Research Issues for Future Cloud Infrastructures

Frédéric Desprez
Frederic.Desprez@inria.fr

Joint work with M. Dias de Assuncao (Inria), I. Chrisment (Telecom Nancy), J. Francois (Inria), A. Lebre (Inria), L. Lefèvre (Inria), D. Margery (Inria), P. Merle (Inria), L. Nussbaum (Université de Lorraine), C. Perez (Inria), G. Pierre (IRISA), D. Saucez (Inria), and P. Valduriez (Inria)

Context

- Cloud offers have reached a certain level of maturity
- Critical infrastructures within the Internet ecosystem
- Convergence of the Internet of Things and the data center worlds
  - Massively distributed federations of smaller datacenters placed at the edge of network backbones (Fog/Edge computing)
  - Advances in the capacities and capabilities of mobile networks and end-user devices
- Several challenges around resource management and efficient use of platforms by applications
- Many research issues given the evolution of the Cloud and its target applications

Outline

1. Infrastructure/Application Management
2. Network “softwarization” and Network as a Service
3. Energy Proportionality
4. Formalization of Cloud Computing solutions
5. Data-Intensive scalable computing
6. Experimental driven research in Cloud Computing

Infrastructure/Application Management

- **Elasticity management**
  - Horizontal replication, vertical scaling, migration
  - Dynamic optimization policies (reactive, predictive)
  - Elasticity = scaling + automation + optimization
- **Challenges**
  - How developers describe the architecture of their application
    - Component models (TOSCA, CAMEL) have limited support for application variability
    - Application composability
  - “A la carte” framework for elasticity
    - Support of any scaling techniques and optimization
    - Automatic and dynamic selection of the right elasticity strategy
    - Multi-level loops of optimization
  - Operating, supervising and managing Fog/Edge infrastructures
    - More advanced systems should be proposed to handle dynamicity, heterogeneity, mobility issues and other specific issues of Fog/Edge Infrastructures

Network “softwarization” and Network as a Service

- Network Function Virtualization (NFV) and Software Defined Networking (SDN)
  - Composition of network services in an elastic manner using APIs
  - Network integrated in applications as another component
- **Challenges**
  - New concept of Service Function Chaining (SFC), building on the fly network services by deploying them on the cloud
    - Variability and outages issues, malicious behavior of operators
    - Develop services that can deliver a comprehensive monitoring of the resources and dynamically determine the number of instances to be deployed for each function
  - Limit of the network virtualization
    - Target network infrastructures for critical services (planes, power plants, factories)
    - Instead of replicating the network to isolate their different functionalities, use virtualization and network services

Energy Proportionality

- Energy consumption is one of the main limiting factor for deploying large sets of physical resources and equipment in Cloud datacenters
- Overprovisioning used in cloud offers to react to possible non predicted usages

⇒ large non energy-proportional infrastructures

- **Challenges**
  - Design of new energy models, algorithms and frameworks to be able to express and support energy proportional consumption of virtualized infrastructures
  - Use of software defined infrastructures, protocols, tools, and models allows new level of flexibility in Clouds
  - Design of metrics and approaches to combine this flexibility with multi-objectives models to support trade-off between performance, energy efficiency, QoS, etc.
  - Use of renewable energy for provisioning large datacenters

Formalization of Cloud Computing Solutions

- High complexity of Cloud platforms
  - Many definitions of Cloud Computing, several characteristics and services and deployment models
  - Hundreds of cloud offers
  - Heterogeneity of provided services, various standards
  - Collection of informal and ambiguous definitions, non interoperable technological offers and poorly deployed standards
- Lack of a theory of Cloud Computing
- **Challenges**
  - Providing a formal definition of cloud computing
  - Specify its foundations mathematically
  - Capture and specify any functional and non-functional characteristics of any computational resources
    - Reason and prove properties
    - Address semantic interoperability in Cloud Computing

Data-Intensive Scalable Computing

- Convergence of High Performance Computing (HPC) and Data-Intensive Scalable Computing (DISC)
  - HPC: compute centric and simulation applications on supercomputers
  - DISC: data-centric and focus on fault-tolerant and scalability of web and cloud applications on cost-effective clusters of commodity hardware
- **Challenges**
  - Developing architectures and methods to combine simulation and data analysis
    - Distinct approaches depending on where analysis is done: post-processing, in situ and in-transit
  - Research around stream processing
    - Operational monitoring of large infrastructures, IoT, smart cities
    - Models using resources available at the edge of the Internet
    - Optimizing the placement of data processing tasks while minimizing the use of network resources and latency, elasticity management

Experimental Driven Research in Cloud Computing

- Observation and control is needed to understand the behavior of algorithms or applications
  - Delivery of services by cloud providers kept secret with at most SLAs
  - No information about the load induced by other users and the way resources are provisioned
  - No easy way to experiment on other infrastructures (decentralized approaches, other kind of processors)
- **Challenges**
  - Invest in platforms to support experiment-driven science
  - Observability and reproducibility are key features
  - Availability of metrics about all level of the infrastructure (network, storage elements, hypervisor, …)
  - Federations of testbeds, Fog/edge clouds, end-user mobile devices, …

THANKS

Joint work with M. Dias de Assuncao (Inria), I. Chrisment (Telecom Nancy), J. Francois (Inria), A. Lebre (Inria), L. Lefèvre (Inria), D. Margery (Inria), P. Merle (Inria), L. Nussbaum (Université de Lorraine), C. Perez (Inria), G. Pierre (IRISA), D. Saucez (Inria), and P. Valduriez (Inria)