Savoir Modéliser les systèmes à grande échelle et Valider leurs simulateurs

Thème 7

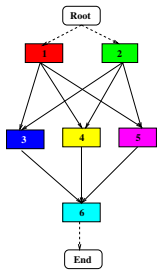
Arnaud Legrand (Grenoble), Martin Quinson (Nancy)

5 octobre 2010

Simulating Distributed Systems

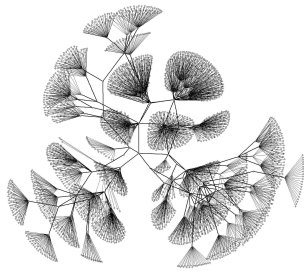
Principle

Idea to test



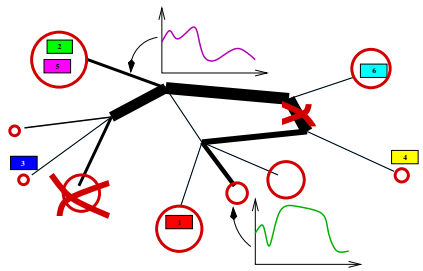
+

System Model



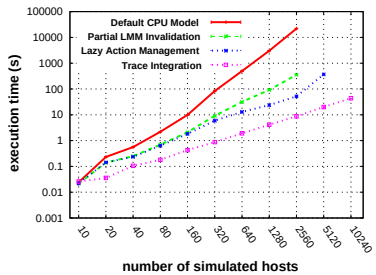
+

Experimental setup



⇒

Scientific results



Advantages

- Less simplistic than proposed theoretical models (which are useful too)
- Better XP control ($\rightsquigarrow$ reproducible) than production systems ($+$ not disruptive)
- Not as tedious, time/labor consuming than experimental platforms
- Plus: Lower technical burden; Quick and easy experiments; What if analysis

Main challenges

- Validity: Get realistic results (controlled and understood experimental bias)
- Scalability: Simulate *fast enough* problems *big enough*
- Usability: Associated Tools; Ease of use; Applicability to context of interest

Modéliser les systèmes à grande échelle

Qu'est ce qu'un modèle

- collection d'éléments + ensemble de règles régissant leurs interactions
- Utilisable pour penser à propos de la réalité et faire des prédictions

$\Rightarrow$ *Graal* de tous les scientifiques

Propriétés souhaités

- Réaliste et Précis: capture la réalité à propos de l'objet d'intérêt
- Manipulable (*tractable*): reste analysable dans les faits
  (taille raisonnable, algorithmes non-exponentiels, etc)
- Instanciable: on sait décrire une plate-forme donnée dans ce formalisme

Intuition: la simulation a trop de biais expérimentaux pour être utile

- On pense qu'il est impossible d'avoir ainsi des prédictions à la nanoseconde

Modéliser les systèmes à grande échelle

Qu'est ce qu'un modèle

- collection d'éléments + ensemble de règles régissant leurs interactions
- Utilisable pour penser à propos de la réalité et faire des prédictions

$\Rightarrow$ *Graal* de tous les scientifiques

Propriétés souhaités

- Réaliste et Précis: capture la réalité à propos de l'objet d'intérêt
- Manipulable (*tractable*): reste analysable dans les faits
  (taille raisonnable, algorithmes non-exponentiels, etc)
- Instanciable: on sait décrire une plate-forme donnée dans ce formalisme

Intuition: la simulation a trop de biais expérimentaux pour être utile

- On pense qu'il est impossible d'avoir ainsi des prédictions à la nanoseconde
- Effectivement, c'est impossible.
- Mais une expérience sur Grid'5000 ne permet pas de prédire le comportement sur Egee. Même pas à la minute près.

Validity Challenge

Context: Models in most simulators are either simplistic, wrong or not assessed

- PeerSim: discrete time, application=automaton; GridSim: naive packet level
- OptorSim, GroudSim: documented as wrong on heterogeneous platforms

Quality Levels of Validity

- Level -1: not validated (probably plainly wrong)
- Level 0 (visually ok): a few curves that look similar (generally hides a lot)
- Level 1 (ratios ok): $A < B$ in Simulation $\Leftrightarrow$ $A < B$ in Reality
- Level 2 (prediction abilities): bounded distance between simul. and reality
- Orthogonal to this: need to assess when the model is valid (validity domain)
- Validity eval: tricky, requires meticulous attention & sound methodology

SimGrid validity 2 years ago: Research focus in SimGrid since 2002

Setting: *Synthetic* App. + *Synthetic* WAN; Compare vs. packet-level simulator

- Error in percents if: TCP steady state (flows $>$ 10Mb), latency-bound (WAN)
- Wrong estimations when capacity-bound (suspect: max-min sharing)

Current Work on Validity in SimGrid

First Step: *Synthetic* App. + *Synthetic* WAN. Compare against *GTNetS*

- Some errors were hunted down + unexpected phenomenon were understood

$\rightsquigarrow$ The model and its instanciation were considerably improved
Widen validity range to flows $>$ 100Kb and WAN with small latencies

- Sharing mechanism from theoretical literature experimentally proved wrong

Current Work on Validity in SimGrid

First Step: *Synthetic* App. + *Synthetic* WAN. Compare against *GTNetS*

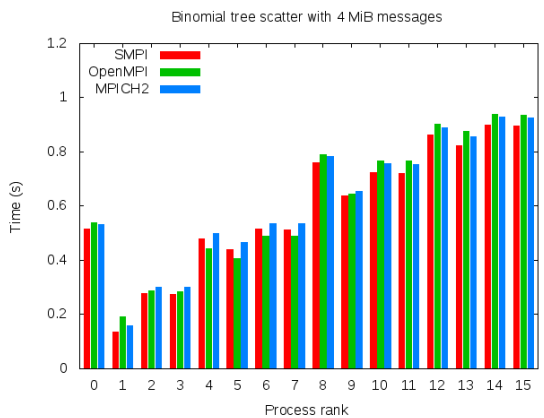
- Some errors were hunted down + unexpected phenomenon were understood

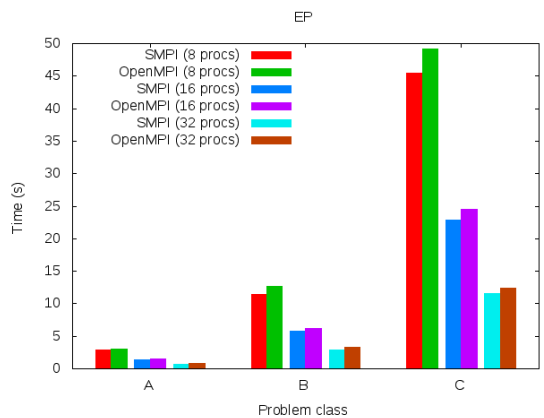
$\rightsquigarrow$ The model and its instanciation were considerably improved
Widen validity range to flows $>$ 100Kb and WAN with small latencies

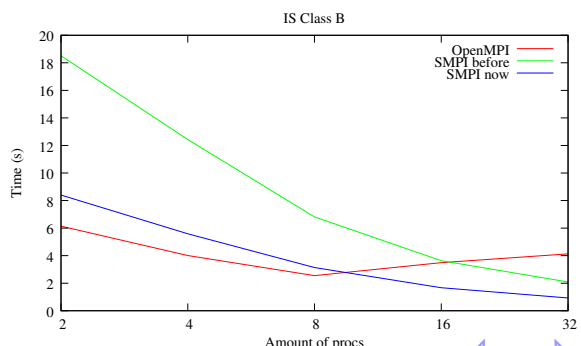
- Sharing mechanism from theoretical literature experimentally proved wrong

Going Further: developed SMPI $\rightsquigarrow$ Real App. (NAS PB) + clusters (LAN)

- Good prediction for short messages is crucial (piecewise linear)
- Need to accurately implement/model collective operation algorithms
- Evaluating weight of computation phases tricky, numerical instabilities deadly
- Need to account for MPI overhead; what is Real with several MPI implems?







Pourquoi est-ce si dur?

En informatique

- Mesures très sensibles aux conditions extérieurs et initiales
- Évolution technologique invalide rapidement les mesures

$\Rightarrow$ Encore plus de précautions nécessaires lors de la manipulation de modèles

$\Rightarrow$ Il faut vérifier la validité des modèles, détailler les conditions de mesures

Pratiques actuelles rarement irréprochables

- Domaine de validité des modèles flou; Usage des modèles à tors et à travers
- Utilisation de simulateur sans réflexion sur la validité du modèle

Modéliser la bonne chose

> *A theory has only the alternative of being right or wrong.*
> *A model has a third possibility: it may be right, but irrelevant.*
> – Manfred Eigen

Axes de recherche

Défi 1: Un instrument scientifique pour les spécialistes du domaine

⇒ Kernel de simulation de systèmes à large échelle

⇒ Modèles validés (ie, marge d'erreur connue selon l'intervale de paramètres)

- Mise au point d'un framework théorique pour valider un modèle
- Grid'5000 est l'outil parfait pour ces études

Défi 2: Reconstruire une description fiable du système Grid'5000

- Parallèle à l'Observatoire des Grilles pour Grid'5000
- (plus des générateurs synthétiques et un cartographe automatique)

Rapports étroits avec le Working Group 6 (expérimentation)

- Mêmes problèmes; la simulation $\rightsquigarrow$ moins de problèmes techniques
- Simterpose = émulation par simulation et non par resource trashing

Par rapport à Héméra (1/2)

Nos besoins vis-à-vis d'Héméra

- Une description précise, dynamique, de niveau applicatif de Grid'5000
- Matériel: aucun (G5K actuel amplement suffisant – hétérogénéité?)
- Humain: aucun (maintenance d'instruments scientifiques lourds, mais ANR, ADT)

Ce que nous apportons

- Un défrichage des plans d'expériences en informatique
- Outils associés mis au point dans des conditions plus simples
  - Visualisation: la suite de Pajé
  - Post-processing: analyse semi-automatique et comparaison de traces
- Des briques expérimentales comme simterpose
- Des contacts avec les grilles de production
  - Dans SimGLite, simulation et émulation sont deux approches complémentaires
  - SimData ouvre les portes du CERN

Autres interactions au sein d'Héméra

SimGrid est un bon terrain de jeu pour certains d'entre vous

- P2P-Churn: SimGrid utilise des traces de panne; intégration/tests de modèles?
- Energy: Ca fait quasi 2 ans qu'on dit que je vais le faire (honte, honte)
- COP: 140,000 processus sur une seule machine, c'est (dur mais) possible

Autres interactions au sein d'Héméra

SimGrid est un bon terrain de jeu pour certains d'entre vous

- P2P-Churn: SimGrid utilise des traces de panne; intégration/tests de modèles?
- Energy: Ca fait quasi 2 ans qu'on dit que je vais le faire (honte, honte)
- COP: 140,000 processus sur une seule machine, c'est (dur mais) possible

Conclusion: l'avenir du WG7 d'Héméra

- Fin WG validation; Intégration de approche expérimentale par simulation?
- Renommer le WG pour redécouper?
- Ignorer les ptits problèmes de nommage et avancer comme ça?

  je n'ai pas vraiment de réponse...