The Architectural Approach of Affordable5G

Lambros Sarakis

National and Kapodistrian University of Athens

Department of Digital Industry Technologies

HIGH-TECH AND AFFORDABLE 5G NETWORK ROLL-OUT TO EVERY CORNER –
Dissemination Workshop, NKUA, Evripos Campus, Psachna, Greece, 20/10/2022

Introduction (1/2)

- Businesses and industries are looking forward to 5G NPNs to get high-level granular views of their operations, service flexibility and spread of deployment possibilities or cost reduction
  - ✓Especially if the network is affordable and easy-to-manage
- 5G NPNs are expected to deliver high speeds and low latency supporting next-generation applications
  - ✓Also ensure that critical civil functions/business processes have access to high-quality and responsive communications even when parts of the system fail due to external factors
- The evolution of 5G is complemented by an industry-wide change towards software defined and cloud technologies, using COTS compute and networking infrastructure to
  - ✓Manage costs and expand the supplier ecosystem
  - ✓Enhance openness, competition and spur innovation in the RAN and CN

Introduction (2/2)

- Industry transition towards an open, disaggregated, intelligent, and highly extensible 5G vRAN architecture meets the O-RAN vision of complementing 3GPP 5G standards with virtualized network elements, white-box hardware and standardized interfaces that support intelligence and openness
- In this talk, we present the architectural approach of Affordable5G for cost-efficient Stand-alone NPN deployment (i.e., NPN that is independent of public mobile network), leveraging
  - ✓5G SA network
  - ✓Disaggregated RAN with open interfaces based on O-RAN
  - ✓Flexible resource orchestration and slicing
  - ✓ AI/ML-based network optimization
- Key enablers for cost-efficient 5G NPN deployment and related challenges are also discussed

Key Enablers for Cost-Efficient 5G NPN Deployment (1/3)

- **Network sharing**
  - ✓Sharing the network infrastructure and adopting network virtualization can help operators to save significant amounts of capital and operational costs
- **Neutral hosting with opportunity for private networks**
  - ✓The idea is to have a single network infrastructure owned by a third party and leased to interested operators
  - ✓Currently, neutral hosts offer mostly infrastructure for public networks, i.e., towers, power, RF front-ends and antennas
    - ✓They can be also involved in the management of 5G equipment (small cells) for private networks
  - ✓The ability to share this infrastructure among multiple public and non-public operators not only reduces relevant costs, but also results in flexible network topologies

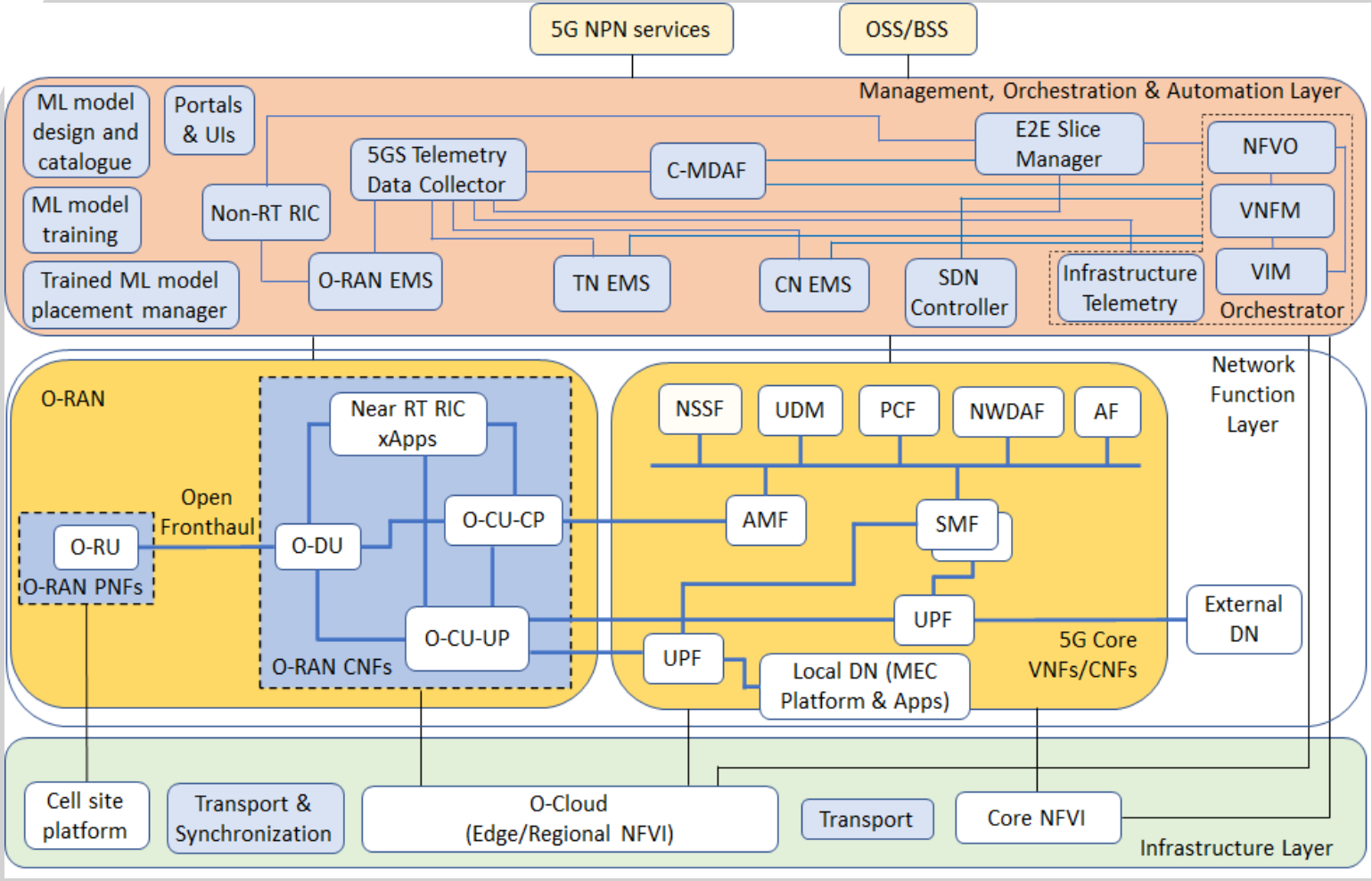
Key Enablers for Cost-Efficient 5G NPN Deployment (2/3)

- **Partitioning edge and cloud**
  - ✓Cost-latency balance choices when positioning computing servers from the network edge/cell site (higher cost and lower latency) to the central office/data center (lower cost and higher latency)
- **Network Slicing**
  - ✓Support and operation of different kind of services (like eMBB, mMTC and uRLLC) with very distinct needs onto the same infrastructure in a cost-efficient manner
  - ✓Needs proper planning and dimensioning

Key Enablers for Cost-Efficient 5G NPN Deployment (3/3)

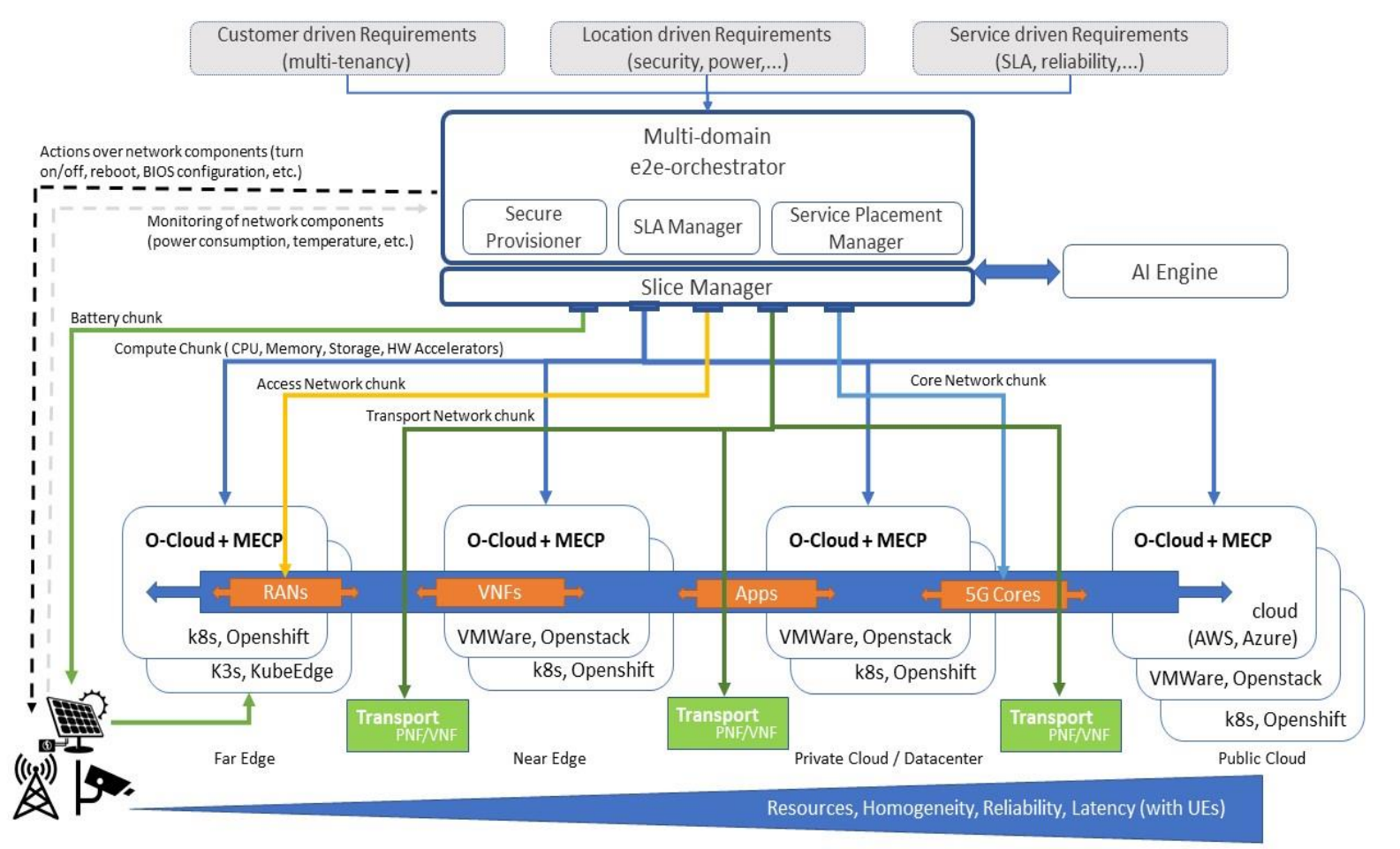
- **Automation for smart operation**
  - ✓Design of algorithms that can autonomously manage all the main phases of the 5G slice and edge resource lifecycle and automate the operations through AI/ML
- **Open software platforms**
  - ✓Such platforms covering network functionalities at RAN, Edge, core and management are essential for a cost-effective and reusable network architecture

Affordable5G Architectural Approach



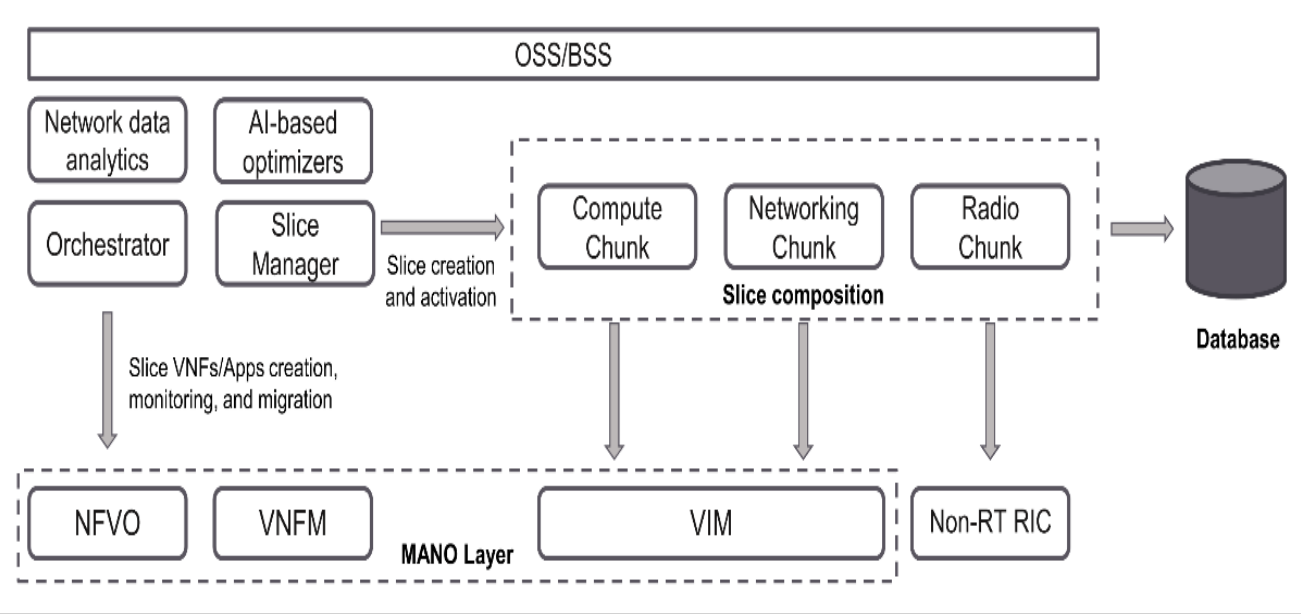
- **Management, Orchestration and Automation Layer**
  - ✓Orchestration, Slicing, Telemetry and Data analytics, AI/ML-based RAN optimization
- **Network Function Layer**
  - ✓NFs related to O-RAN and 5G Core
- **Infrastructure Layer**
  - ✓Core and Edge/Regional NFV Infrastructures, cell site platform and TN segments

Orchestration



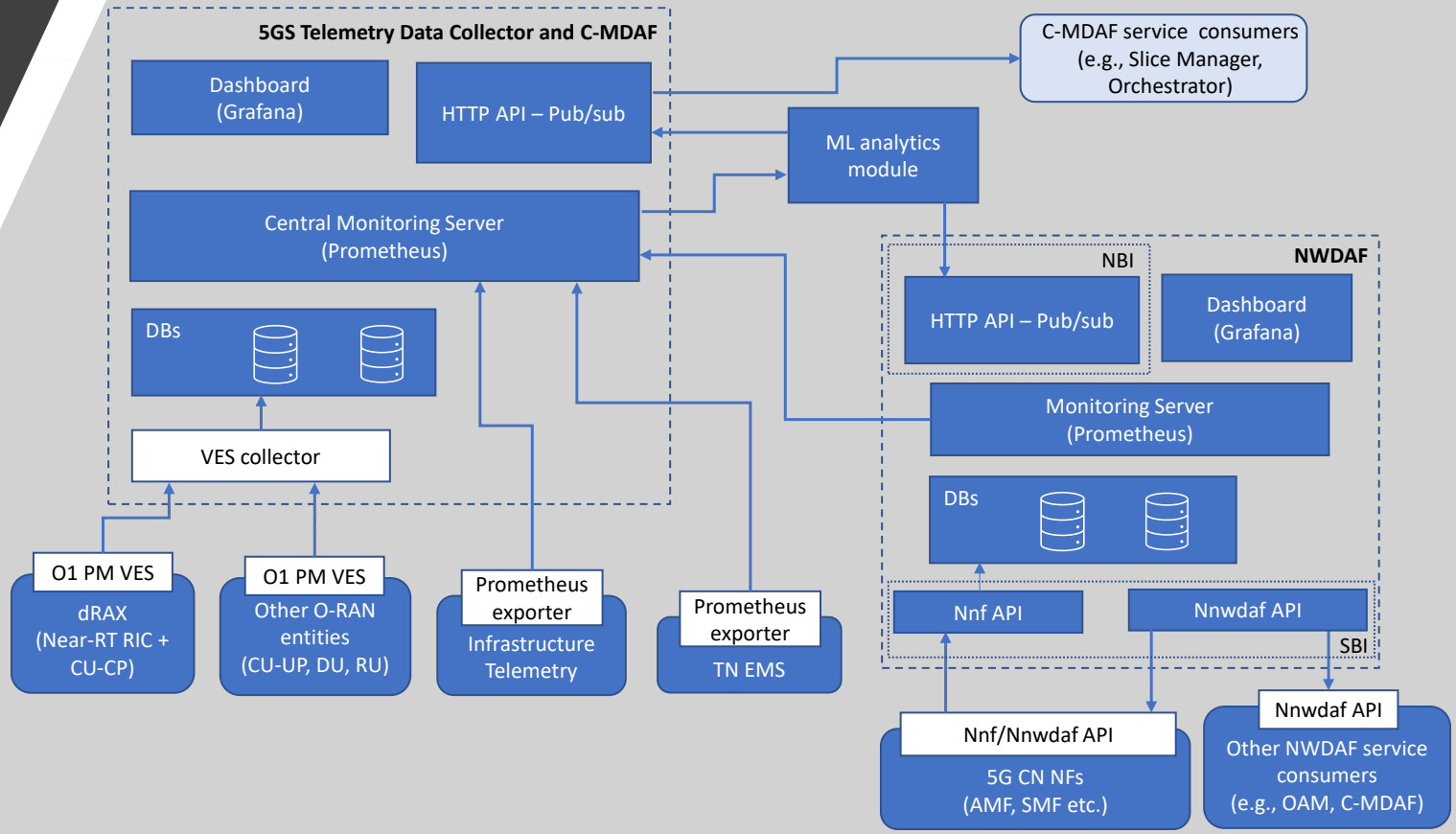
- E2E orchestration solution in real 5G infrastructures, composed of heterogeneous components (from VNFs to MEC resources and hardware devices, spanning across multiple domains with various underlying technologies like Openstack and K8s)
- Responsible for joint network and compute resource orchestration, actuation over network infrastructure and CPU pinning for isolation and guaranteed QoS
- In conjunction with the Slicing solution can support advanced AI algorithms
- Aligned with O-RAN specifications regarding the management of the cloud infrastructure hosting the O-RAN elements (O2)

Slicing



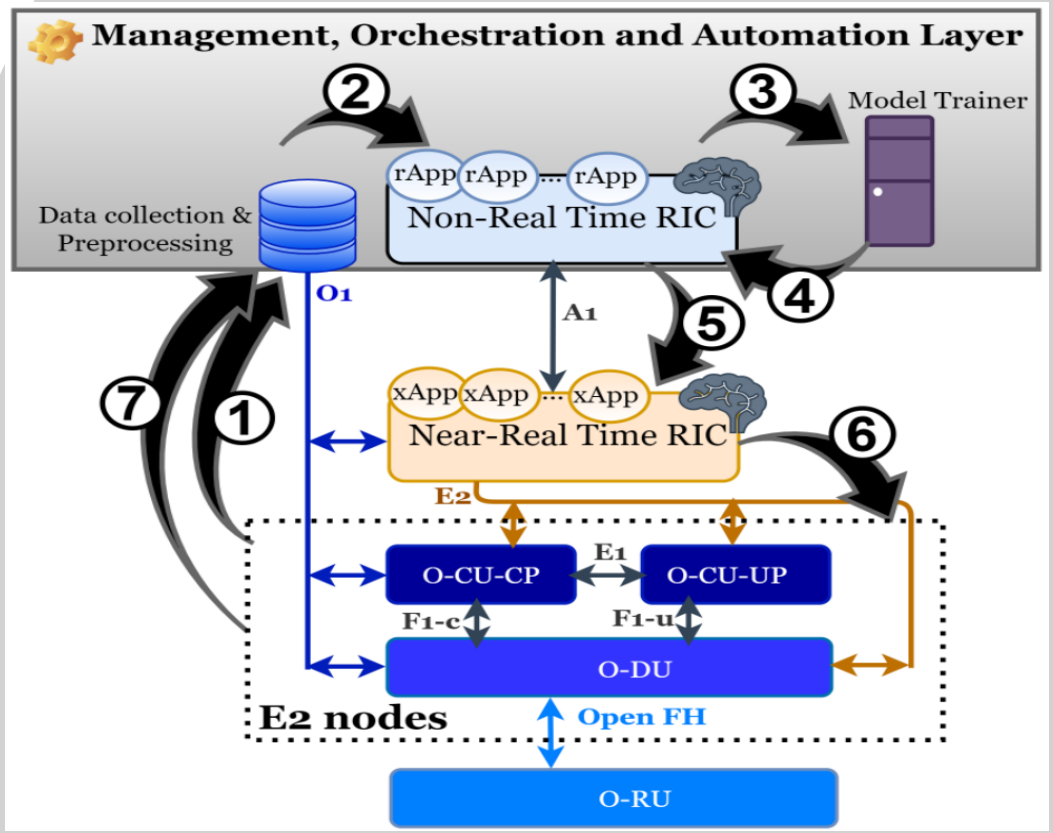
- Provisions E2E slices in the compute, network, and access network domains, allowing several tenants to seamlessly manage the required resources, and deploy services for different verticals within the slices
- The slice manager can exploit AI-based optimizers to predict possible scarcities in the resources of the slices that could prevent the deployment of new services in a slice or even put at risk the performance of the running ones

Telemetry and Data Analytics



- Combines 5GS Telemetry Data Collector, NWDAF and C-MDAF
- The 5GS Telemetry Data Collector gathers
  - ✓ performance measurements from the cloud infrastructure (OpenStack & Kubernetes environments)
  - ✓ “application layer” measurements from O-RAN, CN and TN possibly by utilizing existing EMSs for performance monitoring
- NWDAF and C-MDAF provide data analytics facilitating intelligent decision making
  - ✓ NWDAF focuses on network data analytics
  - ✓ C-MDAF focuses on management data analytics
  - ✓ Possibly utilizing AI/ML
- To support ML-based data analytics and optimizations, the proposed system includes facilities for ML model
  - ✓ Design
  - ✓ Training
  - ✓ Packaging
  - ✓ Placement in the appropriate component

AI/ML-based RAN Optimization



- Involves non-RT RIC and near-RT RIC together with facilities for data collection and preprocessing, and model training
- Example workflow scenario for a DL RRM task
  - ✓ Step 1: data collection and optional pre-processing
  - ✓ Step 2: pre-processed model is forwarded to Non-RT RIC
  - ✓ Step 3: Model training accommodated in the AI platform (e.g., TensorFlow)
  - ✓ Step 4: Trained model is sent back to the Non-RT RIC, and is available for inference purposes
  - ✓ Step 5: Through the A1 interface, the pre-trained model is forwarded to the Near-RT RIC
  - ✓ Step 6: During near-real time control, the model predictions are applied to the O-RAN actors
  - ✓ Step 7: ML-assisted network entities provide feedback for possible model update/re-training

Challenges (1/2)

- Support of O-RAN and placement of NFs
  - ✓ Use of standardized interfaces for O-RAN element management (O1), O-Cloud management (O2), passing policies from the non-RT RIC to the near-RT RIC (A1) and control of O-RAN elements by the near-RT RIC (E2) is important to avoid vendor lock-in.
  - ✓ Analyze the implications of placing O-RAN and UPF NFs in the available cloud-native infrastructures (edge, regional and core cloud)
    - ✓ Depending on factors like available fronthaul capacity, traffic density, number of supported RUs and network slices, existence of local DN etc
- Orchestration and network slicing
  - ✓ Orchestration should equally address the challenges of the network edge where resources are scarcer and the need to support different functions (physical and virtual) and different devices (sensors, cameras etc) makes the network heterogeneous
  - ✓ Benefits from exploiting AI/ML for providing predictions regarding resource utilization impacting the QoS of established slices

Challenges (2/2)

- RAN sharing and neutral hosting
  - ✓ New challenges in 5G have to do with evolving from pure RAN sharing towards multi-tenancy in the NG-RAN enabled by the introduction of PDU sessions, network slicing and NPN
  - ✓ With 5G Core CUPS and URLLC slices, the UPF can now be located at the edge
    - ✓ Creates a need for addressing per tenant ownership or sharing of this UPF, together with the control-plane mappings between the RAN CU-CP and 5G Core AMF and SMF
- AI/ML and analytics for network optimization
  - ✓ Deployment and holistic testing of both DL and DRL models towards resource management of O-RAN, utilizing the learning capabilities of Non-RT and Near-RT RICs
    - ✓ Model deployment and training ➔ model inference and action taking ➔ model evaluation for possible retraining and update
  - ✓ Efficient exploitation of NWDAF and MDAF
- Hardware acceleration
  - ✓ For network operations at the edge (e.g., video processing) relying on AI and DNN, it is challenging to exploit Application-Specific Instruction set Processor-type hardware accelerators combining efficient processing of DNN workloads, low power consumption and performance scalability

Conclusions (1/2)

- The approach of the Affordable5G project regarding the realization of 5G NPNs was presented
  - ✓ Comprising solutions which aim at reduced deployment and operational costs
- The system's RAN follows the specifications provided by the O-RAN and includes components and open interfaces for the management/control of the O-RAN NFs, and the management of the underlying cloud infrastructure
- To facilitate network automation and minimize the human intervention, the system includes components for network and management data analytics, and AI/ML-based O-RAN optimization
- Support for neutral host RAN sharing is based on solutions for multi-tenancy in the NG-RAN enabled by the introduction of PDU sessions, network slicing and NPNs

Conclusions (2/2)

- Resource and service orchestration is undertaken by an E2E solution operating on top of 5G infrastructures, composed of heterogeneous components like virtual/containerized NFs, MEC resources and hardware devices
- The proposed system is being validated in scenarios, including
  - ✓ Industrial use cases with time synchronization constraints (leveraging TSN over 5G)
  - ✓ Mission critical services including voice, video and data
  - ✓ Video analysis at the edge for smart city applications

Thank you for your attention

Lambros Sarakis, National and Kapodistrian University of Athens, lsarakis@uoa.gr