

SILECS

Super Infrastructure for Large-scale Experimental Computer Science

F. Desprez - Inria

INRIA, CNRS, RENATER, CEA, CPU, CDEFI, IMT, Sorbonne Université, Université Strasbourg, Université Lorraine, Université Grenoble Alpes, Université Lille 1, Université Rennes 1, Université Toulouse, ENS Lyon, INSA Lyon

The Discipline of Computing: An Experimental Science

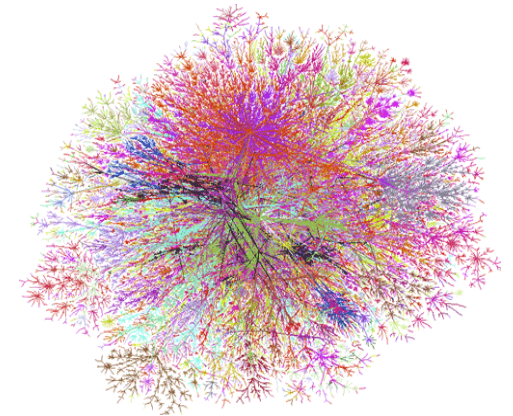
The reality of computer science

- Information
- Computers, networks, algorithms, programs, etc.

Studied objects (hardware, programs, data, protocols, algorithms, networks) are more and more complex

Modern infrastructures

- Processors have very nice features: caches, hyperthreading, multi-core, ...
- Operating system impacts the performance (process scheduling, socket implementation, etc.)
- The runtime environment plays a role (MPICH \neq OPENMPI)
- Middleware have an impact
- Various parallel architectures that can be heterogeneous, hierarchical, distributed, dynamic



Experimental Culture not Comparable with Other Sciences

Different studies

- 1994: 400 papers
 - Between 40% and 50% of CS ACM papers requiring experimental validation had none (15% in optical engineering) [Lukowicz et al.]
- 1998: 612 papers
 - “*Too many articles have no experimental validation*” [Zelkowitz and Wallace 98]
- 2007: Survey of simulators used in P2P research
 - Most papers use an unspecified or custom simulator
- 2009 update
 - *Situation is improving*

Computer science not at the same level than some other sciences

- Nobody redo experiments
- Lack of tool and methodologies

Paul Lukowicz et al. **Experimental Evaluation in Computer Science: A Quantitative Study**. In: *J.I of Systems and Software* 28:9-18, 1994
M.V. Zelkowitz and D.R. Wallace. **Experimental models for validating technology**. *Computer*, 31(5):23-31, May 1998
Marvin V. Zelkowitz. **An update to experimental models for validating computer technology**. In: *J. Syst. Softw.* 82.3:373–376, Mar. 2009
S. Naicken et al. **The state of peer-to-peer simulators and simulations**. In: *SIGCOMM Comput. Commun. Rev.* 37.2:95–98, Mar. 2007

Good Experiments

A **good experiment** should fulfill the following properties

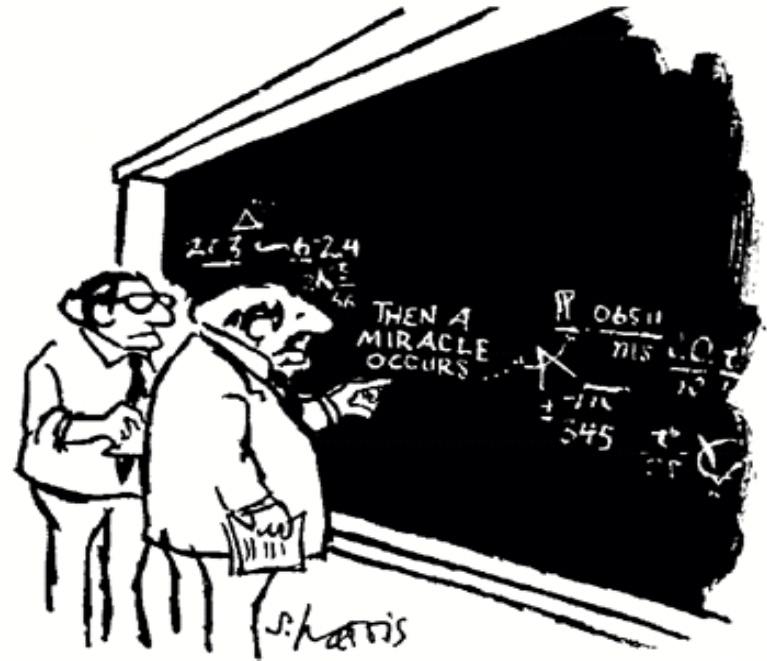
- **Reproducibility**: *must* give the same result with the same input
- **Extensibility**: *must* target possible comparisons with other works and extensions (more/other processors, larger data sets, different architectures)
- **Applicability**: *must* define realistic parameters and *must* allow for an easy calibration
- **“Revisability”**: when an implementation does not perform as expected, *must* help to identify the reasons



Analytic Modeling

Purely analytical (mathematical) models

- Demonstration of properties (theorem)
- Models need to be tractable: over-simplification?
- Good to understand the basic of the problem
- Most of the time ones still perform a experiments (at least for comparison)



"I THINK YOU SHOULD BE MORE EXPLICIT
HERE IN STEP TWO."

For a practical impact (especially in distributed computing): analytic study not always possible or not sufficient

Experimental Validation

A good alternative to analytical validation

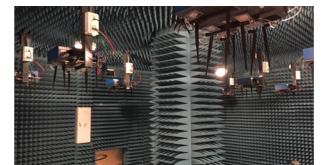
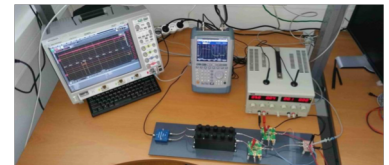
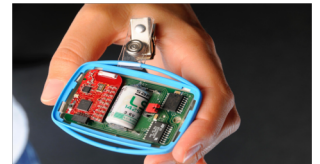
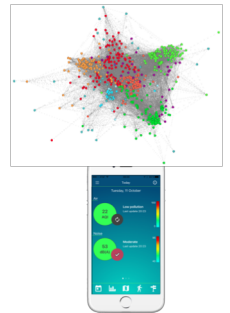
- Provides a comparison between algorithms and programs
- Provides a validation of the model or helps to define the validity domain of the model

Several methodologies

- **Simulation** (SimGrid, NS, ...)
- **Emulation** (MicroGrid, Distem, ...)
- **Benchmarking** (NAS, SPEC, LINPACK,)
- **Real-scale** (Grid'5000, FIT, FED4Fire, Chameleon, OpenCirrus, PlanetLab, ...)

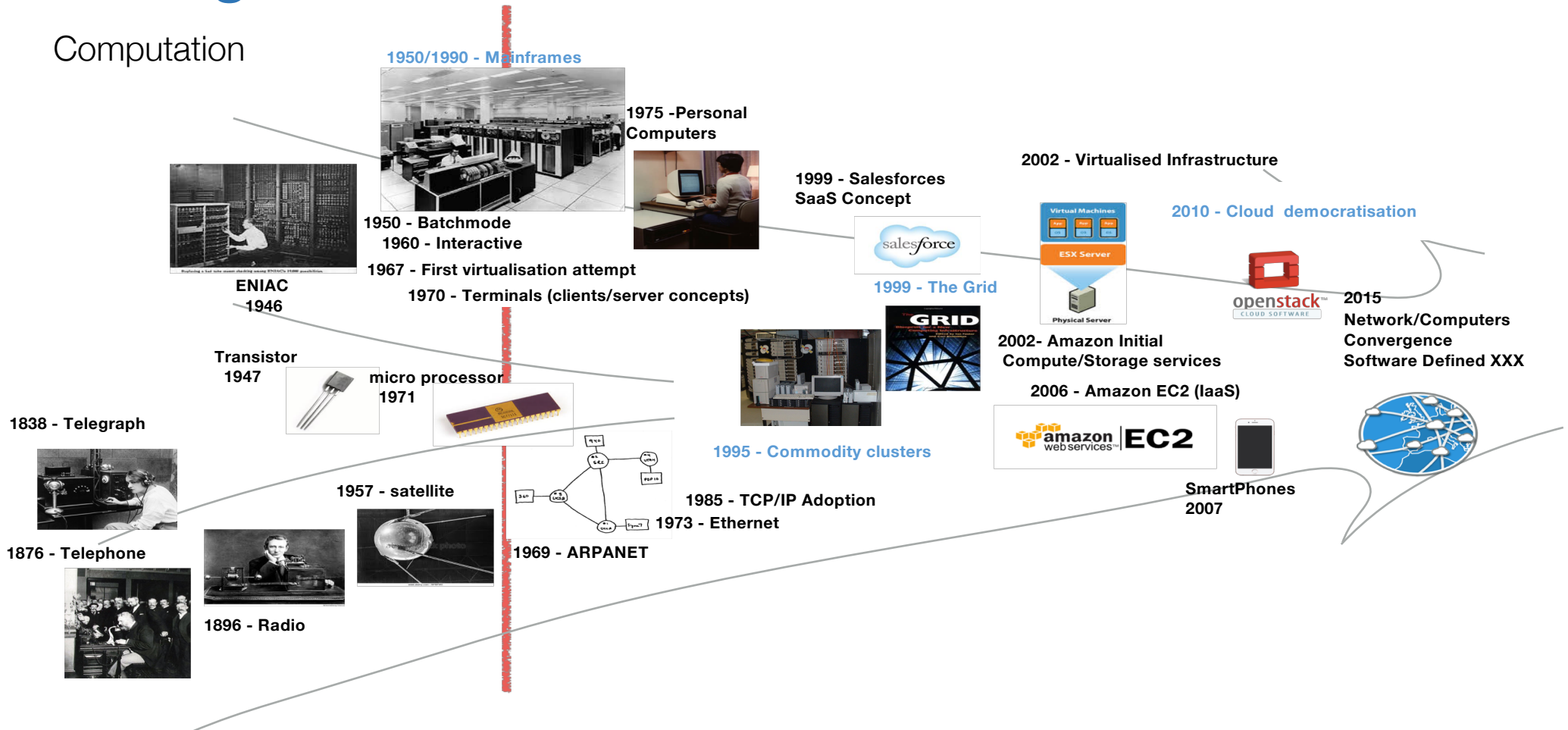
SILECS Motivation

- Exponential improvement of
 - Electronics (energy consumption, size, cost)
- Capacity of networks (WAN, wireless, new technologies)
- Exponential growth of applications near users
 - Smartphones, tablets, connected devices, sensors, ...
 - Prediction of 50 billions of connected devices by 2020 (CISCO)
- Large number of Cloud facilities to cope with generated data
 - Many platforms and infrastructures available around the world
 - Several offers for IaaS, PaaS, and SaaS platforms
 - Public, private, community, and hybrid clouds
 - Going toward distributed Clouds (FOG, Edge, extreme Edge)



Convergence

Computation



Communication

RESCOM - 17/01/2019

F. Desprez - SILECS - Frederic.Desprez@inria.fr

SILECS: based upon two existing infrastructures



- **FIT**

- Providing Internet players access to a variety of fixed and mobile technologies and services, thus accelerating the design of advanced technologies for the Future Internet
- 4 key technologies and a single control point: IoT-Lab (connected objects & sensors, mobility), CorteXlab (Cognitive Radio), wireless (anechoic chamber), Cloud technology including OpenStack, Network Operations Center
- 9 sites (Paris (2), Evry, Rocquencourt, Lille, Strasbourg, Lyon, Grenoble, Sophia Antipolis)

- **Grid'5000**

- A scientific instrument for experimental research on large future infrastructures: Clouds, datacenters, HPC Exascale, Big Data infrastructures, networks, etc.
- 10 sites, > 8000 cores, with a large variety of network connectivity and storage access, dedicated interconnection network granted and managed by RENATER

- **Software stacks dedicated to experimentation**

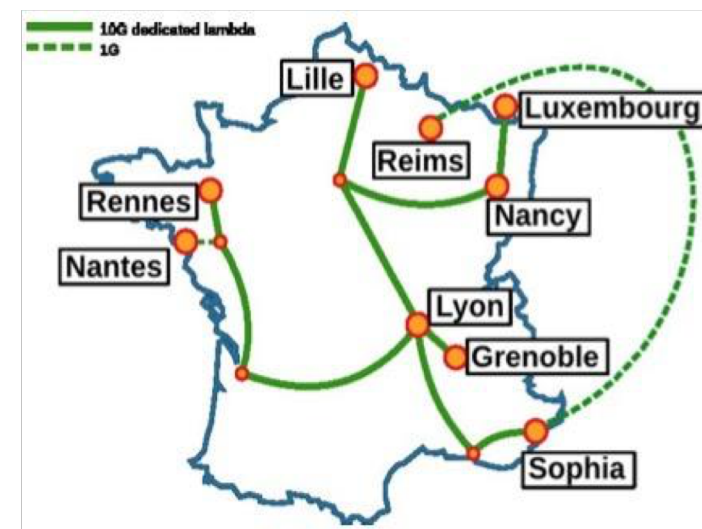
- Resource reservation, disk image deployment, monitoring tools, data collection and storage



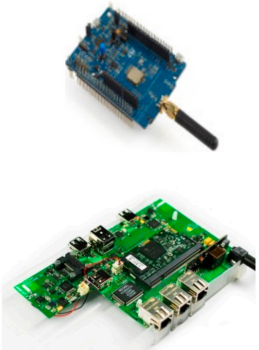
GRID'5000



- **Testbed for research on distributed systems**
 - Born from the observation that we need a better and larger testbed
 - HPC, Grids, P2P, and now Cloud computing and BigData systems
 - A complete access to the nodes' hardware in an exclusive mode (from one node to the whole infrastructure)
 - Dedicated network (RENATER)
 - Reconfigurable: nodes with Kadeploy and network with KaVLAN
- **Current status**
 - 10 sites, 29 clusters, 1060 nodes, 10474 cores
 - Diverse technologies/resources (Intel, AMD, Myrinet, Infiniband, two GPU clusters, energy probes)
- **Some Experiments examples**
 - In Situ analytics
 - Big Data Management
 - HPC Programming approaches
 - Network modeling and simulation
 - Energy consumption evaluation
 - Batch scheduler optimization
 - Large virtual machines deployments



FIT

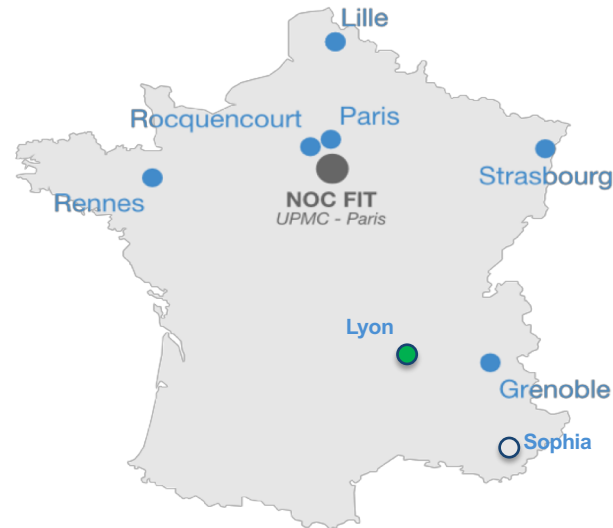


- **FIT-R2Lab:** WiFi mesh testbed (DIANA)



- **FIT-CortexLab:** Cognitive Radio Testbed
40 Software Defined Radio Nodes (SOCRATE)

FIT
FUTURE INTERNET
TESTING FACILITY



FIT-IoT-LAB

- - 2700 wireless sensor nodes spread across six different sites in France
 - Nodes are either fixed or mobile and can be allocated in various topologies throughout all sites

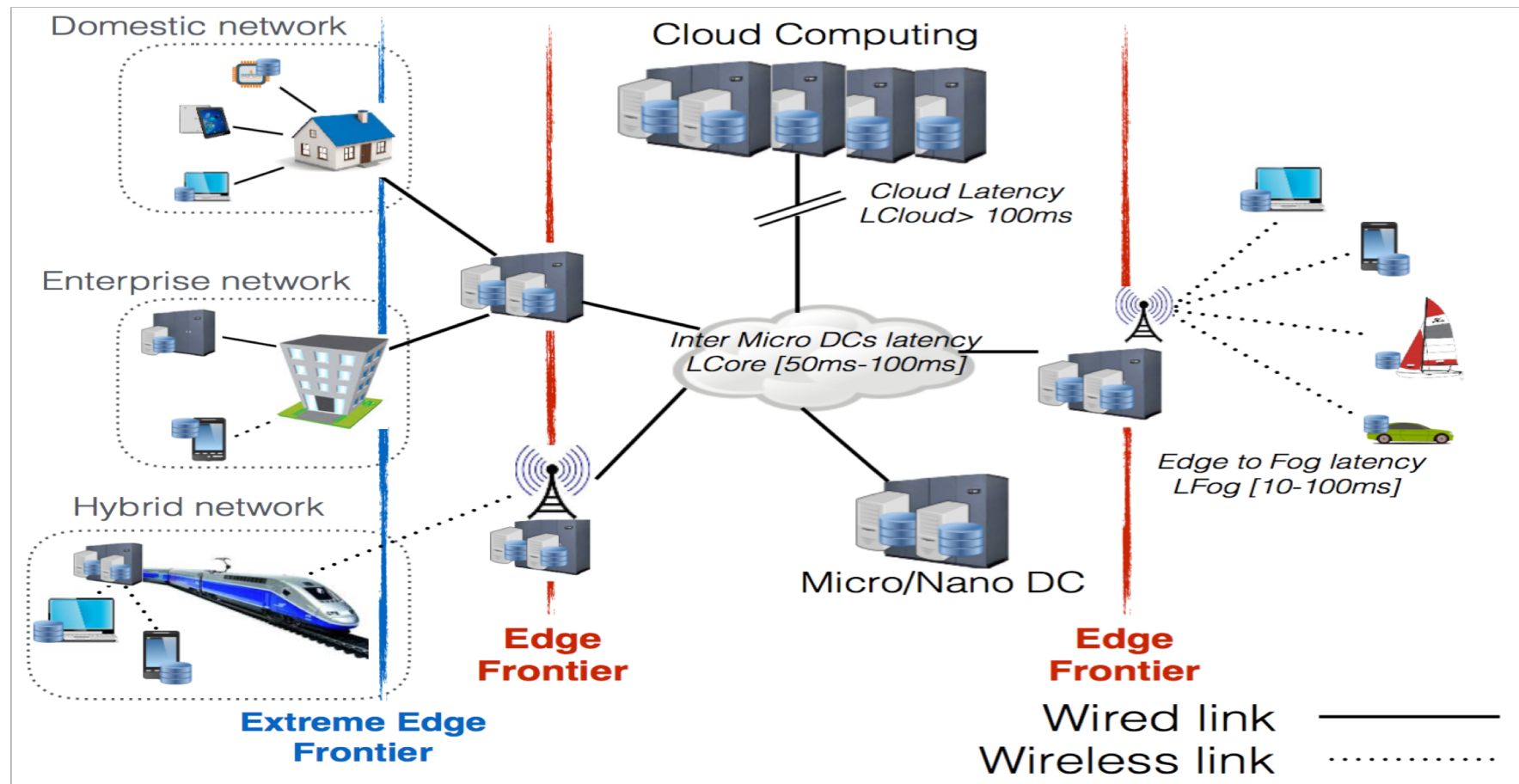
RESCOM - 17/01/2019

F. Desprez - SILECS - Frederic.Desprez@inria.fr

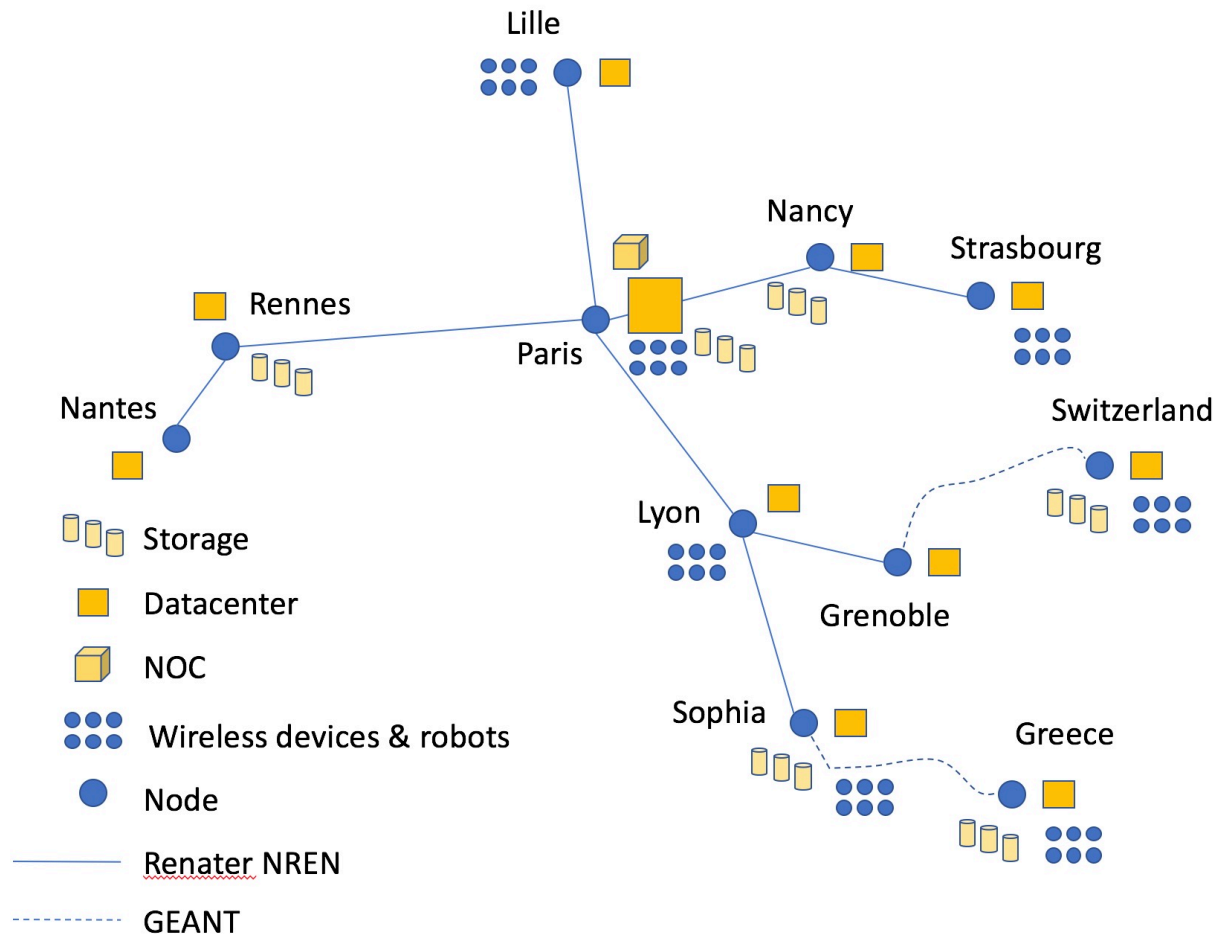
<https://www.iot-lab.info/hardware/>

<https://fit-equipex.fr/>

Envisioned Architecture



Short Term View of the Architecture



Data Center Portfolio

Targets

- Performance, resilience, energy-efficiency, security in the context of data-center design, Big Data processing, Exascale computing, etc.

Hardware

- Servers: x86, ARM64, POWER, accelerators (GPU, FPGA)
- Networking: Ethernet (10G, 40G), HPC networks (InfiniBand, Omni-Path)
- Storage: HDD, SSD, NVMe, both in storage arrays and clusters of servers

Experimental support

- Bare-metal reconfiguration
- Large clusters
- Integrated monitoring (performance, energy, temperature, network traffic)

Wireless Portfolio

Targets

- Performance, security, safety and privacy-preservation in complex sensing environment,
- Performance understanding and enhancement in wireless networking,
- Target applications: smart cities/manufacturing, building automation, standard and interoperability, security, energy harvesting, health care.

Hardware

- Software Defined Radio (SDR), LTE-Advanced and 5G
- Wireless Sensor Network (WSN/IEEE 802.15.4), LoRa/LoRaWAN
- Wifi/WIMAX (IEEE 802.11/16)

Experimental support

- Bare-metal reconfiguration
- Large-scale deployment (both in terms of densities and network diameter)
- Different topologies with indoor/outdoor locations
- Mobility-enabled with customized trajectories
- Anechoic chamber
- Integrated monitoring (power consumption, radio signal, network traffic)

Outdoor IOT testbed

- IoT is not limited to smart objects or indoor wireless sensors (smart building, industry 4.0,)
- Smart cities need outdoor IoT solutions
 - outdoor smart metering
 - outdoor metering at the scale of a neighborhood (air, noise smart sensing,)
 - citizens and local authorities are more and more interested by outdoor metering
- Controlled outdoor testbed
 - (reproducible) polymorphic IoT: support of multiple IoT technologies (long, middle and short range IoT wireless solutions) at the same time on a large scale testbed
 - Agreement and support of local authorities
 - Deployment in Strasbourg city (500000 citizens, 384 km²) and Lyon campus

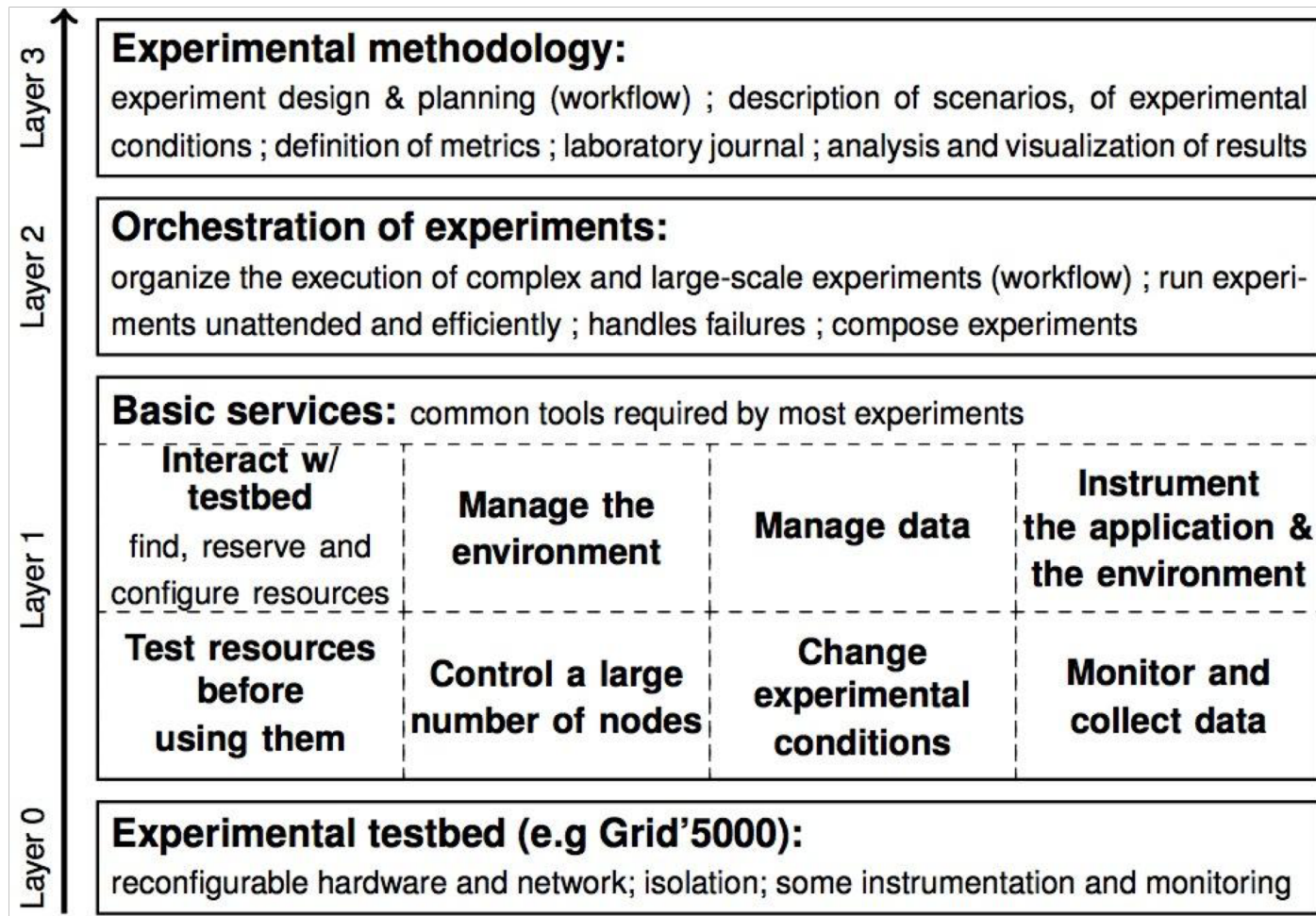
Plans for SILECS: Testbed Services

- **Provide a unified framework that (really) meets all needs**
 - Make it easier for experimenters to move for one testbed to another
 - Make it easy to create simultaneous reservations on several testbeds (for cross-testbeds experiments)
 - Make it easy to extend SILECS with additional kinds of resources
- **Factor testbed services**
 - Services that can exist at a higher level, e.g. open data service, for storage and preservation of experiments data
 - in collaboration with Open Data repositories such as OpenAIRE/Zenodo
 - Services that are required to operate such infrastructures, but add no scientific value
 - Users management, usage tracking

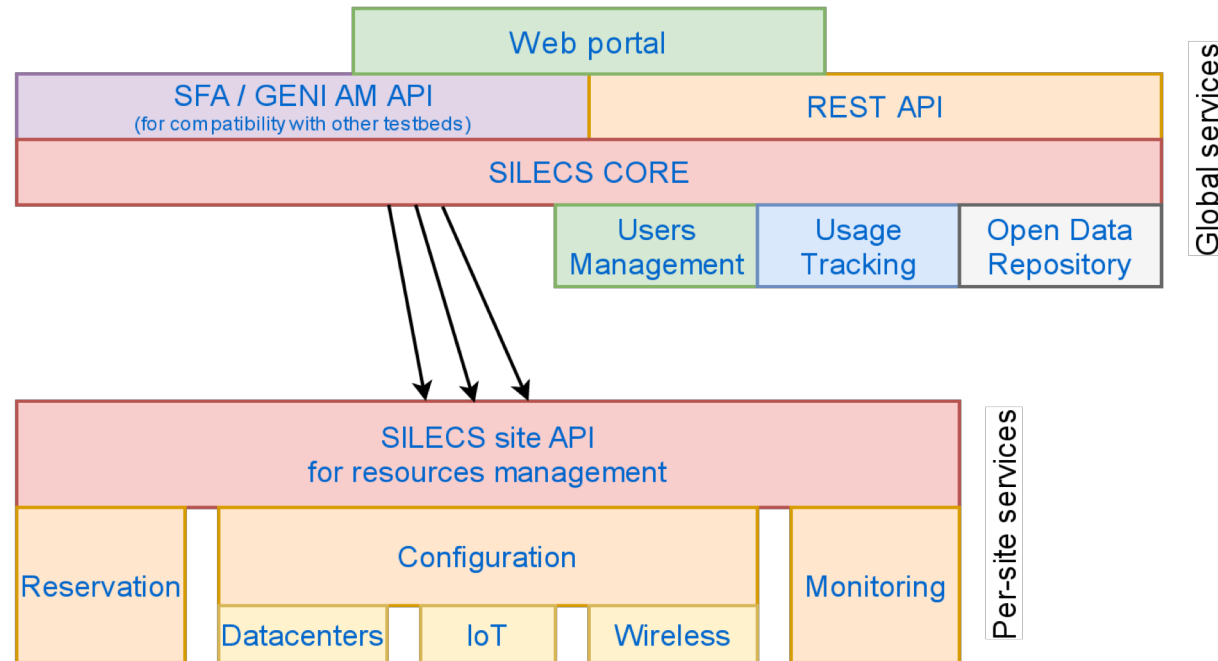
An experiment outline

- Discovering resources from their description
- Reconfiguring the testbed to meet experimental needs
- Monitoring experiments, extracting and analyzing data
- Controlling experiments: API

The GRAIL



Services & Software Stack



Built from already functional solutions



Some recent experiments examples



- **Proxy location selection in industrial IoT**, T.P. Raptis, A. Passarella, M. Conti
- **QoS differentiation in data collection for smart Grids**, J. Nassar, M. Berthomé, J. Dubrulle, N. Gouvy, N. Mitton, B. Quoitin
- **Damaris: Scalable I/O and In-situ Big Data Processing**, G. Antoniu, H. Salimi, M. Dorier
- **KerA: Scalable Data Ingestion for Stream Processing**, O.-C. Marcu, A. Costan, G. Antoniu, M. Pérez-Hernández, B. Nicolae, R. Tudoran, S. Bortoli
- **Frequency Selection Approach for Energy Aware Cloud Database**, C. Guo, J.-M. Pierson
- **Distributed Storage for a Fog/Edge infrastructure based on a P2P and a Scale-Out NAS**, B. Confais, B. Parrein, A. Lebre
- **FogIoT Orchestrator: an Orchestration System for IoT Applications in Fog Environment**, B. Donassolo, I. Fajjari, A. Legrand, P. Mertikopoulos

Proxy location selection in industrial IoT

- Distributed data collection with low latency in Industrial context
- Traditional approach
 - Improving data routing by selecting quicker links
 - Deploying enhanced edge-nodes for fog computing
- Solution
 - Dynamically select sensor nodes to act as proxys and get the information closer to consuming nodes.
- FIT IoT LAB as a validation testbed
 - Access to 95 sensor nodes with IoT features remotely
 - Customizable environment and tools (sniffer, consumption measure, etc)
 - Repeat the experiments later and compare to alternate approaches with the same environment
- The results show that latency is much reduced
- FIT IoT LAB helped validate the approach before real costly deployment

Performance Analysis of Latency-Aware Data Management in Industrial IoT Networks, T.P. Raptis, A. Passarella, M. Conti - MDPI Sensors, 2018, 18(8), 2611

QoS differentiation in data collection for smart Grids

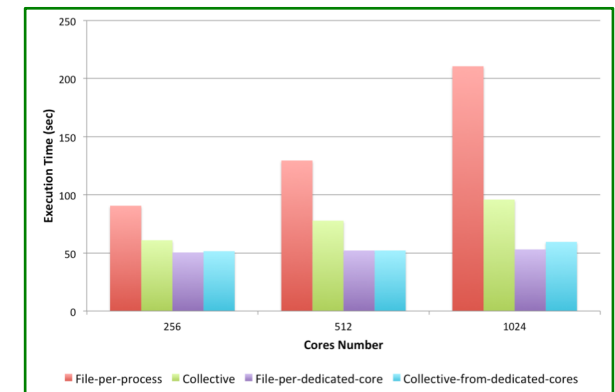
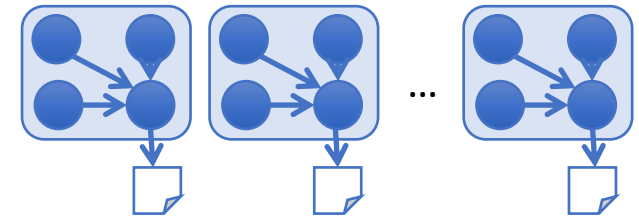
- Data collection with different QoS requirements for Smart Grid applications
- Traditional approach
 - Use of standard RPL protocol which offers overall good performance but no QoS differentiation based on application
- Solution
 - Use a dynamic objective function
- FIT IoT LAB as a validation testbed
 - Access to 67 sensor nodes with IoT features remotely
 - Customizable environment and tools (data size and rate, consumption measure, clock, etc)
 - Repeat the experiments and compare to alternate approaches with the same environment
- The results show that based on the service requested, data from different applications follow different paths, each meeting expected requirements.
- FIT IoT LAB helped validate the approach to go further with standardization

Multiple Instances QoS Routing In RPL: Application To Smart Grids – J. Nassar, M. Berthomé, J. Dubrulle, N. Gouvvy, N. Mitton, B. Quoitin – MDPI Sensors, July 2018

Damaris

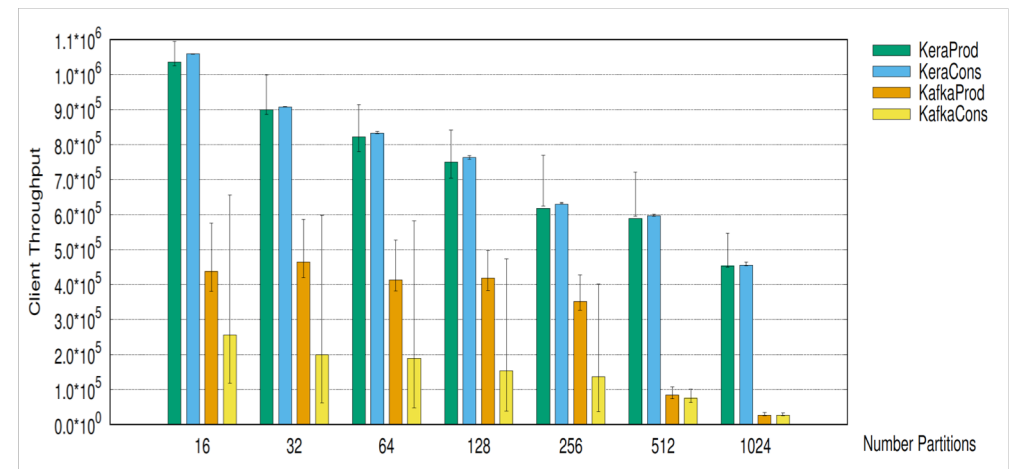


- Scalable, asynchronous data storage for large-scale simulations using the HDF5 format (HDF5 blog at <https://goo.gl/7A4cZh>)
- Traditional approach
 - All simulation processes (10K+) write on disk at the same time synchronously
 - Problems: 1) I/O jitter, 2) long I/O phase, 3) Blocked simulation during data writing
- Solution
 - Aggregate data in dedicated cores using shared memory and write asynchronously
- Grid'5000 used as a testbed
 - Access to many (1024) homogeneous cores
 - Customizable environment and tools
 - Repeat the experiments later with the same environment saved as an image
- The results show that Damaris can provide a jitter-free and wait-free data storage mechanism
- G5K helped prepare Damaris for deployment on top supercomputers (Titan, Pangea (Total), Jaguar, Kraken, etc.)



KerA: Scalable Data Ingestion for Stream Processing

- Goal: increase ingestion and processing throughput of Big Data streams
 - Dynamic partitioning and lightweight stream offset indexing
 - Higher parallelism for producers and consumers
- Grid'5000 Paravance cluster used for development and testing
 - Customized OS image and easy deployment
 - 128GB RAM and 16 CPU cores
 - 10Gb networking
- Next steps: KerA* unified architecture for stream ingestion and storage
 - Support for records, streams and objects
- Collaborations
 - INRIA, HUAWEI, UPM, BigStorage



KerA vs Kafka: up to 4x-5x better throughput

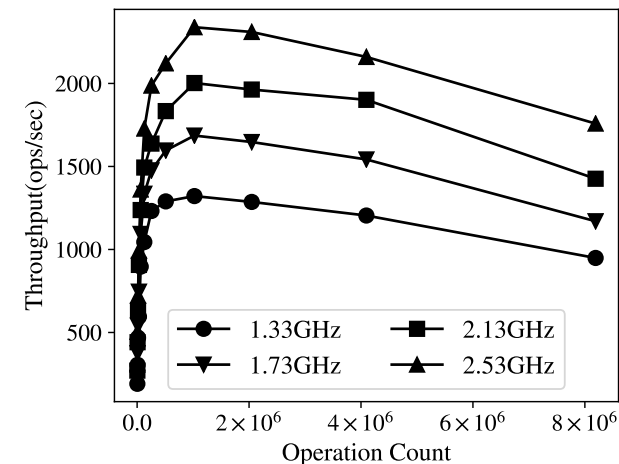
KerA: Scalable Data Ingestion for Stream Processing, O.-C. Marcu, A. Costan, G. Antoniu, M. Pérez-Hernández, B. Nicolae, R. Tudoran, S. Bortoli. In Proc. ICDCS, 2018.

Frequency Selection Approach for Energy Aware Cloud Database

- **Objective:** Study the energy efficiency of cloud database systems and propose a frequency selection approach and corresponding algorithms to cope with resource proposing problem

- **Contribution:** Propose frequency selection model and algorithms.

- Propose a Genetic Based Algorithm and a Monte Carlo Tree Based Algorithm to produce the frequencies according to workload predictions
- Propose a model simplification method to improve the performance of the algorithms



Relationship between Request Amount and Throughput

- **Grid5000 usage**

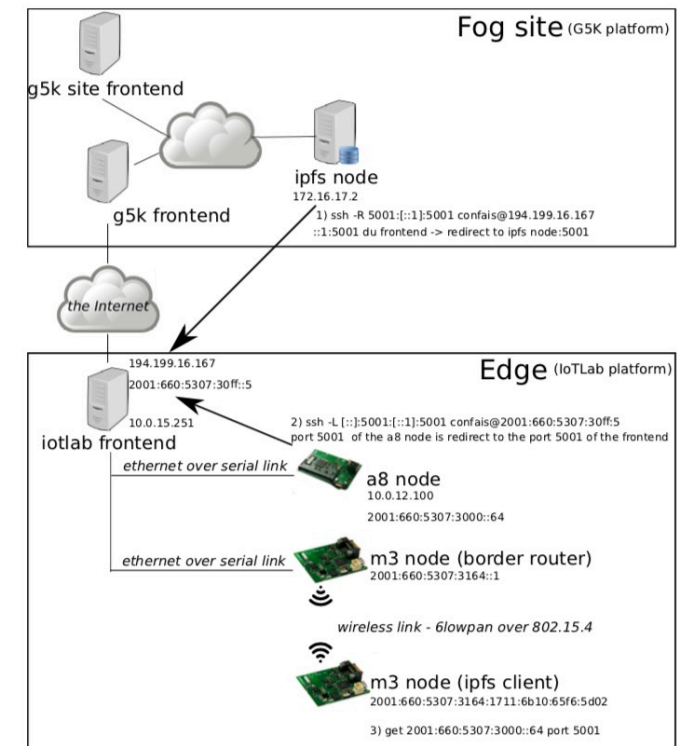
- A cloud database system, Cassandra, was deployed within a Grid'5000 cluster using 10 nodes of Nancy side to study the relationship between system throughput and energy efficiency of the system
- By another benchmark experiment, the migration cost parameters of the model were obtained

Frequency Selection Approach for Energy Aware Cloud Database, C. Guo, J.-M. Pierson. In Proc. SBAC-PAD, 2018.

Distributed Storage for a Fog/Edge infrastructure based on a P2P and a Scale-Out NAS



- Objective
 - Design of a storage infrastructure taking locality into account
 - Properties a distributed storage system should have: data locality, network containment, mobility support, disconnected mode, scalability
- Contributions
 - Improving locality when accessing an object stored locally coupling IPFS and a Scale-Out NAS
 - Improving locality when accessing an object stored on a remote site using a tree inspired by the DNS
- Experiments
 - Deployment of a Fog Site on the Grid'5000 testbed and the clients on the IoTLab platform
 - Coupling a Scale-Out NAS to IPFS limits the inter-sites network traffic and improves locality of local accesses
 - Replacing the DHT by a tree mapped on the physical topology improves locality to find the location of objects
 - Experiments using IoTlab and Grid'5000 are (currently) not easy to perform



An Object Store Service for a Fog/Edge Computing Infrastructure based on IPFS and Scale-out NAS, B. Confais, A. Lebre, and B. Parrein (May 2017). In: 1st IEEE International Conference on Fog and Edge Computing - IC FEC'2017.

FogIoT Orchestrator: an Orchestration System for IoT Applications in Fog Environment

- Objective

- Design a Optimized Fog Service Provisioning strategy (O-FSP) and validate it on a real infrastructure

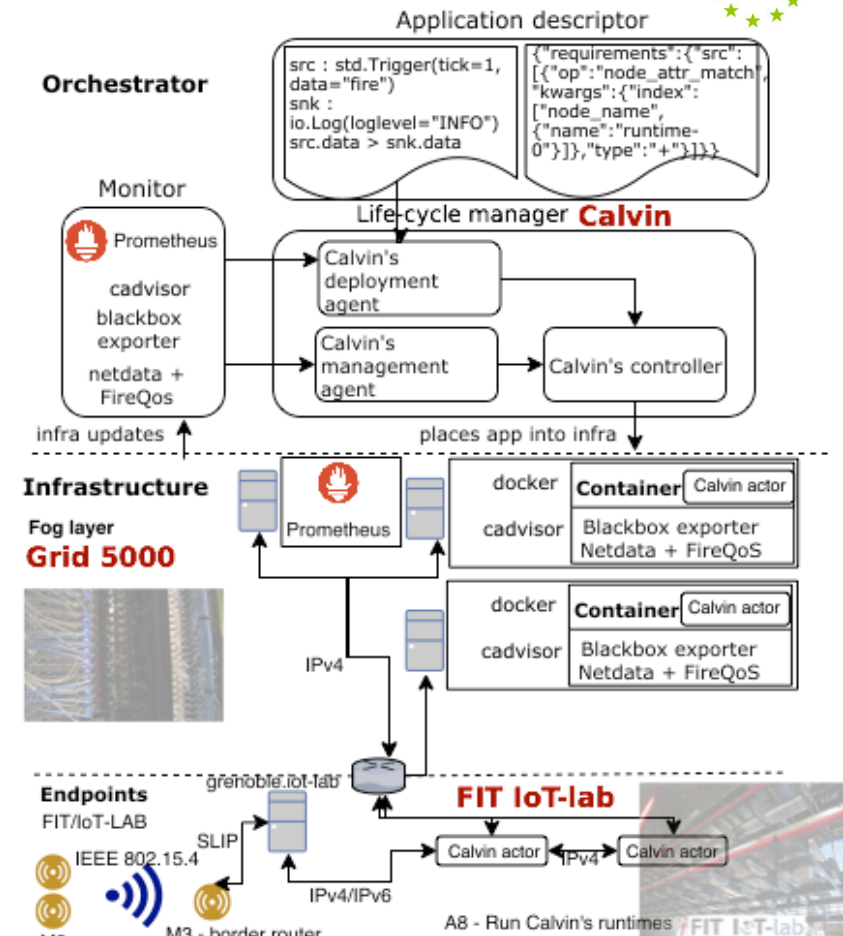
- Contributions

- Design and implementation of FITOR, an orchestration framework for the automation of the deployment, the scalability management, and migration of micro-service based IoT applications
- Design of a provisioning solution for IoT applications that optimizes the placement and the composition of IoT components, while dealing with the heterogeneity of the underlying Fog infrastructure




































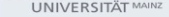
















- Experiments

- Fog layer is composed of 20 servers from Grid5000 which are part of the genepi cluster, Mist layer is composed of 50 A8 nodes
- Use of a software stack made of open-source components (Calvin, Prometheus, Cadvisor, Blackbox exporter, Netdata)
- Experiments show that the O-FSP strategy makes the provisioning more effective and outperforms classical strategies in terms of: i) acceptance rate, ii) provisioning cost, and iii) resource usage

FogIoT Orchestrator: an Orchestration System for IoT Applications in Fog Environment, B. Donassolo, I. Fajjari, A. Legrand, P. Mertikopoulos.. *1st Grid'5000-FIT school*, Apr 2018, Sophia Antipolis, France. 2018.



European dimension

Countries	FR	GR	CH	ES	CY	IT	DE	NL	LU	BE
										
Gov.										
										
Research	    	 	   	       	 	  	  	 	 	
Industry										
NRENs										

Conclusions

- New infrastructure based on two existing instruments (FIT and Grid'5000)
- Design a software stack that will allow experiments mixing both kinds of resources while keeping reproducibility level high
- **Keep the aim of previous platforms** (their core scientific issues addressed)
 - Scalability issues, energy management, ...
 - IoT, wireless networks, future Internet for SILECS/FIT
 - HPC, big data, clouds, virtualization, deep learning ... for SILECS/Grid'5000
- **Address new challenges**
 - IoT and Clouds
 - New generation Cloud platforms and software stacks (Edge, FOG)
 - Data streaming applications
 - Locality aware resource management
 - Big data management and analysis from sensors to the (distributed) cloud
 - Mobility
 - ...



**We need
YOU!**

Thanks, any questions ?

<http://www.silecs.net/>

<https://www.grid5000.fr/>

<https://fit-equipex.fr/>

