EUCloudEdgeIoT.eu

Towards an Open Heterogeneous Smart Cloud

Josep Lluís Berral, Aaron Call, Ramon Canal
Barcelona Supercomputing Center, Universitat Politècnica de Catalunya

The Claridge – Brussels, Belgium | 10-11 May 2023

Concertation and Consultation on Computing Continuum:
From Cloud to Edge to IoT

Organized by: **Open Continuum** | Supported by: Unlock CEI and SWForum

Open Heterogeneous Smart Cloud

Cloud-Edge Scenarios

- Pervasive Computing (IoT, Wearables, Edge)
- Big Data (Cloud and Analytics)
- Artificial Intelligence (Machine Learning applications)

Current Status

- Growth on data/users (sources, storage, transmission, consumers, clients)
- Scalability, reliability and adaptivity vs. Distributed and heterogeneous

Challenges

- Development & deployment at Cloud-Edge on trustworthy and open HW
- Management of hyper-distributed resources
- Migration of AI/Analytics near to data

Aspects to Target

1. Cloud-Edge into Open Architectures
   - Open access to underlying technologies -> RISC-V
   - Transparent design -> Knowledge of underlying HW
   - Transparent management -> Guarantee QoS & privacy

   } EPI & EU technological sovereignty

1. Heterogeneous Resources Integration
   - Impact on energy, availability and trust
   - Use of accelerators -> Variety of devices/providers
   - No "one fits all" policy -> Standardization of methods

   } Sustainable competition

1. AI in the Orchestration Cycle
   - Automation of management -> Human out-of-the-loop
   - Hyper-distribution -> Pay attention to privacy
   - Federated infrastructures -> Leverage SoA in AI/ML

   } Enhance efficiency

EUCloudEdgeIot.eu is supported by the Open Continuum and Unlock CEI and both received funding from the European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement numbers 101070030 and 101070571.

EUCloudEdgeIoT.eu

josep.berral@bsc.es

VITAMIN-V
VIRTUAL ENVIRONMENT AND TOOL-BOXING FOR TRUSTWORTHY DEVELOPMENT OF RISC-V BASED CLOUD SERVICES

EU-HORIZON GA.101093062

The Claridge – Brussels, Belgium | 10-11 May 2023

Concertation and Consultation on Computing Continuum:
From Cloud to Edge to IoT

Organized by: **Open Continuum |** Supported by: Unlock CEI and SWForum

Full-stack Runtime Reconfiguration for Efficient Data Processing

Jens Hagemeyer – Bielefeld University

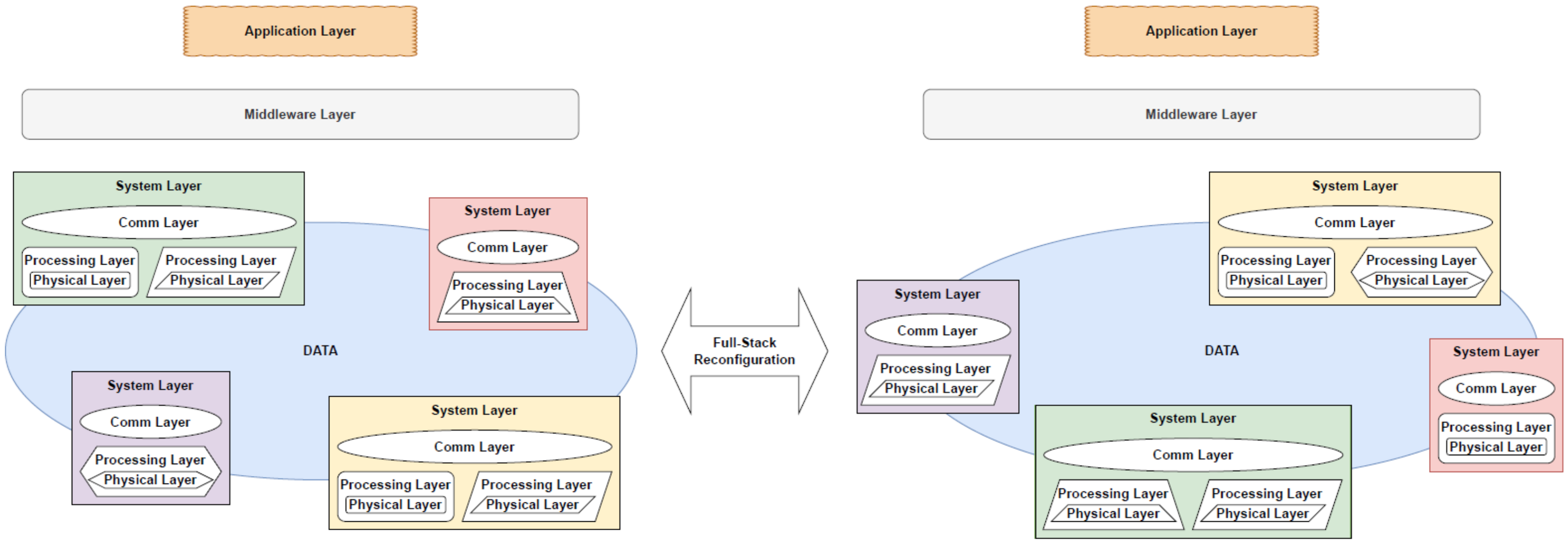
The Claridge – Brussels, Belgium | 10-11 May 2023

Concertation and Consultation on Computing Continuum:
From Cloud to Edge to IoT

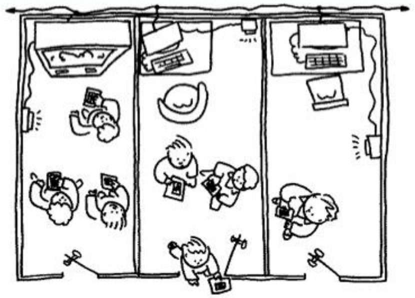
Organized by: **Open Continuum |** Supported by: Unlock CEI and SWForum

Motivation (Why Full-stack Runtime Reconfiguration?)

- Reconfigurability is a powerful tool as it allows to
  - adapt the architecture to the application
  - be applied from physical to application level
  - be a generic tool for optimizing data processing, communication, data movement, etc.
  - use existing research
  - adapt at run-time to changing requirements in application



Status (Run-time reconfigurion is already there)

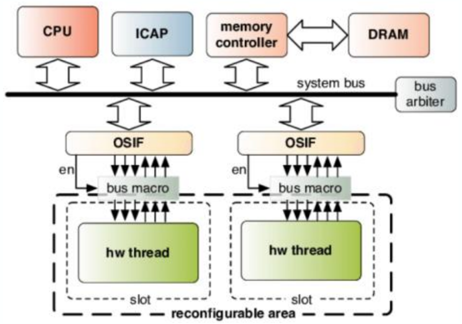


Middleware layer

- Contextaware computing in ubiquitous/pervasive Computing

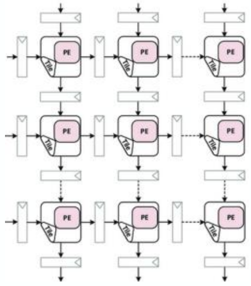
System layer

- Management level
- OS and deployment support



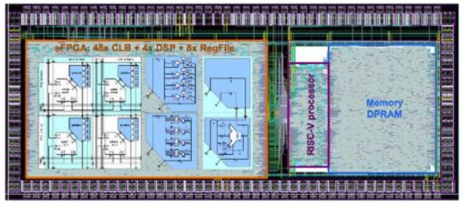
Communication layer

- Dataflow computing
- Data-centric computing



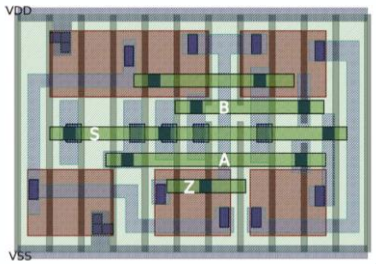
Processing layer:

- FPGAs and Coarse Grain Architectures
- Reconfigurable Processors



Physical layer:

- RFETs (devices that can be configured between an n-channel and p-channel behavior)
- Memristor-based LUT cells



Challenges (How to get Full-stack Runtime Reconfiguration?)

- Develop **full-stack run-time reconfiguration**
  - Foster run-time reconfigurability on all architectural levels – transistor to application
  - Employ co-design methodology for run-time reconfigurability
  - Include dependability aspects
  - Develop a "run-time reconfigurabilty aware" middleware and run-time aware, including abstraction layers and auto-reconfigurabilty features
  - Port and demonstrate on application level

- Resulting in
  - better hardware utilization
  - better fit to changing application demands at runtime
  - improved performance and efficiency
  - improved sustainability and dependability

EUCloudEdgeIot.eu is supported by the Open Continuum and Unlock CEI and both received funding from the European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement numbers 101070030 and 101070571.

EUCloudEdgeIoT.eu

Contacts

Mario Porrmann, Osnabrück University, Germany, mario.porrmann@uni-osnabrueck.de

Marcelo Pasin, University of Neuchâtel, Switzerland, marcelo.pasin@unine.ch

Pedro Trancoso, Chalmers University of Technology, Sweden, ppedro@chalmers.se

Jens Hagemeyer, Bielefeld University, Germany, jhagemey@techfak.uni-bielefeld.de

EUCloudEdgeIoT.eu

Towards next generation digital infrastructures to contribute to a more resilient and sovereign European economy

Isabelle Chrisment – Inria

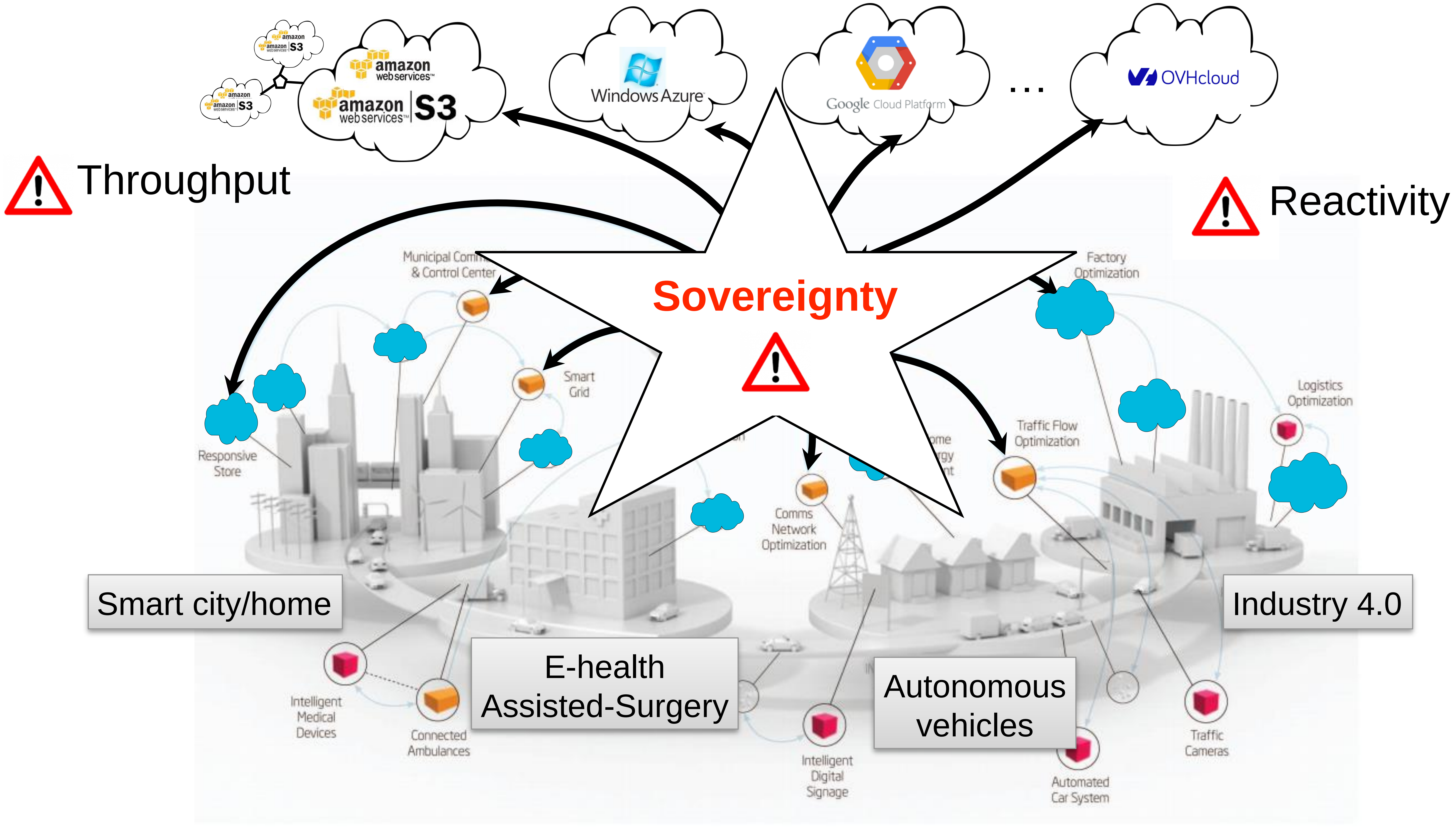
The Claridge – Brussels, Belgium | 10-11 May 2023

Concertation and Consultation on Computing Continuum:
From Cloud to Edge to IoT

Organized by: **Open Continuum |** Supported by: Unlock CEI and SWForum

DIGITAL INFRASTRUCTURE

UBIQUITOUS AND CONSTANTLY EVOLVING



Operators: How to administrate?

Developers: How to program?

RESEARCH CHALLENGES

- **How to manage Compute/Network/Storage resources?**
  - High volatility, heterogeneity, constrained resources (embedded systems, CPS, …),
  - Multi tenant: computing continuum delivered by multiple operators.
- **How to design new storage systems for sharing data?**
  - Multiple failures at network and node level,
  - Untrusted nodes and data protection.
- **How to address energy consumption and environmental impact?**
  - Tradeoff between energy consumption and resilience,
  - Use of renewable energy / intermittent energy (impacting the resilience).
- **How to adapt security mechanisms?**
  - Specific security threats to cloud paradigms (SaaS, PaaS, IaaS),
  - Security policies for more complex and dynamic environments (inconsistencies between tenants or administrators).

DIGITAL INFRASTRUCTURES@INRIA

EUROPE

- **SLICES-RI (Research Infrastructure) in the 2021 ESFRI Roadmap**
  - Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies
  - SLICES-PP (Preparatory Phase): Inria coordinator

NATIONAL

- **PEPR (Priority Research Programs and Equipments) CLOUD**
  - 56 M€, Inria co-leader, 7 years starting in September 2023
- **OTPAAS - Delivering an edge software stack for the industry**
  - 56 M€, Inria partner, October 2021 - 2024

INRIA Challenges (ambitious research projects)

- **Inria-OVHCloud :** End-to-end eco-design of a cloud to reduce its environmental impact
- **Inria-Qarnot Computing:** PUshing Low-carbon Services towards the Edge
- **Inria-Hive :** Large Scale Secure and Reliable Peer-to-Peer Cloud Storage

EUCloudEdgeIot.eu is supported by the Open Continuum and Unlock CEI and both received funding from the European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement numbers 101070030 and 101070571.

EUCloudEdgeIoT.eu

Adrien Lebre : adrien.lebre@inria.fr
Frédéric Desprez: frederic.desprez@inria.fr
Isabelle Chrisment: isabelle.chrisment@inria.fr

EUCloudEdgeIoT.eu

Seamless X-by-Construction Methods towards scalable & reliable Cloud-Edge-IoT Computing

Jürgen Becker – Karlsruhe Institute of Technology (KIT)

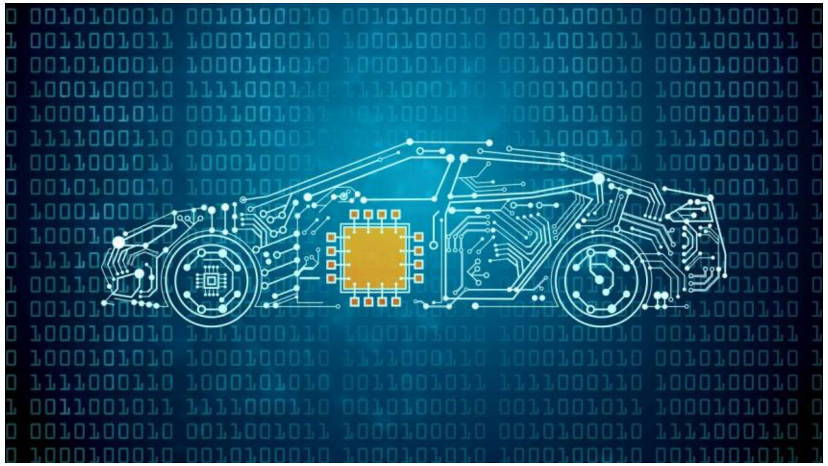
The Claridge – Brussels, Belgium | 10-11 May 2023

Concertation and Consultation on Computing Continuum:
From Cloud to Edge to IoT

Organized by: **Open Continuum |** Supported by: Unlock CEI and SWForum

Motivation and status

- Benefits of a cloud-edge continuum: Energy efficiency, improved customer experience, intellectual property protection, central/decentral trade-offs …
- However, integrating safety-critical nodes into such a continuum is a challenging (real-time) task:
  - Inherently dynamic behaviour of applications
  - Fast integration cycles + over-the-air updates
  - Extreme platform heterogeneity + adaptivity

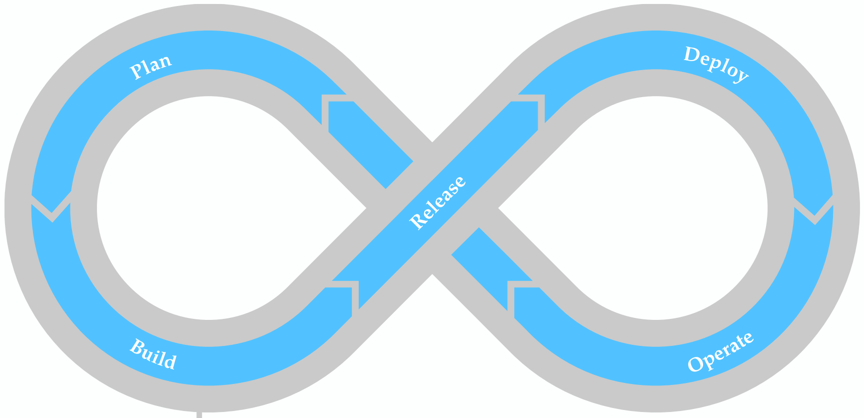


© kaptn – stock.adobe.com

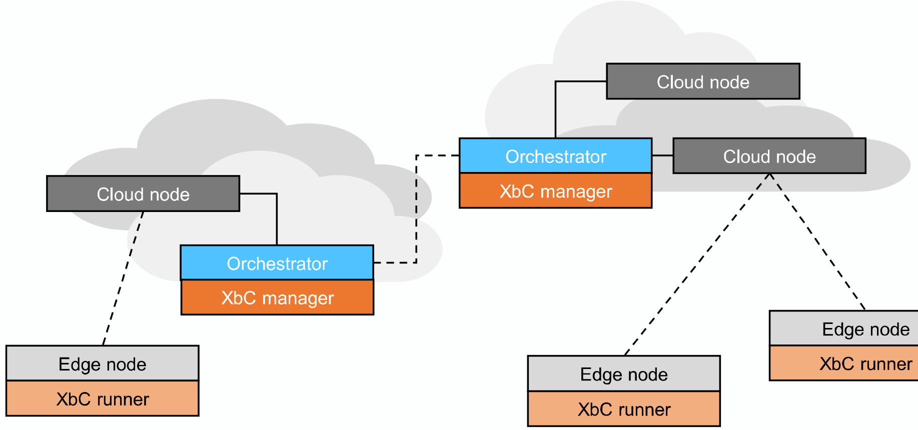
- Future autonomous systems (e.g. in the automotive domain) depend on:
  - A scalable and reliable cloud continuum (esp. from a safety/real-time perspective)
  - First-class support for heterogeneous hardware platforms (incl. AI accelerators)
  - Automatizable on-demand *X-by-Construction* integration (incl. Validation & Deployment)

Research challenges

- How can a reliable federation of cloud ecosystems with support for heavily heterogeneous hardware platforms be realised?
- Promising solution: The novel *X-by-Construction (XbC)* paradigm
  - Key idea: Automatic generation of system implementations with guaranteed properties
  - Currently limited to the design phase ⇒ ill-suited for highly adaptive systems
  - *XbC* Applicability to dynamic systems is a crucial research question



X-by-Construction strategy
(along the entire DevOps cycle)



European Cloud Processor

Roger Espasa, Founder & CEO

Problem: No European Cloud Processors

- ALL cloud architectures (X86 and ARM) are non-European
- ALL cloud processors are non-European
- Hence
  - Concerns on security and sovereignty
  - Innovation happening elsewhere
  - Economic opportunity loss

- Solution(s)
  - Adopt an open architecture (RISC-V) for the long term
  - Design a European processor for the cloud continuum

semidynamics

The R&D Challenges

- Research, definition and design of a high-frequency, very wide (12-wide) out-of-order core, with the following key attributes
  - At least 3 memory operations issued per cycle
  - Native hardware support for virtual machines through a hypervisor layer
  - Extensive protection of processor internal storage through ECC
  - Support for cryptographic instructions
  - Support for security enclaves, either through separate memory spaces or a similar technology
  - Cache coherent
  - Support for efficient synchronization across cores
  - Support for multithreading
  - New forms of energy efficiency, energy allocation and energy rationing
  - Resistant to side-channel attacks
- Research, definition and design of a high-performance “uncore” tailored for the cloud, capable of
  - Supporting 16 to 128 of the above defined core
  - Supporting large second level and last level caches
  - Advanced prefetching techniques
  - Novel resource partitioning algorithms to split the hardware resources to different virtual machines following an administrator-set policy
  - Memory encryption techniques
  - Advanced network-on-chip
  - Advanced reliability techniques, for all “uncore” components
  - Energy and power controller techniques to manage the cores and the “uncore”
  - Secure boot technologies, preferably open-source, for maximum public scrutiny
  - High bandwidth off-ide interfaces, specially to remote memory (such as CXL memory)

semidynamics

EUCloudEdgeIot.eu is supported by the Open Continuum and Unlock CEI and both received funding from the European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement numbers 101070030 and 101070571.

EUCloudEdgeIoT.eu

Thank you

The Claridge – Brussels, Belgium | 10-11 May 2023

Concertation and Consultation on Computing Continuum:
From Cloud to Edge to IoT

Organized by: **Open Continuum |** Supported by: Unlock CEI and SWForum

semidynamics

EUCloudEdgeIoT.eu

UPV-SRIPAS

Expression of Interest on future visions and research directions 2025-27 in the area of Cloud-to-Edge-to-IoT for European Data

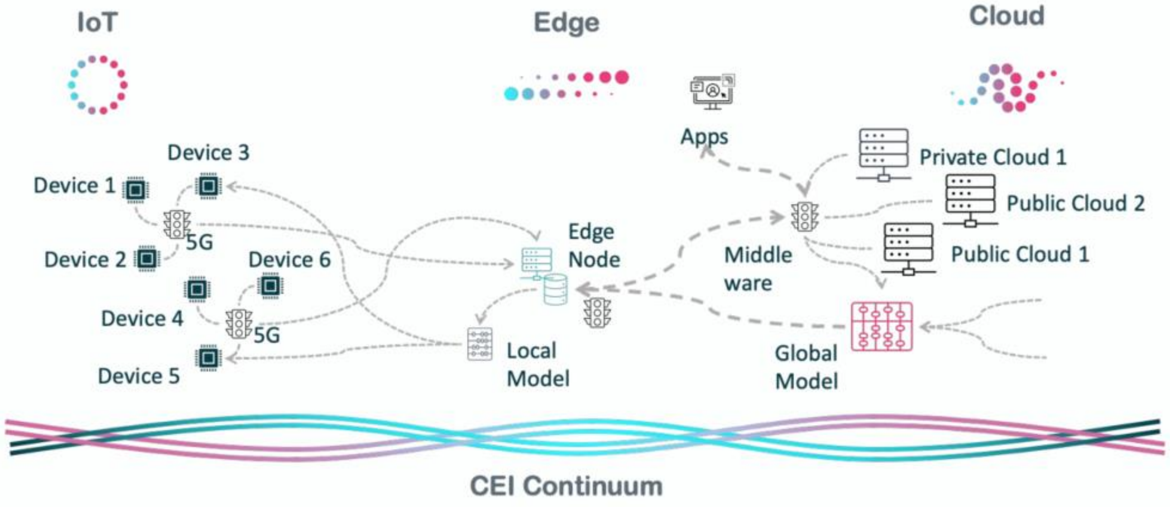
Prof. Carlos E. Palau - UPV

The Claridge – Brussels, Belgium | 10-11 May 2023

Concertation and Consultation on Computing Continuum:
From Cloud to Edge to IoT

Organized by: **Open Continuum |** Supported by: Unlock CEI and SWForum

What we see in the current status



- There is a lack of coverage within the range of computing devices in the continuum. Non-Linux equipment, IoT-resource-constrained elements, FPGAs and others are often disregarded, thus the continuum does not stretch well to those kinds of devices.
- There is no proper open source, end-to-end service pipelines that abstract the complexity and heterogeneity of the underlying equipment.
- The management of the lowest layers of the continuum is not solved at all, with connectivity and interoperability issues even at the level of research actions.
- Privacy and security are not tackled homogeneously across the entire continuum, and all stages of applications' lifecycle and may require a holistic approach from different clusters (i.e. CL4 and CL3).
- The lack of formal structure of the continuum hinders potential of organization and standardization.

Research challenges and mid-term priorities

- **Security, privacy and data spaces** for the continuum.
- Advanced (frugal, explainable, trustworthy) AI mechanisms for improving continuum performance: robustness, **autonomy, traceability, governance.**
- Underlying **network automation**, IP-abstraction and connectivity by service names (eBPF, kernel technologies…).
- Achievement of a true AI-based cognitive mesh with ambient intelligence.
- **Miniaturization** of workloads' containers and packages.
- **Explainable offloading and orchestration,** involving context-aware self-configuration of workloads (from day-0 to day-2).
- Business models for effective **federation** of resource sharing in **multi-stakeholder scenarios**.

The next calls in the programme must target long-term goals:

- European values: GDPR natively and lightweight DLT to govern all continuum transactions (sovereign continuum).
- Europe's own technology must stand out: Alternative CPU architectures completely embedded in the continuum, prominently RISC-V.
- True tactile deployments: Reconfiguration of the continuum in runtime in milliseconds time.
- Metaverse of the continuum, including VR simulation and IoT-powered ubiquitous data. Self-organizing networks with automatic formation, maintenance, and adaptation to ensure optimal data flow, load balancing, and resource allocation in complex, distributed computing environments

EUCloudEdgeIot.eu is supported by the Open Continuum and Unlock CEI and both received funding from the European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement numbers 101070030 and 101070571.

EUCloudEdgeIoT.eu

Prof. Carlos E. Palau (UPV) – cpalau@dcom.upv.es

Dr. Ignacio Lacalle (UPV) – iglaub@upv.es

Prof. Maria Ganzha (SRIPAS) – Maria.Ganzha@ibspan.waw.pl

Dr. Katarzyna Wasielewska-Michniewska (SRIPAS) - wasielk@ibspan.waw.pl

The Claridge – Brussels, Belgium | 10-11 May 2023

Concertation and Consultation on Computing Continuum:
From Cloud to Edge to IoT

Organized by: **Open Continuum |** Supported by: Unlock CEI and SWForum