AIMM

CELTIC-NEXT

Newsletter

July 2021
Issue 1

AI for Network Operation and Management

Executive Summary

AIMM is a two-year CELTIC-NEXT European collaborative research and development project targeting radical performance improvements and efficiency dividends for 5G and beyond Radio Access Network (RAN), through advanced antenna array (Massive MIMO) and Reconfigurable Intelligent Surface (RIS) technologies, powered through and managed by the latest advancements in Artificial Intelligence (AI).

This document provides an outline of the principles and frameworks that are being considered within the AIMM Work-Package 5 (WP5) on AI for Network Operation and Management. The topics of discussion include Network Operation and Management Frameworks, Data Analytics Platforms, Interactive Development Environments, and Reinforcement Learning (RL) Applications.

Introduction

A key objective of the AIMM project is to enhance the performance of 5G and beyond Radio Access Network (RAN) via developing novel Artificial Intelligence (AI) based algorithms. The AIMM project addresses two aspects of AI in the RAN. The first, “bottom-up” approach, is to use AI to optimise the air-interface performance and enable the practical implementation of advanced antenna structures and network architectures. The second, “top-down” approach, is to incorporate data collection features coupled with AI functionalities in order to facilitate RAN intelligence and automation at the system level.

In the context of the “top-down” approach, the traditional human-centric approach to network management and operation will not be capable of providing the dynamic management required to ensure that the service capability and capacity of inherently and increasingly complex future RAN is fully realised.

The intention of AIMM Work-Package 5 (WP5) is to use AI techniques to improve the quality and responsiveness of network management operations. The work of AIMM WP5 focuses on novel approaches to the use of network data in order to drive efficiencies. Access to such data at different levels in the RAN architecture implies the availability of open interfaces where this data can be made available. An assumption within AIMM is that the RAN architecture will follow the principles of the emerging O-RAN standards and potential evolution of that standard.

This document provides an outline of the principles and frameworks that are being considered within AIMM WP5 on AI for Network Operation and Management.

Network Operation & Management Frameworks and Data Sources

5G is an umbrella concept under which multiple technologies and concepts come together in different domains, to implement a seamless integration of mobile and fixed network that allows the interconnection of people, things, and information flows. Network Function Virtualization (NFV) and Software Defined Networking (SDN) are two key enabler technologies of 5G, mainly underpinning the programmability of network functions aimed at the automation of provisioning, network management, operation, and maintenance tasks. The principles of programmability and automation are crucial for the implementation of the most important improvements promised by 5G, mainly at a network performance level such as increased spectrum efficiency or reduced latency. These are also very important for the AIMM project, being the foundations for our work on developing AI algorithms that implement advanced Cognitive Network Management use cases.

At an early stage, the AIMM consortium reviewed the existing technology trends and frameworks to deliver 5G and decided to adopt within the project scope of work the technology concept proposed by the O-RAN Alliance. This decision has a significant impact across all the work-packages, and for the case of WP5 the impact is fundamental in shaping posterior work.

Initial AIMM WP5 research focused on the analysis of existing Network Management Frameworks and identifying the main requirements, features, procedures, and high-level characteristics of a framework that may enable the development of AI based automations upon O-RAN 5G networks in general and may underpin AIMM Use Cases in particular. This determined that FCAPS is the most suitable framework to be used within the project as it is widely accepted and adopted across the industry mainly by 3GPP and on the initial Open RAN standardisation work.

The Service Management and Orchestration (SMO) function using O1 interface, have been defined in the O-RAN Alliance standards, to be responsible for the implementation of all FCAPS interfaces and procedures to the multiple Virtual Network Functions (VNFs) as shown in Figure 1.

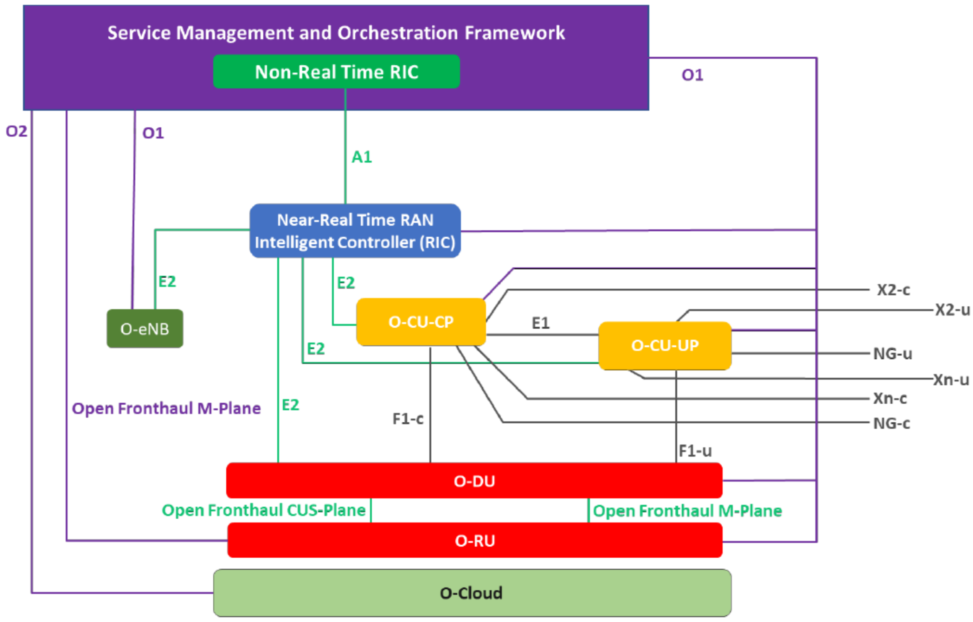

Figure 1. Logical architecture of O-RAN as defined by the O-RAN Alliance.

Another important aspect of this initial research is the identification of the relevant Data Sources and Network Topologies. It is important that at early stages of the development of an AI practice to identify the data sources, the formats, the needs for mediation mechanisms (such as data collection, cleansing, database schemas, storage mechanisms, data linkage, data enrichment, etc.) to inform the project activities dedicated to developing the system-wide simulation environments and the development/training/validation of AIs to implement AIMM Use Cases.

Data Analytics Platforms

The project has deployed a testbed network based on O-RAN standards. The centralized units (CUs), distributed units (DUs), and remote radio units (RRUs) are the main network elements deployed as VNFs, and they produce FCAPS data that describe the network behaviour. The FCAPS data is a fundamental building-block of the AI development cycle, being part of the training and validation process. Once deployed, the AI will further use the collected network data for its general operation.

The various AIMM testbeds offer real-live network traffic allowing to bridge the gap between purely computer simulated environments and a real live network, that being based on standardized interfaces, it presents the typical limitations of the real-world where features and capabilities are delivered as per a staggered development roadmap and market priorities. This helps the AIMM project partners to understand in more detail, the inherent challenges that will be presented when trying to deploy the AIs developed in purely computer simulated environments to real-life networks through a xApp/rApp product solution.

Given the standardisation of the RAN Intelligent Controller (RIC) functionality is still ongoing within the O-RAN Alliance, we have implemented an OSS/BSS and Analytics platform, depicted in Figure 2, for the collection, storage, processing, and analysis of the FCAPS data from all the VNFs and PNFs. This platform offers functionalities like the RIC such as:

- Data pipeline that implements data acquisition, data reliability checks and data cleansing.
- Data transformation and data enrichment processes.
- Storage of raw data according to the defined Network Topology data schemas and metadata.
- Definition of policies for the generation of alarms based on the process of FCAPS data.
- Definition of policies that inform the AI decision-making process.
- Trend analysis of KPIs and performance metrics.
- Data visualization capabilities.
- Notebook to run AI applications.

There are hundreds of FCAPS file types regularly generated, with different schemas, by the different VNFs and PNFs. The total of files processed by the system increases with the network growth.

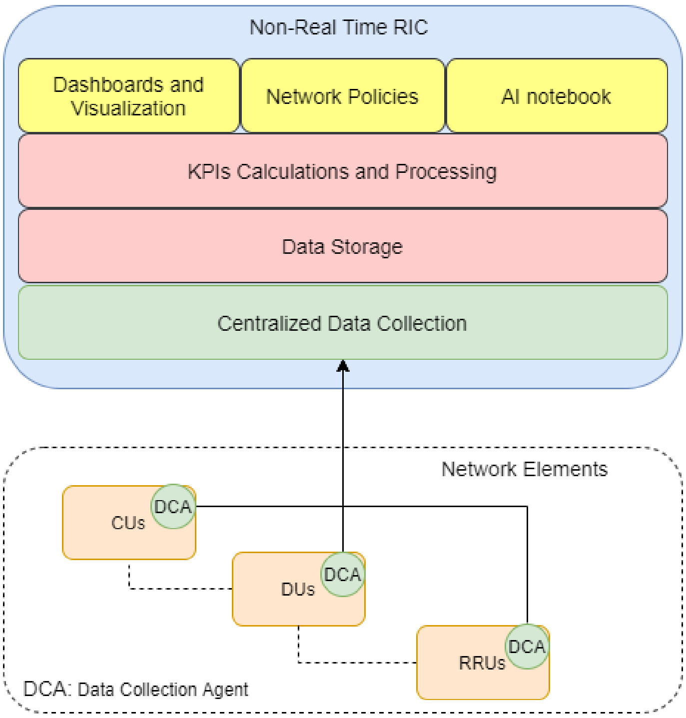

Figure 2. Architecture of the OSS/BSS and Analytics platform implemented for the Live-Network.

Interactive Development Environments

Recent developments in the field of AI provide new capabilities of generating automated solutions for network management functions. Specifically, Reinforcement Learning (RL) is an approach for dynamically controlling and solving Markov Decision Processes. An RL intelligent agent learns to make sequential decisions by interacting with the environment. To gather information and train the intelligent agent, an accurate simulation framework for network management is necessary.

System-wide Simulation Environments

AIMM WP5 partners have been working on two system-level simulators. In one instance, a partner developed an AIMM Network simulation model, known as AIMMSim, to train, test and validate algorithms. On another effort, another consortium partner has been working on repurposing an existing system-wide simulator, known as DRIVE.

AIMMSim

AIMMSim is a system-level simulator which emulates a full cellular radio system following 5G concepts and channel models. The intention is to have an easy-to-use and fast system written in pure Python with minimal dependencies. It is especially designed to be suitable for interfacing to AI engines such as tensorflow or pytorch, and it is not a principal aim for it to be extremely accurate at the level of the radio channel. For the latter task, pre-computed look-up tables (based on simulated channel models) are used to obtain fast run-times. The code has a structure as shown in Figure 3. The code has already been proven successful by the construction of eight test scenarios of varying complexity. Work is currently in progress to next test RIC heuristics for sub-band allocation using Q-learning, a concept we introduce in the next section.

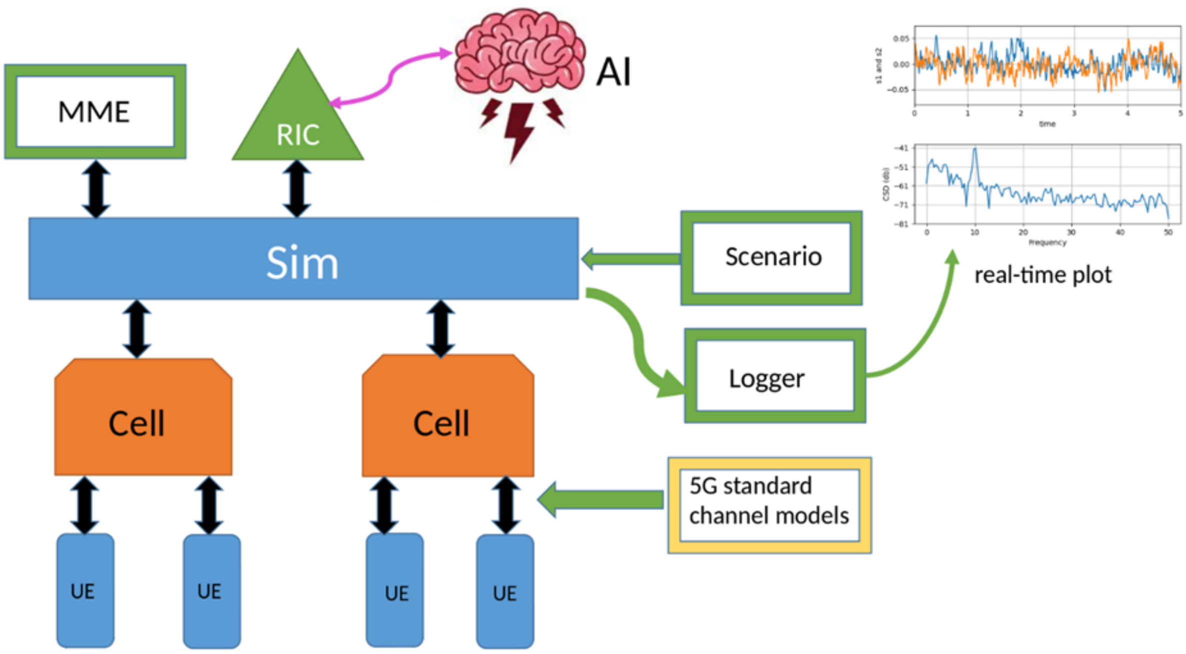

Figure 3. AIMMSim Block Structure.

DRIVE

On another effort, another consortium partner has been working on repurposing an existing system-wide simulator. This work has seen the adaptation of the simulation environment 'DRIVE', originally developed in MATLAB due to its feature-rich toolbox support and excellent data visualization capabilities, to be integrated within the Python-based OpenAI Gym environment. The DRIVE environment provides accurate mobility and radio signal strength simulation results for multi base stations employments at a city-scale level and can generate rich data to support RL algorithms' development. The OpenAI Gym is used, in this case, as the de facto environment wrapper for RL research and development.

The training process of an RL model involves a very large number of environment interactions. In the current design, the 'DRIVE' simulator plays the role of a custom environment for the gym. Hence, to develop RL algorithms with MATLAB simulator, the issue of central importance is MATLAB simulator and Python gym integration. Such API mechanism must be ultra-fast and efficient enough – this is by far the most stringent requirement (typically, tens of millions of interactions are needed to train the RL agent and a single interaction latency of 1 ms might result in a feasible AI training setup, while 100 ms might not).

The overall process of the MATLAB Python interaction is shown in Figure 4.

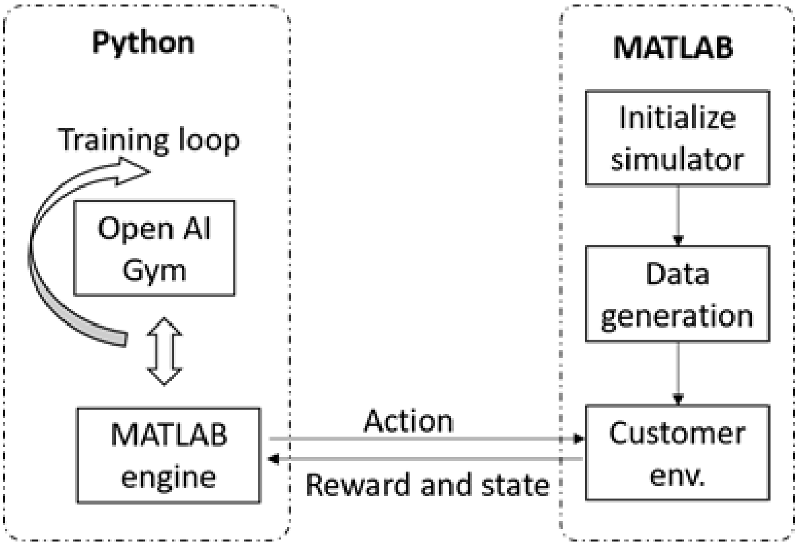

Figure 4. DRIVE Simulator: MATLAB/Python interaction.

Reinforcement Learning

RL for Smart Interference Management

As part of the initial RL research and development work done so far, one partner have selected AIMM Use Case “Smart Interference Management” to develop it first attempt at developing an algorithm using the AIMMSim (AIMM System-wide Simulator) for the training and testing phases. This algorithm aims to implement interference mitigation decisions/actions based on a simplified mathematical model, where the problem is posed as a single control problem using RL techniques, based in the RIC. One popular RL technique in widespread use is Q-learning. Here, a centralized array or look-up table known as the Q-table is maintained. The values in this table are called Q-values, initialized to zero, allowing the agent to start to learn with an equal choice among all available actions. The Q-table, therefore, presents an analysis of the choice of behaviour of all the individual agents, whereas, the Q-value represents the expected cumulative reward the agent receives by learning an action.

In our test scenario, actions are the splitting or not of the channel between two neighbouring cells, which we only do if at least one user equipment (UE) is situated in the cell-edge zone of the two cells. The reward is positive if the splitting results in improved throughput (or other key-performance-indicators). Thus, over time, actions which benefit throughput are more likely to be selected. Further details including numerical results of the solution under development will be provided in subsequent public deliverables of the AIMM project.

Deep RL Applications in AIMM

Real-world problems are incredibly nuanced, with high dimensionality and variation. Massive MIMO management is no exception. To represent the intricacies of the world, Deep Learning is of much use. Deep learning models have shown great promise in approximating functions that represent highly dimensional, continuous domains. Deep RL takes this a step further by applying RL techniques to the deep learning models used to interpret data, effectively combining perception and control into one AI model. However, not all representations of a task would fit the appropriate methodology for solving it. For example, for the task of automotive navigation, Global Positioning System (GPS) coordinates are a better representation for the task as opposed to street names. To that end, a significant portion of the RL research in the AIMM project is going into leveraging Unsupervised Learning/Self-supervised Learning techniques to learn better task-specific representations from the data collected. Such representations will allow our RL agents to be more robust to changes in the environment and the task at hand, as well as increase the training sample efficiency.

In another preliminary test-case, Proximal Policy Optimization (PPO), which is a Deep RL algorithm, is used to solve a handover problem. Using the DRIVE simulator, a target city map is chosen and the base station distributions of corresponding Macro cells and Femto cells are generated. Then, the received signal strength in each location on this map, according to buildings features and signal transmission model, is calculated. Afterwards, the mobility trace generation module is introduced. This module can randomly generate a trace depending on a command from the OpenAI Gym interface. The overall process of PPO training can be formalised as: Gym calls and queries the DRIVE simulator asking for a new mobility trace, prompting the simulated car to move within the trace according to the current timestep. During motion, DRIVE connects the candidate base stations along this trace to the car, following the actions provided by PPO. PPO is trained to pick the best base stations along the trace to maximise the received signal strength and data-rate. Under this formalization, PPO’s goal is: based on the car position, base station positions and RSSI, PPO chooses the most suitable base station to connect to the car.

Additional details including performance results of the Deep RL solutions will be provided in future public deliverables of the AIMM project.

BT

CEMWorks

IMST

interdigital

Loughborough University

NOKIA

thinkRF

University of BRISTOL

Universität Stuttgart

Vilicom
signalling the future

Bundesministerium für Wirtschaft und Energie

UKRI Innovate UK

Canada NRC·CNRC