O-RAN as the Architecture for *Real-time* **Observability and Programmability** for the AI-driven RAN

Institute for the Wireless Internet of Things
Northeastern University

Northeastern University
Reliance Jio
NVIDIA
MITRE
BAH
DeepSig
Digital Catapult
Mavenir
Tiami Networks
Rimedo Labs
HighStreet Technologies
IMDEA
PUT
U Tokyo / Nakao Lab

O-RAN and 6G: What is missing?

Context

6G Use Cases

- Spectrum management
- ISAC
- Energy efficiency
- AI-native

Analysis

O-RAN Gaps

Analysis of limitations and missing capabilities in the O-RAN architecture

Our proposal

6G-ready O-RAN

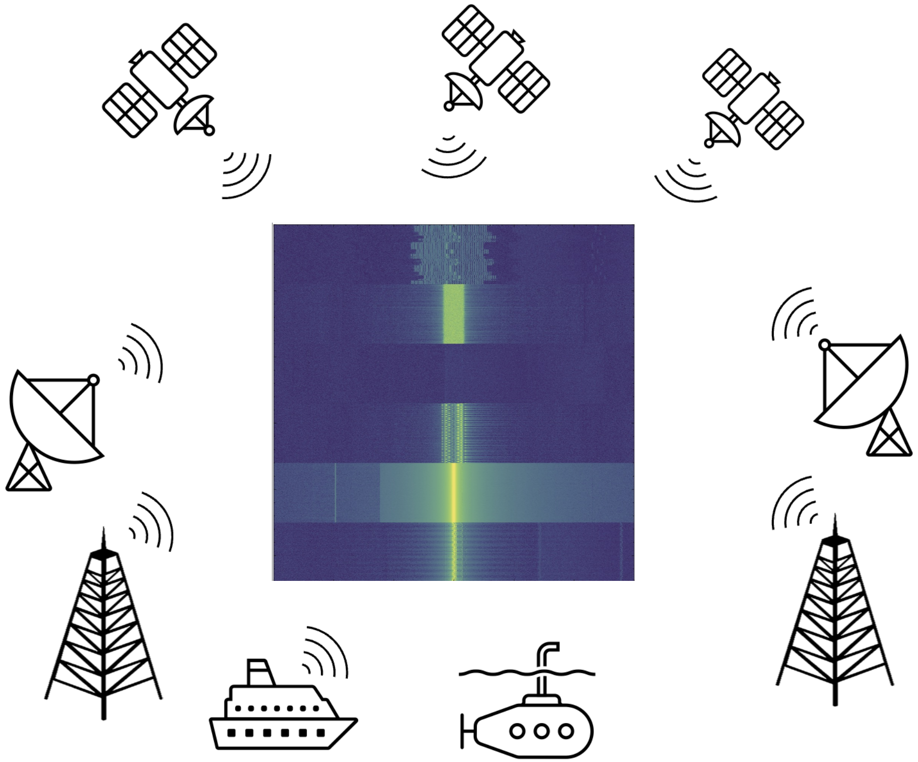
An evolution of the architecture with *real-time observability and programmability* as foundational elements

Gap 1: Real-time Spectrum Management

Significant interest on shared spectrum approaches for 6G:

- FR-3 band (6-24 GHz) with several incumbents
- Lower 3 GHz band

→ Coexistence facilitates allocating additional spectrum to commercial cellular services



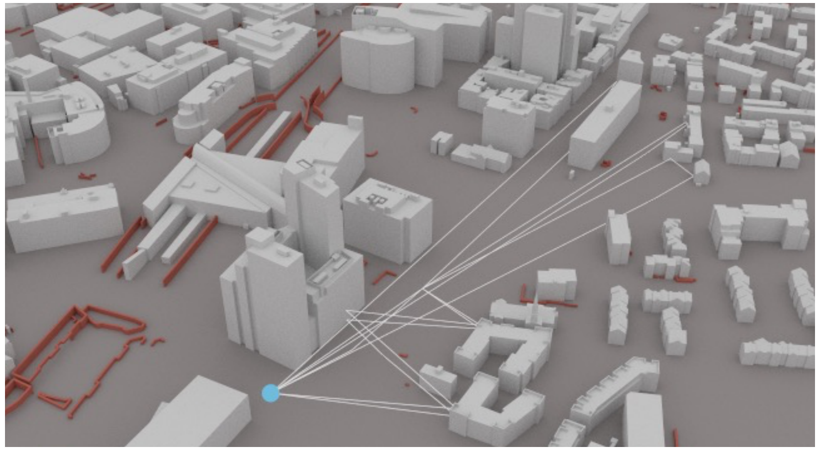
- CBRS is deployed (420000 CBSDs) but is slow and conservative
  - Leave spectrum on the table
  - Need external sensor network
- Pervasive cellular infrastructure can enable real-time

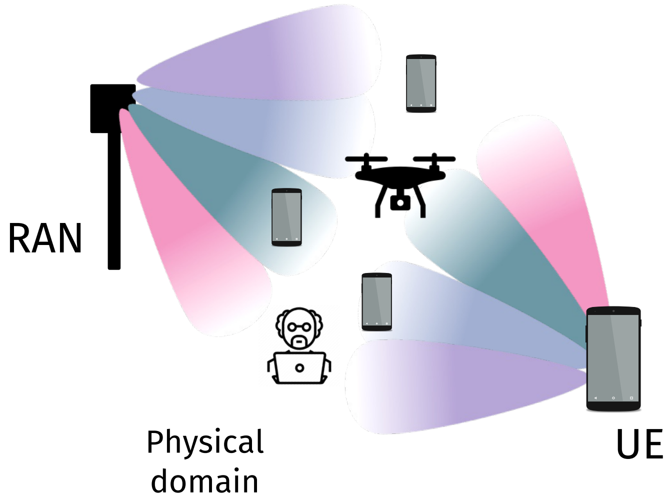
Sensing → Coordination → Sharing →

- **Gap:** Limited support in O-RAN
  - Lack of **real time** observability and programmability
  - Advanced sensing requires interactions with **user plane**
  - Lack of real-time failsafe mechanisms in the architecture

Gap 2: Integrated Sensing and Communications

- Integrated sensing and communications (ISAC) is being considered as a key capability for 6G
- Networks become distributed sensors exposing physical domain into the digital environment
- Commercial and defense applications
- Sensors market estimated at 300B by 2030





- What is the ISAC control plane?
- What is the life-cycle management for ISAC algorithms?
- How can vendor and operators collaborate to advance ISAC operations?
- **Gap:** How can the O-RAN architecture support ISAC?

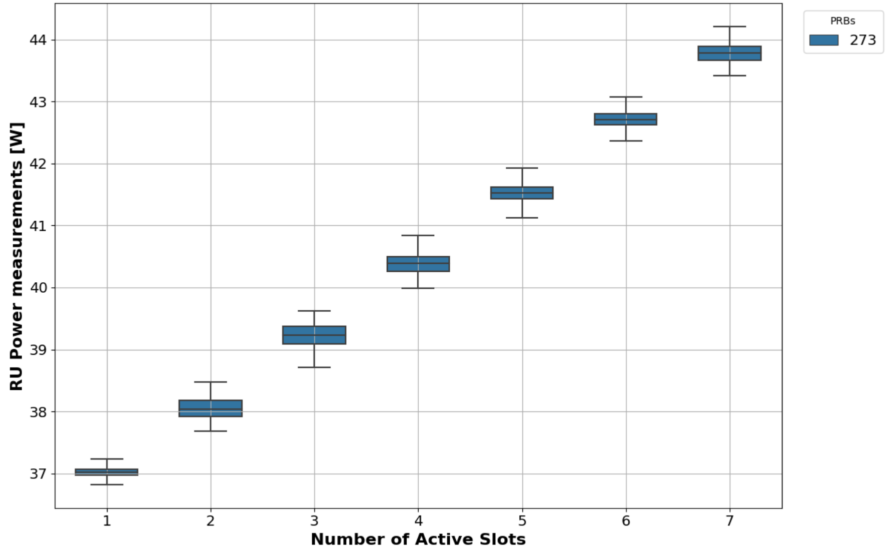
https://www.bccresearch.com/market-research/instrumentation-and-sensors/sensors-technologies-markets-report.html

Gap 3: Energy Efficiency

- Estimates put energy consumption for RAN to 50-80% of energy expenses for operators
- Key to 6G: reduce energy consumption to limit costs
- Energy efficiency is a weak element in the value proposition for AI and software-driven networks

- CU/DU/RU dynamic tuning can improve system energy efficiency

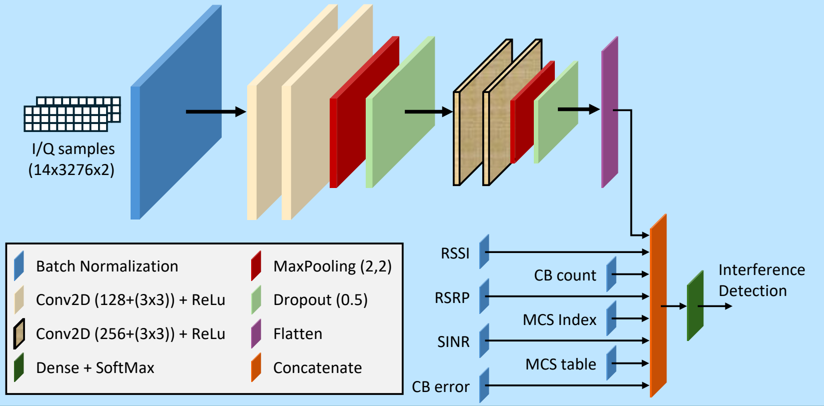
Example: energy-aware dynamic traffic shaping can reduce RU power consumption



- **Gap:** Lack of real-time observability prevents systematic collection and analysis of energy-consumption drivers
- **Gap:** Lack of real-time control prevents dynamic closed-loop granular scaling (symbol, slot, subframe levels)

Gap 4: AI/MLOps for AI-Native RAN

- Leverage AI techniques for inline processing or control procedures
- 3GPP is considering different options for 6G releases
- AI-RAN Alliance is developing proof-of-concepts
- Plug-and-play/swappable AI



- AI/ML for RAN optimization needs data with high granularity, automated labels, workflows and pipelines

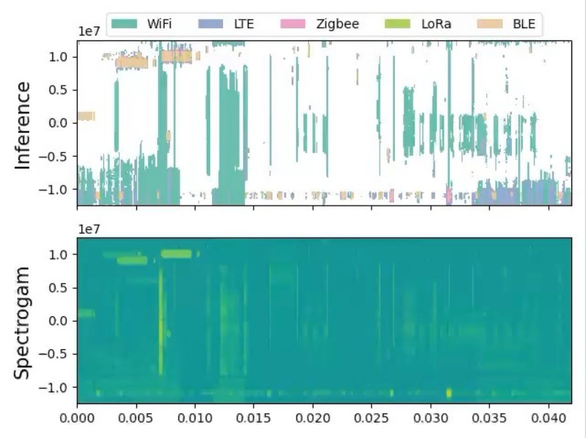
- **Gap:** The granularity is limited to near-real-time reports or non-real-time KPI logs
- **Gap:** Data is limited to the control plane, no or limited access to user plane elements
- **Gap:** Limited support for testing of AI/ML capabilities
- **Gap:** AI/ML lifecycle needs to be driven by software all the way to DU/CU

Recap: O-RAN Architecture Limitations

No interactions with the user plane

- IQ, bits, bytes, packets
- Cannot be used by the RICs because of data rate, security, privacy constraints

Inference on user plane data enables advanced spectrum and RF sensing, analytics



No real-time control plane

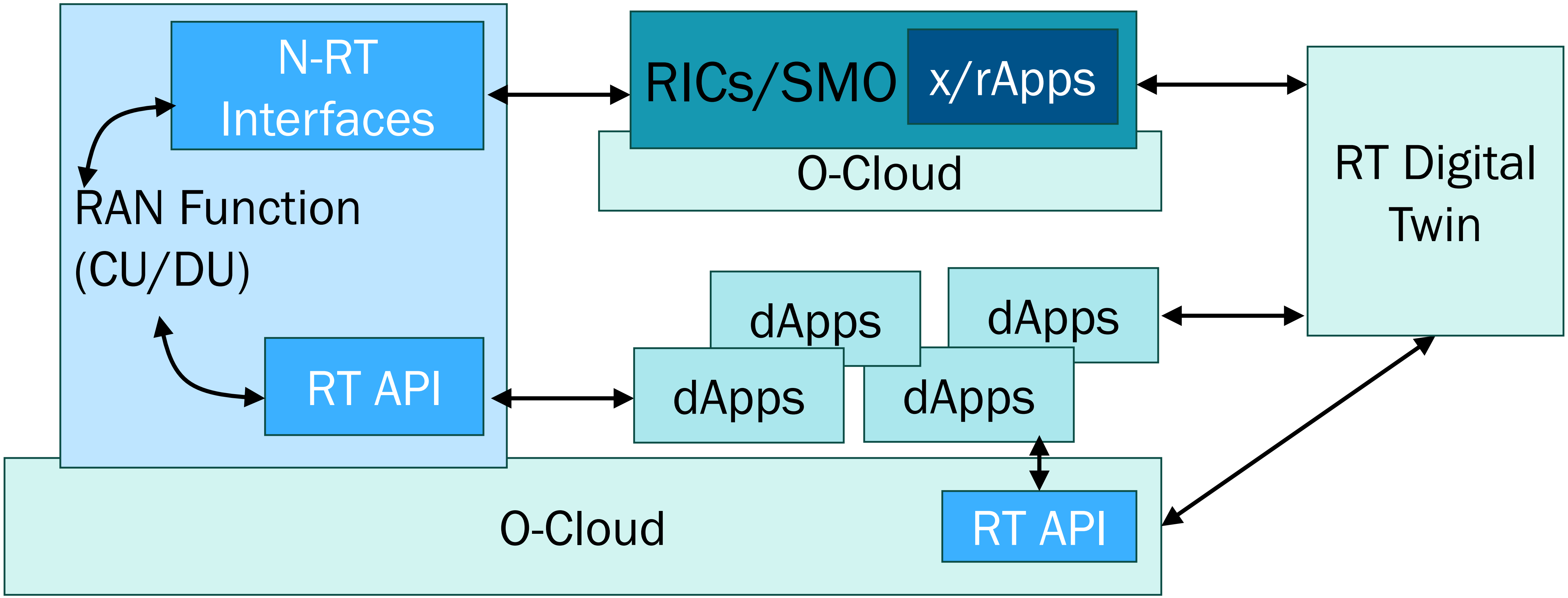
- O-RAN interactions with the control plane only extend to near-real time and above

Limit sensing, dynamic optimization, coordination at multiple time scales, observability

Key to introduce new services to be monetized

Real-Time Observability and Programmability

Proposal: Evolve the O-RAN architecture toward a **platform for real-time** RAN **observability and programmability** across *user and control planes*



dApps - Current Status

2022:

- Original dApp paper - propose the idea, discuss some use cases, present early architecture
  S. D'Oro, M. Polese, L. Bonati, H. Cheng, and T. Melodia, "dApps: Distributed Applications for Real-time Inference and Control in O-RAN," IEEE Communications Magazine, vo. 60, no. 11, Nov. 2022

2023

- Further research and development
- **O-RAN nGRG: proposed and approved research item in RS-02 Architecture** (RI06)

2024

- Architecture and prototype development
- O-RAN nGRG: first research report on dApp use cases

2025

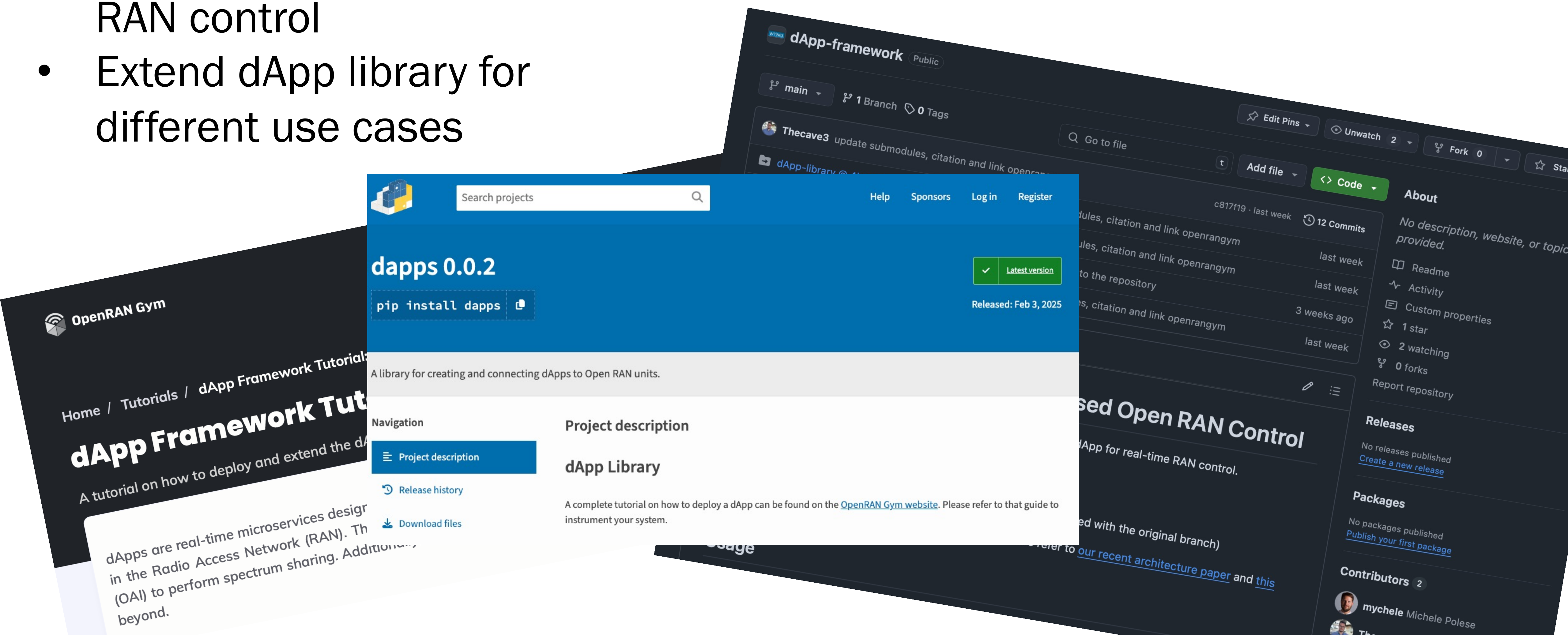
- Released the **first open-source dApp prototype**
- Developed and published dApp use cases
- First paper on dApp architecture
  A. Lacava, L. Bonati, N. Mohamadi, R. Gangula, F. Kaltenberger, P. Johari, S. D'Oro, F. Cuomo, M. Polese, and T. Melodia, "dApps: Enabling Real-Time AI-Based Open RAN Control, Computer Networks, 2025
- O-RAN nGRG: second research report on dApp architecture
- **NVIDIA and others develop and release dApp capabilities as part of their stack**

Institute for the Wireless Internet of Things
Northeastern University

dApp Open-Source Framework

- Enable research on real-time RAN control
- Extend dApp library for different use cases

https://openrangym.com/tutorials/dapps-oai



dApp Research Reports

Use Cases and Requirements

Use Cases Review

ISAC

Interference detection

Beam management

Customizable AI/ML for RAN

Energy efficiency

Real-time scheduling

Spectrum sensing

Physical layer security

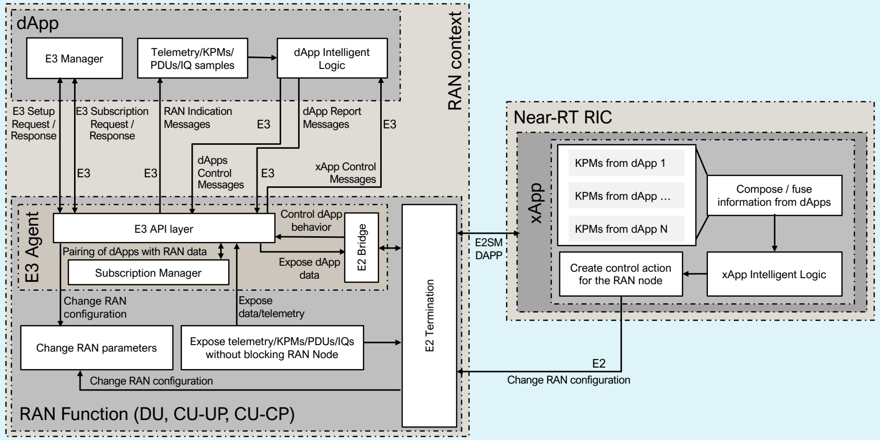
Architectural and Interface Requirements

User and control plane inputs

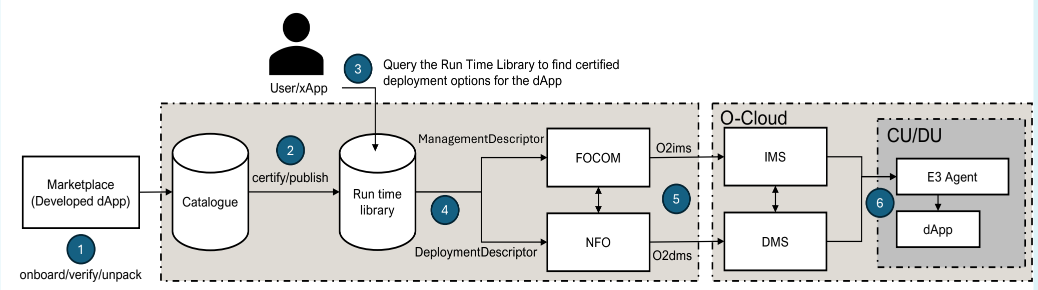
Control knobs and parameters

Architecture

Preliminary Architecture and Procedures



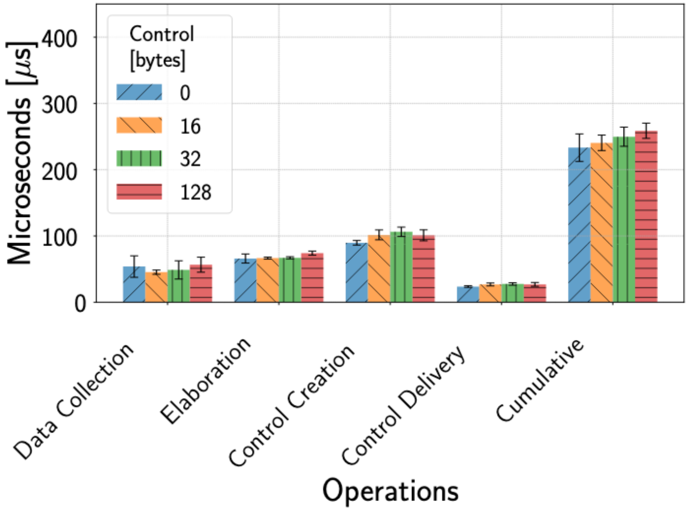
Lifecycle Management



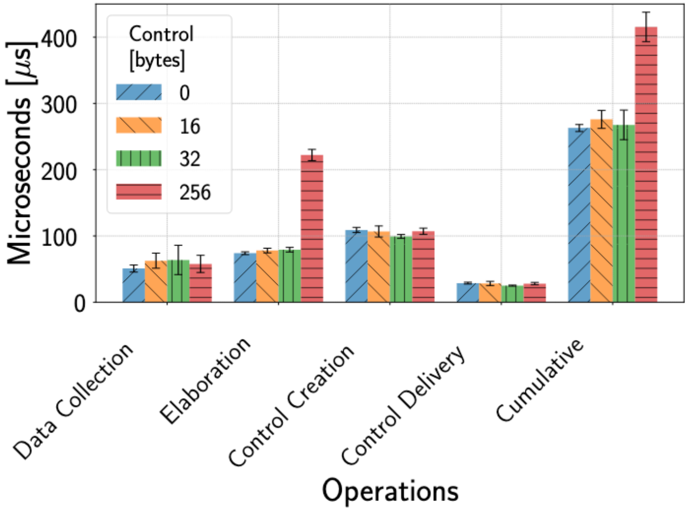
Open-source Prototype

16 contributors, 7+ reviewers, 13 O-RAN members across RS-02, nGRG TOC

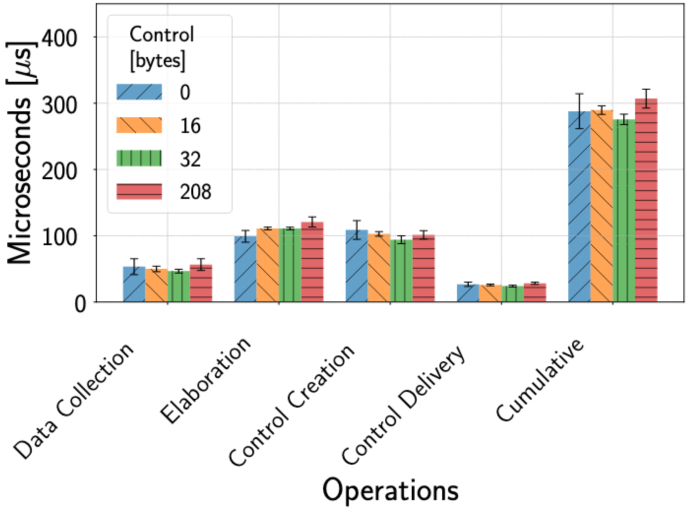
Benchmarking Open-Source dApps



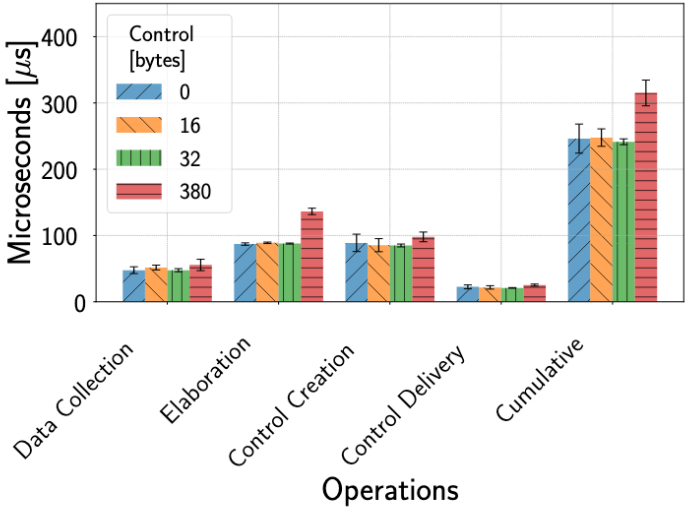
(a) IQ samples 384 Indication size - 1536 bytes



(b) IQ samples 768 Indication size - 3072 bytes



(c) IQ samples 1536 Indication size - 6144 bytes



(d) IQ samples 2048 Indication size - 8192 bytes

Benchmark on 2 different increasing variables:

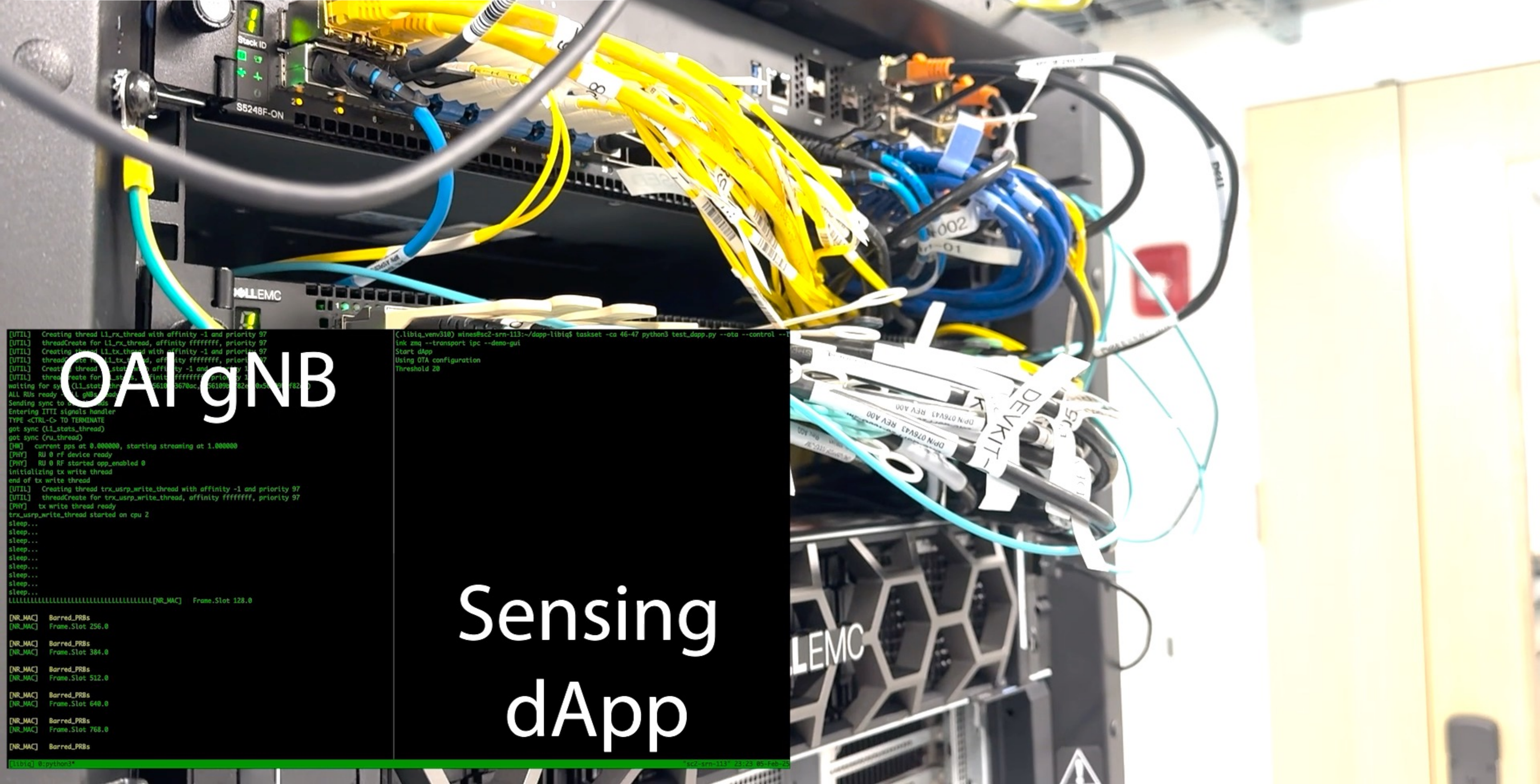
- Size of the indication message
- Size of the payload of the control action

Analyze total and stepwise latency

Key insights:

- Average control loop *always* below 410 μs
- High data/control throughput
- Flexible framework supporting multiple use cases
- Limited overhead (<0.05%)

Spectrum Sensing Demo Video



Real-Time Digital Twin

Real-time RAN data and APIs

Physical to Digital

(incl. ISAC)

Digital to Physical

Digital Twin for real-time RAN production testing

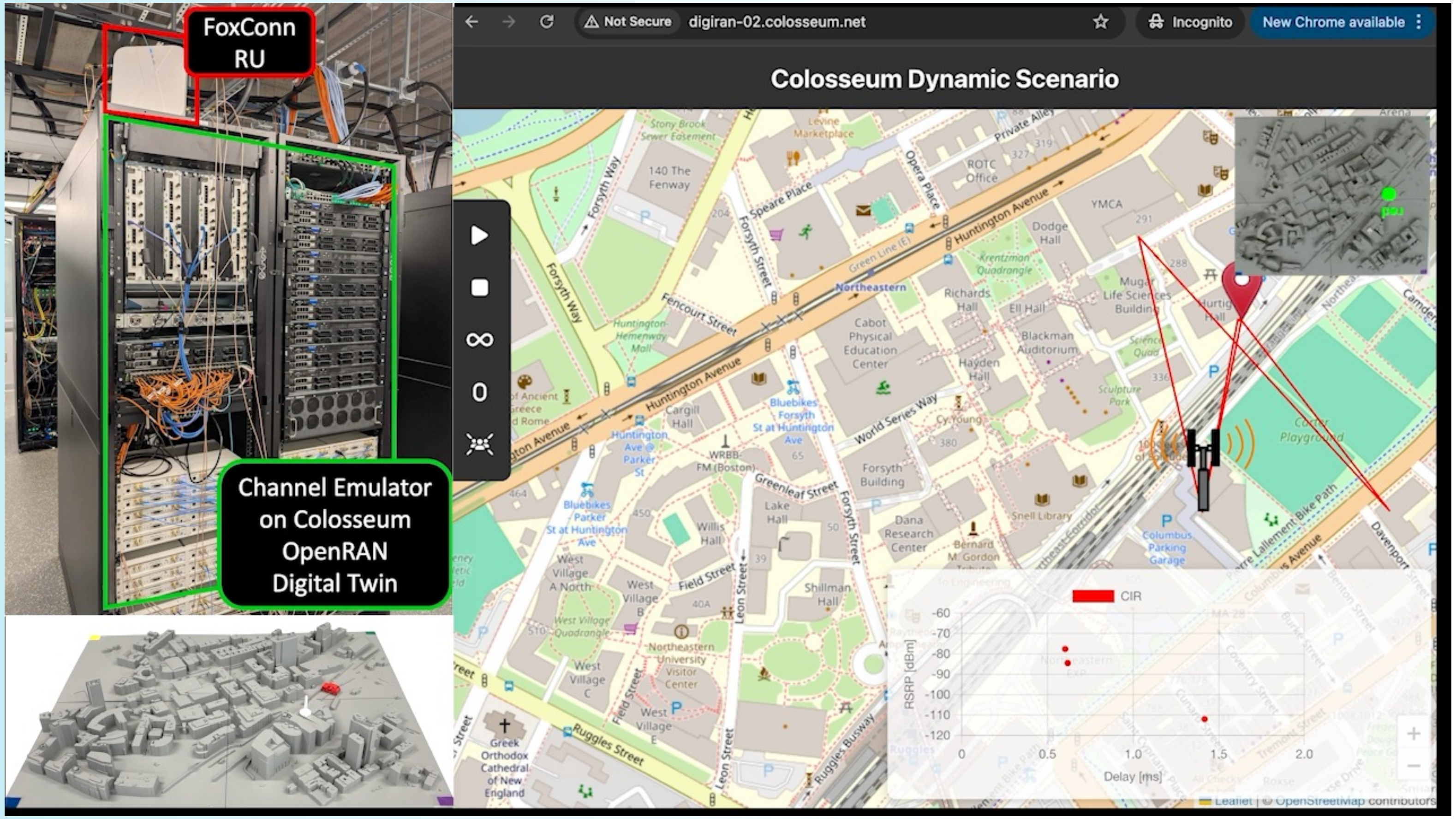
Reproduce network conditions and behavior

Train AI/ML apps

Test configurations, software, and AI

Assist dApp/xApp/rApp lifecycle

Real-time O-RAN digital twin prototype: adapt channel response based on users position and ray tracing in 3D model environment



Proposal

- Study and develop
  - Data exposure framework (at different time scales, including real time)
  - Real-time architectural concepts
    - Focus on reliability and workload isolation
    - Further develop the concept within appropriate WG (or WGs)
  - End-to-end data and AI lifecycle management

- Analyze security and privacy of the framework, aligned with WG11

- Develop digital-twin-driven AI/ML testing and validation

- Open-source prototypes and real-time DT as testing bench

Phase 1		Phase 2
Observability and Sensing Privacy and Security	→	Programmability and Control

Institute for the Wireless Internet of Things
Northeastern University