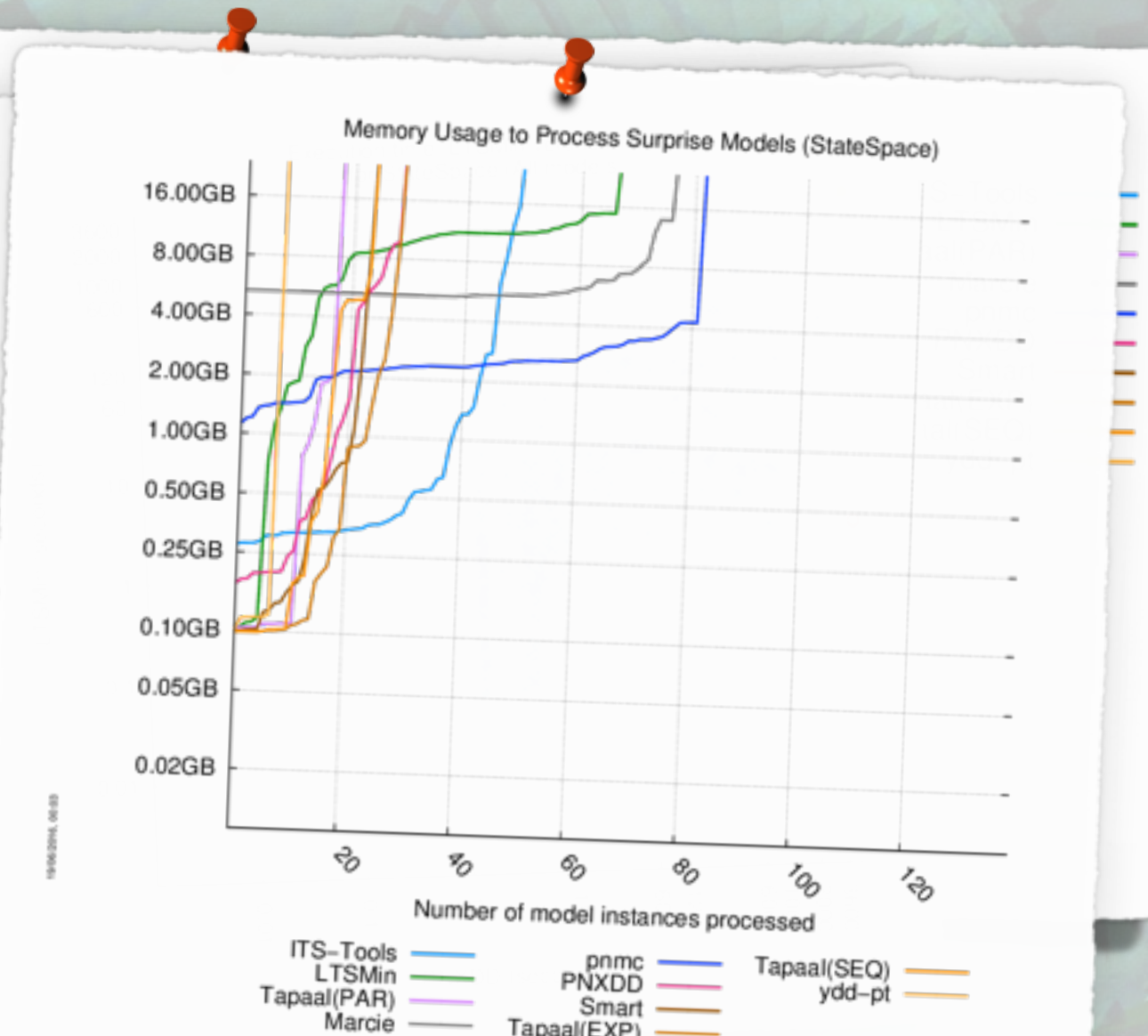


Generated Report

Full HTML report

- 64 481 charts and 58 828 web pages

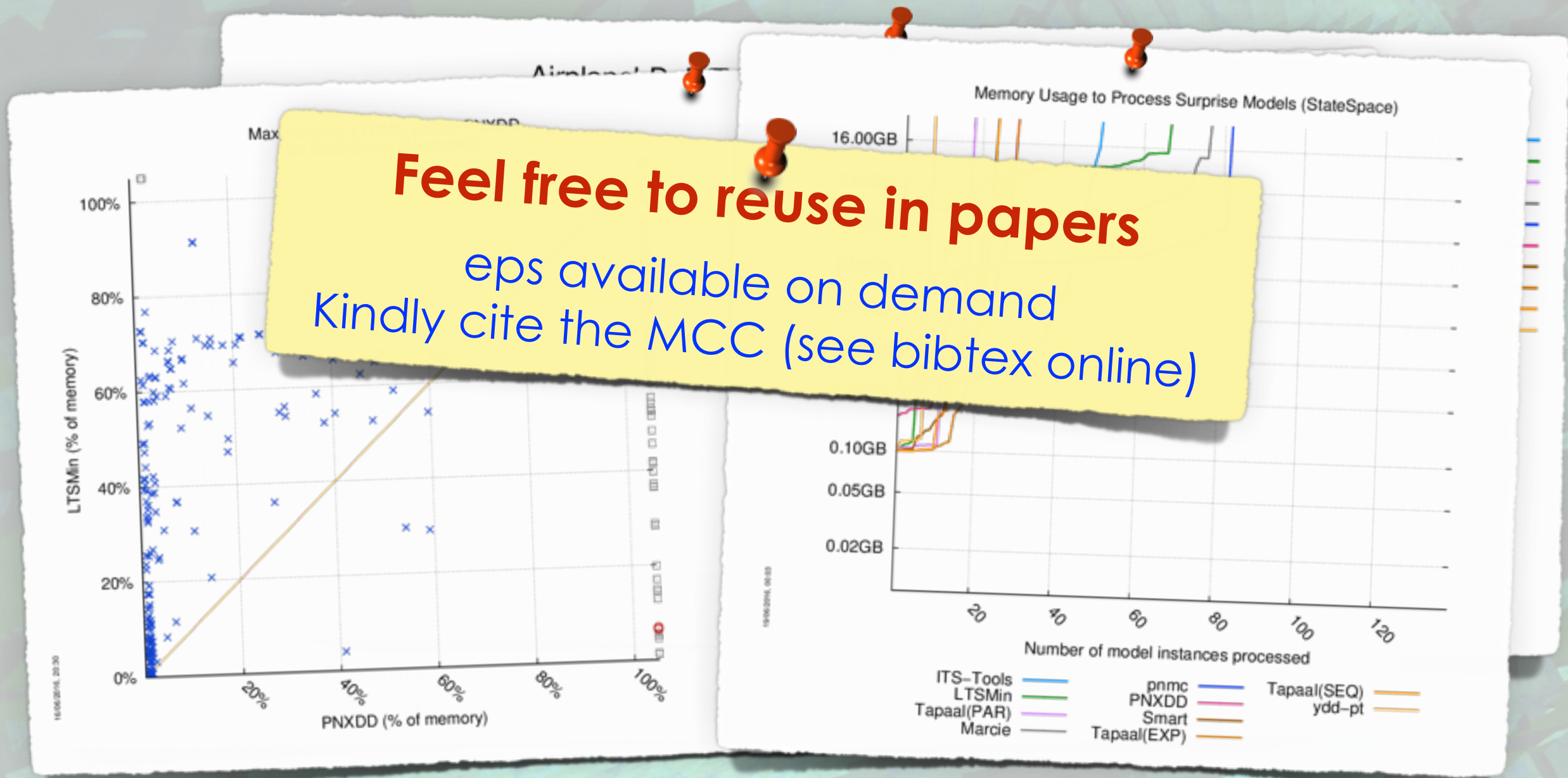




Generated Report

Full HTML report

- 64 481 charts and 58 828 web pages



Some Issues for Next Year

- **Counting transitions for StateSpace**
  - Discussion about semantics (consistency P/T versus Colored)
- **Handling some rare bugs in the benchmark**
  - Possibly on one surprise model
- **Small «almost surprise»**
  - Some instance of GPPP with more than 2^{32} tokens…
- **Better generator for LTL**
  - Possible use of SPOT
- **Please check carefully your logs**
  - Some discussion issues already started

As a Conclusion...

localhost

search | gest publis | HP fko | Equipe/Labo | TA fko | Projets | Conferences/revues | Enseignements | SAR | cloud CNRS | Divers | Images

MCC'2016 - Home

HypertorusGrid — P/T

	ITS-Tools	LTSMin	Tapaal(PAR)	Marcie	pnmc	PNXDD	Smart	Tapaal(EXP)	Tapaal(SEQ)	ydd-pt
Total	260	120	80	260	140	100	200	90	90	0
Bonuses	20	0	0	20	0	0	40	20	20	0
Scores	240	120	80	240	140	100	160	70	70	0
d2k1p8b00	87552 6.6763E+0005 32 36 TTTT ---:-- / 80	87552 ? 32 ? T-T- ---:-- / 60	87552 6.6763E+0005 32 36 TTTT ---:-- / 80	87552 6.6763E+0005 32 36 TTTT ---:-- / 80	87552 ? 32 36 T-TT ---:-- / 70	87552 ? ? ? T--- ---:-- / 50	87552 6.6763E+0005 32 36 TTTT ---:-- / 80	87552 ? 32 36 T-TT ---:M / 70	87552 ? 32 36 T-TT ---:P- / 70	CC 0
d2k2p1b00	5.1737E+0010 9.2167E+0011 4 32 TTTT ---:-- / 80	5.1737E+0010 ? 4 ? T-T- ---:-- / 60	CC 0	5.1737E+0010 9.2167E+0011 4 32 TTTT ---:-- / 80	5.1737E+0010 ? 4 32 T-TT ---:-- / 70	5.1737E+0010 ? ? ? T--- ---:-- / 50	5.1737E+0010 9.2167E+0011 4 32 TTTT ---:PM / 80	DNF 0	CC 0	CC 0
d2k3p2b04	5.4758E+0035 3.0194E+0037 12 144 TTTT ---:P- / 80	DNF 0	CC 0	5.4758E+0035 3.0194E+0037 12 144 TTTT ---:M / 80	DNF 0	CC 0	DNF 0	CC 0	CC 0	CC 0
d3k3p2b06	CC 0	DNF 0	CC 0	DNF 0	DNF 0	CC 0	DNF 0	CC 0	CC 0	CC 0
d4k3p2b08	CC 0	DNF 0	CC 0	DNF 0	DNF 0	CC 0	DNF 0	CC 0	CC 0	CC 0
d5k3p2b10	CC 0	CC 0	CC 0	DNF 0	DNF 0	CC 0	CC 0	CC 0	CC 0	CC 0

PaceMaker — P/T

	ITS-Tools	LTSMin	Tapaal(PAR)	Marcie	pnmc	PNXDD	Smart	Tapaal(EXP)	Tapaal(SEQ)	ydd-pt
Total	100	60	0	80	90	50	0	0	0	0
Bonuses	20	0	0	0	20	0	0	0	0	0
Scores	80	60	0	80	70	50	0	0	0	0
none	3.6803E+0017 8.0565E+0018 949 18085 TTTT ---:M / 80	3.6803E+0017 ? 949 ? T-T- ---:-- / 60	CC 0	3.6803E+0017 8.0565E+0018 949 18085 TTTT ---:-- / 80	3.6803E+0017 ? 949 18085 T-TT ---:P- / 70	3.6803E+0017 ? ? ? T--- ---:-- / 50	DNF 0	CC 0	CC 0	CC 0

As a Conclusion...

localhost

search | gest publis | HP fko | Equipe/Labo | TA fko | Projets | Conferences/revues | Enseignements | SAR | cloud CNRS | Divers | Images

MCC'2016 - Home

DLCshifumi — P/T

	ITS-Tools	LoLa	LTSMin	Tapaal(PAR)	Marcie	PeCan	Tapaal(EXP)	Tapaal(SEQ)
Total	**80**	**1050**	**80**	**5**	**80**	**-15**	**335**	**75**
Bonuses	0	320	0	0	0	0	60	20
Scores	80	730	80	5	80	-15	275	55
2a	TFFFTFFTTFTFTTTF TTTTTTTTTTTTTTTT ---:-- / 80	TFFFTFFTTFTFTTTF TTTTTTTTTTTTTTTT ---:PM / 80	TFFFTFFTTFTFTTTF TTTTTTTTTTTTTTTT ---:-- / 80	DNF 0	TFFFTFFTTFTFTTTF TTTTTTTTTTTTTTTT ---:-- / 80	???????????????? ---------- ---:-- / 0	?F?FT????F?FT??? -T-TT----T-TT--- ---:-- / 30	?F?FT????F?FT??? -T-TT----T-TT--- ---:-- / 30
2b	CC 0	TTFFFTFFFTTFTFFF TTTTTTTTTTTTTTTT ---:PM / 80	???????????????? ---------- ---:-- / 0	???????????T????? ----------T----- ---:-- / 5	DNF 0	???????????????? ---------- ---:-- / 0	???FF??FFTT?TF?? ---TT--TTTT-TT-- ---:-- / 40	???????FFT?????? -------TTT------ ---:-- / 15
3a	CC 0	TTTFTTFTFFFTFTTT TTTTTTTTTTTTTTTT ---:PM / 80	???????????????? ---------- ---:-- / 0	CC 0	DNF 0	T???????????????? T--------------- ---:-- / 5	T????????F??F??? T---------T--T--- ---:-- / 15	????????????F??? ----------T--- ---:-- / 5
3b	CC 0	FFTFFTFFFTTTTTTT TTTTTTTTTTTTTTTT ---:PM / 80	???????????????? ---------- ---:-- / 0	DNF 0	DNF 0	???????????????? ---------- ---:-- / 0	?FTF??FF?T?TTTTT -TTT--TT-T-TTTTT ---:-- / 55	DNF 0
4a	CC 0	FFTTTFTFTTTTTTFT TTTTTTTTTTTTTTTT ---:P- / 80	???????????????? ---------- ---:-- / 0	DNF 0	DNF 0	TT???????????????? XX---------- V--:-- / -20	??????TF?T??T??? ------TT-T--T--- ---:-- / 20	?????????T?????? ---------T------ ---:-M / 5
4b	CC 0	TFFTFFFTFFTTFFFT TTTTTTTTTTTTTTTT ---:PM / 80	???????????????? ---------- ---:-- / 0	DNF 0	DNF 0	???????????????? ---------- ---:-- / 0	TFF??F?TF???FFFT TTT--T-TT---TTTT ---:-- / 50	DNF 0
5a	CC 0	FTTTTFFTTTTTTFFT TTTTTTTTTTTTTTTT ---:PM / 80	???????????????? ---------- ---:-- / 0	DNF 0	DNF 0	???????????????? ---------- ---:-- / 0	???T?????T??T?F? ---T-----T--T-T- ---:-- / 20	DNF 0
5b	CC 0	?FFFTFFFF?F?FF?? ?TTTTTTTT?T-TT?? ---:P- / 55	???????????????? ?--------?----?? ---:-- / 0	CC 0	DNF 0	???????????????? ?--------?----?? ---:-- / 0	??F?????F??T?F?? ?-T-----T?-?-T?? ---:-M / 15	CC 0
6a	CC 0	FTFFFTFTFTTFFTFT TTTTTTTTTTTTTTTT ---:P- / 80	???????????????? ---------- ---:-- / 0	DNF 0	DNF 0	???????????????? ---------- ---:-- / 0	????????F??F???? --------T--T---- ---:-M / 10	DNF 0
6b	CC 0	?F??T?????F?TFTF ?T-?T??????T?TTTT ---:P- / 35	???????????????? ?--?-?????-?---- ---:-- / 0	CC 0	CC 0	???????????????? ?--?-?????-?---- ---:-- / 0	??F????????F?TF?F ?-??-??????T?TT-T ---:-M / 20	CC 0

DNAwalker — P/T

	ITS-Tools	LoLa	LTSMin	Tapaal(PAR)	Marcie	PeCan	Tapaal(EXP)	Tapaal(SEQ)
Total	**1880**	**1025**	**1295**	**625**	**880**	**-480**	**1065**	**780**
Bonuses	520	80	0	40	0	0	60	20
Scores	1360	945	1295	585	880	-480	1005	760

LIP6 UPMC
SORBONNE UNIVERSITÉS

And now…
let's have time for discussion

MCE
2016