O-RAN next Generation Research Group (nGRG)

Contributed Research Report

Research Report on Digital Twin RAN Use Cases

Report ID: RR-2024-07

Contributors:

NVIDIA

Verizon

Qualcomm

Rakuten

Nokia

VTT

Bell Canada

Keysight

Ericsson

Tech Mahindra

Northeastern University

Release date: 2024.05

Authors

Company	Name	Email
Rakuten (Section 3.1)	Kexuan Sun (**Editor**)	kexuan.sun@rakuten.com
VTT (Section 3.1)	Chen Tao	Tao.Chen@vtt.fi
Northeastern University (Section 3.2)	Pedram Johari Salvatore D'Oro Michele Polese Leonardo Bonati Tommaso Melodia	p.johari@northeastern.edu; s.doro@northeastern.edu; m.polese@northeastern.edu; l.bonati@northeastern.edu; t.melodia@northeastern.edu
Qualcomm (Section 3.3)	Geetha Rajendran	geethapr@qti.qualcomm.com
Verizon (Section 3.4)	Nirlay Kundu	nirlay.kundu@verizon.com
Bell Canada (Section 3.4)	Yazan Chakfeh	yazan.chakfeh@bell.ca
Keysight (Section 3.4)	Balaji Raghothaman	balaji.raghothaman@keysight.com
Ericsson (Section 3.4)	Mirko D'angelo	mirko.dangelo@ericsson.com
Tech Mahindra (Section 3.4)	Rajat Agarwal	Rajat.Agarwal@TechMahindra.com
NVIDIA (Section 3.5)	Lopamudra Kundu Xingqin Lin Chris Dick	lkundu@nvidia.com xingqinl@nvidia.com cdick@nvidia.com

Reviewers

Company	Name	Email
Nokia	Daiju Chiriyamkandath Antony	daiju.ca@nokia.com
Dell	George Ericson	geroge.ericson@dell.com
DT	Jan Plachy	jan.plachy@telekom.de
Reliance Jio	Vikas Dixit	vikas1.dixit@ril.com
Ericsson	Mirko D'Angelo	mirko.dangelo@ericsson.com
Dell	Mohammad Alavirad	mohammad.alavirad@dell.com
China Telecom	Qingtian Wang	wangqt08@chinatelecom.cn

Disclaimer

Copyright

Executive summary

The next Generation Research Group (nGRG) is a task force within O-RAN ALLIANCE, which focuses on research of open and intelligent RAN principles in 6G and future network standards.

Research stream (RS)-01 explores the area of Digital Twin RAN (DT-RAN) use cases and performs an analysis of the potential gaps in the O-RAN standards which are then provided as inputs to next DT-RAN research phases and other research streams.

Section 1 provides some background information before DT-RAN is introduced in more details in Section 2. Section 3 provides the use case analysis including background information, motivation, and proposed solutions for each DT-RAN sub use cases including DT-RAN for AI/ML training, evaluation and testing, DT-RAN for network testing automation, DT-RAN for network planning, DT-RAN for network energy Saving, and DT-RAN for site specific network optimization. The use cases addressed in this version of the document come as the top 5 use cases suggested by all participating companies through a comprehensive survey conducted in RS-01. Section 4 provides a conclusion of the use case study results.

Table of Contents

List of abbreviations

3GPP	3rd Generation Partnership Project
6G	6th Generation
AAU	Active Antenna Unit
AI	Artificial Intelligence
API	Application Programing Interface
CI/CD	Continuous Integration and Continuous Delivery
CNF	Cloudified Network Function
DT	Digital Twin
DTN	Digital Twin Network
DT-RAN	Digital Twin RAN
gNB	Next Generation Node B
ITU	International Telecommunication Union
IMT	International Mobile Telecommunication
ML	Machine Learning
NDT	Network Digital Twin
Near-RT	Near-Realtime
Non-RT	Non-Realtime
PA	Power Amplifier
RAN	Radio Access Network
RIC	RAN Intelligent Controller
TRU	Transmitter Receiver Unit
VNF	Virtual Network Function

List of figures

Figure 3.5.1 General DT workflow

1 Background

As the research community across industry and academia continues to shape the scope of the sixth generation (6G) wireless networks, it has become apparent that many novel applications and services will emerge at its onset, ranging from extended reality, immersive multimedia, holographic communication, network and computing convergence, multidimensional sensing, pervasive intelligence, connectivity for industry 4.0 and beyond [1]. The Artificial Intelligence (AI)/ Machine Learning (ML) based technology for Radio Access Network (RAN) automation, management, orchestration, and optimization represents a key factor for the foundations of the O-RAN architecture which is also one of the key enabling technologies for the future 6G evolution. Indeed, Non-Realtime (Non-RT) and Near-Realtime (Near-RT) RAN Intelligent Controllers (RICs) are currently the two main hosts of these technologies enabling RAN intelligence. However, there are still many problems and challenges which need to be addressed in the entire industry before an AI/ML-powered solution can be commercially deployed at scale and start to create real business values in the future. Enabling this wide range of use cases requires addressing a diverse set of requirements, which would be difficult to meet with the previous generations of wireless networks. To that end, various state-of-the-art technologies have emerged as key enablers for 6G use cases, among which the digital twin (DT) has stood out as a highly promising candidate to facilitate the design, analysis, operation, automation, and intelligence of 6G wireless networks [2].

2 Introduction

A Digital Twin network (DTN)* is a digital replica of the full life cycle of a communication network, or part(s) of a communications network, including, for example, any combination(s) of physical network elements and components, virtualized/cloud-native (containerized) network functions (VNFs/CNFs), physical hosts for such VNFs/CNFs etc. Unlike conventional network simulators, the DTN supports communication between the physical network and the virtual twin network to achieve real-time interactive mapping. Inspired by the great potential of DTs for wireless networks, several initiatives have emerged in standards bodies to develop initial guidelines for DTNs. ITU-T has released a recommendation that describes the requirements and architecture of DTNs [3]. ITU-R has published a report on future technology trends of IMT systems towards 2030 and beyond with a list of key emerging use cases for 6G and listed DTNs as one of the important candidates in that category [1]. In particular, ITU-R recommends a top-level design of the DT for RAN to be considered first, before extending the scope to beyond the RAN. Accommodating diverse physical RAN networks, DT-RAN can be designed as a first candidate for DTNs, with "the ability to agilely perceive and adapt to the complex and dynamic environment and achieve network autonomy for its full life cycle in its planning, constructing, monitoring, optimizing, and healing phases" [1].

Creating a realistic and scalable digital replica of the live RF propagation channel between the network and the UEs is also an important part of DT-RAN research. Many of the AI/ML use cases [12] require this, such as for coverage optimization, interference management, mobility and handover optimization, traffic steering, QoS/QoE optimization, mMIMO beam optimization and energy/power management etc. The RF propagation channel as part of the DT-RAN is sometimes called out explicitly in the industry as DT-RF or DT-RAN-RF, since it is an independent research area focusing on the RF propagation and antenna modelling technologies. Research on DT-RAN should include DT-RF due to its importance in the aforementioned wide range of use cases.

Network virtualization/cloudification represents one of the most important technologies and advantages in O-RAN network. Starting from 5G and moving towards 6G, the orchestration and managing of cloud infrastructure, virtual network functions and network slices will be done through advanced AI/ML technologies for the best performance and energy consumption in complicated deployment scenarios of cross domain verticals. Creating an accurate digital replica of the cloud infrastructure that hosts the RAN functions is also an important part of the DT-RAN research.

While the DTN related activities and development in other standards fora are instrumental, application of DTN technologies to O-RAN is vital.. Enabling DTs for generic RAN which may include the RF propagation DT, RAN function DT and RAN Cloud DT could be a good starting point, and extending that work with augmentation of O-RAN specific aspects would make DTNs a reality for the open RAN ecosystem as it steps into the era of 6G.

[*] NOTE: The term “network digital twin” (NDT) has also been used in the industry referring to the same concept as digital twin network (DTN).

3 Use Cases of DT-RAN

3.1 Use Case 1: DT-RAN for AI/ML Training, Evaluation and Performance Assurance

3.1.1 Background Information

Network disaggregation in 5G networks has increased the heterogeneity of the network components, infrastructures, environments, and domains. This heterogeneity poses a challenge for network management. For example, the network components may come from different vendors, the infrastructures may use different hardware types, the environments may span across multiple clouds, and the domains may have different protocols. In 6G networks, the heterogeneity and complexity will be even higher due to further disaggregation and hybrid cloud adoption. To meet the highly diversified service requirements in the future while reducing the network cost and energy consumption in such a heterogeneous scenario, an intelligent solution is demanded. Such a solution should be able to combine various multi-vendor network functions to optimize the trade-offs between performance, cost and energy. AI/ML becomes a powerful tool that can tackle the complex network topologies and problems that may be otherwise unsolvable today.

AI-powered intelligent solutions have the common problem for data generation, training, and solution validation. The use of network DTs which create digital replica of the real network is a promising approach to address these problems.

3.1.2 Motivation

The AI/ML based technology for RAN automation, management, orchestration, and optimization represent a key success factor of today’s O-RAN architecture. However, there are still many problems and challenges that need to be solved in the entire O-RAN ecosystem for the future success of O-RAN technologies.

Challenge1: Data for training and testing AI/ML are key factors determining performance of future networks. However, access to data from operator's radio access network (O-CU, O-DU, O-RU) for AI/ML model training and testing is limited.

Challenge2: Behaviors of AI/ML solutions are difficult to have 100% control on. Performance needs to be assured with run-time validation without impacting the operation of the underlying live physical network.

Challenge3: 3rd party application involvement exposes challenges on solution integration, including solution on-boarding, testing, maintenance, and conflict management. These remain as a key challenge in RIC.

Challenge4: AI/ML performs differently in different network environments. There is no uniformed and reproducible platform for people to evaluate and benchmark the performance of AI/ML solutions from multiple 3rd party vendors and multiple release builds in pre-deployment phase.

AI/ML model training is not simply a one-time offline process. It is normally a continuous optimization and evolution process through continuous interaction of the AI/ML model with the targeted network, and continuously repeating the model training, model inference, performance feedback collection and retaining process. In the current AI/ML workflow design, this entire AI/ML model training, testing, and continuous optimizing process is done based on data collection from and interaction with the real physical network which is:

- **Risky**: the AI/ML could significantly impact the network performance adversely during the continuous optimization process (iterations of training, inference, and retraining) due to lack of effective validation environment prior to provisioning the changes in the network.
- **Slow:** the interaction with the physical network for continuous model optimization is very slow which leaves the network underperforming and at risk with unoptimized AI/ML models for a long time and makes the AI/ML training a long, expensive, and unaffordable process.
- **Costly:** for training the initial AI/ML models for some of the use cases, it needs to build the entire physical network with a large number of mobile users as a test bed first and run it for a long time to collect enough training data.
- **Limited scenarios and data set:** The training and testing scenarios limit to what is seen so far in the real world, but AI/ML models could still make mistakes in other scenarios unseen before. In other words, without sufficient richness in the dataset, the trained AI/ML model may not be generalized enough to cater to diverse scenarios.
- **Untrustworthy:** Even a well-trained AI/ML model could make a serious mistake sometime.

DT-RANs provide a digital replica of a real-world environment and is deeply integrate with the O-RAN network to interact with the AI/ML workflows - model training, testing and continuous optimization. DT(s) will fundamentally overcome the problems mentioned above which will then contribute to delivering more performant AI/ML solutions in future O-RAN network. Furthermore, the performance and reliability of the AI/ML solutions can be continuously monitored and assured during their real-time inference stage with a DT. Any control command and policy generated from AI/ML applications can be validated against the DT first before forwarding it to the real physical network to avoid any potential negative impact to real users' experience.

3.1.3 Proposed solution

The DT concept can be applied to various entities in mobile networks, such as users, services, or network infrastructure. Each entity has its own DT that contains relevant information, updates according to specific rules, and operates on different timescales. There are different ways to realize DT models, for instance, centralized DT, or distributed DT. However, these DTs also have some common features, such as being implemented as software models of the network, not processing actual data flows but mimicking the network behaviors, or multiple DT models being managed by an orchestration system. Therefore, there is a need to develop a method for generalizing and specializing DTs. Different types of DTs require different information, rules, and timescales. A single solution will not be suitable for all of them.

A network DT can be used to facilitate the training and validation of AI/ML models used by network components in 6G networks. As shown in Figure 3.1.1, the Network DT model is used in an AI/ML framework, which forms the cognitive plane to support the needs of AI models in network functions (NF), management and orchestration. The cognitive plane aggregates AI/ML model management, model training and deployment for different network components and entities in a unified, logically separated plane. To avoid an over-complex solution, the framework of framework concept is introduced to build the AI/ML framework for individual domains with similar design principles and toolboxes while providing necessary connections among different AI/ML frameworks for joint processes. For instance, RAN and service orchestration will use separate AI/ML frameworks to support AI/ML functions. The AI functions would be dynamically orchestrated, combining the zero-touch and the CI/CD approaches.

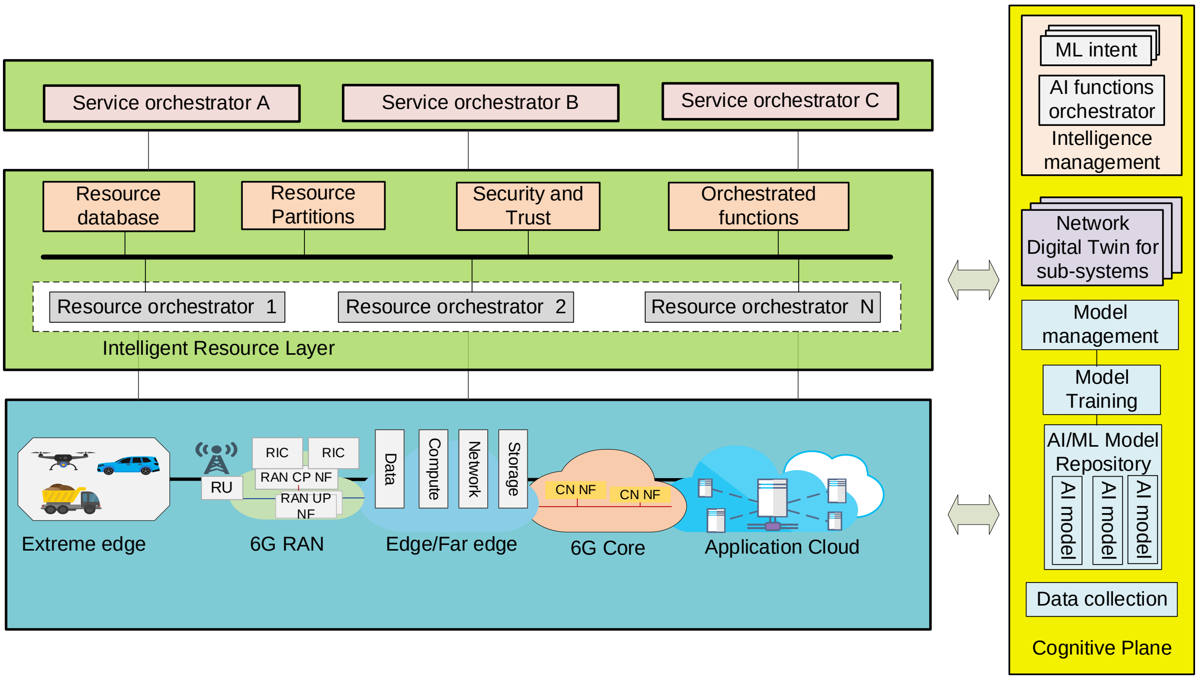

Figure 3.1.1 Network DT assisted AI/ML framework for 6G networks [11]

To understand the AI/ML framework, we explain the main components and their respective functionalities comprising the cognitive plane as follows:

- AI Function Orchestrator is a key component of the AI/ML framework that governs AI/ML models in different network components and ensures the seamless integration, and deployment of AI/ML models within the network functions.
- The Model Management is used to manage various AI/ML models that can be utilized by the network functions within the system.
- The Model Training is a component of the AI/ML framework that handles the training, retraining, or replacement of AI. This component is responsible for fine-tuning the

learning process, first utilizing the DT approach, and then gradually refining the model in the real twins.

- The AI/ML Model Repository is a repository that stores trained AI/ML models and their associated metadata. This database allows for easy access, retrieval, and deployment of AI/ML models within the network functions.
- NDT for Sub-Systems module is used to support AI/ML model training and evaluation. The NDT will provide the representation of network entities at different abstract levels to support the development and validation of AI models. It can generate network settings and training data for pre-training and validation of AI models without the need to interfere with real operation of the network.

The AI/ML framework provides a uniform platform to support AI/ML functionalities in different network domains, segments, and components. Thus, the DTs for sub-systems need different granularities and implementations to serve the purpose of training and validation. These twins have connections and common features/interfaces. Applying object-oriented programming idea, the realization of DTs starts from the hypothesis of a base class NDT. The core APIs are defined to operate twins themselves and to provide information to/from a resource orchestration system. It could be further specialized to NDT for sub-systems (e.g., NDT-RAN, NDT-Core, NDT-Edge, NDT-cloud resource layer) with multiple inheritance being a natural technique to “mix in” different aspects.

In addition to generalizing/specializing, another way that an NDT for Sub-System can manipulate information is by combining data from other NDTs. For example, a NDT that represents a group of users can show information about how all users are moving together in a location, indicating the traffic demand of all users, and providing it to network management. Or a NDT, at the highest level of a combination hierarchy, can show information about a whole network (e.g., average data rate, coverage area, etc.).

The proposed system should evaluate the impact of different aggregation levels, balancing the trade-off between overhead and usefulness. The twins can have different modes of operation. In the simplest mode, they act as static data repositories, receiving updates from their twin and queries from, for example, an orchestrator. In a more advanced mode, they perform semi-active data processing, such as predicting user mobility or service resource requirements for varying load patterns.

Following the general AI/ML framework above, Figure 3.1.2 shows a high-level architecture integrating DT into O-RAN AI/ML framework that can resolve key challenges seen in AI/ML training, evaluation, and real-time performance assurance.

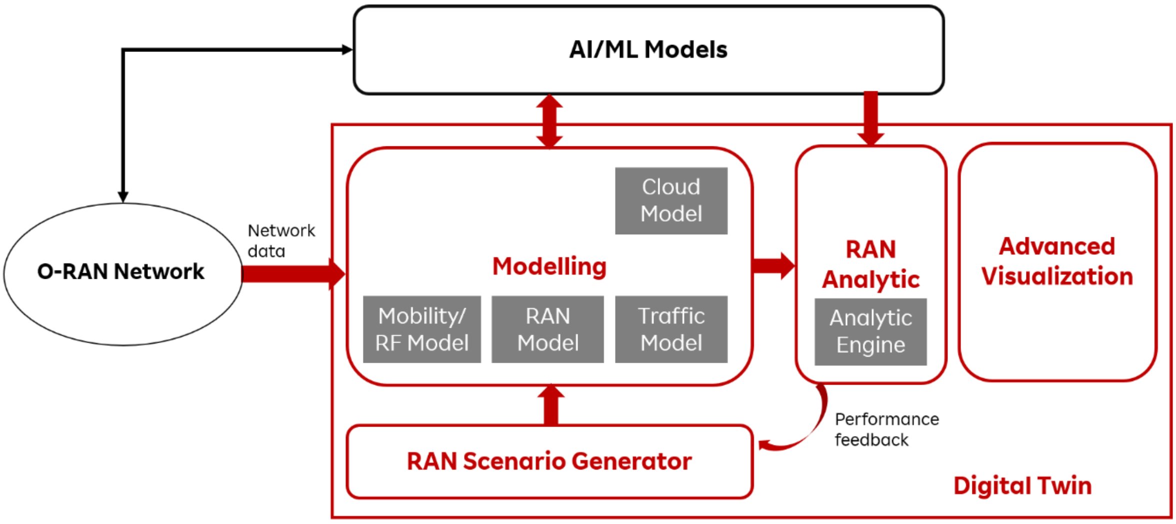

Figure 3.1.2 NDT assisted AI/ML framework in O-RAN

DT is important to address the challenges listed in the section 3.1.2. It is placed here and interacts with both the O-RAN network and AI/ML models.

DT can play several roles:

- Performance monitoring and assurance using sandbox environment provided by DT to avoid AI/ML solution harming the physical network in ways not anticipated otherwise.
- Training data generation when the live data from the O-RAN network is not sufficient.
- Root cause analysis
- Enhanced network health data visualization including the visualization of predicted network health.

There are several building blocks used to construct the DT and make it open for extension as a system-of-systems. These include:

- The **modelling entities** that create accurate digital replicas of different aspects of a RAN network. This includes the mobility/RF model for replicating the propagation channels and user mobilities, the traffic model for replicating the traffic/service demands and their QoS constraints, the RAN model for replicating the RAN functions and the cloud model for replicating the cloud resources.
- The **RAN scenario generator** powered with AI/ML technology which automatically parameterizes the models to generate multitude of training scenarios to challenge the AI/ML models under training. The RAN scenario generator can automatically evolve itself based on the performance feedback from the **RAN analytic module**, generating increasingly challenging training data sets for the AI/ML intelligence and performance to continuously improve.
- **Advanced visualization** to visualize the data needed by the network operators for performance monitoring and system diagnostics.

The DT may be integrated into SMO/Non-RT RIC or Near-RT RIC frameworks where SMO services (SMOSs), rApps or xApps can access the DT function via standardized interface, e.g. SMOS APIs, R1 service APIs or Near-RT RIC APIs, for AI/ML model training, testing/evaluation, and real-time performance assurance. The models are synchronized with the live physical network via the data captured from the O-RAN standardized interfaces such as O1, O2 and E2.

3.2 Use Case 2: DT-RAN for Network Testing Automation

3.2.1 Background Information

There is an urge to develop DT-RAN frameworks for network automation to integrate, deploy and test RAN components reliably and in an automated fashion enabling multi-vendor interoperability and pre-deployment testing and integration of Open RAN equipment, including RUs, DUs, and CUs along with their corresponding interfaces. The DT-RAN will benefit from the flexibility provided by wireless network DTs enabled by the wireless network emulators, e.g., to test equipment in a repeatable way; on heterogeneous propagation environment and conditions, namely Radio Frequency (RF) scenarios; and as part of a larger Open RAN network (e.g., with several virtualized components from multiple vendors) as opposed to tests performed on single functionalities only in isolation. DT-RANs go beyond creating a virtual/DT representation of the physical environment (i.e., high-fidelity 3D models and RF scenarios), and should include protocol stack twinning, as well as integration of commodity equipment in the DT. This allows stakeholders to reliably deploy, integrate and test full protocol O-RAN solutions with hardware-in-the-loop and heterogeneous emulation scenarios. This would also be beneficial, among others, to de-risk AI/ML decisions in the physical/real-world environment by testing them on its digital counterpart first (which acts as a sandbox where “dangerous” configurations can be tested without monetary penalties due to violations of SLAs or due to outages), and at a fraction of the cost compared to similar approaches that rely only on “in real life” data collection and model training.

3.2.2 Motivation

Automated and continuous testing of O-RAN components (e.g., CU, DU, RICs, xApps, to name a few) is a necessary component to unlock the true potential of disaggregated, software-driven, and multi-vendor next-generation wireless networks. As of today, testing wireless systems is a labor-intensive, manual effort. Indeed, while Continuous Integration (CI)/Continuous Deployment (CD) and automation are widely used in cloud systems, existing techniques cannot be directly applied to cellular systems, which come with heterogeneous devices, spectrum and radio requirements, distributed deployments, need for high performance and real-time processing, and complex technical specifications to parse and comply with [13]. DT-RANs can facilitate testing procedures, and this requires automation pipelines that make it possible to easily (and rapidly) instantiate disaggregated gNB components, RICs, xApps/rApps and other network components without any manual intervention. This is particularly important if we consider those cases where DTs are used to provide a real-time twin of the network where different spectrum and network policies can be tested on the twin before being enforced on the production network. For example, one can use the DT to test multiple cell configurations (e.g., small/macro, spectrum policies, ON/OFF cell activation) in parallel and determine the best configuration under the current operational conditions (e.g., network load, UE mobility and traffic profiles, SNR levels, to name a few).

Automated and continuous testing will increase the robustness of the wireless supply chain by allowing swift and continuous integration of new equipment and functionalities from multiple vendors in large Mobile Network Operator (MNO) networks but also in smaller private 5G deployments for new vertical markets. It will also enable a continuous cycle that allows for the validation of new features and code as soon as they are added to the network, before deployments are available, but still in a complete, end-to-end Open RAN sandbox. In this context, DT-RAN will enable reliable, repeatable, cost-efficient, and agile testing of the RAN components.

3.2.3 Proposed Solution

This can be achieved by developing a pipeline for creating and evaluating virtual wireless worlds, and via an automated protocol stack twining framework for continuous testing of standards-based 5G RAN. Moreover, we need to establish a real-time networking infrastructure between the real and virtual worlds with APIs to test fidelity across the twins and develop RAN Intelligent Controller solutions. Finally, we also need to develop a platform for seamless integration of commodity CUs/DUs/RUs to the DT emulation system, perform required performance, security and interoperability testing, and produce testing results. The envisioned DT-RAN can benefit from and expand on the existing Open RAN testing facilities such as the Colosseum, the world's largest wireless network emulator at Northeastern's Open6G Research & Development center.

Example use-case: Colosseum Testbed Extensions (available to O-RAN ALLIANCE community): The Northeastern's Open6G team will plan to expand the capabilities of Colosseum, the Open RAN DT and the world's largest wireless network emulator with hardware-in-the-loop (colosseum.net), and add both dynamic user interactions and near real-time ray-tracing scenarios through newly added GPU hardware, thus enabling high-fidelity Open RAN testing at heterogeneous virtual wireless environments with hardware in the loop. On the protocol stack twinning, the proposed pipeline will follow CI/CD best practices---which are already used to manage the internal components of Colosseum---to replicate software protocol stack of the physical twin in the digital domain, and to perform automated testing of Open RAN components and interfaces through Colosseum automated batch jobs [13].

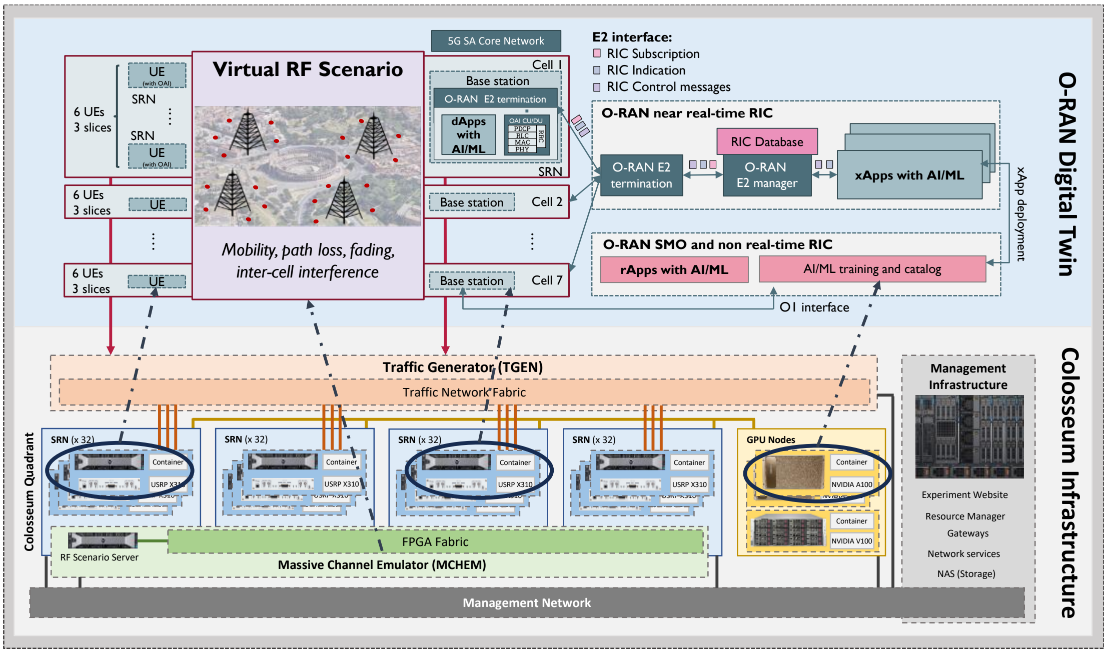

Figure 3.2.1 Emulation-based Open-RAN DT framework for 6G networks, with support of RF, protocol, and AI for DTs

The system will also allow stakeholders to integrate and test commodity Open RAN devices---CUs, DUs, and RUs. We have successfully investigated the integration of commercial products for large-scale testing on Colosseum, for multiple Industry Partners in the past, including a proprietary intelligent jamming system [4], customized Xilinx-based implementation of Software Defined Radios (SDRs) [5], and commodity Cellular Vehicle to Everything (CV2X) radios in a sponsored project with the U.S. Department of Transportation [6].

Finally, this DT framework can be connected to the real-world deployment and stream relevant KPI to the physical twin. This will build on the experience coming from the ongoing integration between the Colosseum emulator and the Arena over-the-air testbed and extend into physical or private virtual segments for integration with stakeholders' sites.

All the above procedures require automated pipelines to simplify the way these configurations are tested, so as to minimize the time needed to determine the best configuration and transfer it to the production network. A possible approach to enable such automation is to leverage microservices, containers and cloud-based software architectures (such as Kubernetes, Docker, Linux Containers (LXC), Virtual Machines, OpenShift) to define automated procedures that process a high-level intent (e.g., specifying the physical deployment to be twinned) and converting it into multiple twin instances that in parallel instantiate and test an individual configuration.

3.3 Use Case 3: DT-RAN for Network Planning

3.3.1 Background Information

A Digital-Twin (DT) is a precise representation of a real-world environment in a virtual model. The virtual model (alias DT) can be used for studying the impacts of geometry of layout, type of objects and material of objects/surfaces on the channel behavior.

In the wireless world this can be used as an effective tool for planning of physical network node (TRU) placements, and projection of expected network performance for a given deployment layout.

A DT can be evolved/updated based on the measurements, learnings, and changes from the real-world environment. A DT supports two-way communication between the real world and virtual world to achieve real-time interactive mapping and closed loop control. A DT can simulate scenarios in real-time or offline.

3.3.2 Motivation

A DT can help in optimal network planning for a given layout, user device density and traffic profiles. It can model various deployments like (not limited to) Indoor, Enterprise, Factory and Warehouses, consisting of 3GPP (FR1, FR2) and/or non-3GPP (Wi-Fi) technology networks.

One can design algorithms which can produce output in the form of network node (TRU) density, node placement locations, configurations, and corresponding Key Performance Indicators (KPI) for a given layout.

Goal (Network Planning) – Finding optimal network deployment options for minimal deployment/operational costs meeting all criteria of Throughput/Latency/Reliability/Energy Efficiency. In addition, a DT can help in studying the impacts of mobility, traffic offloading and Dynamic Tx power adjustment configuration/policies before deploying those configurations & policies in a real network.

3.3.3 Proposed Solution

A real-world environment can be mapped into a virtual model with different level of details.

In a simplistic model one could map the physical characteristics of a layout and simulate/study wave path propagation. Such a twin can be used to develop better channel prediction, link adaptation, beam forming algorithms.

In a more complex model one could map not only the physical characteristics of a layout but also captures the network infra details and simulate/study scenarios of access procedures, beam management, network configurations and application specific behaviors. End-to-end scenarios can be simulated for user experience, performance, and reliability.

A DT can be built comprising of the following:

- Generating virtual models of real-world layouts using advanced computer vision techniques. One can use inputs from various sources like LiDAR, 2D/3D Images, 360 Videos to identify objects, materials, range and create their twins in the virtual model.
- Using existing CAD designs of the layout and adding objects and materials to make the virtual model precisely close to the real-world layout.
- Adding twins of the communication network nodes to the virtual model. For example, UE & Network simulators for 5G and Wi-Fi.
- Adding twins of standard and/or custom traffic generators or the actual traffic generators to complete the DT. E.g., Proprietary Enterprise Applications.
- Communication bridge is defined between the real and virtual world. Network configuration, policy, AI/ML models are deployed from DT to the real network. The real network sends back collected data to DT for virtual model enhancements/updates.

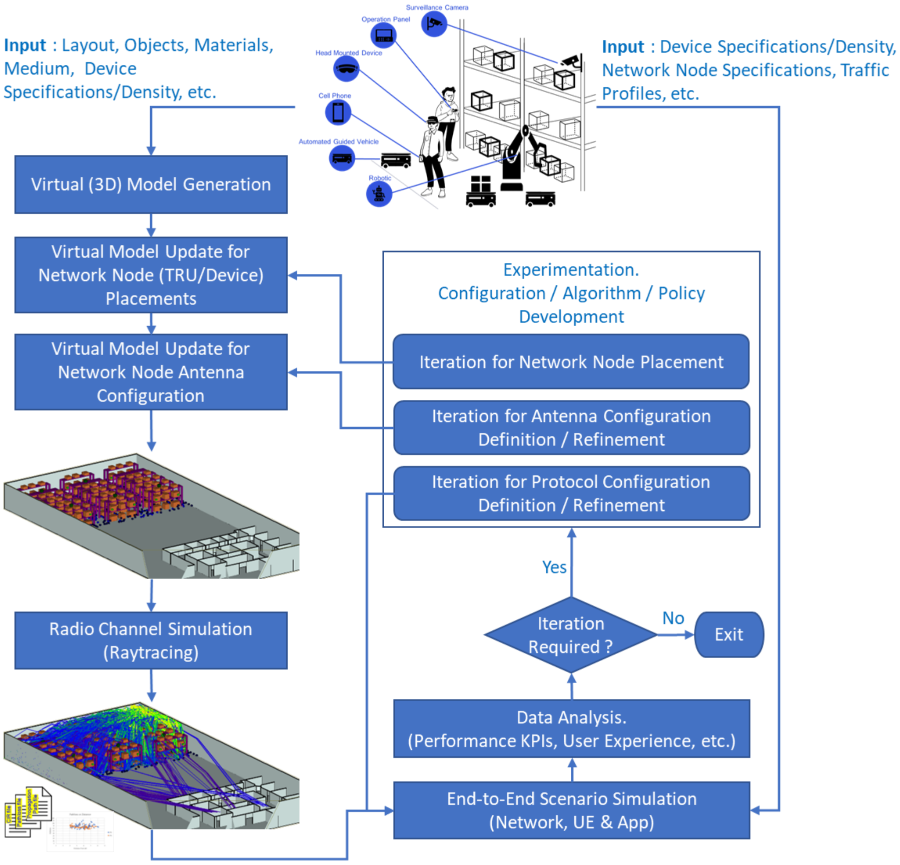

Figure 3.3.1 Virtual model for Warehouse & Office layout

NOTE 1: Virtual model for Warehouse & Office layout in Figure 3.3.1 was created using 3D modelling software, and radio channel simulation was done using Wireless channel modelling software.

NOTE 2: Traffic generators (Apps) for Warehouse & Office layout in Figure 3.3.1 were created using both 3GPP defined (refer [9], [10]) and custom traffic profiles. E.g., Custom AR/MR Head Mounted Devices (HMD) for Warehouse management.

Input Parameters:

DT of a real network requires following inputs to run simulations for defining optimal traffic offloading, enhanced mobility, and dynamic Tx power algorithms,

- Application Requirements
    - Required throughput, Latency/Delay bounds and reliability.
- Network/UE Requirements
    - User device density, traffic profiles and locations
    - Network & UE capability configurations (RATs, Antennas, Band Support, etc.)
    - Network node (5G, Wi-Fi) density and location
    - Monitored network node capacity in terms of throughput & latency.

Output parameters:

DT is used to simulate end-to-end traffic scenarios at every possible location in the network layout with different types of storage materials and storage occupancy. The simulations are used to do a lot of 'What-If' analysis and define thumb rules for designing a wireless communication network for Warehouse geometries. DT also generates a lot of synthetic data on multipath channel responses, pathloss, inter-cell interference, inter-device interference, etc., which are used for training AI/ML algorithms to provide:

- Optimal Handover triggers based on UE location & direction.
- Optimal threshold for network node (Wi-Fi Access Point/5G gNB) loading.
- Triggers for traffic offloading between network nodes (both intra/inter-RAT).
- Tx power recommendations for overall reduction in interference without creating coverage gaps.

These algorithms/models/Thresholds/Tx powers are deployed in the real-world network, and statistics/data are collected to provide feedback to DT.

3.4 Use Case 4: DT-RAN for Network Energy Saving

3.4.1 Background information

In the dynamic landscape of telecommunication networks, the integration of DT emerges as a transformative approach. This section explores the motivations and strategies behind leveraging DT for energy savings and delves into the evolving paradigm of O-RAN ALLIANCE to enhance overall energy efficiency. DTs, heralded as the next generation of network modelling tools, offer virtual replicas of real-world environments. Their applications extend to studying energy consumption, optimizing resource usage, and forecasting behaviour based on traffic fluctuations and user dynamics. As the industry transitions to 6G, sustainability becomes a central tenet, with a keen understanding that energy consumption significantly impacts operational costs, especially in the Radio Access Network (RAN). This section underscores the industry's response to the environmental impact of network operations. Efforts, such as the adoption of renewable energy sources and the implementation of energy-efficient design elements like sleep mode, are acknowledged. However, the study posits that more rigorous strategies are needed to align with industry goals of achieving Net-Zero emissions by 2040-2050. An analysis of wireless network breakdown into key areas, with the RAN consuming a substantial portion (50% in 5G), highlights the pivotal role of RAN in overall energy usage. O-RAN is introduced as a promising avenue for improving energy efficiency, emphasizing its prospects and challenges in design, development, and operational considerations. The European Telecommunications Standards Institute (ETSI) specifications [ETSI ES 203 228 and ETSI ES 202 706-1] focusing on energy efficiency and power consumption measurement, are detailed. Ongoing work on energy efficiency within O-RAN is highlighted, emphasizing its high priority in industry initiatives.

3.4.2 Motivation

The integration of DT in the network represents a critical strategy for reshaping energy efficiency within telecommunication networks. This deployment is spurred by a pressing need to confront the intricate challenges inherent in pursuing optimal energy efficiency without compromising superior network performance.

At the core of this approach is the comprehensive modeling capability of DTs, which creates a detailed digital replica of the physical network. This modeling extends to various aspects of the network environment, offering an emulated version that allows for non-disruptive analysis, evaluation, and optimization of design and operational options, particularly in the context of energy savings.

Furthermore, DTs play a pivotal role in evaluating diverse energy-saving strategies within a virtual model. This mimicking capability is instrumental in exploring what-if scenarios, providing stakeholders with insights into the impact of different strategies on the network. The virtual playground for experimentation empowers decision-makers to refine and optimize energy-saving initiatives.

The motivation for leveraging DTs in energy savings extends to their ability to estimate optimal models and thresholds through simulation and analysis. This continuous analysis enables stakeholders to strike the right balance between energy savings and end-user Quality of Service (QoS).

As the telecommunications industry aligns with global sustainability goals, DTs become integral in driving the sector towards sustainable practices. By offering a platform for virtual experimentation and scenario analysis, DTs empower decision-makers to make informed choices that harmonize the demand for energy savings with the imperative to meet end-user expectations.

Moreover, the integration of DTs in energy savings is underpinned by their inherent capability to enhance network performance. Insights into the impact of energy-saving strategies on QoS enable Communication Service Providers (CSPs) to adopt a conservative yet effective approach, ensuring energy savings without compromising the end-user experience.

Recognizing challenges associated with the implementation of energy-saving solutions, DTs provide a platform for CSPs to evaluate and fine-tune strategies. For example, DTs allow the simulation of underutilized resource shutdowns and their impact on end-user experience, mitigating suboptimal levels of energy savings.

In essence, the motivation behind incorporating DTs in energy savings use cases lies in their transformative capacity to evaluate, analyze, and optimize various strategies.

3.4.3 Proposed solution 1

The DT plays a crucial role in achieving optimal network energy savings, influenced by a blend of environmental, economic, and operational factors. It models diverse energy-saving strategies, simulating them within the DT network to determine optimal models and thresholds without compromising end-user Quality of Service (QoS). Evaluating overall network energy efficiency is essential, weighing the energy consumed by DT platforms against the gains. Literature research conducted assumes DT energy consumption is minimal compared to the achieved benefits, although comparing this consumption is beyond the study's scope. The energy-saving solution involves shutting down underutilized resources, posing challenges even with current AI/ML solutions, as runtime control and guaranteed 100% performance are difficult. Consequently, Communication Service Providers (CSP) opt for a conservative implementation, resulting in suboptimal energy savings.

The DT virtual replica will be mapped with real-network entities, environments, and their behavior. The real-time data is being collected from the network for traffic patterns and current level of QoS to represent the network behavior and patterns.

The DT application will represent the current state of the network. The user should submit its power saving requirement to DT and the DT runs various scenarios iterations to arrive at optimal state of the network for energy saving and analyzes its impact on network configuration changes, coverage, traffic migration and end-user QoS level. The DT application will leverage AI/ML techniques and model representation to arrive at an optimal state and can push the target parameters and configuration changes to the network to realize it through network configurator.

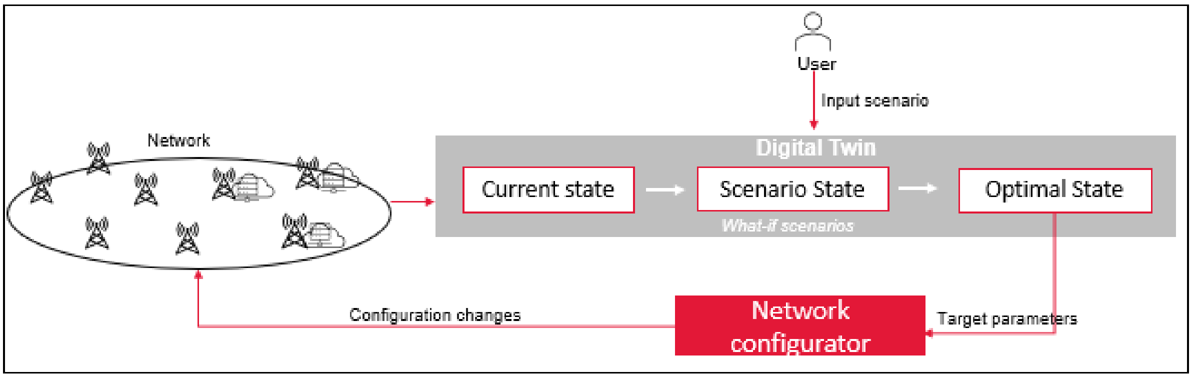

Figure 3.4.1 DT use-case workflow

Input parameters:

The DT may require the following inputs to replicate the real network in virtual model:

- Network configurations and topology.
- User device density, traffic profiles.
- Network & UE capability configurations (RATs, antennas, band Support, etc.).
- Specific service/user level SLAs.
- Traffic patterns/ KPIs and power consumption datasets.

Output parameters:

DT is used to simulate end-to-end energy saving scenarios at every possible location in the network with different strategies, algorithms, and thresholds. The simulations are used to do a lot of 'What-If' analysis and define thumb rules for arriving at optimal strategies, algorithms and thresholds for each system, module, or resources to achieve maximum energy savings without compromising the end-user QoS performance.

- Optimize the thresholds for shutting down resources.
- Optimal traffic/workload migration levels.
- Triggers for traffic offloading between network nodes (both intra/inter-RAT).
- Tx power recommendations for overall reduction in transmit power without creating coverage gaps.
- Criterions for wake-up of resources.

3.4.4 Proposed Solution 2

In this network energy-saving use case, the RAN shall simultaneously fulfil multiple requirements. On one hand, requirements from operator's customers specify a certain level of QoS to be met for specific applications. On the other hand, requirements from the network's operator, aim to optimize energy efficiency, minimizing energy consumption.

The complexity of the multi-objective optimization problem is exacerbated by the dynamicity of the RAN environment under which it shall be solved. In this scenario, users are constantly moving, and applications change their traffic distribution frequently. Hence, techniques that can use available RAN observability to deal with multiple (possibly conflicting) requirements are needed.

The intent is the formal specification of all expectations including requirements, goals, and constraints given to a technical system [7]. Using intent-based management the procedures for configuration of the RAN are transformed. Rather than manually setting technical parameters such as hand-over thresholds, this approach empowers service providers to specify the goals and required characteristics of a connectivity service using intents, including how to prioritize across users and services in conflict situations. Figure 3.4.2 shows the SMO receiving two intents to satisfy: QoS maximization from an application and energy-savings maximization from the CSP.

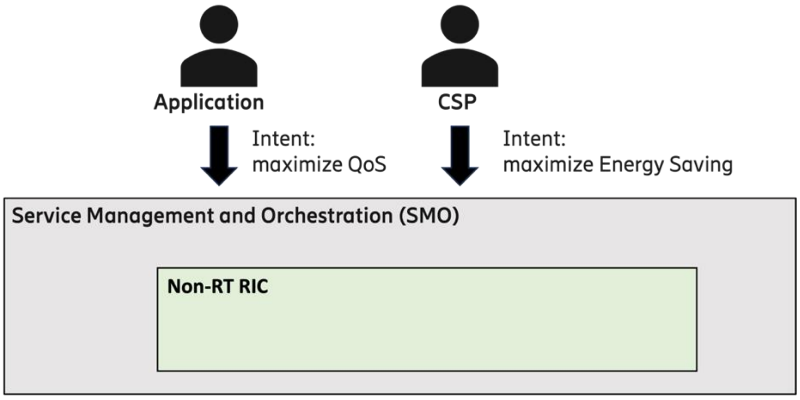

Figure 3.4.2 Intents and SMO: different, possibly conflicting, requirements are specified

Intents are represented in a machine-readable format (e.g., RDF) to allow machine interpretability and machine-to-machine communication.

Following the collection of requirements based on the received intents, intelligent agents within the network may propose actions to fulfil the requirements. These actions typically involve changes in the configuration parameters of the network, for example, adjusting the maximum transmission power of the radio base stations, or fine-tuning microsleep durations, etc. DT technologies provide a safe environment to evaluate the impact that each of these actions on the network. Specifically, DTs can be used to predict, given a certain configuration, the impact of actions on network KPIs, hence allowing informed decision-making for the controller responsible for decisions. This is important not only for determining the ability of a proposed action to fulfil the requirements as set by the intents for the specific application, but also for the impact that this action has on other intents that need to be fulfilled by the network.

In Figure 3.4.3, a DT continuously collects data from various real-world entities (e.g., power usage, resource utilization) but also traffic patterns and the current level of QoS. The observed state is used to update the functional model of the system's representation, allowing for the estimation of a new state based on the model. Consequently, the DT can be queried to respond to "what if scenarios" or be used as a predictor for proposed configuration actions.

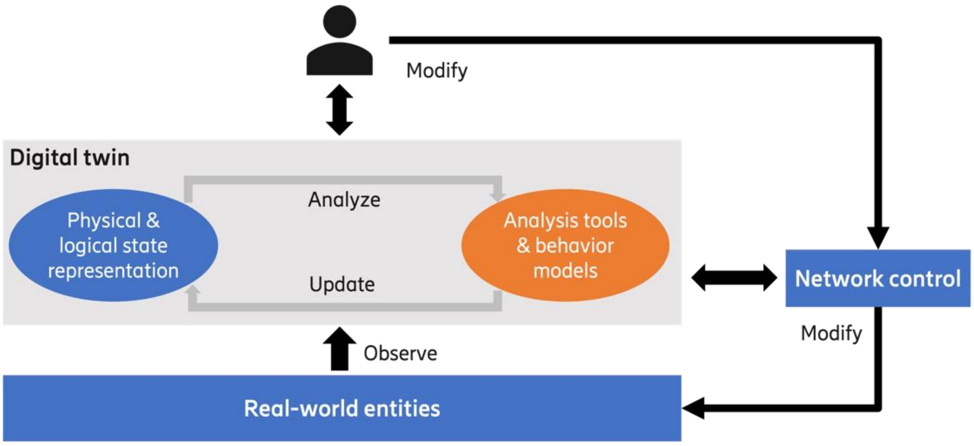

Figure 3.4.3 DT capabilities

Benefit of O-RAN architecture:

The main benefit of the O-RAN architecture is that it already provides a framework for hosting (multi-vendor) applications. Such a framework can be extended to host DTs.

As Figure 3.4.4 shows, DT instances can be hosted in the SMO in the form of rApps as the R1 interface can already support easy onboarding of multi-vendor applications [14].

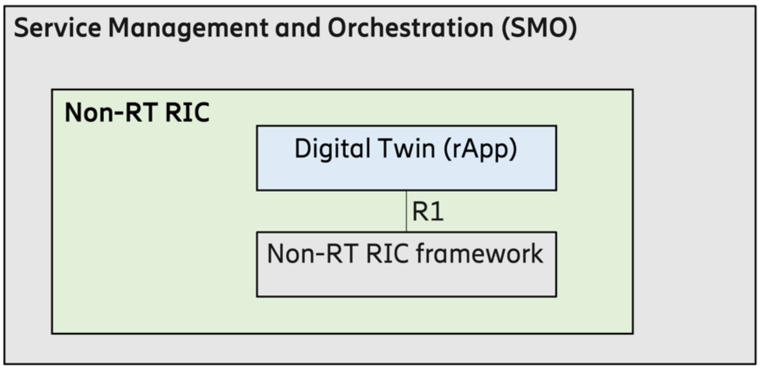

Figure 3.4.4 DT in the Non-RT RIC

There are two ways DTs can play a role.

The first is when the automation App uses internally a DT to perform decision-making. Indeed, O-RAN Apps steer the RAN behavior to certain targets by sending A1 policies. In the Near-RT RIC, xApps use E2 interface to control the RAN functionality (controlling E2 nodes operation) and aim to achieve the targets/objective received in the A1 policies from the Non-RT RIC. In this case, since the DT is used internally by the rApp, no need for standardization is required.

The second case is when the DT exports functionalities to external (management) entities (e.g., other rApps, OSS) as in Figure 3.4.4. In this case, it needs to be investigated what are the interfaces and data to be standardized to allow multi-vendor interoperability of DTs.

3.4.5 Proposed solution 3

The introduction of 5G technology marks a significant shift in mobile communications. This section explores the strategic necessity of optimizing energy efficiency while upholding network performance and user experiences. A major concern is the substantial energy consumption of base stations (BS) in conventional mobile networks according to GSMA reports. To address this, innovations from the 4G era, like carrier shutdown and channel shutdown, prove effective in reducing energy demands for 5G networks. New solutions, such as deep sleep and symbol aggregation shutdown, emerge in the 5G era, offering potent tools for energy efficiency. As global efforts intensify to combat global warming, the telecommunications sector, setting stringent energy standards, plays a leadership role. Despite the impending threefold increase in energy demand with 5G, leveraging insights from previous generations and embracing advanced energy-efficiency technologies provides hope for balancing data demands with sustainable energy practices. Confronting this challenge requires not just technological innovation but a collective global commitment to a greener and more sustainable future from a business perspective.

The development and deployment of 5G networks have ushered in a new era of mobile communication. In this dynamic landscape, it is essential to assess and model the power consumption of critical network components. Two key power consumption models are central to this evaluation: the Base Station (BS) power consumption model and the Carrier Aggregation power consumption model.

The Base Station Power Consumption Model:

At the heart of this model is the power consumption of a non-mMIMO (non-Massive Multiple-Input, Multiple-Output) Base Station. It involves a comprehensive calculation of energy utilization factors, encompassing all antennas within the BS. Each antenna is equipped with its Radio Frequency (RF) transceiver module, complete with a power amplifier (PA). This model further considers the power consumption of the Baseband Unit (BBU) associated with these antennas, the Direct Current to Direct Current (DC-DC) power supply, the active cooling system, and the Alternating Current to Direct Current (AC-DCC) unit facilitating connection to the electrical power grid. Notably, the following parameters are subject to variation:

- The number of RF transceiver modules within a BS.
- The input power of each transceiver.
- The transmit power.
- The power consumption of RF transceiver modules and BBUs.
- Power efficiency.
- Power losses in various components, such as the feeder, DC-DC power supply, mains supply, and active cooling.

The Carrier Aggregation Power Consumption Model:

The Carrier Aggregation power consumption model is indispensable in understanding how power utilization scales with the number of active Component Carriers (CCs). This model takes into account a host of variable parameters, including:

- The effective transmit power used by individual Component Carriers.
- The bandwidth allocated to the Component Carriers.
- The variable circuit power consumption, which scales linearly with the number of active Component Carriers.

In the ever-evolving landscape of 5G networks, power consumption modeling is a fundamental consideration. These models, such as the Base Station power consumption model and the Carrier Aggregation power consumption model, provide invaluable insights into energy utilization patterns and help shape the efficiency and sustainability of our mobile networks. By considering the diverse parameters within these models, we can better understand and optimize the energy footprint of 5G networks, thereby ensuring a greener and more efficient future for mobile communications.

Energy consumption in radio access networks (RANs) can vary based on several key parameters, including:

1. Network Traffic Load: The amount of data being transmitted through the network at any given time affects energy consumption. Higher network traffic often requires more energy to operate.
2. Network Coverage Area: The size of the area that the RAN needs to cover plays a significant role. Larger coverage areas may require more base stations and equipment, resulting in increased energy consumption.
3. Cell Density: The density of cells or base stations in an area can impact energy usage. Higher cell density can increase energy consumption due to the need for more equipment to manage connections.
4. Technology Generation: Different generations of mobile technology (e.g., 2G, 3G, 4G, 5G) have varying energy efficiency levels. Newer technologies like 5G are designed to be more energy-efficient, but they can still consume substantial power when operating at full capacity.
5. Antenna Configuration: The type and configuration of antennas used in the RAN can affect energy consumption. For instance, advanced antenna technologies, such as Massive MIMO, can be more energy-efficient than traditional ones.
6. Device Types: The energy efficiency of user devices (e.g., smartphones, IoT devices) interacting with the RAN can impact energy consumption. More energy-efficient devices can reduce the energy demands on the network.
7. Power Saving Features: The implementation of power-saving features in network equipment and devices can reduce energy consumption during periods of low activity.
8. Network Load Balancing: Effective load balancing among base stations can optimize energy consumption by distributing traffic efficiently.
9. Environmental Conditions: Environmental factors like temperature and humidity can influence the energy needed for cooling systems and equipment, especially in outdoor deployments.
10. Network Topology: The layout and structure of the network, including the arrangement of base stations and their proximity to each other, can impact energy efficiency.
11. Equipment Efficiency: The energy efficiency of network equipment and infrastructure, including base station hardware, power amplifiers, and cooling systems, is a critical factor.
12. Power Sources: The source of power, whether from renewable energy or conventional sources, can affect the carbon footprint and overall energy consumption of the RAN.

Optimizing these parameters and employing energy-efficient technologies and practices can help reduce the energy consumption of radio access networks while maintaining performance and coverage. To do this is a less disruptive scenario the DT modelling of these intertwined functions would help testing the optimal path for energy savings.

Energy Efficiency: Metrics, Measurement and Optimization Strategies

Many strategies are being considered for energy optimization in networks, both at the system and subsystem level, as well as a component and chip level. Switching off a cell or carrier, or some of the RF ports, advanced sleep modes both at the chip level and at the protocol stack, and efficient cloud resource management, use of AI/ML techniques for more efficient RF transceiver design and operation, are all being investigated by the industry.

Advanced Use Cases in Energy Efficiency

The work in energy optimization is in early stages, and a lot of innovation is expected as we move into the next generation of wireless deployments.

One interesting aspect is the interaction between the wireless network, and the energy network/grid. Just like the wireless network, there is a trend of disaggregation in energy sources as well, with local sources such as solar panels gaining wide usage. This includes the use of such local sources for wireless network equipment as well. The management of such sources, including decisions on using the local source vs. drawing power from the grid, involve energy management protocols on the energy, which have parallels to the wireless protocols themselves. The interaction between the energy grid and wireless network is likely to increase in complexity in the future.

Role of DTs

As described above, energy efficiency is a complex topic with many dimensions. The role of DTs will be vital, allowing for

(a) the modelling of various energy efficiency approaches in a virtual environment

(b) Creation of synthetic data using accurate DT models of network equipment and systems

(c) A mechanism for live optimization of the network

3.4.6 Proposed Solution 4

Achieving energy efficiency, coupled with minimization of energy consumption, typically covers the equipment, design and development, and operation, in addition to autonomy and alternative choices for energy, cooling, etc.

Similarly, in the context of O-RAN, this corresponds to energy efficient hardware (AAU/PA, servers, accelerators, etc.), energy-efficient design and deployment, including functional placement, optimal architecture, and integration, and intelligent software, cloud, and automation platforms (RICs) for energy efficient operation, in addition to the use of alternative sources of energy, cooling, etc.

The opportunity is about the ability to conduct scenario analysis, testing, and evaluation, and to draw insights and actions, in an energy efficient, secure, and flexible way. The potentially continuous output can then facilitate the introduction of new models and configurations, rollout of new features, allocation of resources, in addition to prediction and optimization. It should be noted that while the use case is highlighted here, the realization and enablement of this environment require addressing several challenges and requirements related to all aspects, including energy-efficient creation and preservation of the synchronized digital replica, the associated tools and capabilities, interfaces, access, and actuation, as well as data management and modelling.

3.5 Use Case 5: DT-RAN for Site-Specific Network Optimization

3.5.1 Background Information

Site-specific optimization of the radio access network (RAN) refers to the continuous process of fine-tuning network configurations on a per-site basis. The goal is to ensure optimal performance for individual cell-sites in terms of key performance indicators (KPIs) like coverage, capacity, quality-of-service (QoS), quality of experience (QoE) and energy efficiency. Which attributes are tunable and important for site-specific optimization depends on the RAN deployment architecture, as well as the capabilities of cell-site specific network elements. As cellular networks continue to evolve with the emergence and deployment of innovative technologies beyond 5G, site-specific RAN optimization is poised to become more sophisticated due to the increasing number of frequency bands, demanding KPIs (e.g., higher data rates, ultra-low latency, wider coverage, lower energy consumption, and high-accuracy positioning) and innovation of more advanced antenna technologies.

Performance evaluation of intricate ‘site-specific’ optimization operations requires a framework equipped with an automated analysis and feedback loop. DT stands out as one of the most promising technologies to enable that framework. DT for RAN (DT-RAN) is a dynamic digital replica of a real-world RAN that simulates the behavior, performance and characteristics of the physical RAN infrastructure, and aids in improving operations, maintenance, what-if analysis, and decision-making during the network lifecycle. The DT-RAN model for individual cell sites can be built using advanced technologies like machine learning, data analytics and advanced simulation techniques and validated by data collected from various site-specific sources (including network elements and surrounding RF propagation environment from the cell site).

3.5.2 Motivation

Traditional network optimization in existing cellular infrastructure involves conventional optimizers based on Constraint Programming (CP) or Integer Linear Programming (ILP) that typically operate based on rules defined manually by domain experts. These rules are usually static and generalized for the entire network. There are two key limitations of these network optimizers:

1) Translating network-wide optimization policies into site-specific rules tailored toward a particular deployment location is extremely cumbersome if done manually.
2) The optimizers typically restart the optimization process from scratch each time any change in the optimization scenario is triggered. This method is inefficient due to the lack of learning capability from past optimization scenarios and the lack of knowledge transfer capability to the new optimization task.

With recent advancements in AI/ML, several ML techniques have emerged, among which deep reinforcement learning (DRL) has stood out as one of the most promising candidates for advanced network optimization, that can leverage augmentation of general knowledge learned during the pre-training phase with continuous learning from the live network, offering site-specific network optimization policies in an automated and dynamic way. Even though such AI-powered network optimizers can largely overcome the shortcomings of traditional optimizers, they may suffer from the lack of stability at the preliminary phase. As one example, DRL-based network optimizers may show erratic behavior during the initial exploration phase.

To avoid the adverse effect of such unpredictability on a live network, DT-RAN can play a crucial role in providing an experimentation sandbox that faithfully mimics the behavior of the site-specific physical network. Within the safe zone of DT, a DRL agent can conduct the exploration and acquire necessary knowledge about the physical site, without the risk of catastrophic disruption. Post experimentation, the achieved policy can be safely applied to the underlying physical network.

Since the DT-RAN is continually updating its modelling through feedback from the physical site, any change in the site-specific network behavior is timely reflected in the digital replica. Thus, the network optimizer can routinely validate its performance within the twin domain to ensure the validity of its current optimization policies, and if needed, can update, and reevaluate the policies in a safe and timely manner before making critical changes to the live network.

3.5.3 Proposed Solution

3.5.1 General DT workflow

Figure 3.5.1 depicts a general workflow for site-specific network optimization using DT.

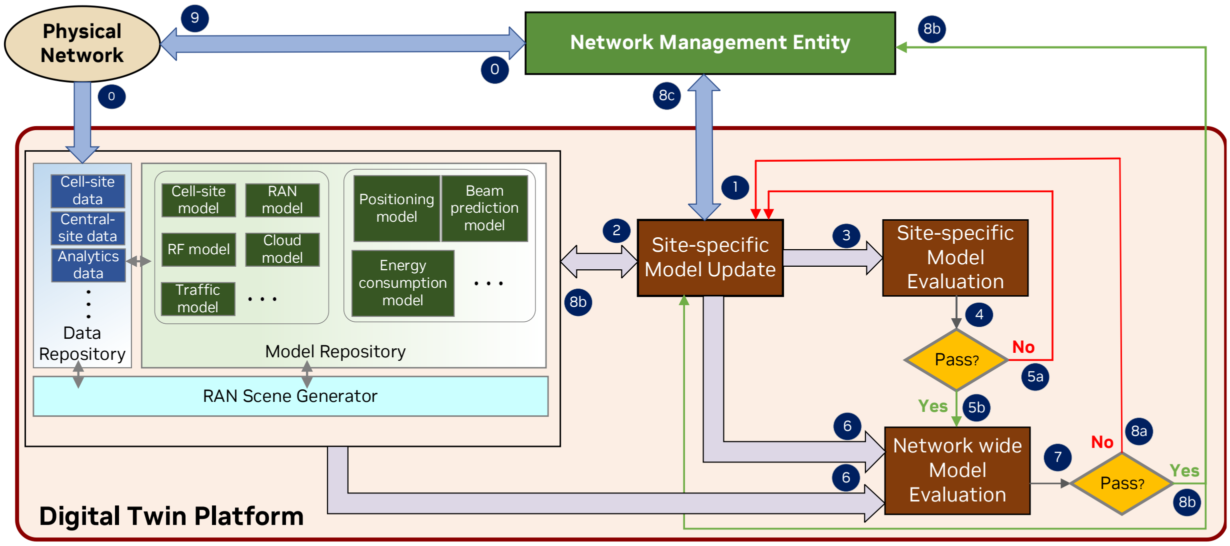

Figure 3.5.1 General DT workflow

- **Step 0:** Network management entity (e.g., SMO) continuously receives information from the live, physical network and monitors network performance. The underlying DT also collects relevant data regularly from the cell sites to keep its twin models up to date with the current physical network.
- **Step 1**: Network management entity may need to change the network policy due to several reasons (e.g., due to changes in the Operator's intent or due to changes in the physical network), which may impact network configuration for specific sites. To evaluate the impact of these changes before applying them to the physical network, the management entity triggers the DT platform to evaluate the impact of policy change on a specific network site.
- **Step 2**: DT-RAN collects necessary data (which could be either data already stored in the data repository or additional data collected from the live network, or a combination thereof depending on the use case), retrieves relevant models from the DT model repository that needs to be updated due to the network policy change and updates the model parameters to mimic the network configuration changes that are intended to be applied to the physical network.
- **Step 3**: DT-RAN evaluates the updated twin models and analyses the behavior of the network site in the twin domain.
- **Step 4**: DT-RAN compares the result of the model evaluation analysis against pre-defined performance threshold (which could be a single value or a range of values) to see whether the updated model meets the site-specific performance criteria as per the new network policy under test.
- **Step 5:** There are two possibilities:
  - **Step 5a:** If the outcome of step 4 is negative, i.e., the updated model does not pass the performance criteria, it goes back to step 3 and repeats from the process of site-specific model update (e.g., by retraining the AI/ML model used for model update). There could be a pre-defined number of iterations through which the model retraining could be retriggered of Step 5a happens. If the model does not meet the performance criteria after that predefined number of

iterations, the updated model is rejected, and the DT falls back to the pre-existing model.
    - **Step 5b**: If the outcome of step 4 is positive, DT-RAN next triggers the evaluation of the site-specific update in the context of network-wide performance impact (e.g., if site-specific beam pattern is changed, whether it would impact neighboring cells in terms of interference).
- **Step 6**: Collecting relevant data and additional models from the DT-RAN repository, the network wide evaluation of the impact of the site-specific model(s) is carried out.
- **Step 7**: DT-RAN compares the result of the model evaluation analysis against the pre-defined performance threshold to see whether the updated model meets the network-wide performance criteria.
- **Step 8:** There are three possibilities:
    - **Step 8a**: If the outcome of step 7 is negative, i.e., the updated model does not pass the performance criteria, it goes back to step 3 and repeats from the process of site-specific model update (e.g., by retraining the AI/ML model used for model update).
    - **Step 8b**: If the outcome of step 7 is positive, DT-RAN sends control signal to network management entity confirming that a new set of network configurations corresponding to the updated twin model can be applied to the real-network. At the same time, it also internally triggers model repository update with the revised model(s) reflecting the network configuration change.
    - **Step 8c**: If the outcome of step 7 is negative, i.e., the updated model doesn't pass the performance criteria, and the number of iterations has reached a pre-defined maximum limit, the updated model is discarded and a control message is sent to the network management entity suggesting that the intended network policy change cannot be implemented in the physical network, since an optimized set of network configurations that can implement the policy change and at the same time meet the performance criteria could not be found in the optimization process (i.e. in the site-specific model update process) .
- **Step 9**: If the previous step is not 8c, the network management entity triggers site-specific configuration changes in the physical network.

3.5.2 Application of DT workflow

3.5.2.1 Site-specific positioning

Site-specific network optimization can play a crucial role in improving positioning accuracy and reliability in wireless networks for 5G and beyond. Positioning in this context often refers to determining the location of mobile devices or Internet of Things (IoT) devices within the coverage area of the network.

The workflow of DT-RAN in site-specific optimization for positioning involves a series of steps and processes to create, maintain, and utilize the DT to enhance positioning accuracy within a specific environment.

Upon reception of a positioning intent from the network management entity, the DT-RAN can run simulations and tests within the DT to evaluate different configuration options and positioning algorithms. Afterwards, the DT-RAN can analyze simulation results to identify optimal network settings and positioning strategies. In this process, AI/ML can be integrated to analyze real-time and historical data from the DT and be used to optimize positioning algorithms, adapt to changing conditions, and predict potential issues.

Input Parameters:

The data for positioning may include the following:

- Various measurements (such as downlink reference signal time difference (DL-RSTD) measurements, uplink relative time of arrival (RTOA) measurements, downlink positioning reference signal received power (RSRP) measurements, uplink sounding reference signal (SRS) angle-of-arrival (AoA) measurements, UE Rx-Tx time difference measurements, gNB Rx-Tx time difference measurements, channel impulse response measurements, and line-of-sight/non-line-of-sight (LOS/NLOS) measurements),
- Time stamps of the measurements,
- Ground-truth labels, and
- Configuration and deployment information (such as transmission and reception point (TRP) coordinates, beam azimuth and elevation angular information, and reference signal configuration).

Output Parameters:

The output of the DT-RAN for positioning may include the following:

- Recommended positioning methods (such as downlink time difference of arrival (DL-TDOA), downlink angle of departure (DL-AoD), uplink time difference of arrival (UL-TDOA), uplink AoA, round-trip time (RTT), direct AI/ML based positioning, AI/ML assisted positioning, or a combination thereof)
- Associated configuration options (such as frequency bands, transmit power levels, antenna characteristics, and reference signal configuration).

With the findings from the DT simulations and analyses, one can implement the optimized positioning strategy in the physical RAN and continuously monitor the positioning performance in the real-world using data from sensors and network equipment. Then the RAN can compare the actual performance with the predictions made within the DT and use the feedback loop between the real-world system and the DT to iteratively improve positioning accuracy and optimize network parameters. In particular, as conditions change or new data becomes available, the DT can be updated, and the configuration options and positioning algorithms can be reevaluated.

3.5.2.2 Site-specific probing beams for mmWave beam alignment

Engaging mmWave frequency bands requires the use of high-dimensional antenna arrays in order to accommodate the harsh wireless channel propagation characteristics associated with high frequency bands. These systems rely on highly directional beamforming in order to maintain a viable signal strength at the receiver. Both the downlink and uplink employ transmit and receive side beamforming respectively. An initial step in opening the communication link is to determine the combination of transmit and receive beams that maximize the signal-to-noise-ratio (SNR) of the link.

One way to accomplish initial access (IA) is via an exhaustive search. That is, every possible combination of beam pairs is tested to find the pair with the best SNR. The large number of beam combinations makes this a time-consuming process. Naturally, as the industry moves to higher frequencies (possibly the sub-THz band for 6G), the dimension of the transmit and receive antenna arrays continues to expand, thereby increasing the beam search time at an even faster pace. Therefore, we are motivated to find a better low-latency strategy for IA.

One such approach is reported in [8]. In this scheme a site-specific probing code book is learned using a complex valued neural network. The learned codebook consists of site-specific probing beams that capture particular characteristics of the propagation environment. Beam sweeping measurements from the probing codebook are used to predict the optimal narrow beam. An end-to-end neural network is used to jointly train the probing codebook and the beam predictor. The narrow beam employed in the communication link itself is formed using weights from a conventional beamforming codebook. This method has the advantage of minimizing the beam pairing latency.

Learning the probing codebook and the weights for the beam prediction neural network (NN) can be performed in DT-RAN. The geometry of the coverage area of a particular gNB is rendered in a DT and radio frequency (RF) ray tracing is employed to reveal the channel propagation characteristics. Periodically, the DT training pipeline can be run to update both the probing codebook and the beam prediction neural network.

Input Parameters:

The input to the DT-RAN for beam alignment may include the following:

- The geometry of the base station coverage region. This would include - descriptions of buildings, foliage, other structures in the scene accompanied by the material properties that influence EM wave propagation (e.g., permittivity, conductivity, and other material properties).
- The location of the base station in the scene.
- Base station antenna configuration and the radiating pattern for the elements that compose the antenna panel.
- Typical UE configurations and location distributions
- Trigger for probe codebook/beam predictor NN re-training

Output Parameters:

The output for the DT-RAN for beam alignment may include the following:

- Channel probing table with a typical number of entries between 10 and 20.
- The neural network weights for the beam predictor network.

4 Conclusion

The report provides an introduction and analysis to the 6G DT-RAN use cases. It is identified that DT-RAN is an important enabler for 6G in many areas, e.g. AL/ML training, evaluation and performance assurance, network automation, network planning, network energy saving and site-specific network optimisation. The motivation and value analysis of each identified DT-RAN use cases are provided and related high-level solutions are proposed which will be considered for the future DT-RAN research phases and normative standardisations.

References

[1] ITU-R, “Future Technology Trends of Terrestrial IMT Systems Towards 2030 and Beyond,” *Report ITU-R M.2516-0*, Nov. 2022.

[2] X. Lin, L. Kundu, C. Dick, E. Obiodu and T. Mostak, "6G Digital Twin Networks: From Theory to Practice", *IEEE Communications Magazine (Early Access)*, *pp.: 1-7, doi: 10.1109/MCOM.001.2200830.*

[3] ITU-T, “Digital Twin Network – Requirements and Architecture”, *Recommendation ITU-T Y.3090*, Feb. 2022.

[4] C. Robinson, L. Bonati, T. van Nieuwstadt, T. Reiss, P. Johari, M. Polese, H. Nguyen, C. Wat- son, T. Melodia, "eSWORD: Implementation of Wireless Jamming Attacks in a Real-World Emulated Network", Proc. of *IEEE Communications and Networking Conference (WCNC)*, Glasgow, Scotland, March 2023.

[5] L. Baldesi, F. Restuccia, T. Melodia, "ChARM: NextG Spectrum Sharing Through Data-Driven Real-Time O-RAN Dynamic Control," *Proc. of IEEE Intl. Conf. on Computer Communications (INFOCOM)*, May 2022.

[6] D. Villa, M. Tehrani-Moayyed, C. Robinson, L. Bonati, P. Johari, M. Polese, S. Basagni, T. Melodia, "Colosseum as a Digital Twin: Bridging Real-World Experimentation and Wireless Network Emulation," arXiv:2303.17063 [cs.NI], pp. 1-15, March 2023.

[7] TM Forum Autonomous Network Project, “Intent in Autonomous Networks”, IG1253, v1.1.0, January 2022.

[8] Y. Heng, J. Mo, J. G. Andrews, “Learning site-specific probing beams for fast mmWave beam alignment”, *IEEE Transactions on Wireless Communications* 21 (8), 5785-5800

[9] 3GPP TR 22.804, “Study on Communication for Automation in Vertical domains”

[10] 3GPP TS 22.104, "Service requirements for cyber-physical control applications in vertical domains"

[11] T. Chen, S. Kuklinski, E. Pateromichelakis, K. Samdanis, A. Kourtis, N. Nikaein, and A. Skarmeta. "Service-oriented Architecture Evolution towards 6G Networks." In 2023 *IEEE Conference on Standards for Communications and Networking (CSCN*), pp. 8-14. IEEE, 2023

[12] O-RAN.WG1.Use-Cases-Detailed-Specification

[13] L. Bonati, M. Polese, S. D'Oro, P. Brach del Prever, T. Melodia, "5G-CT: Automated Deployment and Over-the-Air Testing of End-to-End Open Radio Access Networks," IEEE Communications Magazine, pp. 1-7, April 2024

[14] O-RAN.WG2.Non-RT-RIC-ARCH