White Paper

June 2022

RIC and Connection Management xApp Technical Practice of China Unicom's CUBE-RAN Intelligent Open Platform

—overall scheme and load balancing optimization

Foreword

The wireless cloud network features openness and intelligence. As an important part of the wireless cloud network architecture, the Intelligent Open Platform works to manage and control the radio resources of base stations. By detailing specific business scenarios, deployment plans and practical implementations, this white paper explores China Unicom's applications of the Intelligent Open Platform in the wireless cloud network, elaborates the development of the wireless cloud network, and describes how the Intelligent Open Platform is used in base station handover and load balancing.

Authors:

China Unicom: Yue Wu, Rong Huang, Bin Fan, Youxiang Wang, Shan Liu, Hui Liang, Huanhuan Liu, Ruihua Li, Bo Pang, Xianda Li, YuShan Pei, Qian Huang, Gao Chen, Hanwen Chang

Intel: Jing Zhao, Jiming Gao, Chuang Dong, Hosein Nikopour, Hassnaa Moustafa, Mandar Chincholkar, Sonia Bhattacharya, Shaohe Feng, Krishna J Murthy, Zhongbao Shi, Wei Zhao, Hongli Zeng, Xianghui Sun, Ke Zhang

New H3C: Guangjun He, Danling Huang, Changhua Tong, Xiaoping Zhang, Yanglin hu, Tao Ren, Wei Xu, Xiuzhe Li

Table of Contents

Development of the Wireless Cloud Network

Wireless Network Capability Exposure

Thoroughly tapping into and properly exposing the in-depth capabilities of wireless networks to unleash their potentials and pipeline values will be very useful for operators to remain competitive amidst the overwhelming trend of mobile Internet. Having realized the value and trend of network capability exposure, operators around the world are actively building an open network capability ecosystem. From the perspective of radio access network, capability exposure involves three aspects:

1. Air interfaces exposure: Air interfaces are the most valuable resources for operators. In the face of complex customer groups and business requirements in the future, open air interfaces help allocate time, frequency, and space resources and allow customers to design their own requirements and customize air interface resources.

2. Basic resource exposure: Deploying high bandwidth and low latency businesses on the radio access side helps improve business experience. Third-party applications can be deployed at the same location and on the same platform with the wireless cloud network. As the wireless cloud network platform becomes more common and cloudified, it is possible for third-party applications to share the underlying resources with it. By exposing computing, network, and storage hardware or virtual resources to third parties through sharing or customization, their applications can be deployed at the network edge on the same platform with the wireless protocol software features to satisfy customer demands, while helping operators better take advantage

of the resource pool, potentially reduce business deployment cost, and discover new business models.

3. Network information exposure: Exposing Opening up network information allows user and subscription information to be used for user and business behavior analysis for third-party customers, so that they can autonomously manage their users. In addition, network information can provide channel measurement data and QoE information to help third-party customers optimize their business experience. Moreover, exposing network information also enables developers to innovate applications on the wireless side. For example, AI can be introduced for model training and inferencing of the wireless information and data, and the inference result can then be used for optimizing base station radio resource management and control.

Wireless network capability exposure can be used for various purposes, such as radio access information subscription of local multi-access edge computing (MEC) applications, and radio resource management with the Intelligent Open Platform of wireless cloud network leveraging data reported from base stations.

Business Driving Forces and Requirements

Currently, some standard and open-source organizations have started to study and develop the Intelligent Open Platform architecture, standards, and business applications. Some operators outside China have already conducted integration tests based on the Intelligent Open Platform architecture with device manufacturers, and have preliminarily verified radio resource management and application using an Intelligent Open Platform.

This white paper is about the release of the overall scheme and technical practices of integrating an Intelligent Open Platform with base stations led by a domestic operator in China for the first time. It is mainly driven by the demands of existing network businesses, with special emphasis on connection management optimization for cell handover triggered by mobile user equipment (UE) in multi-cell scenarios. Rapid network traffic increase, mobile UEs, and multi-band cells have made it particularly challenging to achieve load balancing and improve the cell edge performance while addressing the changes in the wireless environment and application requirements. Conventional traffic control includes cell reselection and handover, load calculation, and cell priority setting. These solutions are cell-centered rather than network-centered. Therefore, they rarely involve predicting the future network and UE performance and can hardly adapt to different user requirements and optimization objectives.

To support new businesses and requirements, applications on the wireless side are expected to utilize wireless information in a more efficient and intelligent manner. For example, using AI for wireless data training and modeling at the UE level to support network-centered use and optimization of radio resources.

China Unicom's Intelligent Open Platform

The basic idea of the Intelligent Open Platform is to expose and move some of the RRM capabilities previously placed inside the base stations to the Intelligent Open Platform, which gives a certain level of openness and scalability to radio resource management. Moreover, this concept creates infinite possibilities for the use and management of radio resources, for example, in network intelligence, network autonomy, wireless App application, and interaction with external information sources.

China Unicom's Intelligent Open Platform is based on the Open Networking Foundation (ONF) RAN Intelligent Controller (RIC) platform, an open-source program built on the ONF Open Network Operating System (ONOS). Its open-source building blocks can be integrated with third-party base stations, Apps, and RICs for prototype testing and interoperability verification. In this case study, the new mobile handover (MHO) service model and third-party Apps have also been introduced, while AI technologies such as graph neural network (GNN) have been used for model training and inferencing to help optimize the mobility management and improve the cell edge performance and load balancing.

In early 2022, China Unicom initiated the cross-vendor integration test of the Intelligent Open Platform, Apps, and base stations. The test initiative has involved Intel and H3C. The Intelligent Open Platform is built on the latest version of ONF RIC; the App is Intel's Connection Management (CM) xApp installed on the Intel® Smart Edge for Builders platform; and the base stations and core network are provided by H3C.

The test runs in several phases. Phase I focuses on the basic load balancing functionalities. The main objective is to verify the basic capabilities and overall process of cell handover control with the Intelligent Open Platform. Phase II looks at the load balancing performance. The main objective is to verify the load balancing performance of multiple cells and multiple UEs under more optimization policies such as QoS based policies. As of June 2022, the main objective of the Phase I testing has been primarily accomplished, meaning the end-to-end functional testing and verification for the Intelligent Open Platform based on the latest E2AP 2.0 interface, CM xApp, ONF RIC, base station, and core network have been completed, the basic performance of load balance is verified as well.

Testing environment for Phase I is as follows:
1) Frequency bands of 3.3 GHz–3.6 GHz, allocated among three cells (Cell A, Cell B, and Cell C), 100 MHz per cell. These 3 cells operate on different frequencies to avoid intra-cell interference; 2) The three cells are configured with the same transmit power, the same transmit mode for each UE at each cell; 3) The number of UEs in each cell: UEs of Cell A < UEs of Cell C < UEs of Cell B; Cell B traffic is set close to overload (100% > Cell B traffic > 95%), so that the UE performance when handing over to Cell B deteriorates.

The objective of Phase I is using GNN AI model inferencing with the wireless xApp. With AI, when the Intelligent Open Platform sends the RIC Control Request message to the base station, and a UE in Cell A moves from Point E to Point H (as shown in Figure 1.3.1), there's a bigger chance that it would hand over to Cell C instead of Cell B, thus improving the UE performance at the cell edge and the load balancing performance of all cells.

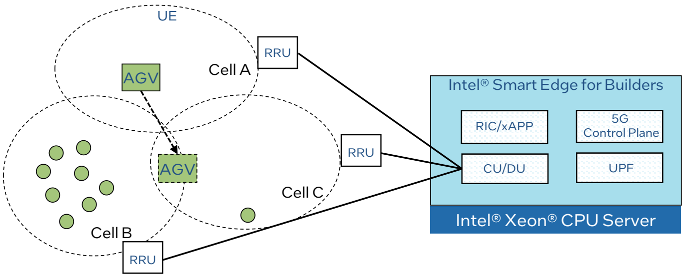


Figure 1.3.1 Phase I test deployment

Integration with the Intelligent Open Platform

Business Scenarios and Deployment

Cellular wireless connections provided by macro cells are plagued by problems such as incomplete coverage and call drop. Though high-density deployment of small cells can be a solution, it gives rise to connection issues. Connection management refers to the management of connection functions between UEs and the wireless network. Conventional connection management revolves around UEs: UEs send handover requests when detecting that the wireless channel quality is poor, and then the base station processes the request. UEs determine the destination cell based on the reference signal received power (RSRP) and establish connections with cells. When the UE moves, the RSRP it receives from the serving cell decreases as it goes farther away from the base station, while the RSRP of the destination cell increases as the UE approaches the destination cell. In the case where multiple destination cells meet the handover requirements, the one with the highest RSRP will be chosen with the conventional connection management approach. It is simple and effective, but it does not take into consideration the status of the current cell and the network as a whole, or if the choice produces the optimal result for load balancing and cell edge coverage. Such a drawback is particularly troublesome when user and traffic distribution is uneven.

How to indicate the best outcome factoring in throughput, coverage, and load balancing of the entire network from a global perspective has become a hot topic for 5G deployment. This test has used the Intelligent Open Platform and AI algorithms to optimize cell-UE connection via controlling handover, with an aim to achieve the optimal network performance and guarantee better service quality for most UEs on the network.

The Overall Scheme

The overall scheme consists of China Unicom's Intelligent Open Platform, Intel® Smart Edge for Builders Cloud Native platform, and H3C Unicell 5G Indoor Wireless Distribution System. The Intelligent Open Platform is responsible for centralized RAN management of the entire or a regional network. With the prediction capability of big data and AI, the platform can control and manage the wireless network with high granularity to improve the overall performance.

The scheme discussed in this paper is designed to optimize connection management of UEs under cell handover scenarios. With the support of an awareness-based connection management framework using machine learning (ML), it abstracts the wireless network into an AI neural network and uses GNN and reinforcement learning (RL) to control UE handover for intelligent and active connection management.

This Intelligent Open Platform has the following features:

1. Openness: When designing the architecture, open southbound and northbound interfaces have been designed to be standardized, so that the platform can integrate with third-party modules as long as they comply with the standards. Northbound interfaces handle business management and the orchestrator, so that RAN capabilities can be opened to user businesses to allow for quick development and adjustment of different businesses, while southbound interfaces connect to E2 nodes to configure and manage base stations from different vendors. The modular design for internal functions and standardized internal interfaces help facilitate the introduction of third-party modules.
2. Intelligence: With AI learning and inferencing capabilities, the overall performance of the wireless network can be improved. AI plays a vital role in global optimization, traffic prediction, network slicing, energy conservation, and emission reduction. This test has used AI to optimize the cell handover policies for network control with higher granularity.
3. Cloud native design: With cloud native design, the software and hardware are decoupled; the hardware is unified; the scaling of software features can be done on demand; hardware resources can be flexibly reused; and the cloud native architecture also provides openness for the platform.

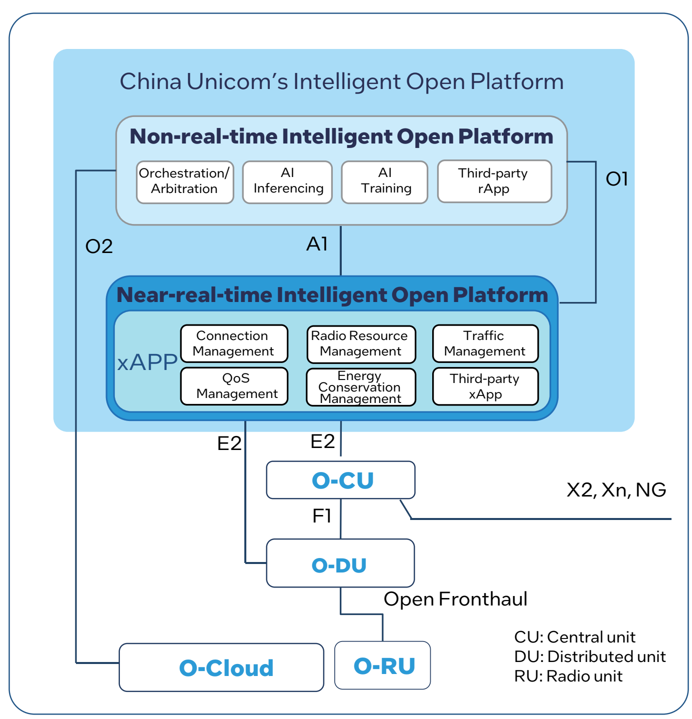


Figure 2.2.1 The overall scheme

Near-real-time Intelligent Open Platform

Deployed on Intel® Smart Edge for Builders Cloud Native platform, the Near-Real-Time (Near-RT) Intelligent Open Platform leveraging ONF SD-RAN project consists of the following functional modules: interface terminations (E2 termination, A1 termination, and O1 termination), databases

(UE-NIB and R-NIB), messaging infrastructure, conflict mitigation, subscription management, management service, API enabling, security, and platform UI. Figure 2.3.1 shows the architecture of the Near-RT Intelligent Open Platform. The functions in blue are already enabled or partially enabled, while the functions in gray are not yet enabled. The modules use gRPC to communicate with one another.

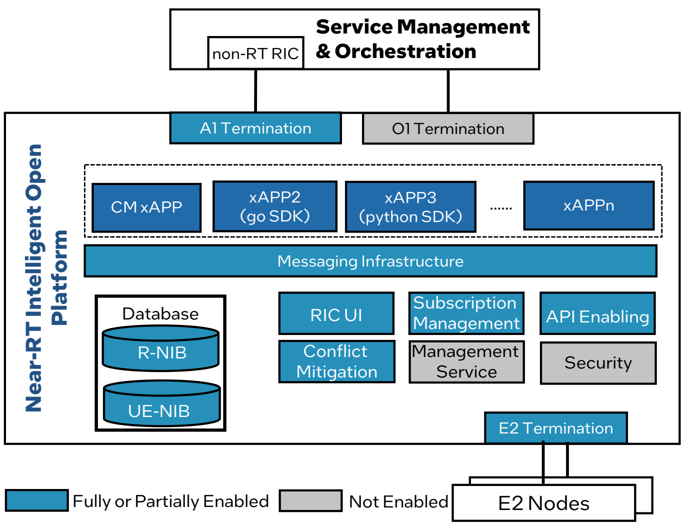


Figure 2.3.1 The functionality architecture of the Near-RT Intelligent Open Platform

This test mainly uses the E2 termination, R-NIB, UE-NIB, API enabling, and xApp. As non-real-time (non-RT) RIC is not involved, the A1 termination is not used. And as only one xApp is used, conflict mitigation is not used either.

The Near-RT Intelligent Open Platform provides a complete set of E2 termination functions. The E2 termination acts as the interface agent and protocol adapter to manage the message exchanges between the components of the Near-RT Intelligent Open Platform and E2 nodes. The E2 termination implements the E2AP specification (ASN.1 over the Stream Control Transmission Protocol [SCTP]) in the southbound direction, and the E2T API of the API enabling module in the northbound direction. Messages that go southbound through the E2 termination nodes are converted from Protobuf into ASN.1, while messages received from the E2 nodes are converted from ASN.1 into Protobuf in the E2 termination. This process can be scaled to the service models (E2SM) that exist as add-ins. E2SM provides the ASN.1 codec capability in the form of LIB, so that the E2 interface management is modularized for easy maintenance. Currently, the Near-RT Intelligent Open Platform supports E2 service models including E2SM-KPM, E2SM-RC, E2SM-NI, E2SM-MHO, and E2SM-RSM, allowing it to adapt to xApp requirements in various scenarios. In this test, the E2SM-MHO service model is used.

E2SM-MHO is an E2 service model designed for intelligent handover. For details about the message exchange process, see Page 12.

Based on the original E2AP, the Near-RT Intelligent Open Platform improves the "RIC Control Failure" message by adding the Cause required by the handover scenario. The message has also been verified from end to end.

The database of the Near-RT Intelligent Open Platform stores the information about UEs and E2 nodes, so that xApp can quickly access data from the local database. UE-NIB stores the meta-information of UEs; R-NIB stores the meta-information of the E2 termination, E2 nodes, and cells, as well as their topology; and the E2 termination stores the meta-information of subscription or channels. They have all exposed gRPC interfaces for data transfer. The xApp calls the gRPC interfaces provided by the E2 termination, R-NIB, or UE-NIB to act on meta-information. The subscription and control operations require protocol conversion from gRPC to ASN.1 through the E2 termination, after which messages are forwarded to E2 nodes. Then, E2 nodes report base station measurements through the corresponding channel. In this test, three channels are used to manage three types of subscription and their corresponding information in RIC INDICATION.

The messaging infrastructure provides low-latency message transfer service between internal endpoints of the Near-RT Intelligent Open Platform. It supports the registration, discovery, and deletion of endpoints. It provides the following APIs: API for sending messages to the messaging infrastructure; API for receiving messages from the messaging infrastructure. It supports a number of messaging patterns, including the point-to-point pattern (e.g. message exchange between endpoints), publish–subscribe pattern (e.g. real-time data distribution from the E2 termination to multiple subscribers' xApps), and message routing (i.e. distributing messages to endpoints based on the message routing information). The messaging infrastructure supports robust messaging, which helps avoid data loss when the messaging infrastructure is interrupted or rebooted, and frees up resources after the message expires.

Here's how a connection between the E2 termination and an E2 node is established: The E2 node sends an E2 SETUP REQUEST to the E2 termination for an SCTP connection. The endpoint of the E2 termination includes the pre-configured E2 termination service IP and SCTP association port. After the E2 termination responds with the E2 SETUP RESPONSE, the first SCTP association (management connection) is established. Through this association, the E2 termination sends the E2 CONNECTION UPDATE messages to the E2 node one by one to notify it of the IP and SCTP association port of each E2 termination Pod. After the E2 node responds with the E2 CONNECTION UPDATE ACKNOWLEDGE, a new SCTP association (E2AP connection) is established. The number of new associations is the same as that of E2 termination instances. Subsequent E2 messages are transferred through the new SCTP associations, while the first SCTP association maintains the heart beat messages.

Multiple E2 termination Pod instances are required if there is more than one E2 termination. In this test, the message structure has been simplified based on the specific deployment scenario. As there is only one E2 termination instance, the Pod IP and service IP both direct to the same instance. Therefore, we can merge the management connection and E2 connection into one, as is shown in Figure 2.3.2. The result turns out that the system has managed to work as expected.

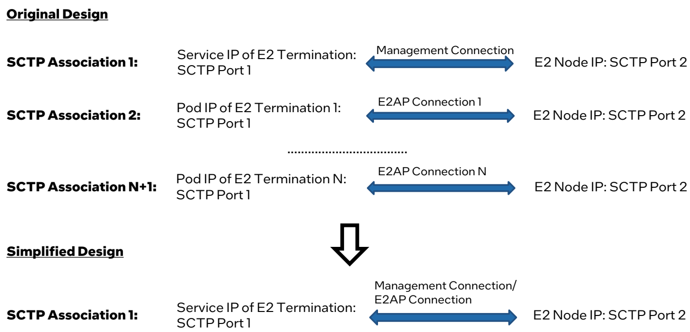


Figure 2.3.2 Simplification of the messaging infrastructure in the test

The Near-RT Intelligent Open Platform supports the xApp SDKs, including the Go SDK and Python SDK. This gives the platform the following advantages: 1) xApp providers do not need to know the details about the internal interactions of the Intelligent Open Platform or customize xApps for the platform, which provides excellent portability; 2) Development is made easier with less dependency on the platform, allowing OEMs to easily integrate a large number of xApps. Figure 2.3.3 on the right top shows the xApp SDK framework. In this test, the CM xApp has used the Go SDK of the Near-RT Intelligent Open Platform, and the code has been adapted.

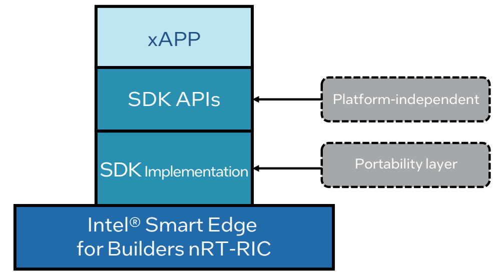


Figure 2.3.3 xApp SDK framework

The Near-RT Intelligent Open Platform has its own command line interface (CLI). Via the CLI, the topology, configuration, and control relationships between subsystems can be accessed, and information on cells, UEs, subscriptions, as well as the topology of E2 termination/E2 nodes and cells can also be checked. In this test, the CLI has been substantially used for debugging.

Connection Management xApp

For any wireless network, it is essential to ensure smooth communication and load balancing through connection management. Conventional connection management connects every user to the candid cell with the highest Reference Signal Receiving Power (RSRP), without considering the overall network performance, UE Quality of Service (QoS), coverage optimization at the cell edge, or load balancing among cells.

Using ML and AI solutions to optimize network performance is now a popular research direction in wireless communication.

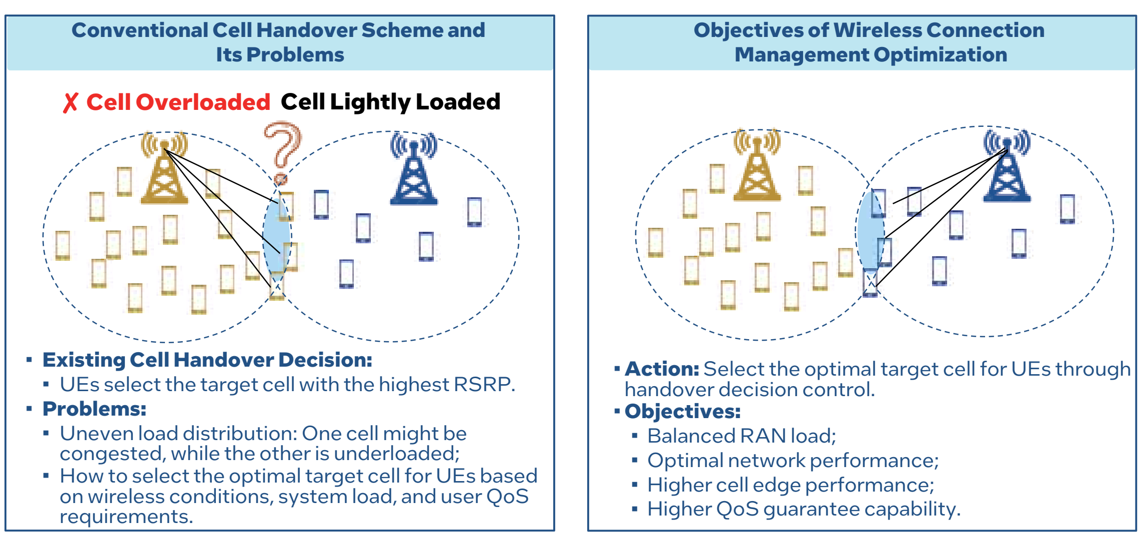


Figure 2.4.1 Objectives of wireless connection management optimization

Intel® Labs' connection management xApp (CM xApp) aims to optimize connection management with ML and AI. As shown in Figure 2.4.1, to resolve the aforementioned problems with conventional wireless communication handover schemes, the CM xApp is designed to support customization based on optimization policies to meet various user requirements, such as RAN load balancing, network throughput optimization, higher UE performance at the cell edge, and selection of handover target cells based on UE QoS.

The CM xApp formulates connection management as a combinatorial graph optimization problem. It provides a ML/AI solution that uses the underlying graph to learn the weights of the graph neural networks (GNN) for optimal user-cell association.

Graph Neural Networks are a framework to capture the dependence of nodes in graphs via message passing between the nodes. Unlike deep neural networks, a GNN directly operates on a graph to represent information from its neighborhood with arbitrary hops. This makes GNN an apt tool to use for wireless networks which have complex features that cannot be captured in a closed form.

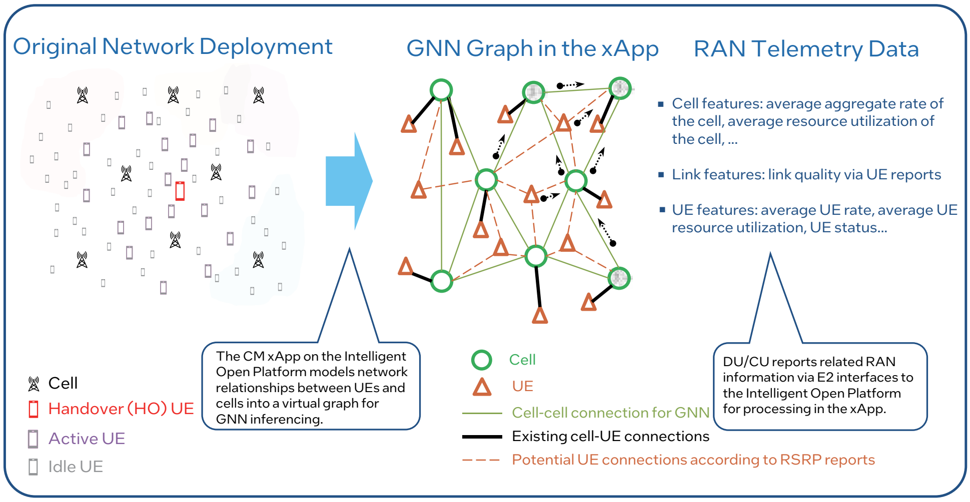


Figure 2.4.2 CM xApp based on GNN and RL

As shown in Figure 2.4.2, UEs and cells are abstracted as nodes in the GNN , and the (virtual) relationship between UEs and cells as well as that between cells are abstracted as edges in the GNN. The relationship between UEs and cells tells whether UEs are attached to cells, and that between cells tells whether they are neighbors. All the relationship information can be obtained from the messages reported to the Near-RT Intelligent Open Platform from the RAN side via E2 interfaces. Cell features (e.g. average aggregate rate of the cell and average resource utilization of the cell), link features (e.g. signal to interference & noise ratio (SINR) and RSRP), and UE features (e.g. average UE rate, average UE resource utilization, and UE status) can all be obtained from the messages reported to the Intelligent Open Platform from the RAN side via E2 interfaces. The information enriches the graph features from all aspects. In Phase I where the CM xApp has undergone simplified implementation, RSRP and other measurements from serving cells and neighboring cells are used as the RAN telemetry data. In Phase II, such data is collected as much as possible to improve the performance of the algorithm for wider application in the wireless environments.

The CM xApp provides intelligent optimization solutions for different optimization policies. It supports the following 3 optimization policies (QoS based Connection Management is under development):

1) Total network throughput — sum of all users' throughput.
2) Cell edge coverage — we usually use the data rate of UEs in the 5th percentile (only greater than 5% of the UEs on the network) as the measurement indicator of cell edge coverage.
3) Load balancing — it indicates distribution of traffic load across cells. The closer the value is to 1, the more balanced the load is.

Each optimization policy corresponds to a unique reward function. Those reward functions are calculated (based on Ran parameters) for training and inferencing.

In this trial, load balancing is used as the handover optimization policy.

The CM xApp is scalable and allows different optimization policies to choose different training methods and then provides inferencing results that RAN needs.

The CM xApp accelerates inferencing with Intel® OpenVINO™ toolkit. Its processing latency is proved to be less than 10 ms with 99% probability. Figure 2.4.3 shows the improvement of Python inferencing latency that the CM xApp enables with the help of OpenVINO™.

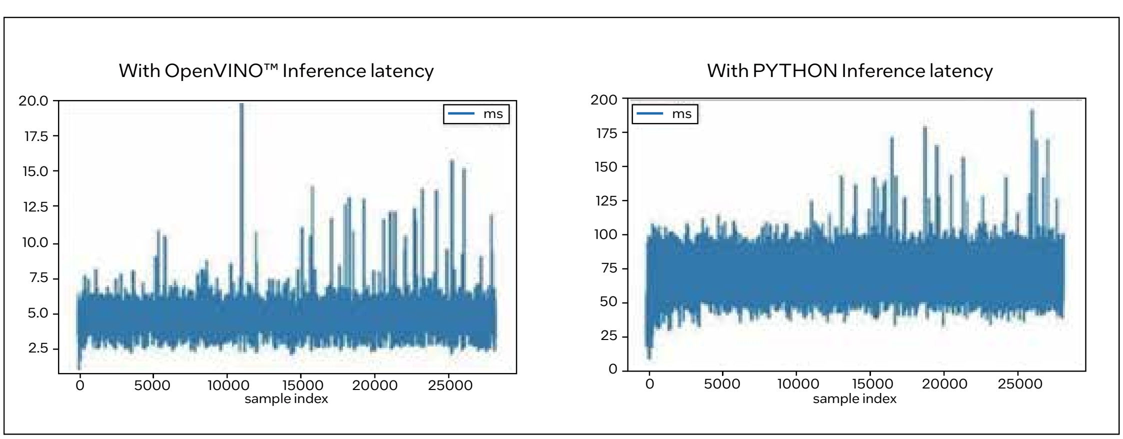


Figure 2.4.3 Inferencing latency improvement with OpenVINO™

In this test, the Phase I CM xApp has been integrated in our implementation while the Phase II CM xApp is still under development. The objective of the Phase II CM xApp is to develop an xApp that provides intelligent connection management in the near commercial wireless environment. To achieve this, the CM xApp needs to collect more RAN measurements. In the meantime, its core algorithm will be enhanced. As a result, the CM xApp will be able to demonstrate its capability to improve the users' QoS with intelligent Connection Management. For the Phase II CM xApp implementation, new service models will be enabled. The Phase II CM xApp will be launched in the second phase of the integration testing.

Edge Computing Platform

Intel® Smart Edge for Builders computing platform is the O-Cloud platform of this CUC trial. Intel® Smart Edge for Builders provides a reference architecture design optimized for typical edge scenarios and considerable modular components. With the support of the reference architecture design, users can quickly deploy edge computing platforms for specific scenarios.

Compared with cloud platforms, edge platforms need to deliver better network performance and more autonomy with constrained resources. They also need to have higher hardware affinity, and to be able to deal with more threat vectors. To address these challenges, Intel® Smart Edge for Builders has made the following enhancements or expansions on the basis of Kubernetes:

1. It provides different experience kits for different deployment scenarios, such as Developer Experience Kit (DEK), Secure Access Service Edge Experience Kit (SASE EK), and Private Wireless Experience Kit (PWEK). Intel® Edge for Builders enables one-click installation of the operating system, Kubernetes, and other basic function add-ins so that users can quickly deploy it, and also allows to select and deploy right modular components for a particular scenario. The PWEK is optimized for local (on-premise) private 5G networks and provides design reference at the edge — an example of European Telecommunications Standards Institute (ETSI) MEC implementations.
2. It works well with the capabilities of Intel® CPUs, including CPU virtualization, CPU pinning, CPU enhanced instruction set, and frequency scaling, and can provide better support for container-based business deployment. For example, CPU pinning enables high bandwidth and low latency of 5G networks; the enhanced instruction set meets the requirements for video processing and AI; and frequency scaling improves the performance or lowers the power consumption.
3. It is compatible with Intel® Network Interface Cards (NICs) and various accelerator cards, including GPU, VPU, Intel® QuickAssist Technology (Intel® QAT) cards, and FEC accelerator cards. Through operators providing various accelerator cards, hardware resources can be automatically discovered, managed, and allocated in edge scenarios.
4. It satisfies the performance requirements of different CNI models of Kubernetes. You can select the optimal CNI model for different service types of edge computing to meet the requirements such as high-performance forwarding between containers, multi-tenant isolation, and support business chains, virtual private clouds (VPCs), and fixed IP.
5. It works well with EdgeX and OpenVINO™ (open-source framework for AI). OpenVINO™ is a pipeline toolkit, compatible with models trained by various open-source frameworks. It offers capabilities for releasing and deploying algorithm models, which enables users to quickly deploy pre-trained models on Intel® CPUs. For AI workload, OpenVINO™ provides the Deep Learning Deployment Toolkit (DLDT) for users to deploy models trained by various open source frameworks. In addition, it provides the image processing toolkit OpenCV and video processing toolkit Media SDK for image and video decoding, preprocessing, inference result post-processing.

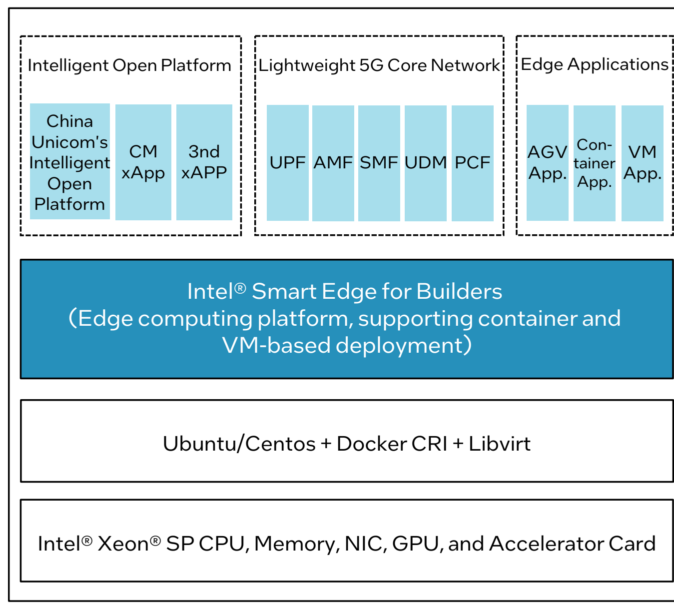


Figure 2.5.1 The Cloud Native platform of Intel® Smart Edge for Builders

In this test, the PWEK is used as the platform, which integrates the CM xApp, China Unicom's Intelligent Open Platform, and H3C's 5G RAN and lightweight 5G core network.

For the integration, the Intel Smart Edge team made some custom enhancements on the Intelligent Open Platform and provided the feedback to the ONF open source community. The community has accepted the feedback. In addition, the Smart Edge team used Intel® OpenVINO™ to optimize the CM xApp based on visual inference and neural network optimization, significantly reducing the inferencing latency of the CM xApp.

The Smart Edge team also designed the test-bed (see Page 11), demonstrating performance benefits of the CM xApp in connection management. From lab to the real world wireless environments, much of algorithm analysis and optimization have also been carried out.

E2 Node

Functions of the E2 Node

The E2 node is a logical node that terminates the E2 interface. It establishes a dedicated E2 connection to the Near-RT Intelligent Open Platform via the E2 interface as the unique identifier that differentiates it from other E2 nodes. One Near-RT Intelligent Open Platform can connect to multiple E2 nodes, while one E2 node can connect to only one Near-RT Intelligent Open Platform. For the new radio (NR) access network, the E2 node can take any form of the following: O-CU-CP, O-CU-UP, O-DU, and their combinations.

The E2 node exposes the RAN functions via E2 interfaces, for example, the X2AP, F1AP, E1AP, S1AP, and NGAP interfaces, as well as the internal functions of RAN processing UEs, cells etc. The E2 node supports all the protocol layers and interfaces defined in the 3GPP radio access network, including eNB for E-UTRAN and gNB/ng-eNB for NG-RAN.

The E2 node provides the following RIC services for the Near-RT Intelligent Open Platform:

- **REPORT service**: The Near-RT Intelligent Open Platform uses an RIC Subscription procedure to request that the E2 node sends a REPORT message to the platform and the associated procedure continues in the E2 node after each occurrence of a defined RIC Subscription procedure event trigger.
- **INSERT service**: The Near-RT Intelligent Open Platform uses an RIC Subscription procedure to request that the E2 node sends an INSERT message to the platform and suspends the associated procedure in the E2 node after each occurrence of a defined RIC Subscription procedure event trigger.
- **CONTROL service**: The Near-RT Intelligent Open Platform sends a CONTROL message to the E2 node to initiate a new associated procedure or resume a previously suspended associated procedure in the E2 node.
- **POLICY service**: The Near-RT Intelligent Open Platform uses an RIC Subscription procedure to request that the E2 node executes a specific policy during the functioning of the E2 node after each occurrence of a defined RIC Subscription procedure event trigger.

5G Base Station Practices

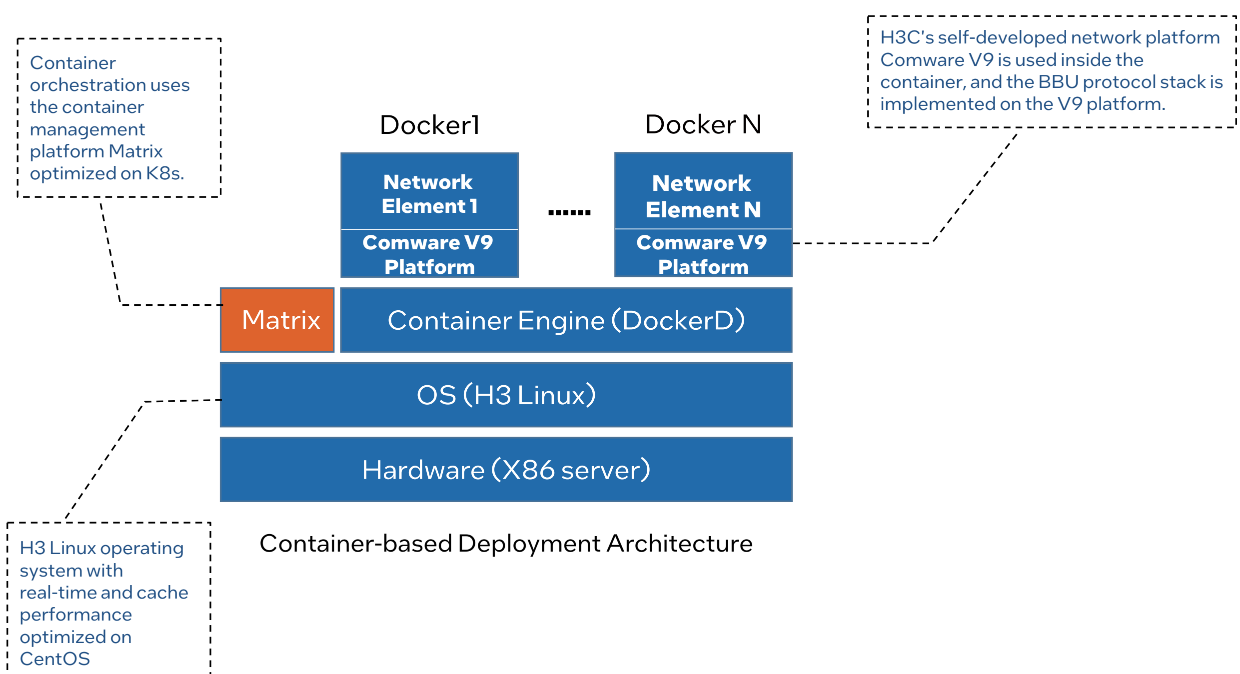
In this practical implementation, the H3C Unicell5100 5G Indoor Wireless Distribution System is used as the E2 node and connects to the Near-RT Intelligent Open Platform via the E2 interface.

The H3C Unicell5100 5G Indoor Wireless Distribution System consists of the baseband unit (BBU), fronthaul switch (FSW), and pico-remote radio unit (pRRU). Two-level cascading of FSWs is allowed , and each FSW can connect to a maximum of eight pRRUs. Optical fibers are used to connect FSWs and BBUs as well as two cascading FSWs while optical and electrical hybrid cables are used to connect FSWs and pRRUs.

The core technology of BBUs is 5G NR. They use customized OTII servers and support container and cloud-based deployment. They can be co-deployed with a user plane function (UPF) and MEC.

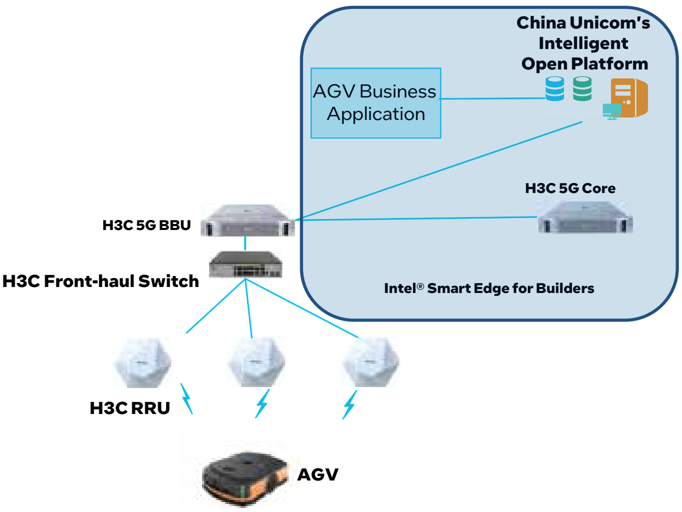
The CU and DU software of BBU is logically separated. For hardware, CUs and DUs can share network devices in the deployment, and CUs can be deployed on the cloud as well. In the future, CUs and DUs can be easily separated through software upgrade with the network evolution.

BBU container-based deployment:



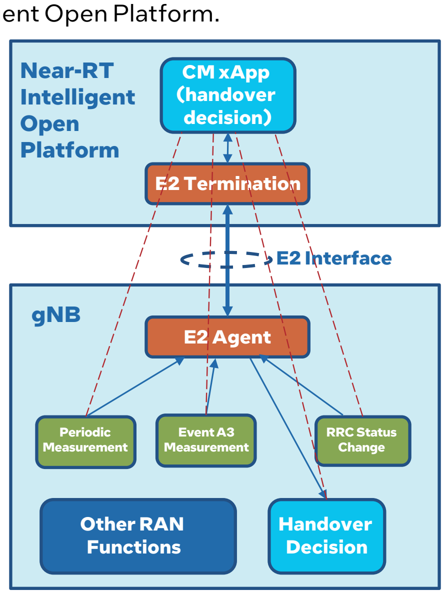
Container-based Deployment Architecture

In this practical implementation, the H3C Unicell5100 5G Indoor Wireless Distribution System is used as the E2 node, and allows the Intelligent Open Platform to control its handover decision. The network deployment is shown below:



The H3C 5G core network and China Unicom's Intelligent Open Platform can be either co-deployed or separately deployed on the Intel® Smart Edge for Builders platform. The BBU has two backhaul interfaces. One connects to the core network, and the other connects to the Intelligent Open Platform. Resources of the two links are independent and do not affect each other.

The E2 interface of the H3C Unicell5100 5G Indoor Wireless Distribution System (base station) complies with the E2AP v2.0 standard. CUs and DUs are not separated on the picocell. They are managed by one E2 Agent. With the E2AP procedure, the picocell exposes the RAN functions to the E2 interface for REPORT and CONTROL services. The E2AP signaling is coded and decoded according to ASN.1 to ensure efficient and safe transmission. E2AP signaling interaction is based on reliable transmission at the SCTP layer. It provides congestion and traffic control, error control, data discarding and replication, and support for selective retransmission mechanism functions for E2AP packets through SCTP. The figure below shows the functionality architecture of the base station and the Near-RT Intelligent Open Platform.



The base station supports the following function modules of the E2 node:

1) E2 Agent

The base station uses the E2 Agent to terminate the E2 interface, and send and receive E2 messages.

2) RAN functions controlled by the Near-RT Intelligent Open Platform

With the REPORT service, the base station reports 3 types of information — periodic measurement, RRC status change, and Event A3 measurement — to the Near-RT Intelligent Open Platform.

With the CONTROL service, the base station allows the Near-RT Intelligent Open Platform to control the handover decision.

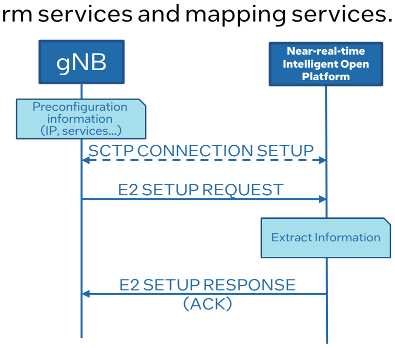
3) RAN functions unrelated to the Near-RT Intelligent Open Platform

The base station provides the RIC service for the Near-RT Intelligent Open Platform through E2AP procedures. In this practical implementation, the following E2AP procedures are supported:

1. E2 Setup procedure

The E2 Setup procedure is not directly relevant to dedicated applications. It is initialized by the base station. It sets up the basic interfaces and sends dedicated configuration of the base station to the Near-RT Intelligent Open Platform. Thus, the signaling link is established between the base station and the Near-RT Intelligent Open Platform.

The base station provides its functions for the Near-RT Intelligent Open Platform to support Near-RT Intelligent Open Platform services and mapping services.



2. RIC Subscription procedure

The RIC Subscription procedure is relevant to the dedicated messages of the application. These messages are transferred between the application (xApp) of the Near-RT Intelligent Open Platform and the target function of the E2 node. The RIC Subscription procedure is used to set up E2 subscription on the E2 node. It consists of event triggers and a set of actions.

In this practical implementation, the Near-RT Intelligent Open Platform requests the REPORT service from the base station through the RIC Subscription procedure. To be specific, there are 3 RIC Subscription procedures to be initiated, requesting 3 types of information: RRC status change, periodic measurement, and Event A3 measurement.

3. RIC Indication procedure

In the RIC Indication procedure, messages related to the

REPORT and/or INSERT services are transmitted to the Near-RT Intelligent Open Platform, which has initiated RIC Subscription to the E2 node at an earlier time, and the E2 node has also detected the event trigger.

In this practical implementation, the base station provides the REPORT service for the Near-RT Intelligent Open Platform with the RIC Indication procedure. The RIC Indication procedures corresponding to the three RIC Subscription procedures report the following information:

1) RRC status change

When the base station detects an RRC status change, for example, the UE status changes from connected to idle, from connected to inactive, or from idle to connected, the RIC Indication procedure is triggered to report the RRC status change to the Near-RT Intelligent Open Platform.

When receiving A3 measurements, the Near-RT Intelligent Open Platform will send the handover control decision to the base station via the RIC CONTROL request messages.

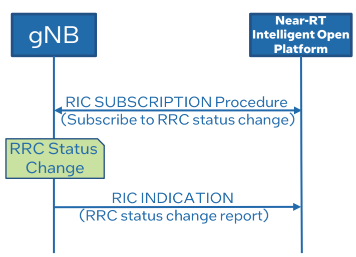
For example, in a successful handover within a base station, the original handover process is as follows:

a) The UE reports A3 measurements;

b) The base station determines the target cell;

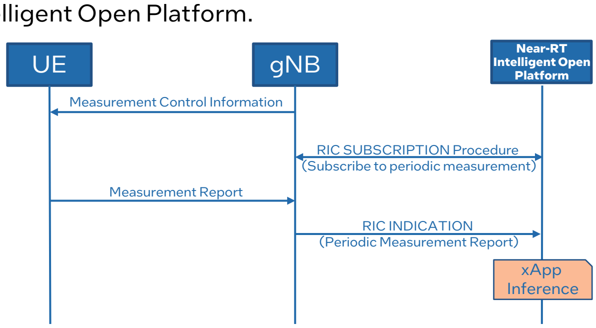
c) The base station sends a handover reconfiguration message to the UE;

d) The UE completes reconfiguration for the handover and sends a reconfiguration completion message to the base station.



2) Periodic RSRP measurement

When the base station receives periodic RSRP measurements of the UE, the RIC Indication procedure is triggered to report periodic measurements to the Near-RT Intelligent Open Platform.



3) Event A3 measurement (report type: RSRP)

When the base station receives Event A3 measurements from the UE (report type: RSRP),

a) the RIC Indication procedure is triggered to report A3 measurements to the Near-RT Intelligent Open Platform.

b) Then, the timer is started to wait for the RIC CONTROL Request message to indicate the handover target.

4. RIC Control procedure

The RIC Control procedure initializes or resumes specific functions on the E2 node.

The RIC CONTROL service can be triggered by the RIC INDICATION messages from the E2 node, or triggered inside the Near-RT Intelligent Open Platform.

In this practical implementation, the RIC CONTROL service is triggered inside the Near-RT Intelligent Open Platform. When receiving A3 measurements, the Near-RT Intelligent Open Platform will send the handover control policy to the base station via the RIC CONTROL request messages.

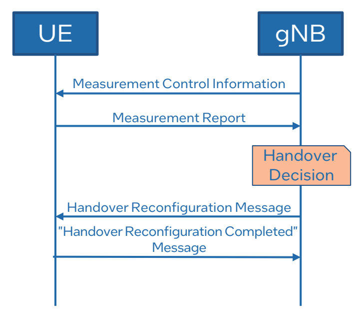
For example, in a successful intra-gNB handover, the original handover process is as follows:

1) The UE reports A3 measurements;

2) The base station determines the target cell;

3) The base station sends a handover reconfiguration message to the UE;

4) The UE completes reconfiguration for the handover and sends a reconfiguration completion message to the base station.



With the Near-RT Intelligent Open Platform used, the handover decision is no longer determined by the base station but is controlled by the platform. Meanwhile, the base station retains the original handover decision in case the Near-RT Intelligent Open Platform fails to control the decision. In this circumstance, the handover will be still determined by the base station.

The following is the process of a successful intra-gNB handover with the Near-RT Intelligent Open Platform:

1) The UE reports A3 measurements;

2) If the base station receives A3 measurements, and satisfy the condition of event triggers for the Subscription procedure, then

a) Send the RIC INDICATION message to the Near-RT Intelligent Open Platform and reports A3 measurements;

b) Start the timer;

3) If the Near-RT Intelligent Open Platform receives A3 measurements, then

a) Trigger the CM xApp on it to infer and determine the target cell of the handover;

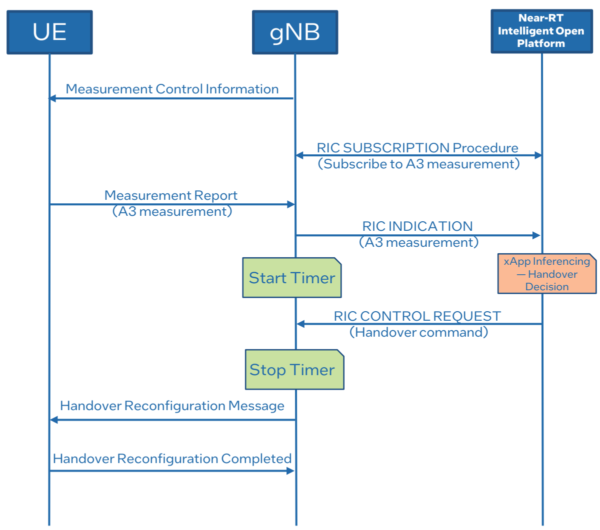
b) Send the RIC CONTROL Request message to the base station and indicate the target cell of the handover;

4) If the base station receives the RIC CONTROL Request message before the timer expires, then

a) The timer stops;

b) Send a handover reconfiguration message to the UE as the RIC CONTROL message indicates.

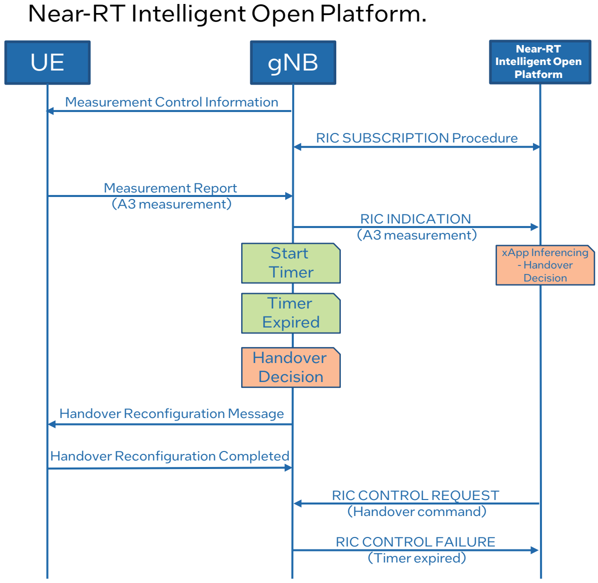
5) The UE receives the handover reconfiguration message and completes the handover.



If the base station timer expires, the RIC Control procedure is deemed as a failure. When this occurs,

a) The base station indicates the UE to perform handover according to the original decision;

b) Send the RIC CONTROL FAILURE message to the Near-RT Intelligent Open Platform.



The base station exposes measurement and status information to the E2 interface through the E2AP procedure. The CM xApp on the Near-RT Intelligent Open Platform can not only detect the signal strength of different cells but also perceive the base station load and business-level performance indicators. This can help the base station select the optimal handover target at the system level and balance the multiple-cell load for better handover success rate and user experience.

Test-bed Design

Design of Test-bed Deployment

1. Test-bed Deployment

As shown in Figure 2.7.1.1, the CM xApp, Near-RT Intelligent Open Platform, E2 node, and 5G core network are all containerized in the deployment. The E2 node and the Near-RT Intelligent Open Platform are connected via the E2 interface, which complies with the E2AP v2.00 standard. The CU and DU are integrated on the E2 node, and the three RRUs serve three cells.

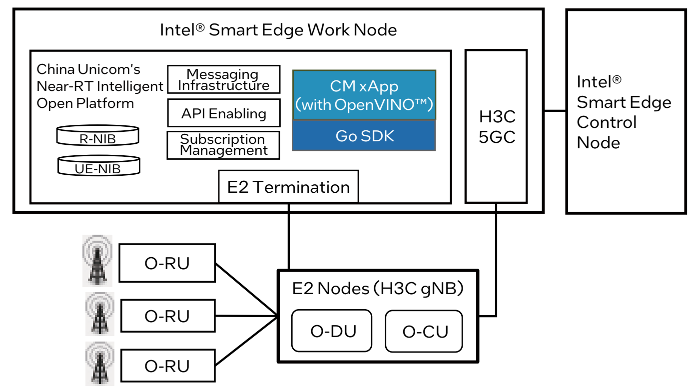


Figure 2.7.1.1 Test platform deployment

2. Test Design

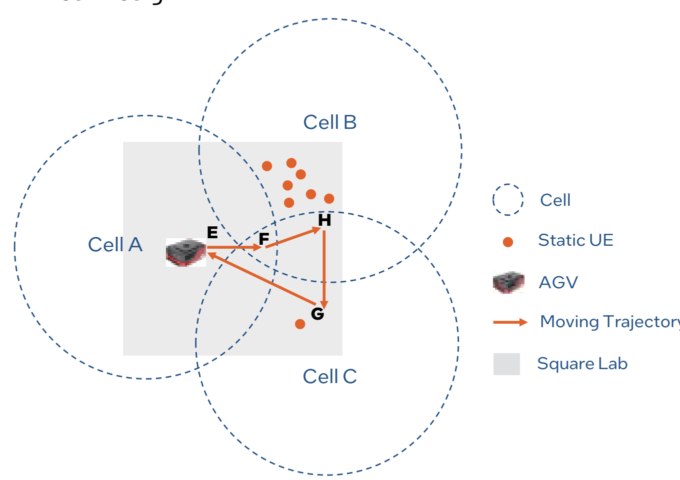


Figure 2.7.1.2 Wireless network deployment

In the same wireless environment, to get reliable performance comparison, the throughput and video playback effect of conventional wireless network schemes with no Intelligent Open Platform and xAPP are compared with those of schemes with a Near-RT Intelligent Open Platform and CM xApp.

As shown in Figure 2.7.1.2, the lab covers a 25 m x 25 m square area, and the RRUs of cells A, B, and C are deployed within the area. The instantaneous bandwidth (IBW) of cells A, B, and C is 100 MHz, and their operation bandwidth (OBW) is 300 MHz (3.3–3.6 GHz). The RRUs of the three cells are at the three vertices of the equilateral triangle. The RRU of Cell B is at the top right vertex of the square, and the RRU of Cell C is at the lower right vertex of the square. The RRUs are 25 meters away from one another. A mobile UE (AGV) and a number of static UEs are deployed in the testing environment.

It is assumed that cells A, B, and C have the same transmit power and transmit mode, and that the base station uses the enhanced proportional fair (EPF) scheduling algorithm. During the test, each UE keeps the service traffic to its full. The number of UEs in Cell A ≤ the number of UEs in Cell C < the number of UEs in Cell B. Cell B should be set close to its full load, but not full yet. Cell B is not in full load, because the AGV must hand over to Cell B when there is no CM xApp and won't be rejected by Cell B; Cell B is close to its full load (above 95% and below 100%), because the resources allocated to the AGV are far less than what it needs when it switches to Cell B, so that the AGV delivers unsatisfactory performance as expected.

The AGV travels at a speed of 1.6 m/s.

The moving trajectory of the AGV is E > F > H > G > E.

Check the result of AGV handover decision when AGV trolley travel from Point F to Point H could show the performance of AI solution on load balance.

3. Testing procedures

Perform the following steps 20 times for the two deployment schemes (with and without the Intelligent Open Platform/CM xApp) respectively:

1) Keep sending UDP packets to all UEs (regardless of static UEs or moving UEs — AGVs).
2) Make the AGV stop at Point E for two minutes.
3) Move the AGV from Point E to Point F and then to Point H.
4) Make the AGV stop at Point H for three minutes.
5) Move the AGV from Point H to Point G.
6) Move the AGV from Point G to Point E.
7) Record the information about which cell the AGV is attached to at point G.

4. Expected Test Results

We expect the test results can prove the CM xApp's advantages over the conventional wireless communication scheme in improving the performance for edge users and in load balancing.

Under the same wireless network environment, the AGV travels along the same trajectory, i.e., E > F > H > G > E. At Point F, Cell B and Cell C start to meet the condition for triggering Event A3. The measurements here tend to be unstable. Point F is in effect an auxiliary point aiming to help the AGV find Point H. Therefore, measurements at this point are not used for comparison. As the AGV reaches Point H, the measurement becomes stable. Thus Point H is used as the observation point where the result of handover from Point F to Point H is measured. The two schemes are expected to deliver different results:

Position	Point E	Point H	Point G	Point E
Conventional Wireless Communication Scheme	AGV is attached to Cell A.	AGV has a bigger chance to switch to Cell B.	AGV is attached to Cell C.	AGV is attached to Cell A.
Scheme with Intelligent Open Platform/CM xApp	AGV is attached to Cell A.	AGV has a bigger chance to switch to Cell C.	AGV is attached to Cell C.	AGV is attached to Cell A.

As shown in the table above, the handover result is the same at all points except at Point H. Therefore, we will highlight the difference at Point H:

Position	Difference
Conventional Wireless Communication Scheme	At Point H, the AGV has a bigger chance to switch to Cell B.
Scheme with Intelligent Open Platform/CM xApp	At Point H, the AGV has a bigger chance to switch to Cell C.

When the AGV trolley moves from cell A to cell B and Cell C, both Cell B and Cell C could fulfill the condition of event A3 (Event A3 is triggered when a neighbor cell measurement becomes better than a special cell by an offset). In our testbed design, the RSRP of Cell B is better than that of Cell C when UE moves from Cell A to the edge of Cell B/Cell C according to the AGV trolley moving route design, then the AGV trolley will handover to Cell B with a big possibility since the RSRP of Cell B is bigger than that of Cell C according to base station's handover decision (Conventional Wireless Communication Scheme). However, with our Intelligent Open Platform/CM xApp, the AGV trolley will be handed over to Cell C with a big possibility.

Interaction Process

Figure 2.7.2.1 illustrates the main message interaction process among E2 node, UE and Near-RT Intelligent Open Platform/CM xApp scheme. The message flow between the UE and base station follows the 3GPP 38.331, and the E2 message procedure follows the E2AP V2.00, E2SM V02.00, and E2SM-MHO service model. The E2 message procedures are mainly as follows: E2 Setup procedure, RIC Subscription procedure, RIC Indication procedure, and RIC Control procedure. The E2 Setup procedure registers the E2 node on the Intelligent Open Platform and establishes connection with the platform. The RIC Subscription procedure helps xApp to subscribe the E2 functions of the E2 node through the Intelligent Open Platform. The RIC Indication procedure allows the E2 node to report the RAN measurements subscribed by the xApp to the Intelligent Open Platform. And the RIC Control procedure happens when the inferencing result of the xApp is sent back to the E2 node through the Intelligent Open Platform, and the node completes the handover according to the content in RIC Control Request message. If any error happens, such as "Control timer expired", "Control failed to execute", the E2 node will send RIC Control Failure message to the Intelligent Open Platform.

The E2SM between the CM xApp and the E2 node follows the E2SM-MHO service model. This service model designed for connection management is developed by ONF, Intel, and other industrial participants.

The E2SM-MHO service model includes reports of the following events: 1) Reporting periodic measurement; 2) Reporting Event A3 measurement; 3) Reporting RRC status change of UEs.

When the E2 node enables the MHO service model and accepts the preceding subscription, it passes the handover control to the CM xApp and takes the xApp's inferencing results as the handover decision. The base station would report the periodic measurement results in the RIC Indication messages at an interval specified in the RIC Subscription

Request message. Base station would report a RIC Indication message per UE, which includes the measurement results of its serving cell and neighboring cells. When detecting a RRC status change of the UE, the base station would report the UE ID and the new status to the Intelligent Open Platform/ CM xApp in a RIC Indication message. When notified of the happening of Event A3 by UEs, the base station would report the event to the Intelligent Open Platform in a RIC Indication message for the CM xApp to trigger the inferencing process.

The inferencing result would then be sent to the E2 node in a RIC Control Request message, based on which the node executes the handover. Upon reporting the RIC Indication that contains the Event A3 information to the Intelligent Open Platform, the E2 node would immediately start a timer to avoid UE call drop. In the case that the Intelligent Open Platform sends a RIC Control Request later than the time when the timer expires, the E2 node would use its own algorithm to make the handover decision and send it to the UE.

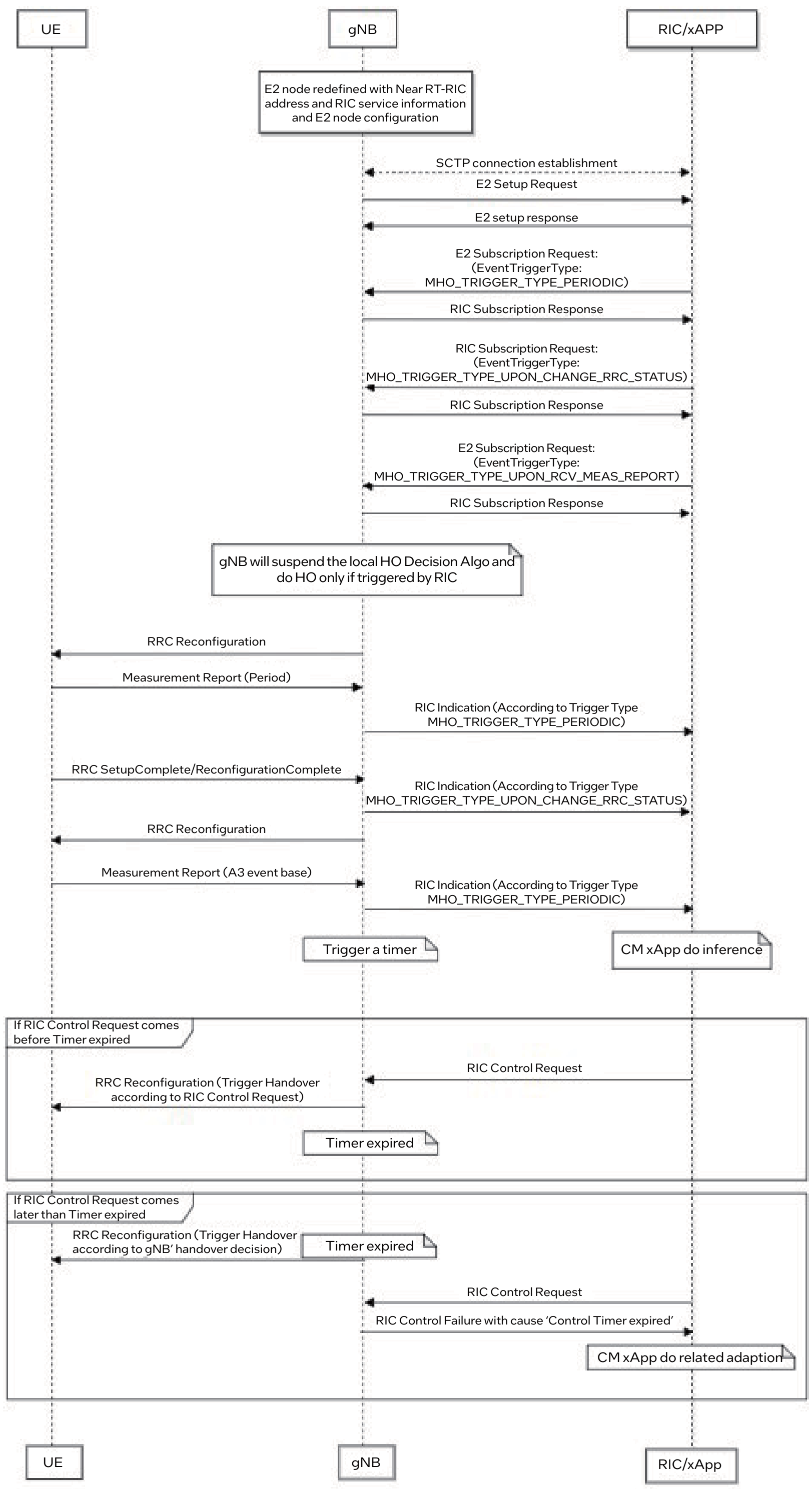


Figure 2.7.2.1 Signaling process

Test Results

1. Actual test results with three cells

According to the test result expectation described in the section above, it makes no difference which cell the AGV is attached to at points E, F, and G under the two test schemes. What matters is to which cell the AGV is attached at Point H. We can examine the performance of the CM xApp by observing how many times that the AGV handed over to Cell B and Cell C at Point H out of 20 handovers under the two schemes respectively.

The following table shows the test results (Here only the inferences, which are triggered by the A3 events that both Cell B and Cell C are reported as candidate cells, are counted):

Scheme	Which Cell the AGV Switched to at Point H
Conventional Wireless Communication Scheme	20 times to Cell B, 0 times to Cell C
Scheme with Intelligent Open Platform/CM xApp	0 times to Cell B, 20 times to Cell C

2. Simulation results with seven cells

Subject to the test environment, the test-bed design only illustrates part of the CM xApp performance in Phase I of the test. To better represent the real-world scenario in the existing network, we conducted another simulation test using the RAN Simulator to create a more complex deployment scenario. The simulation results indicate that the CM xApp delivers better overall performance for the system.

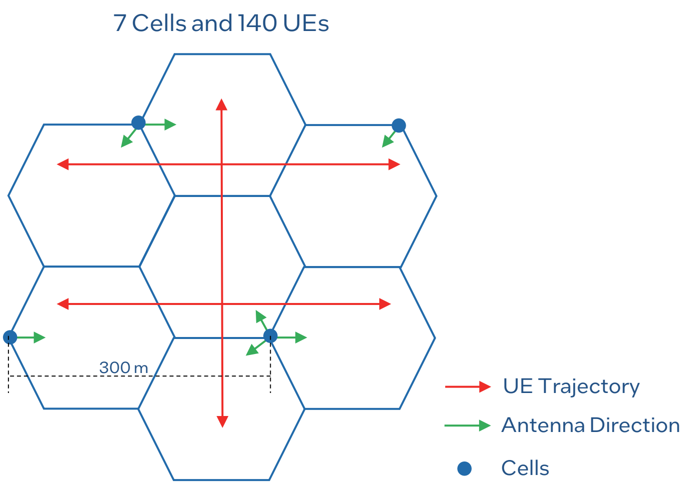


Figure 2.8.1 Network deployment (in simulator) with seven cells and 140 UEs

With the deployment shown in Figure 2.8.1, we obtained the following results (see chapter of Connection Management xApp for detailed explanation of each optimization policy):

Optimization Policy	Performance Gain of CM xApp Algorithm over Conventional Max. RSRP Algorithm
Network Throughput	5.7%
Load Balancing	23%
Cell Edge Coverage	231 %

Summary and Looking Ahead

1. The Intelligent Open Platform and Wireless Applications

As is known to all, wireless environment can be complex and time-varying, making its features hard to capture with an enclosed approach. The Intelligent Open Platform and CM xApp provide an approach based on AI training and inferencing, which is inherently ideal for the wireless network topology and efficient in capturing its changing features. This approach also makes it possible to find the optimal objective function under different optimization policies, which in turn helps address specific connection management issues.

The Phase I test demonstrated the likely performance gain with AI training and inferencing based on the Intelligent Open Platform and CM xApp using a limited amount of measurements (e.g., the RSRP). To better illustrate the results, we specifically designed the wireless deployment scenario. The objective of the Phase II test is to demonstrate the performance of intelligent connection management in near commercial wireless environments. Deployment scenarios will be set closer to commercial wireless environments and more complex test use cases will be used to elaborate the enhanced performance, diverse application scenarios, and extensive optimization policy support of the CM xApp and Intelligent Open Platform.

By having Intel® CM xApp for intelligent handover using ML/AI techniques, and OpenVINO™ optimization for AI inference to significantly reduce latency, the collaboration with Intel helped leveraging Intel® Smart Edge for Builders Cloud Native capabilities to on-board near-RT RIC on an Edge Platform while abstracting the underlying HW complexity.

2. The Value of the Intelligent Open Platform for Base Stations

The construction of cloud-network integrated systems and development of 5G wireless communication technologies are bringing new opportunities for the increasingly specialized verticals. In the meantime, it also raises the bar for flexibility and openness of the radio access network of base stations. Under the conventional wireless communication system architecture, capabilities integrated within base stations are relatively closed. Exposing the capabilities of the radio access network will further unleash the potential of base stations to offer better customer experience and industrial applications, such as business KPI enhancement, SLA satisfaction improvement, boosting intelligent operations and maintenance, and industrial application optimization.

H3C Unicell 5G Indoor Wireless Distribution System could support the deployment of CUs with UPF and MEC on the cloud. By introducing the Intelligent Open Platform, base stations can provide data at higher granularity for third-party software, support applications to be aware of services, and use AI for near-real-time control and optimization of base station functions and resources through big data analytics. In this way, base stations could expose the capabilities of the radio access network to the Intelligent Open Platform to promote white-box base stations, help operators build more intelligent, flexible, and open wireless network, empower various scenarios and help strengthen the competencies of different verticals.

3. Opportunities and Challenges Faced by Operators

For operators, open and intelligent wireless cloud network and its development will bring both opportunities and challenges. In the meantime, operators need to build a dynamic ecosystem and work with different vendors to share strengths. All these have posed higher requirements for operators to develop their own competencies. Moreover, open-source, intelligent, and cloud technologies have provided valuable technical platforms for them. While strengthening their own competencies, operators also need to navigate through these emerging opportunities, and constantly invest in the new business domains.

The idea and development of the Intelligent Open Platform reflect China Unicom's deep insight and practices in this aspect. Exposure of base station's capabilities, interface openness, and AI applications give operators opportunities to work with partners to build a more open and intelligent wireless cloud network. In future, we'll continue taking well-arranged steps to further the openness and intelligence of the wireless cloud network, as is scheduled in the research objectives of and plan for China Unicom's CUBE-RAN. Before the end of this paper, China Unicom would like to sincerely invite all operators, device vendors, system integrators, test instrument manufacturers, and industry customers to join the research and testing of the Intelligent Open Platform, and work together with us to create a more extensive and dynamic new ecosystem for the wireless cloud network.