ORIGAMI

Optimized Resource Integration And Global Architecture For Mobile Infrastructure For 6G

SA5-Origami joint workshop

A view on next 6G system

ORIGAMI Consortium

Co-funded by the European Union

6GSNS

GA:101139270

Call: HORIZON-JU-SNS-2023-STREAM-B-01-01

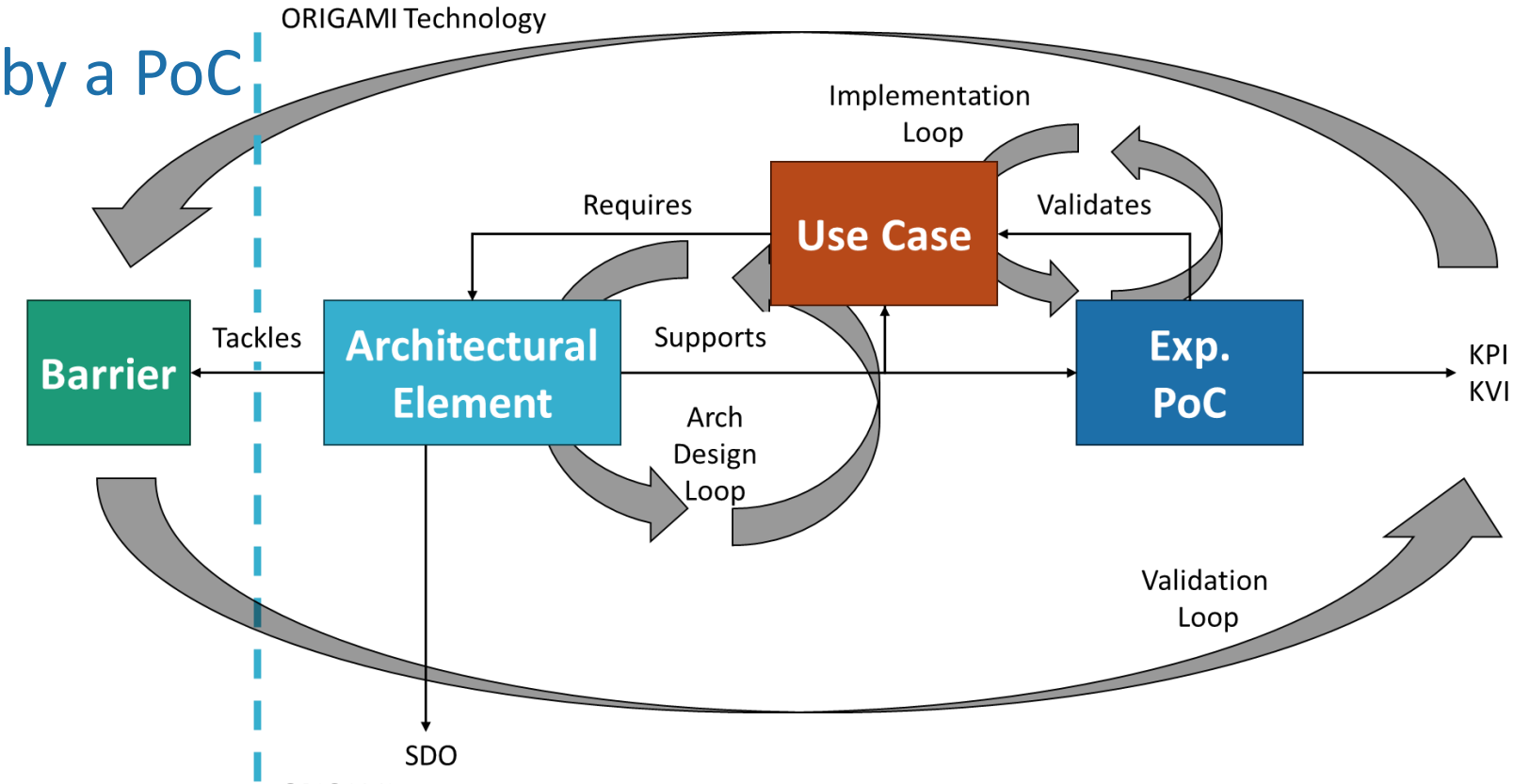
Agenda

- Introduction to the meeting
- Rel-19 SA5 working progress and Rel-20 3GPP SA5 Work Planning
- Enhancement of Service Based Management Architecture.

  GSBA an enhancement of Service Based Approach applied to 6G system: main characteristics and impact on the services, network and management
- Zero-Trust Exposure Layer: a novel Management Service Exposure that enables vertical tenants or other MNOs to interact with the mobile network according to their needs and maintain streamlined global operations.
- Management of network intelligence and the computing resources: Managing Network Intelligence (NI) instances with a simple interface that can be operated at fast timescales, while hiding the complexity of the underlying hardware.
- Q&A

ORIGAMI Overall Methodology

- The ORIGAMI project identified a set of Barriers that currently hinder the success of next generation mobile networks.
- Through the identification of a set of use cases, ORIGAMI propose the introduction of novel architectural models
- Each UC will be validated by a PoC



Barrier description

- Needs to introduce new technologic solutions and new features in the network and in the management system of 6G system

#	Barrier	Description
1	Unsustainable RAN Virtualization	Ineffective deployment and utilization strategies of computing resources at the radio access, as RAN virtualization solutions today rely on expensive and energy-consuming dedicated hardware accelerators.
2	Poor Interoperability of RAN Components	Lack of inter-operability between open RAN components and effective deployment of the RAN bus to enable multi-time scale controllability
3	High Latency and Unreliable Network Intelligence (NI)	We need to solve complex 6G user-plane (UP) processing operations and control-plane (CP) optimization problems with low-latency and robust Network Intelligence (NI).
4	Under-utilized Modern Programmable Transport	Lack of models and tools to transparently exploit the in-band computing capabilities of modern programmable transport domains for the deployment of 6G VNFs into the user plane.
5	Lack of Global Service Application Programming Interfaces (API)	Lack of infrastructure exposure and standardized APIs that go beyond silo-style optimization and can enable global service and monitoring, to support demands by global 6G devices
6	Obsolete Trust Model Hinders Performance	Current trust model hinders performance and growth. Scaling coverage requires pre-established agreements between the end-user and the cellular provider
7	Inadequate Networking Data Representation	Lack of global data provisioning and representation makes service monitoring and root-cause detection challenging.
8	High Volume of Control Plane Signaling	The rapid growth of control plane signaling, and the monolithic implementation of the NFs reduce scaling and resilience capabilities for the network core

EUROPEAN VISION FOR THE 6G NETWORK ECOSYSTEM
10.5281/zenodo.13708424

Origami use case list

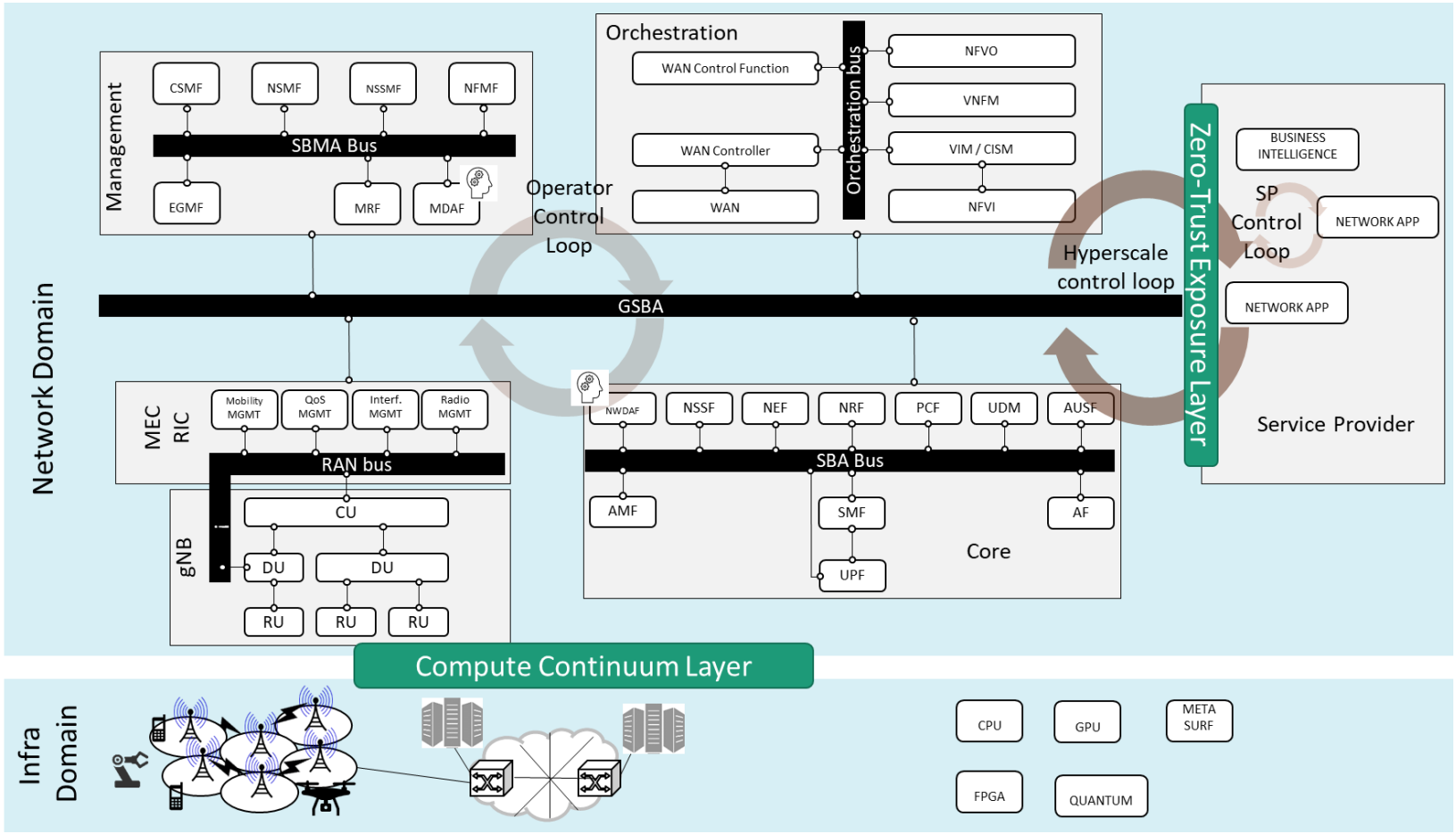
Use case	Id	Architectural innovation	Barrier #	Barrier	Owner
Data-driven task offloading for reliable vRAN acceleration	SRV	CCL	1	Unsustainable RAN Virtualization	NEC, UC3M, i2CAT
Enhancing Management and Stability in the 6G Architecture	EMSA	GSBA	2	Poor Interoperability of RAN Components	ISRD
Interoperable Machine Learning Models Improving RAN Energy Efficiency	IMLE	CCL, GSBA	3	High Latency and Unreliable Network Intelligence (NI)	NETAI
Compute- and Fairness-Aware Radio Resource Allocation Algorithms in Virtualized RANs	CFA	CCL	1, 3	Unsustainable RAN Virtualization High Latency and Unreliable Network Intelligence (NI)	TUD, NEC
Effective, distributed and streamlined access to u-plane computing capabilities	EAUC	CCL, GSBA	4	Under-utilized Modern Programmable Transport	IMDEA
Enabling the Global Operator Model	GMNO	ZTL, GSBA	5, 6	Lack of Global Service Application Programming Interfaces (API) Obsolete Trust Model Hinders Performance	TID
Limited Trust Network Analytics	LTNA	ZTL	6	Obsolete Trust Model Hinders Performance	UC3M
Anomaly Detection	KR	GSBA	7	Inadequate Networking Data Representation	TID, EMN/JMU
Network Core traffic analysis and optimization	NCAM	GSBA	8	High Volume of Control Plane Signaling	FOGUS

Topics of Interest

- Network Resource Model
- Orchestration
- AI/ML
- Energy Efficiency
- Network Performance Management
- Data Analytics
- Network Digital Twins
- Infrastructure Management
- Security Aspects
- Evolution of CAPIF approach

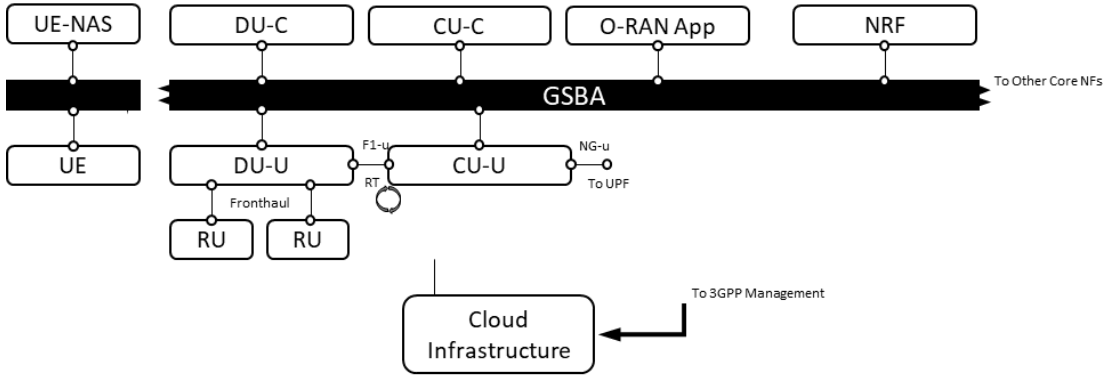
GSBA

- The ORIGAMI architectural design aims to tear down the silo-based design of different network domains, integrating one single bus GSBA to link them all

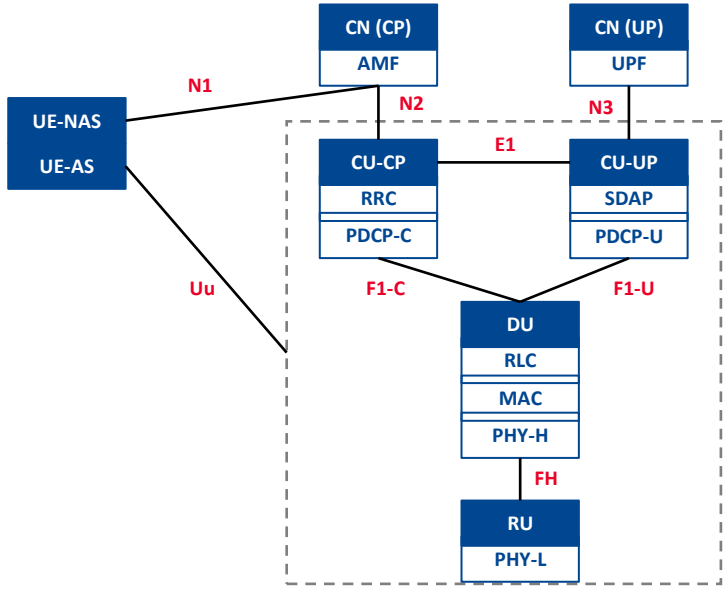


GSBA Highlights

- Among the main highlights of the GSBA there is its extension towards the RAN
- This includes the need to split current monolithic modules into c-plane (that interface with the GSBA) and u-plane functions, analogously to the ones in the 5GC



Possible network interfaces to consider for introducing SBA:
N1, N2, F1-C and E1



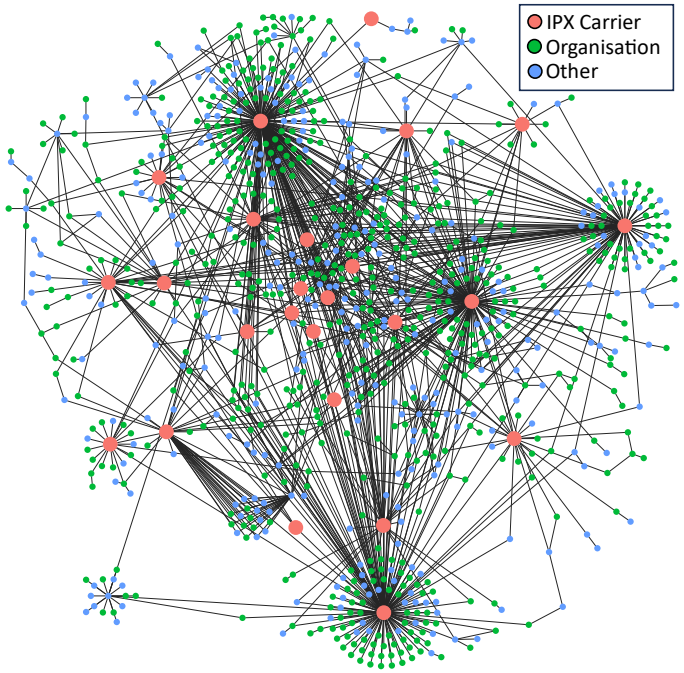
SBA applied to RAN: impacts on Management Services

- Introducing SBA into the RAN brings several benefits to network operations and management.
  - **Improved Scalability**: Service-based architectures (SBA) decouple network functions into modular services, allowing networks to scale dynamically based on demand. This flexibility is essential for handling traffic spikes or expanding capacity without significant hardware investments.
  - **Enhanced Flexibility and Modularity:** The modular design of SBA enables independent development, deployment, and management of network functions. This approach supports agile updates and innovation without disrupting the entire network.
  - **Improved Automation and Orchestration**: SBA supports advanced automation tools and orchestration frameworks. By using programmable interfaces and machine-readable APIs, networks can automate tasks like resource allocation, fault detection, and recovery.
  - **Better Resource Management**: Fine-grained control over individual services enables better monitoring and allocation of resources.
  - **Resilience and Reliability**: SBA supports fault-tolerant designs where failures in one service do not cascade across the network. This architecture improves the overall reliability and uptime of the network.
  - **Future-Proofing**: With its modular and API-driven approach, SBA is well-suited to adapt to evolving standards, protocols, and technologies, ensuring long-term relevance and reduced need for overhauls

Network Resource Model

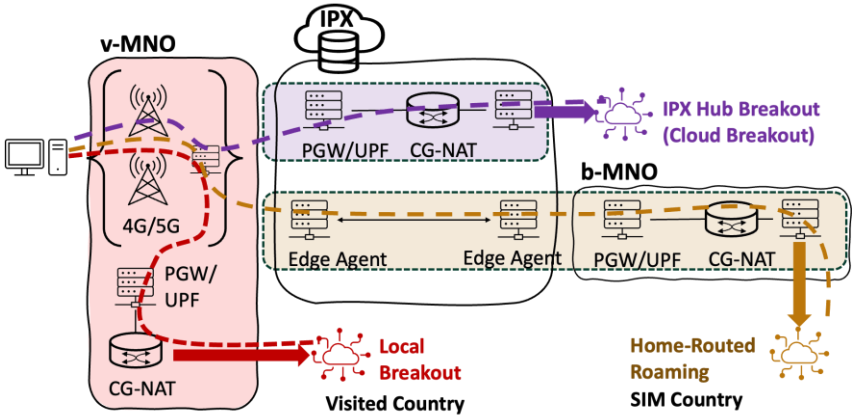
- GSBA enable Resource Abstraction
  - The term "resource abstraction" is used to describe a data model that is capable of representing network and service-related resources in a manner that is independent and general, regardless of the specific solutions that have been implemented.
- ORIGAMI involved Use Cases
  - Enabling the Global Operator Model (GMNO)
  - Effective, distributed and streamlined access to u-plane computing capabilities (EAUC)

Enabling the Global Operator Model [GMNO]



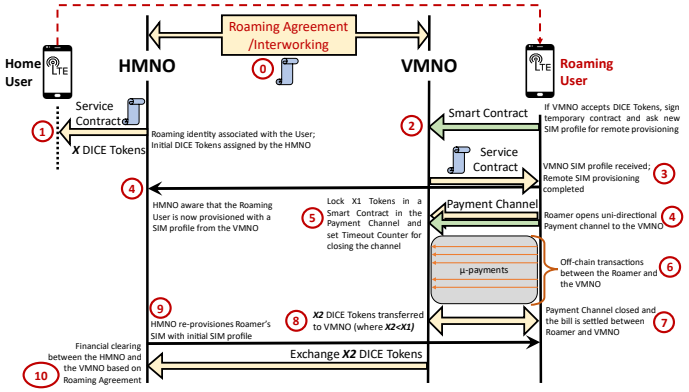
Topology Mapping for the Cellular Ecosystem

- Radio Resources (MNOs)
- Cloud Resources
- Interconnection Infrastructure



The GSBA support the design of the ***Optimal Breakout Selection*** (Cloud Breakout):

- Standards-compliant architecture that achieves the performance benefits of local breakout
- Constraints-aware cost model for deploying the global infrastructure



The ZTL will support the exploration of ***permissioned consortium blockchain*** for global operations through the DICE protocol:

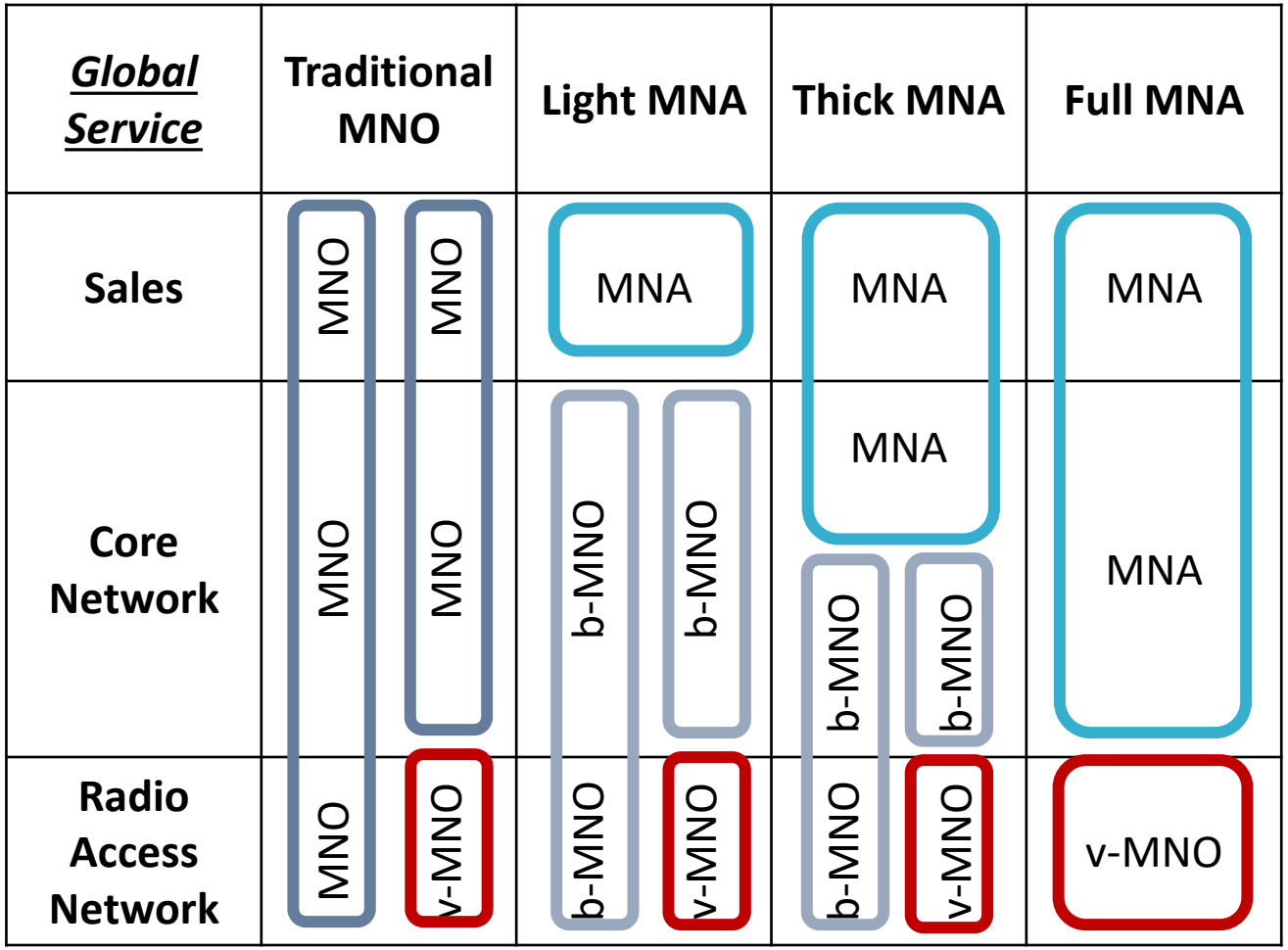
- Zero-trust interactions between resource providers
- Bypass the use of a third party (DCH) for data and financial clearing
- Immutability of global transactions (tamper-proof ledger)
- Confidentiality of the agreements and the transactions

[TMA'24] Viktoria Vomhoff, Marleen Sichermann, Stefan Geißler, Andra Lutu, Martin Giess, and Tobias Hoßfeld. "A Shortcut through the IPX: Measuring Latencies in Global Mobile Roaming with Regional Breakouts."

Co-funded by the European Union

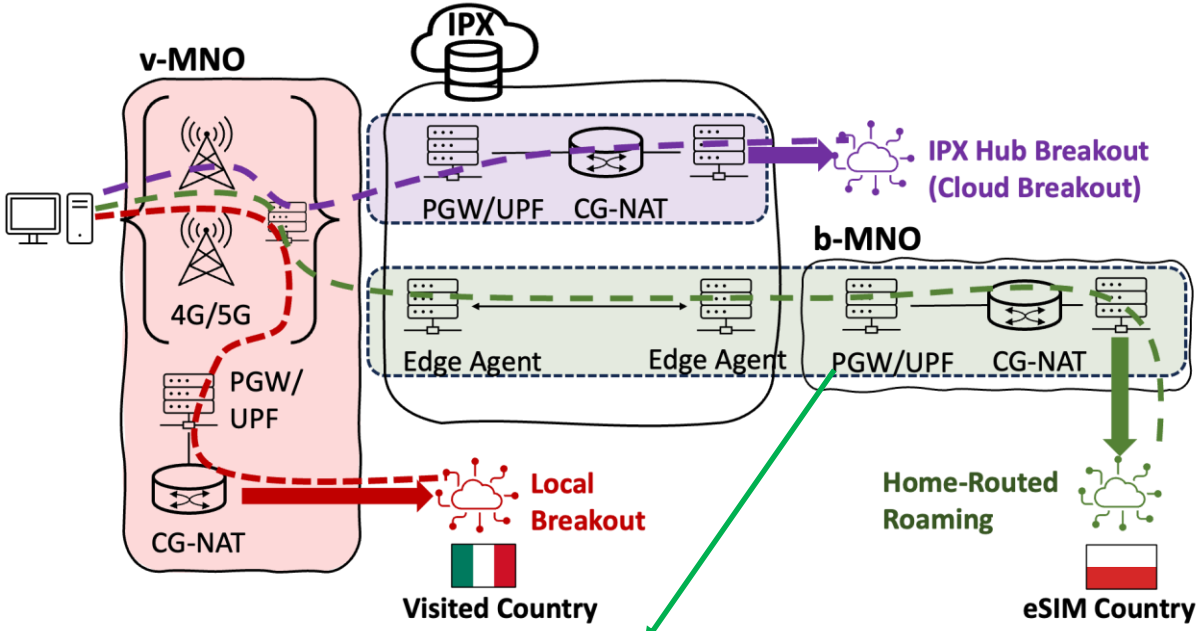
Enabling the Global Operator Model [GMNO]

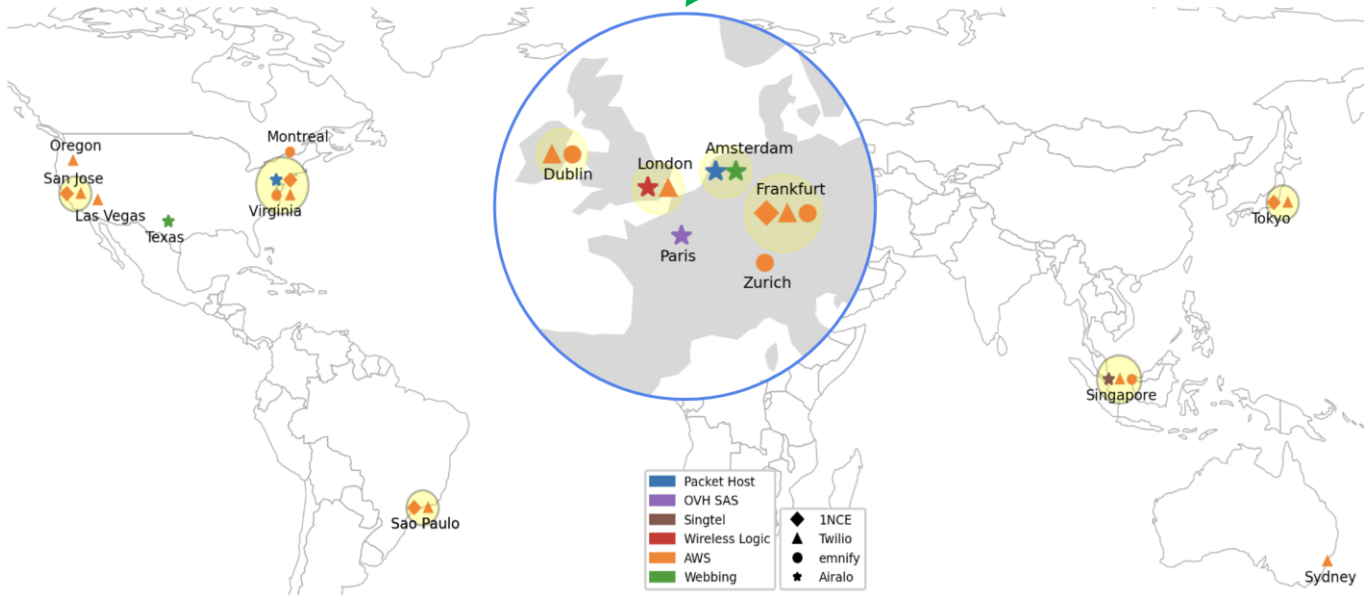
- The GSBA and ZTL enable a new global model of operators that tackle the limitations of the current architectures
  - Though global operators already operate in the cellular ecosystem (e.g., emnify for IoT), they operate by overloading the roaming function and thus suffer for performance impairments in terms of data path performance and billing efficiency
- In a global setting, the **Thick Mobile Network Aggregator** (MNA) builds on resources from different entities:
  - The v-MNO provides RAN resources
  - The b-MNO provides Core Network resources
  - Cloud providers host specific Core Network functions that the MNA operates (e.g., the UPF)



Enabling the Global Operator Model [GMNO]

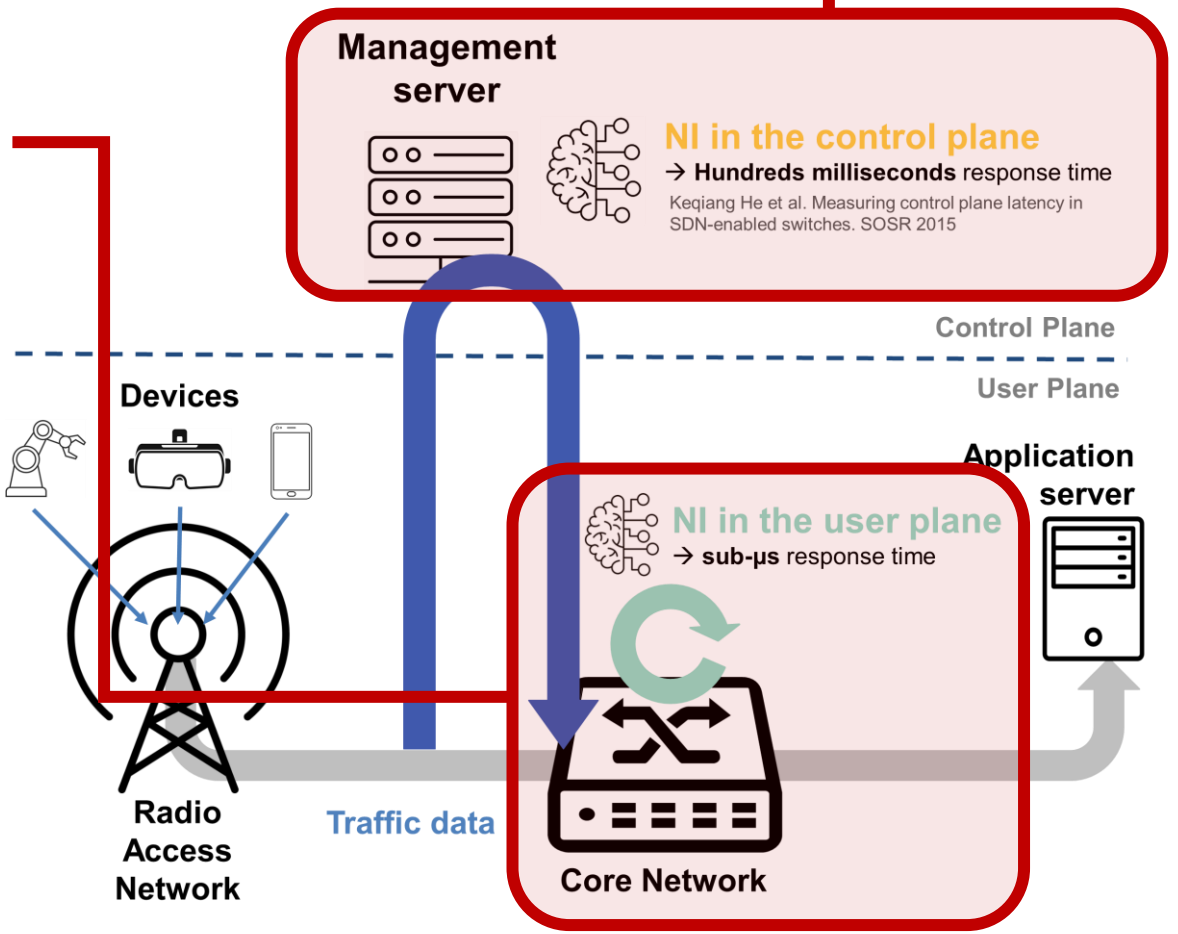
- Through the GMNO use-case, ORIGAMI generates new solutions for:
  - Optimal Breakout Selection (exploiting the GSBA)
  - Dynamic Real-time Billing in a Global Operator (exploiting ZTL)
- **Optimal Breakout Selection** via the GSBA answers to the need of the GMNO to build the best data path for its end-users
  - ORIGAMI modifies the UPF logic to dynamically deploy it close to the end-user, exploiting multi-cloud orchestration





Distributed inference in the user plane [EAUC]

- Traditional Network Intelligence (NI) resides in the control plane
  - AI/ML models perform inference on data received from the user plane
  - Cross-plane loop causes 10-100 ms latency
- **User-plane inference removes the loop**
  - AI/ML models are deployed directly in programmable network hardware with P4
    - E.g., Intel Tofino switches, smartNICs, DPUs
  - NI operates at line rate at sub-µs latency
- Current state of the art
  - Random Forest (RF) on 20+ traffic classes
  - Demonstrations on real-world traffic classification, device identification, or intrusion detection



Distributed inference in the user plane [EAUC]

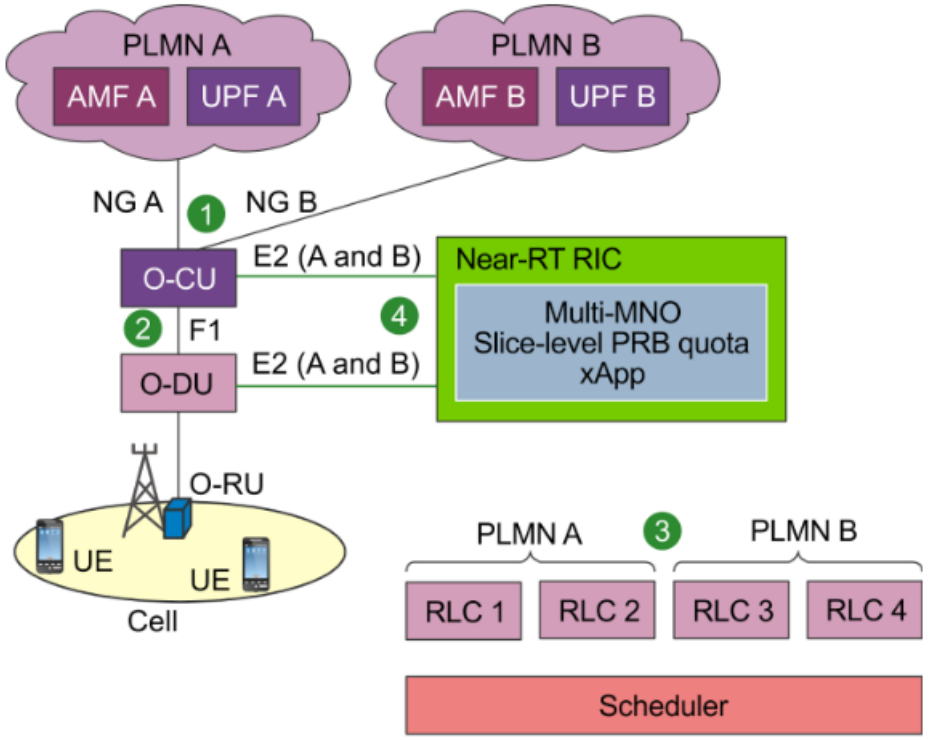
- User-plane NI is presently very ad-hoc
  - Requires manual designing of the model and lots of tinkering for the P4 implementation
  - Is deployed in a single device and does not leverage the distributed nature of networks
  - Needs starting from scratches for each different inference task or application use case
- **ORIGAMI will streamline user-plane NI programming**
  - Defines strategies to efficiently distribute a single AI/ML model across the network
    - Decomposes the target task in AI/ML sub-models each solving a subset of the original task
    - Deploys the AI/ML sub-models across multiple devices
  - Defines pipelines to automatically design and deploy P4 models for any inference task
  - Defines an abstraction layer to expose the programmable user-plane resources
    - Formalizes the capabilities of the (multiple, heterogeneous) programmable devices
    - Exposes such capabilities to the owners of any inference task
- **ORIGAMI makes user-plane capacity easy to use for line-rate tasks**
  - Favors intra-domain orchestration of cutting-edge infrastructure resources

Orchestration

- GSBA enable better Orchestration
  - The term 'orchestration' is used to describe the entire data set, the full range of services and the complete set of functionalities that are involved in the management of networks and the associated services. These include, for example, the processes of network creation, network and service assurance, and the optimisation of resources
- Related ORIGAMI Use Cases
  - Enhancing Management and Stability in the 6G Architecture (EMSA)

Enhancing Management and Stability in the 6G Architecture [EMSA]

- EMSA will leverage the management of networks and the associated services
- The optimization of the resources in by sharing a single RAN by two different core networks will optimize the slice-level PRB quota's
- EMSA for orchestrating the resources can enable faster and more realistic allocation of the resources to manage 6G network
- EMSA will strategically leverage the orchestration of resources by collecting KPMs and data exposed by the GSBA to provide in real-time network management



O-CU and O-DU supporting Multi-MNO

1. O-CU supporting two NG interfaces
2. F1 interface modification
3. For scheduler, associate the RLC entity with PLMN

Near-RT RIC

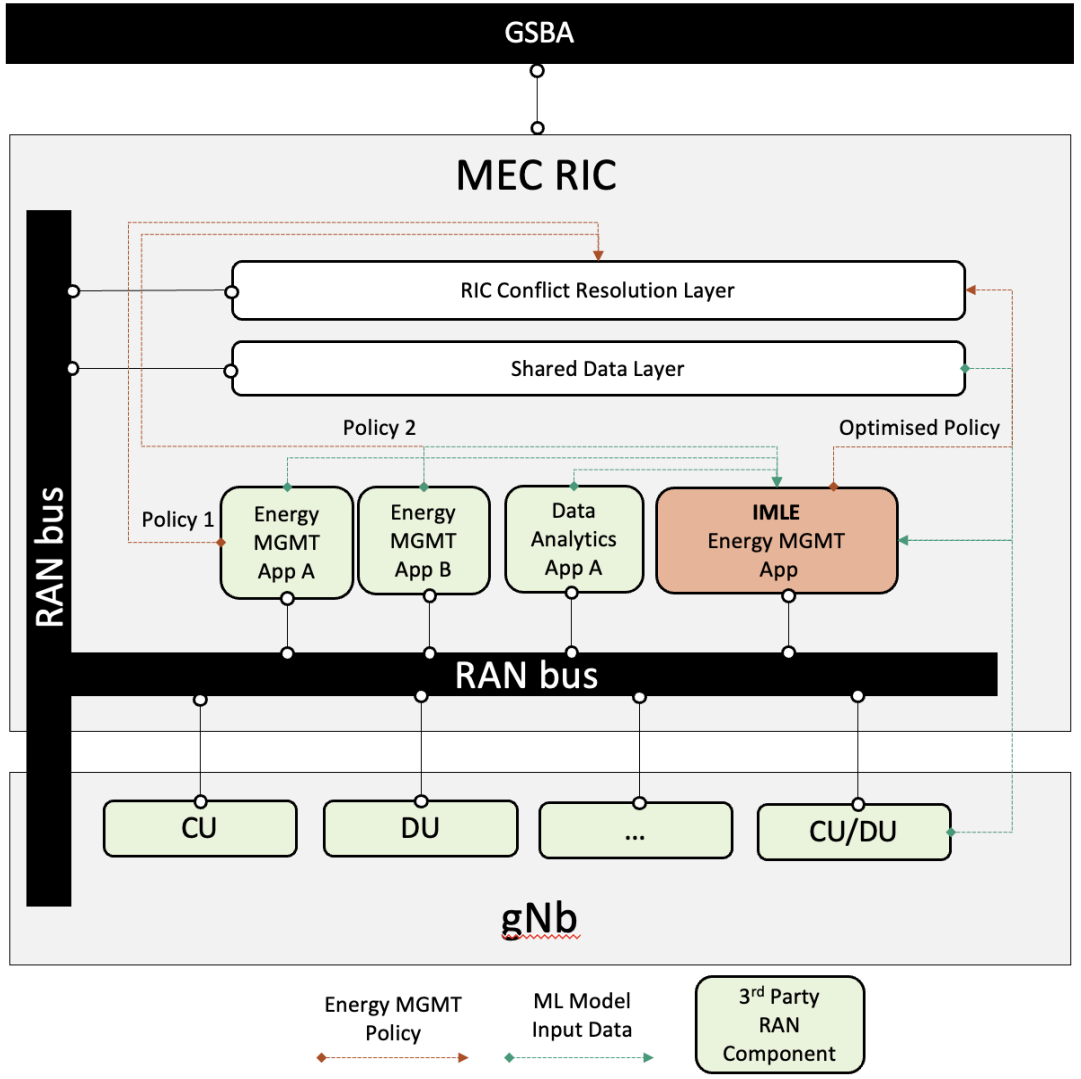
4. E2 interfaces for multiple PLMN and xApp

AI/ML

- GSBA enable faster AI/ML
  - The term AI/ML encompasses the data models, services, and primitives utilized for the life cycle management of ML algorithms, irrespective of their intended use and position within the network and the associated management domains.
- Related ORIGAMI Use Cases
  - Interoperable Machine Learning Models Improving RAN Energy Efficiency (IMLE)
  - Knowledge Representation (KR) for Anomaly Detection

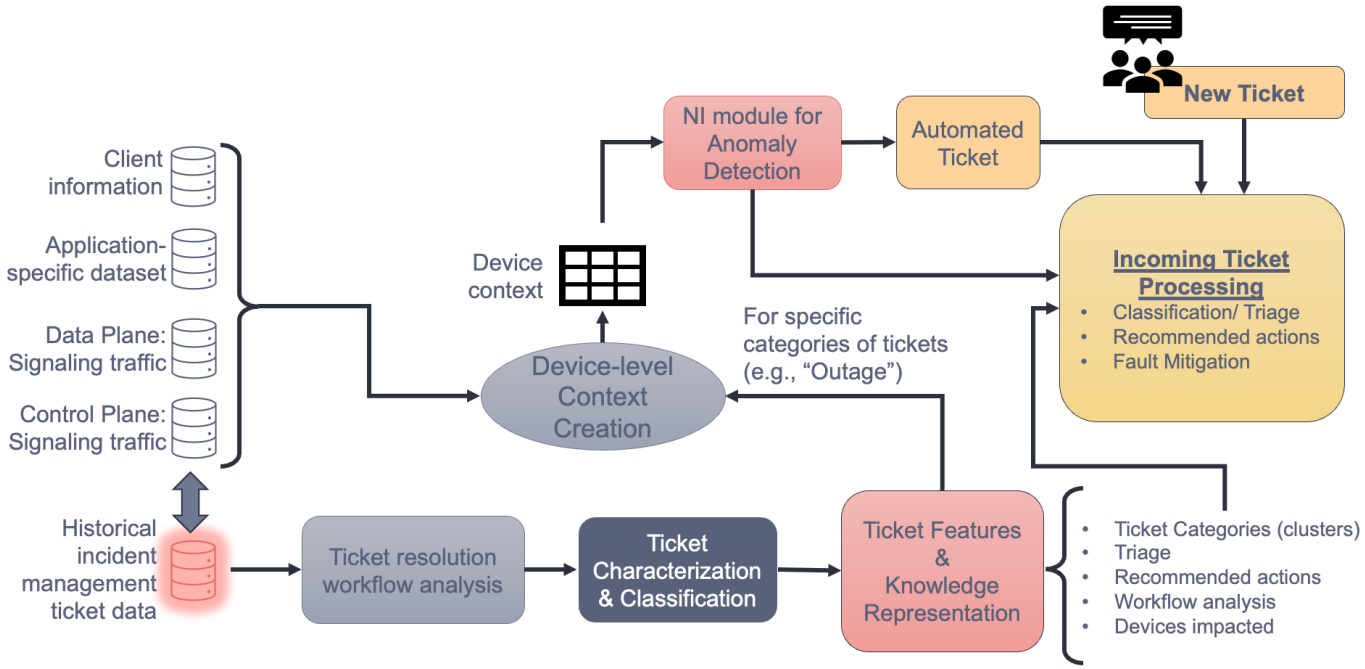
Interoperable Machine Learning Models Improving RAN Energy Efficiency [IMLE]

- IMLE will leverage the GSBA for AI models and services lifecycle management enabling faster and more reliable Network Intelligence to manage 6G network operations effectively.
  - IMLE offers highly accurate and optimized ML models and network NI solutions for energy saving within the ORIGAMI framework.
  - IMLE will strategically leverage the conflict resolution mechanisms, KPMs and data exposed by the GSBA to provide in real-time energy management policies.



Knowledge Representation [KR]

- In the context of the global operator model, KR will leverage the monitoring data model and propose AI/ML models for (proactive) **anomaly detection**
  - Leverage the GSBA for data collection
  - Focus on data processing, and represent the expert knowledge in features to guide the anomaly detection approaches
  - Closed-loop approach: anomaly detection solution to generate Incident Tickets within the incident management system of the GMNO

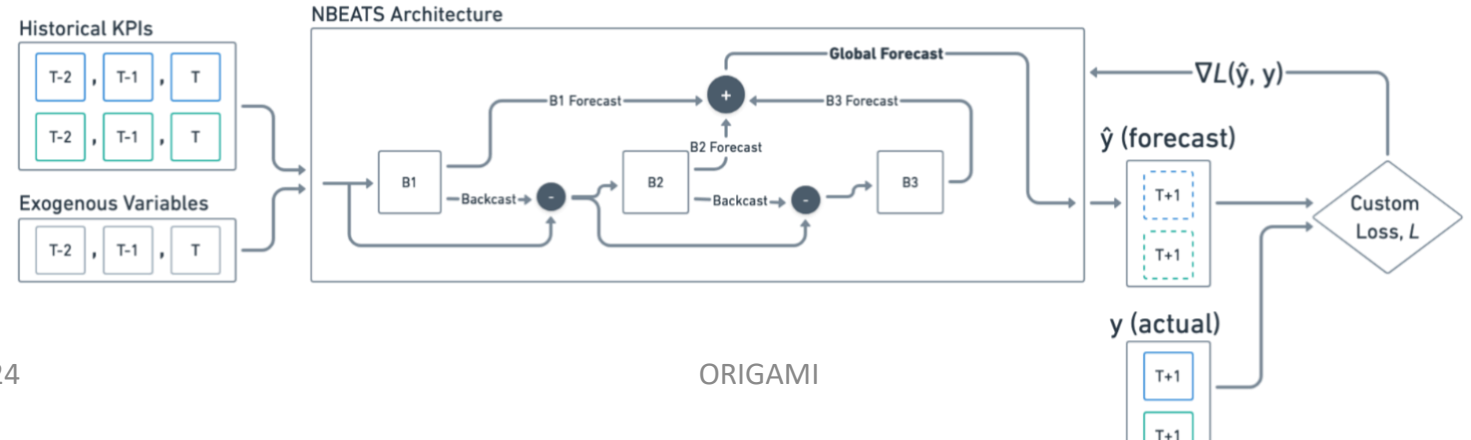


Energy Efficiency

- GSBA can be used to control Energy Efficiency
  - The terms "energy efficiency" and "energy saving" collectively encompass all functionalities and measurements employed for the purpose of monitoring and reducing energy consumption within the network domain.
- Related ORIGAMI Use Cases
  - Interoperable Machine Learning Models Improving RAN Energy Efficiency (IMLE)
  - Data-driven task offloading for reliable vRAN acceleration (SRV)

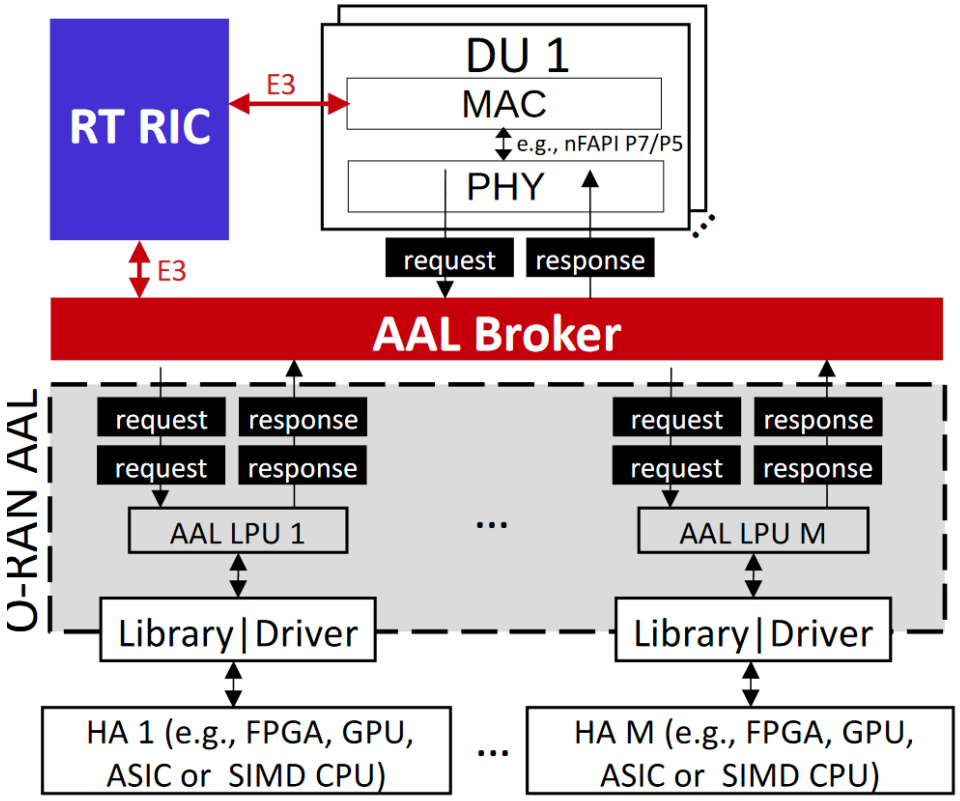
Interoperable Machine Learning Models Improving RAN Energy Efficiency [IMLE]

- IMLE will leverage the GSBA monitoring functionalities and measurements to provide NI solutions reducing energy consumption within the network domain
  - IMLE NI solution will rely on the NBEATS architecture and **interpretability** constraints will be incorporated.
  - Exogenous variables, such as additional KPIs or timestamps, will be incorporated as additional inputs to the network.
  - During training, a **custom loss function** that biases the neural network towards over-estimation will be employed to balance the trade-off between **energy saving**, and **resource under-provisioning** while **avoiding potential conflicts**.



Data-driven task offloading for reliable vRAN acceleration [SRV]

- Current softwarization strategies for vRANs are based on a very unflexible and monolithic approach, based on inline accelerators
- An opportunistic usage of Hardware Accelerators, based on the actual traffic characteristics (e.g., the size of the frame, or the channel conditions) can yield to a more efficient resource utilization.
- The interaction of the GSBA with the CCL allows the management system to set the operation of the AI/ML algorithms responsible for the



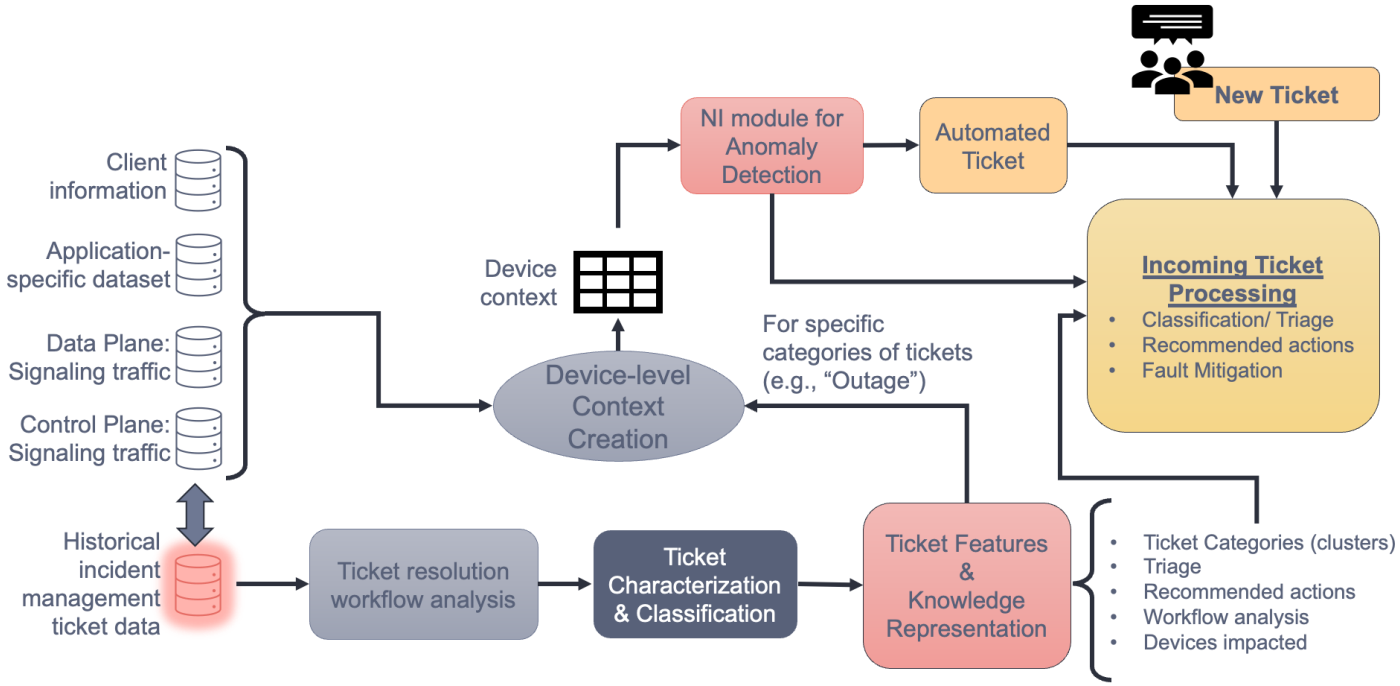
Network Performance Management

- GSBA eases the Network Performance Management
  - The term 'Service and Network Performance' is used to describe the functionalities and measurements that are useful for the assurance process at the service and network domains.
- Related ORIGAMI Use Cases
  - Knowledge Representation (KR) for Anomaly Detection
  - Network Core traffic analysis and optimization (NCAM)

Knowledge Representation [KR]

- KR enables network performance management via the GSBA
  - Focus on integration with the incident management system within the GMNO
  - Dimension the data collection and knowledge representation to achieve high efficiency in terms of processing overhead

- In the context of the global operator model, this use-case will leverage the monitoring data model and propose AI/ML models for (proactive) **anomaly detection**
  - We aim to focus on the data processing, and represent the expert knowledge in features to guide the anomaly detection approaches
  - The anomaly detection solution will generate Incident Tickets within the incident management solution of the GMNO



Network Core traffic analysis and optimization [NCAM]

Control plane performance monitoring and management

- ***Problem***: Rapid increase of the signaling load sets the need for advanced management and monitoring for the interactions among the Network Functions, including optimizations towards adaptive routing and traffic prioritization.
  - The rapid increase of signaling is a result of multiple factors, e.g.,:
    - Due to the explosion in the number of connected devices (UEs) and the need for multiplication of the NF instances to support the Network Slicing (NS) concept.
    - Due to the SBA which necessitates direct interactions among the Network Functions (NFs) of the network core, in a text-based manner (using HTTP).
- ***Proposal:*** Enable control plane monitoring to feed adaptive traffic management. Monitoring can be realized as integral part of the underlay infrastructure while automation in network core traffic management can be enabled with the AI toolbox.

Data Analytics

- GSBA facilitates Data Analysis
  - The term "data analysis" encompasses all functionalities and services employed for the purposes of network data collection, data storage, data elaboration, dataset production, and data visualisation, among other activities.
- Related ORIGAMI Use Cases
  - Knowledge Representation (KR) for Anomaly Detection

Knowledge Representation [KR]

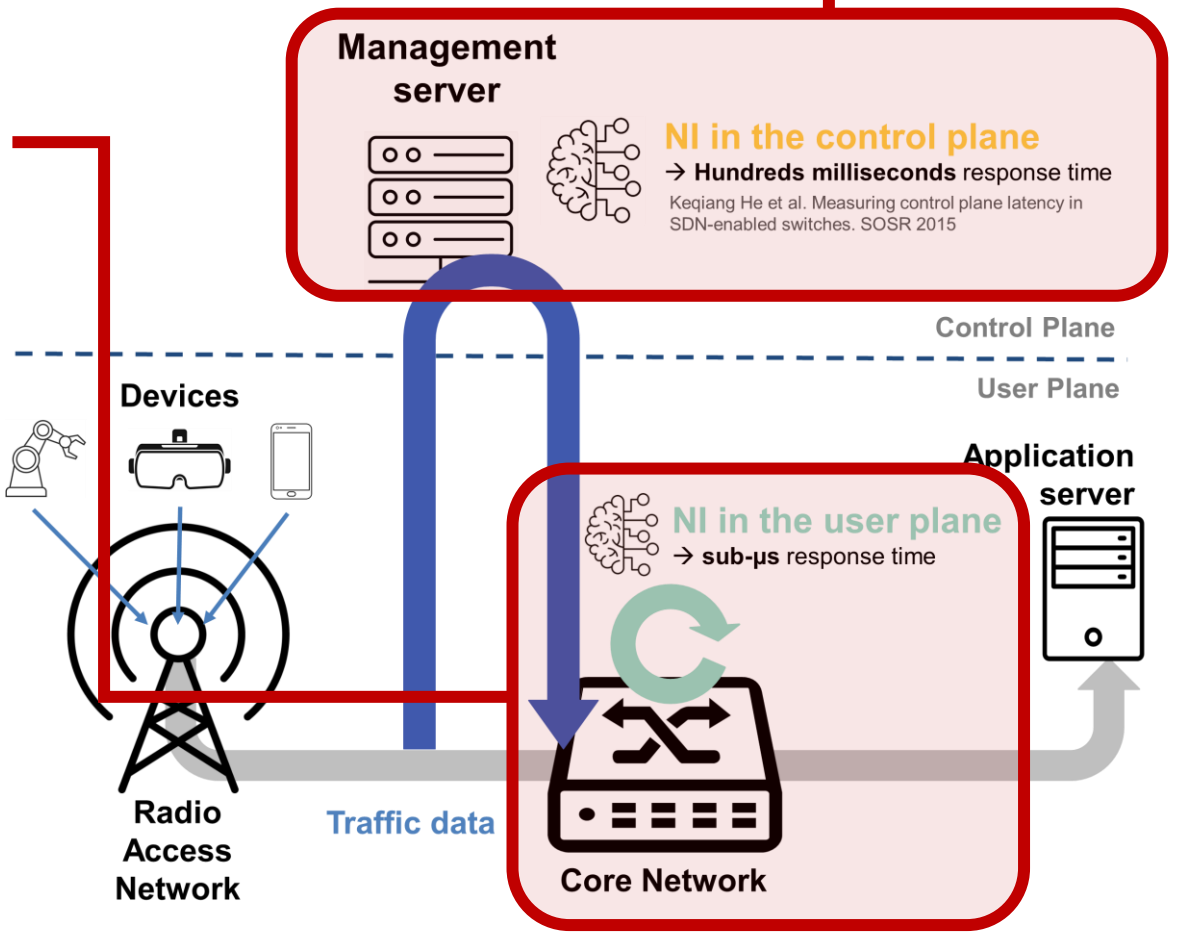
- The GSBA enables the collection of key (global) datasets and enables knowledge representations for anomaly detection in the context of a global operator
  - In the context of a global operator model, a data path that was once confined to one operator realm now builds on top of infrastructure from multiple distinct providers
- ORIGAMI aims to proactively detect anomalies within this global setting by monitoring the signaling traffic between entities controlling the resources that support the end-to-end path for the end-user
  - Transform these datasets and build an effective data representation
  - Engineer features based on expert knowledge
  - Strike optimal tradeoff between AI/ML model complexity and inference speed for proactive anomaly detection and incident management

Infrastructure Management

- GSBA enables Infrastructure Interworking
  - Infrastructure interworking can be defined as the set of data, functionalities, data models and architectures that facilitate the inter-domain orchestration activities, specifically within the context of network and infrastructure domains.
- Related ORIGAMI Use Cases
  - Effective, distributed and streamlined access to u-plane computing capabilities
  - Network Core traffic analysis and optimization

Distributed inference in the user plane [EAUC]

- Traditional Network Intelligence (NI) resides in the control plane
  - AI/ML models perform inference on data received from the user plane
  - Cross-plane loop causes 10-100 ms latency
- **User-plane inference removes the loop**
  - AI/ML models are deployed directly in programmable network hardware with P4
    - E.g., Intel Tofino switches, smartNICs, DPUs
  - NI operates at line rate at sub-µs latency
- Current state of the art
  - Random Forest (RF) on 20+ traffic classes
  - Demonstrations on real-world traffic classification, device identification, or intrusion detection



Distributed inference in the user plane [EAUC]

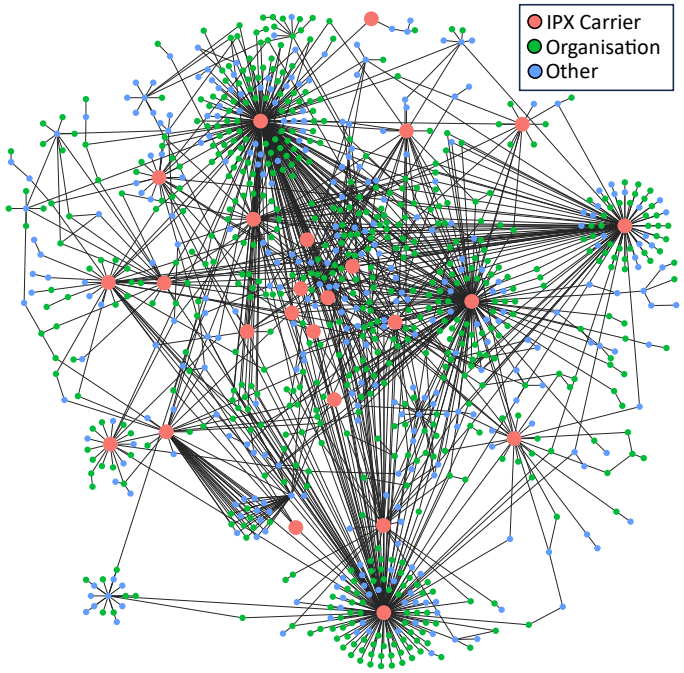
- User-plane NI is presently very ad-hoc
  - Requires manual designing of the model and lots of tinkering for the P4 implementation
  - Is deployed in a single device and does not leverage the distributed nature of networks
  - Needs starting from scratches for each different inference task or application use case
- **ORIGAMI will streamline user-plane NI programming**
  - Defines strategies to efficiently distribute a single AI/ML model across the network
    - Decomposes the target task in AI/ML sub-models each solving a subset of the original task
    - Deploys the AI/ML sub-models across multiple devices
  - Defines pipelines to automatically design and deploy P4 models for any inference task
  - Defines an abstraction layer to expose the programmable user-plane resources
    - Formalizes the capabilities of the (multiple, heterogeneous) programmable devices
    - Exposes such capabilities to the owners of any inference task
- **ORIGAMI makes user-plane capacity easy to use for line-rate tasks**
  - Favors intra-domain orchestration of cutting-edge infrastructure resources

Network Core traffic analysis and optimization [NCAM]

Service mesh infrastructure for the network core.

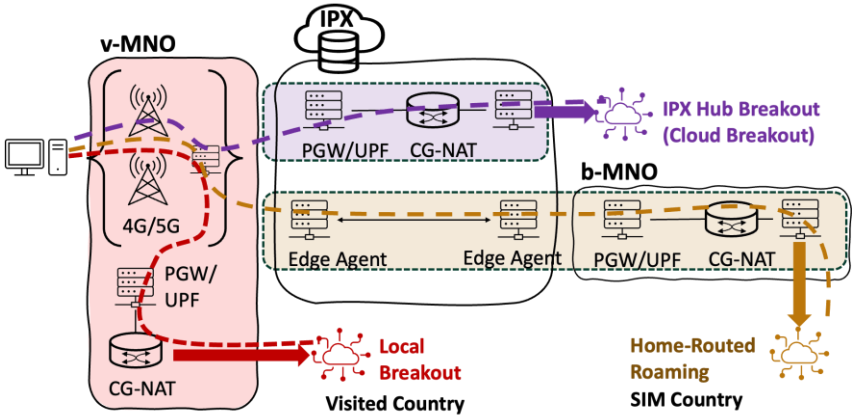
- ***Problem***: Each NF of the control plane is currently implemented as a monolithic application, which encapsulates several functions into one entity (e.g., the NF service implementation is waved with service exposure and networking functionality). This is one of the factors promoting failure points, affecting resilience and scalability of the network core communication model.
- ***Proposal:*** Implement a service-mesh architecture that realizes a distributed Service Communication Proxy (SCP) model for the network core. The service exposure and networking functionality of each NF is offloaded to a sidecar. The sidecars are subject to a service mesh infrastructure guaranteeing resilience and scalability

Enabling the Global Operator Model [GMNO]



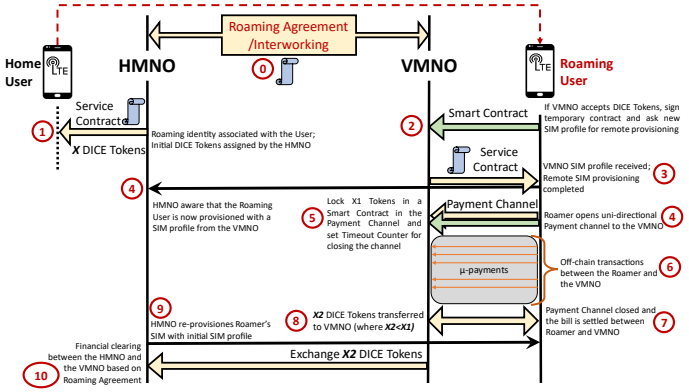
Topology Mapping for the Cellular Ecosystem

- Radio Resources (MNOs)
- Cloud Resources
- Interconnection Infrastructure



The GSBA support the design of the ***Optimal Breakout Selection*** (Cloud Breakout):

- Standards-compliant architecture that achieves the performance benefits of local breakout
- Constraints-aware cost model for deploying the global infrastructure



The ZTL will support the exploration of ***permissioned consortium blockchain*** for global operations through the DICE protocol:

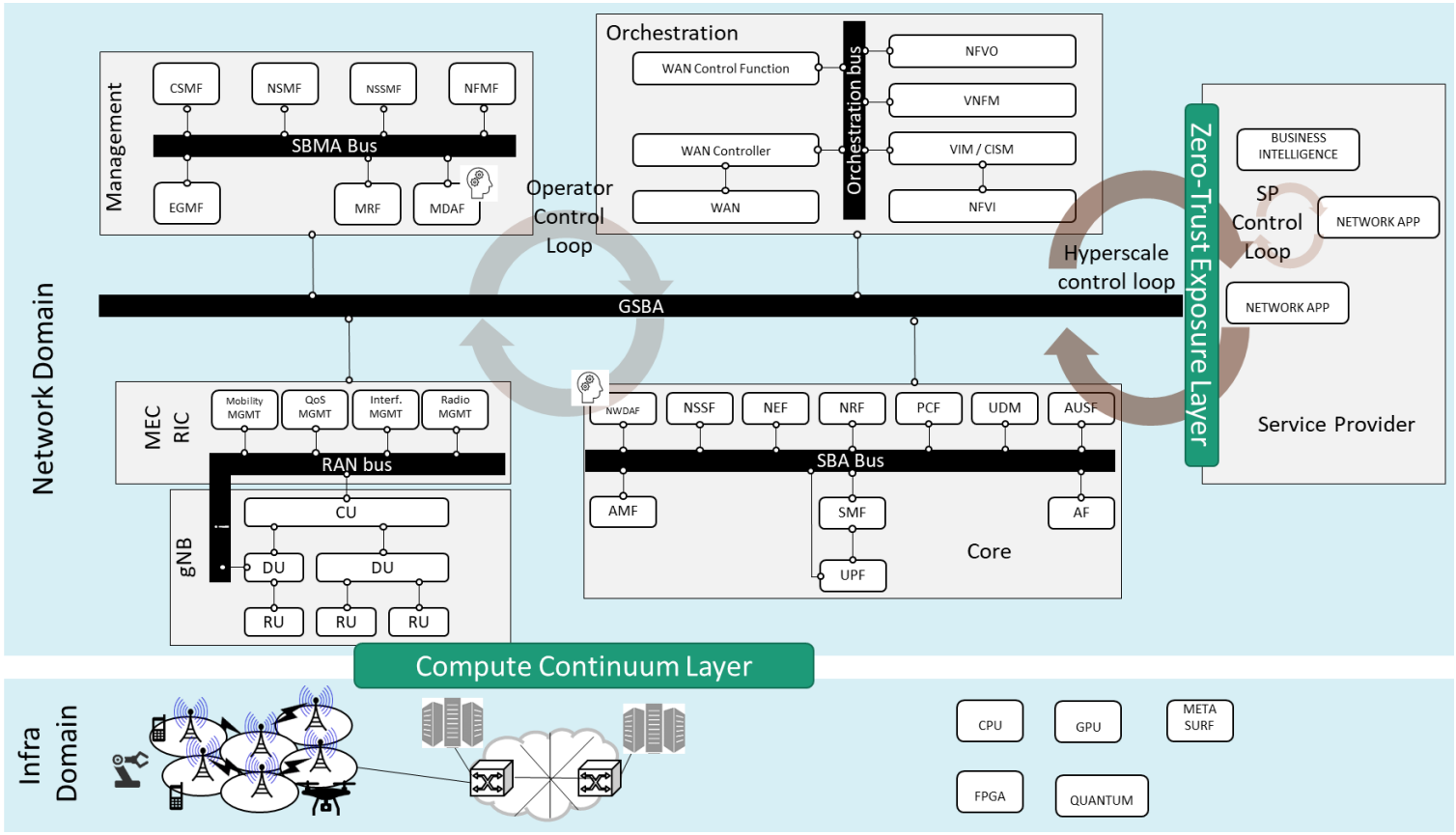
- Zero-trust interactions between resource providers
- Bypass the use of a third party (DCH) for data and financial clearing
- Immutability of global transactions (tamper-proof ledger)
- Confidentiality of the agreements and the transactions

[TMA'24] Viktoria Vomhoff, Marleen Sichermann, Stefan Geißler, Andra Lutu, Martin Giess, and Tobias Hoßfeld. "A Shortcut through the IPX: Measuring Latencies in Global Mobile Roaming with Regional Breakouts."

Co-funded by the European Union

Zero Trust Layer

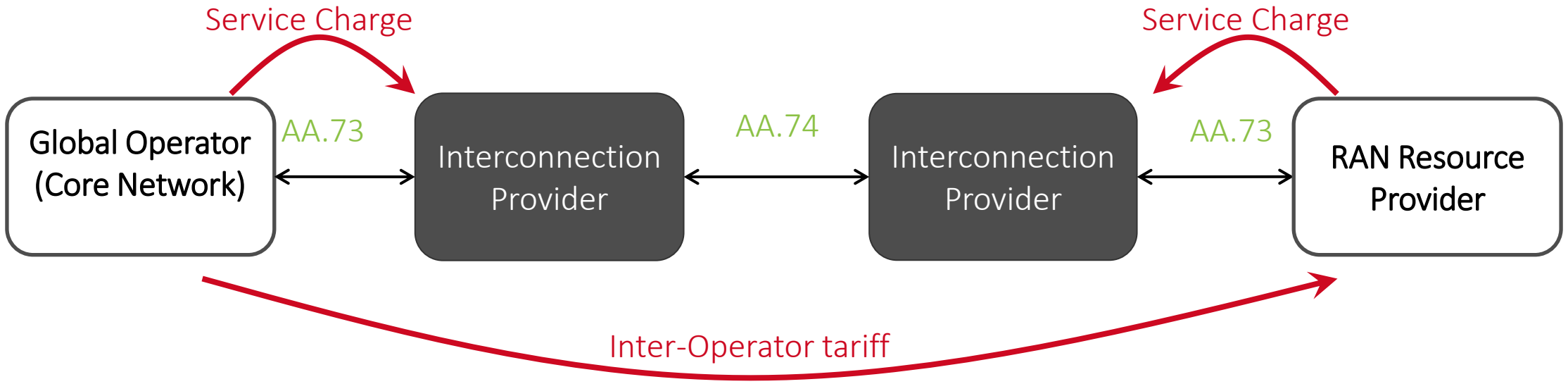
- The ORIGAMI architecture aims to enable zero-trust interactions through the ZTL between entities that support global services, and tackle the limitations of the current trust model within the ecosystem



Zero Trust Layer

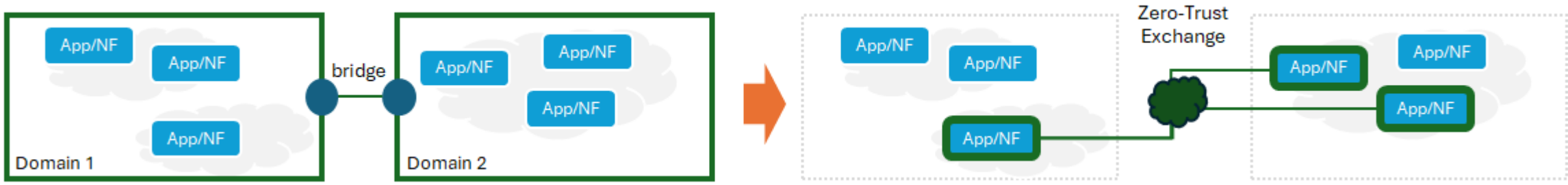
The ORIGAMI ZTL brings two major changes to the ecosystem:

- Enables MNOs to have a dynamic cooperation and enable emerging models of operators (e.g., MNAs such as emnify) establish relationships and perform financial clearing in almost real time
  - The current process using Data Clearing Houses is slow, and money move slowly across geographies
- Enables the end-user to have control over its mobile connection
  - Currently, the end–user is locked-in only with one operator and migrating (though technically possible) is cumbersome and risky



ZTL

- Through Zero-Trust exposure Layer ZTL layer ORIGAMI tackles the following barriers:
  - #5 (“Lack of global service APIs”)
    - Exposure APIs / Frameworks (there are multiple APIs that refer to third party service providers e.g., CAMARA, but the transformation of those to technology specific APIs at the network domain is still a gap)
  - #6 (“Obsolete trust model hinders performance”)
    - pre-provisioned global connectivity (IoT-like model is needed where connectivity is guaranteed, and the differentiation is realized service layer – Each application/service operates under the zero-trust model connectivity does not guarantee trust)
  - #7 (“Inadequate networking data representation”)
    - data representation and data governance: pipeline from the (global) infrastructure providers to the tenants that is fully focused on “data” provisioning. This enables automation processes through AI solutions.

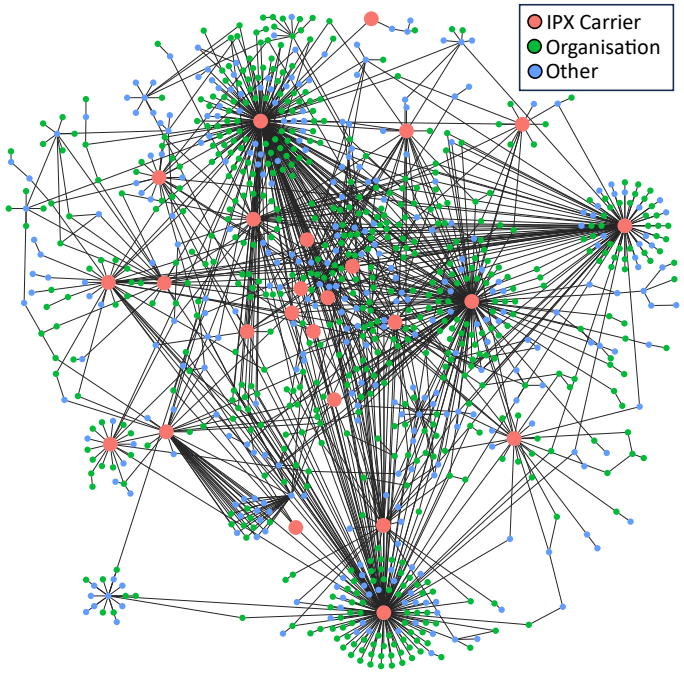


Proposal: Move from traditional firewall-centric architectures that bridge the users/third party applications and the network functions at connectivity layer (trusted and untrusted domains are defined) towards → a full TLS/SSL inspection at scale, with connections brokered peer-to-peer between users/ third party applications and the network functions based on identity, context, and business policies

RAN Architecture

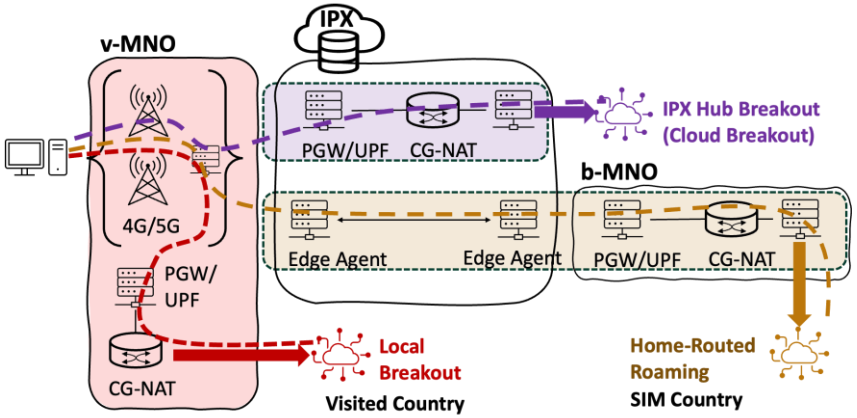
- The ZTL has a large potential impact on the network architecture
  - The term 'network architecture' is used to describe the various functions, interfaces, protocols and services that collectively define the network domain. In particular, the radio access network, the core network and the transport network are all examples of network architecture.
- Related ORIGAMI Use Cases
  - Enabling the Global Operator Model

Enabling the Global Operator Model [GMNO]



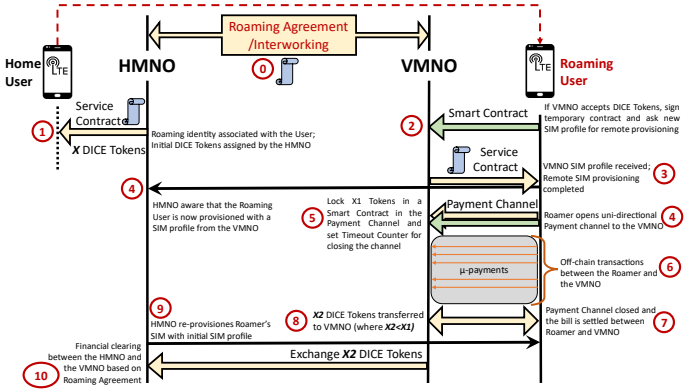
Topology Mapping for the Cellular Ecosystem

- Radio Resources (MNOs)
- Cloud Resources
- Interconnection Infrastructure



The GSBA support the design of the ***Optimal Breakout Selection*** (Cloud Breakout):

- Standards-compliant architecture that achieves the performance benefits of local breakout
- Constraints-aware cost model for deploying the global infrastructure



The ZTL will support the exploration of ***permissioned consortium blockchain*** for global operations through the DICE protocol:

- Zero-trust interactions between resource providers
- Bypass the use of a third party (DCH) for data and financial clearing
- Immutability of global transactions (tamper-proof ledger)
- Confidentiality of the agreements and the transactions

[TMA'24] Viktoria Vomhoff, Marleen Sichermann, Stefan Geißler, Andra Lutu, Martin Giess, and Tobias Hoßfeld. "A Shortcut through the IPX: Measuring Latencies in Global Mobile Roaming with Regional Breakouts."

Co-funded by the European Union

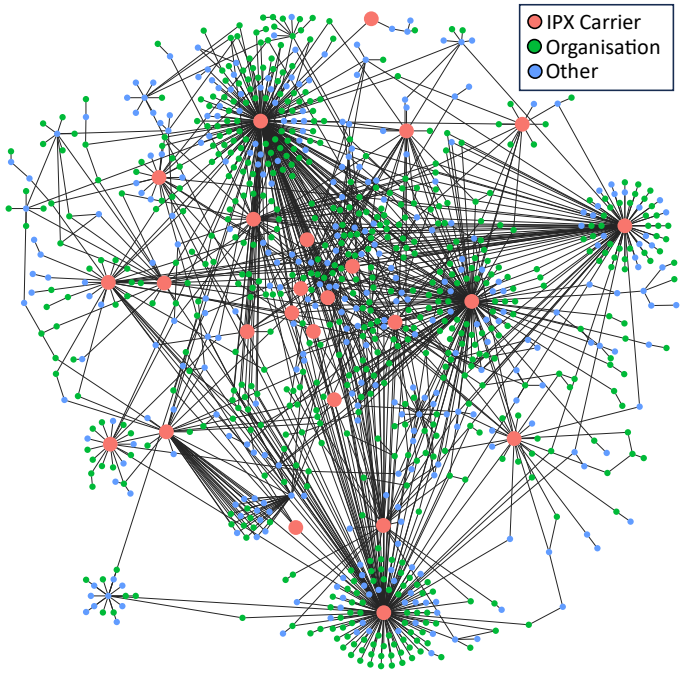
Enabling the Global Operator Model [GMNO]

- The 5G local breakout wholesale billing model is an after-though
  - In the context of a global IoT managed connectivity provider: IoT device data communications ***should*** break out in the local country where they operate
  - Visited networks push to charge each IoT Provider based on the **5G connection and mobility usage of their IoT devices**
- Local breakout 5G wholesale billing models still suffer from the **lack of trust between 5G roaming partners**
  - In an LBO setup, the IoT provider cannot verify the validity of the billing records from the visited operator
- ZTL enables the global operator to interact with any RAN resource provider, without the need of a centralized unit to manage the interconnection (and the access to the radio resources) and arbitrate cases of billing disputes

Network Resource Model

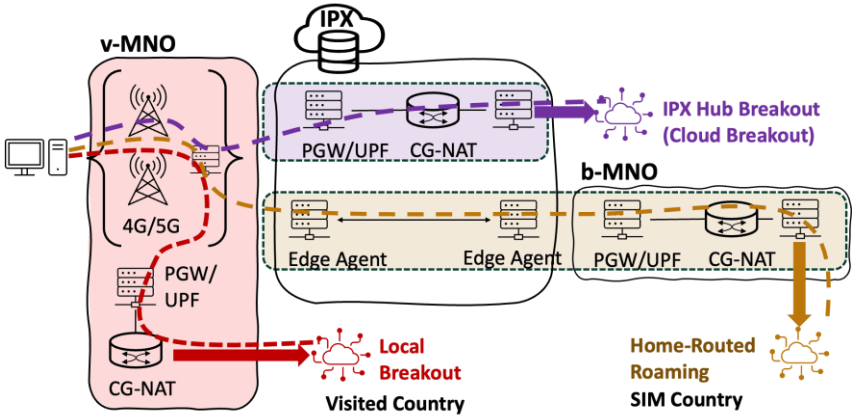
- ZTL enables new Infrastructure Interworking models
- Related ORIGAMI Use Cases
  - Enabling Global Operator Model

Enabling the Global Operator Model [GMNO]



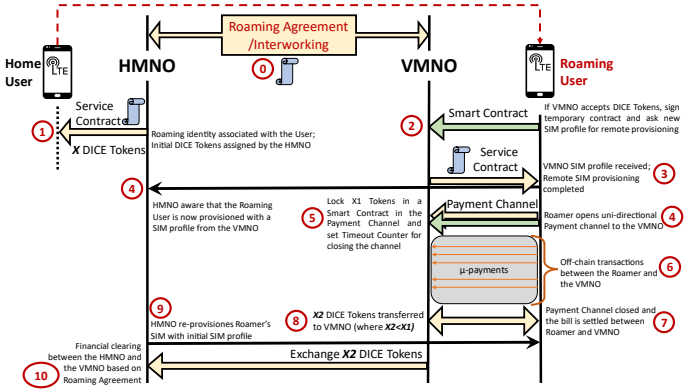
Topology Mapping for the Cellular Ecosystem

- Radio Resources (MNOs)
- Cloud Resources
- Interconnection Infrastructure



The GSBA support the design of the ***Optimal Breakout Selection*** (Cloud Breakout):

- Standards-compliant architecture that achieves the performance benefits of local breakout
- Constraints-aware cost model for deploying the global infrastructure



The ZTL will support the exploration of ***permissioned consortium blockchain*** for global operations through the DICE protocol:

- Zero-trust interactions between resource providers
- Bypass the use of a third party (DCH) for data and financial clearing
- Immutability of global transactions (tamper-proof ledger)
- Confidentiality of the agreements and the transactions

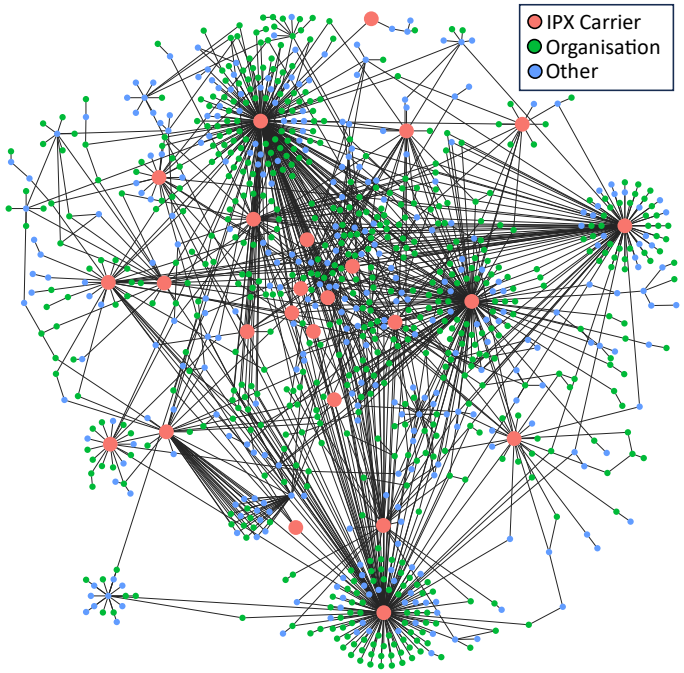
[TMA'24] Viktoria Vomhoff, Marleen Sichermann, Stefan Geißler, Andra Lutu, Martin Giess, and Tobias Hoßfeld. "A Shortcut through the IPX: Measuring Latencies in Global Mobile Roaming with Regional Breakouts."

Co-funded by the European Union

Security aspects

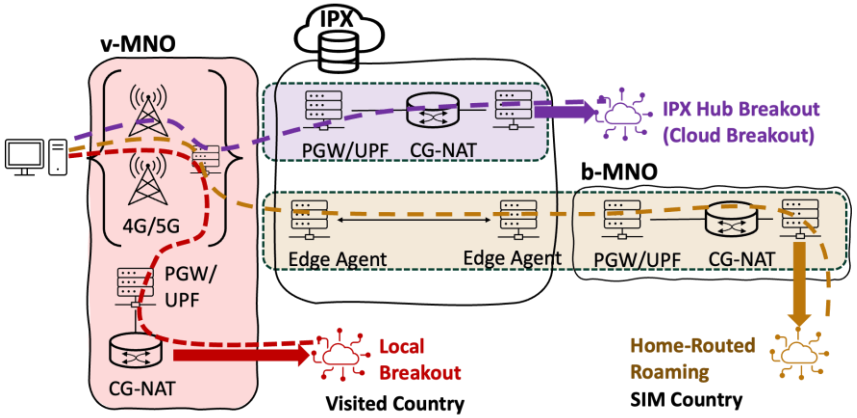
- ZTL enhances Security for UCs
- Related ORIGAMI Use Cases
  - Limited Trust Network Analytics
  - Global Operator Model

Enabling the Global Operator Model [GMNO]



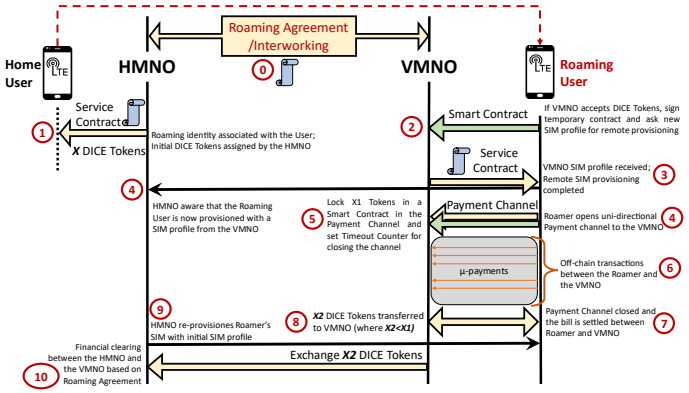
Topology Mapping for the Cellular Ecosystem

- Radio Resources (MNOs)
- Cloud Resources
- Interconnection Infrastructure



The GSBA supports the design of the ***Optimal Breakout Selection*** (Cloud Breakout):

- Standards-compliant architecture that achieves the performance benefits of local breakout
- Constraints-aware cost model for deploying the global infrastructure



The ZTL will support the exploration of ***permissioned consortium blockchain*** for global operations through the DICE protocol:

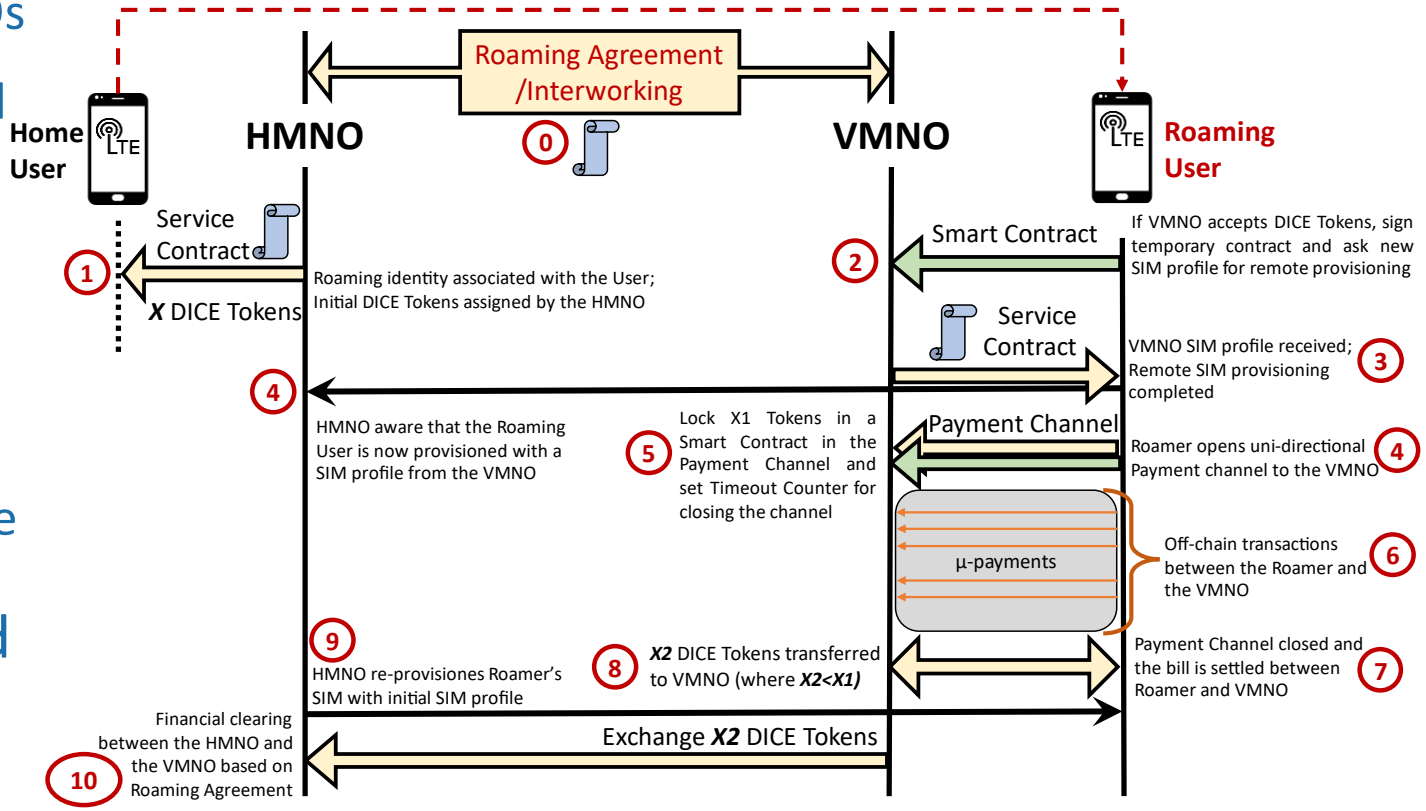
- Zero-trust interactions between resource providers
- Bypass the use of a third party (DCH) for data and financial clearing
- Immutability of global transactions (tamper-proof ledger)
- Confidentiality of the agreements and the transactions

[TMA'24] Viktoria Vomhoff, Marleen Sichermann, Stefan Geißler, Andra Lutu, Martin Giess, and Tobias Hoßfeld. "A Shortcut through the IPX: Measuring Latencies in Global Mobile Roaming with Regional Breakouts."

Co-funded by the European Union

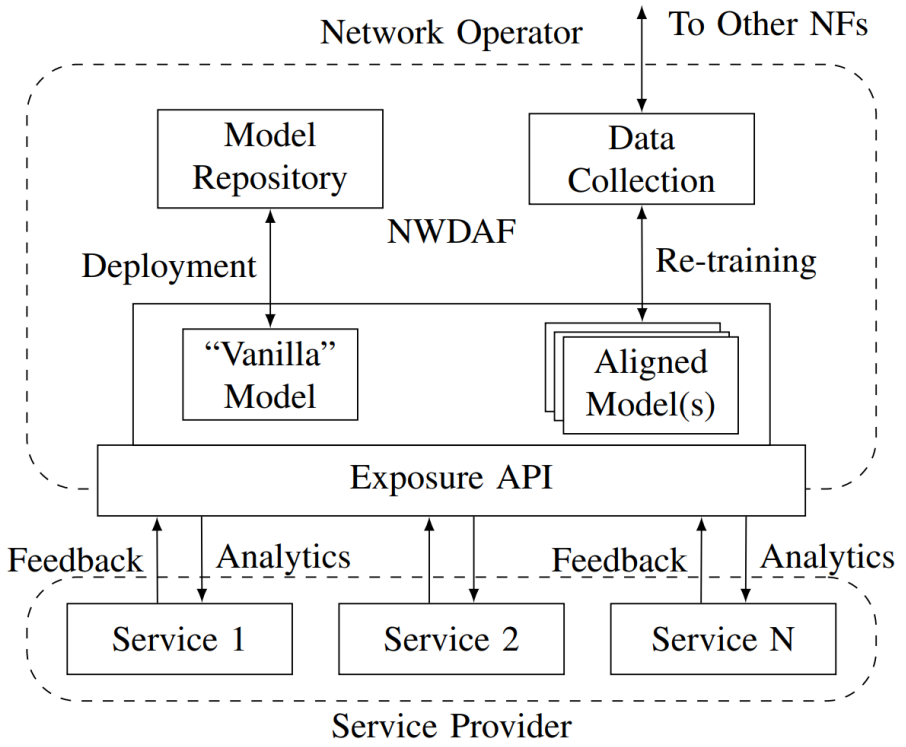
GMNO (focus view on the message exchange)

- Build trust between entities within the cellular ecosystem
  - Integrate the agreements between MNOs in the blockchain, enable the ability to renegotiate agreements dynamically and find roaming partners
- Avoid the use of a third party (DCH) for data and financial clearing
- Immutability of transactions (tamper-proof ledger)
  - The HMNO cannot delete or modify the interactions between the roamer and the VMNO
- Confidentiality of the agreements and the transactions
  - Permissioned blockchain, where we can also inject fake transactions in order to hide information from other participants



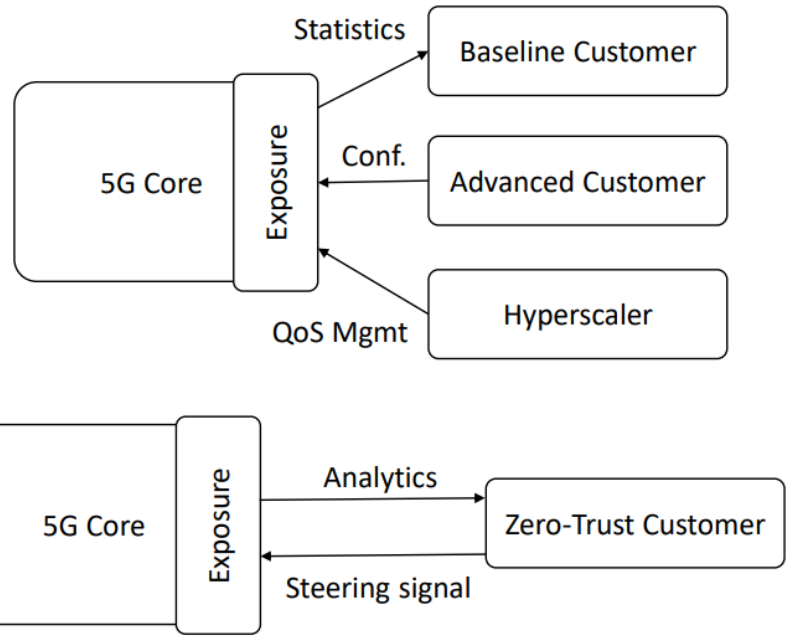
Limited Trust Network Analytics [LTNA]

- The Limited Trust Network Analytics Use Case, enforce the cooperation between service provider and network operators, providing an API for the production of network analytics without the exchange of raw data



Limited Trust Network Analytics [LTNA]

- Through the API, the 5GC can evolve the current interactions (Statistics, Configuration, and QoS Management) to a third class, where the Zero Trust Customer can help the Operator to provide more refined analytics
- This behaviour increases the security of the system, allowing third parties to interact with the 5G Core without directly integrating probes or softwares.

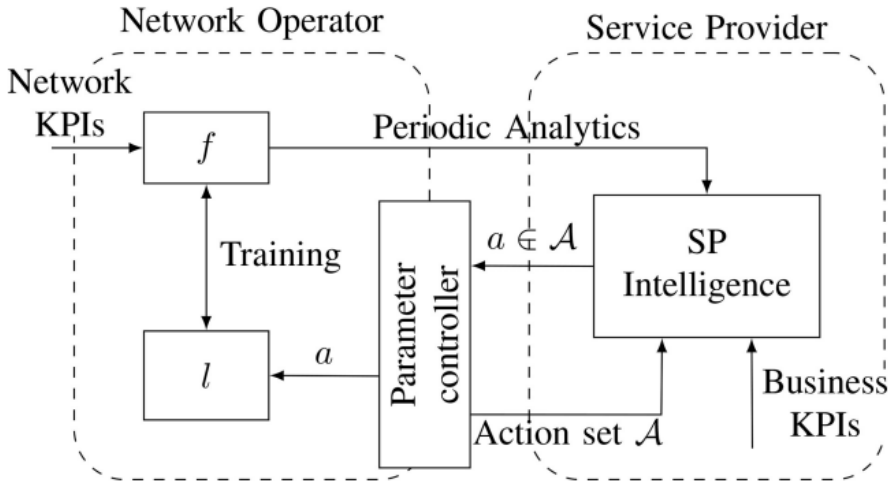


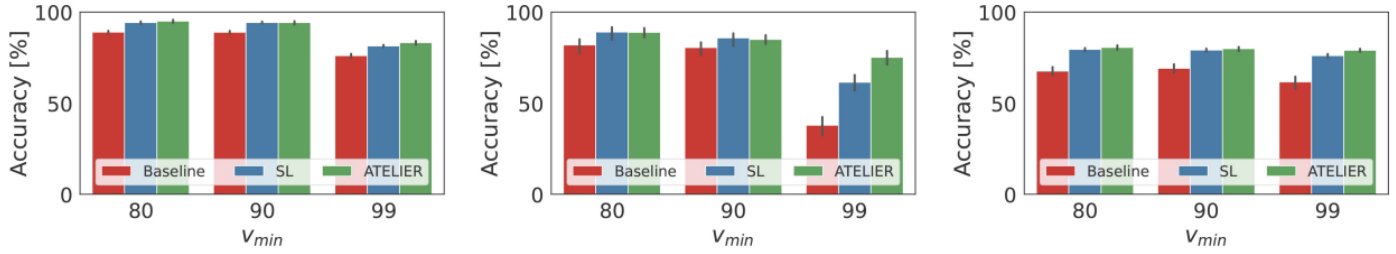
Data Analytics

- ZTL enables enhanced Data Analytics
- Related ORIGAMI Ucs
  - Limited Trust Network Analytics

Limited Trust Network Analytics [LTNA]

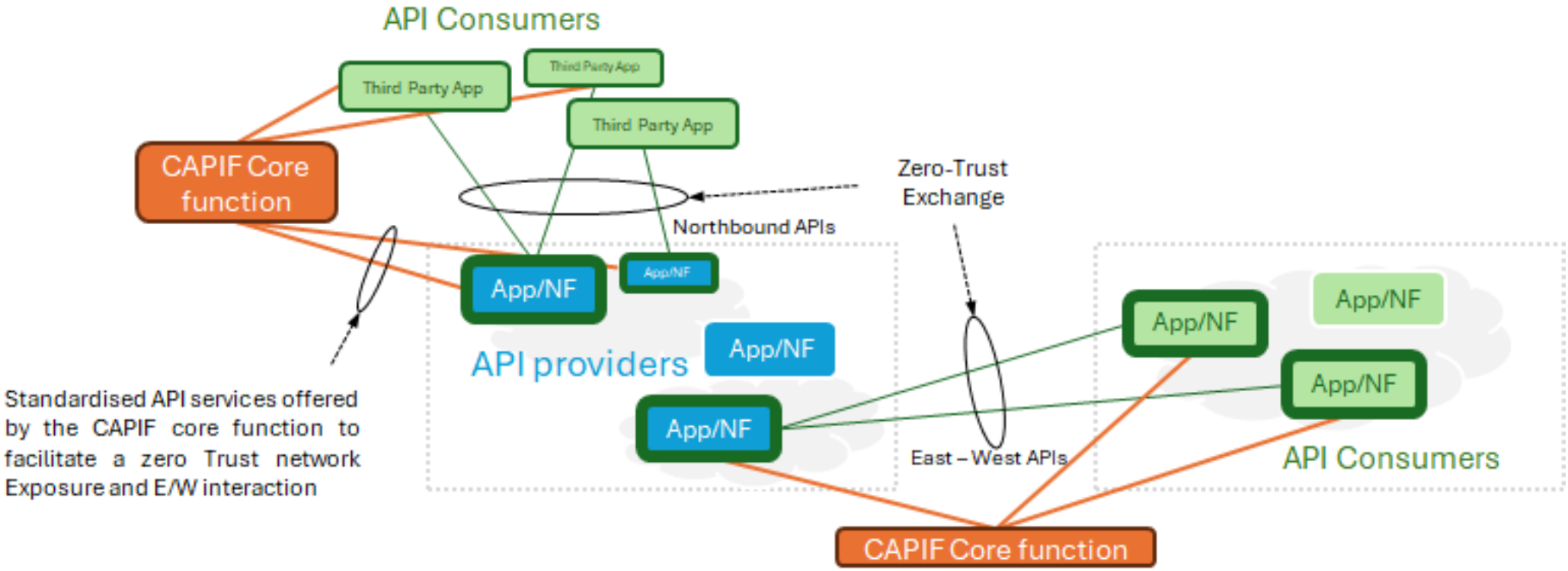
- Through the introduction of the LTNA, the Network Operator can provide better network analytics thanks to the integration of the Service Provider point of view
- In our test, we could improve the performance by a factor of 10% compared to a baseline solution





(a) Bunny (b) Bottle (c) Scarlet

CAPIF (Dimitris)



ZTL approach: Full TLS/SSL inspection at scale, with connections brokered peer-to-peer between users/ third party applications and the network functions based on identity, context, and business policies.

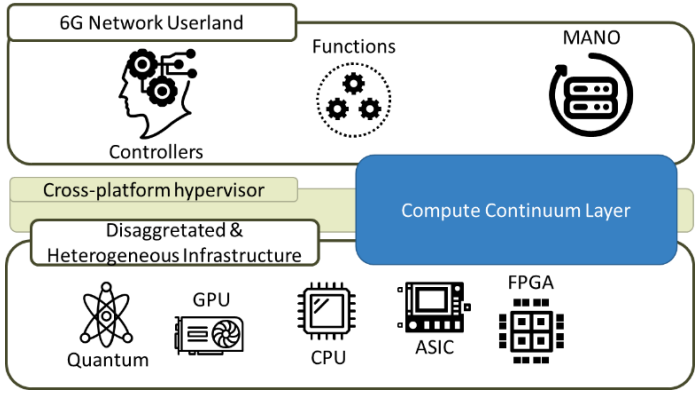
Common API Framework - CAPIF is a key enabler for ZTL adoption in mobile networks.

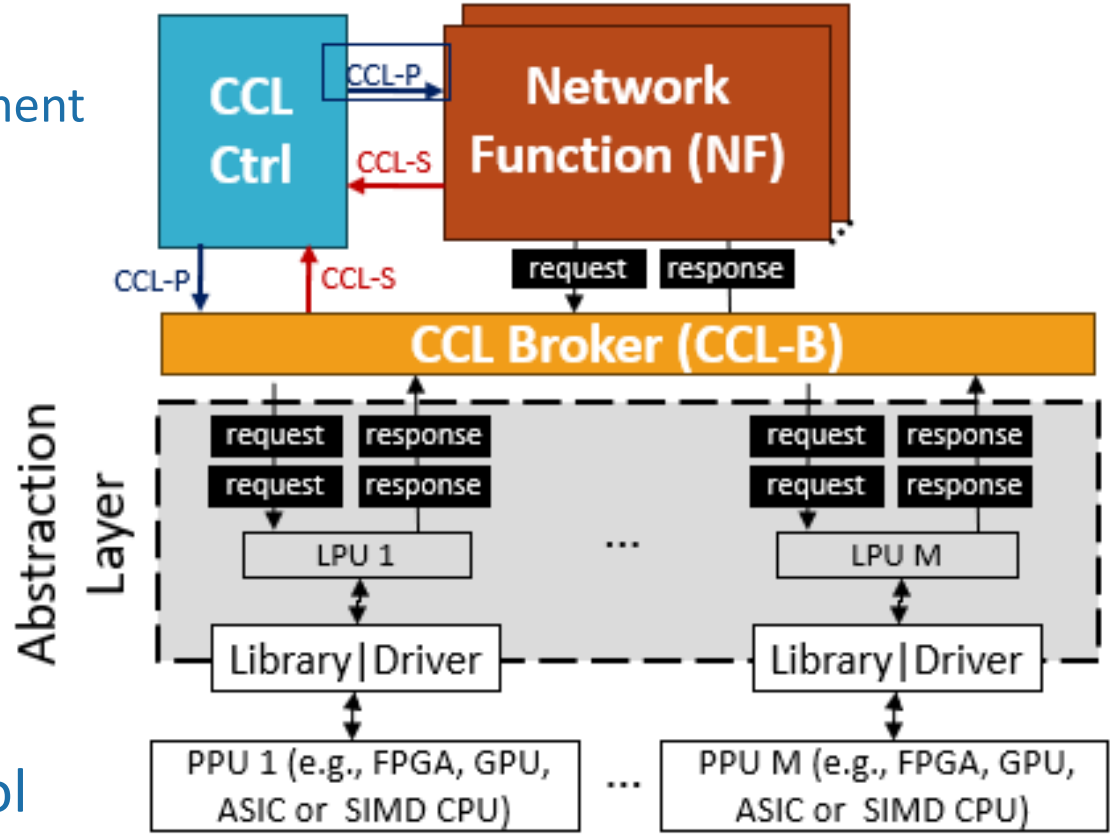
CAPIF (as specified in 3GPP, TR 23.222, TR 29.222, TR 33.122 and implemented by ETSI SDG OpenCAPIF) is recommended as a standardized API manager to facilitate Northbound and East-West API services since it offers:

- Interoperability (the functionality is standardized)
- Discoverability (visibility of the offered service API from the providers to any CAPIF-register API Consumer)
- A valuable toolbox for the API provider (API usage logs, monitoring and auditing capabilities are provided)

CCL

- The Compute Continuum Layer (CCL)
  - CCL Broker: Abstractions of heterog. computing resources
  - CCL ctrl: Compute-aware NF-control & management





- Real-time CCL-x interfaces
  - Sub-ms latency is expected
- Compute awareness is required in NF control
  - To avoid saturation of computing resources
  - To provide compute latency guarantees

Energy Efficiency

- CCL can improve energy efficiency
- Related ORIGAMI Use Cases
  - Interoperable Machine Learning Models Improving RAN Energy Efficiency
  - Data-driven task offloading for reliable vRAN acceleration

Interoperable Machine Learning Models Improving RAN Energy Efficiency [IMLE]

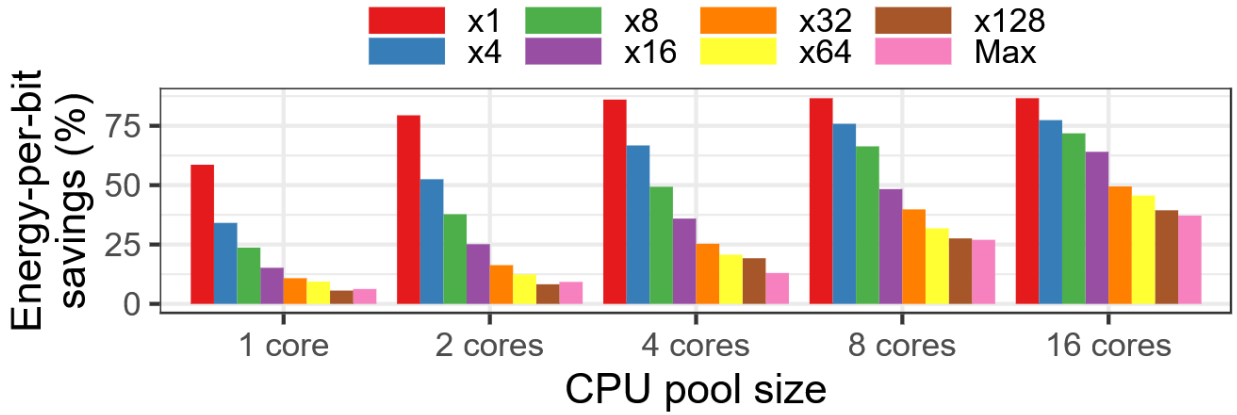
- IMLE will leverage the CCL hardware abstraction (GPU/TPU/CPU) to provide interoperable NI solutions for the purpose of reducing energy consumption within the RAN domain.
- This will enable efficient model training and real-time inference.

Data-driven task offloading for reliable vRAN acceleration [SRV]

- Current softwarization strategies for vRANs are based on a very unflexible and monolithic approach, based on inline accelerators

- CCL broker enables opportunistic usage of heterogeneous hardware accelerators to save energy

- CCL ctrl enables DU pooling to save energy (less servers)
  - Bounds the amount of radio resources available for DU scheduling, controlling the demand of computing resources
  - Avoids saturation of computing resources
  - Guarantees reliability



AI/ML

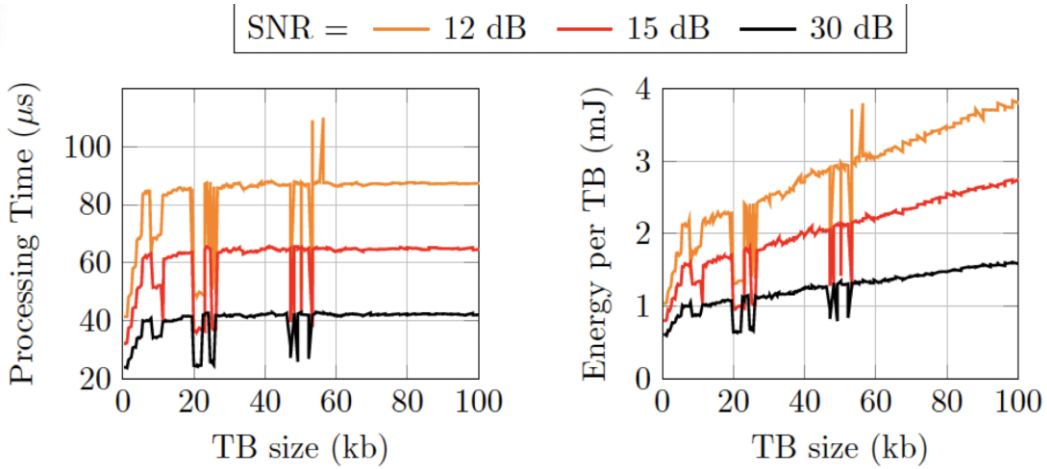
- CCL enables AI/ML Mechanisms
- Related ORIGAMI Use Cases
  - Interoperable Machine Learning Models Improving RAN Energy Efficiency
  - Compute- and Fairness-Aware Radio Resource Allocation Algorithms in Virtualized RANs

Interoperable Machine Learning Models Improving RAN Energy Efficiency [IMLE]

- IMLE will leverage the CCL hardware abstraction (GPU/TPU/CPU) to provide interoperable NI solutions for the purpose of reducing energy consumption within the RAN domain.
- This will enable efficient model training and real-time inference.

CFA

- CFA defines thresholds for minimum TB sizes in near-RT RIC.
- Processing time and energy per bit decreases as TB size grows.



Bundle smaller TBs together by adding a TB size threshold for transmission. → **Reduced energy consumption, Increased delay.** → **Our framework ensures the fair distribution of delays!**

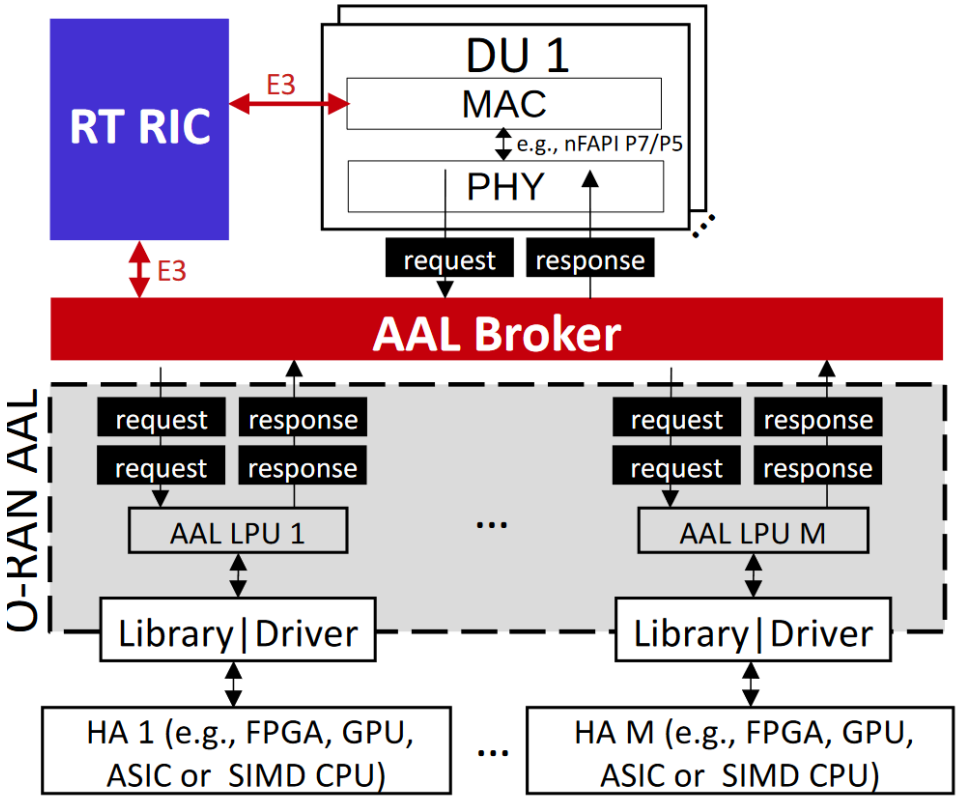
- CFA leverages AI/ML interface of CCL to produce the decisions using Online Convex Optimization.

RAN architecture

- CCL is a novel component in the Network Architecture
- Related ORIGAMI Use Cases
  - Effective, distributed and streamlined access to u-plane computing capabilities
  - Data-driven task offloading for reliable vRAN acceleration

Data-driven task offloading for reliable vRAN acceleration [SRV]

- Current softwarization strategies for vRANs are based on a very unflexible and monolithic approach, based on inline accelerators
- CCL broker (AAL broker): Enables opportunistic usage of heterogeneous hardware accelerators
- CCL ctrl (RT RIC): Enables DU pooling to save CAPEX/OPEX
  - Bounds the amount of radio resources available for DU scheduling, controlling the demand of computing resources
  - Avoids saturation of computing resources
  - Guarantees reliability



Network Digital Twin

- CCL enables DT of the RAN for Infrastructure Interworking
- Related ORIGAMI Use Cases
  - Effective, distributed and streamlined access to u-plane computing capabilities
  - Data-driven task offloading for reliable vRAN acceleration

Data-driven task offloading for reliable vRAN acceleration [SRV]

- The task offloading strategies implemented through the CCL require hardware abstraction to efficiently exploit resource pooling at the edge
  - Different Hardware Accelerators (x86 CPUs, ARM CPUs, GPUs, Quantum Computing Platform) require a common layer for basic wireless operation, decoupled from the actual execution platforms.
- To train the AI/ML algorithm that empower the Use Case, the usage of Network Digital Twins of the specific Hardware Accelerators
- The management system can deal with this aspects, triggering online training requirements or re-training of the solutions

Conclusion

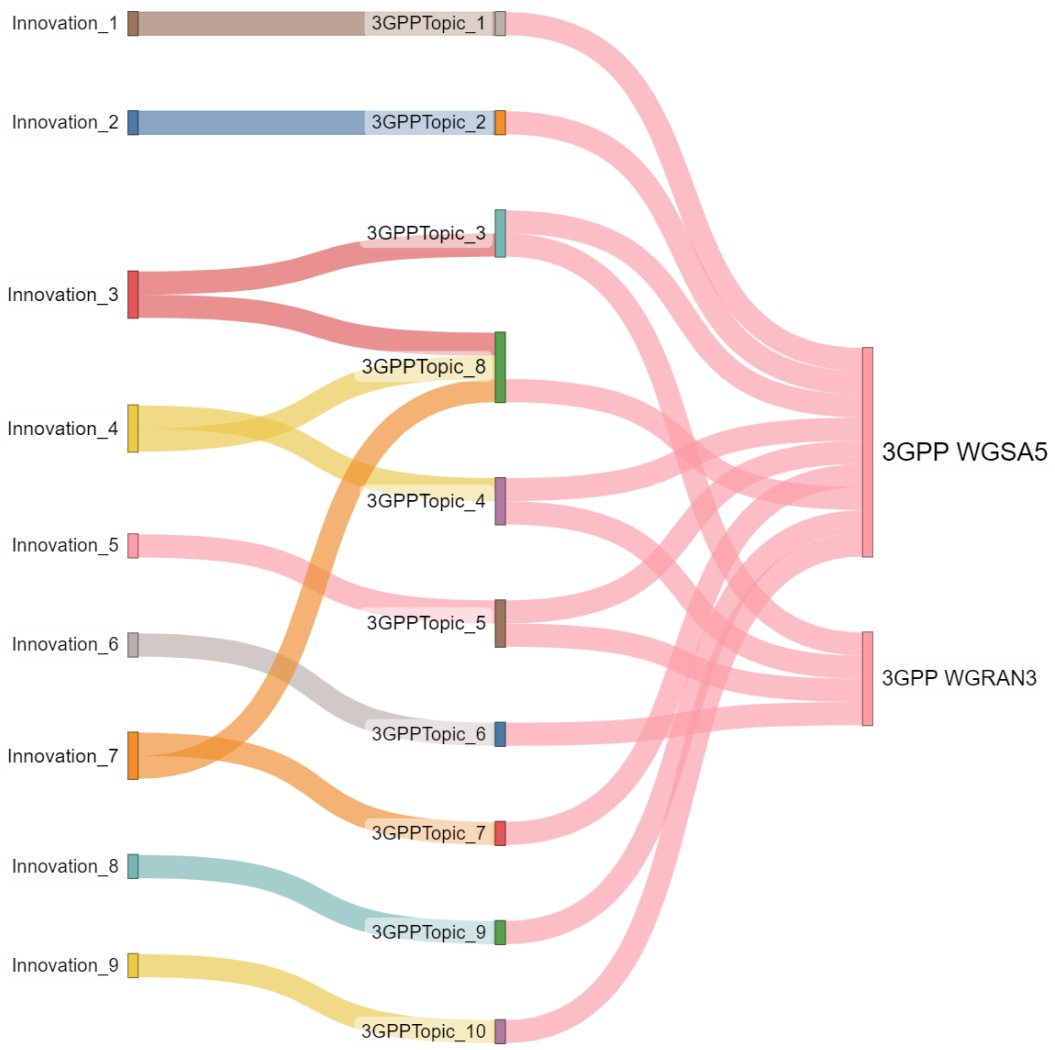
- This is the vision ORIGAMI has for 6G architectural framework as a whole, not only the algorithmic solutions
- Motivated by the use cases we are working towards the architectural requirements introduced by the three novel components
  - GSBA: Global SBA Architecture, bringing the SBA paradigms to all network domains, including RAN
  - CCL: Compute Continuum Layer, to efficiently handle fast scale
  - ZTL: Zero-Trust Layer, to handle interaction between network operators and third parties in a seamless way
- We believe that these modules have the potential to impact on relevant 3GPP topic, as discussed previously

Origami Innovation Mapping to 3GPP contributions [1/2]

#	Innovation Type	Definition
Innovation_1	**Resource Abstraction**	The term "resource abstraction" is used to describe a data model that is capable of representing network and service-related resources in a manner that is independent and general, regardless of the specific solutions that have been implemented.
Innovation_2	**Orchestration**	The term 'orchestration' is used to describe the entire data set, the full range of services and the complete set of functionalities that are involved in the management of networks and the associated services. These include, for example, the processes of network creation, network and service assurance, and the optimisation of resources.
Innovation_3	**AI/ML**	The term AI/ML encompasses the data models, services, and primitives utilized for the life cycle management of ML algorithms, irrespective of their intended use and position within the network and the associated management domains.
Innovation_4	**Energy Efficiency**	The terms "energy efficiency" and "energy saving" collectively encompass all functionalities and measurements employed for the purpose of monitoring and reducing energy consumption within the network domain.
Innovation_5	**Service and Network Performance**	The term 'Service and Network Performance' is used to describe the functionalities and measurements that are useful for the assurance process at the service and network domains.
Innovation_6	**Network Architecture**	The term 'network architecture' is used to describe the various functions, interfaces, protocols and services that collectively define the network domain. In particular, the radio access network, the core network and the transport network are all examples of network architecture.
Innovation_7	**Data Analysis**	The term "data analysis" encompasses all functionalities and services employed for the purposes of network data collection, data storage, data elaboration, dataset production, and data visualisation, among other activities.
Innovation_8	**Infrastructure Interworking**	Infrastructure interworking can be defined as the set of data, functionalities, data models and architectures that facilitate the inter-domain orchestration activities, specifically within the context of network and infrastructure domains.
Innovation_9	**Security**	Network security innovation focuses on developing advanced technologies and strategies to protect data and systems from cyber threats. One of more interesting aspects refers to Zero Trust Architecture. This security model assumes that threats could be internal or external, requiring strict verification for every user and device attempting to access resources

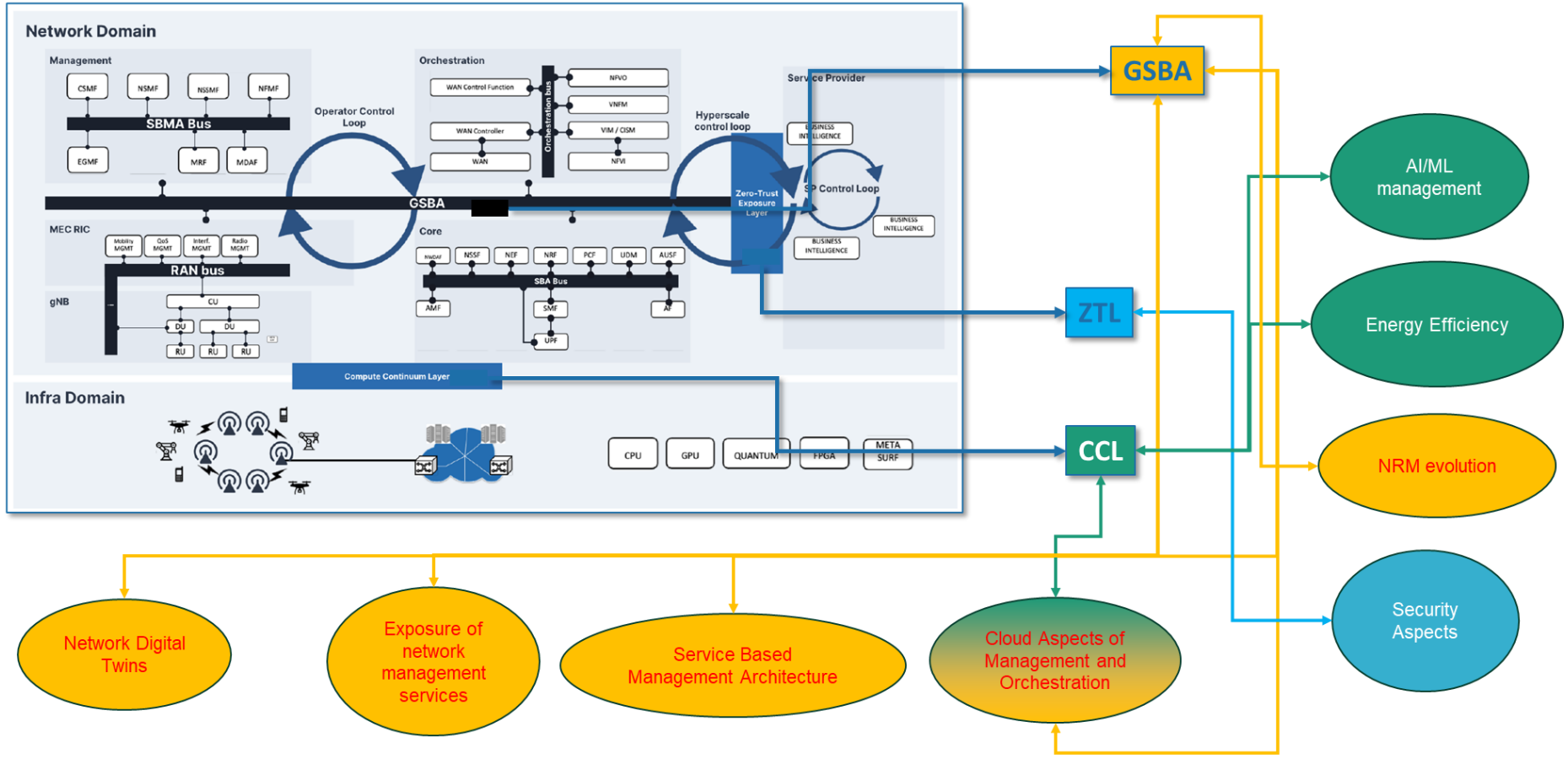
Origami Innovation Mapping to 3GPP contributions [2/2]

#	3GPP Topic	Description	3GPP WGs	Origami Innovations
3GPPTopic_1	**Network Resource Model**	Network Data Model as the underlying model for all network management processes	SA5	Innovation_1
3GPPTopic_2	**Orchestration**	Network management process coordination activities (policy management, intent-based, AI/ML LCM, etc.)	RAN3, SA5	Innovation_2
3GPPTopic_3	**AI/ML**	AI/ML LCM management and all supporting functions (data production, data collection, data analysis, energy efficiency related to AI/ML)	RAN3, SA5	Innovation_3
3GPPTopic_4	**Energy Efficiency**	Performance measurements and related functionalities useful for network energy efficiency	RAN3, SA5	Innovation_4
3GPPTopic_5	**Network Performance Management**	Define performance measurement, KPIs and KQIs and related management processes	RAN3, SA5	Innovation_5
3GPPTopic_6	**RAN Architecture**	Evolution of RAN architectures with the introduction of SBA to 6G	RAN3	Innovation_6
3GPPTopic_7	**Data Analytics**	Management processes related to the collection and processing of data produced by the network.	SA5	Innovation_7
3GPPTopic_8	**3GPP Network Digital Twins**	Role of digital twins in network management processes	SA5	Innovation_3, Innovation_4, Innovation_7
3GPPTopic_9	**Infrastructure Management**	Network management processes Relationship to infrastructure environment (NFV, Cloud, etc.)	SA5	Innovation_8
3GPPTopic_10	**Security aspects**	Security applied to OAM services	SA5	Innovation_9



Discussion

- Discussion items:
- How can ORIGAMI contribute to the SA5 activities?



ORIGAMI

Thanks!

The ORIGAMI project has received funding from the
Smart Networks and Services Joint Undertaking (SNS JU)
under the European Union's Horizon Europe research and
innovation programme under Grant Agreement No 101139270
www.smart-networks.europa.eu

Co-funded by the European Union

6GSNS