Experimental Methodologies for Large-Scale Distributed Systems

Martin Quinson (team AlGorille)

Université de Lorraine

April 24, 2012

My Research Context

Scientific Objects

Large Scale Distributed Systems

- Scientific Computing
- High Performance Computing
- Grids
- Peer-to-peer Systems
- Volunteer Computing
- Cloud Computing

Scientific Questions

My Research Context

Scientific Objects

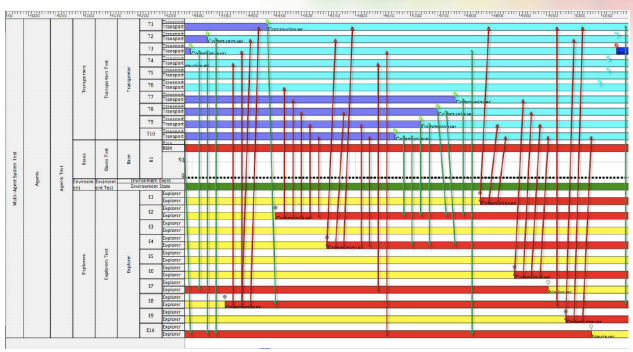
Large Scale Distributed Systems

- Scientific Computing
- High Performance Computing
- Grids
- Peer-to-peer Systems
- Volunteer Computing
- Cloud Computing

Scientific Questions

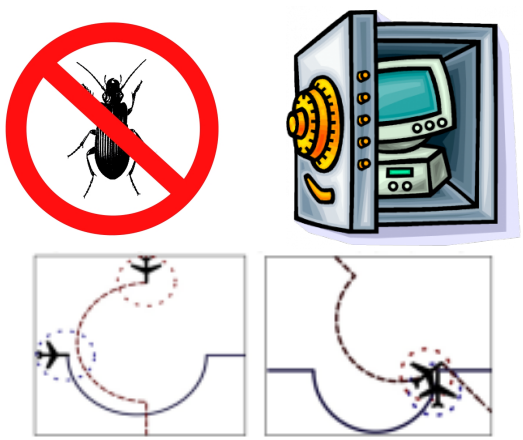
Performance

- User/Provider
- Time/Energy
- Throughput/Makespan/#Msg
- Worst case/Avg/Amortized



Correction

- Safety: bad things don't happen
- Liveness: good things do happen



My Research Context

Scientific Objects

Large Scale Distributed Systems

- Scientific Computing
- High Performance Computing
- Grids
- Peer-to-peer Systems
- Volunteer Computing
- Cloud Computing

Scientific Questions

Performance

- User/Provider
- Time/Energy
- Throughput/Makespan/#Msg
- Worst case/Avg/Amortized

Correction

- Safety: bad things don't happen
- Liveness: good things do happen

Methodo.

- Theoretical proofs
- Direct execution
- Experimental facilities
- Simulation
- Emulation

- Tests (manual/automated)
- Theorem proving
- Model checking
- Dynamic verification

My Research Context

Scientific Objects

Large Scale Distributed Systems

- Scientific Computing
- High Performance Computing
- Grids
- Peer-to-peer Systems
- Volunteer Computing
- Cloud Computing

Scientific Questions

Performance

- User/Provider
- Time/Energy
- Throughput/Makespan/#Msg
- Worst case/Avg/Amortized

Correction

- Safety: bad things don't happen
- Liveness: good things do happen

Methodo.

- Theoretical proofs
- Direct execution
- Experimental facilities
- Simulation
- Emulation

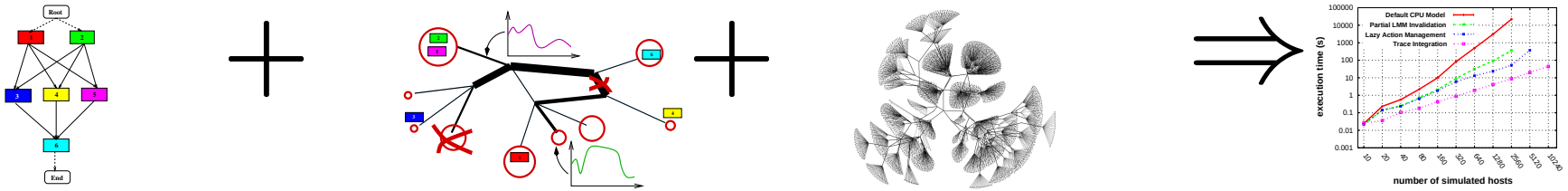
- Tests (manual/automated)
- Theorem proving
- Model checking
- Dynamic verification

My Research interests: Experimental Methodologies

- Meta-research about how to produce scientifically sound research
- Strive at developing ready-to-use tools addressing methodological challenges

SimGrid: Versatile Simulator of Distributed Apps

Idea to test | Experimental setup | System Model | Scientific Results



Scientific Instrument

- Allows studies of Grid, P2P, HPC, Volunteer Computing and other systems
- Validity limits studied and pushed further for years
- Scalable to death: 10M nodes on a single node; 1000× faster than others
- 100+ papers; collab CERN, IBM; 100+ users on ML; 5+ contributed tools

Scientific Object (and lab)

- Allows comparison of network and middleware performance models
- Experimental (but on par with SotA) Model Checker; Soon an emulator

Scientific Project since 12 years

- Collaboration Loria / Inria Rhône-Alpes / CCIN2P3 / U. Hawaii
- ANR's USS SimGrid (7 labs,13 P.M,0.8M€) and SONGS (7 labs,17 P.M,1.8M€)
- INRIA funding of engineers: ODL (06/08) and ADT (10/12)

Professional Production

Scientific Publications

- 1 book chapter; 8 journals and highly selective conferences ($< 33\%$)
  - Parallel Simulation of Peer-to-Peer. CCGrid'12
  - Scalable Multi-Purpose Network Representation for LSDA Simulation. CCGrid'12
- 19 other conferences and workshops; 4 research report; 7 tutorials (+1)
  - Assessing the Perf. of MPI Apps w/ Time-Independent Trace Replay. PSTI'11
  - SimGrid MC : Verification Support for a Multi-API Simulator. FORTE'11
  - Towards Scalable, Accurate, and Usable Simulations of Distributed Apps.
  - JLM: A Learning Management System Dedicated To Computer Science Education.

Scientific Expertise and Collective Duties

- Advisor of 2 PhD ('11: 1 defense, 1 start), 2 post-docs (+1) 11 masters (+2)
- Coordinator of 2 big ANR projects (+1); co-Coordinator of Grid'5000@Nancy
- Regular member of PhD Committees and Program Committees

Technical Production

- SimGrid (leading role since 2003); GRAS; Simulacrum; ALNeM
- JLM: Teaching of Java Programming; CSIRL: gentle unplugged intro to CS
  ISN-live: practical support to ISN teachings; IDEES: working group on CSE
  Castor: CS contest targeting the younger; CLE: System Programming in C

Current State of my HDR

Méthodologies d'expérimentation pour l'informatique distribuée à large échelle

- Contexte et état de l'art
  - Expérimentation et informatique
  - Taxonomie des systèmes distribués
  - Méthodologies expérimentales
- Simulation d'applications distribuées à large échelle
  - Simulation efficace de systèmes distribués, vers un système d'exploitation simulé
  - Nouveaux horizons pour la simulation
  - Vérification dynamique de propriétés de sécurité et vivacité
- Modélisation de l'environnement des applications
  - Génération réaliste de plates-formes pour la simulation d'application
  - Cartographie automatique d'un réseau d'interconnection en vue de simulation
  - Caractérisation des performances de communications des middlewares MPI
- Conclusions et perspectives
  - Principales contributions et collaborations
  - Vers un environnement intégré d'étude d'applications distribuées

15 pages of detailed outline

Conclusion

Why asking a délégation now? (verbatim from last year's conclusion)

- I need to finish what's ongoing, and get published the ideas that emerged
  Publication file may not really reflect my production yet
- I need to write the manuscript
- I'm short on time with 200-250 hours of teaching duty per year

This Year's Achievements

- 2 major publications, and 2 other accepted, 2 more submitted;
  1 PhD defended, 1 PhD started
- 1 huge ANR project warming up; 2 major SimGrid releases
- Plus, several mediation and CS Edu projects in line with ISN option

Expected Achievements Next Year

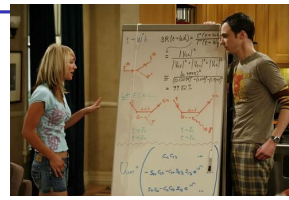
- Redaction effort to go from 15 to 150 pages on the HDR
- Starting ANR project, with huge coordination needs (5 labs out of 7 are INRIA)
- Follow up on ISN option and corresponding local and national efforts
- Seriously considering starting a new EPI on experimental methodologies

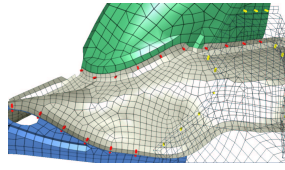
Agenda

- Experimental Methodologies for Large-Scale Distributed Systems
  - My Research Context
  - SimGrid: Versatile Simulator of Distributed Apps
  - Professional Production
  - Current State of my HDR
  - Conclusion

- Annexes
  - Assessing Distributed Applications Performance & Correction
  - Emulation
  - Other Contributions

Assessing Distributed Applications Performance





Classical Scientific Pillars Apply

- Theoretical Approach: Mathematical study of algorithms
- Experimental Science: Study applications on scientific instrument
- Computational Science: Simulation of a system model

Performance Study $\rightsquigarrow$ Experimentation

- Theory still mandatory, but everything's NP-hard
- Experimental Facilities: Real applications on Real platform *(in vivo)*
- Emulation: Real applications on Synthetic platforms *(in vitro)*
- Simulation: Prototypes of applications on system's Models *(in silico)*

	Experimental Facilities	Emulation	Simulation
Experimental Bias	☺☺	☺	☹
Experimental Control	☹☹	☺	☺☺
Ease of Use	☹	☹☹	☺☺

Assessing Distributed Applications Correction

- Absence of crash / data corruption (like always)
- Absence of race condition / deadlocks / livelocks (classic in multi-entities)
- Feal with lack of central time and central memory (specific to distributed)

Correction Assessment $\rightsquigarrow$ Formal Methods

- Facilities: Experience plans limited, by abilities or by time
- Simulation: How to decide if coverage is sufficient?
- Proof assistants: semi-automated proof demonstration (tedious for users)
- Model checking: Exhaustive state space exploration, search counter examples

	Experimental Facilities	Emulation	Simulation	Proofs	Model Checking
Performance Assessment	☺☺	☺☺	☺☺	☹☹	☹☹
Experimental Bias	☺☺	☺	☹	(n/a)	(n/a)
Experimental Control	☹☹	☺	☺☺	(n/a)	(n/a)
Ease of Use	☹	☹☹	☺☺	☹☹	☺
Correction Assessment	☹☹	☹	☹	☺☺	☺
Result if failed	(n/a)	(n/a)	(n/a)	☹	☺☺

Emulation as an Experimental Methodology

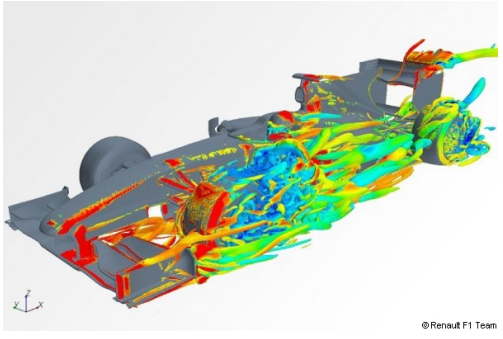
Execute your application in a perfectly controlled environment

- Real platforms are not controllable, so how to achieve this?
- Let's look at what engineers do in other fields

When you want to build a race car...

...adapted to wet tracks

...in a dry country ...



...you can simulate it.

Courtesy of Lucas Nussbaum

But then, you have

- To assess models
- Technical burden
- No real car

Why don't you... just control the climate? or tweak the car's reality?

Emulation as an Experimental Methodology

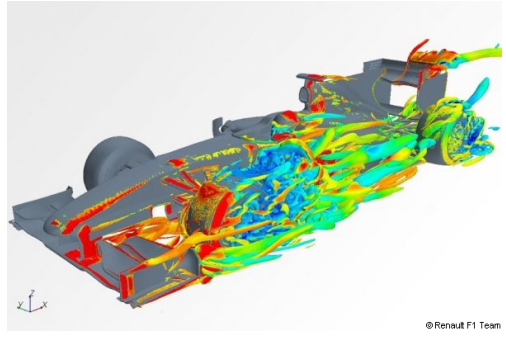
Execute your application in a perfectly controlled environment

- Real platforms are not controllable, so how to achieve this?
- Let's look at what engineers do in other fields

When you want to build a race car...

...adapted to wet tracks

...in a dry country ...



...you can simulate it.

But then, you have

- To assess models
- Technical burden
- No real car

Why don't you...

just control the climate?

or tweak the car's reality?

Courtesy of Lucas Nussbaum

Emulation in other Sciences

Studying earthquake effects on bridges

Studying tsunamis

Studying Coriolis effect and stratification vs. viscosity

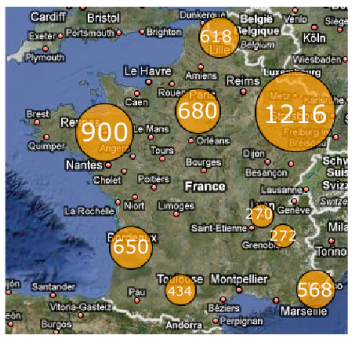
Studying climate change effects on ecosystems

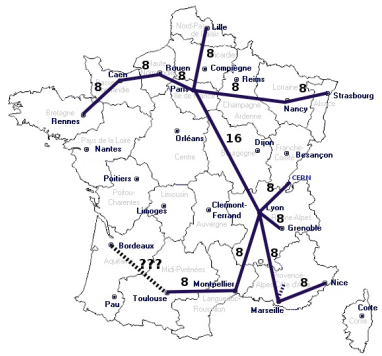
(who said that science is not fun??)

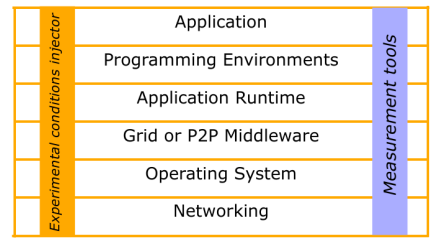
Contributions to Experimental Facilities (in vivo)

Grid'5000 Project: world leading scientific instrument for dist. apps

- Instrument for research in computer science (*deployment* of customized OSes)
  1500 nodes (2800 cpus, 7200 cores), 9 sites; dedicated 10Gb network







Personal Contributions

- National steering committee; Local project co-leader (CPER, Aladdin, Hemera)
- Scientific animation, event co-organization: Nancy is a leading site
- Collaboration: Production grids (IdG), CEA, Arcelor-Mittal

Project: Experimentation Process Industrialization (with L. Nussbaum)

- Open science: ensure that experiments can be shared, reviewed, improved
- Convergence of simulation and direct execution
- Methodological framework and practical tools (+administrative duties)

Other Contributions

Model-Checking (collaboration with S. Merz and C. Rosa)

- **Goal:** democratize Formal Methods to non specialists through SimGrid
- Achievements:
  - Model-checking mode in SimGrid; Generic modeling of communications API
  - DPOR implementation fighting combinatorial explosion regardless of used API
- Projects:
  - Integrate Liveness Properties; Automatically bridge code $\leftrightarrow$ model variables
  - Long Term: *semantic* debugger of distributed applications within SimGrid
  - Very Long term: Performance checking (time discrete at best in MC)

Simulated MPI (collaboration with S. Genaud, H. Casanova, F. Suter, P.N. Clauss)

- **Goal:** study real applications based on MPI within SimGrid
- Achievements: Partial implem of MPI; Assessment of LAN models
- Projects: Modeling collectives' *Semantic* ($\rightsquigarrow$ MPI-3); Trace based simulation

Study of Real Applications: SimTerpose (collaboration with L. Nussbaum)

- **Goal:** intercept every actions of the application, and study them online
- Achievements: Prototype of interceptor; Projects: TBC, and used

\+ PlusCal (MC$\rightarrow$Sim with Lamport); GRAS, Alnem; Energy, DistSim; JLM, CLE