RESEARCH CENTRE

Inria Centre at Rennes University

IN PARTNERSHIP WITH:

Ecole Nationale Supérieure Mines-Télécom Atlantique Bretagne Pays de la Loire, Nantes Université, Orange SA

2024
ACTIVITY REPORT

Project-Team

STACK

Software Stack for Massively Geo-Distributed Infrastructures

IN COLLABORATION WITH: Laboratoire des Sciences du numérique de Nantes

DOMAIN

Networks, Systems and Services, Distributed Computing

THEME

Distributed Systems and middleware

Contents

Project-Team STACK

Creation of the Project-Team: 2019 January 01

Keywords

Computer sciences and digital sciences

- A1.1.8. – Security of architectures
- A1.1.10. – Reconfigurable architectures
- A1.1.13. – Virtualization
- A1.2.1. – Dynamic reconfiguration
- A1.2.2. – Supervision
- A1.2.4. – QoS, performance evaluation
- A1.2.5. – Internet of things
- A1.2.8. – Network security
- A1.3.4. – Peer to peer
- A1.3.5. – Cloud
- A1.3.6. – Fog, Edge
- A1.5.1. – Systems of systems
- A1.6. – Green Computing
- A2.1.7. – Distributed programming
- A2.1.10. – Domain-specific languages
- A2.5.2. – Component-based Design
- A2.6. – Infrastructure software
- A2.6.1. – Operating systems
- A2.6.2. – Middleware
- A2.6.3. – Virtual machines
- A2.6.4. – Ressource management
- A3.1.2. – Data management, quering and storage
- A3.1.3. – Distributed data
- A3.1.8. – Big data (production, storage, transfer)
- A4.1. – Threat analysis
- A4.4. – Security of equipment and software
- A4.9. – Security supervision

Other research topics and application domains

- B2. – Health
- B4. – Energy
- B4.5.1. – Green computing
- B5.1. – Factory of the future
- B6.3. – Network functions
- B6.4. – Internet of things
- B6.5. – Information systems
- B7. – Transport and logistics
- B8. – Smart Cities and Territories

1 Team members, visitors, external collaborators

Research Scientists

- Adrien Lebre [Team leader, INRIA, Professor Detachement, HDR]
- Daniel Balouek [INRIA, ISFP]
- Sébastien Bolle [Orange SA, Industrial member]
- Abdelhadi Chari [Orange SA, Industrial member]
- Thierry Coupaye [Orange SA, Industrial member, HDR]
- Thomas Hassan [Orange SA, Industrial member]
- Philippe Raipin Parvedy [Orange SA, Industrial member]

Faculty Members

- Hélène Coullon [IMT ATLANTIQUE, Associate Professor]
- Carlos Javier Gonzalez [IMT ATLANTIQUE, Associate Professor, from Sep 2024]
- Remous Koutsiamanis [IMT ATLANTIQUE, Associate Professor]
- Thomas Ledoux [IMT ATLANTIQUE, Professor, HDR]
- Jean-Marc Menaud [IMT ATLANTIQUE, Professor, HDR]
- Jacques Noyé [IMT ATLANTIQUE, Associate Professor]
- Kandaraj Piamrat Lerebours [Nantes University, Associate Professor]
- Guillaume Rosinosky [IMT ATLANTIQUE, Associate Professor, from Apr 2024]
- Mario Südholt [IMT ATLANTIQUE, Professor, HDR]

Post-Doctoral Fellows

- Yasmina Bouizem [INRIA, Post-Doctoral Fellow, until Aug 2024]
- Clément Mommessin [IMT ATLANTIQUE, until Jun 2024]
- Eloi Perdereau [IMT ATLANTIQUE, Post-Doctoral Fellow, from Oct 2024]
- Eloi Perdereau [IMT ATLANTIQUE, Post-Doctoral Fellow, until Aug 2024]
- Jolan Philippe [IMT ATLANTIQUE, Post-Doctoral Fellow, until Jul 2024]

PhD Students

- Tengfei An [INRIA]
- Geo Johns Antony [IMT ATLANTIQUE, from Apr 2024 until Jun 2024]
- Geo Johns Antony [INRIA, until Mar 2024]
- Hiba Awad [SMILE, CIFRE]
- Samia Boutalbi [Ericsson, CIFRE]
- Divi De Lacour [ORANGE, CIFRE]

- Houssem Jmal [Nantes University, from Apr 2024]
- Mohammed Abdrrahim Lahmar [IMT ATLANTIQUE, from Oct 2024]
- Wedwang Romial Menra [INRIA, from Nov 2024]
- Martin Molli [IMT ATLANTIQUE, from Nov 2024]
- Duc Thinh Ngo [ORANGE, CIFRE]
- Antoine Omond [IMT ATLANTIQUE, until Nov 2024, On leave at Tromso University (Norway)]
- Abdou Seck [IMT ATLANTIQUE]
- Cherif Si Mohammed [ADEME]
- Lylian Siffre [Kapela, CIFRE, from Nov 2024]

Technical Staff

- Farid Arfi [IMT ATLANTIQUE, Engineer, until Jul 2024]
- Marie Delavergne [IMT ATLANTIQUE, Engineer, until Aug 2024]
- Tayeb Diab [IMT ATLANTIQUE, Engineer, until May 2024]
- Baptiste Jonglez [INRIA, Engineer]
- Sidi Mohammed Kaddour [INRIA, Engineer, from Dec 2024]
- Sidi Mohammed Kaddour [IMT ATLANTIQUE, Engineer, until Nov 2024]
- Jabran Khan [IMT ATLANTIQUE, Engineer, until Aug 2024]
- Anas Mokhtari [IMT ATLANTIQUE, Engineer]
- Clément Mommessin [IMT ATLANTIQUE, Engineer, until Jun 2024]
- Matthieu Rakotojaona Rainimangavelo [IMT ATLANTIQUE, Engineer, until Oct 2024]

Interns and Apprentices

- Severin Bradley Anzie [IMT Atlantique, Intern, from May 2024 until Aug 2024]
- Alexandre Boissel [Nantes University, Intern, from Jun 2024 until Aug 2024]
- Abdoulaye Diarisso [IMT ATLANTIQUE, Intern, from May 2024 until Sep 2024]
- Celeste Precil Guimapi Guefack [LS2N, Intern, from May 2024 until Aug 2024]
- Wedwang Romial Menra [IMT ATLANTIQUE, Intern, from Apr 2024 until Aug 2024]

Administrative Assistant

- Anne-Claire Binetruy [INRIA]

Visiting Scientist

- Lorenzo Carnevale [University of Messina (Italia), from Jun 2024 until Sep 2024]

2 Overall objectives

2.1 STACK in a Nutshell

The STACK team addresses challenges related to the management and advanced usages of the Cloud to IoT continuum (infrastructures on the Cloud, Fog, Edge, and IoT). More specifically, the team is interested in delivering appropriate system abstractions to operate and use massively geo-distributed infrastructures, from the lowest to the highest levels of abstraction (i.e. system to application development, respectively), and addressing crosscutting dimensions such as energy or security. These infrastructures are critical for the emergence of new kinds of applications related to the digitalization of the industry and the public sector (a.k.a. the Industrial Internet, smart cities, e-medecine, etc.).

2.2 Context & Objectives

Initially proposed to interconnect computers worldwide, the Internet has significantly evolved to become in two decades a key element in almost all our activities. This (r)evolution mainly relies on the progress that has been achieved in the computation and communication fields which in turn has led to the well-known and widely spread Cloud Computing paradigm. Nowadays most Internet exchanges occur between endpoints ranging from small-scale devices, such as smart-phones, to large-scale facilities, *i.e.*, cloud computing platforms, in charge of hosting modern information systems.

With the emergence of the Internet of Things (IoT), stakeholders expect a new revolution that will push, once again, the limits of the Internet, in particular by favouring the convergence between physical and virtual worlds into an *augmented world* or *cyber-physical world*. This convergence is about to be made possible thanks to the development of minimalist sensors as well as complex industrial physical machines that can be connected to the Internet through edge computing infrastructures. Edge computing is an extension of the cloud computing model that consists in deploying a federation (or cooperation) of smaller data centers at the edge of the network, thus closer to sensors, devices, machines, and end-users that produce and consume data [86, 84]. This new kind of digital infrastructure, which covers resources from the "center" to the extreme edge of the network, is expected to improve almost all aspects of daily life and the decision processes in various domains such as industry, transportation, health, training and education. The corresponding applications target the control and optimization of the business processes of most companies thanks to the intensive use of ICT systems and real-time data collected by geo-distributed physical devices (video, sensors, ...).

Among the obstacles to this new generation of Internet services is the development of a convenient and powerful software stack, *i.e.*, a set of system mechanisms and software abstractions capable of operating and exposing a significant number of diverse computational resources in a unified, efficient and sustainable way.

In other words, this framework should allow operators, and devops, to manage the life-cycle of both the digital infrastructure and the applications deployed on top of this infrastructure, throughout the **cloud to IoT continuum**. These include operations such as the initial configuration but also all the reconfigurations that can be required in response to particular events (maintenance operations, equipment failures, application load variation, user mobility, energy shortage, etc.).

The existing software stacks that have been proposed to manage Cloud Computing platforms are not appropriate for handling the specifics of the next generation of digital infrastructure (in terms of scale, heterogeneity, dynamicity, security threats, and energy opportunities). For example, this infrastructure have to be operated remotely and automatically as much as possible as it will impossible to have human presence on all locations. Due to their number, it will be necessary to allow operations not on a single site but on sets defined on the fly as needed. Moreover, the management mechanisms must have been designed to cope with intermittent network access to the sites. That is to say, offering on the one hand safety properties and on the other hand autonomy in order to allow each site to remain as operational as possible in the event of network partitioning. Finally, currently existing interfaces (APIs) should be extended to turn location into a first-class citizen. In particular, the locality aspects should be reified from the core system building blocks to the high-level application programming interfaces.

The STACK activities cover the full Cloud to IoT continuum, including recent challenges related to the network dimension and urgent computing. An elargement of STACK core activities has started with the arrivals of Ass. Prof Koutsiamanis, Ass. Prof Piamrat and Dr. Balouek who respectively joined the team in 2021, 2022 and 2023. In 2024, the expertise of STACK has been further strengthened with the arrival of Orange members, and Ass. Prof Rosinosky and Ass. Prof Gonzalez.

2.3 Scientific Foundations

Through the ongoing integration of Orange members, STACK consolidates its expertise in distributed systems, networks, cyber-physical systems, IoT, device management, and software programming as well as combining significant skills in the design, practical development and evaluation of large-scale systems. More precisely, our research activities mainly rely on a set of scientific foundations detailed below.

- **(Distributed) Systems.** The first scientific foundation of the team is related to our strong expertise in resource management and capacity planning of large-scale infrastructure [66, 79, 70, 88, 82, 83]. This includes the design and evaluation of system mechanisms and algorithms to operate and use computation, network, storage, and IoT resources in an efficient and sustainable manner. Our knowledge is based on traditional as well as distributed system fundamentals, covering virtualization technologies, storage, security, energy, and distributed/parallel algorithms.

- **Networks.** Another set of expertise in the team concerns network related topics. This includes intelligent analysis and management in wireless and mobile networks using artificial intelligence and machine learning techniques with particular focus on energy efficiency [2, 12], graph neural networks [36], active learning [11] for network security and traffic prediction [1]. It also includes the optimisation of wireless low-power and lossy networks (LLN), typically wireless Industrial IoT networks, through energy-aware network resource and communications scheduling and routing [69, 67, 47, 65, 50]. Additionally, using SDN enhances network security through fine-grained security policies and efficient control plane/data plane management of complex routing decisions. This enables high-performance networks and scalable management of large-scale environments, ensuring optimized resource usage and robust communication cite Flauzac2020, Gonzalez2023.

- **Digital Twins, Network and Device Management.** Based on initial expertise in IoT platforms and Cyber-Physical-Systems, the management of connected devices and sensor data, and especially on distributed and autonomic architectures of such platforms [80, 41, 40], the team has developed a broader vision of network management and operation with a strong expertise in *digital twins* as a pivotal technology. This includes graph-based modelling of digital twins [78], semantic modelling and ontology mapping [42], graph storage distribution, federation and historisation [63, 64] — and the application of these concepts and technologies in different use cases in the domain of smart building (e.g., localisation [50], dynamic wireless IoT network resource allocation[65]), smart industry (e.g., support for reliable and low-latency wireless Industrial IoT networks [69, 67]), logistics [74] around the Thing'in digital twin platform [56] (Thing'in the future).

- **Autonomic and Self-Adaptive Systems.** Considering the high (and ever increasing complexity) of ICT systems, autonomic and self-adaptive policies have become the *de facto* standard for designing and building large-scale systems. This second family includes, for example, research approaches that have been harnessed to tackle system modularity, configuration and reconfiguration of dynamic and distributed systems, as well as retroaction and autonomic loops. All these concepts enable administrators and developers to deal with various objectives such as performance, high availability, low energy consumption, etc. STACK members have provided several relevant contributions in the last couple of years [59, 60, 73, 61, 72, 46, 48, 49].

- **Software Engineering and Programming.** Similarly, software engineering and advances in programming are highly valuable to correctly design complex systems such as the software stack we target. Leveraging the expertise of software programming of the team, STACK contributions leverage various techniques including component-based programming models [57, 45, 58, 71, 55, 6], event-driven [62, 87], data-driven and workflow models, as well as models for Utility Computing (Service Level Agreement, aka. SLA) [85], and more generally, distributed and parallel programming models.

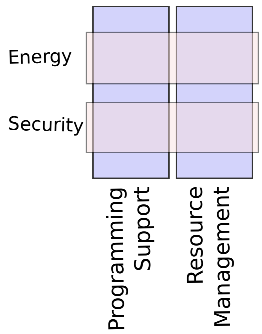
- **Experiment-Driven Research.** Finally, the last important domain of expertise of the future team consists in the evaluation of complex software stacks at large scale through simulations and in-vivo experiments. This includes knowledge on experimental methodology, measuring/monitoring/tracing tools [43] and more recently aspects related to software-defined experiments and reproducible research [52, 51]. Team members are also in charge of the animation of the LASCARE working group (LArge SCale ARchitecture Experimentation and Simulation) of the IOLab.

We aim at strengthening the knowledge in these different areas through two kinds of contributions: First through scientific articles as a regular project team, and second, through concrete pieces of software that can be transferred to major opensource communities.

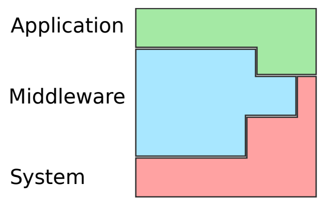
3 Research program

3.1 Overview

STACK activities have been focused on the management and programming of geo-distributed data centers with a work program defined around four research topics as depicted in Figure 1a. The first two ones are related to the resource management mechanisms and the programming support that are mandatory to operate and use ICT geo-distributed resources (compute, storage, network). They are transversal to the three software layers that generally compose a software stack (System/Middleware/Application in Figure 1b) and nurture each other (*i.e.*, the resource management mechanisms will leverage abstractions/concepts proposed by the programming support axis and reciprocally). The third and fourth research topics are related to the Energy and Security dimensions (both also crosscutting the three software layers). Although they could have been merged with the first two axes, we identified them as independent research directions due to their critical aspects with respect to the societal challenges they represent.



(a) Stack Research Topics



(b) Software stack layers

This scientific roadmap to address challenges related to the management and programming of geo-distributed infrastructures applies to the Cloud to IoT continuum and continues to have significant scientific and socio-economic impact. Hence, STACK organizes its activities around these four crosscutting lines that form a unique approach.

Additionally, our activities extend to the management of IoT devices with the ultimate goal of covering the entire Cloud-to-IoT continuum through a common software stack.

Our vision is to base this computing continuum software stack on control loops following the MAPE-K model[1], which can be seen as an infinite loop that monitors the infrastructure as well as the state of the applications in order to maintain in an autonomous manner the expected objectives (in terms of performance, robustness, etc.).

Although it is largely adopted in Cloud orchestrators such as Kubernetes, delivering a MAPE-K software stack for the computing continuum faces multiple challenges. The first one is related to the diversity of resources to consider. KubeEdge [89], for instance, proposes extensions to integrate servers

[1] MAPE-K stands for Monitoring, Analyzing, Planning, Execution and Knowledge [68].

and IoT devices under the same framework. However, the supported operations are rather limited as they only cover communication between software components running on servers at the edge and the connected devices. In other words, the IoT devices are not considered in the control loops. From our viewpoint, this weak integration is linked to an incomplete understanding of the needs that such a system must take into account (in particular on the IoT device management side). To favor the integration of both dimensions into a common system, we aim to identify major structural properties as well as management operations necessary at the operator and DevOps levels and to implement them when needed. A second important challenge is related to the geo-distribution property (and so the intermittent network connectivity) of this type of infrastructure. This increased complexity implies revising the way control loops are designed in order to handle frequent disconnections that can occur at any time (for instance due to the low energy level of IoT devices). Here, our approach is to combine autonomous loops with well-adapted formal methods to guarantee their verification or to synthesize correct-by-design decisions. Additionally, we study their performance models to be able to automatically, safely, efficiently and in a timely manner adapt Cloud-to-IoT infrastructures and their hosted applications according to different objectives (performance, energy, security, etc.). Finally, we are working towards partitioning a Cloud-to-IoT infrastructure into several areas and delivering the illusion of a single system through a federated approach: each area is managed by an independent controller, and collaborations between areas are done through dedicated middleware. The innovative aspect relies in the way of developing this middleware so that it is reliable in spite of increasing scale and faults (collaborations will be triggered only on-demand without maintaining a global knowledge base of the entire infrastructure).

Some of the research questions we address in the medium term are:

- *How to specify and model the dynamics of Cloud-to-IoT infrastructures and the dynamics of associated systems and applications in a generic way and how to leverage this dynamics for reconfiguration purposes?* In particular, this is done though studying how existing languages such as SysML, ThingML or TOSCA may be revised for this purpose. In addition, we focus on the exploration of the properties of safety, separation of concerns and efficiency when reconfiguring distributed systems with Concerto [49] extended to IoT devices and to network resources. Finally, we address the dynamicity of systems described by such Architecture Description-like languages (ADL) considering a convergence between the *Model@Runtime* and Digital Twins approaches, i.e. implementing models@runtime as digital twins.

- *How to design and deploy decentralized autonomic loops, from monitoring to the execution of reconfiguration plans?* An important challenge is related to the development of mechanisms capable of rebuilding, on-demand, a knowledge base according to the functional and non-functional properties to be satisfied. From our viewpoint, it is crucial to propose alternative approaches to avoid maintaining such a global knowledge base through time and at the scale of a Cloud-to-IoT infrastructure. Regarding the monitoring/supervision of the infrastructure, we study the latest results on complex event processing as well as machine learning techniques. The former enables the triggering of actions based on predefined events while the latter allow the management to evolve from reactive to predictive strategies. Regarding the analysis and planning phases we study how to combine and leverage satisfiability solvers (SMT and SAT solvers), constraint programming, and distributed algorithms. We are also interested in the formal verification of reconfiguration procedures.

- *How to enhance legacy distributed software in an easy and non-instrusive manner*. In particular, we focus on handling the geo-distribution of applications, optimize their processing or include security features using service meshes and code injection techniques.

- *How to increase the responsiveness of data analysis algorithms and accelerate responses of AI-enabled scenarios across the Cloud-to-IoT continuum?* Another important challenge is related to the steering of computation considering data and events measured by the IoT infrastructure, coupled with historical information and the Quality of Service (QoS) needed. We investigate application formulations that allow developers to balance requirements and costs, along with programming abstractions to define policies that can react to unforeseen events and constraints.

- *On the energy dimension, the main questions are related to the generalization of the usage of renewable energies in the Cloud-to-IoT continuum while guaranteeing availability and reliability properties.* We investigate, in particular, whether energy harvesting devices could be used at the extreme edge and how they complicate the placement challenge that we largely studied in a multi-cloud context. Besides, we are working on extending our work to include green energy awareness for users (e.g., DevOps engineers, web application end-users, etc.).
- *How can telemetry metrics collected by P4 be integrated into a distributed architecture for both local and global decision-making?* Integrating telemetry metrics collected by P4 into a distributed architecture enables real-time monitoring and adaptive decision making. P4's programmable data plane allows the extraction of fine-grained metrics such as latency, packet loss, and throughput. These metrics can be fed into a centralized or hierarchical control system, enabling local decisions to optimize immediate performance and global decisions for long-term resource allocation and policy enforcement. By leveraging SDN or ICN principles, these telemetry insights can improve traffic management, enhance fault tolerance, and support scalable and efficient operation across distributed environments.
- *Finally, on the security side, we investigate the new threats resulting from an externalized management of geo-distribution.* This includes, in particular, the identification of new possible attack channels as well as counter measures to guarantee a satisfactory level of security through the whole continuum. Furthermore, we are making efforts to extend our work on kernel security policies [44] in order to also take into account the network dimension and ensure strong isolation from Cloud/Edge servers to IoT devices.

All the aforementioned research questions are addressed through several application fields: telecommunications operators and smart buildings in the first place through this privileged partnership with the Orange colleagues who are going to join this new team but also in health, in particular, biomedical research in order to allow the execution of analyses, currently emerging, in large-scale geo-distributed environments.

4 Application domains

Industrial/Tactile Internet/Cyber-Physical applications highlight the importance of the computing continuum model. Hence, the use-cases of STACK activities are driven and nurtured by these application domains. Besides, it is noteworthy to mention that Telecom operators such as Orange have been among the first ones to advocate the deployment of Fog/Edge infrastructure. The initial reason is that a geo-distributed infrastructure enables them to virtualize a large part of their resources and thus reduce capital and operational costs. As an example, several researchers have been investigating through the IOLab, the joint lab between Orange and Inria, how 5G networks can be managed. We highlight that while our expertise does partially include the network side, the main focus is rather on how we can deploy, locate and reconfigure the software components that are mandatory to operate next generation of network/computing infrastructure. The main challenges are related to the high dynamicity of the infrastructure, the way of defining Quality of Service of applications and how it can be guaranteed. We expect our contributions will deliver advances in location based services, optimized local content distribution (data-caching) and Mobile Edge Computing [2]. In addition to bringing resources close to end-users, massively geo-distributed infrastructures should favor the development of more advanced networks as well as mobile services.

4.1 Overview

Supporting industrial actors and open-source communities in building an advanced software management stack is a key element to favor the advent of new kinds of information systems as well as web applications. Augmented reality, telemedecine and e-health services, smart-city, smart-factory, smart-transportation and remote security applications are under investigations. Although, STACK does not

[2] ETSI MEC specifications.

intend to address directly the development of such applications, understanding their requirements is critical to identify how the next generation of ICT infrastructure should evolve and what are the appropriate software abstractions for operators, developers and end-users. STACK team members have been exchanging since 2015 with a number of industrial groups (notably Orange Labs and Airbus), a few medical institutes (public and private ones) and several telecommunication operators in order to identify both opportunities and challenges in each of these domains, described hereafter.

4.2 Industrial Internet

The Industrial Internet domain gathers applications related to the convergence between the physical and the virtual world. This convergence has been made possible by the development of small, lightweight and cheap sensors as well as complex industrial physical machines that can be connected to the Internet. It is expected to improve most processes of daily life and decision processes in all societal domains, affecting all corresponding actors, be they individuals and user groups, large companies, SMEs or public institutions. The corresponding applications cover: the improvement of business processes of companies and the management of institutions (*e.g.*, accounting, marketing, cloud manufacturing, etc.); the development of large "smart" applications handling large amounts of geo-distributed data and a large set of resources (video analytics, augmented reality, etc.); the advent of future medical prevention and treatment techniques thanks to the intensive use of ICT systems, etc. We expect our contributions favor the rise of efficient, correct and sustainable massively geo-distributed infrastructure that are mandatory to design and develop such applications.

4.3 Internet of Skills

The Internet of Skills is an extension of the Industrial Internet to human activities. It can be seen as the ability to deliver physical experiences remotely (*i.e.*, via the Tactile Internet). Its main supporters advocate that it will revolutionize the way we teach, learn, and interact with pervasive resources. As most applications of the Internet of Skills are related to real time experiences, latency may be even more critical than for the Industrial Internet and raise the locality of computations and resources as a priority. In addition to identifying how an Utility Computing infrastructure can cope with this requirement, it is important to determine how the quality of service of such applications should be defined and how latency and bandwidth constraints can be guaranteed at the infrastructure level.

4.4 e-Health

The e-Health domain constitutes an important societal application domain of the two previous areas. The STACK teams is investigating distribution, security and privacy issues in the fields of systems and personalized (aka. precision) medicine. The overall goal in these fields is the development of medication and treatment methods that are tailored towards small groups or even individual patients.

We have been working on different projects since the beginning of STACK (e.g., PrivGen CominLabs). In general, we are applying and developing corresponding techniques for the medical domains of genomics, immunobiology and transplantalogy (see Section 10).

The STACK team continue to contribute to the e-Health domain by harnessing advanced architectures, applications and infrastructure for the Fog/Edge, Cloud/Edge, and Cloud/Edge/IoT continuum.

4.5 Network Virtualization and Mobile Edge Services

Telecom operators have been among the first to advocate the deployment of massively geo-distributed infrastructure, in particular through working groups such as the Mobile Edge Computing at the European Telecommunication Standards Institute. The initial reason is that a geo-distributed infrastructure enables Telecom operators to virtualize a large part of their resources and thus reduces capital and operational costs. Through the Sylva project, we aim to expand our collaborative efforts with Orange and other key stakeholders in the network community. We focus on exploring how Cloud-Native Functions (CNFs) and telco cloud infrastructure can support innovative use cases over the SLICES-FR experimental infrastructure.

4.6 Urgent Computing

Urgent Computing refers to a class of time-critical scientific applications that leverage distributed data sources to facilitate important decision-making in a timely manner. The overall goal of Urgent Computing is to predict the outcome of scenarios early enough to prevent critical situations or to mitigate their negative consequences. Motivating use cases refers to rapid response scenarios across the Cloud-to-IoT Continuum, such as in natural disaster management, which implies to gather the local state of each device, transform it into a global knowledge of the network, characterize the observed phenomenon according to an applied model, and finally, trigger appropriate actions. The STACK team investigates Urgent Computing through the realization of a fluid ecosystem where distributed computing resources and services are aggregated on-demand to support delay-sensitive and data-driven workflows.

5 Social and environmental responsibility

5.1 Footprint of research activities

In addition to the international travels, the environmental footprint of our research activities is linked to our intensive use of large-scale testbeds such as Grid'5000 (STACK members are often in the top 10 list of the largest consumers). Although the access to such facilities is critical to move forward in our research roadmap, it is important to recognize that they have a strong environmental impact as decribed in the next paragraph.

5.2 Impact of research results

The environmental impact of digital technology is a major scientific and societal challenge. Even though the software looks to be virtual objects, it is executed on very real hardware contributing to the carbon footprint. This impact materializes during the manufacturing and destruction of hardware infrastructure (estimated at 45% of digital consumption in 2018 by the Shift Project) and during the software use phase via terminals, networks and data centers (estimated at 55%). Stack members have been studying various approaches for several years to reduce the energy footprint of digital infrastructures during the use phase. The work carried out revolves around two main axes: (i) reducing the energy footprint of infrastructures and (ii) adapting the software applications hosted by these infrastructures according to the energy available. More precisely, this second axis investigates possible improvements that could be made by the end-users of the software themselves. At scale, involving end-users in decision-making processes concerning energy consumption would lead to more frugal Cloud computing.

In 2024, the team has taken part in two Inria Challenges (*i.e.*, “Défis”):

- The first one is built around a partnership between Inria and OVHCloud. It aims to study end-to-end eco-design of a cloud in order to reduce its environmental impact.
- The second one involved Inria and Qarnot Computing, with the support of ADEME. Entitled Pulse, this challenge aims to develop and promote best practices in terms of reducing and recycling emissions of intensive computing infrastructures.

Last but not least, STACK has been working since 2022 on a new platform project to design an innovative hardware infrastructure for the scientific study of the cross-cutting issues of computing infrastructures supporting artificial intelligence and their energy autonomy (see the Samurai project Section in 10).

6 Highlights of the year

Regarding scientific contributions, the team has produced major results on the management of large-scale infrastructures, in particular the team continued to consolidate its presence in the network and IoT communities [1, 3]

On the software side, the team has pursued its efforts on the development of the EnosLib library and the resulting artifacts to help researchers perform experiment campaigns with direct contributions from several research engineers of the team.

On the platform side, we continued our effort and took part in the different actions around the SLICES and SLICES-FR, see Section 7.2.

Finally, we would like to highlight our continued efforts to maintain the strong visibility of the team in cloud and edge-related topics. Recent projects acquired by the team, such as the French "Etoile Montante" award granted to Daniel Balouek, and the visiting researcher grant secured by Kandaraj Piamrat for a one-year stay at the National Institute of Informatics in Japan, are prime examples of the recognition of our work, alongside our significant involvement in the French Cloud PEPR initiative.

6.1 Awards

The team has received the best paper award at the 39th ACM/SIGAPP Symposium On Applied Computing for its paper [4] “VeriFog: A Generic Model-based Approach for Verifying Fog Systems at Design Time” by in the category of “Software Design and development” by Hiba Awad, Abdelghani Alidra, Hugo Bruneliere, Thomas Ledoux, and Jonathan Rivalan.

The complementary activities led by our collegues at Orange on digital twin have been honored with two additional awards. The paper "Cognitive Digital Twin for Freight Parking Management in Last Mile Delivery under Smart Cities Paradigm" published in Computers in Industry, volume 153, won the Best Paper Award in Sustainable Supply Chain at the ARIL-SCM 2024 conference, awarded by France Supply Chain. The paper focuses on integrating resources from various stakeholders for logistics operations via a digital twin platform, aiming to optimize urban resource allocation and logistic assets to meet urban logistics challenges.

Additionally, their activity was recognized at the IISE 2024 Annual Conference in North America, where they won the "Innovative Practices in Service Systems Engineering" competition, acknowledging the innovation and impact of their work in service systems engineering.

Finally, Orange received the “Outstanding Catalyst - Beyond Telco Award” at DTW24 Ignite (TMForum) for the Autonomous Networks Hyperloops project[3].

7 New software, platforms, open data

7.1 New software

7.1.1 ENOS

Name: Experimental eNvironment for OpenStack

Keywords: OpenStack, Experimentation, Reproducibility

Functional Description: A typical experiment workflow using Enos is the sequence of several phases:

- enos up : Enos will read the configuration file, get machines from the resource provider and will prepare the next phase - enos os : Enos will deploy OpenStack on the machines. This phase rely highly on Kolla deployment. - enos init-os : Enos will bootstrap the OpenStack installation (default quotas, security rules, ...) - enos bench : Enos will run a list of benchmarks. Enos support Rally and Shaker benchmarks. - enos backup : Enos will backup metrics gathered, logs and configuration files from the experiment.

Release Contributions: - Install a fixed version of Docker on nodes - Add support for kolla-ansible 12 (Openstack Wallaby), which is now the default - Add support for Debian 11 base image - Add support for Python 3.10 and 3.11 - Drop support for Python 3.7 - Update to Enoslib 8 - When creating a configuration template, use a fixed version of kolla-ansible (so that templates are not affected when future versions of Enos update the default version of kolla-ansible)

URL: `https://github.com/BeyondTheClouds/enos`

[3] TM Forum Beyond telco award 2024, valid on Jan 2025, the 15th

Publication: hal-01664515

Contact: Baptiste Jonglez

Participants: Mathieu Simonin, Marie Delavergne, Adrien Lebre, Baptiste Jonglez

Partner: Orange Labs

7.1.2 EnOSlib

Keywords: Distributed Applications, Distributed systems, Evaluation, Grid Computing, Cloud computing, Experimentation, Reproducibility, Linux, Virtualization

Functional Description: EnOSlib is a library to help you with your distributed application experiments on bare-metal testbeds. The main parts of your experiment logic is made reusable by the following EnOSlib building blocks:

- Reusable infrastructure configuration: The provider abstraction allows you to run your experiment on different environments (locally with Vagrant, Grid'5000, Chameleon, IoT-LAB and more) - Reusable software provisioning: In order to configure your nodes, EnOSlib exposes different APIs with different level of expressivity - Reusable services: Install common services such as Docker, monitoring stacks, network emulation... - Reusable experiment facilities: Tasks help you to iterate faster on your experimentation workflow

EnOSlib is designed for experimentation purpose: benchmark in a controlled environment, academic validation . . .

Release Contributions: To reduce dependencies, the default pip package no longer includes Jupyter support.

Add support for Ansible 8, 9 and 10

URL: `https://discovery.gitlabpages.inria.fr/enoslib/`

Publications: hal-01664515, hal-01689726

Contact: Mathieu Simonin

Participants: Alexis Bitaillou, Volodia Parol-Guarino, Thomas Badts, Matthieu Rakotojaona Rainimangavelo, Mathieu Simonin, Baptiste Jonglez, Marie Delavergne

7.1.3 Concerto

Name: Concerto

Keywords: Reconfiguration, Distributed Software, Component models, Dynamic software architecture

Functional Description: Concerto is an implementation of the formal model Concerto written in Python. Concerto allows to : 1. describe the life-cycle and the dependencies of software components, 2. describe a components assembly that forms the overall life-cycle of a distributed software, 3. automatically reconfigure a Concerto assembly of components by using a set of reconfiguration instructions as well as a formal operational semantics.

URL: `https://gitlab.inria.fr/VeRDi-project/concerto`

Publications: hal-03103714, hal-02535077, hal-01897803

Contact: Hélène Coullon

Participants: Christian Perez, Hélène Coullon, Maverick Chardet, Simon Robillard

Partners: IMT Atlantique, LS2N, LIP

7.1.4 StreamBed

Name: StreamBed capacity planning for steam processing

Keywords: Big data, Data stream, Performance measure

Scientific Description: StreamBed is a capacity planning system for stream processing. It predicts, ahead of any production deployment, the resources that a query will require to process an incoming data rate sustainably, and the appropriate configuration of these resources. StreamBed builds a capacity planning model by piloting a series of runs of the target query in a small-scale, controlled testbed. We implement StreamBed for the popular Flink DSP engine. Our evaluation with large-scale queries of the Nexmark benchmark demonstrates that StreamBed can effectively and accurately predict capacity requirements for jobs spanning more than 1,000 cores using a testbed of only 48 cores.

Functional Description: StreamBed is a capacity planning system for stream processing. It predicts, ahead of any production deployment, the resources that a query will require to process an incoming data rate sustainably, and the appropriate configuration of these resources. StreamBed builds a capacity planning model by piloting a series of runs of the target query in a small-scale, controlled testbed. We implement StreamBed for the popular Flink DSP engine. Our evaluation with large-scale queries of the Nexmark benchmark demonstrates that StreamBed can effectively and accurately predict capacity requirements for jobs spanning more than 1,000 cores using a testbed of only 48 cores.

News of the Year: Publication and artefact available on GitHub.

URL: `https://github.com/CloudLargeScale-UCLouvain/StreamBed`

Publication: hal-04708354

Contact: Guillaume Rosinosky

Participant: Guillaume Rosinosky

Partner: Université Catholique de Louvain (UCL), Louvain-la-Neuve,Belgium

7.1.5 Cheops

Name: Cheops for the edge

Keywords: Edge Computing, Geo-distribution, Infrastructure software

Functional Description: Cheops handles the task of synchronizing and replicating your application resources (in the sense of REST).

The user will interact with Cheops to perform operations on resources, and Cheops will then interact with replicas of your application to make sure that all resource copies eventually converge to the same state.

It is assumed that operations are always associated to a specific resource. Cheops gives the possibility to specify the exact distribution of each resource manually so that operators can define how they want them to be spread.

URL: `https://cheops-for-the-edge-discovery-e233bd9558e1939580edccc727171f286.gitlabpages.inria.fr/`

Publications: hal-04886168, hal-04522961, tel-04081084

Contact: Adrien Lebre

7.1.6 Edge-to-cloud video processing

Keywords: Edge Computing, Video analysis

Functional Description: The software performs distributed video processing to identify animals in video feeds, in order to warn populations if a dangerous animal is detected.

The primary goal is to provide a use-case to perform research on edge-to-cloud infrastructure. To achieve this, the software is instrumented with many metrics.

The software is composed of three components:

- video capture: either from a camera, or by replaying a pre-recorded video file that can be parametrized

- motion detection: detects motion in video feeds. If a motion is detected, the feed is forwarded to the recognizer component

- object recognizer: uses a YOLO model to determine the kind of animal visible in the video feed

These three components are designed to execute in different places in the edge-to-cloud infrastructure, so that researchers can explore the trade-off between performance, latency, transferred data, and quality.

URL: https://gitlab.inria.fr/STACK-RESEARCH-GROUP/software/edge-to-cloud-video-processing

Contact: Baptiste Jonglez

Partner: IMT Atlantique

7.2 New platforms

7.2.1 Grid'5000

Participants: Remous-Aris Koutsiamanis, Baptiste Jonglez, Adrien Lebre, Jean Marc Menaud.

Grid'5000 is a large-scale and versatile testbed for experiment-driven research in all areas of computer science, with a focus on parallel and distributed computing including Cloud, HPC and Big Data. It provides access to a large amount of resources: 12000 cores, 800 compute-nodes grouped in homogeneous clusters, and featuring various technologies (GPU, SSD, NVMe, 10G and 25G Ethernet, Infiniband, Omni-Path) and advanced monitoring and measurement features for traces collection of networking and power consumption, providing a deep understanding of experiments. It is highly reconfigurable and controllable. STACK members are strongly involved into the management and the supervision of the testbed, notably through the steering committee or the SeDuCe testbed described hereafter.

7.2.2 PiSeDuCe

Participants: Remous-Aris Koutsiamanis, Baptiste Jonglez, Jean Marc Menaud.

We continue to manage and extend the PiSeDuCe platform, a deployment and reservation system for Edge Computing infrastructures composed of multiple Raspberry Pi Cluster started in 2020. The platform is typically composed of a cluster of 8 Raspberry Pi, which costs less than 900 euros and only needs an electrical outlet and a wifi connection for its installation and configuration Funded by the CNRS through the Kabuto project, and in connection with the SLICES-FR initiative, we have extended PiSeduce to propose a device to cloud deployment system (from devices on Fit IoTLab to servers in Grid'5000). PiSeDuCe and SeDuce led us to submit the Samurai CPER proposal. Recent developments have made the platform more performant, being able to manage 45 worker Raspberry Pis using just one controller Raspbery Pi. The platform has also been demonstrated and used in a hands-on tutorial at the "Green IT - Numérique responsable" 2024 summer school in Nantes.

7.2.3 SAMURAI

Participants: Remous-Aris Koutsiamanis, Baptiste Jonglez, Jean Marc Menaud.

The SAMURAI (Sustainable And autonoMoUs gReen computing for AI) project is currently being financed as a part of the energy and digital transition theme. The project aims at reinforcing an innovative hardware infrastructure for the scientific study of the intersecting problems of computing infrastructure that supports artificial intelligence and its energy autonomy. SAMURAI is focused on extending SeDuCe into energy autonomy by adding a smart and clean energy storage system. Additionally, it has extended the capabilities of the platform by adding AI computing nodes (servers with GPUs) for the scientific study of AI tools. Finally, it will also add new sensor nodes within the Nantes connected object platform (Nantes site of the national SLICES-FR platform) to support future work on embedded AI as well as moregenerally on the Cloud-Edge-IoT continuum. As the majority of the hardware for the IoT, Edge and Cloud with GPUs has been procured, the project is focusing on building the platform and network integration with the SeDuCe and SLICES-FR platform.

7.2.4 SLICES-FR/SLICES

Participants: Remous-Aris Koutsiamanis, Baptiste Jonglez, Adrien Lebre, Jean Marc Menaud.

In 2024, STACK participated in the SLICES-PP project (Preparatory Phase) as part of the SLICES-RI European infrastructure initiative. STACK was particularly active within SLICES-FR, the French component of the infrastructure. STACK Members have been involved in the definition and bootstrapping of the SLICES-FR infrastructure. This infrastructure can be seen as a merge of the Grid'5000 and FIT testbeds with the goal of providing a common platform for experimental Computer Science (Next Generation Internet, Internet of things, network functions, clouds, HPC, big data, etc.). Adrien Lèbre and Remous-Aris Koutsiamanis are part of the SLICES-FR Board (currently provisional, pending the official creation of the GIS legal structure). Additionally, Remous-Aris Koutsiamanis and Baptiste Jonglez are members of the Architects Committee of SLICES-FR, where they contributed key design documents focused on networking and interoperability, aligning with the overall SLICES-RI design. Finally, Adrien Lèbre serves as the French scientific representative on the SLICES Interim Supervisory Board, a provisional body awaiting the formal establishment of the ERIC legal structure.

8 New results

8.1 Resource Management

Participants: Daniel Balouek, Hélène Coullon, Duc Thinh Ngo, Kandaraj Piamrat, Thomas Ledoux, Pierre Jacquet, Guillaume Rosinosky.

The evolution of the cloud computing paradigm in the last decade has amplified the access to on-demand services (economically attractive, easy-to-use manner, etc.). However, the current model, built upon a few large datacenters (DCs), may not be suited to guarantee the needs of new use cases, notably the boom of the Internet of Things (IoT). To better respond to the new requirements (in terms of delay, traffic, etc.), compute and storage resources should be deployed closer to the end-user, forming with the national and regional data centers a new computing continuum. The question is then how to manage such a continuum to provide end-users the same services that made the current cloud computing model so successful. In 2024, we have continued our effort to answer this question and delivered multiple contributions, including new activities around Urgent Computing (see Section 4).

IaaS Consolidation with Resource Oversubscription: Given the surge in energy consumption in data centers, optimizing their infrastructures is crucial. However, focus should also be on the server itself and its tasks, with a key area being the utilization rate. Low utilization rates present significant potential for gain. Pierre Jacquet's thesis proposal [30] is to analyze these utilization rates on the basis of several complementary contributions. First, the calculation of individual server oversubscription ratio. By considering the individual stability of servers, it is possible to fine-tune the calculation of this ratio without causing additional violations [7]. Second, the introduction of a new oversubscription paradigm. By first demonstrating that VM vCPUs are not uniformly used in a real-world context, we expose to VMs cores of different powers (by oversubscribing them to different amounts) and demonstrate that this paradigm can improve overall performance [21]. Third, the complementarity of oversubscription techniques to reduce unallocated resources. Comparing so-called premium VMs and oversubscribed VMs identifies that they tend to saturate their hosts' resources differently. By hosting them on the same servers, it is thus possible to benefit from synergies and reduce the number of servers by up to 9.6% [20].

Network resource management: The team continues to explore the Cloud-IoT continuum, integrating edge computing, cloud computing, and IoT for rapid IIoT development. Despite its promise, challenges like robustness, latency, and model convergence persist due to system and data heterogeneity. Federated Learning (FL) partially addresses these but has limitations. To improve scalability and energy efficiency, we propose combining Hierarchical FL (HFL) and Spiking Neural Networks (SNN) [2] for building a scalable and energy-efficient solution for the industrial CEI continuum. In telecommunications, growing users and devices in next-generation networks (beyond 5G) intensify traffic demands, stressing limited resources. Machine Learning (ML) and Deep Learning (DL) offer advanced traffic prediction and network optimization. We review recent DL-based Network Traffic Prediction (NTP) advances, conducts experiments, and discusses challenges and future directions in [1]. Additionally, cellular traffic prediction faces challenges like dynamic base station deployment. Spatio-temporal forecasting using Graph Neural Networks (GNNs) addresses this with a novel inductive learning scheme [36]. The proposed model processes diverse graphs with one-time training and adapts easily via transfer learning, showing up to 9.8% performance improvement in rare-data scenarios.

Urgent Computing: Urgent computing scenarios describe challenges in promptly responding to changes in the Edge-Cloud Continuum or adapting the quality of service (QoS) constraints of the application. This line of work is highly motivated by decision-making systems or natural disaster case studies such as earthquake early warning and wildfires.

In 2024, we proposed a set of collaborative services, the CrowdMesh platform [17], to allow for processing, analysis, and sharing of crowd monitoring datasets constructed over the Edge-Cloud Continuum. Comprehensive tools and best practices are needed to ensure that analysis pipelines are robust, and that results can be replicated. This approach is complementary to mechanisms for monitoring resources at the edge in real-time, orchestrating service provisioning, and coordinating sensing tasks as proposed in [16]. This work has been associated to a live demonstration showcasing autoscaling of containers and distributed sensing, while considering utility metrics to help achieve a fluid workload in Kubernetes clusters.

Resources for biomedical analyses: In the course of our long-running cooperation with researchers from Nantes University Hospital (CHU Nantes), we have worked on clinical data for lung transplantations in order to develop medical analyses for the survival rate after such transplantations.

The main limitation to long-term lung transplant (LT) survival is chronic lung allograft dysfunction (CLAD), which leads to irreversible lung damage and significant mortality. Individual factors can impact CLAD, but no large genetic investigation has been conducted to date. We have established the multicentric

Genetic COhort in Lung Transplantation (GenCOLT) biobank from a rich and homogeneous sub-part of COLT cohort [5]. Overall, GenCOLT is an accurate snapshot of LT clinical practice in France and Belgium between 2009 and 2018. It currently represents one of the largest genetic biobanks dedicated to LT with data available simultaneously for donors and recipients. This unique cohort will empower to run comprehensive GWAS investigations of CLAD and other LT outcomes.

8.2 Programming Support

Participants: Farid Arfi, Daniel Balouek, Hélène Coullon, Marie Delavergne, Mohamed Graiet, Thomas Ledoux, Jacques Noyé, Antoine Omond, Eloi Perdereau, Jolan Philippe, Hiba Awad, Mario Südholt, Divi De Lacour.

Fog Modeling: Fog Computing decentralizes the Cloud by distributing computation, storage, and network resources geographically, reducing bottlenecks and unnecessary data movements. However, managing Fog systems can be complex and costly. To address this, a model-based approach called VeriFog is proposed for verifying Fog systems during the design phase [14] . This approach uses a customizable Fog Modeling Language (FML) and enables the automatic generation of deployment configuration files [4]. Experiments with three use cases from different application domains demonstrated the ability to generate usable configuration files for various deployment tools. Collaborating with the industrial partner Smile, this approach aims to support the entire lifecycle of Fog systems.

Ensuring correct communications during the execution of IoT applications is not guaranteed. Moreover, IoT devices exchange the data derived from various Cloud providers and in accordance with different protocols. In [19], we propose a formal approach to model and verify the applications deployed over IoT-Cloud environments. The proposed model encompasses four verification levels: the Structural, Functional, Operational and Behavioral levels.

Configuration languages: In 2023, as part of the OTPaaS project, we started to study the anatomy of configuration languages, with an initial focus on CUE [76]. In 2024, we have resumed this work in the context of the PEPR Taranis.

Software deployment/reconfiguration: For a few years, the team has been working on deployment and dynamic reconfiguration of distributed software systems through the tool Concerto [49]. Compared to the literature, Concerto is a component model closer to Infrastructure-as-Code (IaC) approaches of the DevOps community (as Aeolus [54]). In Concerto, the lifecycle of pieces of software (application or infrastructure) are modeled in a programmable manner with fine grain dependencies, enhancing the flexibility and speed of deployments and management procedures compared to existing approaches.

The accumulated knowledge on component-based distributed software reconfiguration has made possible the publication in 2024 of a survey in ACM Computing Surveys [6], a great success. In particular, this survey analyzes the state of the art on reconfiguration through the spectrum of verification. Indeed, given the complexity of distributed software and the adverse effects of incorrect reconfigurations, a suitable methodology is needed to ensure the correctness of reconfigurations in component-based systems. The survey distinguishes different ways in which formal methods can improve the reliability of reconfigurations, and lists techniques that contribute to solving each of these particular scientific challenges.

In 2023, the team has worked on the decentralized version of Concerto, namely Concerto-D [75]. Such a decentralized reconfiguration tool is very useful when building the global knowledge on the state of the system is difficult, impossible, or very costly because of faults, disconnection or scale. Concerto-D has been developed in the context of the DAO project at the University of Tromso, Norway, where a Tundra observatory is geographically deployed and composed of highly constrained nodes equipped with batteries (only) and small devices.

In 2024, we have worked on the formal specfication (under the form of an operational semantics) of Concerto-D. As being decentralized, Concerto-D is more complex than Concerto. In particular, it has to synchronize n reconfiguration programs distributed accross distant nodes without a central coordinator, hence requiring communications. The semantics of Concerto-D has been mechanized with the Maude rewriting system [53] and published in a workshop [13]. This mechanical semantics offers an interpreter of the semantics, a model-checker to prove properties on Concerto-D examples, and the possibility to extract the semantics to Lean[4] to make more general proofs on the semantics itself. This work will be extended and submitted to a journal in 2025.

In 2024, we have also extended Concerto-D with a decentralized planner namely Ballet [24]. A planner is a decision entity responsible for generating correct-by-construction Concerto-D reconfiguration programs from the current and the target state. This is comparable to declarative approches of IaC tools where the desired state is specified, not the way to reach it. A planner has already been designed for Concerto [81] but is fully centralized, requiring a global knowledge on the system. Ballet instead generates n reconfiguration programs for Concerto-D in a decentralized manner. To this purpose it combines constraint programming and distributed algorithms to spread additional constraints in local constraint programming models. Ballet has been evaluated on synthetic cases and the real use-case of reconfiguring OpenStack.

In 2024, while not being published yet, we have started to work more directly on Infrastructure-as-Code tools and languages. After working on Concerto, Concerto-D and Ballet (i.e., well constructed research models and prototypes), we think it is important to start from existing production and more complex languages, with the aim of bringing together both sides. First, in collaboration with Daniel Sokolowski (who visited the team in December 2023) and Guido Salvaneschi, and in the context of the PEPR Cloud (taranis project), we are working on a submission arround the formal semantics and verification of Terraform and Pulumi (provisioning tools). Second, in collaboration with Orange, the PhD of Tengfei An has started on convergence issue when combining custom operators in Kubernetes. Third, the Inria transfer action (ADT) project CoAnsible has been accepted to develop an Ansible extension that uses Concerto as a coordination backend. Finally, the ANR project For-CoaLa (coordinated by Hélène Coullon) has been accepted and will focus on the formal semantics of Ansible and CoAnsible (including Concerto).

Reifying geo-distribution at the software level: One question to answer in the shift from the Cloud to the Edge computing paradigm is: how distributed applications developed for Cloud platforms can benefit from the opportunities of the Edge while dealing with inherent constraints of wide-area network links? Our solution to this question is to give the illusion of “single service images” spreading over the Edge infrastructure. Thanks to the modularity of micro-service based applications, one can deploy multiple instances of the same service (one per edge site) and deliver collaborations between them according to each request. This non-invasive approach is made possible by (i) a DSL that extends the application API and allows DevOps to program where/how the execution of each request should be executed, (ii) and its runtime, Cheops, a service that interprets and orchestrates each request in order to satisfy the geo-distribution parameters, allowing collaborations in a transparent manner for the underlying application. While previous efforts focused on resource sharing and replication, we introduced this year a novel collaboration method that extends resources across multiple instances, going beyond simple replication. Our approach employs a shard-like strategy, enabling the creation of a distributed resource with a unified state view while mitigating a synchronization overhead. The effectiveness of our mechanism has been demonstrated through a proof-of-concept implemented on top of the Kubernetes ecosystem [10]

8.3 Energy-aware computing

Participants: Remous-Aris Koutsiamanis, Thomas Ledoux, Anas Mokhtari, Jean-Marc Menaud, Clément Mommessin, Guillaume Rosinosky, Thierry Coupaye.

[4] Theorem proving with Lean

The activities on this axis are mainly related to the design, development and deployment of the SAMURAI project(7.2.3), a testbed that will allow researchers to investigate energy related challenges over the computing continuum (from the Cloud to IoT devices/cyber physical systems).

In [9], we present an in-depth case study exploring the application of federated learning (FL) for cellular traffic forecasting in mobile networks. This work extends the findings of [77] by delving deeper into the potential of FL as a distributed, privacy-preserving approach that addresses data heterogeneity across base stations while enabling high-quality predictions. Our study evaluates various facets of the federated framework, including model aggregation techniques, personalized learning, and the integration of external data sources, while also considering the impact of outliers and individual client performance. By analyzing both predictive accuracy and environmental sustainability, we underline FL's dual benefits of being an effective and eco-friendly solution for mobile traffic prediction. The results demonstrate FL's robustness in achieving privacy-conscious, responsive, and efficient resource allocation in real-world telecommunications scenarios.

In [29], we explore the problem of scheduling precedence-constrained tasks on an edge-cloud platform powered by green energy sources. This work addresses the challenges of minimizing the reliance on conventional, non-renewable energy while efficiently managing computational tasks. The study introduces a novel scheduling model that incorporates renewable energy dynamics and precedence constraints, emphasizing short time windows where energy availability is assumed to be constant. By leveraging constraints-based programming and heuristic approaches, the proposed solutions aim to optimize task execution across distributed computing sites connected to renewable energy sources such as solar panels and wind turbines.

In [26], we provide novel methods to estimate the computing needs of big data stream processing jobs, using cheap small-scale testbeds with a regression-based approach. Our models permit to estimate the needs in terms of CPU cores and memory for a given event rate. We prove that we can predict the needs of a rate needing 20 times the size of the small-scale testbed, even on non-trivial jobs showing non-linear behaviours on the relation between rate and computing needs - typically stateful jobs.

In [28], we propose to address the inherent complexities of effective application scheduling in cloud environments by focusing on reducing operational costs more specifically in terms of energy consumption and application migration expenses. In this intent, we propose a two-stage decision-making process supplemented by heuristics and metaheuristics. Extensive experiments executed on a public dataset and industry-standard contests demonstrate the significant advancements provided by our approach.

In [23], we adress the role of end-users in their IT enery consumption. Cloud services are big energy consumers, leading to high CO2 emissions. Current strategies focus on infrastructure, but this paper suggests involving end-users to reduce the carbon footprint. Users can choose a mode to control the carbon impact of their cloud applications. A dynamic algorithm then adjusts the application based on user choice and energy source carbon intensity. Tests on an image-resizing app show significant energy savings and a good balance between quality and energy use, especially when using green energy.

8.4 Security and Privacy

Participants: Wilmer Edicson Garzon Alfonso, Houssem Jmal, Kandaraj Piamrat, Mario Südholt.

This year the STACK team has provided new results on security and privacy issues in the networking and biomedical domains. We have developed new AI-based methods that support new means of property analysis and dynamic adaptation.

Network security and privacy: The Internet of Things (IoT) has revolutionized communication technology and customer services, with AI enhancing IoT operations in modern applications. The convergence of IoT and AI has created Intelligent IoT (IIoT), transforming businesses and industries. In [3], we survey IIoT's applications in mobile networks, focusing on key domains like smart healthcare, cities, transportation, and industries. We examine security issues, including network attacks, confidentiality, and intrusion, alongside countermeasures. Privacy concerns such as data and location leakage are also analyzed. Lastly, we identify challenges and suggest future research directions in this evolving field. Moreover, a security mechanism, which is intrusion detection system, has been investigated in dynamic distributed and decentralized system using FL. It is known that FL empowers edge devices with AI capabilities but faces challenges like single-point failures and client heterogeneity. Semi-decentralized FL (SDFL) addresses this by using multiple servers, though it requires careful coordination. To tackle these issues, the proposed TUNE-FL [22] introduces an adaptive semi-synchronous mechanism that ensures consensus across network nodes while accounting for device heterogeneity. Evaluated on intrusion detection system datasets, TUNE-FL outperformed baseline models in accuracy and reduced training time by approximately 97-fold.

Secure federated learning. This year, we presented Ti-Skol, a model and prototype implementation for securing federated learning algorithms, at the main international conference IEEE BigData. [18]

Federated Learning (FL) is a growing technology that enables training of Deep Learning models on private data. Many FL enhancements have been proposed, notably for better security and privacy. Current architectures and frameworks focus on specific sets of enhancements with little extensibility and do not support composition of enhancements.

Ti-skol supports the composition of security and privacy countermeasures, including countermeasure incompatibilities. Ti-skol also enables modular management of FL enhancements beyond security, being compatible with most enhancements. Ti-skol is enables the assessment of the cost of countermeasures, individually or in combination.

The framework is evaluated on a use-case of Volunteer Deep Learning – applying Volunteer Computing to reduce the cost of large model training by harnessing idle resources of single machines into the required massive distributed computing power. Experimental results show that Ti-skol is scalable as the network size increases. While adding security countermeasures such as Byzantine protections or secure aggregation substantially increase computing overheads, they do not change their order of magnitude, individually or in combination.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

Kelio (formely Bodet Software)

Participants: Thomas Ledoux.

The ArchOps 2 Chair (for Architecture, Deployment and Administration of Agile IT Infrastructures) is an industrial chair of IMT Atlantique, in partnership with Kelio, an SME specialized in solutions for time and attendance management. It is dedicated to all IMT Atlantique students in the field of IT. It is also a channel for the transfer of high-level skills: researchers, experts and industrials.

In 2024, several activities were carried out, such as a conference on Shift-left testing in software life-cycle (with around 30 participants), an event promoting the integration of international students at IMT Atlantique (with around 150 participants) and preliminary discussions on the topic of a joint PhD thesis with the Stack team.

9.2 Bilateral grants with industry

Alterway/Smile

Participants: Thomas Ledoux, Hiba Awad.

In 2020, during the preparation of the ANR SeMaFoR project, we started a cooperation with Alter-way/Smile, an SME specialized in Cloud and DevOps technologies. This cooperation resulted in a joint PhD thesis (called Cifre) entitled "Quality of Service Assurance Before Deployment of Fog Systems with Model-Based Engineering and DevOps" started in Nov. 2021.

Several research papers [14] ,[4, 15] have resulted from this collaboration (including a best paper at SAC 2024 [14]) and the thesis is due to be defended in March 2025.

Kapela

Participants: Thomas Ledoux, Lylian Siffre.

Lylian Siffre started his PhD in Nov. 2024 on the subject "Impacts and Uses of Local-First Software for Energy Optimization of IT Services", under a co-supervision with Kapela (Constellation group), an SME specialized in IT eco-design.

OVHcloud

Participants: Yasmina Bouziem, Helene Coullon, Thomas Ledoux, Pierre Jacquet.

In 2021, INRIA and OVHcloud have signed an agreement to jointly study the problem of a more frugal Cloud. They have identified 3 axes : (i) Software eco-design of Cloud services and applications; (ii) Energy efficiency leverages; (iii) Impact reduction and support for Cloud users. The Stack team obtained a PhD grant (with the Spirals team) and a 24-month post-doc grant.

Pierre Jacquet defended his doctoral thesis "Enhancing IaaS Consolidation with Resource Oversubscription" [30] in July 2024. He has also published a number of high-quality publications [7, 20, 21]. He is a candidate for this year's Gilles Kahn Thesis Prize.

Yasmina Bouziem started her postdoc in September 2022 for two years, under a co-supervision with the Avalon team in Lyon with the subject "pulling the energy cost of a reconfiguration execution within reconfiguration decisions".

Orange

Participants: Paul Bori, Divi de Lacour, Adrien Lebre, Thomas Ledoux, Jean-Marc Menaud, Duc-Thinh Ngo, Kandaraj Piamrat, Mario Südholt.

Since 2022, Orange Labs and the Stack team have launched several PhD grants.

Paul Bori started his PhD in January 2023, with the subject "Container application security: a programmable OS-level approach to monitoring network flows and process executions".

Duc-Thinh Ngo started his PhD in December 2022, on the subject "Dynamic graph learning algorithms for the digital twin in edge-cloud continuum", under a co-supervision with Orange team in Rennes.

Divi de Lacour started his PhD in January 2022, on the subject "Architecture et services pour la protection des données pour systèmes coopératifs autonomes", under a co-supervision with the Orange team in Chatillon (Paris south region).

Please note that, since April 2024, the STACK research group has officially become a joint team between Inria, IMT Atlantique, Nantes University, and Orange Labs.

Ericsson

Participants: Samia Boutalbi, Mario Südholt, Remous-Aris Koutsiamanis.

Samia Boutalbi started her PhD in January 2022, on the subject "Secure deployment of micro-services in a shared Cloud RAN/MEC environment", under a co-supervision with the Ericsson team in Paris.

10 Partnerships and cooperations

10.1 International research visitors

10.1.1 Visits of international scientists

Other international visits to the team

Lorenzo Carnevale

Status Postdoc

Institution of origin: University of Messina

Country: Italy

Dates: 01/06/2024 - 31/08/2024

Context of the visit: Collaboration on european project (TEMA) and joint papers

Mobility program/type of mobility: Research stay

Ruben Opdebeeck

Status Postdoc

Institution of origin:

Country: Belgium

Dates: 2024-11-18 to 2024-11-21

Context of the visit: Invited talk to the PEPR Cloud and GDR GPL and RSD communities and collaboration

10.1.2 Visits to international teams

Sabbatical programme

Kandaraj Piamrat

Visited institution: JFLI/NII

Country: Japan

Dates: 01/09/2024-31/08/2025

Context of the visit: Collaboration with researchers at NII

Mobility program/type of mobility: CNRS research leave

Research stays abroad

Daniel Balouek

Visited institution: SCI Institute, University of Utah

Country: USA

Dates: 19/02/2024-01/03/2024

Context of the visit: Collaboration on NSF project

Mobility program/type of mobility: Research stay and lectures

10.2 European initiatives

10.2.1 Horizon Europe

SLICES-PP / SLICES-RI (Scientific Large Scale Infrastructure for Computing/Communication Experimental Studies - Research Instrument)

Participants: Adrien Lebre, Remous-Aris Koutsiamanis.

The STACK Research team is actively involved in SLICES-RI, the European initiative aimed at developing a large-scale research infrastructure for digital sciences. This project brings together partners from 16 countries, including France, Greece, Poland, Norway, the Netherlands, Italy, and Switzerland, all committed to contributing resources and expertise.

Members of our team are currently actively participating in the Preparatory Phase (SLICES-PP) of the project, helping to shape its direction and implementation. Our goal is to continue this involvement at both the European level and the national level via SLICES-FR, contributing to the creation of a flexible platform that supports large-scale experimental research in areas like networking protocols, radio technologies, and cloud and edge computing architectures.

Once established, SLICES-RI aims provide researchers across Europe with long-term access to advanced computing, storage, and network resources. This infrastructure will facilitate experiments in various domains, including information theory, networking, distributed systems, and software engineering, thereby enabling experimentation and innovation in multiple domains such as smart cities, e-health, industrial internet, transport, and energy solutions.

SEED: Training the next generation of technological scientists to achieve Societal, Energy, Environmental, industrial and Digital transitions

Participants: Mario Südholt *(Coordinator)*.

SEED, which stands for Societal, Energy, Environmental, industrial and Digital transitions, is a 60-month interdisciplinary, international and intersectoral doctoral training programme offered by IMT Atlantique and co-funded by the European Union. The programme itself is designed to nurture four key dimensions: thesis interdisciplinarity, internationality, cross-sector experience, and promotion of innovation. It offers 40 fully funded early-stage researcher (ESR) positions within three different tracks. Each track builds on the same fundamental excellence trainings implementing a 4i approach (Interdisciplinarity, Internationality, Intersectorial, Innovation), while providing a different degree of mobility and focus. Further information at:Cordis website.

10.2.2 Other european programs/initiatives

DI4SPDS (Distributed Intelligence for Enhancing Security and Privacy of Decentralised and Distributed Systems)

Participants: Houssem Jmal, Kandaraj Piamrat.

Decentralised systems face challenges from sophisticated cyber-attacks that evolve and propagate to disrupt different parts. Additionally, communication overhead makes implementing authentication and access control complex. Existing approaches unlikely provide effective access control and multi-stage attack detection due to limited event capture and information correlation. This project offers a framework to improve security and privacy of decentralised systems through cross-domain access control, collaborative intrusion detection, and dynamic risk management considering resource consumption. It facilitates subsystem collaboration to prevent widespread disruption from attacks and share threat awareness. The project will develop methods and prototypes utilizing blockchain, federated learning, and multi-agent architecture to enhance access control, detection, risk management, and response capabilities. The consortium is composed of four partners: Nantes University, LUT University (Finland), Universidad de Castilla - La Mancha (Spain), and Firat University (Turkey). LUT is the coordinator of the project DI4SPDS has been accepted in July 2023, started from March 2024 for 36 months, with an allocated budget of 874k€ (230K€ for Stack).

DISCOVER-US: Collaboration with NSF on fundamental research on new concepts for distributed computing and swarm intelligence

Participants: Daniel Balouek.

DISCOVER-US represents a collaboration between EU and US research institutions, focusing on advancing distributed computing and swarm intelligence. At the core of DISCOVER-US is the commitment to develop a practical infrastructure that supports a dynamic research ecosystem. The /stack team has significantly contributed to a vision white paper highlighting requirements and potential approaches for next-generation distributed computing. This work is aligned with key objectives from Horizon Europe and the US National Science Foundation, driving progress in cloud-to-edge processing technologies, AI, and cybersecurity.

10.3 National initiatives

10.3.1 ANR

For-CoaLa (Formalization of Configuration Languages)

Participants: Hélène Coullon *(coordinator)*.

Large distributed software systems (applications or infrastructures) are now ubiquitous, with component-based systems (e.g., service-oriented architectures or microservices) offering a convenient way to structure large systems, in particular distributed systems deployed in the Cloud, in the core, or at the edge of the network. DevOps operations, that include system configurations and reconfigurations, are required to handle various kinds of scenarios such as fault tolerance, scalability, software updates, or various optimizations, etc. Such changes may lead to faults. A study of 597 unplanned outages that affected popular cloud services between 2009 and 2015 found that 16% of them were caused by a system upgrade.

On the one hand, many configuration tools and languages exist in the DevOps community, some of them being specific to the provisioning of resources in Cloud providers, packaging problems, containerized deployments, configuration of applications or infrastructures, etc. The main advantage of these tools is their full integration and adoption in the DevOps community. Their disadvantage is they lack both formal and textual specifications. Moreover, their contours are blurred. On the other hand, many initiatives have been studied in academia to contribute to the deployment, configuration, and reconfiguration of distributed software, bringing improvements such as expressivity, speed, safety, etc. Many come with precise and sometimes formal definitions. However, they lack the breadth of the mainstream DevOps tools.

The goal of For-CoaLa is twofold: (1) understand and bridge the gap between a popular tool from the DevOps community (Ansible) and a tool from academia (Concerto); (2) improve the understanding of these languages based on mechanized formal semantics and develop verified semantic-preserving cross-language transformations. See the For-CoaLa web site for more information.

SeMaFoR (Self-Management of Fog Resources)

Participants: Thomas Ledoux *(coordinator)*, Hélène Coullon, Jolan Philippe.

Fog Computing is a paradigm that aims to decentralize the Cloud at the edge of the network to geographically distribute computing/storage resources and their associated services. The objective of the SeMaFoR project is to model, design and develop a generic and decentralized solution for the self-management of Fog resources.

The consortium is composed of three partners: LS2N-IMT Atlantique (Stack, NaoMod, TASC), LIP6-Sorbonne Université (Delys), Alter way/Smile (SME). The Stack team supervises the project. SeMaFoR is running for 42 months (starting in March 2021 with an allocated budget of 506k€, 230K€ for Stack). See the Semafor web site for more information.

The main result of the year was the publication [24] dealing with Choreography of Cross-DevOps Reconfiguration.

PicNic (Transfert de grands volumes de données entre datacenters)

Participants: Jean-Marc Menaud *(STACK representative)*, Remous-Aris Koutsiamanis, Adrien Lebre, Abdou Seck, Guillaume Rosinosky.

Large dataset transfer from one datacenter to another is still an open issue. Currently, the most efficient solution is the exchange of a hard drive with an express carrier, as proposed by Amazon with its SnowBall offer. Recent evolutions regarding datacenter interconnects announce bandwidths from 100 to 400 Gb/s. The contention point is not the network anymore, but the applications which centralize data transfers and do not exploit parallelism capacities from datacenters which include many servers (and especially many network interfaces – NIC). The PicNic project addresses this issue by allowing applications to exploit network cards available in a datacenter, remotely, in order to optimize transfers (hence the acronym PicNic). The objective is to design a set of system services for massive data transfer between datacenters, exploiting distribution and parallelisation of networks flows.

The consortium is composed of several partners: Laboratoire d'Informatique du Parallélisme, Institut de Cancérologie de l'Ouest / Informatique, Institut de Recherche en Informatique de Toulouse, Laboratoire des Sciences du Numérique de Nantes, Laboratoire d'Informatique de Grenoble, and Nutanix France.

PiCNiC will be running for 42 months (starting in Sept 2021 with an allocated budget of 495k€, 170k€ for STACK).

Taranis

Participants: Thomas Ledoux, Helene Coullon *(STACK Representative)*, Adrien Lebre, Daniel Balouek.

New infrastructures, such as Edge Computing or the Cloud-Edge-IoT computing continuum, make cloud issues even more complex, as they add new challenges linked to the diversity and heterogeneity of resources (from small sensors to data centers/HPCs, from low-power networks to core networks), geographical distribution, as well as increased requirements for dynamicity and security, all under constraints such as energy consumption.

To exploit these new infrastructures efficiently, the Taranis project is based on a strategy aimed at abstracting the description of the structure of applications and resources in order to automate their management even further. In this way, it will be possible to globally optimize the resources used with regard to multi-criteria objectives (price, deadline, performance, energy, etc.) on both the user side (applications) and the resource provider side (infrastructures). Taranis also addresses the challenges of abstracting application reconfiguration and dynamically adapting resource usage.

The consortium is composed of 6 partners (Inria, CNRS, IMT, University of Grenoble Alpes, CEA and Universtity of Rennes) for a budget of 7.2M€ (470K€ for STACK) overall.

Spirec

Participants: Mario Sudholt *(coordinator)*, Remous-Aris Koutsiamanis *(STACK Representative)*, Kandaraj Piamrat, Carlos Gonzalez, Jean-Marc Menaud.

The SPIREC project aims to meet the challenges of monitoring the cloud-Edge-IoT continuum services, detecting execution anomalies and predicting resource utilization. The consortium is composed of 6 partners (Inria, CNRS, CEA, IMT Atlantique, Télécom SudParis and Université de Versailles, Saint Quentin) for a budget of $\tilde{2}$,7Millions€ (580K€ for STACK) overall.

Steel

Participants: Daniel Balouek, Carlos Gonzalez *(STACK Reresentative)*, Adrien Lebre.

The STEEL project aims to provide solutions for efficient and secure data storage and processing on cloud-based infrastructures. The consortium is composed of 8 partners (Inria, CNRS, University of Grenoble, University of Bordeaux, University of Rennes, IMT Atlantique, IMT TeraLab and IN2P3) for a budget of $\tilde{3}$Millions€ (300K€ for STACK) overall. activies are organized around three technical work packages. A fourth work package is dedicated to management, communication and dissemination of results. STACK members are involved in the second wp, addressing the challenges related to the management of data sets in presence of node failures and network partitions.

CareCloud

Participants: Jean-Marc Menaud *(STACK representative)*, Remous-Aris Koutsiamanis.

At a time when climate change is a growing concern, with serious consequences for people and the planet worldwide, all sectors (transport, construction, agriculture, industry, etc.) must contribute to the effort to reduce GHG emissions. Clouds, despite their ability to optimize processes in other sectors, are no exception to this observation: the increasing slope of their GHG emissions must be reversed, or their potential benefits in other sectors will be wiped out. The CARECloud project aims to drastically reduce the environmental impact of cloud infrastructures. The consortium is composed of 4 partners (CNRS, IMT Atlantique, Inria and Univeristy of Paul Sabatier - Toulouse), for a budget of 5.5M€ (600k€ for STACK) overall.

SILECS

Participants: Baptiste Jonglez, Remous-Aris Koutsiamanis *(STACK representative)*, Adrien Lebre, Jean-Marc Menaud.

Digital transformation relies on a sophisticated infrastructure of networks, computing and services. The availability, reliability, performance, interoperability and energy efficiency of these systems are major challenges that the digital sciences must meet to foster innovation, sovereignty and industrial competitiveness.

SILECS, the Cloud/Fog/Edge/IoT part of the SLICES-FR platform, enables prototyping and reproducible experiments at all levels of the Cloud IoT continuum. It meets the experimental needs of researchers in networks, systems, telecoms, IoT and other fields. The main objective of SILECS and SLICES-FR is to build a tool for experimentation that fosters the design of new services and applications in distributed computing, edge computing, reprogrammable wired or wireless networks and IoT, using a diversity of technologies on all aspects of the data chain, software or hardware, to meet the needs of the community.

The consortium is composed of 3 partners (Inria, CNRS, and IMT) for a budget of Ĩ2Millions€ (272k€ for STACK) overall .

10.3.2 PIA 4

OTPaaS

Participants: Farid Arfi, Daniel Balouek, Hélène Coullon, Marie Delavergne, Tayeb Diab, Mohamed Graiet, Houssem Jmal, Sidi Mohammed Kaddour, Jabran Khan, Remous-Aris Koutsiamanis, Adrien Lebre *(STACK representative, until Feb 2024)*, Thomas Ledoux, Jean-Marc Menaud, Anas Mokhtari, Clément Mommessin, Jacques Noyé *(STACK representative, from Feb 2024)*, Kandaraj Piamrat, Eloi Perdereau, Matthieu Rakotojaona Rainimangavelo, Mario Südholt.

The OTPaaS project targets the design and development of a complete software stack to administrate and use edge infrastructures for the industry sector. The consortium brings together national and user technology suppliers from major groups (Atos / Bull, Schneider Electric, Valeo) and SMEs / ETIs (Agileo Automation, Mydatamodels, Dupliprint, Solem, Tridimeo, Prosyst, Soben), with a strong support from major French research institutes (CEA, Inria, IMT, CAPTRONIC). The project started in October 2021 for a period of 36 months with an overall budget of 56M€ (1.2M€ for STACK). The project has been extended until April 2025.

The OTPaaS platform objectives are:

- To be built on national and sovereign technologies for the edge cloud.
- To be validated by industrial demonstrators of multisectoral use cases.
- To be followed and supported by ambitious industrialization programs.
- To be accompanied by a massive campaign to promote its use by SMEs / midcaps.
- To integrate solutions for controlling energy consumption.
- To be compliant with the Gaia-X ecosystem.

10.3.3 CPER

SAMURAÏ

Participants: Jean-Marc Menaud *(coordinator)*, Remous-Aris Koutsiamanis.

The SAMURAI (Sustainable And autonoMoUs gReen computing for AI) infrastructure aims to design an innovative hardware infrastructure for the scientific study of the cross-cutting issues of computing infrastructures supporting artificial intelligence and their energy autonomies.

This project paves the way toward a larger infrastructure at the natonial level in the context of the SLICES-FR initiative.

The project started in 2022 for a period of 5 years with an overall budget of 730K€ (500K€ for STACK).

10.3.4 Local and regional projects

SysMics network

Participants: Mario Südholt.

SysMics is an integrated cluster of research that is part of the Nantes Excellence Initiative in Medecine and Engineering. Its main objective is the development of new methods for precision medecine, in particular, based on genomic analyses. In this context, we have worked on new large-scale distributed biomedical analyses and provided several results on how to distributed popular statistical analyses, such as FAMD-based and EM-based analyses.

10.3.5 Inria Challenges

FrugalCloud (Inria-OVHCloud)

Participants: Yasmina Bouziem, Hélène Coullon, Thomas Ledoux, Pierre Jacquet.

A joint collaboration between Inria and OVHcloud company on the challenge of frugal cloud has been launched in October 2021 with a budget of 2 M€. It addresses several scientific challenges on the eco-design of cloud frameworks and services for large scale energy and environmental impact reduction, across three axes: i) Software eco-design of services and applications; ii) Efficiency leverages; iii) Reducing the impact and supporting users of the Cloud.

Pulse (Inria-Qarnot Computing)

Participants: Adrien Lebre.

The joint challenge between Inria and Qarnot computing is called PULSE, for "PUshing Low-carbon Services towards the Edge". It aims to develop and promote best practices in geo-repaired hardware and software infrastructures for more environmentally friendly intensive computing.

The challenge is structured around two complementary research axes to address this technological and environmental issue:

Axis 1: "Holistic analysis of the environmental impact of intensive computing".

Axis 2: "Implementing more virtuous edge services"

The STACK group is mainly involved in the second axis, addressing the challenges related to data management.

10.3.6 Inria transfer projects (ADT)

CoAnsible

Participants: Hélène Coullon, Baptiste Jonglez.

The transfer actions CoAnsible is a two year project to develop an Ansible extension able to use the Concerto language [49] as an efficient backend tasks coordinator to handle the lifecycle of infrastructure resources. A 24-months software engineer is hired starting in December 2024.

10.4 Regional initiatives

10.4.1 A Cloud-Edge-IoT continuum site for SLICES-FR

Participants: Remous-Aris Koutsiamanis, Jean-Marc Menaud, Baptiste Jonglez.

The SLICES-FR site in Nantes, comprising the STACK Research team and the LS2N and IETR UMRs, is advancing the development of a Cloud-Edge-IoT continuum-supporting site. This infrastructure envisions to enable seamless, low-latency experimentation across the continuum, contributing to the SLICES-FR and SLICES-RI platforms.

The current efforts are focused on integrating new hardware and collaborating within the architecture and technical teams to align with the project's vision. The ANR Samurai project has been instrumental in funding hardware acquisition and providing engineering resources for management, integration, and operations, ensuring the sites's robustness and readiness for experimentation.

11 Dissemination

Participants: Daniel Balouek, Hélène Coullon, Carlos Javier Gonzalez, Baptiste Jonglez, Remous-Aris Koutsiamanis, Thomas Ledoux, Adrien Lebre, Jean-Marc Menaud, Jacques Noyé, Kandaraj Piamrat, Guillaume Rosinosky, Mario Südholt.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

General chair, scientific chair

- Daniel Balouek: QUICK 2024 (CCgrid Workshop) - Urgent Computing for the Continuum

Member of the steering committees

- Adrien Lebre IEEE ICFEC (International Conference on Fod and Edge Computing)

Member of the organizing committees

- Adrien Lebre took part to the organisation of the PEPR Cloud kickoff and the first edition of the PEPR Cloud days (respectively in April and Decembre 2024 with more than 100 participants for each event).

11.1.2 Scientific events: selection

Chair of conference program committees

- Daniel Balouek: ICT-DM 2024

Member of the conference program committees

- Daniel Balouek: PAISE (IPDPSW) 2024, SioTEC 2024, SSDBM 2024, UCC 2024
- Adrien Lebre: IEEE IC2E, IEEE CloudCom, IEEE ICFEC
- Kandaraj Piamrat: IEEE ICC (CSM symposium) 2024, IEEE CCNC (IoT: From Sensors to Vertical Applications Track) 2024, IEEE WiMob 2024, IEEE ICCE 2024, IEEE ISCC 2024, IEEE ICC Workshop on Data Driven Intelligence for Networks and Systems (DDINS), Algotel/Cores 2024
- Guillaume Rosinosky: ACM/IFIP International Middleware Conference 2024, The Tenth International Workshop on Container Technologies and Container Clouds 2024, IEEE CloudNet 2024
- Mario Südholt: CloudCom 2024, ICISSP 2024
- Remous-Aris Koutsiamanis: Algotel/Cores 2024, IEEE ISCC 2024, IEEE CSCN 2024, ICCS (International Conference on Computational Science) 2024, Compas 2024

11.1.3 Journal

Member of the editorial boards

- Mario Südholt is a member of the advisory board of ”The Programming Journal”

Reviewer - reviewing activities

- Kandaraj Piamrat: IEEE Transactions on Network and Service Management, Elsevier Computer Network (2), Adhoc Networks, Internet of Things (3), Computer Science Review.
- Carlos Javier Gonzalez: IEEE Transactions on Computers, IEEE Access, IEEE Sensors Letters, IEEE Transactions on Mobile Computing, Cyber-Physical Systems, IETE Journal of Research.
- Daniel Balouek: IEEE Transactions on Parallel and Distributed Systems, Future Generation Computer Systems, IEEE Transactions on Sustainable Computing, IEEE Computing
- Thomas Ledoux : IEEE Internet Computing
- Remous-Aris Koutsiamanis: IEEE Transactions on Network and Service Management, IEEE Transactions on Parallel and Distributed Systems, Wiley Transactions on Emerging Telecommunications Technologies, Elsevier Internet of Things

11.1.4 Invited talks

- Daniel Balouek: Enabling Urgent Science across the Edge-Cloud-HPC Continuum, ISC HPC, Hamburg, May 2024
- Daniel Balouek: Engineering the Computing Continuum for (Urgent) Science, ICIT 2024 Keynote, Islamabad, Pakistan, October 2024
- Remous-Aris Koutsiamanis: "Maîtrise énergétique des centres de données", Summer school Green IT 2024
- Adrien Lebre: Cheops, How to Blow Clouds to the Edge, keynote at CEA List days, Paris, France, June 2024

- Adrien Lebre: Blowing Cloud Computing to the Edge and Beyond, CISIRH (Centre Interministériel de Services Informatiques relatifs aux Ressources Humaines), Paris, France, Oct 2024.
- Thomas Ledoux: "Intégrer les enjeux Numérique responsable dans l'enseignement supérieur", Summer school Green IT 2024
- Kandaraj Piamrat: AI-enhanced Intrusion Detection System for Cloud-Edge-IoT continuum, JFLI seminar, NII, 20 November 2024

11.1.5 Scientific expertise

Remous-Aris Koutsiamanis has acted as a reviewer for the Scientific Council of CEFIPRA (Funding Agency for India and French Research Fostering).

11.1.6 Research administration

- Kandaraj Piamrat is the international coordinator of the LS2N (Laboratoire des sciences du numérique de Nantes).
- Hélène Coullon is the vice-president of the ACM SigOps France
- Hélène Coullon is co-chair of the working group YODA (trustworthY and Optimal Dynamic Adaptation) in the national GDR GPL (software engineering and languages)
- Remous-Aris Koutsiamanis is co-chair of the IETF ROLL (Routing Over Low power and Lossy networks) Working Group
- Remous-Aris Koutsiamanis : Member of the executive committee of the Grid'5000 GIS (Groupement d'intérêt scientifique), Member of the SLICES-FR temporary executive board, Member of the SLICES-FR Architects committee, Member of the SLICES-FR Site managers, Scientific representative of IMT to SLICES.
- Adrien Lebre: Member of the executive committee of the Grid'5000 GIS (Groupement d'intérêt scientifique), Member of the SLICES-FR temporary executive board, Scientific French representative of the SLICES ISB, Co-director of the <I/O> Lab, a joint lab between Inria and Orange Labs, co-leader the French CLOUD PEPR, and since December 2023, head of the Cloud, Network and System program at the French agency "Algorithmes, Logiciels et Usages" operated by Inria.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

As a team mainly composed of Associate Prof. and Prof., the amount of teaching activities is significant (around 150hours per person). We present here only the management activities.

- Hélène Coullon: responsible for the LOGIN training in computer science at IMT Atlantique (last year)
- Guillaume Rosinosky: head of the apprenticeship program in Software Engineering FIL. This 3-year program leads to the award of a Master degree in Software Engineering from the IMT Atlantique.
- Thomas Ledoux: member of the teaching committee at IMT Atlantique
- Thomas Ledoux: co-pilot of the core curriculum of the 1st year of the IT program at IMT Atlantique
- Thomas Ledoux: head of the Filière informatique nantaise since Sept. 2020. This entity, created by Nantes University, Centrale Nantes and IMT Atlantique, aims to bring together the main players in Computer Science training in Nantes to ensure a coherent and ambitious training offer that meets the present and future challenges of Computer Science. It is organized around a Council made up of representatives from the academic and socio-economic worlds.

- Thomas Ledoux : member of the board of directors of Talents du numérique.
- Jacques Noyé: deputy head of the Automation, Production and Computer Sciences Department of IMT Atlantique.
- Mario Südholt: representative for MSc-level and PhD-level studies of the API department of IMT Atlantique.
- Mario Südholt: coordinator of the international PhD program SEED (see 10).
- Kandaraj Piamrat : elected member of scientific council at Faculty of Sciences and Techniques, Nantes University

11.2.2 Supervision

- PhD in progress: Duc Thinh Ngo, Dynamic graph learning algorithms for the digital twin in edge-cloud continuum, Dec.2022 - Nov.2025, Director: A. Lebre, Advisor:: K. Piamrat
- PhD in progress: Houssem Jmal, Federated Learning for Enhancing Security and Privacy of Decentralized and Distributed Systems, Apr.2024 - Mar.2027, Director: JM. Menaud, Advisor:: K. Piamrat
- PhD in progress: Martin Molli, Decision models for the Edge-Cloud Computing Continuum, Nov.2024 - Sept.2027, Director: T. Ledoux, Advisor:: D. Balouek
- PhD in progress: Lylian Siffre, Impacts and Uses of Local-First Software for Energy Optimization of IT Services, Nov.2024 - Sept.2027, Director: T. Ledoux
- PhD in progress: Antoine Omond, Safe, efficient and low-energy self-adaptation for Cyber Physical Systems - Application to a scientific observatory in the Arctic tundra, Dec. 2021 - Nov. 2024 (extended to May 2025), Director: T. Ledoux, Advisor:: H. Coullon
- PhD in progress: Tengfei An, Modeling and Studying Self-Stabilization within Kubernetes, Oct. 2023 - Sept. 2026, Director: A. Lebre, Advisor: H. Coullon, J. Noyé
- PhD in progress: Samia Boutalbi, Secure deployment of microsservices in a shared RAN/MEC Cloud environment, Jan 2022 - Jan 2025, Director: M. Südholt, Coen De Roover, Advisor: R.-A. Koutsiamanis.
- PhD in progress: Mohammed Abdrrahim Lahmar, Contracts for Distributed ML-Intensive Systems, Oct 2024 - Oct 2027, Director: M. Südholt, Advisor: R.-A. Koutsiamanis.
- PhD in progress: Abdou Seck, Parallel transfer service for the exchange of large volumes of data between Datacenters, June 2022 - June 2025, Director: J.-M. Menaud, Noel De Palma, Advisor: R.-A. Koutsiamanis.
- PhD in progress: Divi Delacour, Architecture and security of cooperative autonomous systems, Jan 2022 - Jan 2025, Director: M. Südholt, Marc Lacoste (Orange).
- PhD in progress: Cherif Si Mohammed, Eco-responsible management of data storage, Dec 2023, Dec 2026, Director: A. Lebre, Advisor: A. Van Kempen (Qarnto Computing)
- PhD in progress: Hiba Awad, A Model-based Approach for Multi-Scale and Dynamic Distributed Systems, Nov. 2021 - March 2025. Director: T. Ledoux.
- PhD in progress: Hiba Awad, A Model-based Approach for Multi-Scale and Dynamic Distributed Systems, Nov. 2021 - March 2025. Director: T. Ledoux.
- PhD in progress: Wedwang Romial Menra, Deployment and updating electronic equipment, Nov. 2024 - Nov. 2027. Director: J.-M. Menaud

- PhD defended: Pierre Jacquet, Enhancing IaaS Consolidation with Resource Oversubscription, July 2024, Director: Romain Rouvoy, T. Ledoux
- PhD defended:Geo Johns Antony, Cheops reloaded : Further steps in Decoupling Geo-Distribution from Application business logic, Dec 2024, Director: A. Lebre
- Post-doc: Jabran Khan, Placement in Cloud, Edge and Far Edge devices in the IoT to Cloud continuum, Feb. 2023 - Aug. 2024, Advisor: R.-A. Koutsiamanis, J.-M. Menaud
- Post-doc: Clément Mommessin, Model of Energy consumption in Cloud, Edge and Far Edge devices in the IoT to Cloud continuum, Jan. 2023 - Jun. 2024, Advisor: J.-M. Menaud, R.-A. Koutsiamanis
- Post-doc: Anas Mokhtari, A Holistic Approach for Designing Carbon-Aware and Energy-Aware Cloud applications, Sept. 2023 - Janv. 2025, Advisor: T. Ledoux, B. Jonglez.
- Post-doc: Tayeb Diab, Generic Transfer Learning, March. 2023 - July 2024, Advisor: M. Südholt.

11.2.3 Juries

- Kandaraj Piamrat was a member of the selection committee for an associate professor positions at University of Rennes.
- Kandaraj Piamrat was an examiner of the PhD defense of Menuka PERERA JAYASURIYA KURANAGE at IMT Atlantique - Rennes - Solutions du type Zero-Touch basées sur l'IA pour la gestion des ressources dans les réseaux 5G cloud-native, 18 November 2024.
- Kandaraj Piamrat was a jury member of the PhD defense of Sid Ali Hamideche at University of Rennes - Automatisation au-delà de la 5G : Apprentissage Automatique appliqué au profilage du comportement de l'utilisateur de services mobiles, 17 December 2024.
- Hélène Coullon was a member of the selection committee for a temporary associate professor position at the University of Orléans, France.
- Thomas Ledoux was a member of the selection committee for an associate professor positions at University of Rennes and INSA Rennes.
- Thomas Ledoux was a jury member of the PhD defense of Benjamin Somers at IMT Atlantique - IT infrastructure modeling for risk identification and prevention, may 2024.
- Thomas Ledoux was reviewer of the PhD of Vincent Lannurien at ENSTA Bretagne- Allocation et placement dynamiques sur ressources hétérogènes pour le cloud serverless, nov. 2024.
- Thomas Ledoux was reviewer of the PhD of Houssam Hajj Hassan, Télécom SudParis - Enabling Autonomous IoT Systems: A Middleware-based Hybrid AI Approach to Self-adaptation, dec. 2024.

11.3 Popularization

11.3.1 Education

- Thomas Ledoux: participated in the creation of the RNCP certification "Responsable Green IT" by France Compétences at the invitation of Numeum.

12 Scientific production

12.1 Publications of the year

International journals

[1] ***Best Paper***
O. Aouedi, L. van An, K. Piamrat and J. Yusheng. 'Deep Learning on Network Traffic Prediction: Recent Advances, Analysis, and Future Directions'. In: *ACM Computing Surveys* (12th Oct. 2024), pp. 1–35. URL: https://hal.science/hal-04811081 (cit. on pp. 6, 11, 17).

[2] O. Aouedi and K. Piamrat. 'Toward a Scalable and Energy-Efficient Framework for Industrial Cloud-Edge-IoT Continuum'. In: *IEEE Internet of Things Magazine* 7.5 (Sept. 2024), pp. 14–20. DOI: 10.1109/IOTM.001.2300229. URL: https://hal.science/hal-04811034 (cit. on pp. 6, 17).

[3] ***Best Paper***
O. Aouedi, T.-H. Vu, A. Sacco, D. Nguyen, K. Piamrat, G. Marchetto and Q.-V. Pham. 'A Survey on Intelligent Internet of Things: Applications, Security, Privacy, and Future Directions'. In: *Communications Surveys and Tutorials, IEEE Communications Society* (2024), pp. 1–1. DOI: 10.1109/COMST.2024.3430368. URL: https://hal.science/hal-04811049 (cit. on pp. 11, 20).

[4] H. Awad, A. Alidra, H. Bruneliere, T. Ledoux and J. Rivalan. 'VeriFog: A Generic Model-based Approach for Verifying Fog Systems at Design Time and Generating Deployment Configurations'. In: *ACM SIGAPP applied computing review : a publication of the Special Interest Group on Applied Computing* 24.3 (8th Oct. 2024), pp. 18–36. DOI: 10.1145/3699839.3699841. URL: https://hal.science/hal-04727206 (cit. on pp. 12, 18, 22).

[5] S. Brocard, M. Morin, N. dos Santos Brito Silva, B. Renaud-Picard, B. Coiffard, X. Demant, L. Falque, J. Le Pavec, A. Roux, T. Villeneuve, C. Knoop, J.-F. Mornex, M. Salpin, V. Boussaud, O. Rousseau, V. Mauduit, A. Durand, A. Magnan, P.-A. Gourraud, N. Vince, M. Südholt, A. Tissot and S. Limou. 'Description and first insights on a large genomic biobank of lung transplantation'. In: *European Journal of Human Genetics* (20th Aug. 2024), pp. 1–8. DOI: 10.1038/s41431-024-01683-y. URL: https://inserm.hal.science/inserm-04712234 (cit. on p. 18).

[6] H. Coullon, L. Henrio, F. Loulergue and S. Robillard. 'Component-Based Distributed Software Reconfiguration: a Verification-Oriented Survey'. In: *ACM Computing Surveys* 56.1 (Jan. 2024), pp. 1–37. DOI: 10.1145/3595376. URL: https://inria.hal.science/hal-04067909 (cit. on pp. 6, 18).

[7] P. Jacquet, T. Ledoux and R. Rouvoy. 'ScroogeVM: Boosting Cloud Resource Utilization with Dynamic Oversubscription'. In: *IEEE Transactions on Sustainable Computing* (2024), pp. 1–13. DOI: 10.1109/TSUSC.2024.3369333. URL: https://hal.science/hal-04466538 (cit. on pp. 17, 22).

[8] C. Mommessin, T. Erlebach and N. Shakhlevich. 'Classification and evaluation of the algorithms for vector bin packing'. In: *Computers and Operations Research* 173 (Jan. 2025), p. 106860. DOI: 10.1016/j.cor.2024.106860. URL: https://inria.hal.science/hal-04769128.

[9] N. Pavlidis, V. Perifanis, S. Yilmaz, F. Wilhelmi, M. Miozzo, P. Efraimidis, R.-A. Koutsiamanis, P. Mulinka and P. Dini. 'Federated Learning in Mobile Networks: A Comprehensive Case Study on Traffic Forecasting'. In: *IEEE Transactions on Sustainable Computing* 2377-3782 (21st Nov. 2024), pp. 1–12. DOI: 10.1109/TSUSC.2024.3504242. URL: https://hal.science/hal-04885759 (cit. on p. 20).

International peer-reviewed conferences

[10] G. J. Antony, M. Delavergne, A. Lebre and M. Rakotojaona Rainimangavelo. 'Thinking out of replication for geo-distributing applications: the sharding case'. In: ICFEC 2024 - 8th IEEE International Conference on Fog and Edge Computing. Philadelphia, United States, 2024, pp. 1–8. URL: https://inria.hal.science/hal-04522961 (cit. on p. 19).

[11] O. Aouedi, G. Jajoo and K. Piamrat. 'METALS : seMi-supervised fEderaTed Active Learning for intrusion detection Systems'. In: ISCC 2024 - 29th IEEE Symposium on Computers and Communications. Paris, France: IEEE, 2024, pp. 1–5. DOI: 10.1109/ISCC61673.2024.10733565. URL: https://hal.science/hal-04811225 (cit. on p. 6).

[12] O. Aouedi and K. Piamrat. 'SURFS: Sustainable IntrUsion Detection with HieraRchical Federated Spiking Neural Networks'. In: ICC 2024 - IEEE International Conference on Communications. Denver, United States: IEEE, 2024, pp. 2173–2178. DOI: 10.1109/ICC51166.2024.10622560. URL: https://hal.science/hal-04811219 (cit. on p. 6).

[13] F. Arfi, H. Coullon, F. Loulergue, J. Philippe and S. Robillard. 'An Overview of the Decentralized Reconfiguration Language Concerto-D through its Maude Formalization'. In: ICE 2024 - 17th Interaction and Concurrency Experience. Electronic Proceedings in Theoretical Computer Science. Groningen, Netherlands, 2024, pp. 1–18. URL: https://inria.hal.science/hal-04572043 (cit. on p. 19).

[14] ***Best Paper***
H. Awad, A. Alidra, H. Bruneliere, T. Ledoux, E. Leclerq and J. Rivalan. 'VeriFog: A Generic Model-based Approach for Verifying Fog Systems at Design Time'. In: *SAC '24: Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*. SAC '24: Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing. Avila, Spain: ACM, 2024, pp. 1252–1261. DOI: 10.1145/3605098.3635973. URL: https://hal.science/hal-04332046 (cit. on pp. 18, 22).

[15] H. Awad, T. Ledoux, H. Bruneliere and J. Rivalan. 'VeriFogOps: Automated Deployment Tool Selection and CI/CD Pipeline Generation for Verifying Fog Systems at Deployment Time'. In: *SAC '25: Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing*. SAC '25: Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing. Catania, Italy, 31st Mar. 2025. DOI: 10.1145/3672608.3707854. URL: https://hal.science/hal-04833623 (cit. on p. 22).

[16] D. Balouek. 'Performance-cost trade-offs in service orchestration for edge computing'. In: SSDBM 2024 - 36th International Conference on Scientific and Statistical Database Management. Rennes, France: ACM, 10th July 2024, pp. 1–4. DOI: 10.1145/3676288.3676307. URL: https://hal.science/hal-04775133 (cit. on p. 17).

[17] D. Balouek and J. Pettré. 'Collaborative services for Crowd Safety systems across the Edge-Cloud Computing Continuum'. In: CCGridW 2024 - IEEE 24th International Symposium on Cluster, Cloud and Internet Computing Workshops. Philadelphia, France: IEEE, 6th May 2024, pp. 92–97. DOI: 10.1109/ccgridw63211.2024.00017. URL: https://hal.science/hal-04775145 (cit. on p. 17).

[18] D. De Lacour, M. A. Lacoste, M. Südholt and J. Traoré. 'Ti-skol: A Modular Federated Learning Framework Supporting Security Countermeasure Composition'. In: Big Data 2024 - Special Session on Federated Learning on Big Data. Washington DC, United States, 2024. URL: https://hal.science/hal-04760901 (cit. on p. 21).

[19] Y. G. Hellal, L. Hamel, M. Graiet and D. Balouek. 'A Formal Modeling and Verification Approach for IoT-Cloud Resource-Oriented Applications'. In: CCGrid 2024 - IEEE 24th International Symposium on Cluster, Cloud and Internet Computing. Philadelphia, United States: IEEE, 6th May 2024, pp. 347–356. DOI: 10.1109/CCGrid59990.2024.00047. URL: https://hal.science/hal-04775142 (cit. on p. 18).

[20] P. Jacquet, T. Ledoux and R. Rouvoy. 'SLACKVM: Packing Virtual Machines in Oversubscribed Cloud Infrastructures'. In: 2024 CLUSTER - IEEE International Conference on Cluster Computing. Vol. Proceedings of the IEEE International Conference on Cluster Computing (CLUSTER). Kobe, Japan: IEEE, 2024, pp. 1–12. DOI: 10.1109/CLUSTER59578.2024.00024. URL: https://hal.science/hal-04636648 (cit. on pp. 17, 22).

[21] P. Jacquet, T. Ledoux and R. Rouvoy. 'SweetspotVM: Oversubscribing CPU without Sacrificing VM Performance'. In: *Proceedings of the 24th IEEE/ACM international Symposium on Cluster, Cloud and Internet Computing (CCGrid'24)*. CCGrid'24 - 24th IEEE/ACM international Symposium on Cluster, Cloud and Internet Computing. Philadelphia, United States: IEEE, 2024, pp. 1–10. DOI: 10.1109/CCGrid59990.2024.00026. URL: https://hal.science/hal-04454043 (cit. on pp. 17, 22).

[22] H. Jmal, K. Piamrat and O. Aouedi. 'TUNE-FL: adapTive semi-synchronoUs semi-deceNtralizEd Federated Learning'. In: IEEE Consumer Communications & Networking Conference (CCNC). Las Vegas (Nevada), United States, 10th Jan. 2025. URL: https://hal.science/hal-04752941 (cit. on p. 21).

[23] A. Mokhtari, B. Jonglez and T. Ledoux. 'Towards Digital Sustainability: Involving Cloud Users as Key Players'. In: IC2E 2024 - 12th IEEE International Conference on Cloud Engineering. Paphos, Cyprus: IEEE; IEEE, 2024, pp. 126–132. DOI: 10.1109/IC2E61754.2024.00021. URL: https://hal.science/hal-04633237 (cit. on p. 20).

[24] J. Philippe, A. Omond, H. Coullon, C. Prud'Homme and I. Raïs. 'Fast Choreography of Cross-DevOps Reconfiguration with Ballet: A Multi-Site OpenStack Case Study'. In: SANER 2024 - IEEE International Conference on Software Analysis, Evolution and Reengineering. Rovaniemi, Finland: IEEE; IEEE, 2024, pp. 1–11. DOI: 10.1109/SANER60148.2024.00007. URL: https://hal.science/hal-04457484 (cit. on pp. 19, 26).

[25] S. Pour, D. Balouek, A. Masoumi and M. Brynskov. 'An urgent computing oriented architecture for dynamic climate risk management framework'. In: *Proceedings of the IEEE/ACM 16th International Conference on Utility and Cloud Computing*. UCC '23: IEEE/ACM 16th International Conference on Utility and Cloud Computing. Taormina, Italy, 4th Apr. 2024, pp. 1–4. DOI: 10.1145/3603166.3632539. URL: https://inria.hal.science/hal-04387718.

[26] G. Rosinosky, D. Schmitz and E. Rivière. 'StreamBed: Capacity Planning for Stream Processing'. In: DEBS 2024 - 18th ACM International Conference on Distributed and Event-based Systems. Lyon, France: ACM, 22nd Nov. 2024, pp. 90–102. DOI: 10.1145/3629104.3666034. URL: https://hal.science/hal-04708354 (cit. on p. 20).

[27] C. Sicari, D. Balouek, M. Villari and M. Parashar. 'Event-Driven FaaS Workflows for Enabling IoT Data Processing at the Cloud Edge Continuum'. In: *Proceedings of the IEEE/ACM 16th International Conference on Utility and Cloud Computing*. UCC 2023: 16th International Conference on Utility and Cloud Computing. Taormina, Italy, 4th Apr. 2024, pp. 1–10. DOI: 10.1145/3603166.3632125. URL: https://inria.hal.science/hal-04387675.

[28] Y. Xia, X. Etchevers, L. Letondeur, T. Coupaye and F. Desprez. 'Optimizing Cloud Application Scheduling: A Dual-Stage Heuristic Approach'. In: ICCCBDA 2024 - 9th International Conference on Cloud Computing and Big Data Analytics. Chengdu, China: IEEE, 28th June 2024, pp. 134–140. DOI: 10.1109/ICCCBDA61447.2024.10569682. URL: https://hal.science/hal-04886644 (cit. on p. 20).

National peer-reviewed Conferences

[29] R.-A. Koutsiamanis, J.-M. Menaud and C. Mommessin. 'Ordonnancement de Tâches avec Précédences sur Plateforme Edge-Cloud Alimentée par des Sources d'Energie Verte'. In: ROADEF 2024 - 25ème congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision. Amiens, France, 4th Mar. 2024. URL: https://hal.science/hal-04885928 (cit. on p. 20).

Doctoral dissertations and habilitation theses

[30] P. Jacquet. 'Enhancing IaaS Consolidation with Resource Oversubscription'. Université de Lille, 19th July 2024. URL: https://theses.hal.science/tel-04685771 (cit. on pp. 17, 22).

Reports & preprints

[31] G. J. Antony, M. Delavergne and M. Rakotojaona Rainimangavelo. *An update on Cheops: new implementation report*. RT-0524. Inria, Oct. 2024. URL: https://inria.hal.science/hal-04886168.

[32] D. De Lacour. *D-ECS: Towards decentralising video games*. 14th Jan. 2025. URL: https://hal.science/hal-04886531.

[33] D. De Lacour. *D-Lambda: a fully decentralised serverless model*. 2024. URL: https://hal.science/hal-04805193.

[34] D. De Lacour, M. Lacoste, M. Südholt and J. Traoré. *Towards Volunteer Deep Learning: Security Challenges and Solutions*. 10th Jan. 2025. URL: https://hal.science/hal-04879559.

[35] D.-T. Ngo, O. Aouedi, K. Piamrat, T. Hassan and P. Raipin-Parvédy. *Graph Neural Network-based Models for Mobile Network Traffic Prediction*. 2024. URL: https://hal.science/hal-04588238.

[36] D.-T. Ngo, K. Piamrat, O. Aouedi, T. Hassan and P. Raipin. *FLEXIBLE: Forecasting Cellular Traffic by Leveraging Explicit Inductive Graph-Based Learning*. 2024. DOI: 10.1109/PIMRC59610.2024.10817182. URL: https://hal.science/hal-04573168 (cit. on pp. 6, 17).

Other scientific publications

[37] E. Abisset-Chavanne, T. Coupaye, F. Golra, D. Lamy, A. Piel, O. Scart and P. Vicat-Blanc. 'A Digital Twin use cases classification and definition framework based on Industrial feedback'. In: *Computers in Industry* 161 (2024), p. 104113. DOI: 10.1016/j.compind.2024.104113. URL: https://cea.hal.science/cea-04607213.

[38] M. I. Ghamri, M. Lacoste and D. De Lacour. 'Disaggregation Patterns for Secure AI Systems'. In: HPEC 2024 - 28th Annual IEEE High Performance Extreme Computing Virtual Conference. Boston (MA), United States, 2024, pp. 1–2. URL: https://inria.hal.science/hal-04717976.

Software

[39] [SW] F. Arfi, H. Coullon, F. Loulergue, J. Philippe and S. Robillard, *Concerto-D in Maude* version ice24-final-version, 20th July 2024. LIC: GNU General Public License v3.0 or later. DOI: 10.5281/zenodo.12786127, HAL: ⟨hal-04654817⟩, URL: https://hal.science/hal-04654817, VCS: https://github.com/Concerto-D/concertod-maude/tree/ice24-final-version, SWHID: ⟨swh:1:dir:32a2fe14ad750c01d73d2bd101ef0c2f2e440f62;origin=https://hal.archives-ouvertes.fr/hal-04654817;visit=swh:1:snp:ad21f1e90f6cf2549821eb2142ab6657817241f1;anchor=swh:1:rel:9689c8ed0577d56bfb7885c6ed94e9ce0020eca3;path=/⟩.

12.2 Cited publications

[40] N. Ayeb, É. Rutten, S. Bolle, T. Coupaye and M. Douet. 'Coordinated autonomic loops for target identification, load and error-aware Device Management for the IoT'. In: *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems, FedCSIS 2020, Sofia, Bulgaria, September 6-9, 2020*. Ed. by M. Ganzha, L. A. Maciaszek and M. Paprzycki. Vol. 21. Annals of Computer Science and Information Systems. 2020, pp. 491–500. DOI: 10.15439/2020F154. URL: https://doi.org/10.15439/2020F154 (cit. on p. 6).

[41] N. Ayeb, É. Rutten, S. Bolle, T. Coupaye and M. Douet. 'Towards an Autonomic and Distributed Device Management for the Internet of Things'. In: *IEEE 4th International Workshops on Foundations and Applications of Self* Systems, FAS*W@SASO/ICCAC 2019, Umea, Sweden, June 16-20, 2019*. IEEE, 2019, pp. 246–248. DOI: 10.1109/FAS-W.2019.00065. URL: https://doi.org/10.1109/FAS-W.2019.00065 (cit. on p. 6).

[42] N. Bakerally, C. Bareau, F. Blache, S. Bolle, C. Ecrepont, P. Folz, N. Hernandez, T. Monteil, G. Privat and F. Ramparany. 'Semi-automatic RDFization Using Automatically Generated Mappings'. In: *The Semantic Web: ESWC 2020 Satellite Events - ESWC 2020 Satellite Events, Heraklion, Crete, Greece, May 31 - June 4, 2020, Revised Selected Papers*. Ed. by A. Harth, V. Presutti, R. Troncy, M. Acosta, A. Polleres, J. D. Fernández, J. X. Parreira, O. Hartig, K. Hose and M. Cochez. Vol. 12124. Lecture Notes in Computer Science. Springer, 2020, pp. 25–31. DOI: 10.1007/978-3-030-62327-2_5. URL: https://doi.org/10.1007/978-3-030-62327-2%5C_5 (cit. on p. 6).

[43] D. Balouek, A. C. Amarie, G. Charrier, F. Desprez, E. Jeannot, E. Jeanvoine, A. Lèbre, D. Margery, N. Niclausse, L. Nussbaum et al. 'Adding virtualization capabilities to the Grid'5000 testbed'. In: *International Conference on Cloud Computing and Services Science*. Springer. 2012, pp. 3–20 (cit. on p. 7).

[44] M. Belair, S. Laniepce and J.-M. Menaud. 'SNAPPY: Programmable Kernel-Level Policies for Containers'. In: *SAC 2021: 36th ACM/SIGAPP Symposium On Applied Computing*. Gwangju / Virtual, South Korea, Mar. 2021, pp. 1636–1645. DOI: 10.1145/3412841.3442037. URL: https://hal.inria.fr/hal-03108231 (cit. on p. 9).

[45] G. Blair, T. Coupaye and J.-B. Stefani. 'Component-based architecture: the Fractal initiative'. In: *Annals of telecommunications* 64.1 (Feb. 2009), pp. 1–4. DOI: 10.1007/s12243-009-0086-1. URL: https://doi.org/10.1007/s12243-009-0086-1 (cit. on p. 6).

[46] H. Bruneliere, Z. Al-Shara, F. Alvares, J. Lejeune and T. Ledoux. 'A Model-based Architecture for Autonomic and Heterogeneous Cloud Systems'. In: *CLOSER 2018 - 8h International Conference on Cloud Computing and Services Science*. Vol. 1. Best Paper Award. Funchal, Portugal, Mar. 2018, pp. 201–212. DOI: 10.5220/0006773002010212. URL: https://hal.archives-ouvertes.fr/hal-01705248 (cit. on p. 6).

[47] A. Bruniaux, R.-A. Koutsiamanis, G. Z. Papadopoulos and N. Montavont. 'Defragmenting the 6LoWPAN Fragmentation Landscape: A Performance Evaluation'. In: *Sensors* 21.5 (2021), p. 1711 (cit. on p. 6).

[48] M. Chardet, H. Coullon, D. Pertin and C. Pérez. 'Madeus: a formal deployment model'. In: *2018 International Conference on High Performance Computing & Simulation (HPCS)*. IEEE. 2018, pp. 724–731 (cit. on p. 6).

[49] M. Chardet, H. Coullon and S. Robillard. 'Toward Safe and Efficient Reconfiguration with Concerto'. In: *Science of Computer Programming* 203 (Mar. 2021), pp. 1–31. DOI: 10.1016/j.scico.2020.102582. URL: https://hal.inria.fr/hal-03103714 (cit. on pp. 6, 8, 18, 30).

[50] M. Charlier, R.-A. Koutsiamanis and B. Quoitin. 'Scheduling UWB Ranging and Backbone Communications in a Pure Wireless Indoor Positioning System'. In: *IoT* 3.1 (2022), pp. 219–258 (cit. on p. 6).

[51] R.-A. Cherrueau, M. Delavergne, A. van Kempen, A. Lebre, D. Pertin, J. Rojas Balderrama, A. Simonet and M. Simonin. 'EnosLib: A Library for Experiment-Driven Research in Distributed Computing'. In: *IEEE Transactions on Parallel and Distributed Systems* 33.6 (June 2022), pp. 1464–1477. DOI: 10.1109/TPDS.2021.3111159. URL: https://inria.hal.science/hal-03324177 (cit. on p. 7).

[52] R.-A. Cherrueau, D. Pertin, A. Simonet, A. Lebre and M. Simonin. 'Toward a Holistic Framework for Conducting Scientific Evaluations of OpenStack'. In: *Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*. IEEE Press. 2017, pp. 544–548 (cit. on p. 7).

[53] M. Clavel, F. Durán, S. Eker, P. Lincoln, N. Martí-Oliet, J. Meseguer and J. F. Quesada. 'The Maude System'. In: *Proceedings of the 10th International Conference on Rewriting Techniques and Applications*. RtA '99. Berlin, Heidelberg: Springer-Verlag, 1999 (cit. on p. 19).

[54] R. D. Cosmo, J. Mauro, S. Zacchiroli and G. Zavattaro. 'Aeolus: A component model for the cloud'. In: *Information and Computation* 239.Supplement C (2014), pp. 100–121. DOI: https://doi.org/10.1016/j.ic.2014.11.002. URL: http://www.sciencedirect.com/science/article/pii/S0890540114001424 (cit. on p. 18).

[55] H. Coullon, J. Bigot and C. Pérez. 'Extensibility and Composability of a Multi-Stencil Domain Specific Framework'. In: *International Journal of Parallel Programming* (Nov. 2017). DOI: 10.1007/s10766-017-0539-5. URL: https://doi.org/10.1007/s10766-017-0539-5 (cit. on p. 6).

[56] T. Coupaye, S. Bolle, S. Derrien, P. Folz, P. Meye, G. Privat and P. R. Parvedy. 'A graph-based cross-vertical digital twin platform for complex cyber-physical systems'. In: *The Digital Twin*. Ed. by R. M. Noel Crespi Adam T. Drobot. In press, to appear. Springer, Cham, 2022 (cit. on p. 6).

[57] T. Coupaye and J. Stefani. 'Fractal Component-Based Software Engineering'. In: *Object-Oriented Technology, ECOOP 2006 Workshop Reader, ECOOP 2006 Workshops, Nantes, France, July 3-7, 2006, Final Reports*. Ed. by M. Südholt and C. Consel. Vol. 4379. Lecture Notes in Computer Science. Springer, 2006, pp. 117–129. DOI: 10.1007/978-3-540-71774-4_13. URL: https://doi.org/10.1007/978-3-540-71774-4%5C_13 (cit. on p. 6).

[58] P.-C. David, T. Ledoux, M. Léger and T. Coupaye. 'FPath and FScript: Language support for navigation and reliable reconfiguration of Fractal architectures'. In: *annals of telecommunications - annales des télécommunications* 64.1 (Feb. 2009), pp. 45–63. DOI: 10.1007/s12243-008-0073-y. URL: https://doi.org/10.1007/s12243-008-0073-y (cit. on p. 6).

[59] F. A. De Oliveira, T. Ledoux and R. Sharrock. 'A framework for the coordination of multiple autonomic managers in cloud environments'. In: *Self-Adaptive and Self-Organizing Systems (SASO), 2013 IEEE 7th International Conference on*. IEEE. 2013, pp. 179–188 (cit. on p. 6).

[60] X. Etchevers, G. Salaün, F. Boyer, T. Coupaye and N. D. Palma. 'Reliable self-deployment of cloud applications'. In: *Symposium on Applied Computing, SAC 2014, Gyeongju, Republic of Korea - March 24 - 28, 2014*. Ed. by Y. Cho, S. Y. Shin, S. Kim, C. Hung and J. Hong. ACM, 2014, pp. 1331–1338. DOI: 10.1145/2554850.2554951. URL: https://doi.org/10.1145/2554850.2554951 (cit. on p. 6).

[61] X. Etchevers, G. Salaün, F. Boyer, T. Coupaye and N. D. Palma. 'Reliable self-deployment of distributed cloud applications'. In: *Softw. Pract. Exp.* 47.1 (2017), pp. 3–20. DOI: 10.1002/spe.2400. URL: https://doi.org/10.1002/spe.2400 (cit. on p. 6).

[62] V. Gasiūnas, L. Satabin, M. Mezini, A. Núñez and J. Noyé. 'EScala: Modular Event-Driven Object Interactions in Scala'. In: *Proceedings of the 10th International Conference on Aspect-Oriented Software Development, AOSD 2011*. Porto de Galinhas, Pernambuco, Brazil: ACM Press, Mar. 2011. URL: http://doi.acm.org/10.1145/1960275.1960303 (cit. on p. 6).

[63] C. Gottstein, P. R. Parvedy, M. Hurfin, T. Hassan and T. Coupaye. 'Inverse Space Filling Curve Partitioning Applied to Wide Area Graphs'. In: *DMS 2020 - 11th International conference on Database Management Systems*. Zurich, Switzerland, Nov. 2020, pp. 223–241. DOI: 10.5121/csit.2020.101417. URL: https://hal.inria.fr/hal-03137952 (cit. on p. 6).

[64] C. Gottstein, P. R. Parvedy, M. Hurfin, T. Hassan and T. Coupaye. 'Partitioning Wide Area Graphs Using a Space Filling Curve'. In: *International Journal of Data Mining & Knowledge Management Process* 11.1 (Jan. 2021), pp. 13–31. DOI: 10.5121/ijdkp.2021.11102. URL: https://hal.inria.fr/hal-03513456 (cit. on p. 6).

[65] D. Hauweele, R.-A. Koutsiamanis, B. Quoitin and G. Z. Papadopoulos. 'Thorough performance evaluation & analysis of the 6TiSCH minimal scheduling function (MSF)'. In: *Journal of Signal Processing Systems* 94.1 (2022), pp. 3–25 (cit. on p. 6).

[66] F. Hermenier, X. Lorca, J.-M. Menaud, G. Muller and J. Lawall. 'Entropy: a consolidation manager for clusters'. In: *Proceedings of the 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments*. ACM. 2009, pp. 41–50 (cit. on p. 6).

[67] T. L. Jenschke, R.-A. Koutsiamanis, G. Z. Papadopoulos and N. Montavont. 'ODeSe: On-Demand Selection for multi-path RPL networks'. In: *Ad Hoc Networks* 114 (2021), p. 102431 (cit. on p. 6).

[68] J. Kephart and D. Chess. 'The vision of autonomic computing'. In: *Computer* 36.1 (2003), pp. 41–50. DOI: 10.1109/MC.2003.1160055 (cit. on p. 7).

[69] R.-A. Koutsiamanis, G. Z. Papadopoulos, B. Quoitin and N. Montavont. 'A Centralized Controller for Reliable and Available Wireless Schedules in Industrial Networks'. In: *2020 16th International Conference on Mobility, Sensing and Networking (MSN)*. IEEE. 2020, pp. 1–9 (cit. on p. 6).

[70] A. Lebre, J. Pastor, A. Simonet and M. Südholt. 'Putting the Next 500 VM Placement Algorithms to the Acid Test'. In: *IEEE Transactions on Parallel and Distributed Systems* (2018). URL: https://hal.inria.fr/hal-01816248 (cit. on p. 6).

[71] M. Léger, T. Ledoux and T. Coupaye. 'Reliable Dynamic Reconfigurations in a Reflective Component Model'. In: *Proceedings of the 13th International Conference on Component-Based Software Engineering*. CBSE'10. Prague, Czech Republic: Springer-Verlag, 2010, pp. 74–92. DOI: 10.1007/978-3-642-13238-4_5. URL: http://dx.doi.org/10.1007/978-3-642-13238-4_5 (cit. on p. 6).

[72] J. Lejeune, F. Alvares and T. Ledoux. 'Towards a Generic Autonomic Model to Manage Cloud Services'. In: *Proceedings of the 7th International Conference on Cloud Computing and Services Science (Best Paper Award)*, INSTICC. SciTePress, 2017, pp. 175–186. DOI: 10.5220/0006302801750186 (cit. on p. 6).

[73] L. Letondeur, X. Etchevers, T. Coupaye, F. Boyer and N. D. Palma. 'Architectural Model and Planification Algorithm for the Self-Management of Elastic Cloud Applications'. In: *2014 International Conference on Cloud and Autonomic Computing, London, United Kingdom, September 8-12, 2014*. IEEE Computer Society, 2014, pp. 172–179. DOI: 10.1109/ICCAC.2014.29. URL: https://doi.org/10.1109/ICCAC.2014.29 (cit. on p. 6).

[74] Y. Liu, P. Folz, S. Pan, F. Ramparany, S. Bolle, E. Ballot and T. Coupaye. 'Digital Twin-Driven Approach for Smart City Logistics: The Case of Freight Parking Management'. In: *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems - IFIP WG 5.7 International Conference, APMS 2021, Nantes, France, September 5-9, 2021, Proceedings, Part IV*. Ed. by A. Dolgui, A. Bernard, D. Lemoine, G. von Cieminski and D. Romero. Vol. 633. IFIP Advances in Information and Communication Technology. Springer, 2021, pp. 237–246. DOI: 10.1007/978-3-030-85910-7_25. URL: https://doi.org/10.1007/978-3-030-85910-7%5C_25 (cit. on p. 6).

[75] A. Omond, H. Coullon, I. Raïs and O. Anshus. 'Leveraging Relay Nodes to Deploy and Update Services in a CPS with Sleeping Nodes'. In: *Proceeding of the 2023 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*. Dec. 2023. DOI: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics60724.2023.00102. URL: https://hal.science/hal-04372320 (cit. on p. 18).

[76] E. Perdereau and J. Noyé. *The LIFE of CUE*. Presentation to the CONFLANG workshop, SPLASH 2023. Oct. 2023 (cit. on p. 18).

[77] V. Perifanis, N. Pavlidis, S. Yilmaz, F. Wilhelmi, E. Guerra, M. Miozzo, P. Efraimidis, P. Dini and R.-A. Koutsiamanis. 'Towards Energy-Aware Federated Traffic Prediction for Cellular Networks'. In: *2023 Eighth International Conference on Fog and Mobile Edge Computing (FMEC)*. Tartu, Estonia: IEEE, Sept. 2023, pp. 93–100. DOI: 10.1109/FMEC59375.2023.10306017. URL: https://hal.science/hal-04381142 (cit. on p. 20).

[78] G. Privat, T. Coupaye, S. Bolle and P. Raipin Parvedy. 'WoT Graph as Multiscale Digital-Twin for Cyber-Physical Systems-of-Systems'. In: *Proc., 2nd W3C Web of Things Workshop. Grenoble, France*. June 2019 (cit. on p. 6).

[79] F. Quesnel, A. Lebre and M. Südholt. 'Cooperative and reactive scheduling in large-scale virtualized platforms with DVMS'. In: *Concurrency and Computation: Practice and Experience* 25.12 (2013), pp. 1643–1655 (cit. on p. 6).

[80] F. Ramparany, M. Thalgott, S. Bolle and S. Martin. 'Exploiting Data Analytics for Home Automation Services'. In: *4th IEEE International Conference on Future Internet of Things and Cloud, FiCloud 2016, Vienna, Austria, August 22-24, 2016*. Ed. by M. Younas, I. Awan and W. Seah. IEEE Computer Society, 2016, pp. 228–234. DOI: 10.1109/FiCloud.2016.40. URL: https://doi.org/10.1109/FiCloud.2016.40 (cit. on p. 6).

[81] S. Robillard and H. Coullon. 'SMT-Based Planning Synthesis for Distributed System Reconfigurations'. In: *Lecture Notes in Computer Science*. Lecture Notes in Computer Science. Munich, Germany, Apr. 2022. DOI: 10.1007/978-3-030-99429-7_15. URL: https://inria.hal.science/hal-03536643 (cit. on p. 19).

[82] F. A. Salaht, F. Desprez, A. Lebre, C. Prud'Homme and M. Abderrahim. 'Service placement in fog computing using constraint programming'. In: *2019 IEEE International Conference on Services Computing (SCC)*. IEEE. 2019, pp. 19–27 (cit. on p. 6).

[83] F. A. Salaht, F. Desprez and A. Lebre. 'An overview of service placement problem in fog and edge computing'. In: *ACM Computing Surveys (CSUR)* 53.3 (2020), pp. 1–35 (cit. on p. 6).

[84] M. Satyanarayanan. 'The Emergence of Edge Computing'. In: *Computer* 50.1 (Jan. 2017), pp. 30–39. DOI: 10.1109/MC.2017.9 (cit. on p. 5).

[85] D. Serrano, S. Bouchenak, Y. Kouki, F. A. de Oliveira Jr., T. Ledoux, J. Lejeune, J. Sopena, L. Arantes and P. Sens. 'SLA guarantees for cloud services'. In: *Future Generation Computer Systems* 54.Supplement C (2016), pp. 233–246. DOI: https://doi.org/10.1016/j.future.2015.03.018. URL: http://www.sciencedirect.com/science/article/pii/S0167739X15000801 (cit. on p. 6).

[86] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu. 'Edge computing: Vision and challenges'. In: *IEEE Internet of Things Journal* 3.5 (2016), pp. 637–646 (cit. on p. 5).

[87] J. M. Van Ham, G. Salvaneschi, M. Mezini and J. Noyé. 'JEScala: Modular Coordination with Declarative Events and Joins'. In: *Proceedings of the 13th International Conference on Modularity (MODULARITY '14)*. Lugano, Switzerland: ACM Press, 2014, pp. 205–216. URL: http://doi.acm.org/10.1145/2577080.2577082 (cit. on p. 6).

[88] Y. Xia, X. Etchevers, L. Letondeur, A. Lebre, T. Coupaye and F. Desprez. 'Combining heuristics to optimize and scale the placement of iot applications in the fog'. In: *2018 IEEE/ACM 11th International Conference on Utility and Cloud Computing (UCC)*. IEEE. 2018, pp. 153–163 (cit. on p. 6).

[89] Y. Xiong, Y. Sun, L. Xing and Y. Huang. 'Extend Cloud to Edge with KubeEdge'. In: *2018 IEEE/ACM Symposium on Edge Computing (SEC)*. 2018, pp. 373–377. DOI: 10.1109/SEC.2018.00048 (cit. on p. 7).