Clouds/Big Data @ Inria

Frédéric Desprez
Frederic.Desprez@inria.fr

May 20, 2013

Outline

1. Inria Strategy in Clouds
2. HPC, Clouds: Where within Inria?
3. Inria Large-Scale initiatives

Introduction

Cloud computing has emerged as a new paradigm for many commercial and scientific venues

- Starts to be widely adopted by the industries
- Many platforms available around the world
- Several offers for IaaS and PaaS platforms

– Still many applications left that could benefit from such platforms

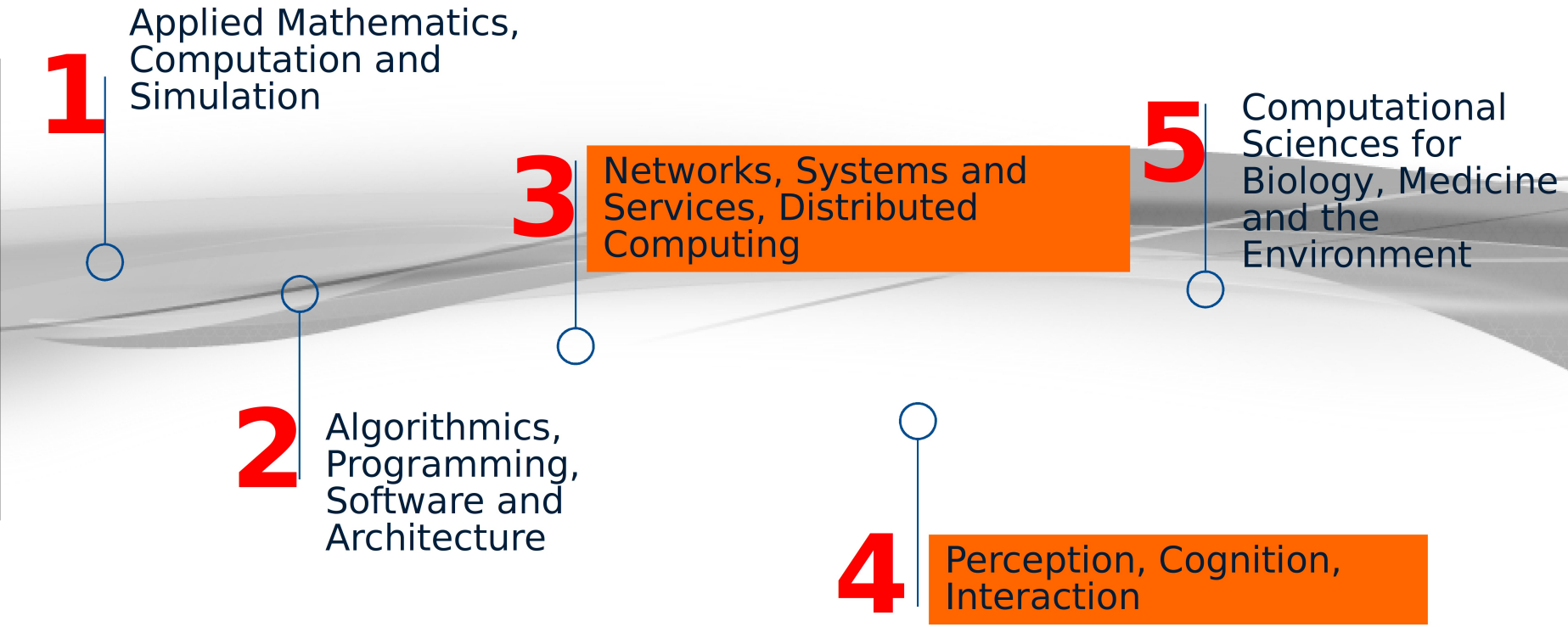
• Elasticity, availability, self-configuration, heterogeneous computing and storage capacities

– Several challenges remain to be addressed and transfered into industrial products

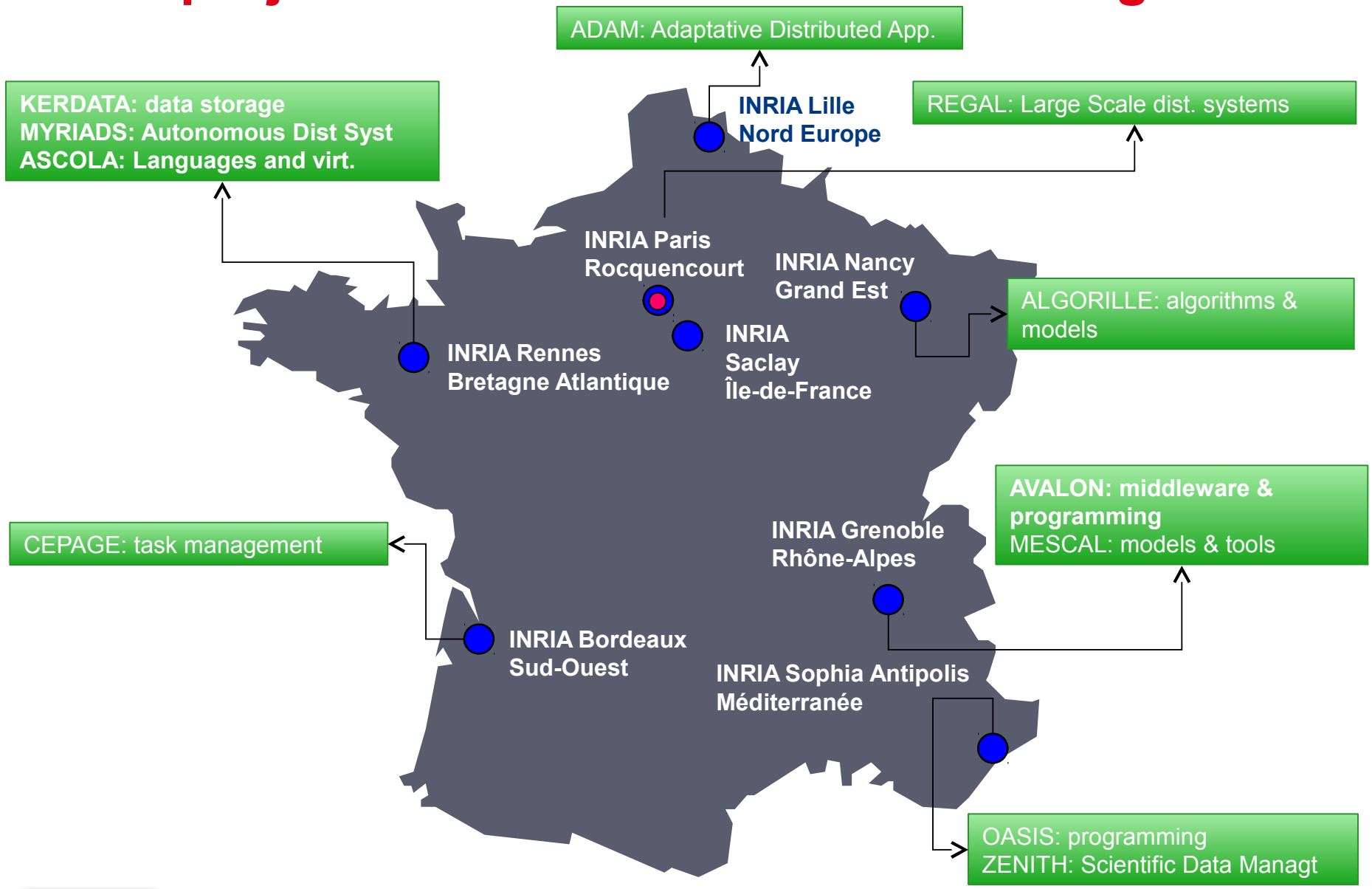
Some research issues in Cloud Computing

- Reliability / Resilience / Fault-tolerance
- Trust, Security and Privacy
- New economical models for computing
- Service Level Agreement / Quality of Service
  - From Best Effort to SLA
- Building cloud-aware applications from legacy applications
- Cloud federation
- Autonomic behaviors / Self-*
- Brokering / Scheduling
- Programming models (MapReduce, ...)
- Interactions between legal aspects (laws) and computer science
  - privacy and liability

Clouds/Big Data: where within Inria ?



Some project-teams involved in Clouds/BigData



Current activities around Clouds@ INRIA

- Resource allocation and management
  Avalon, Cepage, Regal
- Energy management
  Avalon, Myriads, Ascola
- VM management
  Ascola, Myriads
- Big Data management
  KerData, Zenith
- MapReduce paradigm
  Avalon, Kerdata
- OS
  Ascola, Myriads
- Simulation
  Algorille, Ascola, Avalon, Mescal

- Programming models
  Avalon, OASIS
- Unconventional paradigms
  Myriads
- Model Driven Engineering
  Triskell
- Adaptative Dist. Applications
  ADAM
- PaaS
  Avalon, Myriads, ADAM
- IaaS
  Ascola, Myriads
- Scalable data analysis
  Zenith

Some other joint research activities

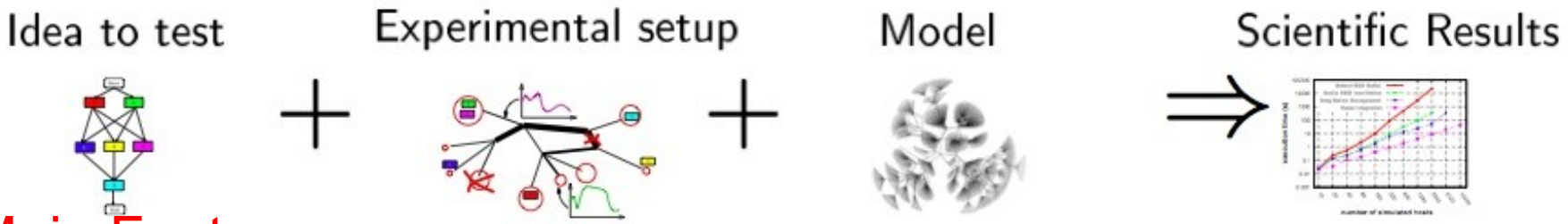
- FP7 projects
  - CONTRAIL (Myriads)

    Development of an integrated approach to virtualization, offering Infrastructure as a Service (IaaS), services for federating IaaS Clouds, and Platform as a Service (PaaS) on top of federated Clouds.
  - BONFIRE, FED4FIRE (Myriads)

    Cloud Computing testbeds
- Many ANR projects
- MSR-INRIA collaboration using Microsoft Azure (Kerdata)
- ARGONNE/URBANA Champaign/INRIA joint lab
- FUTUREGRID

SimGrid: Simulator of Distributed Applications

Scientific instrument for the study of large scale distributed computing



Main Features

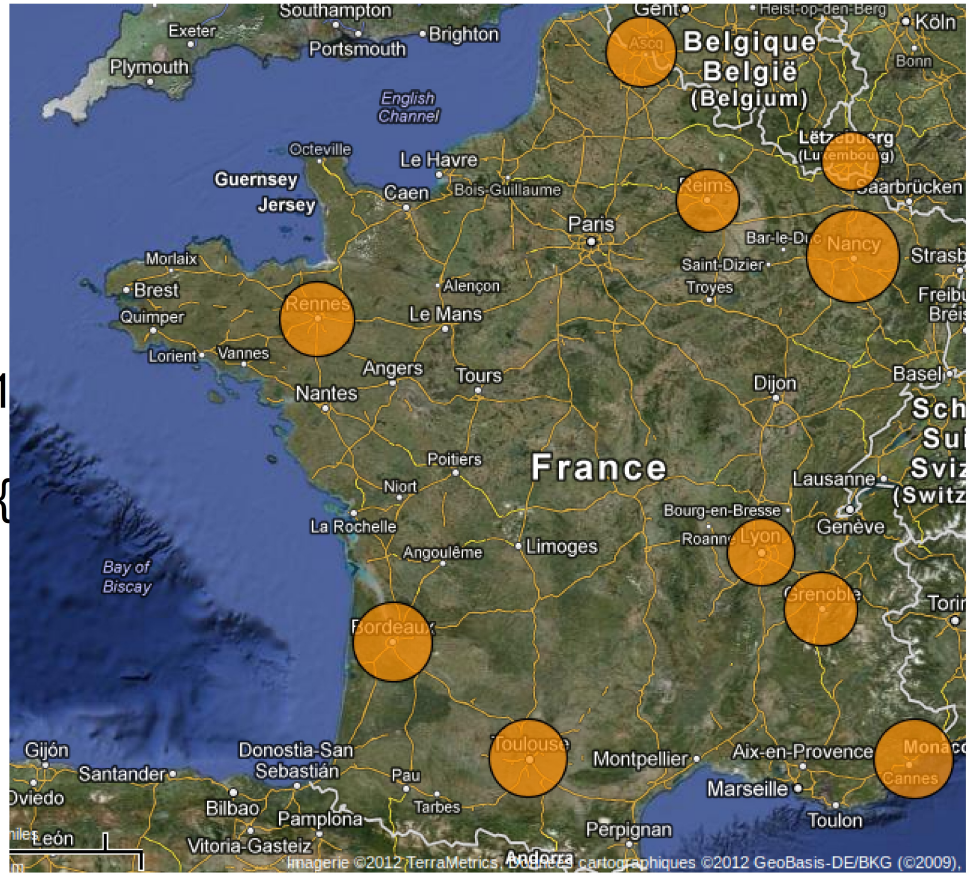
- Versatile: Grid, P2P, HPC, Volunteer Computing, Clouds, . . .
- Valid: Accuracy limits studied and pushed further for years
- Scalable: 3M chord nodes; 1000× faster than other (despite precise models)
- Usable: Tooling (generators, runner, vizu); Open-source, Portable, . . .

GRID'5000

- **Testbed for research on distributed systems**
  - Born from the observation that me need a better and larger testbed
  - High Performance Computing, Grids, Peer-to-peer systems, Cloud computing
  - A complete access to the nodes' hardware in an exclusive mode (from one node to the whole infrastructure)
  - RIaaS : Real Infrastructure as a Service ! ?
- **One rule:** only for research on distributed systems
  - → no production usage
- Free nodes during daytime to prepare experiments
- Large-scale experiments during nights and week-ends
- Software stack: Resource management (OAR), system reconfiguration (Kadeploy), network isolation (KaVLAN), monitoring (Ganglia, Kaspied, Energy), GRID'5000 API

Current Status

- 10 sites (1 outside France)
  - New sites are joining the infrastructure (Nantes, Porto-Allegre)
  - 1195 machines, 8184 cores
- **Diverse technologies**
  - Intel (60%), AMD
  - CPUs from one to 1
  - Myrinet, Infiniband {
  - Two GPU clusters
- More than **500 users** per year
- **1826 users** since 2004



Recherche en Cloud à Inria:
Focus sur la sécurité

Jonathan Rouzaud-Cornabas

Inria – CNRS – CC-IN2P3 / LIP (UMR 5668)

Motivation

- Les clouds sont de plus en plus utilisés pour tout (et n'importe quoi);

Motivation

- Les clouds sont de plus en plus utilisés pour tout (et n'importe quoi);
- Pour l'utilisateur, le cloud est une boite noire:
  - L'utilisateur ne peut pas savoir qui et quoi accéde à ses données;
  - L'utilisateur n'a pas de moyen de surveiller les actions sur ces données;
  - L'utilisateur ne peut pas être sûr que les actions qu'il effectue sont réellement exécuter e.g. suppression des données;

Motivation

- Les clouds sont de plus en plus utilisés pour tout (et n'importe quoi);
- Pour l'utilisateur, le cloud est une boite noire:
  - L'utilisateur ne peut pas savoir qui et quoi accéde à ses données;
  - L'utilisateur n'a pas de moyen de surveiller les actions sur ces données;
  - L'utilisateur ne peut pas être sûr que les actions qu'il effectue sont réellement exécuter e.g. suppression des données;
- La sécurité est critique pour les fournisseurs de Cloud: réputation;
  - Le sabotage est un risque aussi voir plus grand que n'importe quelle grande société;
  - La perte de confiance ne peut pas être composée: faillite;

Multi-utilisateurs

- Conflit entre les utilisateurs: objectifs non-compatible
  - Les utilisateurs partagent un ensemble de ressources et ont des objectifs opposés

Multi-utilisateurs

- Conflit entre les utilisateurs: objectifs non-compatible
  - Les utilisateurs partagent un ensemble de ressources et ont des objectifs opposés

- Comment gérer des utilisateurs ayant des conflits d'intérêts ?
  - Est-ce que les utilisateurs peuvent partager une infrastructure et ne pas s'attaquer ?
  - Si ils ne peuvent pas, comment les isoler ?

Multi-utilisateurs

- Conflit entre les utilisateurs: objectifs non-compatible
  - Les utilisateurs partagent un ensemble de ressources et ont des objectifs opposés
- Comment gérer des utilisateurs ayant des conflits d'intérêts ?
  - Est-ce que les utilisateurs peuvent partager une infrastructure et ne pas s'attaquer ?
  - Si ils ne peuvent pas, comment les isoler ?
- Comment fournir une séparation entre les utilisateurs ?

La virtualisation n'est pas la sécurité !

- L'isolation n'est pas aussi forte que ce qu'on pourrait penser;

  Whispers in the Hyper-space: High-speed Covert Channel Attacks in the Cloud, Z. Wu and al., Usenix Security 2012

- Observer la consommation de ressources des autres machines virtuelles (side-channel attack);

  $\Rightarrow$ Fonctionne sur public Cloud

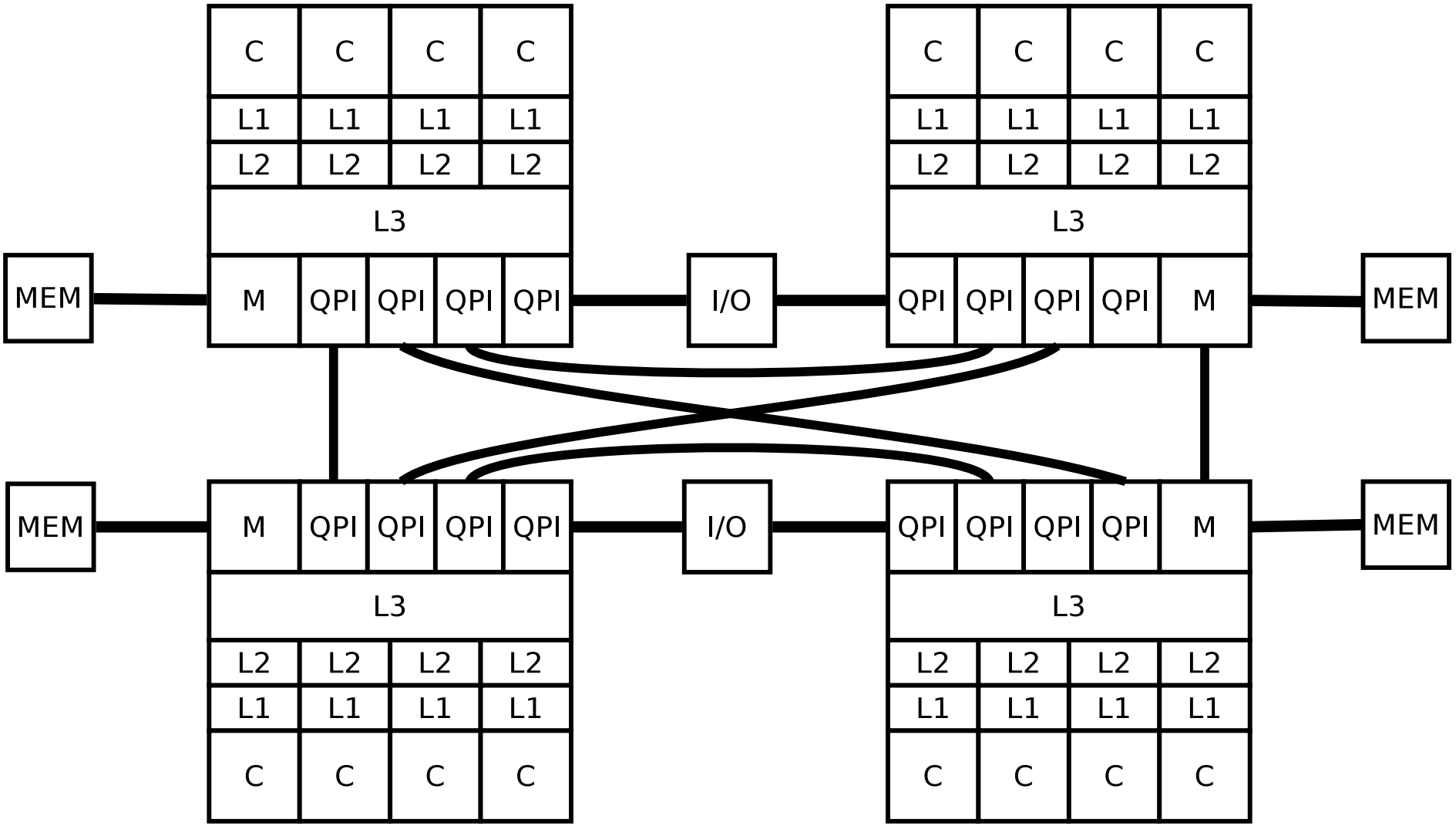
  Resource-Freeing Attacks: Improve your Cloud Performance (At your Neighbor's Expense), V. Varadarajan and al, ACM CCS 2012

- Récupérer et/ou modifier des données dans les ressources partager e.g. L2 Cache (covert channel);

  $\Rightarrow$ Vol de clef de cryptographie sur public Cloud

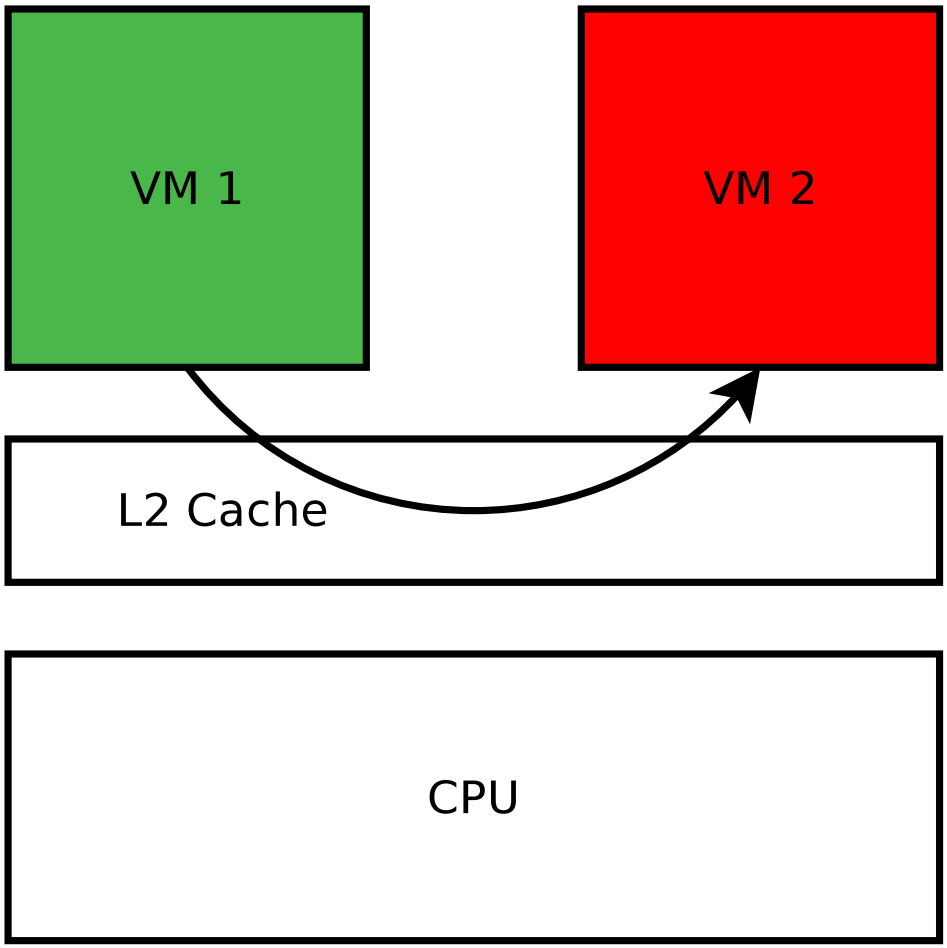
  Cross-VM Side Channels and their Use to Extract Private Keys, Y. Zhang and al., ACM CCS 2012

Micro-architecture des processeurs modernes



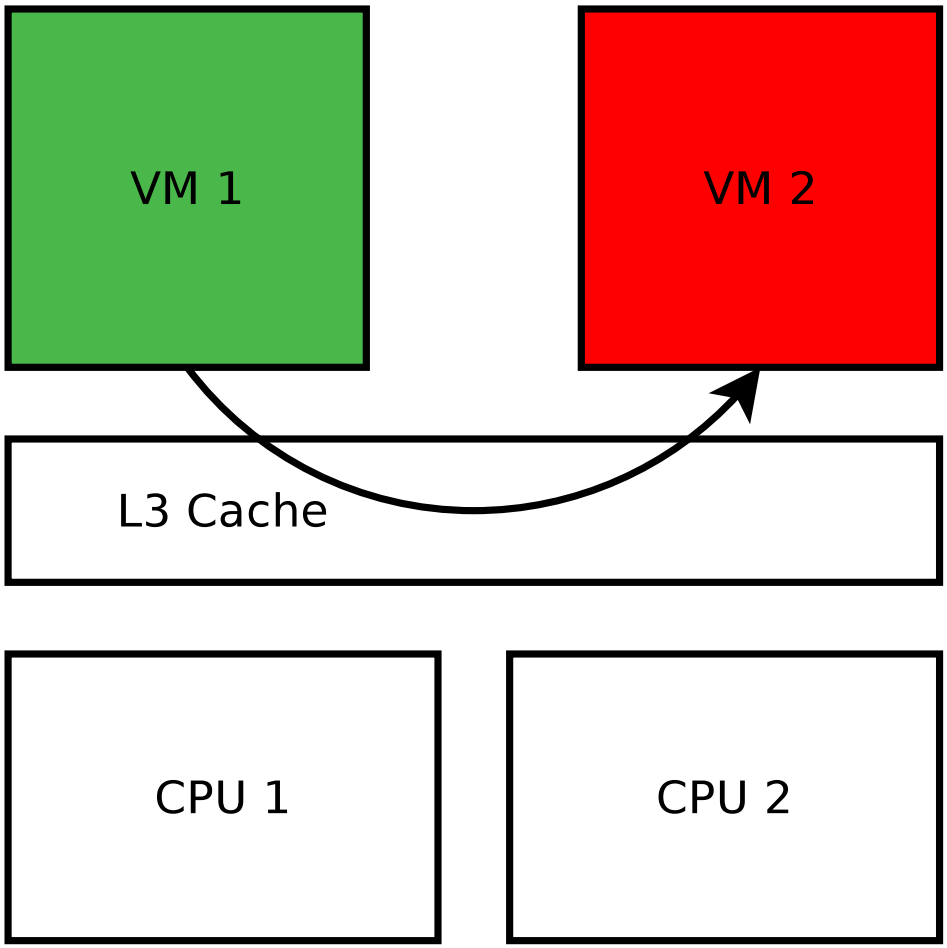
La virtualisation n'est pas la sécurité !

- Example de faille de virtualisation



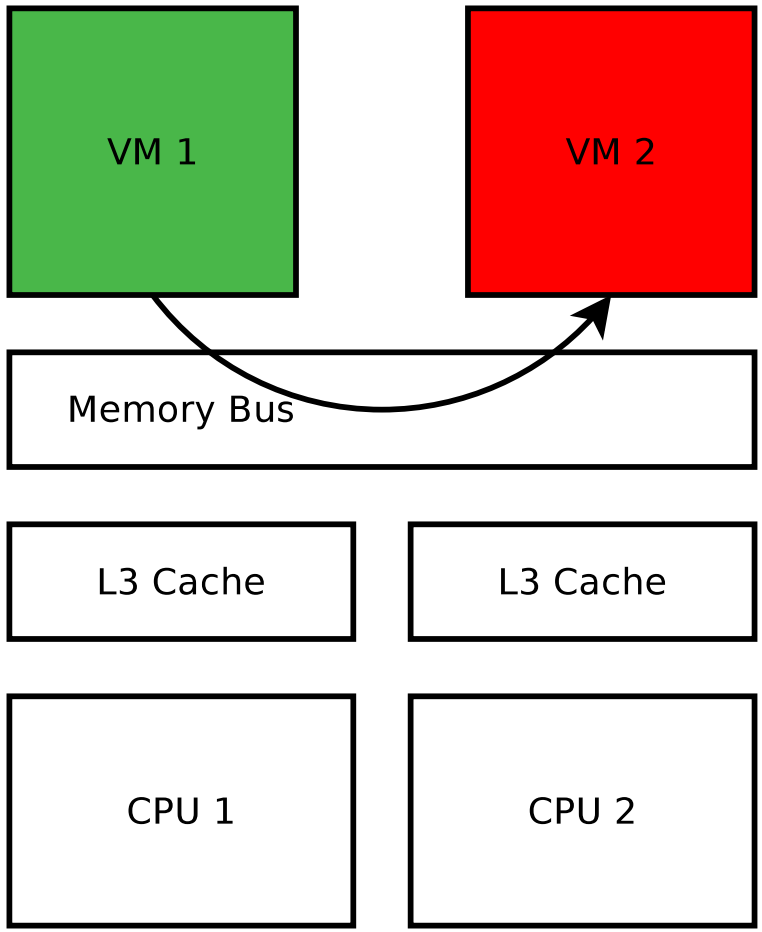
La virtualisation n'est pas la sécurité !

- Example de faille de virtualisation



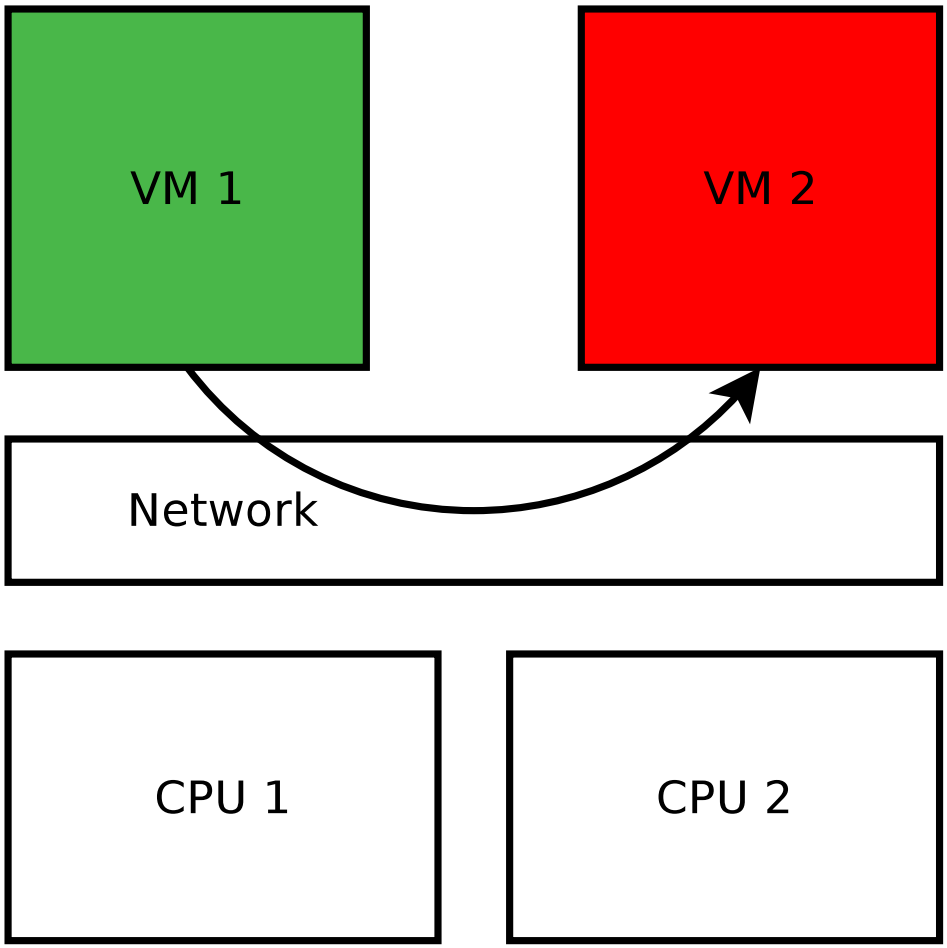
La virtualisation n'est pas la sécurité !

- Example de faille de virtualisation



La virtualisation n'est pas la sécurité !

- Example de faille de virtualisation

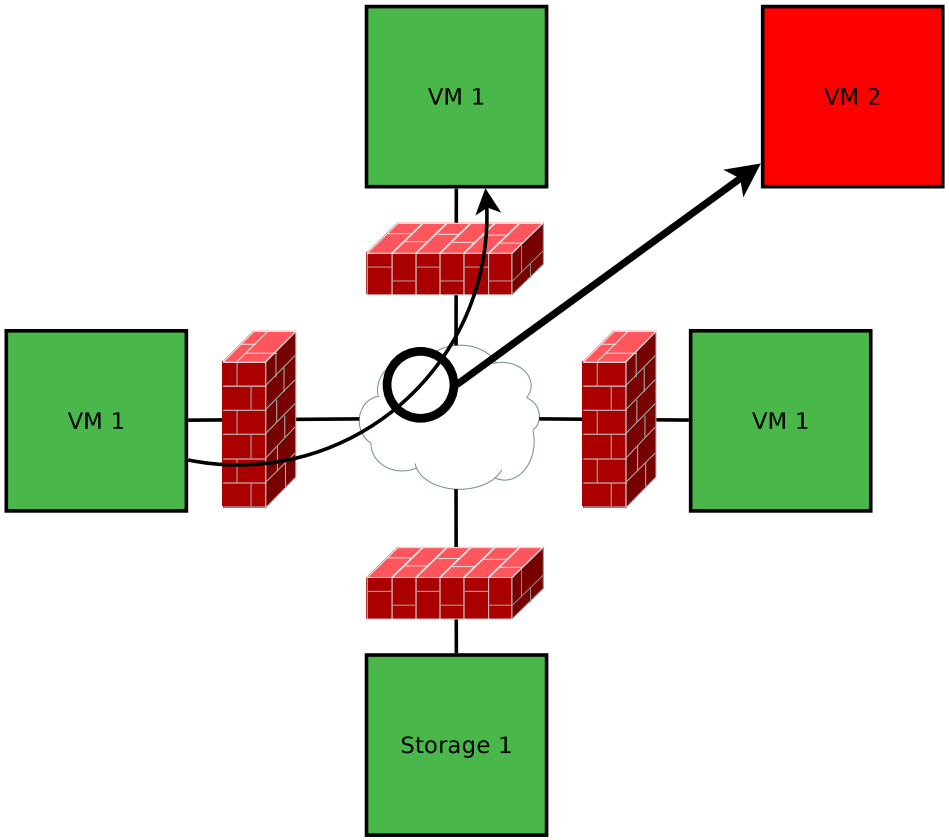


La virtualisation n'est pas la sécurité !

- Example de faille de virtualisation $\rightarrow$ **Sécurité dans les Clouds**;
- Pas seulement sécuriser les logiciels qui tournent dans les machines virtuelles

La virtualisation n'est pas la sécurité !

- Example de faille de virtualisation $\rightarrow$ **Sécurité dans les Clouds**;
- Pas seulement sécuriser les logiciels qui tournent dans les machines virtuelles



La virtualisation n'est pas la sécurité !

- Example de faille de virtualisation $\rightarrow$ **Sécurité dans les Clouds**;
- Pas seulement sécuriser les logiciels qui tournent dans les machines virtuelles $\rightarrow$ **Administrer et garantir la sécurité d'un ensemble d'ordinateurs vu comme une entité unique**;

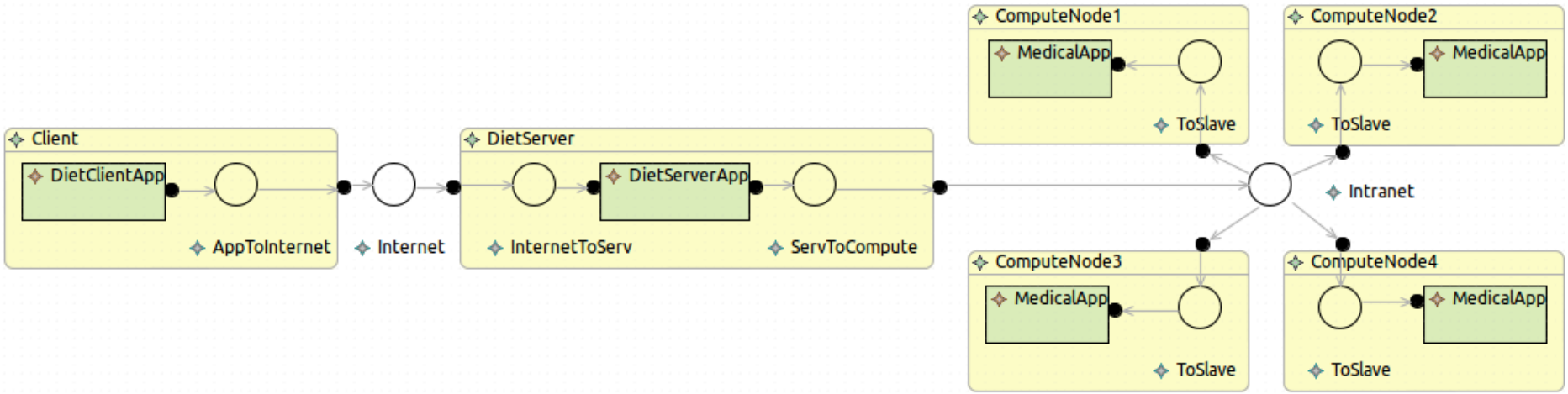
La virtualisation n'est pas la sécurité !

- Example de faille de virtualisation $\rightarrow$ **Sécurité dans les Clouds**;
- Pas seulement sécuriser les logiciels qui tournent dans les machines virtuelles $\rightarrow$ **Administrer et garantir la sécurité d'un ensemble d'ordinateurs vu comme une entité unique**;
- La centralisation de la sécurité n'est pas une solution: Problème de passage à l'échelle;

Hot Topics en sécurité & Clouds

- Stockage distribué et Sécurité
- Big Data/Analytics et Sécurité
- Contrôle d'accès dans le Cloud
- Localisation et application des lois sur les Clouds
- Sécurité de la virtualisation: calcul, réseau et stockage
- Securité et Privacy
- Sécurité des fédérations de Clouds
- Application des avancées en cryptographie aux Clouds: calcul sur des données chiffrées: Fully Homomorphic Encryption
- Authentification et Identification dans le Cloud

Security-Aware Models for Clouds




```
// Sec properties for intranet
isolation({ctx_intranet})
confidentiality({ctx_intranet})
integrity({ctx_intranet})

// Sec property for compute nodes
isolation({ctx_computeNodes})
```

Méthodologie: Simulation

- **Défis scientifiques**: Comment évaluer la recherche sur les Clouds ?
  - Plate-formes difficile d'accès ou pas encore existantes
  - Coût élevé (budget et temps)
  - Exactitude des simulations
- **Objectifs**
  - Simulateur pour les plate-formes mutualisées, fortement distribuées et hétérogènes
  - Modèles de comptabilité, facturation, performance et comportements
  - Validation des simulations $\Rightarrow$ Exactitude
- Travaux commencés (ANR SONGS)