Beyond 5G Multi-Tenant Private Networks Integrating Cellular, Wi-Fi, and LiFi, Powered by Artificial Intelligence and Intent Based Policy

# 5G-CLARITY Deliverable D3.3

# Complete Design and Final Evaluation of the Coexistence, Multi-Connectivity, Resource Management, and Positioning Frameworks

| | |
|---|---|
| **Contractual Date of Delivery:** | **April 30, 2022** |
| **Actual Date of Delivery:** | **June 6, 2022** |
| **Editor(s):** | **Hamada Alshaer (USTRATH)** |
| **Author(s):** | **Hamada Alshaer, Anil Yesilkaya, Ardimas Purwita (USTRATH), Daniel Camps-Mur, Ferran Cañellas, Miguel Catalan, August Betzler (i2CAT), Antonio Garcia (ACC), Jesús Gutiérrez, Vladica Sark (IHP), Jorge Navarro-Ortiz, Juan Jose Ramos-Munoz (UGR), Srinivasan Raju, Rui Bian (PLF), Tezcan Cogalan (IDCC), Carlos Colman Meixner, Amin Emami, Hilary Frank (UNIVBRIS), Mir Ghoraishi (GIGASYS)** |
| **Work Package:** | **WP3** |
| **Target Dissemination Level:** | **Public** |

This document has been produced during 5G-CLARITY Project. The research leading to these results received funding from the European Commission H2020 Programme under grant agreement No. H2020-871428. All information in this document is provided “as is", there is no guarantee that the information is fit for any particular purpose. The user thereof uses the information at its own risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors’ views.

# Revision History

| Revision | Date | Editor /Commentator | Description of Edits |
|---|---|---|---|
| 0.1 | 24.11.2021 | Anil Yesilkaya (USTRATH), Mir Ghoraishi (GIGASYS) | Created master documents and Sections |
| | 15.12.2021 | Mir Ghoraishi (GIGASYS)/ Antonio Garcia (ACC), Jorge Navarro-Ortiz, Juan Jose Ramos-Munoz (UGR), Daniel (I2CAT), Jesús Gutiérrez, Vladica Sark (IHP), Rui Bian (PLF), Anil Yesilkaya, Hamada Alshaer (USTRATH), Tezcan Cogalan (IDCC) | Added KPI table sheet/ ToC by partners |
| | 09.02.2022 | Daniel Camps-Mur, Ferran Cañellas, Miguel Catalan, August Betzler (I2CAT) | First draft of section 4.2 available |
| 0.2 | 23.02.2022 | Antonio Garcia (ACC), August Betzler (i2CAT) | First draft of section 2 available |
| 0.3 | 30.03.2022 | Jorge Navarro-Ortiz, Juan Jose Ramos-Munoz (UGR), Ardimas Purwita (USTRATH), Daniel Camps-Mur(i2CAT) | First draft of subsections 3<br>UGR: subsections 3.1, 3.1.2, 3.2.3, 3.2.4, 3.3.2<br>USTRATH: subsection 3.1.1<br>I2CAT: Subsection 3.2.2 |
| 0.31 | 06.04.2022 | Anil Yesilkaya (USTRATH) | Merged individual section files into a single main file |
| 0.4 | 13.04.2022 | Hamada Alshaer (USTRATH), Rui Bian (PLF) | First draft of subsections available<br>USTRATH: Subsection 4.1<br>PLF: Subsection 5.3 |
| 0.5 | 20.04.2022 | Daniel Camps-Mur(i2CAT) | Sections 2, subsections 3.1, 3.1.1, 3.1.2, 3.2.2, 3.2.3, 3.2.4, 3.3.2, 5.3 external review |
| 0.51 | 20.04.2022 | Hamada Alshaer (USTRATH) | Reviewed section 4 external |
| 0.6 | 28.04.2022 | Vladica Sark, Jesús Gutiérrez (IHP), Anil Yesilkaya (USTRATH) | First draft sections available<br>IHP: Subsections 5.1, 5.2<br>USTRATH: section 6 |
| 0.61 | 28.04.2022 | Anil Yesilkaya (USTRATH) | Added the executive summary and Section 1 |
| 0.62 | 0.4.05.2022 | Daniel Camps-Mur(i2CAT) | Reviewed sections 5 and 6 |
| 0.7 | 11.05.2022 | Tezcan Cogalan (IDCC), Carlos Colman Meixner, Amin Emami, Hilary Frank (UNIBRIS) | First draft subsections available<br>IDCC : Subsection 3.3.1<br>UNIBRIS : Subsections 3.2.1, 3.3.3 |
| 0.8 | 18.05.2022 | Daniel Camps Mur (I2CAT) | Sections 1-6 external review |
| 0.81 | 19.05.2022 | Hamada Alshaer (USTRATH) | Revised the executive summary, section 1, introduction<br>Added conclusion |
| 0.9 | 20.05.2022 | Hamada Alshaer (USTRATH) | Full review and proof reading |
| 1.0 | 28.05.2022 | Mir Ghoraishi (GIGASYS) | Full review and proof reading to final version |

# Table of Contents

List of Acronyms ..... 9
Executive Summary ..... 11
1 Introduction ..... 13
1.1 Scope of this document ..... 13
1.2 Objectives of this document ..... 14
1.3 Document structure ..... 15
1.4 Fulfilment of 5G-CLARITY functional requirements and KPIs ..... 15
2 Spectrum Sharing Framework with CBRS ..... 18
2.1 Introduction ..... 18
2.2 Baseline legacy CBRS architecture and newer specifications ..... 18
2.3 Industry status in CBRS alliance, O-RAN alliance and 3GPP ..... 22
2.4 5G-CLARITY coexistence framework ..... 22
2.4.1 5G-CLARITY co-existence framework high level architecture ..... 22
2.4.2 Testbed description ..... 23
2.4.3 Dependencies on disaggregated third-party DU/RU components ..... 25
2.4.4 Solution design ..... 26
2.4.5 Test results against WINNFORUM test harness ..... 27
2.4.6 Example testcase (WINNF.FT.C.HBT.1) ..... 29
2.4.7 Section summary ..... 29
3 Advanced Multi-Connectivity and Multi-WAT Aggregation ..... 30
3.1 5G-CLARITY final architectural design ..... 30
3.1.1 5G-CLARITY multi-connectivity eAT3S testbed REST API ..... 30
3.1.1.1 System model ..... 30
3.1.1.2 Examples of endpoints ..... 32
3.1.2 Validation of the 5G-CLARITY multi-connectivity eAT3S REST-API ..... 32
3.2 Benchmarking 5G-CLARITY multi-connectivity framework ..... 34
3.2.1 Multi-WAT throughput benchmarking ..... 34
3.2.1.1 Multi-connectivity testbed setup ..... 35
3.2.2 Latency benchmarking ..... 41
3.2.2.1 Experimental methodology ..... 41
3.2.2.2 Results analysis ..... 43
3.2.2.3 Discussion on 5G-CLARITY O3.2 ..... 45
3.2.3 eAT3S mobility benchmarking ..... 45
3.2.3.1 Experimental setup ..... 46
3.2.3.2 Experimental results ..... 46
3.2.4 Benchmarking E2E TCP connection over 5G-CLARITY multi-connectivity framework ..... 47
3.3 5G-CLARITY eAT3S framework ..... 50
3.3.1 5G-CLARITY utility-based eAT3S algorithm ..... 51
3.3.2 5G-CLARITY eAT3S bandwidth prediction algorithm validation ..... 55

3.3.2.1 Estimation of the available throughput ..... 55
3.3.2.2 MPTCP weights estimation ..... 57
3.3.3 5G-CLARITY low-level eAT3S user plane function ..... 60
3.4 Section summary ..... 65
4 Advanced Resource Management ..... 66
4.1 Autonomic LiFi attocellular network slicing ..... 66
4.1.1 Background and motivation ..... 66
4.1.2 System model ..... 67
4.1.3 Problem formulation ..... 68
4.1.4 MDP formulation at SDN controller ..... 69
4.1.5 DRL model design ..... 70
4.1.5.1 State ..... 70
4.1.5.2 Actions ..... 70
4.1.5.3 Reward calculation ..... 70
4.1.6 Performance evaluation ..... 74
4.2 Network-wide slicing of Wi-Fi networks with variable load in space and time ..... 76
4.2.1 Motivation ..... 76
4.2.2 G-ADRR design ..... 76
4.2.2.1 System model ..... 76
4.2.2.2 SLA contract definition ..... 77
4.2.2.3 G-ADRR Cost Function ..... 78
4.2.3 G-ADRR Performance evaluation ..... 81
4.2.3.1 G-ADRR testbed ..... 81
4.2.3.2 G-ADRR experimental validation against a static scheduling policy ..... 82
4.3 Section Summary ..... 85
5 Multi-WAT Based Localization Optimization and Implementations ..... 86
5.1 Sub-6 GHz based indoor localization ..... 86
5.2 mmWave based localization ..... 90
5.3 LiFi based localization ..... 91
5.4 Section summary ..... 94
6 Enhanced System Level Simulations and Resource Management ..... 95
6.1 Enhanced simulation platform functionalities ..... 96
6.1.1 Content and request generation ..... 98
6.1.2 Random mobility and orientation models for the UEs ..... 100
6.1.3 Handover protocol ..... 101
6.1.4 Single point and coordinated multi-point transmission techniques ..... 102
6.2 Obtaining the system performance ..... 102
6.2.1 Link quality model ..... 103
6.2.2 Link performance model ..... 103
6.3 Simulation results ..... 106

6.3.1 SIR distribution results ..... 108
6.3.1.1 Generic deployment with SP transmission ..... 108
6.3.1.2 Dense deployment with SP transmission ..... 110
6.3.1.3 Generic deployment with 3-CoMP transmission ..... 111
6.3.1.4 Dense deployment with 3-CoMP transmission ..... 112
6.3.2 Achievable area traffic capacity results ..... 113
6.4 Section summary ..... 114
7 Summary and Concluding Remarks ..... 116
8 Bibliography ..... 118

# List of Figures

Figure 2-1 CBRS overall architecture ..... 18
Figure 2-2 Admission control system architecture CBRS ..... 19
Figure 2-3 Overall CBRS architecture in LTE system (using Accelleran small cell as example) ..... 20
Figure 2-4 Mandatory NR-TDD TDD UL/DL Configurations for the CBRSA CxG ..... 20
Figure 2-5 Shared/Common TCCSs with multiple SAS/CxMs ..... 22
Figure 2-6 5G-CLARITY spectrum sharing and co-existence architectural concept ..... 23
Figure 2-7 Enhanced 4G/5G CBRS solution ..... 24
Figure 2-8 5G CBRS testbed with Effnet/Phluido DU and Benetel and NI B210 RUs ..... 25
Figure 2-9 5G CBRS testbed in the lab with B210 RU ..... 25
Figure 2-10 5G CBRS DU/RU configuration ..... 26
Figure 2-11 5G CBRS client components ..... 26
Figure 2-12 5G CBRS client typical successful sequence ..... 27
Figure 3-1 Diagram of the REST API for the multi-connectivity testbed ..... 31
Figure 3-2 Virtual testbed with OpenVSwitch. ..... 33
Figure 3-3 Visualization of performance metrics (throughput and latency) in the virtual testbed ..... 34
Figure 3-4 5G-CLARITY multi-WAT framework testbed setup architecture. ..... 35
Figure 3-5 5G-CLARITY multi-WAT framework testbed setup at the Smart-Internet Lab -UoB ..... 36
Figure 3-6 5GNR test with two types of UE setups: (a) Throughput in Mbps, (b) Latency in ms ..... 37
Figure 3-7 Wi-Fi-6 Throughput and latency performance results ..... 38
Figure 3-8 LiFi Performance test results (a) Throughput in Mbps (b) Latency in ms ..... 38
Figure 3-9 5GNR download throughput when MPTCP is in place ..... 39
Figure 3-10 Wi-Fi-6 download throughput while MPTCP is running ..... 39
Figure 3-11 LiFi download throughput while MPTCP is running ..... 40
Figure 3-12 Aggregate download throughput over all radio access technologies while MPTCP is running ..... 40
Figure 3-13 Testbed setup ..... 41
Figure 3-14 Experimental methodology ..... 42
Figure 3-15 RTT delay PDFs of 5GNR plus Wi-Fi latency testbed under different MPTCP schedulers ..... 44
Figure 3-16 RTT delay CDFs for different schedulers under different intereference situations ..... 44
Figure 3-17 Delay difference under redundant scheduler. A positive value indicated that the packet in the Wi-Fi path arrived first ..... 44
Figure 3-18 Histogram of transmission resume delay ..... 46
Figure 3-19 TCP (between clients and server) over TCP (MPTCP tunnel between CPE and proxy) ..... 48
Figure 3-20 Updated eAT3S algorithm to consider different traffic types ..... 51
Figure 3-21 Requested file popularity distribution and file size (left) and arrival time of user requests (right) based on Poisson arrival process ($r$=16, $T$=10s) ..... 52
Figure 3-22 A snapshot from the simulator to depict dynamic change of telemetry data and eAT3S scheduling decisions ..... 53
Figure 3-23 Comparison of 5G SINR, weight assignment and steering score for load balancing steering mode when a steering threshold value of 0.6 is used ..... 54
Figure 3-24 a) Link capacity histogram used for the estimation of the available throughput experiment, b) absolute error between the available throughput estimation and the real capacity of a link ..... 56
Figure 3-25 Relative estimation error CDF ..... 57
Figure 3-26 Example of round robin weights which obtains ratio values greater thar 0.96 for a case with links whose capacities are 11500kbps, 9500kbps and 8500kbps ..... 58
Figure 3-27 Pseudocode for the “GCD criterion” ..... 59
Figure 3-28 Results of achieved versus maximum achievable throughput ratio with different weight assignation strategies ..... 59
Figure 3-29 Resulting throughput for different capacities sets and weight assignment algorithms ..... 60
Figure 3-30 Resulting ratio for different capacities sets and weight assignment algorithms ..... 60
Figure 3-31 5G-CLARITY Low-level eAT3S Architecture Overview ..... 61

Figure 3-32 Optimal ATSSS and MA rules update flow ..... 62
Figure 3-33 Selection and scheduling transport protocol ..... 63
Figure 3-34 Aggregated Throughput in 300 minutes Video Streaming ..... 64
Figure 3-35 Protocol operation overhead and the deviation ..... 64
Figure 4-1 Conceptual schematic of the proposed autonomic LiFi attocellular network slicing framework ..... 67
Figure 4-2 Interaction of the DQN-based SDN controller with the sliced LiFi attocellular network environment ..... 71
Figure 4-3 DQN Training for Multi-agent Learning Optimum Policy ..... 72
Figure 4-4 DQN enabled-autonomic matching of the Offered load with their allocated bandwidth for the eMBB and URLLC network slices in Mbps ..... 75
Figure 4-5 URLLC network slice service satisfaction ratio ..... 75
Figure 4-6 eMBB network slice service satisfaction ratio ..... 75
Figure 4-7 Average eMBB network slice data rate in Mbps ..... 76
Figure 4-8 G-ADRR system model (a) and SLA contract (b) ..... 77
Figure 4-9 G-ADRR experimental testbed ..... 81
Figure 4-10 Scenario 1: Network slices DAA ($\rho^{SLA}_0$ = 0.7 and $\rho^{SLA}_1$ = 0.3) ..... 83
Figure 4-11 Scenario 2: Network slices DAA ($\rho^{SLA}_0$ = 0.2 and $\rho^{SLA}_1$ = 0.8) ..... 84
Figure 4-12 Guaranteed network slices quota-based SLA violation ..... 85
Figure 5-1 Sub-6 GHz positioning system using 10 MHz/1 PPS synchronization ..... 87
Figure 5-2 PPS output of the ANs; a) unsynchronized; b) synchronized; c) leading edge of PPS pulse - synchronized ANs ..... 88
Figure 5-3 Sub-6 GHz positioning system using WR synchronization ..... 88
Figure 5-4 Distribution of TDOA from different APs ..... 89
Figure 5-5 mmWave positioning system architecture ..... 90
Figure 5-6 mmWave localization showing the anchor nodes and the estimated position of the UE ..... 91
Figure 5-7 CFDs of the estimated position of the UE ..... 91
Figure 5-8 LiFi based localisation system ..... 92
Figure 5-9 Test result showing RSSI reading and communication with server ..... 93
Figure 6-1 Flow diagram of the OOP based system level simulator ..... 96
Figure 6-2 Re-scanning and re-association handshake diagram ..... 97
Figure 6-3 Zipf content popularity distribution ..... 98
Figure 6-4 Probability density functions of the content requests for $\boldsymbol{\lambda = 102}$ events/sec ..... 99
Figure 6-5 Trapped random walk-based UE mobility diagram ..... 100
Figure 6-6 Random orientation model for the mobile terminals ..... 101
Figure 6-7 5G-CLARITY PHY layer link abstraction model ..... 103
Figure 6-8 Scenario description and random UE orientation ..... 106
Figure 6-9 (a) mobility map and AP deployment/labelling, (b) PDF and CDF distributions of the SIR values at the UEs ..... 109
Figure 6-10 (a) Mean, variance and median values, (b) Max and min values of the SIR distribution in dB ..... 110
Figure 6-11 The UE SIR distribution statistical properties when a random walk for 100 steps with LiFi, Wi-Fi, 5GNR-I and 5GNR-II APs are considered in the dense deployment scenario; (a) mean, variance and median values of SIR in dB, (b) max and min values of SIR in dB ..... 110
Figure 6-12 The UE SIR distribution statistical properties when a random walk for 100 steps with LiFi, Wi-Fi, 5GNR-I and 5GNR-II APs are considered in the dense deployment scenario; (a) mean, variance and median values of SIR in dB, (b) max and min values of SIR in dB. ..... 111
Figure 6-13 The UE SIR distribution statistical properties when a random walk for 100 steps with LiFi, Wi-Fi, 5GNR-I and 5GNR-II APs are considered in the dense deployment scenario; (a) mean, variance and median values of SIR in dB, (b) max and min values of SIR in dB ..... 112
Figure 6-14 Achievable throughput results for UEs with 100 steps random walk; (a) “Generic” deployment with SP transmission, (b) “Dense” deployment with SP transmission, (c) “Generic” deployment with CoMP-3 transmission, and (d) “Dense” deployment with CoMP-3 transmission ..... 113

# List of Tables

Table 1-1 Requirements and KPIs Mapping ..... 15
Table 2-1 WINNFORUM Testcases Passed ..... 27
Table 3-1 Details of Experimental Setup for Wi-Fi 6 Performance Experiments ..... 38
Table 3-2 MPTCP Parameters on CPE and MPTCP Proxy ..... 39
Table 3-3 Different RATs Throughput Measurement Performances ..... 40
Table 3-4 5GNR and Wi-Fi Access Network Configuration ..... 42
Table 3-5 Performance of the Virtual Testbeds ..... 48
Table 3-6 Comparison of Decisioning of Load Balancing and Smallest Delay/Redundant Schedulers for the Same AGV Position ..... 54
Table 4-1 LiFi AP Channel Simulation Parameters ..... 73
Table 4-2 Services and Traffic Simulation Parameters ..... 73
Table 4-3 DNN Training and Simulation Setup ..... 73
Table 6-1 MCS index, SINR, and Throughput Mapping Function, $\mathcal{M}(\cdot)$ [73] ..... 105
Table 6-2 Simulation Parameters ..... 106
Table 6-3 Adopted WAT Operation Frequencies [2] ..... 107
Table 6-4 Average Achievable Area Traffic Capacity Values for the Considered Scenarios ..... 114

# List of Acronyms

| | |
|---|---|
| 3GPP | 3rd Generation Partnership Project |
| 5GNR | 5G New Radio |
| AP | AP |
| AMF | Access and Mobility Management Function |
| ANR | Automatic Neighbour Relations |
| ARFCN | Absolute Radio Frequency Channel Number |
| AT3S / ATSSS | Access Traffic Steering, Switching, and Splitting |
| BBU | Baseband Unit |
| CBRS | Citizens Broadband Radio Service |
| CCGs | Common Channel Groups |
| CPE | Customer Premises Equipment |
| CPRI | Common Public Radio Interface |
| DL | Downlink |
| DMARL | Deep Multi-Agent Reinforcement Learning |
| DU | distributed unit |
| E2E | End to End |
| eAT3S / eATSSS | enhanced Access Traffic Steering, Switching, and Splitting |
| eMBB | Enhanced Mobile Broadband |
| FCC | Federal Communications Commission |
| G-ADRR | Global-Airtime Deficit Round Robin |
| GAA | General Authorised Access |
| GAL | General Access License |
| ICGs | Interference Coordination Groups |
| LLS | Low Level Split |
| LoS | Line of Sight |
| MAC | Medium Access |
| MAR | Multi-Access Rules |
| MLB | Mobility Load Balancing |
| mMTC | massive Machine Type Communication |
| mmWave | Millimieter Wave |
| MPTCP | Multipath Transmission Control Protocol |
| multi-WAT | Multiple Wireless Access Technology |
| Near-RT | Near Real Time |
| NHN | Neutral Host Networks |
| OANF | Optimal Access Network framework |
| OOP | Object Oriented Programming |
| O-RAN | Open RAN (Alliance) |
| PAL | Priority Access License |
| PBCH | Physical Broadcast Channel |
| PCC | Policy and Charging Configuration |
| PCI | Physical Cell ID |
| PLMN | Public Land Mobile Network |
| PN | Pseudo Noise |
| RAN | Radio Access Network |
| RAT | Radio Access Technology |

| | |
|---|---|
| RIC | RAN Intelligent controller |
| RSRQ | Reference Signal Received Quality |
| RSSI | Received Signal Strength Indicator |
| RTT | Round Trip Time |
| RU | Radio Unit |
| SAS | Spectrum Access System |
| SDF | Service Data Flow |
| SDR | software defined radios |
| SIB | System Information Broadcast |
| SINR | Signal to Interference plus Noise Ratio |
| SIR | Signal to Interference Ratio |
| SMF | Session Management Function |
| SRTT | Smooth Round Trip Time |
| SSID | Service Set Identifier |
| SSF | Special Subframe Format |
| TAC | Tracking Area Code |
| TCCS | TDD Configuration Connected Sets |
| TDD | Time Division Duplex |
| TDoA | Time Difference of Arrival |
| TTI | transmit time interval |
| TWR | Two-Way Ranging |
| UDP | User Datagram Protocol |
| UL | Uplink |
| URLLC | ultra-reliable low latency communication |
| URSP | UE Route Selection Policy |
| WR | White Rabbit |

# Executive Summary

This document, 5G-CLARITY D3.3, aims to provide evaluation results and final refinements on the proposed and refined 5G-CLARITY user and control plane architecture, which is reported in 5G-CLARITY D3.1 [1], and 5G-CLARITY D3.2 [2]. The proposed network structure, its components and related solutions reported in this document are aligned with the "Network Function and Application Stratum" technical requirements, which presented in 5G-CLARITY D2.2 [3].

In a nutshell, 5G-CLARITY D3.3 provides performance evaluations, implementational details and final refinements for the following items:

- **Spectrum sharing framework using citizens broadband radio service (CBRS)**: The proposed framework is based on an end-to-end (E2E) cloud native and fully disaggregated 5G New Radio (5GNR) solution. This employs 3rd-party distributed units (DUs) and radio units (RUs) with multi-wireless access technology (WAT) support and enablement of near real-time (near-RT) radio access network (RAN) intelligent controller (RIC) xApps to deliver to the pilots.
- **Multi-connectivity and 5G-CLARITY eAT3S frameworks:** A complete design and evaluation of the 5G-CLARITY multi-connectivity framework is provided. The throughput, reliability and latency metrics are investigated for a multi-WAT based network. An investigation study is also introduced on the operational flows and algorithms to implement high- and low-level enhanced access traffic steering, switching, and splitting (eAT3S).
- **Advanced Resource management framework:** Deep multi-agent reinforcement learning (DMALR)-based autonomic LiFi attocellular network slicing framework is introduced. Wi-Fi network slicing based on Wi-Fi APs' airtime sharing scheme is presented, which considered the spatial and temporal users and traffic load across the network.
- **Positioning framework:** The multi-WAT based localisation architecture is optimized and synchronization capabilities are also included. WAT-specific positioning schemes, which operates on sub-6GHz, mmWave and LiFi bands and their performance evaluation are presented.
- **Integrated 5GNR/Wi-Fi/LiFi network performance evaluation:** Enhanced system level simulator with additional features and functionalities is presented. The link quality and performance models are employed to capture the potential reflection of the link-level simulations on the higher levels.

In Section 2, the spectrum sharing network that benefits from the CBRS Alliance dynamic spectrum access paradigm is given. Accordingly, the baseline legacy architecture as well as the newer specifications are provided. Then, the cloud native, disaggregated radio access networks (RAN) and 4G/5G generalised CBRS solution is presented. The obtained results are compared with the Wireless Innovation Forum (WINNFORUM) test harness, where an example testcase is also presented.

In Section 3, the final evaluation of the multi-connectivity with a novel application programming interface (API) to modify traffic steering policies and its demonstration has been presented. Moreover, benchmarking of the 5G-CLARITY multi-connectivity framework is obtained by using real-world 5GNR, Wi-Fi and LiFi infrastructure. Several algorithms for the 5G-CLARITY proposed enhanced access traffic steering, switching, and splitting (eAT3S) framework are presented and evaluated.

In Section 4, the intelligent Wi-Fi and LiFi slicing schemes to provide custom resource utilization flexibility are introduced. The network slicing problem is formulated, and the decision-making problem is solved by the developed algorithms. A DMARL-based autonomic LiFi attocellular network slicing framework is introduced along with the computer simulation results. Results are also presented on a sliced Wi-Fi network based on the Wi-Fi APs' airtime interface sharing.

In Section 5, the 5G-CLARITY proposed 5GNR, Wi-Fi and LiFi WATs based localization framework is presented. The WATs sub-6GHz, mmWave bands and LiFi based advanced positioning framework is evaluated and the results for real-world measurements are presented. A solution for the synchronization problem in the downlink time difference of arrival (DL-TDoA) is also presented.

In Section 6, an object-oriented programming (OOP) based system level evaluation simulator with enhanced simulation capabilities is developed. The enhanced simulator could capture more realistic scenarios as well as higher level architectural processing on top of the link level simulations. The section reports achievable signal-to-interference-ratio (SIR) distributions along with the throughput results.

# 1 Introduction

This document, 5G-CLARITY D3.3, provides the final refinements on the initially designed and proposed 5G-CLARITY control and user plane architecture, which is reported in 5G-CLARITY D3.1 [1]. This deliverable covers the last tier of refinements and potential deviations from the initial 5G-CLARITY control and use plane architecture. The design and implementation details of the integrated coexistence, multi-connectivity, resource management, and positioning/localization frameworks are aligned with the 5G-CLARITY architectural principles and technical requirements for "Network Function and Application Stratum" provided in 5G-CLARITY D2.2 [3] .The final refinements and evaluations in this deliverable can be summarized as follows:

1. The Citizens Broadband Radio Service (CBRS) based spectrum sharing framework that enables the co-existence to the 5G-CLARITY network structure is investigated and the testbed results are provided.
2. Advanced multi-connectivity framework with multi radio access technology (RAT) implementation and validation are provided. Enhanced access traffic steering, switching, and splitting (eAT3S) along with the achievable throughput, latency and reliability performances are investigated, and the results are presented.
3. Advanced resource management frameworks are developed, which incorporate interface selection, dynamic resource allocation, and machine learning (ML) based network slicing to support diverse services. The overall performance enhancement of sliced LiFi and Wi-Fi networks by using DMARL and bandwidth estimation algorithms are investigated, considering the network variable load in spatial and temporal spaces.
4. The performance of the advanced multi-WAT based localization framework is reported. The WATs sub-6 GHz, mmWave and LiFi are investigated, where a synchronization framework is also proposed, and the performance evaluation is reported. The positioning accuracy and reliability key performance indicators (KPIs) of the 5GNR/Wi-Fi/LiFi based network is obtained by the real-world measurement results.
5. The achievable 5G-CLARITY system area capacity KPIs via a new iteration of the 5G-CLARITY system level simulator is provided. The performance of the 5GNR/Wi-Fi/LiFi WATs based 5G-CLARITY network is also investigated by obtaining the signal-to-interference ratio (SIR) distribution and aggregated area capacity values, demonstrating that the project KPIs are achieved.

## 1.1 Scope of this document

5G-CLARITY D3.3 takes inputs from the previous WP3 deliverables, 5G-CLARITY D3.1 [1] and 5G-CLARITY D3.2 [2] as well as the previous and present WP2 5G-CLARITY D2.1 [4], 5G-CLARITY D2.2 [3], 5G-CLARITY D2.3 [5], and 5G-CLARITY D2.4 and 5G-CLARITY D2.5 [6]. The output of this document is going to be utilized by the 5G-CLARITY WP2 (D2.5), 5G-CLARITY WP4, 5G-CLARITY WP5, and 5G-CLARITY WP6. The preparation and a successful delivery of the 5G-CLARITY D3.3 marks the accomplishment of the milestone MS3.3 "*Complete control and user plane design to support coexistence of private and public networks, operates in licensed and/or unlicensed bands, multi-connectivity to 5G/Wi-Fi/LiFi, with enhanced aggregate area capacity and positioning performance is ready and evaluated*".

In 5G-CLARITY D3.1 [1], the state-of-the-art technologies on integrated network coexistence, multi-connectivity, resource management, and positioning are investigated, where the initial 5G-CLARITY network design with the 5GNR/Wi-Fi/LiFi wireless access technologies (WATs) is also proposed. Accordingly, the 5G-

CLARITY network user and control plane structure is focused on: (i) non-3GPP access to the 5G core network, (ii) coexistence of public and private networks via spectrum databases, (iii) multi-connectivity via 5GNR/Wi-Fi/LiFi, (iv) resource management and the traffic routing on the access network level as well as resource scheduling on the access point (AP) level, and (v) indoor positioning with cm-level localization and ns level synchronization capabilities.

In 5G-CLARITY D3.2 [2], the refinements and improvements on the state-of-the-art network design that is initially proposed in 5G-CLARITY D3.1 [1] are captured, and a preliminary performance evaluation is included. Especially, the non-3GPP access, which is an integral part of the multi-connectivity framework, as well as the control and user plane functionalities/algorithms have been emphasized. Moreover, refinements and evaluation regarding the resource management is investigated both by computer simulations and laboratory measurements in the developed testbed environment. Lastly, the evaluation of the multi-WAT based localization performance was also inspected and initial results were reported. Note that a potential way forward is provided for coexistence, multi-connectivity, resource management, and positioning frameworks to finalize the design, which is envisaged to meet the related key performance indicators (KPI) and technical objectives of the 5G-CLARITY project.

In 5G-CLARITY D3.3, the last tier of the refinements and potential diversions from the initial design that has been proposed in 5G-CLARITY D3.1 [1] are covered. Therefore, the final 5G-CLARITY system architecture, which includes both the theoretical and practical considerations for user and control plane functionalities, algorithms, and implementations are provided in the current deliverable, 5G-CLARITY D3.3. Moreover, the technical objectives OBJ-TECH-2 to OBJ-TECH-5 are addressed via development of the advanced spectrum sharing, multi-connectivity, resource management, coexistence, and localization frameworks. The inherited frameworks from 5G-CLARITY D3.1 and 5G-CLARITY D3.2 are enhanced and the results for the final design are presented. It is important to note that the findings and final design details obtained in 5G-CLARITY D3.3 will be reflected on 5G-CLARITY D2.4, D2.5, D4.3 and D5.4.

## 1.2 Objectives of this document

In the following, the specific objectives of this 5G-CLARITY D3.3 will be presented. Accordingly, the deliverable objectives could be mapped to the overall technical objectives of the 5G-CLARITY project (in the format of **OBJ-TECH-X**), which are summarized in the project's description of work (DoW) as follows:

- **D3.3 OBJ-1:** Further validation of the multi-technology coexistence framework that enables efficient spectrum sharing between the private and public networks (**OBJ-TECH-2**).
- **D3.3 OBJ-2:** Development of the multi-connectivity framework, which integrates the 5GNR/Wi-Fi/LiFi WATs by evolving the 3GPP Release 16 capabilities. The developed network will be able to provide: (i) more than 1Gbps downlink (DL) user experienced data rate, (ii) less than a millisecond latency in the air interface through multi/parallel access, (iii) at least six 9s reliability through smart interface selection, and (iv) vertical handover time less than 5 ms (**OBJ-TECH-3**).
- **D3.3 OBJ-3:** Provision of system area capacity of more than 500 Mbps per meter square by the utilization of smart radio resource management (RRM) algorithms (**OBJ-TECH-4**).
- **D3.3 OBJ-4:** Simultaneous support of positioning and related synchronization[1] infrastructure over the proposed 5GNR/Wi-Fi/LiFi WAT network, which can provide; (i) less than a centimetre peak positioning along with 99% availability for less than a meter positioning accuracy, (ii) ns level

[1] The synchronization related design and evaluations are not included in this deliverable because they have been reported in 5G-CLARITY D2.4.

synchronization by using the wireless transport of clock distribution protocols (**OBJ-TECH-5**).

## 1.3 Document structure

The rest of the document is structured as follows:

- In Section 2, the **D3.3 OBJ-1** is covered, where the spectrum sharing network that benefits from the CBRS Alliance dynamic spectrum access paradigm is presented.
- Section 3 covers the **D3.3 OBJ-2**, where the final evaluation of the multi-connectivity framework is provided.
- In Section 4, both the **D3.3 OBJ-2** and **D3.3 OBJ-3** are partly covered. Accordingly, intelligent Wi-Fi and LiFi slicing schemes that provide custom resource utilization flexibility to the designed network is given.
- In Section 5, the objective **D3.3 OBJ-4** is addressed. The 5G-CLARITY proposed 5GNR, Wi-Fi and LiFi WATs based localization framework is provided.
- In Section 6, the **D3.3 OBJ-3** is covered. An object-oriented programming (OOP) based system level evaluation simulator with enhanced simulation capabilities is presented.
- and finally, the Section 7 concludes the deliverable.

## 1.4 Fulfilment of 5G-CLARITY functional requirements and KPIs

In this section, the mapping between the project objectives, as defined in the DoW, related requirement KPIs, as defined in previous 5G-CLARITY D2.1 and D2.2, and verification of those elements by the specific component that has been developed during the lifetime of the project as well as the verification within the sections in this deliverable is provided in the following table.

**Table 1-1 Requirements and KPIs Mapping**

| Project Objectives | Requirement/KPI ID [D2.1, D2.2, KPI Tracking Sheet] | Requirement/ KPI Description | Component | Means of Verification / Status [D3.3 Section] |
|---|---|---|---|---|
| **OBJ-TECH-3** | D22-CLARITY-NFAS-R16, KPI-FUNC-REQ-18, KPI-FUNC-REQ-19 | The 5G-CLARITY network function and application stratum shall decouple downlink and uplink transmissions and shall have the capability to schedule downlink and uplink traffic to different WATs. | MPTCP | Sections 3.2.1, 3.2.2, 3.2.3, 3.2.4 provide functional validation of default, redundant, round robin, and 5G-CLARITY MPTCP schedulers. |
| **OBJ-TECH-4** | D22-CLARITY-NFAS-R18, KPI-FUNC-REQ-20, KPI-FUNC-REQ-21, KPI-FUNC-REQ-22 | The 5G-CLARITY network function and application stratum shall allow controlling physical resources of 5GNR gNBs, Wi-Fi and LiFi APs. | Wi-Fi airtime-based scheduler, LiFi airtime scheduling, LiFi spectrum resource scheduling | Sections 4.1 and 4.2 provide evaluation results of network-wide slicing of Wi-Fi networks with variable loads and autonomic LiFi attocellular network slicing, respectively. |
| **OBJ-TECH-3** | D22-CLARITY-NFAS-R19, KPI-FUNC-REQ-19, | The 5G-CLARITY network function and application stratum shall | dRAX (near/non-real-time RIC); Positioning | Section presents how dRAX can host a spectrum related microservice and hosts |

| Project Objectives | Requirement/KPI ID [D2.1, D2.2, KPI Tracking Sheet] | Requirement/ KPI Description | Component | Means of Verification / Status [D3.3 Section] |
|---|---|---|---|---|
| | KPI-FUNC-REQ-23 | support hosting xApps to provide value added services such as spectrum access system, localization server, real-time access traffic controller, integrated Wi-Fi/ LiFi network controller, etc. | server | xApps. Sections 2, 3, 4 and 5 provide details and evaluations on spectrum access, multi-access, integrated Wi-Fi/LiFi network controller and localization frameworks, respectively. |
| **OBJ-TECH-3** | D22-CLARITY-NFAS-R20 | The 5G-CLARITY network function and application stratum shall provide necessary telemetry data to the hosted xApps. | dRAX (near/non -real-time RIC) | The xApp deployment in the 5G-CLARITY RAN cluster is discussed in Sections 2.4.2 and 3.1. Telemetry is demonstrated in the demonstration described in section 3.1. |
| **OBJ-TECH-3** | D21-5GC.KPI-2, KPI-FUNC-REQ-5 | Reducing latency in the air interface < 1 ms for uplink and downlink through parallel access across various technologies | MPTCP | The evaluation results for the E2E latency achievable by the 5G-CLARITY multi-connectivity framework is provided in Section 3.2.2. |
| **OBJ-TECH-3** | D21-5GC.KPI-3, D21-TECH-UC2.1-01, D21-TECH-UC2.1-02, D21-TECH-UC2.1-03, D21-TECH-UC2.1-04, D21-TECH-UC2.1-05, D21-TECH-UC2.2-08, KPI-FUNC-REQ-6 | Providing reliability of at least six 9s through smart interface selection | MPTCP | The reliability evaluation and further details are provided in Section 3.2. |
| **OBJ-TECH-3** | D21-5GC.KPI-4, KPI-FUNC-REQ-7 | Supporting vertical handover between wireless technologies with handover times < 5 ms | MPTCP | Section 3.2.3 provides further information and evaluation results for less than 5 ms vertical handover between wireless technologies. |
| **OBJ-TECH-4** | D21-5GC.KPI-5 KPI-FUNC-REQ-8 | Demonstrate aggregate system area capacity in relevant indoor scenarios > 500 Mbps/m^2 through smart RRM algorithms and SDN control frameworks that fully exploit the capacity of the combined 5G/Wi-Fi/LiFi access | Simulation platform | Section 6 provides a simulator development process and presents obtained link quality and performance results. |
| **OBJ-TECH-5** | D21-5GC.KPI-8 KPI-FUNC-REQ-9 | Positioning to a peak accuracy < 1 cm, and availability of < 1 meter accuracy 99% of the time. | Positioning server | The multi-WAT positioning framework development and related results are provided in Section 5. |
| **OBJ-TECH-5** | KPI-FUNC-REQ-10 | Synchronization to the ns-level via wireless transport of clock distribution protocols. | Simulation Platform | Synchronization results are demonstrated in D2.4, section 5 and 6. |
| **OBJ-TECH-2** | KPI-FUNC-REQ-2 | Validation of slice aware | dRAX (near/non | The details of a co- |

| Project Objectives | Requirement/KPI ID [D2.1, D2.2, KPI Tracking Sheet] | Requirement/ KPI Description | Component | Means of Verification / Status [D3.3 Section] |
|---|---|---|---|---|
| | KPI-FUNC-REQ-14 | spectrum broker | -real-time RIC) | existence and spectrum sharing architecture based on CBRS are presented in Section 2. |
| **OBJ-TECH-2** | KPI-FUNC-REQ-15 | Integrated LTE SAS client used in CBRS Small Cell context will be implement in O-RAN dRAX content as an xApp | dRAX (near/non -real-time RIC) | Disaggregated and cloud native CBRS client is implemented, and results are provided in Section 2. |
| **OBJ-TECH-5** | KPI-FUNC-REQ-24 | 5GNR location report interface setup | Positioning server | Localisation server implementation will be provided in 5G-CLARITY D5.2. |
| **OBJ-TECH-5** | KPI-FUNC-REQ-26 | mmWave position interface | Positioning server | mmWave based positioning interface details and implementation results are provided Section 5.2. |
| **OBJ-TECH-5** | KPI-FUNC-REQ-27 | LiFi position interface | Positioning server | LiFi positioning server implementation and related results are provided in Section 5.3. |
| **OBJ-TECH-5** | KPI-FUNC-REQ-28 | Location management of 3GPP and non-3GPP access | Positioning server | Theoretical and simulation-based investigations and related results are provided in Sections 3.2.3, 5.1, 5.2 and 5.3. |

# 2 Spectrum Sharing Framework with CBRS

## 2.1 Introduction

The spectrum sharing framework for 5G-CLARITY 5GNR access leverages CBRS Alliance dynamic spectrum access paradigm, as described in 5G-CLARITY D3.1. This innovative regulatory regime, albeit was initially the focus of the United States (US) market, might potentially be replicated elsewhere. Its intrinsic advantage is that since it is based on a 3-tier approach, it can be mapped to different regulatory spectrum regimes. These include from the usual traditional licensed ones to others where incumbent protection or local vertical licenses are granted to private network deployments.

## 2.2 Baseline legacy CBRS architecture and newer specifications

In April 2015 [7] and May 2016 [8] the Federal Communications Commission (FCC) established the regulatory grounds for the CBRS, involving the shared commercial use of the 3.5 GHz (3550-3700 MHz) with incumbent military radars and fixed satellite stations. The derived CBRS specifications and certification framework were defined together between the Wireless Innovation Forum [9] and the CBRS Alliance [10]. At that time, Long Term Evolution (LTE) was the 3rd Generation Partnership Project (3GPP) mainstream technology in use, since the 5GNR specifications were only defined in the 3GPP Release 15 in 2019. As a result, the CBRS system was based on 3GPP LTE, for which a specific B48 band was defined in the 3GPP specifications.

At that time, most base stations were based on either fully integrated enhanced NodeBs (eNBs) architectures (Centralised Unit (CU), Distributed Unit (DU) and Radio Unit (RU) functionality in the same box) or Baseband Unit (BBU)/RU splits using Common Public Radio Interface (CPRI) interface. It was still very early days for a fully disaggregated RAN as later on defined by 3GPP with a normative High Layer Split based on F1 interface between CU and DU, and a Low Layer Split based on Fronthaul 7.2x interface defined by Open RAN (O-RAN) Alliance.

CBRS was defined to make additional spectrum available for flexible wireless broadband use while ensuring interference protection and uninterrupted use by the incumbent users. The main components of the CBRS system are depicted in Figure 2-1.

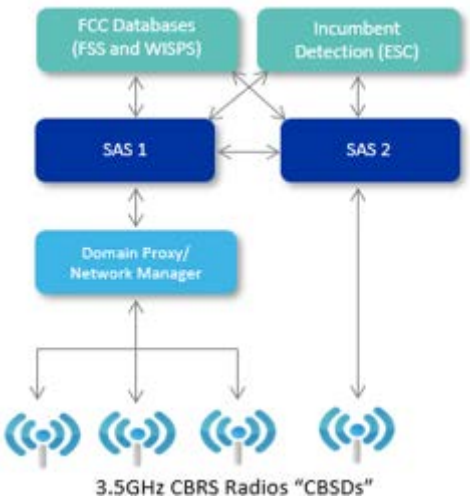



**Figure 2-1 CBRS overall architecture**

In CBRS, a novel three-tier sharing paradigm coordinates spectrum access among the incumbent military radars, satellite ground stations and temporarily protected Fixed Wireless Access (FWA) legacy stations and new commercial users.

The three tiers sharing the 150 MHz of CBRS spectrum are:

- **Protected Incumbents**: this tier is protected from interference caused by Priority Access License (PAL) and General Access License (GAA) users. These are primarily radars and Fixed Satellite Service (FSS) stations (coastal areas mainly) and temporarily the Wireless ISPs in 3650-3700 MHz.
- **Priority Access License (PAL):** this tier is protected from interference caused by GAA users. Up to 70 MHz where available per county. This priority spectrum was awarded via auction during 2020 [11]. In principle a maximum of 40 MHz could be offered to a given licensee.
- **General Authorised Access (GAA):** this tier can use any portion of spectrum not assigned to Protected Incumbents or PALs in an area. A minimum of 80 MHz is reserved for this tier per county, although up to the full 150 MHz can be made available by the Spectrum Access System (SAS) if no Protected Incumbents or PAL users are present in the area.

As described in [12] the CBRS framework enables the use of sensing network inputs to enable real time awareness of naval radars and allows dynamic interference protection managed by the SAS, as shown in Figure 2-2.

Each SAS has a "map" of all deployments on the PAL channels and can facilitate opportunistic GAA use of vacant PAL spectrum in discrete geographic areas on a "use-it-or-share-it" basis. In the CBRS band, licenses (PALs) ensure interference protection for deployed nodes but confer no right to exclude opportunistic users (GAA) when and where the spectrum is not in use. Because the SAS has awareness of the transmit power, bandwidth and other characteristics of each device authorized to operate in a local area, it can make assignments to GAA users that optimize performance and mitigates mutual interference.

Figure 2-3 shows the overall CBRS architecture within an LTE system and the example deployment of Accelleran E1012 CBRS fully integrated small cell as starting point for the enhancements needed to:

- Have CBRS functionality supporting both LTE and 5GNR deployments
- Have CBRS functionality based on a cloud native paradigm
- Have CBRS functionality working in a fully disaggregated RAN with separate CU, DU, and RU

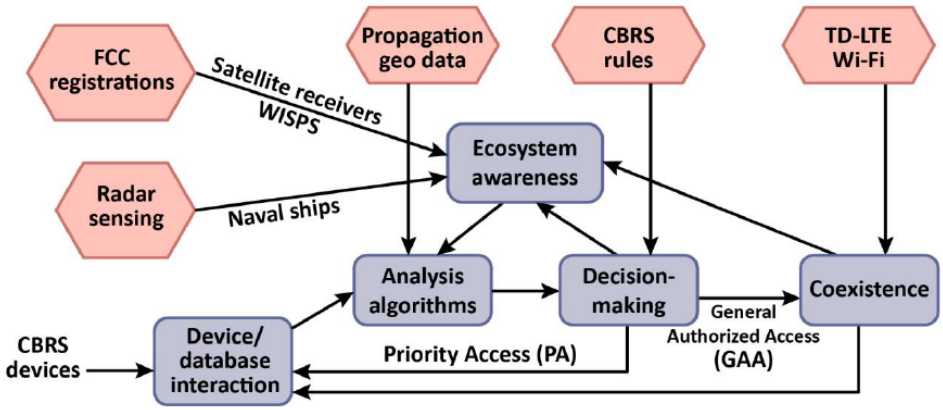



**Figure 2-2 Admission control system architecture CBRS**

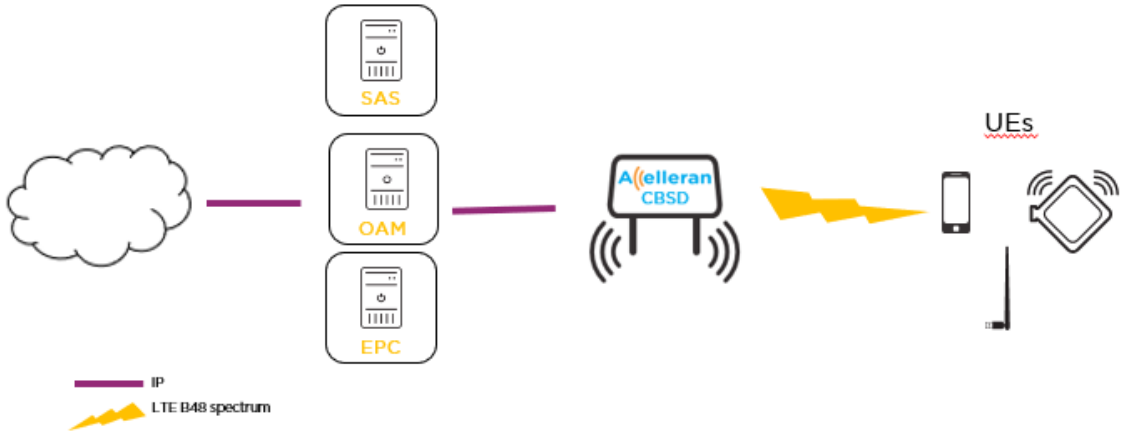



**Figure 2-3 Overall CBRS architecture in LTE system (using Accelleran small cell as example)**

The CBRS Alliance Release 1 standard, published in February 2018, described the extensions required to 3GPP standards to enable LTE operation in the US 3.5 GHz CBRS band. It encompassed radio, networks service and coexistence specification. The network services standard defines Neutral Host Networks (NHN) and private networks operation. The CBRS Alliance Release 2 standard, published in April 2019, added Multi-Service Operator (MSO) and fixed wireless use cases, non-SIM access mode (non-EPS-AKA) and User Equipment (UE) profiles. The standard included new LTE network identifiers for private or NHN networks, referred to as a shared Home Network Identifier (HNI), which are administered by the CBRS Alliance in conjunction to ATIS IOC [13] oversight. SAS operation is extended to facilitate coexistence between GAA devices.

The CBRS Alliance Release 3 standard, published in February 2020, included for the first-time support for 5GNR. In [14] and [15] CBRS specifications defined the 5GNR E-UTRAN New radio - Dual Connectivity (EN-DC) Use Case based on Non-Standalone (NSA) mode for Private and NHNs, initially as derived from 3GPP Release 15. In [16] CBRS defined the coexistence between and among multiple LTE and NR networks where Coexistence between Citizen Broadband Radio Service Devices (CBSD) belonging to the CBRS Alliance Coexistence Group is coordinated by one or multiple Coexistence Managers (CxMs). The specification included GAA coexistence requirements for CBSDs including cell phase synchronization, Time Division Duplex (TDD) Configuration for LTE-TDD and NR-TDD CBSDs amongst others. That version of the document focused on Band 48 [15] [14] LTE-TDD using Frame Structure 2 (FS2) and limited support for n48 NR-TDD deployment.

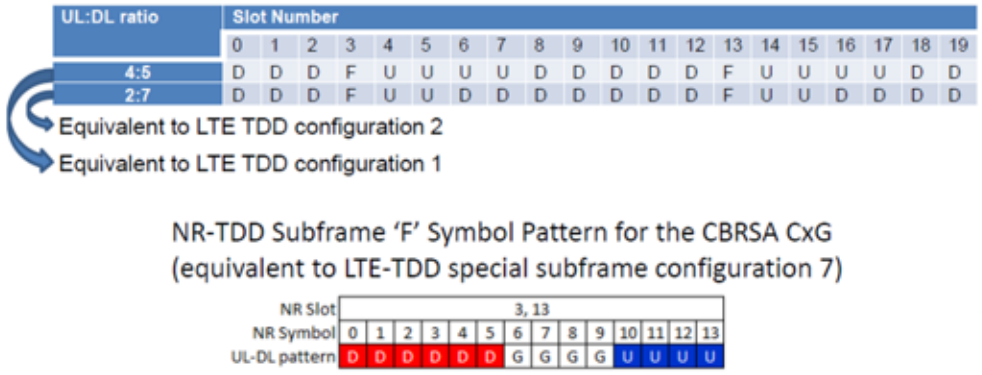



**Figure 2-4 Mandatory NR-TDD TDD UL/DL Configurations for the CBRSA CxG**

Additionally, all NR TDD CBSDs in a CBRS Alliance (CBRSA) Figure 2-4 "connected set" need to use the same or equivalent TDD configurations (G symbols are guard symbols for compatibility with F slots in LTE Special Subframe Format (SSF) 7).

The CBRS Alliance Release 4 standard, published in August 2021, and later updated to Release 4.1 in November 2021, included important enhancements in the architecture such as the introduction of support for 5G Standalone (SA) and Non-Public Networks and others related to Coexistence handling.

- Network Architecture ( [17], [18] and [19])
    - Extended Authentication for 5GNR SA
    - 5GNR SA Public Network support
    - 5GNR SA Non-Public Network support
- Coexistence ( [20]and [21])
    - Inter-CxM operation for TDD configuration coordination
    - Enhanced support for 5GNR TDD configuration
    - PAL TDD configuration coordination
    - TDD configuration coordination with CBSDs using non-3GPP technologies

For coexistence the SAS/CxM associated to a particular domain or network creates TDD Configuration Connected Sets (TCCS) with CBSDs having overlapping coverage contours according to specific propagation models. All LTE TDD and 5GNR TDD CBSDs belonging to the same TCCS are mandated to use the same or NR equivalent TDD configuration to avoid interference and are the basis for the primary channel assignments. Different CBSDs can also request the SAS/CxM to be part of Interference Coordination Groups (ICGs) and Common Channel Groups (CCGs) in which interference coordination may be done within themselves on the same assigned channel by the SAS/CxM. Otherwise, the SAS/CxM will assign separate channels. In case where multiple CxMs manage CBSDs in the same area, the CxMs exchange the list of CBSDs that are part of their respective TCCS and agree on the final common/shared TCCS to be used. If the CBSDs in the TCCS have different desired TDD configurations, there is a protocol based on majority voting to decide which of the default TDD configurations 1 or 2 are finally used. While in LTE TDD the configuration is specified by only 2 parameters, in 5GNR the configuration depends on 11 parameters (Subcarrier spacing, number of resource blocks, pattern periodicity, downlink (DL) slots, downlink symbols, uplink (UL) slots, uplink symbols…), but not all of them may influence the alignment of UL/DL symbols which will cause interference. There are certain combinations of each of those 11 5GNR parameters have equivalent UL/DL symbol alignment, which means that all the CBSDs in a TCCS have the same or equivalent desired TDD configuration. Therefore, they can use their desired configuration. Figure 2-5 shows how CBSDs, belonging to different domains, i.e., networks, and sharing the same geographical area, may potentially be interfering with each other according to the intra-SAS/CxM propagation models. They are assigned non-overlapping common TCCSs with non-interfering configurations and channels, because of the inter-SAS/CxM coordination.

The CBRS system has a great flexibility, as it supports different spectrum sharing tiers, protects incumbents and general priority licensed allocations from general use. It also manages the coexistence, coordination of TDD configurations, allocation of spectrum blocks, and transmission power across CBSDs in the same geographical area and manages co-existence across domains. These flexibility features of the CBRS system are the most important aspects to conceive the spectrum sharing framework based on CBRS in the 5G-CLARITY project.

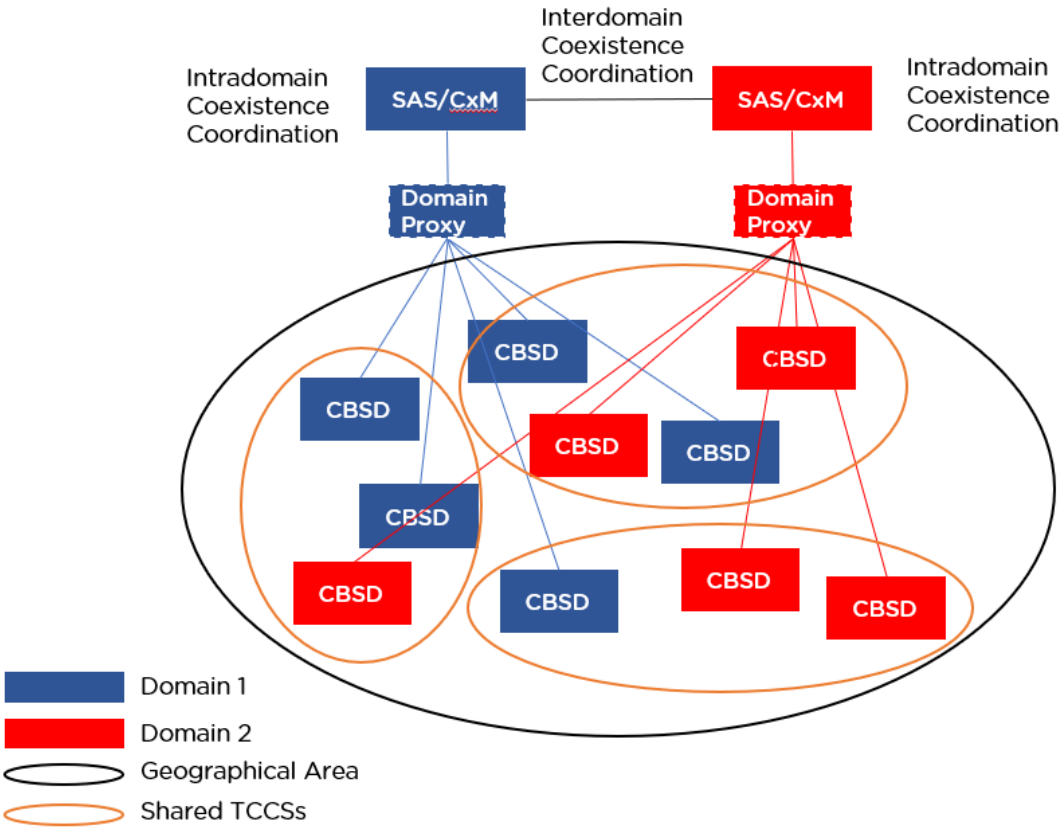



**Figure 2-5 Shared/Common TCCSs with multiple SAS/CxMs**

## 2.3 Industry status in CBRS alliance, O-RAN alliance and 3GPP

3GPP establishes the normative specifications for LTE and 5GNR evolution with interfaces that enable amongst others a disaggregated RAN based on a High-Level Split (HLS) Split 2 between the CU and DU. CBRS Alliance leverages on what 3GPP specifies, but with a focus on the additional specifications needed to enable CBRS operation in a dynamic spectrum access framework. O-RAN Alliance also leverages 3GPP specifications for the normative interfaces and enhances and extends what 3GPP defines with newer interfaces and specifications to enable a fully interoperable, open, disaggregated cloud native and intelligent RAN. One of these important interfaces is the Low-Level Split (LLS) or Fronthaul 7.2 between DU and RU which is not normative in 3GPP. Currently, across the industry there is not a cloud native, open and fully disaggregated RAN blueprint definition for CBRS.

Accelleran is involved in discussions with different companies and industry fora to enable the definition of an open RAN blueprint in CBRS context. The work done in 5G-CLARITY to show a cloud native and disaggregated CBRS testbed as described in the next sections can be easily aligned with the future direction the industry may take for such blueprint.

## 2.4 5G-CLARITY coexistence framework

### 2.4.1 5G-CLARITY co-existence framework high level architecture

Since the status of the disaggregated open RAN CBRS blueprint, latest releases of CBRS specifications and O-RAN interface alignment of DU/RU ecosystem are not currently mature enough or readily available, there are certain concepts that could not be validated in the testbed. However, as industry evolves these

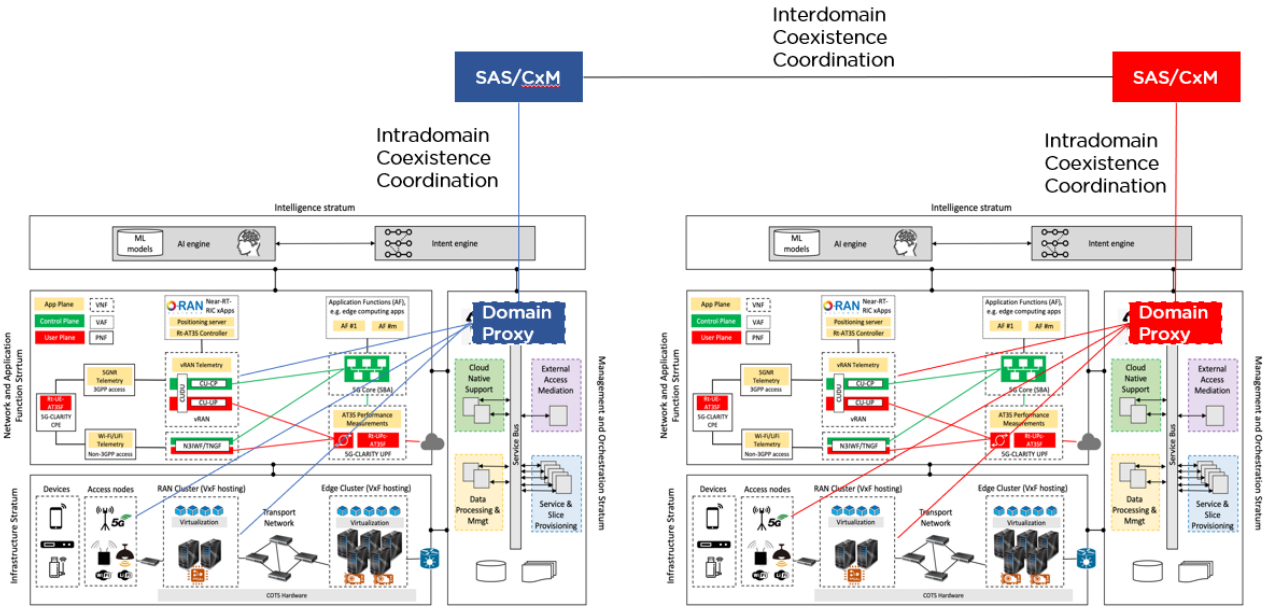



**Figure 2-6 5G-CLARITY spectrum sharing and co-existence architectural concept**

supporting concepts could be enabled. The following are some of the main aspects and concepts that could contribute to coexistence and slice awareness:

- PAL versus GAA allocations implemented in the SAS Server enable the protection of licensed spectrum blocks available at the same time as GAA ones in a particular geographical area. This can be used to avoid interference between macrocells deployed outdoors and small cells deployed in a non-public network (NPN), whether indoor or outdoor
- Co-existence management done by the CxM components within GAA coordinates co-existence between CBSD deployed in the same geographical area, so that compatible TDD configurations are used across CBSD with overlapping coverage areas according to propagation models. Channel separation in a geographical area within GAA is guaranteed by CxM to manage co-existence, while provisioning the requirements for channel separation for CBSDs. These can manage interference between themselves with their own proprietary methods, or mandate co-channel assignment for CBSDs that consider themselves as part of a common co-existence group.
- The use of DU schedulers capable of supporting slicing in terms of physical resource block usage or different bandwidth parts can be used. This realises slice awareness based on the actual availability of spectrum blocks in particular geographical areas as communicated via spectrum inquiries.
- The use of xApps/rApps responsible for Automatic Neighbour Relations (ANR), Coverage and Capacity Optimisation (CCO), Mobility Load Balancing (MLB), proprietary Interference Management (IM) across a group of CBSDs. Others can be leveraged so that co-existence and interference coordination provided by CBRS are enriched.

The architectural concept for 5G co-existence and spectrum sharing in 5G-CLARITY is shown in Figure 2-6.

### 2.4.2 Testbed description

In addition to the implementation of an E2E cloud native and fully disaggregated 5GNR solution using 3rd-party DUs (Effnet/Phluido) and RUs (Benetel) with multi-WAT support and enablement of nRT-RIC xApps to deliver to the pilots (the lion share of Accelleran contribution to 5G-CLARITY), Accelleran has also enabled a generalised CBRS Solution based on an E2E architecture. However, this testbed had some limitations in

terms of test harness and third-party components to support most recent releases of Winnforum/CBRS Alliance/O-RAN Alliance interfaces and procedures.

For the generalised CBRS solution the following aspects had to be developed:

- Generalisation of Radio Access Technology (RAT) to support both 4G and 5G
- Enabling CBRS client in a cloud native architecture
- Definition of a framework so that future interference management, co-existence, and self-optimisation 4G and 5G xApps can be developed appropriately
- Design the solution to align with the O-RAN architecture as much as possible. Enable flexibility to use any future interfaces that will be defined in the CBRS blueprint while still using the current proprietary interfaces available in the 3rd-party DUs and RUs used in the testbed

Figure 2-7 shows the evolution of the architecture in the solution towards a generalised cloud native open and disaggregated blueprint.

The idea, in the initial 5G-CLARITY proposal, was to enable a CBRS client as an xApp running on the nRT-RIC. However, the final decision for the CBRS testbed was to enable that functionality as a cloud native microservice deployed along the different dRAX microservices in the platform. This enables more flexibility for deployment and placement. The CBRS blueprint discussed by Accelleran and other stakeholders in the industry have different options that need to be studied. For example, whether a Domain Proxy or a single CBSD controller approach should be used, and where this component should be deployed (outside of O-RAN, in Service Management and Orchestration (SMO), in Non-RT RIC context, in near-RT-RIC context, etc.), and which interfaces should be used to manage the disaggregated RAN components. For the validation of the CBRS client functionality in the testbed the official WINNFORUM Test Harness is used in addition to the use of Open5GS as shown in

Figure 2-8. The WINNFORUM test harness is officially used to prove CBSD to SAS conformance as part of FCC certification process. The current version of this test harness is based on former CBRS Release 1 and WINNFORUM Release 1 specifications.

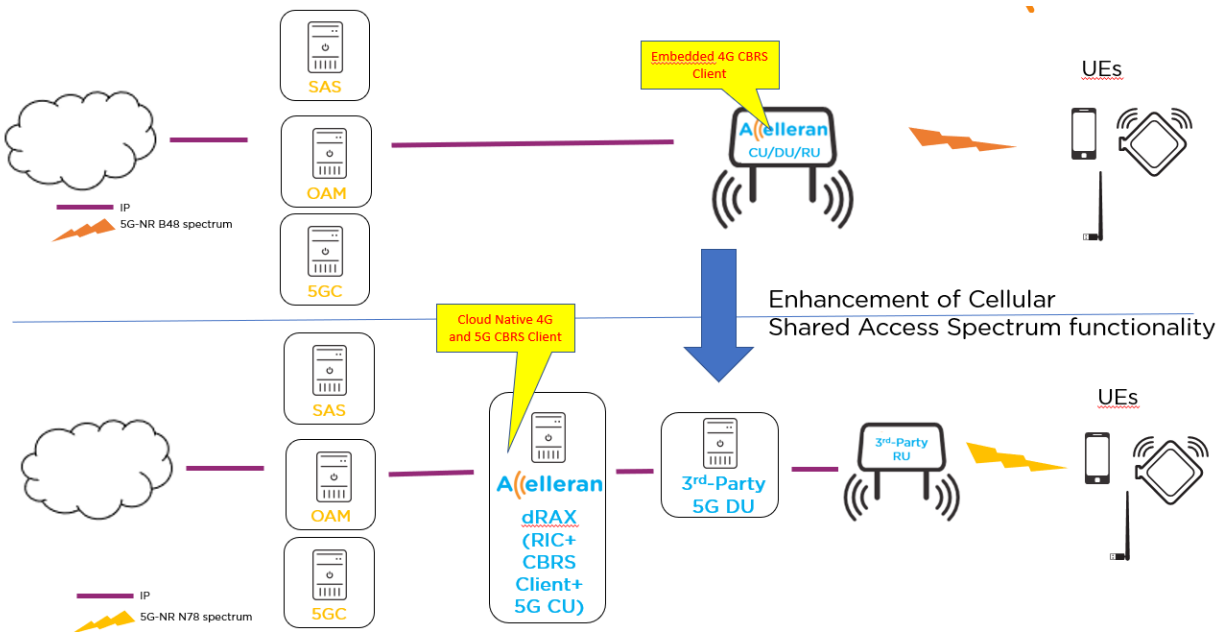



**Figure 2-7 Enhanced 4G/5G CBRS solution**

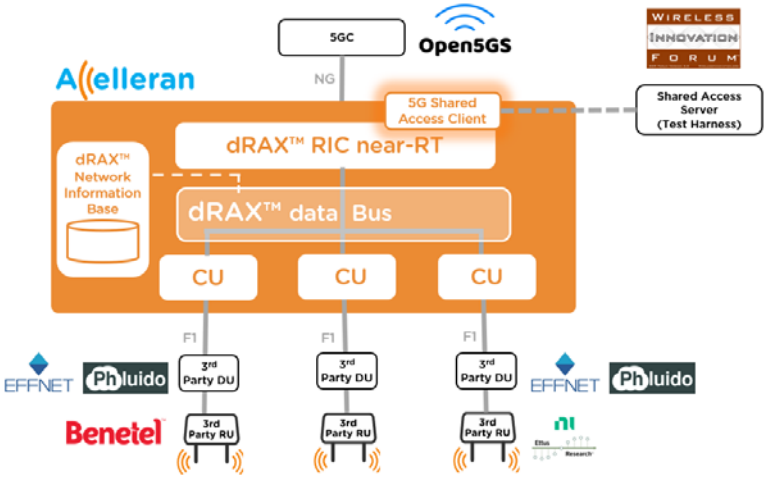



**Figure 2-8 5G CBRS testbed with Effnet/Phluido DU and Benetel and NI B210 RUs**

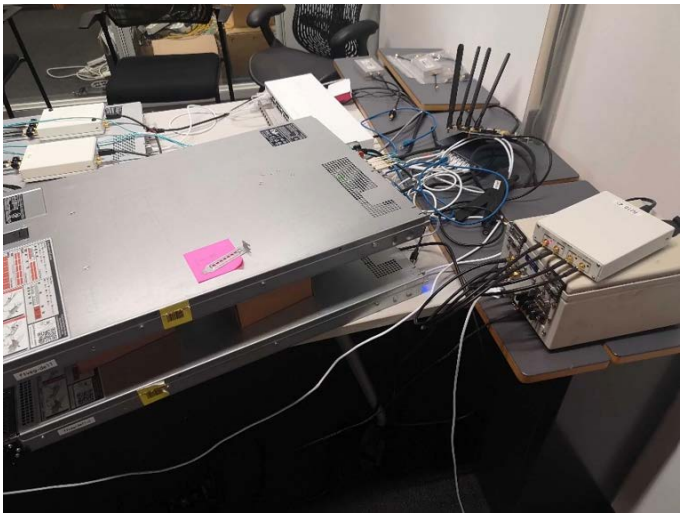


**Figure 2-9 5G CBRS testbed in the lab with B210 RU**

The testbed validation was done initially using the research RU National Instruments B210 Software Defined Radio (SDR), as shown in Figure 2-9. It was later validated with the commercial Benetel RUs that are also part of the multi-connectivity pilots.

### 2.4.3 Dependencies on disaggregated third-party DU/RU components

The initial E2E disaggregated dRAX 5G reference platform delivered by Accelleran with the dRAX 5G is based on the use of 3rd party DU (Effnet and Phluido) and RU (Benetel and National Instruments SDR) components. The CBSD acquires the grant from the SAS server with the Radio Frequency (RF) parameters that are allowed to be used by the base-station in a particular geographical area using the CBSD to SAS protocol. These parameters need to be configured appropriately and consistently across the DU and RU using the available interfaces. The RF parameters to configure are:

- Carrier frequency
- Transmission power
- Bandwidth

In the current testbed, the 3rd-party DU (Effnet and Phluido) is configured via a JavaScript Object Notation

(JSON) configuration file until the appropriate O-RAN O1 interface based in Netconf is available.

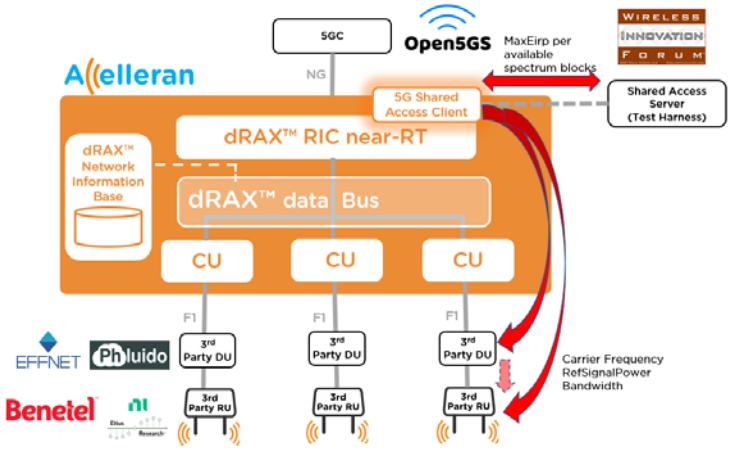



**Figure 2-10 5G CBRS DU/RU configuration**

The 3rd-party RU (Benetel) is configured via a Command Line Interface (CLI) until the appropriate O-RAN O1 interface based on Netconf is supported (hybrid architecture) or the O-RAN fronthaul management interface (hierarchical architecture) is supported between the DU and RU.

The carrier frequency, transmission power and bandwidth need to be consistently configured in the DU and RU. The carrier frequency and bandwidth in the RU need to be consistent with the New Radio Absolute Radio Frequency Channel Number (NR-ARFCN) and bandwidth configured in the DU. The transmission power configured in the RU needs to be consistent with the one advertised in the System Information Broadcast 1 (SIB1) by the DU for the Physical Broadcast Channel (PBCH). This is used by the UEs to calculate the actual pathloss experienced by the wireless link. Figure 2-10 shows the testbed setup with some of the main interfaces/parameters used from Winnforum Test Harness. The scope of this validation was limited to the current capabilities of Winnforum Test Harness based on Winnforum Release 1 baseline.

### 2.4.4 Solution design

Figure 2-11 shows the design of the CBRS client components and interfaces used. The main goal of the CBRS client is to implement the protocol with the SAS to obtain the RF parameters for the DU/RU and configure them for operation.

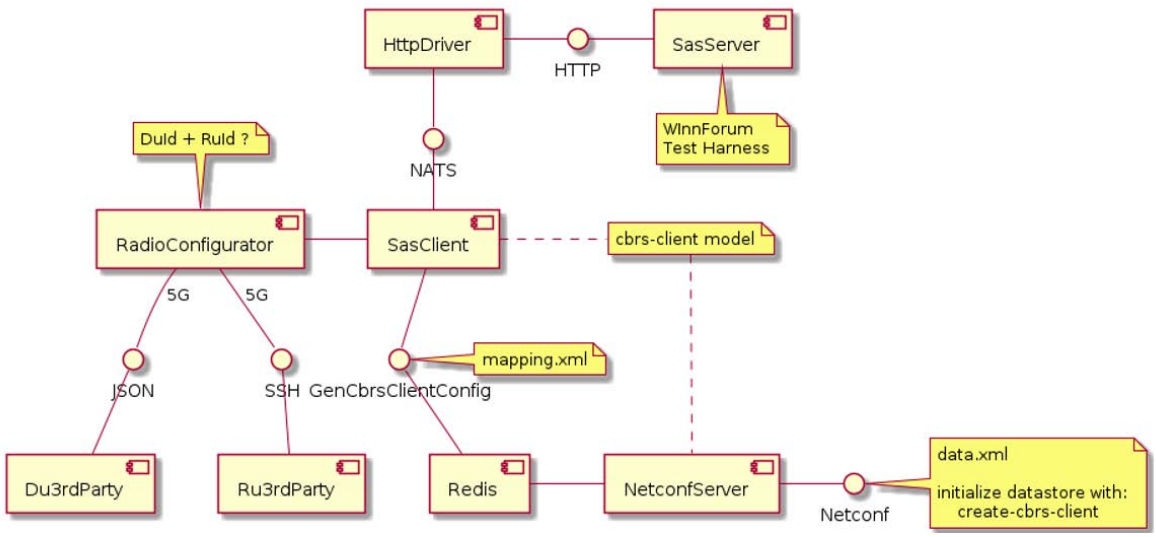



**Figure 2-11 5G CBRS client components**

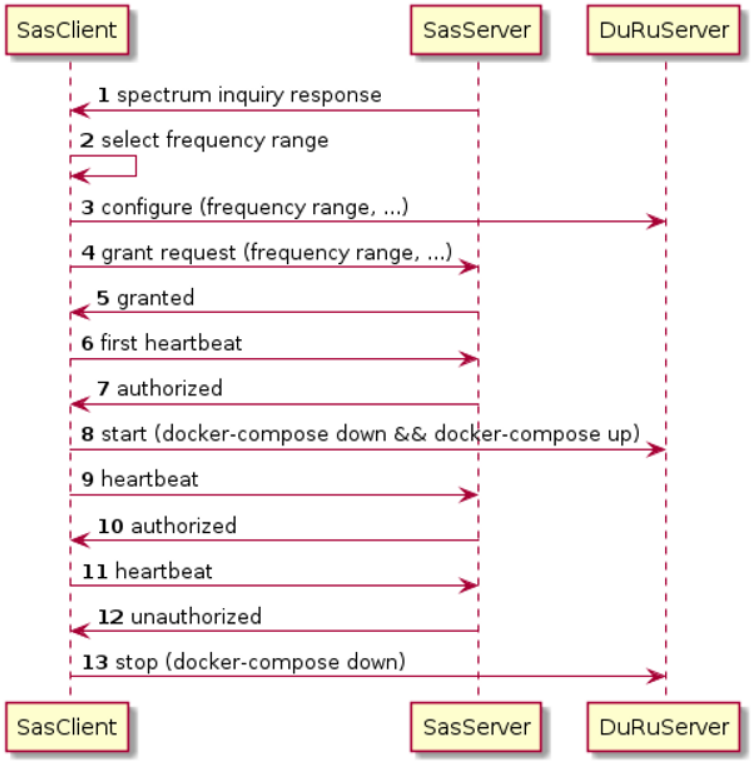



**Figure 2-12 5G CBRS client typical successful sequence**

Figure 2-12 shows the typical success sequence from the point where the CBSD (SasClient) receives a response to spectrum inquiry from the SAS (SasServer), requests the grant, acquires the grant, starts transmitting by configuring the DU/RU (DuRuServer) and stops transmitting when the grant is terminated as part of a heartbeat.

### 2.4.5 Test results against WINNFORUM test harness

The validation was done against the WINNFORUM test cases (official protocol conformance test harness based on WINNF-TS-0122), which run successfully with a pass verdict for different test cases.

**Table 2-1 WINNFORUM Testcases Passed**

| Testcase Group | Testcase ID | Testcase Title | Description |
|---|---|---|---|
| **Registration** | WINNF.FT.C.REG.1 | Multi-Step registration | This test validates that each of the required parameters appear within the registration request message |
| | WINNF.FT.C.REG.7 | Registration due to change of an installation parameter | This test validates that the CBSD sends notification to the SAS when an installation parameter has been changed |
| | WINNF.FT.C.REG.8 | Missing Required parameters (responseCode 102) | This test validates that the CBSD does not transmit because of the SAS reporting an unsuccessful registration due to missing required parameters |
| | WINNF.FT.C.REG.10 | Pending registration (responseCode 200) | This test validates that the CBSD does not transmit because of the SAS reporting the registration as pending |
| | WINNF.FT.C.REG.12 | Invalid parameter (responseCode 103) | This test validates that the CBSD does not transmit because of the SAS reporting an unsuccessful registration due to an invalid |

| Testcase Group | Testcase ID | Testcase Title | Description |
|---|---|---|---|
| | | | parameter |
| | WINNF.FT.C.REG.14 | Blacklisted CBSD (responseCode 101) | This test validates that the CBSD does not transmit because of the SAS reporting an unsuccessful registration due to a blacklisted CBSD |
| | WINNF.FT.C.REG.16 | Unsupported SAS protocol version (responseCode 100) | This test validates that the CBSD does not transmit because of the SAS reporting an unsuccessful registration due to an unsupported SAS protocol invalid parameter |
| | WINNF.FT.C.REG.18 | Group Error (responseCode 201) | This test validates that the CBSD does not transmit because of the SAS reporting an unsuccessful registration due to an error in the groupingParam optional field |
| **Grant** | WINNF.FT.C.GRA.1 | Unsuccessful Grant responseCode=400 (INTERFERENCE) | This test validates that the CBSD does not transmit because of the SAS reporting an unsuccessful grant response due to interference |
| | WINNF.FT.C.GRA.2 | Unsuccessful Grant responseCode=401 (GRANT_CONFLICT) | This test validates that the CBSD does not transmit because of the SAS reporting an unsuccessful grant response due to a conflict with the grant requested |
| **Heartbeat** | WINNF.FT.C.HBT.1 | Heartbeat Success Case (first Heartbeat Response) | This test case incorporates validation of successful Spectrum Inquiry messaging (if present) and successful Grant messaging into the Heartbeat Success case. It validates that the CBSD requests heartbeats with the heartbeat interval requested by the SAS and does not start transmitting until after the first heartbeat, with the RF parameters provided by the SAS |
| | WINNF.FT.C.HBT.3 | Heartbeat responseCode=105 (DEREGISTER) | This testcase validates that the CBSD stops transmitting within 60 seconds after being requested by the SAS due to a deregistration. The CBSD stays in deregistered state |
| | WINNF.FT.C.HBT.4 | Heartbeat responseCode=500 (TERMINATED_GRANT) | This testcase validates that the CBSD stops transmitting within 60 seconds after being requested by the SAS due to a grant termination. The CBSD stays registered but in unauthorised state |
| | WINNF.FT.C.HBT.5 | Heartbeat responseCode=501 (SUSPENDED_GRANT) in First Heartbeat Response | This testcase validates that the CBSD stops transmitting within 60 seconds after being requested by the SAS due to a grant suspension. The CBSD either relinquishes the grant or issues a first heartbeat and stays in a granted state |
| | WINNF.FT.C.HBT.6 | Heartbeat responseCode=501 (SUSPENDED_GRANT) in Subsequent Heartbeat Response | This testcase validates that the CBSD stops transmitting within 60 seconds after being requested by the SAS due to a grant suspension. The CBSD either relinquishes the grant or issues a subsequent heartbeat and stays in a granted state |
| | WINNF.FT.C.HBT.7 | Heartbeat responseCode=502 (UNSYNC_OP_PARAM) | This testcase validates that the CBSD sends a grant relinquishment and stops transmitting within 60 seconds after being requested by the SAS due to an inconsistent system state |
| | WINNF.FT.C.HBT.9 | Heartbeat Response | This testcase validates that the CBSD sends a |

| Testcase Group | Testcase ID | Testcase Title | Description |
|---|---|---|---|
| | | Absent (First Heartbeat) | first heartbeat within the requested heartbeat interval by the SAS while it is in granted state |

### 2.4.6 Example testcase (WINNF.FT.C.HBT.1)

This testcase shows the completion of the allocation of 100 MHz for 5GNR for GAA type deployment with 17 dBm Effective Isotropic Radiated Power (EIRP). The procedure starts with the registration of the CBSD, the spectrum inquiry procedure with available spectrum block, grant procedure to operate at a particular maxEirp over 100 MHz and subsequent heartbeat procedures. These can be used by the SAS server to request the change of RF operational parameters (maxEirp, frequency, …) due to co-existence or request the shutdown of the transmission because of incumbent or PAL operation.

### 2.4.7 Section summary

Shared spectrum access frameworks, such as the one defined for CBRS, creates a new regulatory approach for new stakeholders. It allows them to have dynamic access to spectrum under reasonable licensing/cost conditions to deploy private networks. They can also leverage the co-existence frameworks for different networks to share the spectrum in an efficient manner. However, since the last specifications from the CBRS Alliance containing enhanced co-existence and 5G SA updates were released very recently (CBRS Alliance Release 4 during 2021 timeframe) the availability of commercial solutions incorporating these enhancements are not yet a reality. Furthermore, from the point of view of the Open RAN the industry does not have yet a multi-vendor and interoperable blueprint such as the one that can be defined as part of consortia such as the O-RAN Alliance. More work is required in the industry to develop commercial solutions, which can be combined with CBRS Alliance spectrum sharing and co-existence frameworks with O-RAN Alliance disaggregated and open RAN intelligence.

# 3 Advanced Multi-Connectivity and Multi-WAT Aggregation

In 5G-CLARITY D3.2 [2], an architecture validation and an initial evaluation of the 5G-CLARITY multi-connectivity framework were performed. This section completes the design and evaluation of the 5G-CLARITY multi-connectivity framework. The section is organised in three main contributions:

- In Section 3.1, we complete the design of the 5G-CLARITY multi-connectivity framework by introducing a novel API that allows an external process to modify traffic steering policies in an MPTCP based AT3S user plan function in real-time. We include a demonstration of this API. We note that all the developed enablers of the 5G-CLARITY multi-connectivity framework have been made available as open source under Apache license in [22].
- In Section 3.2, a benchmarking of the 5G-CLARITY multi-connectivity framework is described using real 5GNR, Wi-Fi and LiFi infrastructure. Target benchmarking KPIs include throughput, latency, performance during transitions between technologies, and a performance comparison when considering E2E tunnelling solutions required.
- Finally, in Section 3.2 the 5G-CLARITY several algorithmic implementations for the eAT3S framework are described and evaluated.

## 3.1 5G-CLARITY final architectural design

5G-CLARITY D3.2 [2] introduced the architectural design of the 5G-CLARITY multi-connectivity framework, which is based on the following design principles: i) integrating Wi-Fi and LiFi networks with the 5GCore through an N3IWF/TNGF function, and ii) using an MPTCP-based AT3S function to enable steering, splitting, and switching of traffic from UEs equipped with multiple access interfaces. A key innovation of this framework is to enable an external network function, such an xApp deployed in the 5G-CLARITY RAN cluster, to control the portion of traffic transmitted by each UE through each interface based on near-real time access network telemetry. In Section 3.1.1 we describe the implementation of an API that enables this external control. In Section 3.1.2 we report on a demonstrator that we build to illustrate the capabilities of this API. This demonstrator is publicly available in [23].

### 3.1.1 5G-CLARITY multi-connectivity eAT3S testbed REST API

In this section, details of application programming interface (API) for the 5G-CLARITY multi-connectivity testbed are discussed. Specifically, we develop a representation state transfer (REST) API to manage resources in the 5G-CLARITY virtual testbed [23]. First, the system model explaining how it works will be presented. Then, details of endpoints will be discussed.

#### 3.1.1.1 System model

To better visualize the interface and the interaction of the REST API, please first visit the link in [23]. The architecture of the system is illustrated in Figure 3-1. It contains a client that acts as a user agent, a web server that acts as a middleware between the user agent and the testbed application, and the application that runs the testbed. A user can initiate a HTTP/1.1 request to the web server. The interface and interaction are handled by Swagger [24], which is a description language to develop a REST API that is used JavaScript object notation (JSON) as the data format.

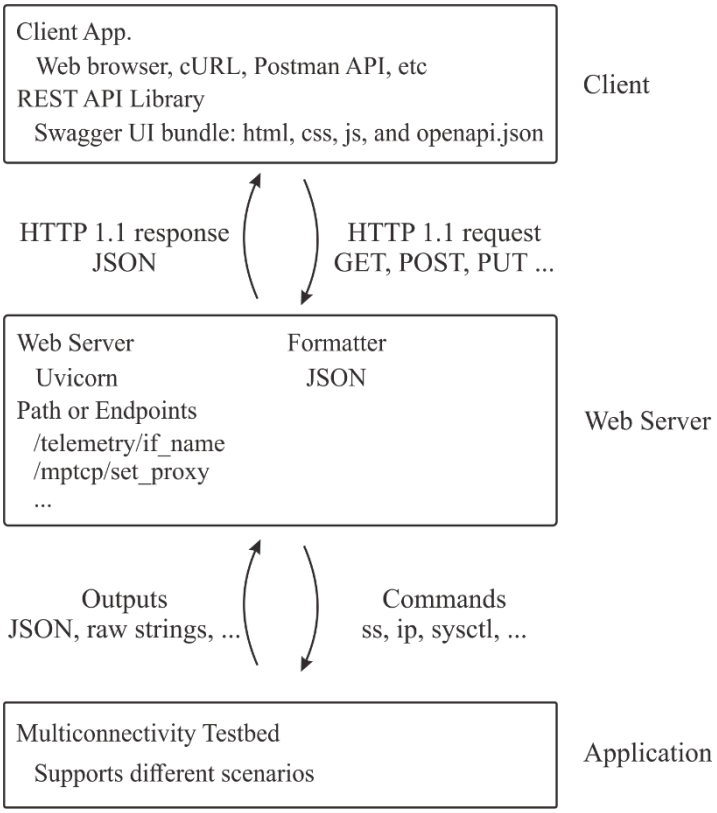



**Figure 3-1 Diagram of the REST API for the multi-connectivity testbed**

In the web server side, the Uvicorn, which is an asynchronous server gateway (ASGI) server [25], is used. It implies that Python (at least Python version 3.4) is used to run the web server. The web server handles the HTTP/1.1 requests based on the method and the request target (which will be later referred to as an endpoint). In addition, the web server can handle different ways of receiving input, i.e., path parameters (e.g., '/items/{item_id}'), query parameters (e.g., e.g., '/items/?id=10'), and a request body data in the form of JSON.

By using combinations of the HTTP request method, request target, and the user input, the web server proceeds the request accordingly to the multi-connectivity testbed by using corresponding Linux commands. For example, if user requests information regarding the interface name, then the web server can use the ip command[2]. Furthermore, the command sysctl is also used to change the MPTCP kernel during runtime, for example, to change the weights of different MPTCP subflows. Upon obtaining outputs from the application having different data formats (e.g., JSON, raw strings, or table-like outputs), then the web server also performs a data formatting to JSON as a respond the user's request.

As described in the multi-connectivity Github repository[3], the application supports multiple scenarios, such as a point-to-point scenario, a multiple UEs scenario, or a client with multiple proxies having a CPE in the middle are shown in the documentation. Practically, the REST API web server can be installed in all virtual machines (VMs) acting as different entities, such as a CPE, a UE, or a proxy. Therefore, endpoints will differ according to in which VMs that they are deployed. For example, in the CPE having an OVS switch, there will be specific configurations regarding the switch, while in the UE it does not have the endpoint. The current implementation supports the scenario having multiple proxies and a CPE. Other developers are free to

[2] https://man7.org/linux/man-pages/man8/ip.8.html
[3] https://github.com/jorgenavarroortiz/multitechnology_testbed_v0

develop their own endpoints according to their needs based on examples given in the Github repository.

#### 3.1.1.2 Examples of endpoints

In this subsection, we focus on endpoints that are already implemented and presented in the Github repository[4]. Referring to this, it lists the available endpoints, their HTTP methods, input, output, and their short descriptions. Specifically, for the output this is only listed if the request is handled successfully. If there is an error, then the server returns either the 4xx HTTP status code or 5xx HTTP status code. A user can query regarding many resources, e.g., interfaces, MPTCP, OVS, CPE, or simulate a delay.

### 3.1.2 Validation of the 5G-CLARITY multi-connectivity eAT3S REST-API

We have extended the functionality of the virtual testbed used for architecture validation in [2]. More precisely, we have added functionality that allows to control the type of MPTCP scheduler used for a particular connection, i.e., default, redundant or (weighted) round-robin. This is a useful functionality because different services may favour different AT3S implementations. For example, a latency sensitive may want to use a redundant MPTCP scheduler, whereas a bandwidth hungry service the round-robin one.

In Figure 3-2 , we can see an extension of the 5G-CLARITY virtual testbed featuring three independent MPTCP proxies. Also, we can see that the server and the client are connected in a layer 2 network. This allows us to dynamically change the MPTCP proxy during a TCP connection, since the endpoints (both the server and the client) are in the same IP network. Since each MPTCP proxy employs a different MPTCP scheduler, we can select the most suitable one in real time without requiring re-establishing current TCP connections.

The CPE and the MPTCP proxies execute ethernet bridges (using the `brctl` command). Additionally, the CPE also executes an OpenVSwitch (configurable using the `ofctl` command or using OpenFlow messages), which allows us to select one VLAN for the packets coming from the client. Each MPTCP proxy is associated with a VLAN, so choosing a VLAN allows us to select a particular MPTCP proxy, and therefore the type of service received by the connection.

An example of the operation of this testbed can be watched as a video in the YouTube channel of the 5G-CLARITY project [26] , including the REST API to select the proxy (and therefore the employed scheduler) and the scheduler parameters (e.g. the weights in the Weighted Round-Robin scheduler (WRR)), as well as obtaining the MPTCP telemetry.

Figure 3-3 presents an example of the operation of this virtual testbed, visualized in Grafana. In this example three proxies are employed, and the CPE dynamically changes between them; from the MPTCP default scheduler to the 5G-CLARITY scheduler (based on WRR) and finally to the MPTCP redundant scheduler.

The experiment starts with an *iperf* session between the server and the client. Initially, the CPE was configured to send traffic through proxy 1, which employs the MPTCP default scheduler. This scheduler selects the path with the lowest smooth round-trip time (SRTT), which is shown on the second right graph (labelled with "*1. RTT on the default scheduler (path with lowest RTT is selected*"). The variability of the SRTT causes instantaneous variations on the selected path, as shown on the first left graph (pointed with the label "*1. Traffic received from default sch.*"). After approx. 30 seconds, the CPE (using our REST API) selects proxy 2 (labelled with "*Switch from default sch. to 5G-CLARITY sch.*" between the second and third left graphs), which employs the 5G-CLARITY scheduler, so proxy 1 stops sending data (second left graph) and proxy 2 starts sending data (third left graph).

[4] https://github.com/jorgenavarroortiz/multitechnology_testbed_v0

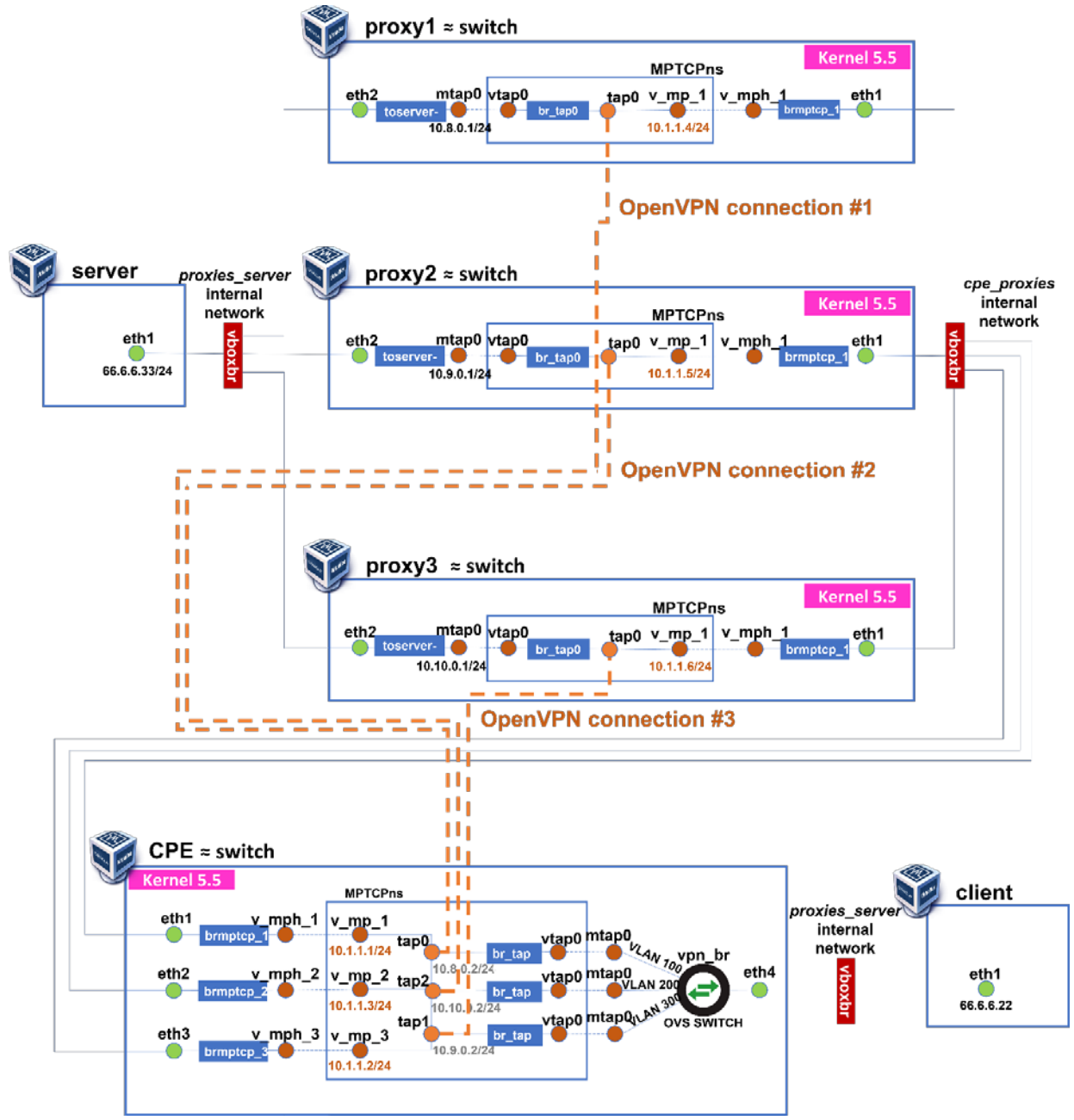



**Figure 3-2 Virtual testbed with OpenVSwitch.**

In this example, the weights for the different paths are identical (33.3%) so the data received through each path are also identical, as shown on the first left graph (labelled with "*2. Traffic received from 5G-CLARITY sch.*"). Finally, after approx. 35 seconds, the CPE selects proxy 3 (labelled with "*3. Switch from 5G-CLARITY sch. to redundant sch.*"), which employs the MPTCP redundant scheduler, so proxy 2 stops sending data (third left graph) and proxy 3 starts sending data (fourth left graph). Using this scheduler, all the paths send the same information to achieve redundancy, which is shown on the first left graph (labelled with "*3. Traffic sent to redundant scheduler*"). After approx. 35 seconds, the experiment concludes. All the traffic is generated by the same *iperf* session, meaning that the switch from one proxy/scheduler to another occurs seamlessly without breaking the TCP connection.

This virtual testbed achieves around 300 Mbps, which can be considered as very good performance since all the entities (client, server, CPE and the three proxies) are executed as virtual machines in the same host computer.

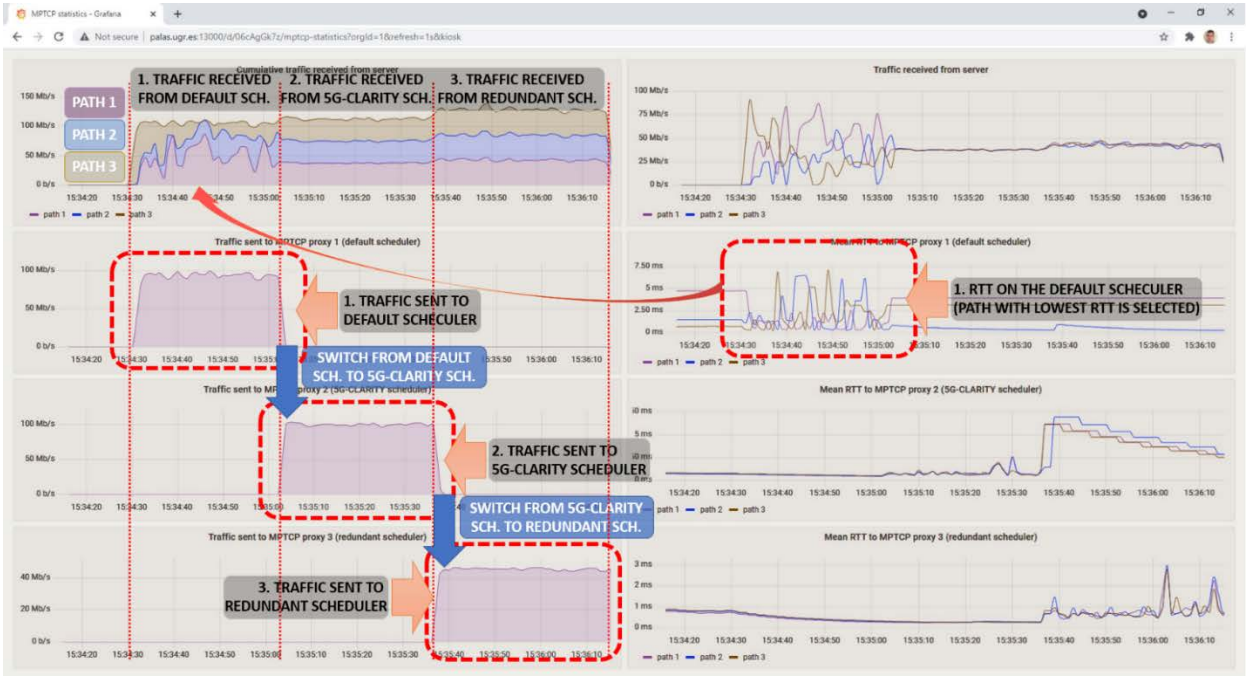



**Figure 3-3 Visualization of performance metrics (throughput and latency) in the virtual testbed**

## 3.2 Benchmarking 5G-CLARITY multi-connectivity framework

In this section, we validate in a virtual and real laboratory environment the 5G-CLARITY multi-connectivity framework using KPIs associated with 5G-CLARITY Objective 3.2 (O3.2), which reads as follows:

- 5G-CLARITY **O3.2**: Design and development of a multi-connectivity framework integrating 5G/Wi-Fi/LiFi evolving 3GPP R16 capabilities by:
    - Achieving downlink user experienced data rates > 1 Gbps through interface aggregation
    - Reducing latency in the air interface < 1 ms for uplink and downlink through parallel access across various technologies
    - Providing reliability of at least six 9s through smart interface selection
    - Supporting vertical handover between wireless technologies with handover times < 5 ms

We discuss in each section whether 5G-CLARITY has managed to achieve the targeted KPIs

### 3.2.1 Multi-WAT throughput benchmarking

The multi-WAT throughput validation against 5G-CLARITY O3.2 KPIs was completed based on experimental measurements obtained from the 5G-CLARITY multi-connectivity framework setup implemented at the Smart Internet Lab, the University of Bristol. The experimental setup consists of the 5G-CLARITY CPE, 5G, Wi-Fi, and LiFi radio and optical access technologies. To benchmark the obtained throughput of multi-WAT connectivity against the KPIs, link utilization analysis is performed on the individual access technologies. The throughput of 5GNR, Wi-Fi-6, and LiFi is compared to their aggregated throughput when MPTCP is deployed. In what follows, a short description of the Lab setup, network connectivity and results from the different access technologies is provided.

### 3.2.1.1 Multi-connectivity testbed setup

The multi-connectivity infrastructure setup consists of a multi-WAT CPE, which involves hardware configurations, deployment of network services and the CPE integration with MPTCP [27]. This integration enables the aggregation of traffic parameters such as throughput.

The 5G-CLARITY CPE setup consists of 5G modem, Wi-Fi-6 module, and LiFi dongle, which are all integrated into a single computer board running MPTCP enabled Linux kernel. The multi-WAT radio setup comprises of Nokia 5GNR RRH connected to a Open5Gs (5G-StandAlone setup) core, Ruckus Wi-Fi-6 access point as well as University of Bristol developed Wi-Fi-6 AP based on QCN9074, and LiFi access points. Each WAT is connected in an independent VLAN and IP subnet. To establish multi-connectivity, the MPTCP enabled 5G-CLARITY CPE is connected to an MPTCP proxy installed in a Virtual Machine (VM) hosted in an edge server as shown in Figure 3-4. The VM enabling the proxy for MPTCP has one interface connected to the VLANs and subnetworks of the 5GNR, Wi-Fi-6, and LiFi. The Robot virtual network functions (VNFs) connect the CPE indirectly only through a direct connection to the MPTCP proxy.

Monitoring and measurement modules are developed using iperf3, which enable the CPE to monitor radio parameters such as reference signal received power (RSRP) and signal-to-noise ratio (SNR), and network KPIs, namely throughput, latency, and jitter.

Figure 3-4 presents the architectural design of the multi-connectivity testbed used in the validation of the 5G-CLARITY multi-connectivity framework.

Figure 3-5 shows images of the 5G-CLARITY CPE used and the Nokia 5GNR radio deployed in the Lab environment at the Smart Internet Lab – University of Bristol.

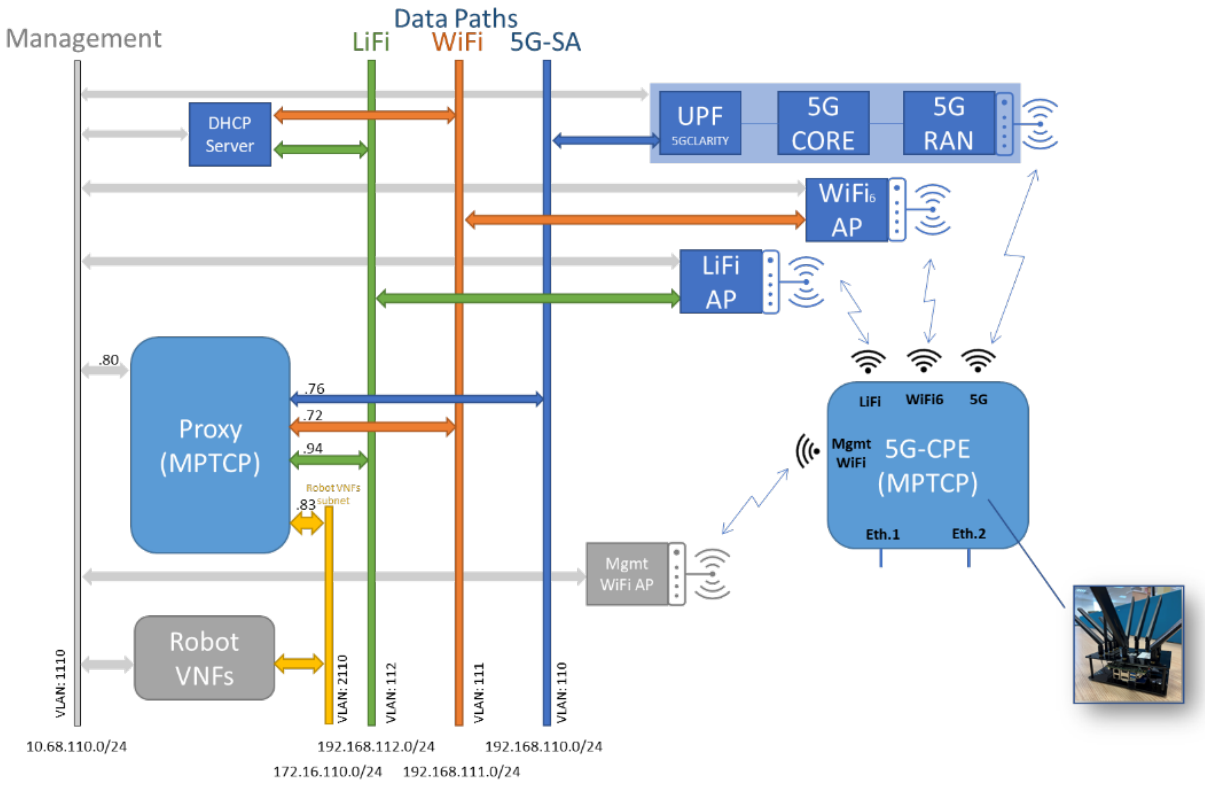



**Figure 3-4 5G-CLARITY multi-WAT framework testbed setup architecture.**


**Figure 3-5 5G-CLARITY multi-WAT framework testbed setup at the Smart-Internet Lab -UoB**

The hardware configuration of the CPE comprises of the following components:

- Processor:
  - Pentium Quad Core
- Memory:
  - 8GB
  - 5G Modems
  - 1x SIM8200ea-m2
  - 1x Quectel RM500Q
- Antennas:
  - 8x 5G antennas
  - 2x Wi-Fi Antennas
  - 1x GPS antenna
- Ports
  - 2x Gigabit Eth.
  - 3x USB3 Ports
  - 1x HDMI
  - 1x DP
  - RS232 and GPIO
- Power input
  - 12 or 24 VD

The following are the 5GNR Configuration deployed in the setup:

- RAN: Nokia AWHQM
- Band 77, 100 MHz BW, MIMO 4(tx)x2(rx)
- Subcarrier spacing: 30 KHz
- Frame structure type: Semi Static
- Guard Period length: 2 Symbols
- TDD configuration: 7DS2U (7/4 Nokia configuration)
- Environment: Lab

The Wi-Fi-6 configuration:

- Wi-Fi AP: Ruckus R850
- SSID: 5G-CLARITY Wi-Fi 6
- Frequency: 5.600 GHz
- Signal: -40 dBm
- RX bitrate: 1200.9 MBit/s 80MHz HE-MCS 11 HE-NSS 2 HE-GI 0 HE-DCM 0
- TX bitrate: 1200.9 MBit/s 80MHz HE-MCS 11 HE-NSS 2 HE-GI 0 HE-DCM 0
- Environment: Lab

The LiFi configuration:

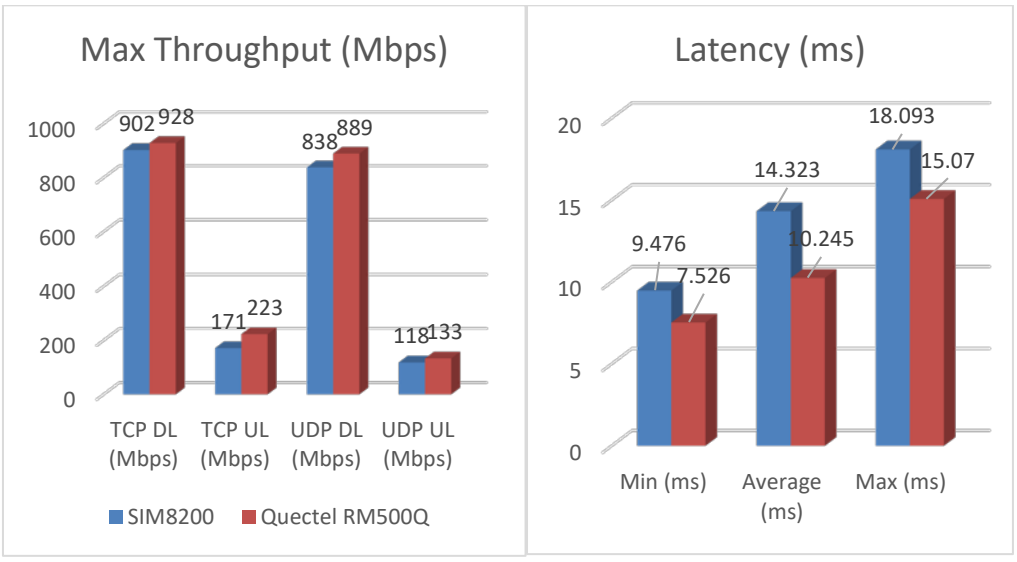
- LiFi AP: pureLiFi-X
- LiFi Client: pureLiFi USB dongle
- Environment: Lab

### 3.2.1.2 Multi-connectivity test

To measure throughput, the monitoring and measurement tool is deployed on the CPE. The throughput and latency were first measured for each individual radio or wireless technology connection before obtaining the multi-connectivity throughput aggregation. Each conducted experiment measures the throughput of Downlink (DL) and Uplink (UL) traffic in Mbps. Figure 3-6 presents the results of an experiment performed with 5GNR**.**

Figure 3-6 (a) shows the throughput results for two different 5G modems, SIM8200ea-m2 and Quectel RM500Q. It shows both modems could achieve up to 900 Mbps maximum throughput for TCP DL and up to 800 Mbps for UDP DL. In terms of the minimum latency, both 5G modems achieved 9.476 ms and 7.526 ms, as shown in Figure 3-6 (b). The Wi-Fi-6 throughput performance is presented in Figure 3-7 (a). These results were obtained by using the details which are presented in Table 3-1. The minimum, average and maximum Wi-Fi-6 latencies are presented in Figure 3-7(b).

Figure 3-8 presents the LiFi technology throughput and Latency performance results for DL and UL performance for both TCP and UDP. The best TCP DL throughput performance for LiFi is 34.23 Mbps as captured in Figure 3-8 (a), while the minimum Latency performance is 3.28 ms, as shown in Figure 3-8 (b).

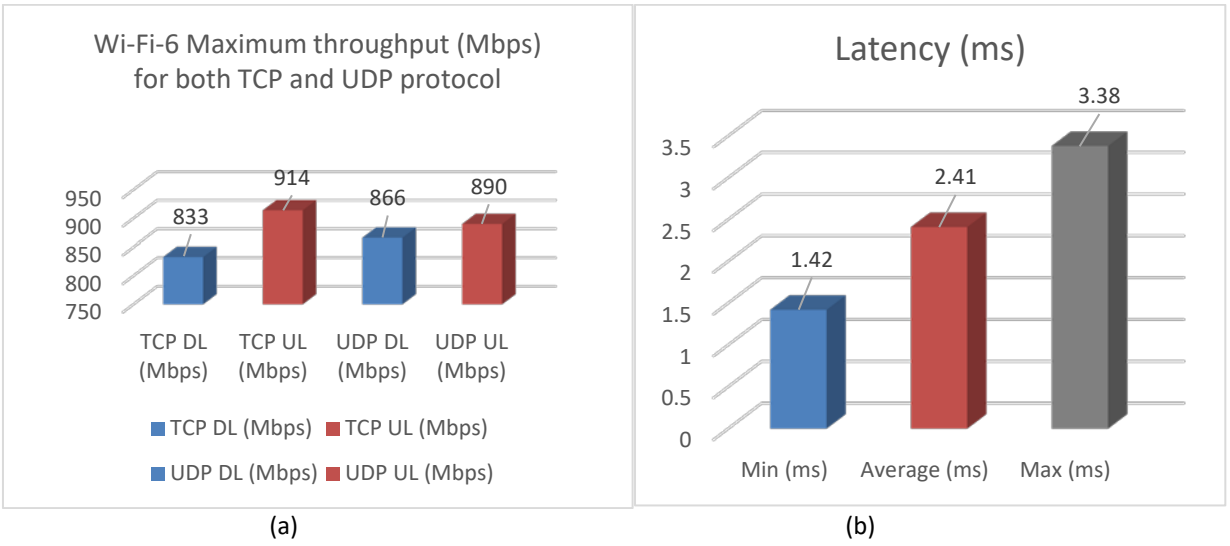



(a) (b)

**Figure 3-6 5GNR test with two types of UE setups: (a) Throughput in Mbps, (b) Latency in ms**

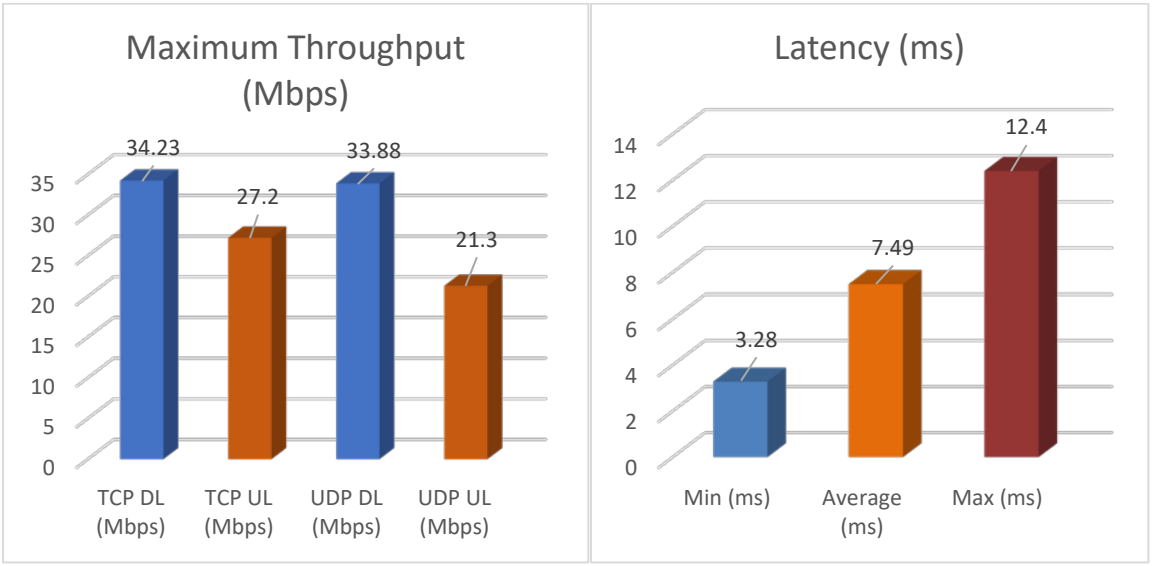
**Table 3-1 Details of Experimental Setup for Wi-Fi 6 Performance Experiments**

| AP | Wi-Fi Module | Kernel | Firmware Version | Tx/Rx bandwidth | Frequency |
|---|---|---|---|---|---|
| Ruckus R850 | Intel AX210 | 5.5.0 | 48.4fa0041f.0 | 80/80 MHz | 5.60 GHz |




(a) (b)

**Figure 3-7 Wi-Fi-6 Throughput and latency performance results**




(a) (b)

**Figure 3-8 LiFi Performance test results (a) Throughput in Mbps (b) Latency in ms**

The above results are individual link performance tests for 5GNR, Wi-Fi-6 and LiFi radio access technologies. As they are all integrated in the 5G-CLARITY CPE, using MPTCP, all the links are then aggregated to deliver high throughput to the end user. In this case, Table 3-2 presents the MPTCP parameters on both the CPE and MPTCP Proxy which has been described in section 3.2.1.1. The subsequent figures present the DL throughput (Mbps) for about 1 minute while iperf3 is running.

Figure 3-10 presents the download throughput over the Wi-Fi-6 link while iperf3 test is running. The average speed over the Wi-Fi-6 is about 750 Mbps.

**Table 3-2 MPTCP Parameters on CPE and MPTCP Proxy**

| Parameter | Value |
|---|---|
| kernel.osrelease | 5.5.0-mptcp |
| net.mptcp.mptcp_binder_gateways | - |
| net.mptcp.mptcp_checksum | 1 |
| net.mptcp.mptcp_debug | 0 |
| net.mptcp.mptcp_enabled | 0 |
| net.mptcp.mptcp_path_manager | fullmesh |
| net.mptcp.mptcp_scheduler | roundrobin |
| net.mptcp.mptcp_syn_retries | 3 |

**Figure 3-9 Download throughput over the 5GNR link while iperf3 test is running: the average speed over this link is about 800 Mbps**

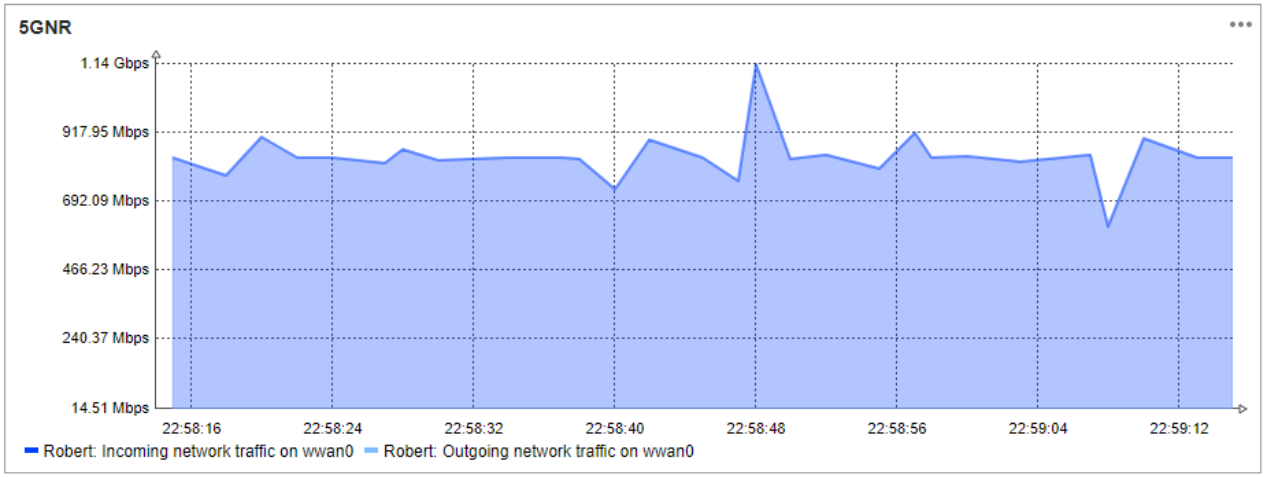



**Figure 3-9 5GNR download throughput when MPTCP is in place**

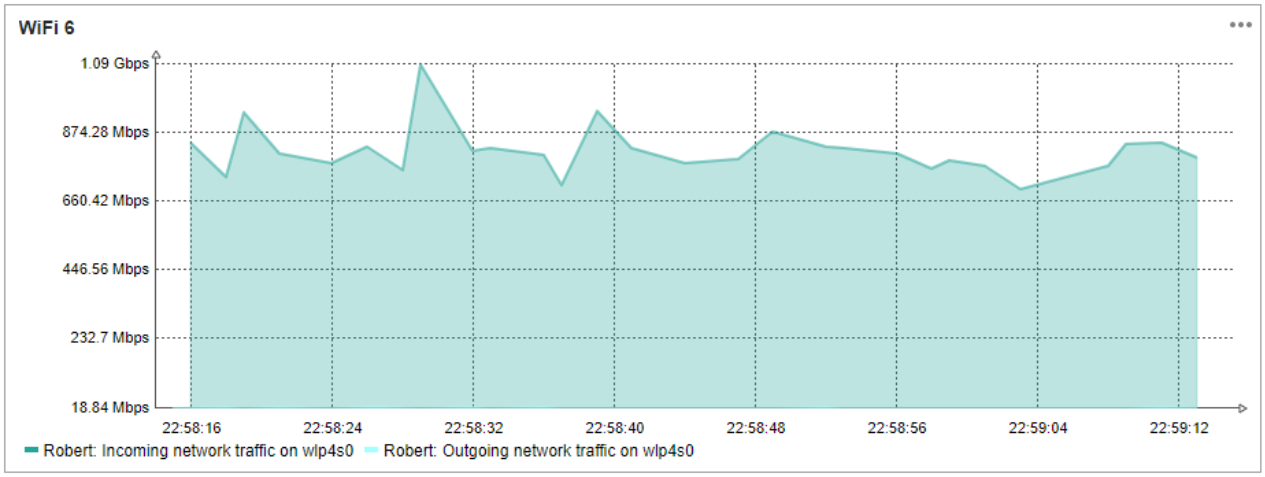



**Figure 3-10 Wi-Fi-6 download throughput while MPTCP is running**

Figure 3-11 presents the download throughput over the LiFi link while iperf3 test is running. The average speed over this link is about 34 Mbps. Figure 3-12 shows the aggregated throughput over all the three RATs and this is what iperf3 measured during the performance test.

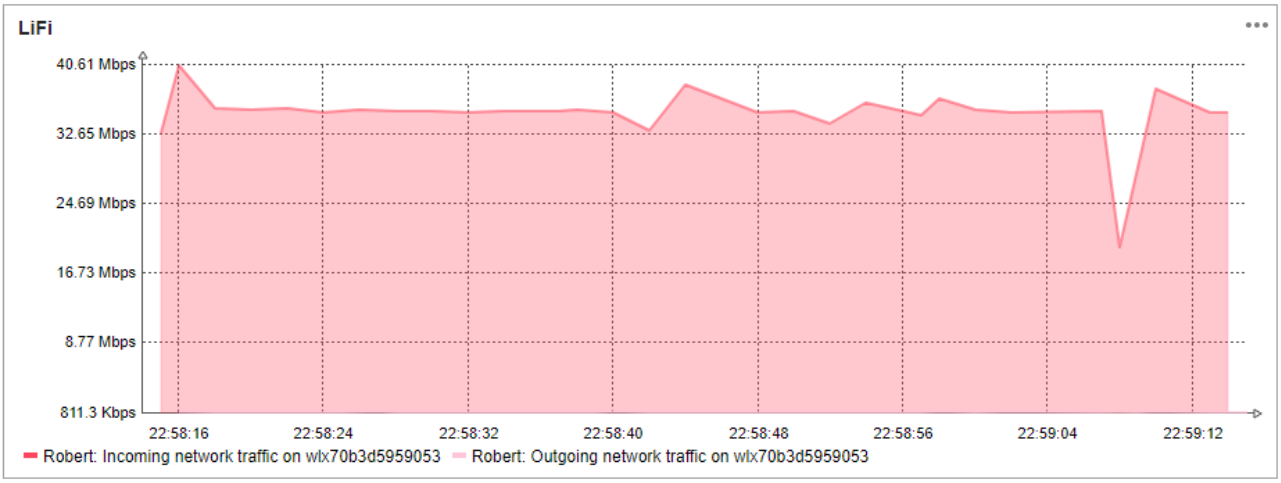



**Figure 3-11 LiFi download throughput while MPTCP is running**

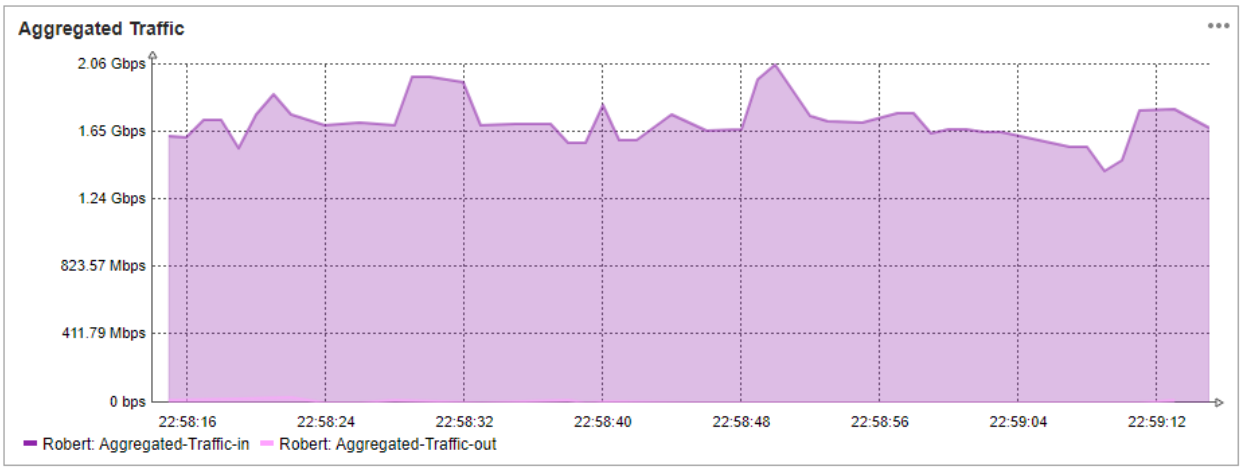



**Figure 3-12 Aggregate download throughput over all radio access technologies while MPTCP is running**

**Table 3-3 Different RATs Throughput Measurement Performances**

| Radio Access Technology | Average Max. Throughput Performance | | | Average Min. Measured Latency |
|---|---|---|---|---|
| | Measured Individually | Measured with MPTCP in place | Aggregated Throughput with MPTCP in place | |
| 5GNR | 900 Mbps | 800 Mbps | 1.45 Gbps | 7.50 ms |
| Wi-Fi-6 | 830 Mbps | 750 Mbps | | 1.40 ms |
| LiFi | 34 Mbps | 34 Mbps | | 3.30 ms |

The various results of the different radio access technologies, when measured individually compared to when MPTCP is running, show differences in throughput performance. Individual measurements showed better performance, compared to the same measurement while MPTCP is running. LiFi is the only exception, where the throughput performance results remain around the same average. However, when the throughput is aggregated over MPTCP, the result shows a significantly improved performance. Table 3-3 presents the average throughput measurements of the different radio access technologies in comparison when measured individually, with MPTCP in place and when aggregated with MPTCP. The average aggregated download throughput is around 1.45 Gbps. This value meets the KPI requirement of >1 Gbps throughput using multi-WAT aggregation.

Regarding latency, Table 3-3 presents the minimum latency for the three selected radio access technologies.

Wi-Fi-6 has the lowest latency performance measured at 1.42 ms, which is the closest latency to the predefined KPI of 1 ms.

### 3.2.2 Latency benchmarking

In this section, we evaluate the E2E latency achievable by the 5G-CLARITY multi-connectivity framework, which is related to the 5G-CLARITY Objective 3.2.

Figure 3-13 depicts a lab-based experimental setup that we have used at I2CAT to measure the achievable E2E latency. This is defined between the 5G-CLARITY CPE, which features a 5GNR and a Wi-Fi 5 interface, and the 5G-CLARITY AT3S user plane function deployed as a virtual network function within the edge compute server. The AT3S user plane functions in the edge compute and the 5G-CLARITY CPE are based on the MPTCP implementation in Linux kernel-5.4.132+.

We notice that LiFi is not included in the measurement setup, due to the unavailability of a pureLiFi AP at the time these experiments were carried out. However, during previous experimentation with the current LiFi product [2], we observed that the access delays are higher in LiFi than in Wi-Fi. Therefore, the current setup including only 5GNR and Wi-Fi is suitable to benchmark the minimum achievable latency.

In addition to benchmarking the latency achievable with the 5G-CLARITY multi-connectivity framework, a side goal of this study is to understand how 5GNR and Wi-Fi compare in terms of latency, and how does MPTCP react to the latency experienced in each access network. To this aim we design a specific experimental methodology that we describe in the next section.

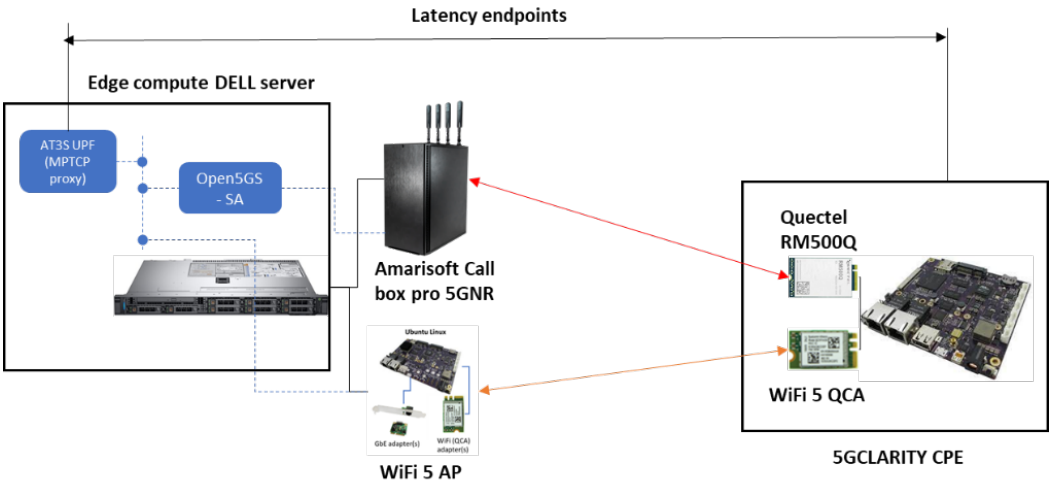



**Figure 3-13 Testbed setup**

#### 3.2.2.1 Experimental methodology

Looking at the experimental setup described in Figure 3-13 we can see that E2E latency is composed of:

- The access latency of the 5GNR or Wi-Fi networks, i.e the latency between the access modem and the gNB or AP
- The backhaul latency between the gNB and the virtual core network, based on open5gs, for the case of the 5GNR path
- The latency between the AP or the open5gs core network and the AT3S user plane function

The previous latency components are in turn affected by the following elements:

- The carrier bandwidth used in the access network, i.e., the higher the bandwidth the lower the

access network latency. Here we have that Wi-Fi can use carrier bandwidths of 20, 40, 80 and 160MHz, whereas 5GNR in the 3.5 GHz band can use carrier bandwidths between 20MHz and 100MHz. Our 5GNR SDR can however only sustain a maximum carrier bandwidth of 40MHz. Therefore, we choose for our experiment a carrier bandwidth of 40MHz for both Wi-Fi and 5GNR. This configuration will allow to have a fair comparison between Wi-Fi and 5GNR, although not resulting in the lowest possible latency.

- The scheduler used by MPTCP, which will determine how packets are sent through each access network. For our analysis we consider the three standard MPTCP schedulers, namely: i) the default scheduler, which should transmit packets always through the lowest latency path, ii) the redundant scheduler, which sends always packets through all paths in parallel, and iii) the round-robin scheduler, configured to send one packet through each path.
- The amount of interference presents in each access network. WiFi and 5GNR highly differ in the design of their channel access mechanisms, hence their latency characteristics are affected differently when multiple devices are present in the network [28]. For this purpose, we measure latency performance under the presence of different degrees of interference.

Figure 3-14 depicts our measurement methodology, which consists of:

- To measure latency, we use an open-source network diagnostics tool called lagscope [28], which provides a rich measurement output that allows to quickly analyse results in terms of histograms or CDFs.
- In addition to the 5G-CLARITY CPE connected to the Wi-Fi and the 5GNR networks, we consider an additional device in each network that acts as interferer. The interferer receives a data-stream in downlink, which we vary between 150 Mbps and 300 Mbps.

Finally, Table 3-4 describes the configuration of the two access networks in our experiment. To collect statistically meaningful results, we leave each delay measurement session running for 60 seconds and collect 10 independent samples for each configuration.

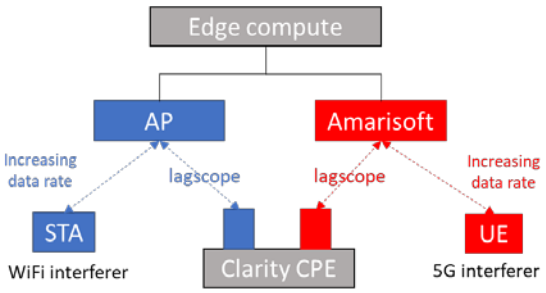



**Figure 3-14 Experimental methodology**

**Table 3-4 5GNR and Wi-Fi Access Network Configuration**

| | 5GNR | Wi-Fi 5 |
|---|---|---|
| **Channel/Band** | n78 | 149 (5745 MHz) |
| **Carrier Bandwidth** | 40MHz | 40MHz |
| **MIMO configuration** | DL: 2x2, UL: 1x1 | DL: 2x2, UL: 2x2 |
| **Subcarrier spacing** | 30 KHz | 312.5 KHz |
| **TDD pattern** | period: 2.5ms, dl_slots:3, ul_slots: 1 | N/A |

#### 3.2.2.2 Results analysis

Figure 3-15 depicts the empirical delay probability density functions (PDF) measured in our testbed, where each plot depicts the results of one MPTCP scheduler, i.e., default scheduler (left plot), redundant scheduler (middle plot) and round-robin scheduler (right plot). We benchmark each MPTCP scheduler under different data rates of the downlink stream received by the interferer, namely 150 Mbps (blue line), 200 Mbps (yellow line), 250 Mbps (green line) and 300 Mbps (red line). The same results are shown in a complementary way in the form of cumulative distribution functions (CDFs) in Figure 3-16, where this time the plots are split according to the interferer data rate and within each plot we depict all MPTCP schedulers.

In addition to the PDF plots, the legend in each figure indicates the percentage of lagscope packets from the 5G-CLARITY CPE that according to the MPTCP scheduler were transmitted through the Wi-Fi access network, where the remaining packets were transmitted through the 5GNR access network.

The following conclusions can be extracted from our results:

- As expected, all MPTCP schedulers see their latency performance degraded as the interferer data rate increases.
- The default and redundant scheduler achieve significantly lower latencies than the round-robin scheduler for all interferer conditions, which is seen more clearly in Figure 3-16. The reason is that under low interference conditions the Wi-Fi path achieves a lower latency than the 5GNR one. The default and redundant schedulers benefit from the low latency Wi-Fi, however the round-robin scheduler is forced to send 50% of the packets through the 5GNR path, even if the Wi-Fi path would be better.
- Wi-Fi achieves lower latency when the interference is low, but the trend reverts as the interference increases. We can see this clearly looking at the percentage of packets transmitted through the Wi-Fi path by the default scheduler, which should always take the lowest latency path based on socket-level measurements [29]. We see in the left plot of Figure 3-15 that when the interferer is 150 Mbps the default scheduler sends 99% of the packets through Wi-Fi. However, when the interferer grows to 300 Mbps only 4% are transmitted through Wi-Fi. We can validate how well the default scheduler is choosing the lowest latency path by looking at the redundant scheduler. The redundant scheduler sends each packet replicated through the two paths and delivers the first packet that arrives to the receiving application socket. Therefore, the redundant scheduler is guaranteed to have the lowest delay, at the cost of incurring a higher resource utilisation. Figure 3-16 shows how indeed the redundant scheduler always results in the lowest delay. To understand in detail the performance of the redundant scheduler, Figure 3-17 depicts for the redundant scheduler and for each packet the time difference between the 5GNR path and the Wi-Fi path and indicates in the legend the percentage of packets where the Wi-Fi path is faster. Interestingly, even in the case of a 300 Mbps interferer we find in the redundant scheduler that 64% of the times the Wi-Fi path is faster, meaning that the MPTCP default scheduler is too aggressive in choosing the 5G path (96% of the packets for the 300 Mbps interferer). Looking at the 300 Mbps interferer in Figure 3-16 we can see how the redundant scheduler is clearly better than the default one. Looking at these results we can see that there is still room for improvement in MPTCP scheduler design, defining a mechanism that can deliver the same latency than the redundant scheduler, without the resulting network overutilisation.

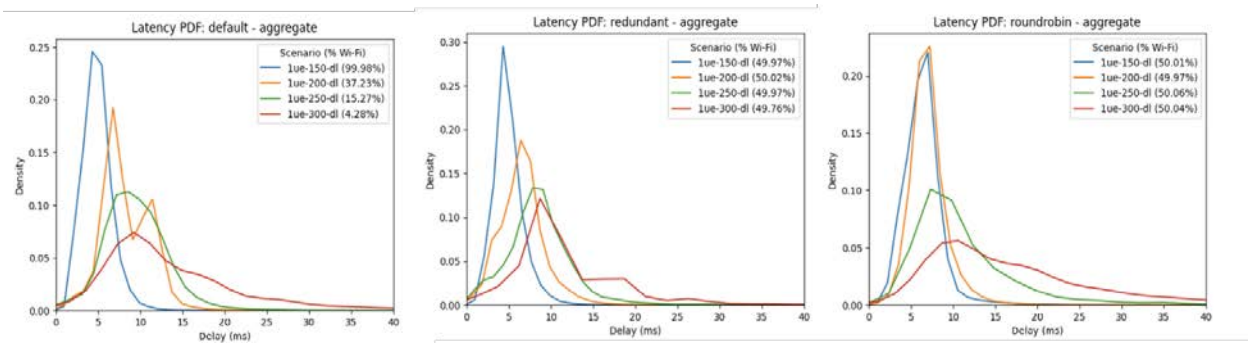



**Figure 3-15 RTT delay PDFs of 5GNR plus Wi-Fi latency testbed under different MPTCP schedulers**

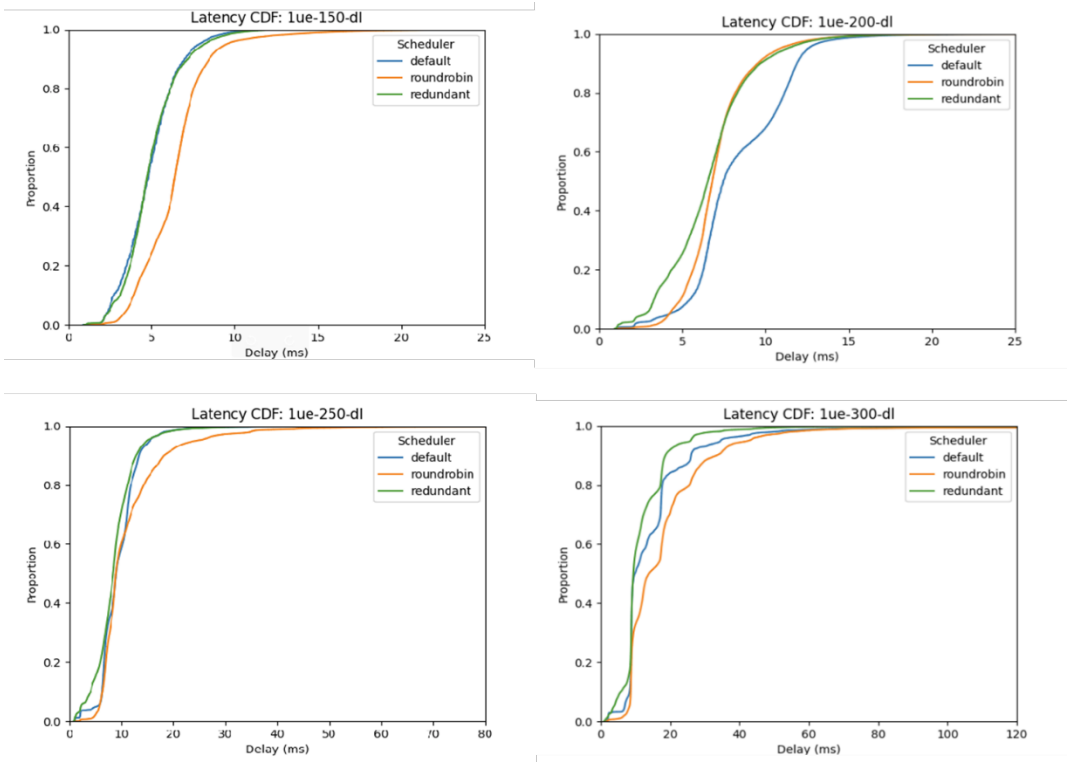



**Figure 3-16 RTT delay CDFs for different schedulers under different intereference situations**

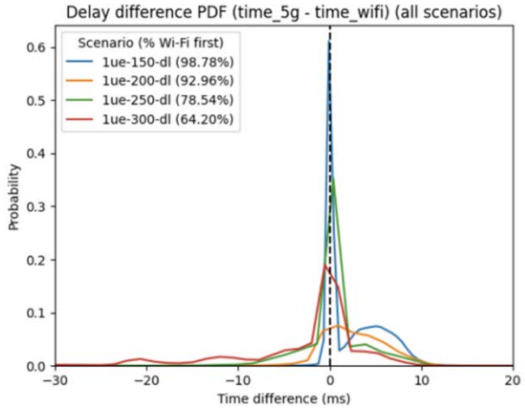



**Figure 3-17 Delay difference under redundant scheduler. A positive value indicated that the packet in the Wi-Fi path arrived first**

#### 3.2.2.3 Discussion on 5G-CLARITY O3.2

We discuss in this section the previous results considering the following two 5G-CLARITY requirements:

- *O3.2-ii: Air interface latency < 1ms for UL and DL through parallel access*
- *O3.2-iii: Air interface reliability > 0.999999 through smart interface selection*

Regarding the air interface reliability requirement, we conclude that it is achieved by design. The 5G-CLARITY multi-connectivity framework is based on MPTCP, which ensures reliable packet delivery.

The requirement on one-way delay 1 ms air interface latency was not generally achieved based on the previous latency measurements, but we provide the following considerations:

- Whereas the nominal requirement referred to air interface, the results in the previous sections refer to the E2E connection, including the AT3S user plane function in the edge cluster. We have decided to adopt this setup as benchmark, because it is more meaningful in the sense of the multi-connectivity framework design adopted in 5G-CLARITY.
- Under low interference conditions we see E2E delays below 5 ms, mostly in the case of Wi-Fi. This value is close to the target KPI, considering that 1 ms referred only to one-way delay and that the 5 ms also include the latency to the AT3S VNF in the edge cluster. However, such low latencies have not been possible in our setup using 5GNR. Factors that can contribute to enhance access network latency include:
    - Using higher carrier bandwidth, i.e., 160 MHz in Wi-Fi and 100 MHz in 5GNR. Recall this was not possible in our testbed due to limitations of the Amarisoft SDR.
    - For 5GNR, using a subcarrier spacing above 30 KHz, which has not been tested in the project due to limited equipment availability.
- Therefore, we conclude that the 5G-CLARITY multi-connectivity architecture is suitable to achieve the target low latency KPI, once more mature 5GNR equipment becomes available in the market.

### 3.2.3 eAT3S mobility benchmarking

This subsection is devoted to evaluating *Objective O3.2 iv) Supporting vertical handover between wireless technologies with handover times < 5 ms.* Given that performing mobility experiments in a laboratory environment is complex, we decided to use a virtual testbed based on the open-source implementation of the 5G-CLARITY multi-connectivity testbed.

Since 5G-CLARITY's multi-WAT connectivity is supported by using MPTCP, which sends data through the different available network interfaces, there is technically no handover between technologies. Instead, when a link is enabled, MPTCP will detect it as an available network interface, and will consider it in its packet scheduling. Therefore, the assessment of "*O3.2 iv)*" will consist in measuring the elapsed time since an interface transition from disabled to enabled and MPTCP starts sending data through the new interface. This time will be called the transmission resume delay. Notice therefore that technically our original objective "*O3.2 iv)*" is achieved by design, because there is no break-before-make handover in the 5G-CLARITY multi-connectivity framework, instead we can understand that MPTCP enables a sort of soft-handover where new access networks can be incorporated to a running connection as soon as they become available. Our goal thus is to quantify the time required to benefit from a new access network, once it becomes available.

#### 3.2.3.1 Experimental setup

To perform the experiments under a controlled environment, we have configured two machines virtualized with VirtualBox version 6.1 and connected to two virtual networks via two network interfaces. Each machine runs an Ubuntu 20.4 image with the modified kernel which includes 5G-CLARITY MPTCP module.

Test traffic is generated using the `iperf` tool. Traffic is sent from one virtual machine (VM1) to the other virtual machine (VM2). VM1 runs a script called `detect` which detects when the network interfaces become enabled or disabled. To this end, this script uses NETLink sockets. This way, as soon as an interface changes, the event is captured and notified.

VM1 also runs the tool `tcpdump` to capture traffic from the interface that will be enabled and disabled. Since `detect-interfaces` and `tcpdump` run in the same machine, they both use the same clock. This way, the timestamp of the event of a new available interface generated by `detect-interfaces` and the one from the tcpdump network capture can be used to determine the transmission resume delay. The delay is estimated by an AWK script which processes the output of `detect-interfaces` and `tcpdump`.

To simulate the outage of a link, one of the network interfaces of VM1 is disabled and enabled from the VirtualBox user interface several times during the test. A total 100 temporal outages were simulated.

#### 3.2.3.2 Experimental results

As it can be seen in Figure 3-18, more than the 60% of the outages resume the transmission after 0.065s since the interface is reenabled. Two ARP resolution requests are made before sending a TCP segment, one for each destination IP address. Nevertheless, this ARP resolutions lasts 0.6ms in the worst case.

The results of these experiments demonstrate that the MPTCP approach for the multi-WAT connection allows to provide continuous transmission even when a link fails. This “always-on” type of connection replaces the need for a vertical handover across access networks. Therefore, the time with no connectivity due to a handover is zero.

From the experiments, we have measured that the time to resume the transmission through a new link, is nearly 60ms. To identify the source of this delay, we have conducted several experiments. To start with, we forged Ethernet frames with the source and destination MAC addresses of some of the MPTCP packets captured during the previous experiments. The payload of these frames was randomly created. We created a Python script which sends these frames continuously. The time to resume was measured as in previous experiments. The resulting measurements shown that the average delay was near to 5ms. Additionally, the captured traffic shown that a Multicast Listener Report Message v2 packet is sent just 2 ms after the interface is available.

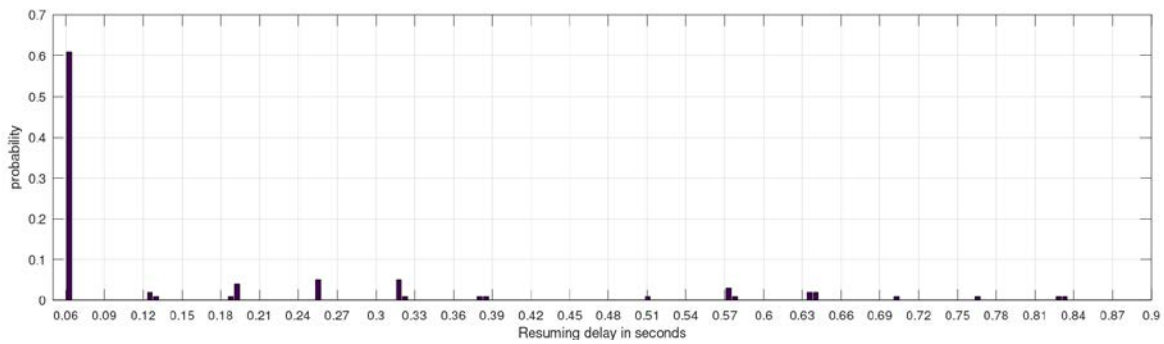



**Figure 3-18 Histogram of transmission resume delay**

In the following experiments we used the iperf tool to generate UDP traffic. In this case, the average delay was equal or higher than the one of the TCP cases. Our hypothesis from these results is that this delay is related to the mechanism of notifying transport layer sockets that an interface can be used for transmission. This mechanism would be based on state polling with a sampling period.

To validate our hypothesis, we analysed the MPTCP and sockets' related code to find they are notified of lower network layers change of state. However, there are some references to a `SOCK_DEAD` state and `sock_wake_async` functions in sections of code devoted to enable or disable sending data, we have not found the polling mechanism which may trigger these changes. As future work we will check lower levels of the network modules of the kernel, to identify the notification mechanism, as well as the polling period time to see if our measured “transmission resume time” can be reduced.

### 3.2.4 Benchmarking E2E TCP connection over 5G-CLARITY multi-connectivity framework

Section 3.2.1 benchmarks the throughput achievable by the 5G-CLARITY multi-connectivity framework. However, a drawback of our current implementation is that it requires MPTCP support in the endpoint (the device). Unfortunately, MPTCP is not yet integrated upstream in the Linux kernel, although there are discussions on this direction [30] Therefore, a custom kernel needs to be deployed in the device if we want to support 5G-CLARITY multi-connectivity. To overcome this issue, we propose and evaluate an alternative in this section, which is to use the 5G-CLARITY CPE as gateway to connect end-devices to the 5G-CLARITY network, where end devices would remain unmodified.

This problem was addressed in the virtual testbed used in Section 3.1 and in 5G-CLARITY D3.2 [2] using an `openvpn` tunnel between the CPE and the MPTCP proxy, which allows to transport any traffic between endpoints that are connected to the 5G-CLARITY network at layer two. The tests using the virtual testbed showed a proper functioning and high throughput values, however initial tests using real equipment showed a poor performance, achieving only a few Mbps, which is not acceptable for the 5G-CLARITY demonstrators of WP5. This poor performance is because the TCP connections are transmitted over an OpenVPN connection, leading to a well-known TCP-over-TCP problem (see Figure 3-19).

To overcome this problem, we have tested several alternative solutions. The target solution shall allow to create a TCP connection between the CPE and one proxy, which will employ MPTCP so that several network interfaces with different radio technologies can be used simultaneously. A tunneling solution is required since data is sent between the client and the server, which do not require to implement MPTCP. This data may use TCP or UDP, will be transported over the MPTCP connection between the CPE and the proxy. In order to avoid the TCP-over-TCP problem, we have identified and tested the following options:

- An SSH tunnel using SSHuttle [31] assembles the TCP stream locally, multiplexes it statefully over an SSH session, and disassembles it back into packets at the other end. Thus, it never ends up doing TCP-over-TCP. SSHuttle may employ NAT or transparent proxy (tproxy). The NAT method only supports TCP, whereas tproxy supports both TCP and UDP.
- A SOCKS5 proxy, ShadowSocks [32]. ShadowSocks is a lightweight secured SOCKS5 proxy for embedded devices and low-end boxes. To employ ShadowSocks, a tool is required to create a network interface so that we can continue using our virtual testbed with minor modifications. For that purpose, we have found two possible options:


**Figure 3-19 TCP (between clients and server) over TCP (MPTCP tunnel between CPE and proxy)**

- *tun2socks* [33] , which "socksifes" TCP connections at the network layer. It implements a TUN device which accepts all incoming TCP connections (regardless of destination IP) and forwards the connections through a SOCKS server. This allows you to forward all connections through SOCKS, without any need for application support.
- *Ip2socks* [34], similar to *tun2socks* but with *tun* (L3 device) and *tap* (L2 device) mode.

Table 3-5 presents the performance evaluation considering throughput (for TCP and UDP, where applicable, with one and several concurrent flows) and latency (using ICMP ping if supported over the MPTCP connection or *tcpping* [35] otherwise).

**Table 3-5 Performance of the Virtual Testbeds**

| | Max. thr. TCP 1 flow | Max. thr. TCP 10 flows | Max. thr. UDP 1 flow | Max. thr. UDP 10 flows | Min. latency (ICMP) | Min. latency (tcpping) |
|---|---|---|---|---|---|---|
| **No MPTCP, connection between client and server** | 881 Mbps | 790 Mbps | 501 Mbps | 454 Mbps | 1.16 ms | 0.998 ms |
| **No MPTCP, connection between CPE and proxy** | 515 Mbps (maybe duplicated packets?) | 447 Mbps (maybe duplicated packets?) | 906 Mbps | 867 Mbps | 0.464 ms | 0.410 ms |
| **Between CPE and proxy[5]** | 288 Mbps | 296 Mbps | 914 Mbps (no MPTCP) | 852 Mbps (no MPTCP) | 0.622 ms | 0.486 ms |
| **OpenVPN with OVS (proxy router)[6]** | 55.9 Mbps | 51.9 Mbps | 51.8 Mbps | 59.7 Mbps | 1.42 ms | 1.378 ms |
| **SSHuttle with NAT (no cipher)** | 64.3 Mbps | 56.9 Mbps | - | - | - | 0.442 ms |
| **SSHuttle with *tproxy* (no cipher)** | 75.2 Mbps | 274 Mbps | 40.7 Mbps | 41.4 Mbps | - | 0.406 ms |

[5] In this scenario, the test is performed between the CPE and the proxy (although the 4 VMs are running) because the client and the server do not support MPTCP and there is no tunneling between CPE and proxy to forward packets. Thus, the *iperf* connections have to be established between CPE and proxy in order to use MPTCP. Thus, there are less bridges 5(executed by VirtualBox) processing the packets (*client_cpe* and *proxies_server* bridges).

[6] With namespaces and bridges (*brctl* and *OVS*), which may also affect the performance.

| | | | | | | |
|---|---|---|---|---|---|---|
| **SSHuttle with NAT (default cipher)** | 62.6 Mbps | 57.2 Mbps | - | - | - | 0.453 ms |
| **SSHuttle with *tproxy* (default cipher)** | 70.1 Mbps | 270 Mbps | 43.6 Mbps | 43.5 Mbps | - | 0.436 ms |
| **ShadowSocks with tun2socks** | 90.2 Mbps | 102 Mbps | 117 Mbps | 123 Mbps | - | 0.429 ms |
| **ShadowSocks with ip2socks (tun interface)** | 110 Mbps | 120 Mbps | - | - | 0.465 ms | 0.415 ms |
| **ShadowSocks with ip2socks (tap interface)** | 111 Mbps | 118 Mbps | - | - | 0.553 ms | 0.466 ms |

As shown in Table 3-5 , SSHuttle with tproxy stands out as a possible solution for the testbeds with real equipment, since it avoids the TCP-over-TCP problem and achieves a performance slightly lower to that of OpenVPN in the virtual testbed. In addition, it allows using both TCP and UDP between the client and the server.

Additionally, we have analysed Wireshark traces for these options. Traces have been captured on the selected proxy and several parallel connections have been generated between the client and the server using `iperf`. The following summarizes the conclusions from this analysis:

- SSHuttle using NAT (only TCP)
  - Proxy always listens on TCP port 22 (SSH server) for all the paths.
  - CPE listens on three different TCP ports, one for each path (SSH client).
  - This means that there is **only one MPTCP connection**.
- SSHuttle with *tproxy*
  - Using TCP: same conclusions as SSHuttle with NAT
  - Using UDP: UDP over TCP, using the same ports than the test with TCP
- ShadowSocks with tun2socks and ip2socks, tested with TCP
  - Proxy always listens on TCP port 8388 (SOCKS5 server) for all the paths.
  - CPE listens on a large amount of TCP ports (30 TCP connections for 3 paths x 10 parallel connections)
  - This means that there are **many MPTCP connections** (10, one per parallel connection) **established between CPE and proxy**.
- ShadowSocks with tun2socks with UDP
  - Proxy always listens on TCP port 8388 (SOCKS5 server) for all the paths.
  - CPE listens on a large amount of UDP ports.
  - UDP is used, i.e. UDP over TCP. This means that **MPTCP is not being used**. Despite this fact, it could be interesting for cases where UDP may be needed (e.g. for DNS).

## 3.3 5G-CLARITY eAT3S framework

A key component of the 5G-CLARITY multi-connectivity framework is the eAT3S policy. This is an intelligent function that controls for each 5G-CLARITY enabled device how many packets can be transmitted through each available access network.

This section describes different eAT3S algorithms and provides simulation/testbed-based validations of the proposed eAT3S algorithms. The algorithms provided in this deliverable are either enhanced or different from the ones provided in 5G-CLARITY D3.2 [2], 5G-CLARITY D4.2 [36] and 5G-CLARITY D4.3 [37]. It is worth to note that in WP4, the algorithms proposed for traffic steering are ML-based algorithms, whereas in WP3 the algorithms are either based on thresholding or estimating available throughput/scheduler weights. Below is a summary of the proposed algorithms in the previous deliverables and those provided in this deliverable:

- 5G-CLARITY utility-based eAT3S algorithm (5G-CLARITY D3.2/D3.3) described in Section 3.3.1:
    - In 5G-CLARITY D3.2, the SINR is used as the main parameter to decide on the weight factor assignment for load-balancing steering mode.
    - In 5G-CLARITY D3.3, an update to the proposed eAT3S algorithm in 5G-CLARITY D3.2 [2] is proposed. In Section 3.3.1 in this deliverable, other parameters, including mainly the signal strength, gNB/AP buffer status/resource utilization, RTT/transmission delay measurements are considered by the weight factor assignment decision mechanism. This manages not only eMBB traffic but also URLLC traffic.
- RT-RIC: AT3S traffic routing/handover (5G-CLARITY D4.2/D4.3):
    - In 5G-CLARITY D4.2 [36], a model-based predictor is described, which is used to predict user position. By using the terminology from the RL algorithm, the model-based predictor receives a state at t from the environment and predicts an output of a function of position at t+1. The predicted position information is planned to be utilized to determine the RSSI of LiFi interface at the next time step. Then, an RL agent could anticipate accordingly how to steer the traffic to other wireless access interfaces with a specific focus on the LiFi network.
    - In 5G-CLARITY D4.3 [37], evaluations of the model-based as well as model-free RL algorithms are provided. A set of observations including emulated snapshots of images from CCTVs, received signal strength of all the UE interfaces and congestion window size from all MPTCP sub-flows is collected and used by DRL agent.
- 5G-CLARITY eAT3S bandwidth prediction algorithm in Section 3.3.2 (D3.3):
    - In D3.3, this algorithm is proposed to select the round robin scheduler weights to maximize the aggregated throughput. This allows to overcome the throughput limitation that the MPTCP experiences due to the slowest link. A strategy to estimate the link capacity by means of active probing is described. For selecting the weights for each interface to get the highest throughput, the estimated capacity of the network links is used to determine the scheduler weights for each access technology.
- 5G-CLARITY eAT3S lower-level user plane function in Section 3.3.3 (D3.3):
    - This subsection introduces the proposed 5G-CLARITY Low-Layer eAT3S approach, which allows flexible multi-connectivity through its integration with the Optimal Access Network framework to provide multi-transport protocol scheduling and low-layer access control and traffic steering. The system implements the multi-layer ATSSS rules and Multi-Access rules (MAR) obtained by the optimal access network framework while scheduling one or multiple

transport protocols, single path (e.g., UDP, TCP, QUIC) and multi-path (e.g., MPTCP and MPQUIC [38] [39]).

### 3.3.1 5G-CLARITY utility-based eAT3S algorithm

This section provides an update on the proposed eAT3S algorithm in 5G-CLARITY D3.2 [2]. In 5G-CLARITY D3.2, it is reported that the SINR is used as the main parameter to decide on the weight factor assignment in the load-balancing steering mode. In 5G-CLARITY D3.3, more parameters are considered by the eAT3S algorithm described in 5G-CLARITY D3.2 [2], such as signal strength, gNB/AP buffer status/resource utilization, RTT/transmission delay measurements. These are included in the weight factor assignment decision mechanisms. Also, not only eMBB traffic but also URLLC traffic is considered in the updated algorithm. Accordingly, based on the user requested traffic type, the algorithm decides the steering mode to be used, where different threshold levels are used to identify the satisfaction of various parameters. More specifically, if the eMBB traffic is requested, the load-balancing steering mode will be chosen with a set of specific threshold weights for signal strength, gNB/AP buffer status, SINR and RTT/transmission delay, to decide the assigned load level on 3GPP and non-3GPP access networks. In case, the URLLC traffic is requested, smallest-delay or redundant steering mode will be applied with a specific set of threshold weights for the same parameters to decide what WATs out of all available WATs should be used to duplicate the traffic. Figure 3-20 depicts a flow for the updated eAT3S algorithm for eMBB and URLLC traffic types.

To enable a random spatial traffic distribution among the deployed APs/gNBs to mimic more realistic network deployment, different content sizes, different popularities among available contents and Poisson arrival process for UE requests are considered.

Regarding the content popularity, the popularity within the library of $N_F$ files is characterized as a Zipf distribution [40] [41]. Therefore, the probability $P_F$ of a file $f \in \{1, \ldots, N_F\}$ being requested can be written as:

$$P_F(f) = Z f^{-\alpha_z} \qquad (3\text{-}1)$$

where $\alpha_z$ is the Zipf parameter that characterizes the probability of content reuse; and $Z = 1/\sum_{i=1}^{N_F} i^{-\alpha_z}$ is a constant for a given number of files in the library and $\alpha_z$. When $\alpha_z$ is chosen as a value larger than zero, the most popular contents have a lot higher chance to be requested. The content popularity distribution is only used to decide what content should be considered for a given user request.

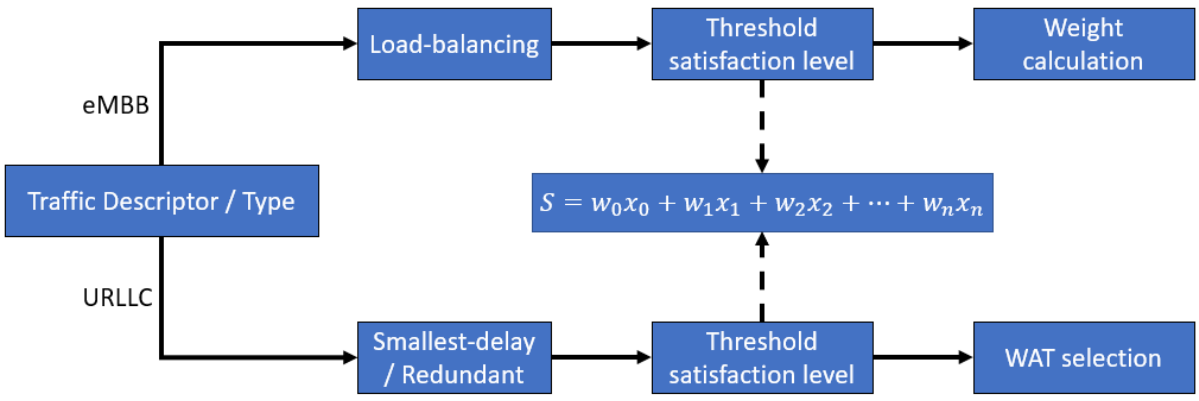



**Figure 3-20 Updated eAT3S algorithm to consider different traffic types**

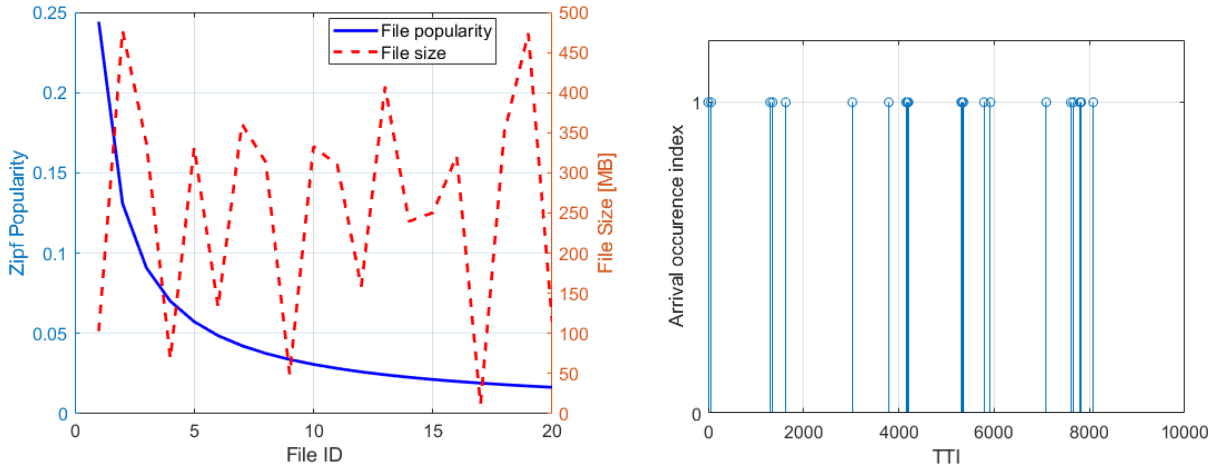



**Figure 3-21 Requested file popularity distribution and file size (left) and arrival time of user requests (right) based on Poisson arrival process (*r*=16, *T*=10s)**

In addition to content popularity distribution, a random distribution is considered for the file size. This is independent from the content popularity. Each file f in the library of $N_F$ files has a randomly allocated file size as shown in Figure 3-21 (left).

The UE request is modeled as a Poisson arrival process where the sequence of inter-arrival times for UE requests are independent and identically distributed (i.i.d) random variables. The Poisson distribution probability mass function $P_R$ of a *r* number of requests/events in a given time interval *T* is given as:

$$\boldsymbol{P_R(\mathrm{r}) = \left((\lambda \mathrm{T})^{\mathrm{r}} \mathrm{e}^{-\lambda \mathrm{T}}\right)/\mathrm{r}!} \qquad \textbf{(3-2)}$$

where $\lambda$ is the Poisson distribution parameter, which is also known as the average number of requests/events during a unit of time; and (.)! is the factorial operation. In the considered system model, a randomly selected UE is activated in each Poisson-based request arrival event, and a content *f* is requested based on the considered Zipf content popularity distribution $P_F$. Figure 3-21 (right) illustrates arrival time of user requests based on the described Poisson arrival process with a number of requests *r*=16 and a time interval *T*=10s.

Figure 3-22 shows an instance from the system level simulator that mimics the factory environment defined in D5.1 [42]. In the simulation scenario, an AGV that follows the path shown in red in the figure and 100 other users randomly distributed in the considered factory environment exist. The randomly distributed 100 users are considered as static users that include factory workers, connected production machinery and sensors. Along the AGV's path, each access technology has a different level of coverage and provides different level of signals from various access nodes. In the simulation environment, there are six 5G NR gNBs (shown by green circles) deployed based on factory scenario defined in TR 38.901 [43], five Wi-Fi APs (shown by blue circles), and LiFi APs (shown by yellow circles along the AGV's path) deployed with a inter-site distance of 3 meters along the AGV's predefined path. While the AGV moves along the path, the system level simulator provides telemetry readings such as AGV's connected access node, its received signal power, SINR, access node's buffer status for the downlink traffic and UE-specific transmission delay measurements from all three WATs. The proposed eAT3S algorithm makes use of the available telemetry data and depending on the UE's requested traffic type, firstly decides on the scheduler type as load-balancing or smallest delay/reduntant, then allocates weights onto 3GPP and non-3GPP access networks. The used scheduling type and allocated weight information is also captured in Figure 3-22 below. During the simulation, a different parameter/steering scoring weight, $w_i$, for each available telemetry data is considered for load-balancing and reduntant schedulers. For the load-balancing scheduler, the steering

score weights of SINR ($w_{SINR}$), gNB/AP buffer status ($w_{Buffer}$) and transmission delay ($w_{Delay}$) are set to 1, 0.7 and 0.2, respectively. For the reduntant scheduler, $w_{SINR}$, $w_{Buffer}$ and $w_{Delay}$ are set to 0.5, 0.2 and 1, respectively. Therefore, it can be expected that for the same telemetry readings for the considered parameters, the scheduling metric/threshold level for different schedulers will be different. For example, for the given system level simulator instance in the figure below, the scheduling metric of the load-balancing scheduler can be different from the redundant scheduler. In the considered instance, Wi-Fi access is not considered for load balancing due to mainly its buffer status. However, it will be considered for smallest-delay/redundant scheduler as it has the minimum delay performance among the available WATs. Hence, weight allocation and WAT selection can differ for each steering mode, depending on the operator's preference on the steering scoring weights as well as threshold value to consider any of the access technology as a candidate for steering decisions.

Figure 3-23 shows a comparison of 5G SINR, weight assignment and steering score for load balancing steering mode for the whole journey of the AGV when a steering threshold value of 0.6 is used. The AGV's traffic is considered as eMBB. The figure shows that the weight assignment of the load balancing steering mode mostly follows the SINR performance as $w_{SINR}$ has the heights weight in the steering scoring. The figure also shows that on some occasions, having a good SINR performance is not enough to steer traffic onto the WAT. For example, for a period from 3000s to 3500s, although the SINR changes from 0 dB to 10 dB, the steering score is below the threshold. Hence, the load balancing scheduler does not steer any traffic onto 5G during that period. This is also the case for other WATs.

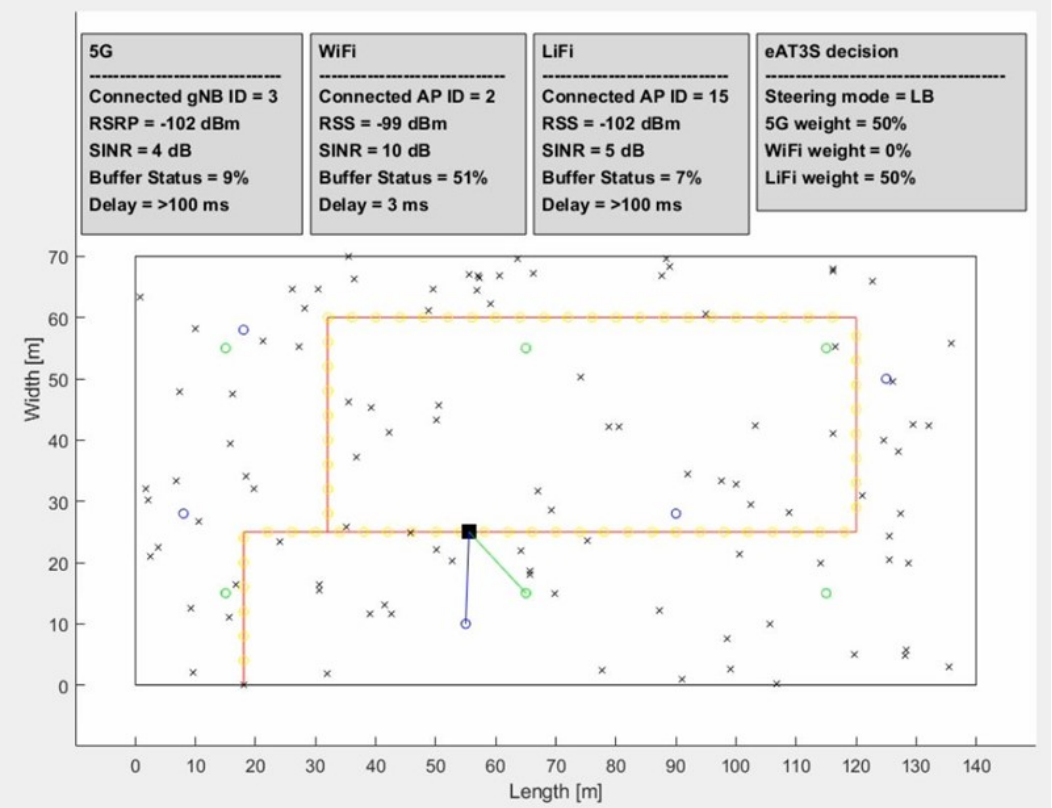



**Figure 3-22 A snapshot from the simulator to depict dynamic change of telemetry data and eAT3S scheduling decisions**

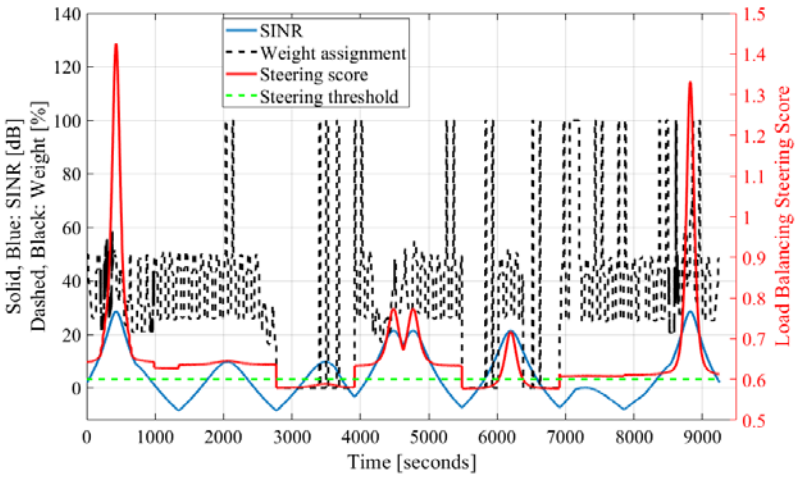



**Figure 3-23 Comparison of 5G SINR, weight assignment and steering score for load balancing steering mode when a steering threshold value of 0.6 is used**

Table 3-6 compares the decisioning of both schedulers for the same AGV position. It is worth to note that, although the radio conditions/telemetry is same for different simulation runs with different traffic type requested by AGV (eMBB for load balancing and URLLC for smallest delay/redundant), the buffer status and transmission delay performance are different. This is due to the Poisson-based user request arrival as well as popularity-based content selection. In each run, the user requests and selected content are generated based on the described procedures. In the simulation time of 00:59, the load balancing scheduler divides the access traffic to two and assigns it to 5G and LiFi. However, at the same simulation time, at the same AGV position, the smallest delay/redundant scheduler only selects Wi-Fi to transmit URLLC traffic as it achieves a transmission delay of 10ms whereas 5G and LiFi achieve 89ms and >100ms, respectively. Moreover, although Wi-Fi achieves a higher SINR performance compared to 5G and LiFi for the same AGV position, its buffer status degrades the steering score of Wi-Fi and the load balancing scheduler discards Wi-Fi in the weight assignment decisioning. The similar conclusions can be derived for the other considered simulation times/AGV positions.

**Table 3-6 Comparison of Decisioning of Load Balancing and Smallest Delay/Redundant Schedulers for the Same AGV Position**

| Simulation time | Telemetry | Load balancing $w_{SINR}$:1 $w_{Buffer}$: 0.7 $w_{Delay}$: 0.2 | | | Smallest Delay/Redundant $w_{SINR}$:0.5 $w_{Buffer}$: 0.2 $w_{Delay}$: 1 | | |
|---|---|---|---|---|---|---|---|
| | | 5G | Wi-Fi | LiFi | 5G | Wi-Fi | LiFi |
| 00:14 | Connected node ID | 1 | 1 | 4 | 1 | 1 | 4 |
| | RSRP/RSS [dBm] | -77 | -101 | -100 | -77 | -101 | -100 |
| | SINR [dB] | 29 | 9 | 9 | 29 | 9 | 9 |
| | Buffer Status [%] | 8 | 27 | 8 | 17 | 64 | 8 |
| | Delay [ms] | 14 | 6 | 2 | 18 | 9 | 2 |
| | Weight [%] / Selection | 49 | 22 | 29 | 1 | 1 | 1 |
| 00:59 | Connected node ID | 3 | 2 | 15 | 3 | 2 | 15 |
| | RSRP/RSS [dBm] | -102 | -99 | -102 | -102 | -99 | -102 |
| | SINR [dB] | 4 | 10 | 5 | 4 | 10 | 5 |
| | Buffer Status [%] | 9 | 51 | 7 | 14 | 65 | 8 |
| | Delay [ms] | >100 | 3 | >100 | 89 | 10 | >100 |
| | Weight [%] / Selection | 50 | 0 | 50 | 0 | 1 | 0 |

| Simulation time | Telemetry | Load balancing $w_{SINR}$:1 $w_{Buffer}$: 0.7 $w_{Delay}$: 0.2 | | | Smallest Delay/Redundant $w_{SINR}$:0.5 $w_{Buffer}$: 0.2 $w_{Delay}$: 1 | | |
|---|---|---|---|---|---|---|---|
| | | 5G | Wi-Fi | LiFi | 5G | Wi-Fi | LiFi |
| 02:24 | Connected node ID | 6 | 4 | 37 | 6 | 4 | 37 |
| | RSRP/RSS [dBm] | -90 | -79 | -100 | -90 | -79 | -100 |
| | SINR [dB] | 16 | 30 | 10 | 16 | 30 | 10 |
| | Buffer Status [%] | 10 | 68 | 10 | 21 | 100 | 7 |
| | Delay [ms] | 41 | 3 | 3 | 26 | 4 | 2 |
| | Weight [%] / Selection | 23 | 48 | 28 | 0 | 1 | 1 |

In this section, it is shown that the updated eAT3S algorithm enables the inclusion of multiple network parameters in the steering decisioning. It would enable private/public network operators to derive user or environment specific network policy configurations to efficiently utilize 3GPP and non-3GPP networks. The structure of the updated eAT3S algorithm also enables an inclusion of AI-based approaches that predict the performance of different network parameters and proactively update the steering weights/WAT selections to further improve the user QoS/QoE.

### 3.3.2 5G-CLARITY eAT3S bandwidth prediction algorithm validation

A problem that has been detected during the benchmarking of the 5G-CLARITY multi-connectivity framework in Section 3.2, is that when using MPTCP with a round robin scheduler, the slowest link limits the throughput of the rest. This limitation can be overcome by using the weighted round-robin MPTCP scheduler developed in 5G-CLARITY if appropriate weights are allocated to each access network, according to its bandwidth. A problem though is how to get a reliable estimation of the throughput achievable through a wireless access network, such as 5GNR, Wi-Fi and LiFi, where achievable throughput depends on the current occupancy of the cell, or the channel conditions. This section describes an algorithm to address this problem.

#### 3.3.2.1 Estimation of the available throughput

To calculate the weights of the weighted round robin scheduler, the estimation of the available throughput in each link is required. This estimation should be dynamic, since network conditions can change during an MPTCP session, especially for a mobile device.

In this section we propose and test a strategy to estimate the link capacity by means of active probing. The main idea behind this approach is to establish MPTCP connections during periods of the session, sending at maximum throttle, and measuring the data sent by each interface. These measurements include both the data generated by the probing agent, as well as all the other data sent through the MPTCP enabled network interfaces.

To this end, we have programmed a monitoring agent that is launched at the 5G-CLARITY CPE side (c.f. Section 3.1). The agent records which scheduler is configured and sets the MPTCP scheduler to “redundant”. This way, applications started from this moment will try to send all the data via the different interfaces. MPTCP sockets opened before this moment will keep their MPTCP scheduler. The probing agent operates in the following way:

- The agent periodically starts “`iperf -c <destination-ip> -t <probing-duration>`”,

where <destination-ip> is the address of the proxy, in this experiment, and <probing-duration> is the time `iperf` will be sending data at full throttling through the interfaces. <probing-duration> is set to 10 seconds in this experiment.

- Before and after starting the `iperf` operation, the statistics of bytes sent by each interface are collected. This information is obtained as a JSON document by "`ip -j -s link show`".
- After that, the agent resets the MPTCP scheduler, and waits for the next probing period.

This scheme affects the operation of the applications running on the device. The probing period and duration should be selected/adapted depending on the application requirements and expected network behaviour (e.g., the variability of the available throughput). The reading of the bytes transmitted through each interface includes every protocol, so it may reflect a more accurate estimation of the real throughput.

To test the performance of the monitoring agent, two VirtualBox virtual machines (VM) are configured with two interfaces each. Each interface is connected to a virtual link, so an interface from a VM is connected to an interface of the other VM. A set of 470 different link capacities are configured by using the network traffic control tool "`tc`". The distribution of the link capacities used in this experiment are shown in Figure 3-24a.

After the executions of the monitoring agent, estimated throughputs available for each of the two interfaces are compared to the real throughput, which is the bit rate configured by means of the "tc" tool. Figure 3-24b depicts the resulting absolute difference between the estimated values and the real ones. As it can observe, the errors grow proportionally to the capacity of the link.

Nevertheless, the relative errors, calculated as the absolute difference of the estimated and real capacity, normalized by the real capacity, are below 0.003 points. The cumulative distribution function of the relative error of estimation for the experiments carried out is shown in Figure 3-25.

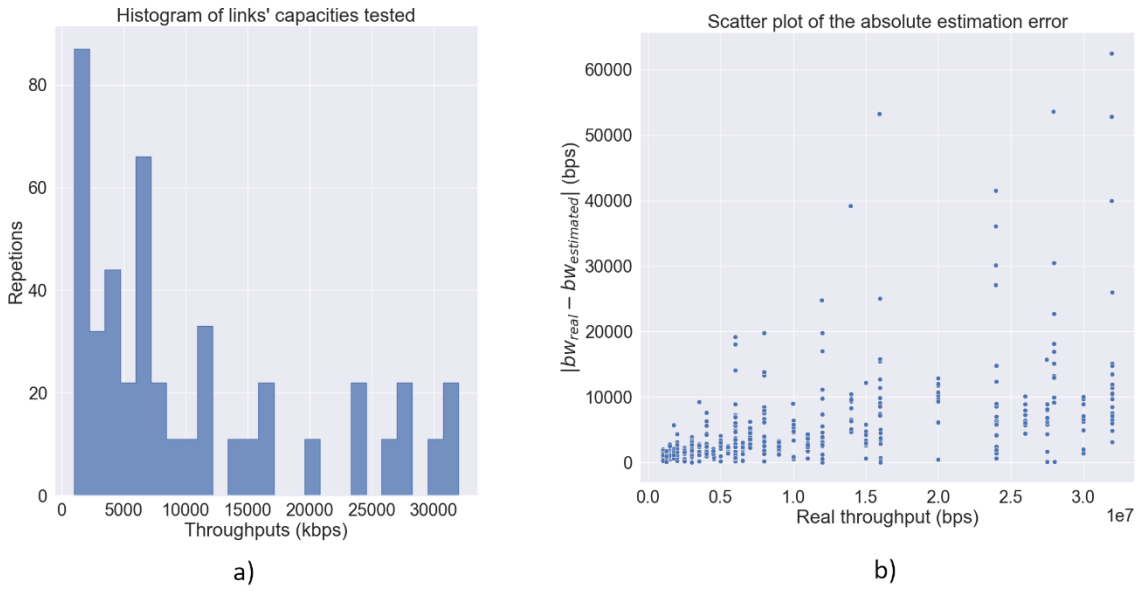



**Figure 3-24 a) Link capacity histogram used for the estimation of the available throughput experiment, b) absolute error between the available throughput estimation and the real capacity of a link**


**Figure 3-25 Relative estimation error CDF**

From the previous results, we can argue that the proposed method can be used for experimentally assess the available throughput of the MPTCP enabled interfaces with a negligible relative error. Since this is an active probing procedure, it may interfere with the performance of other running applications which use the multi-WAT connections during the probing period. During that testing period, the links will share their capacity between the traffic of the MPTCP active sessions.

For this reason, future work will include a procedure to correlate the telemetry of the network links and the MPTCP sockets to estimate the available bit rate.

### 3.3.2.2 MPTCP weights estimation

From the experimentation carried out with the round robin scheduler of the MPTCP sessions we can conclude that the sending throughput achieved in one link may be limited by the configuration of other links: the duration of a round of the round robin scheduler is the largest of each link's transmission time, and the MPTCP session needs to reorder the packets at the receiver before delivering them to the application socket. Therefore, the allocation of the weight of each links should be carefully selected for the case in which we need to achieve the maximum aggregated throughput.

In this section we propose an algorithm to select the weights that maximize the aggregated throughput, given the capacity of the network links. To do this, we consider the ratio between the achieved throughput versus the link capacity as the parameter to maximize. This ratio is calculated as "bw_aggregated" (the maximum achievable throughput given a set of weights) divided by the sum of the real capacity of each link, labelled as "bw_real". To get a general picture of the problem to solve, let us use the example of Figure 3-26. After running a simple simulator which calculates the maximum achievable aggregated throughput of 3 links with a weighted round robin scheduler, the corresponding ratio for each triplet of weights is calculated. In this figure, we represent the weights that obtain ratios higher than 0.96. As it can be seen, there is a space of possible solutions which follow a pattern for this case.

Given a set of available bitrates corresponding to the capacity of several links, we aim at providing an algorithm which maximizes the resulting aggregated throughput. To that end, a set of integer weight values, i.e., turns within a scheduler round, should be provided. A simple approach would be to set a minimum weight 1 to the lower link capacity, and to set the other weights related to this allocation. This leads to 3 related naïve algorithms:

- Ceil criterion: for link_i, where i ranges from 1 to the number of links, the weights are calculated as the upper bound integer resulting after dividing the capacity of link_j by the capacity of the lowest

link.

- Floor criterion: for link_i, where i ranges from 1 to the number of links, the weights are calculated as the lower bound integer resulting after dividing the capacity of link_j by the capacity of the lowest link.
- Round criterion: for link_i, where i ranges from 1 to the number of links, the weights are calculated as the nearest bound integer resulting after dividing the capacity of link_j by the capacity of the lowest link.

However, these algorithms do not perform well in all the cases. The truncation operation leads to limit even the slowest links since the proportionality of the turns and the real capacity are not contemplated. Hence, we propose another algorithm called "GCD criterion" based on calculating the greater common divisor of the different links' capacities. This approach maintains the proportional relation between the different links and can be easily calculated with the pseudocode shown in Figure 3-27.

The calculation of the GCD can be performed with the recursive `calculateGCD` function described in Figure 3-27. To show the performance of the previous algorithms, Figure 3-28 represents the execution of a simple simulator which calculates the maximum throughput achievable within a set of links capacities and their corresponding set of weights. In this case, only two links are considered. In the figure, the darker the colour of the plot, the closer to the maximum ratio (1.0). The x and y axis represent the capacity of each link. As it can be observed, Ceil, Floor and Round criteria obtain areas which values ratios below 0.75, whereas the GCD criterion always achieves the maximum ratio.

To verify these results, obtained by simulation, we perform 12 executions of the different algorithms within two VirtualBox VMs. In this case, 3 network interfaces are attached in each VM. Each interface relates to a different interface of the other VM.

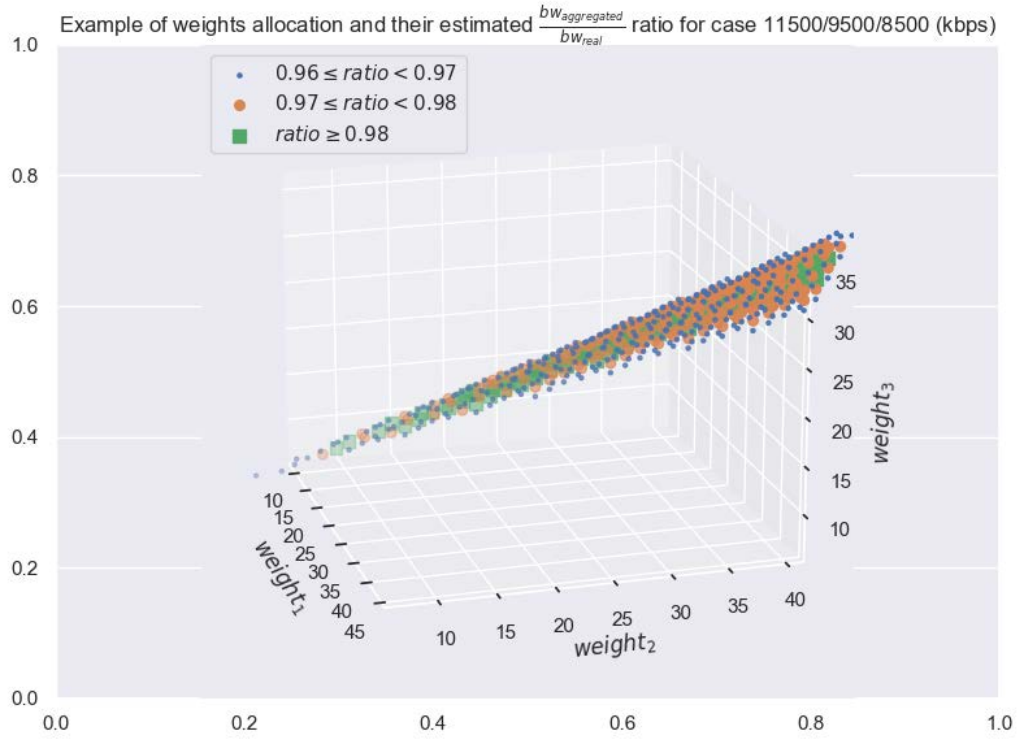



**Figure 3-26 Example of round robin weights which obtains ratio values greater thar 0.96 for a case with links whose capacities are 11500kbps, 9500kbps and 8500kbps**

```
// The input of the algorithm is the list of the capacities
 for each of the n_links links, labelled as linkCapacity_i.
Input: linkCapacity_i, for i = 0 ... n_links − 1

// The output of the algorithm is the ordered list of
 weights. Namely, for each interface i, the algorithm
 calculate its weight weight_i.
Output: weight_i, for i = 0 ... n_links − 1

// we initialize the GCD variable total_gcd with the capacity
 of first link.
total_gcd ← linkCapacity_0

// The GCD of a set of numbers can be calculated by
 repeatedly taking the GCDs of pairs of numbers:
for i ← 1 to n_links do
    total_gcd ← calculateGCD(linkCapacity_i, total_gcd)

// We calculate weight as the link capacity divided by the
 GCD:
for i ← 0 to n_links do
    weight_i ← linkCapacity_i / total_gcd

// Auxiliar function for calculating the GCD recursively.
Function calculateGCD(a, b):
    if a = 0 then
        return b
    else
        return calculateGCD(b mod a, a)
```

**Figure 3-27 Pseudocode for the "GCD criterion"**

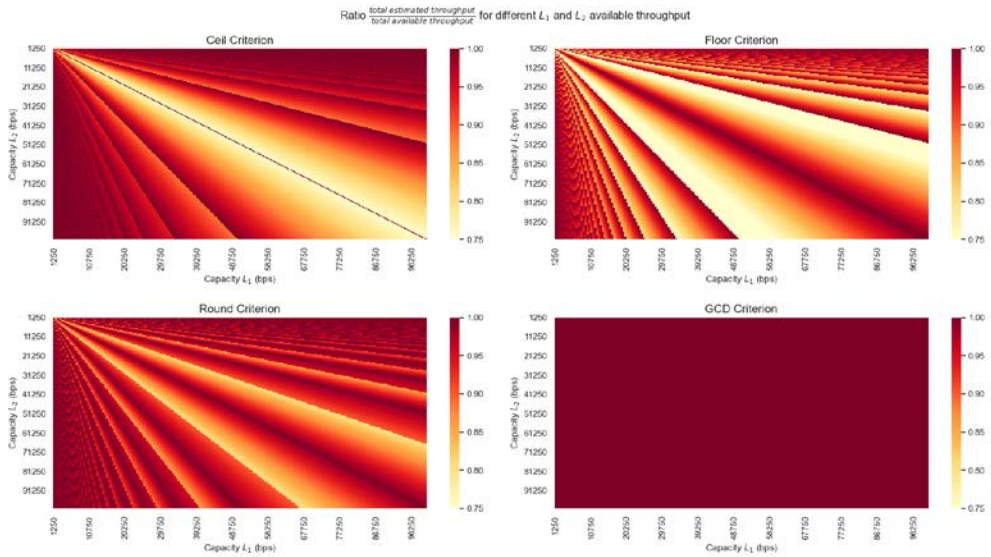



**Figure 3-28 Results of achieved versus maximum achievable throughput ratio with different weight assignation strategies**

In each experiment we configure the bit rate of each network interface with the `tc` tool and we launch the iperf tool at full throttle. Each experiment lasts 60s, and we use the aggregated average output of the `ifstat` tool to assess the achieved throughput.

Figure 3-29 shows the resulting performance achieve for different link's configurations and algorithms applied to select the weights' set. As it can be stated, in these experiments, the resulting throughput obtained with the assignation of weights of the GCD criterion is the highest, as the simulations predicted.

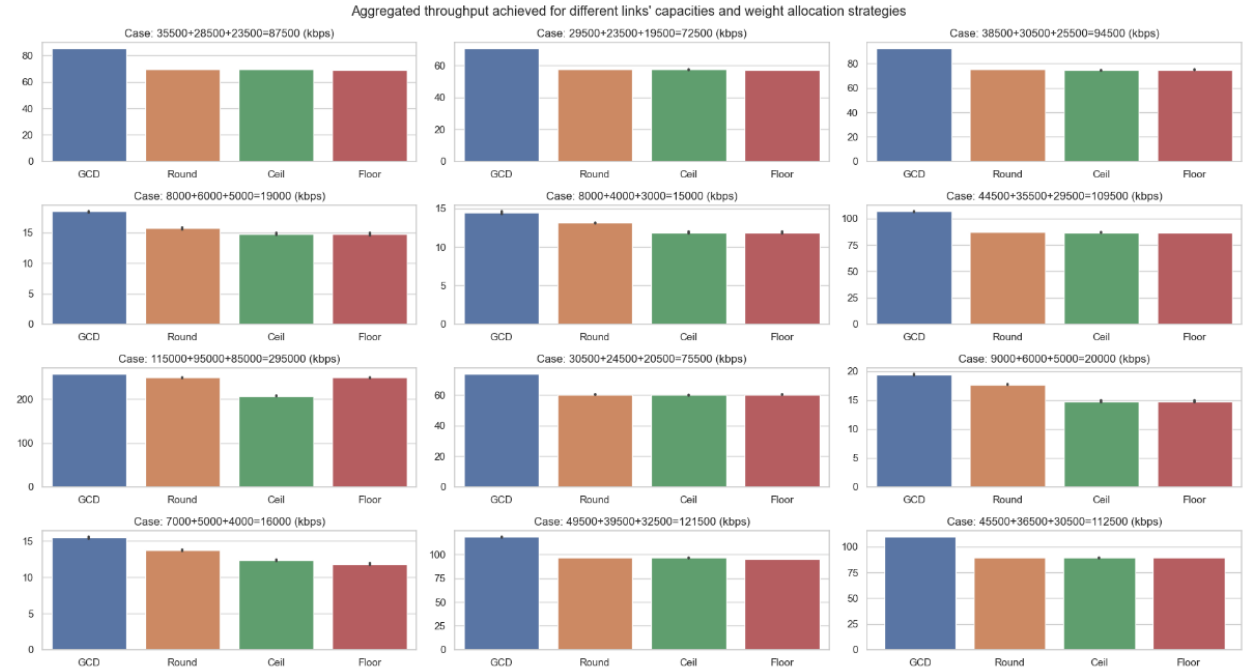



**Figure 3-29 Resulting throughput for different capacities sets and weight assignment algorithms**

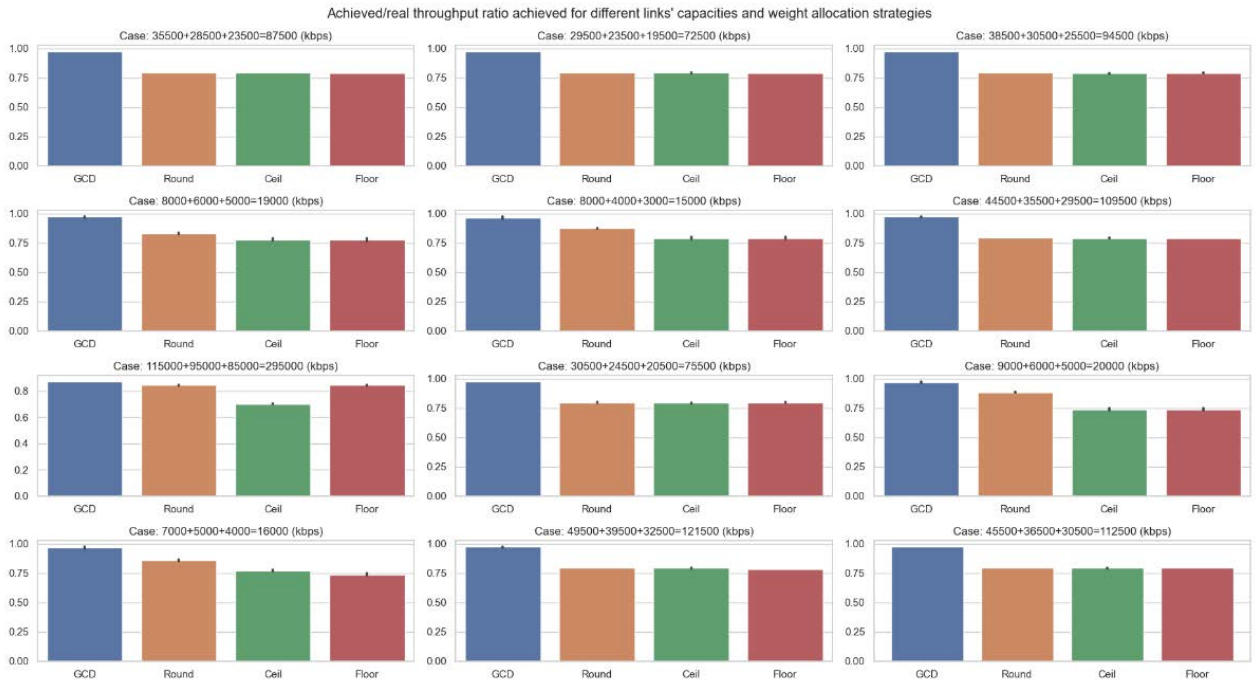



**Figure 3-30 Resulting ratio for different capacities sets and weight assignment algorithms**

Figure 3-30 illustrates these results in terms of the throughput ratio, the parameter that we aim to maximize. In that case, we can observe that the GCD criterion may achieve a 25% higher ratio than the others.

### 3.3.3 5G-CLARITY low-level eAT3S user plane function

The proposed 5G-CLARITY Low-level eAT3S or LL-ATSSS user plane function is designed to enable the Optimal Access Network framework (OANF) proposed in D4.1. It provides flexible and efficient multi-connectivity. The solution is based on the data received from the multiple sources. These sources are 1) UE

telemetry, 2) access policy from the OANF, and 3) the quality of experience or application requirements enabling the ATSSS through single or aggregated multiple connectivity. In this section, the 5G-CLARITY Low-Level eAT3S solution and components are introduced with an illustrative example. The example emulates the 5G-CLARITY Multi-WAT architecture.

#### 3.2.4.1. Architecture and modules

The 5G-CLARITY Low-level eAT3S is a dedicated user plane function assisted by the OANF and UE Telemetry application. It supports flexible multi-connectivity ATSSS functionalities with 3GPP and non-3GPP access technologies (e.g., any-connectivity through B5G any-WAT technology). The solution operates in the Multi-Access (MA) PDU layer (N3) defined in the 3GPP rel-16. However, it is also designed to operate in full non-3GPP multi-WAT environments and with any-multi-connectivity frameworks. It summarizes the architecture embedded into the 3GPP rel-16 standard. Following the standard for each Service Data Flow (SDF) the ATSSS-enabled Policy Control Function (PCF) must generate rules for Policy and Charging Configuration (PCC). Rules translated by the ATSSS-enabled Session Management Function (SMF) both into the N4 rules of the UPF. PCC rules include routing and QoS information (FAR, PDR, etc.) and dedicated Multi-Access Rules (MAR). Finally, the ATSSS rules must be delivered via the Access and Mobility Management Function (AMF) to the UE. This goes with the UE Route Selection Policy (URSP) rules via NAS (N1). The Optimal Access Network framework can generate global and flexible ATSSS and MAR rules for schedule multi-connectivity protocol (traffic steering) over any 3GPP and non-3GPP wireless access technology. In Figure 3-31 we present the components of the 5G-CLARITY Low-Level eAT3S function. This is formed by the LL-ATSSS Transport Protocols Proxy modules. These modules are connected to the Optimal Access Network Framework and telemetry functions deployed in the UE and with UPF.

- **Transport Protocol Proxy and Scheduler Module (T2PSM):** It is a cluster of single and multi-path transport protocols, scheduler, and proxy controlled by LL-ATSSS module. The cluster includes, TCP, UDP, and RTP for single connectivity and MPTCP [27], MPUDP, MPQUIC [39], and MP-RTP (RTCP) [38] for multi-connectivity. T2PSM is deployed in conjunction with the LL-ATSSS and the Optimal Access Network Framework on both sides - in the UPF or Application side and the UE.

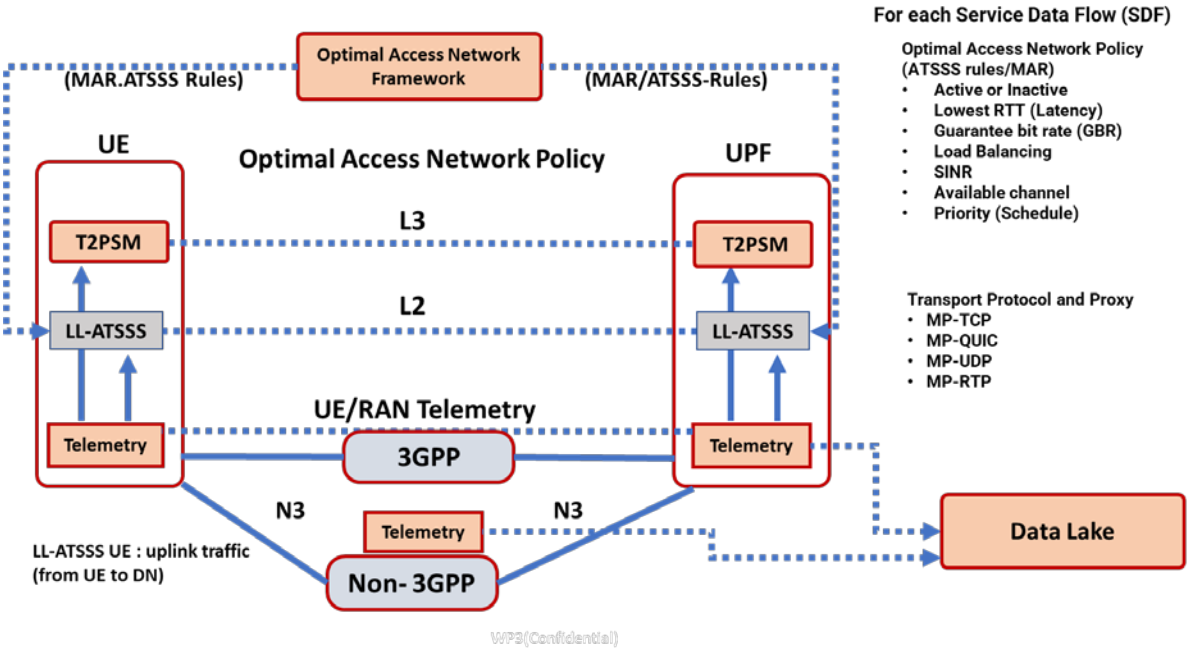



**Figure 3-31 5G-CLARITY Low-level eAT3S Architecture Overview**

- **Low-Layer ATSSS (LL-ATSSS) Module:** It is responsible for execution of the appropriated ATSSS rules based on the policies provided by the Optimal Access Network framework. This module applies the given policy by enabling uplink traffic to be transferred through an optimal combination of channels and access nodes (e.g., aggregating). This is done while selecting and scheduling single or multi-path transport protocols based on the application demands and network conditions. For downlink traffic, the LL-ATSSS deployed in the UPF or in the Edge (application side) performs selection and scheduling. This consist of single or multi-path protocol bases on the application demand and the access network policy provided by the Optimal Access Network Framework.

Optimal Access Network Framework is the implementation of the algorithm Optimal Access Network based on Deep Reinforcement Learning described in the D4.1, D4.2 [36], and D4.3 [37]. The main role of the frameworks is to generate the ATSSS and MA rules using the data sets generated by the telemetry functions. Telemetry: It is a set of frameworks introduced and described in the 5G-CLARITY project focused on the provision of necessary analytics for the Optimal Access Networks Framework and LL-ATSSS modules

### 3.2.4.1. Flow and process

5G-CLARITY Low-level eAT3S main flow enabled three key processes – 1) optimal ATSSS and MA rules update, 2) selection and schedule of transport protocols, and 3) optimal traffic steering through multiple radio/wireless access networks. An example of the flow of the optimal ATSSS and MA rules update is shown in Figure 3-32.

In the example of Figure 3-32 we observe two UEs regularly updating the state of the signal. These signals are the connectivity availability, and traffic to the Data Lake through their UE Telemetry module. On the RAN side, a telemetry module also provides regular updates of the global state of the multiple wireless access networks (nultiWAT) nodes. Then the new state of the two UEs and the RAN is updated to the Optimal Access Network Framework. This action begins the update of ATSSS and MA rules for each UE. The process requires some processing before ATSSS, and MA Rules are sent to the LL-ATSSS functions hosted in the UEs and UPF or edge. In Figure 3-33, the flow of the selection and schedule of transport protocols is presented. It introduces with example, the optimal traffic steering through multiple radio and wireless access networks.

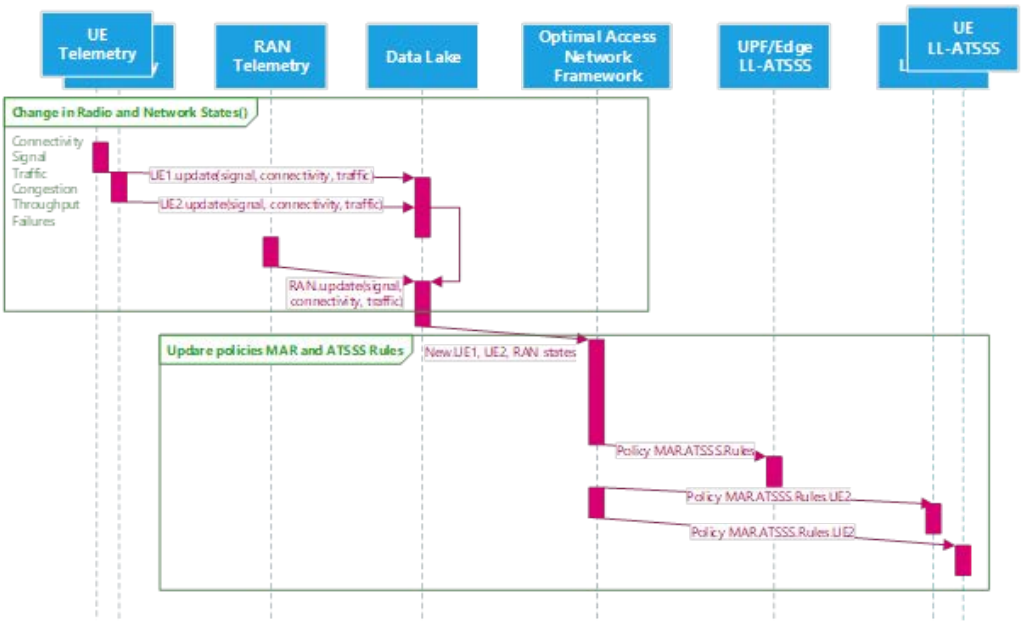



**Figure 3-32 Optimal ATSSS and MA rules update flow**

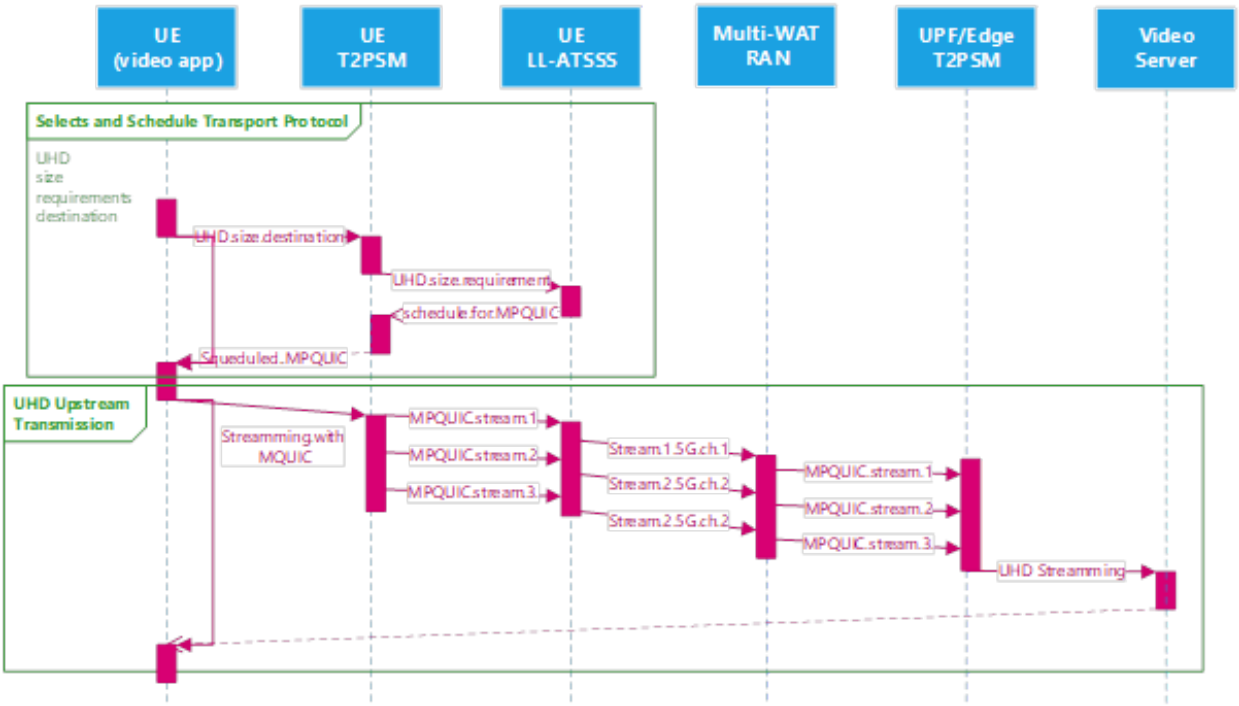



**Figure 3-33 Selection and scheduling transport protocol**

The process of selection and schedule of transport protocols is shown in Figure 3-33, which uses the ATSSS and MA rules updated previously and the Application or Service requirements. In this example, Ultra High-Definition Video streaming will need to be sent to a video server, the chosen transport protocol is the MP-QUIC protocol is scheduled based on the ATSSS and MA rules. Then, the optimal traffic steering through multiple radio/wireless access networks is executed, as shown in Figure 3-33, the UHD streaming traffic is steered, split, and transmitted using MP-QUIC through 5GNR and Wi-Fi channels. Examples of the ATSSS and MA rules application in the UE are:

- **For single video streaming,** Single-UDP path, Dest Addr", "Steering Mode: Active-Standby, Active=5G, Standby=Wi-Fi" "IF Wi-Fi SINR > 20 dB, and 5G SINR < 20 and Wi-Fi GBR > 20Mbps and 5G < 20 Mbps", Set active = Wi-Fi, Set Active-Standby = 5G, and send UDP to destination using Wi-Fi.
- **For a single file transfer**, Single-TCP, DestPort 8080", "Steering Mode: Smallest Delay": "steer TCP traffic with destination port 8080 to the access with the smallest delay". UE collect RTT value from UE/RAN Telemetry and as previous rule using SINR and GBR decide to activate either 5G or Wi-Fi.
- **For application-1,** "Steering Mode: Load-Balancing, 3GPP=20%, non-3GPP=80%", "Steering Functionality: MPTCP": This rule means "send 20% of the traffic of Application-1 to 3GPP access and 80% to non-3GPP access by using the MPTCP scheduler".
- **For Application-2**, "Steering Mode: Load-Balancing, 3GPP=40%, non-3GPP=60%", "Steering Functionality: MPQUIC": This rule means schedule "Stream 1" with 40% of the traffic of Application-2 to 5G and schedule "Stream 2" with 60% of the traffic to Wi-Fi by using the MPQUIC scheduler.

#### 3.2.4.1. Illustrative Example

Using the example of Figure 3-34 we implement a test with an early release of the LL-AT3S solution enabling UDP, TCP, MPTCP, and MPQUIC. In this illustrative example, two UEs are connected to a Video Server hosted in the UPF or edge location.

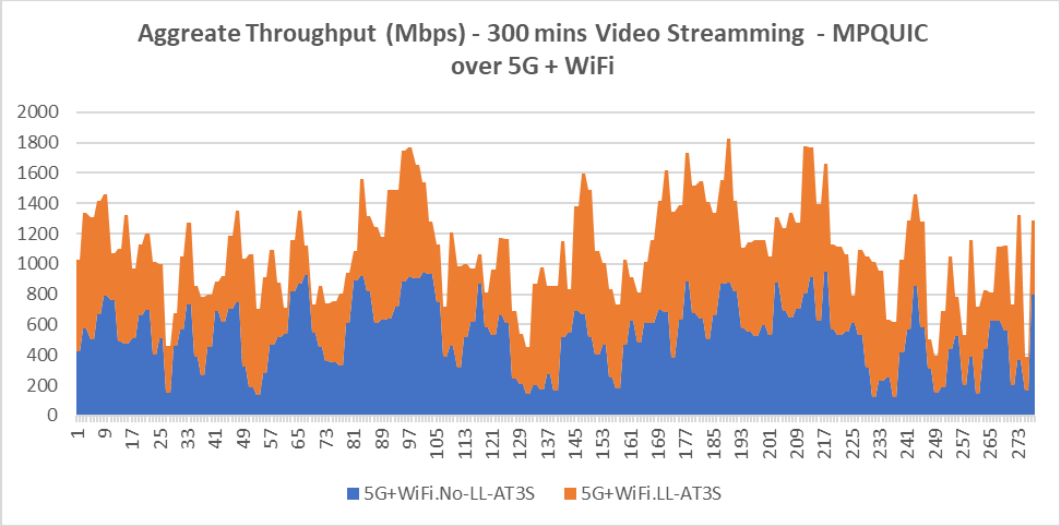



**Figure 3-34 Aggregated Throughput in 300 minutes Video Streaming**

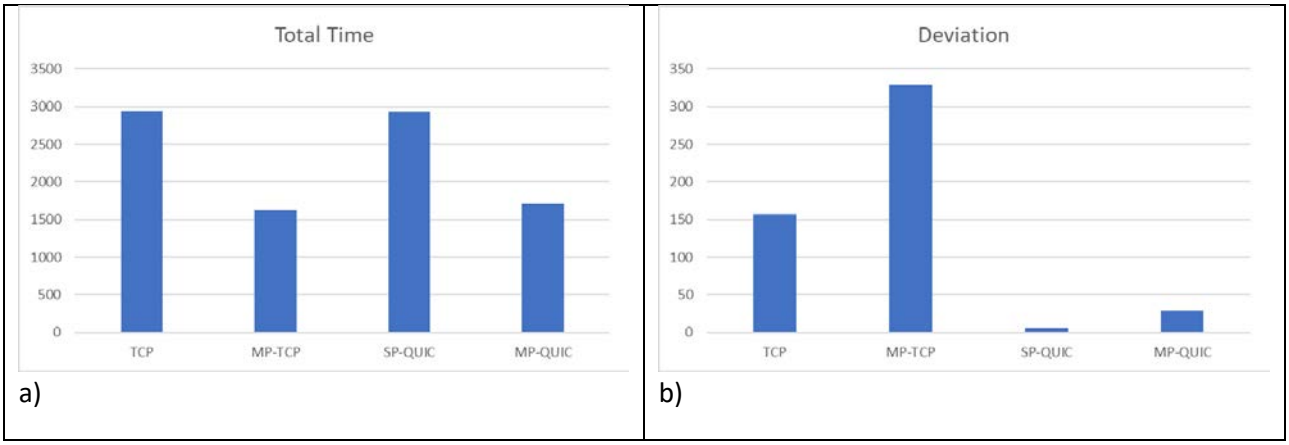



**Figure 3-35 Protocol operation overhead and the deviation**

We tested 300 minutes of video streaming from the two UEs, enabling LL-AT3S selection and scheduling of the transport protocol as well as the application of the ATSSS and MU rules. Figure 3-34 presents the results from the 300 minutes of UHD video streaming. This test used MPQUIC (Application 2) with UDP (additional single video streaming) to be able to adapt the aggregated video streaming. Random, reduction and fluctuation on the SNIR (e.g., between 25 and 26 minutes) increase in congestion (e.g., between min 10 and 80 minutes) and multiple handovers where simulate minutes (e.g., at minutes 25, 52, 140, 156,231, 238, 257, 265, and 273)). As observed, the usage of the proposed LL-AT3S might lead to nearly 80% increase in the aggregated throughput. Also observed is the minimization of the negative impact due to the reduction of the SNIR and the increase of the congestion and WAT handover. In this example the UHD Video streaming selected and scheduled simultaneously the MPQUIC with UDP, adapting it to the changes in the network state. Similar test was conducted with MPTCP and TCP. Figure 3-35 shows the total time of protocol operation overhead and the standard deviation or changes required during the streaming.

Figure 3-35 shows the TCP and single point QUIC required more overhead to transmit the video streaming compared their Multipath versions. And in terms of changes or reconfiguration required, MP-QUIC outperform MPTCP when implemented with AT3S. As results were obtained in a controlled environment, further study and test must be conducted to confirm our findings.

## 3.4 Section summary

The 5G-CLARITY advanced multi-connectivity and multi-RAT Aggregation framework has been extended from D3.2. The extensions and architecture enhancement mainly include (i) a functional testbed with an API to manage the steering policies in an MPTCP based AT3S user plan function in real-time, (ii) integration a multi-weight factor assignment in the eAT3S algorithm presented in D3.2 to support adaptive 5G URLLC and eMBB traffic load-balancing steering mode according to their QoS requirements, (iii) adaptive and probing based available bandwidth estimation on the mWAT interfaces to maximize their capacity utilization, (iv) implementation of the 5G-CLARITY multi-connectivity framework architecture with real 5GNR, WiFi and LiFi access networks.

The proposed 5G-CLARITY multi-connectivity framework architecture allowed to achieve an aggregated throughput of 1.45 Gbps, which mainly referred to the integration of MPTCP and AT3S operations, the designed reliability requirements based the air interface of 5GNR. The air interface latency was near the KPI objective, though new measurements are needed with mature 5GNR equipment. A demo can be viewed by public, and the development repository of the testbed can be accessed by other researchers around the world. The updated eAT3S algorithm enables the inclusion of multiple network parameters in the steering decisioning. It would enable private/public network operators to derive user or environment specific network policy configurations to efficiently utilize 3GPP and non-3GPP networks. Another AT3S solution is also introduced, which enables flexible selection and scheduling of one or multiple transport protocols. The initial experimental results show the benefit of this proposed solution in enabling multi-connectivity over a B5G Multi-WAT network.

# 4 Advanced Resource Management

This section introduces intelligent LiFi and Wi-Fi networks slicing schemes, which provide network slice customization parameters to mobile network operators (MNOs) and mobile virtual network operators (MVNOs). The resource utilization of network slices is customised and optimized, while satisfying their heterogeneous quality-of-service (QoS) requirements. The admitted traffic on network slices may vary over time at the different network cells, where dynamic resources allocation to network slices becomes challenging and cannot be accurately predicted. Hence, LiFi or Wi-Fi network slicing problems can be formulated as Markov decision problems (MDP), which can be solved by using deep reinforcement learning (DRL) approaches. Advanced resource allocation mechanisms are developed, which solve the decision-making problem of network slicing at the network level and locally at the different cells. A two-level algorithm is introduced, which employs a DRL approach. This dynamically shares the downlink channel bandwidth of network access points (APs) among a set of network slices while controlling traffic packets scheduling at the different network cells. In essence, a centralized scheduling policy, Global-Airtime Deficit Round Robin (G-ADRR), is developed, which tunes the airtime weights of APs distributed over a geographical area according to their traffic load and network slices requirements. The proposed network sharing solutions for the interface airtime of Wi-Fi APs and the downlink channel capacity of network LiFi APs are designed as self-organizing network (SON) functions. These can automatically and dynamically adjust the capacity provided to each Wi-Fi and LiFi network slice across the different network cells. This work agrees with the study conducted by the 3GPP standard on the SON functionalities for 5G, which identifies the so-called cross-slice network resource optimization use case [44]. This is intended to optimise the allocation of physical and virtual resources across multiple network slice instances.

## 4.1 Autonomic LiFi attocellular network slicing

A hierarchical software defined autonomic LiFi access network slicing enables MNOs to customize network slices through providing differentiated networks slices. It also enables MVNOs to customize their quota guarantees offered to their network slices through dynamically offering heterogenous services with diverse QoS requirements. The proposed soft network resource slicing framework consists of network-level slicing and LiFi AP-level slicing. In the network level, network slices are allocated resource or data rate quota guarantees at LiFi attocells. While in the LiFi AP level, the allocated resource units are dynamically scheduled to the traffic packets of 5G services requested by the users subscribed to the different network slices.

### 4.1.1 Background and motivation

A utility scheduler based LiFi attocellular network slicing is introduced in D3.2 [2]. In this work, utility, and marginal weight utility functions-based resource to network slices, $\mathcal{M}$ , scheduling policies are developed to support multiple MVNOs. These dynamically offer 5G ultra-reliable low latency communication (uRLLC) (u), massive machine type communication (mMTC) (m) and enhanced mobile broadband (eMBB) (e) services on resource customized LiFi network slices [45]. The service guarantees of uRLLC, mMTC and eMBB are encoded in a tuple, [u m e], which should be enforced per LiFi AP, *n*, to meet the percentage of maximum data rate or resources to be guaranteed for each MVNO.

The total utility of sliced LiFi attocellular network is maximized with respect to the packets' transmission throughput or head-of-line (HoL) delay, according to the considered policies for resource scheduling to the users of network slices. In this work D3.2 [2] and [45], the slice-specific throughput and delay utility weight parameters, $\phi_m^n = \{\phi_m^{th}, \phi_m^d\}$, are set arbitrarily to investigate their impact on the performance of 5G services offered on the network slices.

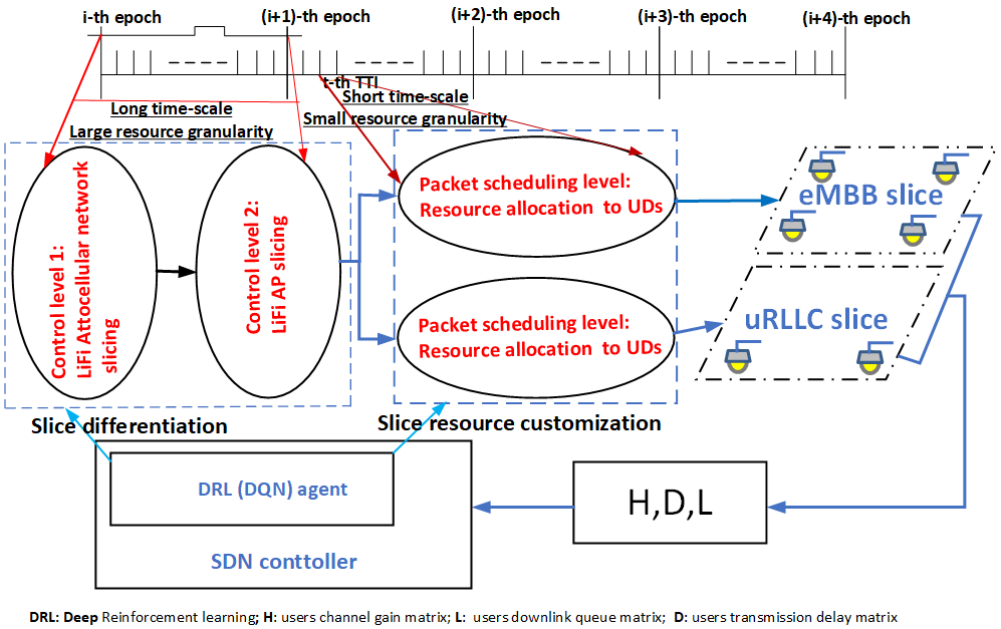



**Figure 4-1 Conceptual schematic of the proposed autonomic LiFi attocellular network slicing framework**

LiFi attocellular network APs with small coverage and channel dynamics require an automation for network resource slicing operations to meet the dynamic QoS requirements of MVNOs in the different network states in an efficient and autonomic manner. In this network slicing automation process, MVNOs state intents that define the expected performance of their network slices, which should then be intuitively translated into real-time network actions. In essence, a LiFi attocellular network resource slicing automation framework is introduced, as shown in Figure 4-1 . A centralized DRL scheme is run by the SDN controller to manage resource allocation to network slices and traffic packets scheduling to their users at the LiFi APs.

### 4.1.2 System model

A LiFi attocellular network is composed of a set of LiFi APs, $\mathcal{N} = \{1,2,\cdots,|\mathcal{N}|\}$, which is run by a single MNO. A LiFi AP has a maximum transmit power, $p_{n,\max}, \forall n \in \mathcal{N}$. The network management runs in discrete time, $t \in \{0,1,\cdots\}$. The AP downlink channel bandwidth, $B_n$ Hz, is divided into set of Direct Current-biased Optical Orthogonal Frequency-Division Multiplexing (DCO-OFDM) sub-carriers (resource units), $\mathcal{O} = \{1,2,\cdots,|\mathcal{O}|\}$, which each has a bandwidth b Hz. A LiFi AP uses half of the available sub-carriers to realize the Hermitian conjugate of the complex-valued symbol after modulation, as only real-valued signals can be transmitted to users [46]. Let $\beta_{n,k}^{o}$ denotes a binary decision variable, which equals 1 if a resource unit o is allocated to user k associated with AP n; and otherwise, it equals to 0. The LiFi AP downlink channel is managed by an orthogonal frequency-division multiple access protocol (OFDMA) [47]. A set of network slices, $\mathcal{M} = \{1,2,\cdots,|\mathcal{M}|\}$, are configured on the downlink channel resources of network LiFi APs. Network slices are allocated resource quotas to LiFi attocells, where they are instantiated. Network slices have a priority, $\varsigma \in [0,1]$, such that $\sum_{m\in\mathcal{M}} \varsigma_m = 1$, which reflects their provided network service priority. A tenant (MVNO) of network slice provides a single service. We denote $\vartheta^{\mathrm{t}} = \{\vartheta_1^{\mathrm{t}}, \vartheta_2^{\mathrm{t}}, \cdots, \vartheta_{\mathrm{K}}^{\mathrm{t}}\}$ as control variables for scheduling users at time slot t. A user k subscribed to MVNO m and associated with LiFi AP *n* has a scheduling control variable $\vartheta_{\mathrm{k}}^{\mathrm{t}} \in \{0,1\}$, where $\sum_{\mathrm{k}\in\mathcal{K}_m^n} \vartheta_{\mathrm{n,k}}^{\mathrm{t}} = 1$, $\mathcal{K}_m^n$ denotes the set of users associated to AP *n* and subscribed to MVNO *m*. A tenant (MVNO) of network slice provides a single service. Tenants have SLAs that define their minimum and maximum target rate $[\eta_{n,m}^{\min}, \eta_{n,m}^{\max}]$ or resource guarantees, $[Y_{\min}^{n,m}, Y_{n,m}^{\max}]$ in each LiFi attocell *n*.

### 4.1.3 Problem formulation

The process of LiFi AP downlink channel resource slicing allocates maximum quota guarantees to MVNOs as follows:

$$\boldsymbol{Y}_{\boldsymbol{n,m}}^{\max} = \lfloor \boldsymbol{\psi}_{\boldsymbol{n,m}} \boldsymbol{O}_{\boldsymbol{n}} \rfloor, \qquad \textbf{(4-1)}$$

where $O_n$ is the total downlink channel resource units of LiFi AP *n*; $\psi_{n,m} \in [0,1]$ is a control parameter which the MNO uses to customize $O_n$ among the network slices by scaling up or down their allocated maximum quota guarantees at AP *n* according to the SLA with MVNO *m*, $\forall n \in \mathcal{N}, \forall m \in \mathcal{M}$. Note that, $\psi_{n,m}$, is managed by the MNO on a long to medium time scale according to the availability and quality of network resources or emerging business requirements of MVNOs.

Network slice utility based LiFi AP downlink channel resources scheduling policies are developed in [45] , which schedule resource quota of network slices to their users associated with the different LiFi APs. The 5G eMTC, eMBB and URLLC services consume different amount of resource units to meet their requirements in terms of delay, throughput, and reliability. Furthermore, the traffic characteristics of URLLC and eMBB services are different. The URLLC traffic is sporadic with short packet sizes, whereas the eMBB traffic arrive in batches with larger packet sizes. This requires to dynamically update the quota guarantees of each slice *m* at each LiFi AP *n* according to the AP channel conditions, type and volume of traffic demands.

However, this is a challenging task, because of the heterogeneous QoS requirements of network slices and their traffic as well as network resource quality and availability change on a short-time scale. This requires to optimally allocate the available resources among the traffic flows *f* of user *k* subscribed to a slice *m*. The traffic flow arrival and departure may result in utility disparity of network slices at the different LiFi APs, degrading the overall service quality of network slices. Furthermore, the MVNOs may also have different number of users at the network LiFi APs. To this end, the MVNOs uses the weight control parameter $\phi_{n,m} \in [0,1]$ to scale (reset) their $\mathrm{Y}_{\mathrm{n,m}}^{\max}$ to $\max\{0, \zeta_{n,m}\}$ at LiFi AP *n*, as follows

$$\boldsymbol{\zeta}_{\boldsymbol{n,m}} = \lfloor \boldsymbol{\phi}_{\boldsymbol{n,m}} \boldsymbol{Y}_{\boldsymbol{n,m}}^{\max} \rfloor, \qquad \textbf{(4-2)}$$

where $\phi_{n,m}$ should be managed on a short-time scale at the different network LiFi APs. It is also used to enforce the scheduler to schedule traffic packets according to their delay or throughput, which may enable the MVNOs to transfer their resource quotas among the different LiFi APs and make some profit. The MVNOs aim to match their resource quota to the resource units demanded by their users and their traffic at the different LiFi APs.

The optimization objective in Eq.(4-3) is to maximize the capacity assigned to each user $\mathrm{k} \in \mathcal{K}_m^n$, subject to the priority and capacity share of their slice at LiFi AP *n*, $\forall n \in \mathcal{N}$.

$$\max_{\boldsymbol{\eta}_{\boldsymbol{n,k}}^{\boldsymbol{o}}} \sum_{o \in \mathcal{O}} \sum_{k \in \mathcal{K}_m^n} \boldsymbol{\varsigma}_{\boldsymbol{m}}^{\boldsymbol{n}} \boldsymbol{U}\left(\overline{\boldsymbol{\eta}}_{\boldsymbol{n,k}}\right) \boldsymbol{\eta}_{\boldsymbol{n,k}}^{\boldsymbol{o}} \qquad \textbf{(4-3)}$$

$$\boldsymbol{s.t.} \sum_{k \in \mathcal{K}_m^n} \boldsymbol{\beta}_{\boldsymbol{n,k}}^{\boldsymbol{o}} \boldsymbol{\vartheta}_{\boldsymbol{n,k}}^{\boldsymbol{t}} \leq \mathbf{1}, \quad \forall \boldsymbol{n, m, o, t} \qquad \textbf{(4-4)}$$

$$\sum_{o \in \mathcal{O}} \sum_{k \in \mathcal{K}_m^n} \boldsymbol{\eta}_{\boldsymbol{n,k}}^{\boldsymbol{o}} \boldsymbol{\beta}_{\boldsymbol{n,k}}^{\boldsymbol{o}} \boldsymbol{\vartheta} \leq \boldsymbol{\zeta}_{\boldsymbol{n,m}} \boldsymbol{b}, \quad \forall \boldsymbol{n, m} \qquad \textbf{(4-5)}$$

$$\boldsymbol{\eta}_{\boldsymbol{n,k}}^{\boldsymbol{min}} \leq \sum_{o \in \mathcal{O}} \boldsymbol{\eta}_{\boldsymbol{n,k}}^{\boldsymbol{o}} \leq \boldsymbol{\eta}_{\boldsymbol{n,k}}^{\max}, \quad \forall \boldsymbol{n, k} \qquad \textbf{(4-6)}$$

$$\boldsymbol{d}_{\boldsymbol{n,k}} \leq \boldsymbol{d}_{\boldsymbol{n,k}}^{\boldsymbol{max}}, \quad \forall \boldsymbol{n, k} \qquad \textbf{(4-7)}$$

$$\boldsymbol{\beta}_{\boldsymbol{n,k}}^{\boldsymbol{o}} \in \{\mathbf{0}, \mathbf{1}\}, \quad \forall \boldsymbol{n, k, o} \qquad \textbf{(4-8)}$$

$$\boldsymbol{\vartheta}_{\boldsymbol{n,k}}^{\boldsymbol{t}} \in \{\mathbf{0}, \mathbf{1}\}, \quad \forall \boldsymbol{n, k, o, t} \qquad \textbf{(4-9)}$$

$$\boldsymbol{\eta}_{\boldsymbol{n,k}}^{\boldsymbol{o}} \geq \mathbf{0}, \forall \boldsymbol{n, k} \qquad \textbf{(4-10)}$$

Constraint 1 Eq.(4-4) restricts user association to a single attocell and subscription to a single MVNO per time slot *t*. Constraint 2 Eq.(4-5) imposes an upper bound on the total data rate received by MVNO $m, \forall m \in$

$\mathcal{M}$ in each attocell. Constraint 3 Eq.(4-6) impose lower and upper bounds on the total data rate achieved by each user $\mathrm{k} \in \mathcal{K}_m^n$. Constraint 4 Eq. (4-7) imposes an upper bound on the delay experienced by user $\mathrm{k} \in \mathcal{K}_m^n$, which may vary according to the delay sensitivity of their network slice. Constraint 5 Eq. (4-8) requires a unique resource allocation to only one user *k* in an attocell *n*, $\forall n \in \mathcal{N}$. Constrain 6 Eq. (4-9) requires a unique user *k* scheduling in a time slot *t* under an attocell *n*, $\forall \mathrm{n} \in \mathcal{N}$. Constraint 7 Eq. (4-10) imposes a lower bound on the power assigned to a resource or its achieved data rate allocated to a user *k*, $\mathrm{k} \in \mathcal{K}_m^n$, in each attocell.

### 4.1.4 MDP formulation at SDN controller

A Markov decision problem (MDP) formulation enables the LiFi network APs slicing to consider the expected rewards of all the possible actions in network LiFi APs slicing decision. Since closed-form expressions typically do not exist for the expected reward of each possible action at each system state in a real-world problem, RL is commonly used to empirically learn the optimum policy for the MDP formulation. The RL agent (LiFi AP) learns to maximize the expected reward by trial and error. The requested resource units $o_n(t) \in \mathcal{O}_n$ by the different users of MVNOs are received by the LiFi AP agent.

As a result of the action $a^t$ taken by the agents, the LiFi APs $n \in \mathcal{N}$ will observe a reward $r^t$ and the system state $s^t$ will transition to $s^{t+1}$. A MDP is defined by the tuple $(\mathcal{N}, \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$ where $\mathcal{N}$ is the set of APs; $\mathcal{S}$ is the set of states, $\forall s^t \in \mathcal{S}$; $\mathcal{A}$ is the set of actions $\forall a^t \in \mathcal{A}$; $P(s, s') = P(s^{t+1} = s' | s^t = s, a^t = a)$ is the transition probability from state $s$ to $s'$ when action *a* is taken; *R*(*s*, *s'*) is the reward received by taking action *a*, in state *s*, which is transited to state *s'*, $\forall r^t \in R^a(s, s')$, which is accumulated over the different time slots (epochs), given the cumulative reward as follows:

$$\boldsymbol{G^t = \sum_{j=0}^{T-1} \gamma^j r^{t+j+1}}, \qquad \textbf{(4-11)}$$

Where, $\gamma \in [0,1]$ is the discount factor in computing the return. It represents how much weight is given to the future rewards compared to the immediate reward. For $\gamma = 1$, future rewards are of equal importance as the immediate reward, whereas $\gamma = 0$ completely ignores future rewards. The objective in MDP is to maximize the expected cumulative reward starting from $t = 0$ i.e., $\max_{a^t} E\left[G_{t=0} | s_{t=0}\right]$, where $\mathrm{G^t}$ is given above, by choosing the actions $a^t$. The different DQN agents of tenants compose the state space of system shown in Figure 4-2, which is a pre-processed data observed from the sliced LiFi network environment. The temporal and spatial correlation of service demands for the different network slices is considered in the local observation for AP *n* at time *t* by considering its past and current service demands. Deep Q-network (DQN) pre-processing module contains a network monitoring module which maintains the performance measurement of the following metrics per tenant *m* at the previous time slot *t*-1 and calculates them at the current time slot *t*.

- The requested traffic load of each tenant *m* at each LiFi AP *n*, $\chi_{n,m}$, is the aggregate of the resource units requested by their subscribers, given as follows:

$$\boldsymbol{\chi_{n,m} = \frac{\sum_{k \in \mathcal{K}_m^n} \mathrm{l}_{n,k}^m}{b}}, \qquad \textbf{(4-12)}$$

  where $\mathrm{l}_{n,k}^m$ denotes the packet length of users $k \in \mathcal{K}_m^n$ at time *t* slot.

- The resource quota utilisation of each tenant *m* at LiFi AP *n*, $\rho_{n,m}$, can be calculated as follows:

$$\boldsymbol{\rho_{n,m} = \frac{\min(\chi_{n,m}, \zeta_{n,m})}{o_n}} \qquad \textbf{(4-13)}$$

- The aggregate throughput of tenant *m* across the network is the aggregate throughput of their users served at the different network LiFi APs, which is calculated per LiFi AP *n*, $\eta_{n,m}$, as follows:

$$\boldsymbol{\eta_{n,m} = \min(\chi_{n,m}, \zeta_{n,m})\, b} \qquad \textbf{(4-14)}$$

### 4.1.5 DRL model design

#### 4.1.5.1 State

At each time slot *t*, an agent of LiFi network slice tenant *m* presents the state of network slice environment, as follows [48]:

$$s_m^t = \{s_{1,m}^t, \cdots, s_{N,m}^t, \zeta_{n,m}, \frac{\sum_{m' \in \mathcal{M}-m} \zeta_{n,m'}}{\sum_{n \in \mathcal{N}} O_n}\} \quad \textbf{(4-15)}$$

Let $\rho_n$ denotes the available resource units at LiFi AP *n*, given as follows:

$$\rho_n^t = 1 - \sum_{m \in \mathcal{M}} \rho_{n,m}^t. \quad \textbf{(4-16)}$$

Let $\psi_n$ denotes the available share of resource units that are not allocated to any tenant at the LiFi AP *n*, given as follows

$$\psi_n^{t-1} = 1 - \sum_{m \in \mathcal{M}} \psi_{n,m}^{t-1} \quad \textbf{(4-17)}$$

The agent of network slice tenant *m* presents the state of network slice at LiFi AP *n*, $n \in \mathcal{N}$, as follows:

$$s_{n,m}^t = \{\rho_{n,m}^t, \rho_n^t, \psi_{n,m}^{t-1}, \psi_n^{t-1}, \zeta_{n,m}(t), Y_{n,m}^{\max}\} \quad \textbf{(4-18)}$$

#### 4.1.5.2 Actions

At time *t* the network slice *m* (agent) takes actions at the different LiFi AP $n \in \mathcal{N}$ presented as follows:

$$a_m^t = [a_{1,m}^t, \cdots, a_{N,m}^t], \quad n \in \mathcal{N}, \quad \textbf{(4-19)}$$

where, according to the state of network slice, an action at LiFi AP $n = 1, a_{1,m}^t$, may increase or decrease gradually with a small increment or maintain the capacity share for the slice *m*.

#### 4.1.5.3 Reward calculation

The reward function should reflect the ability of the taken action to fulfil the targets of the optimization problems. The reward evaluates the action $a^{t-1}$ at the system state $s^{t-1}$. The network slice *m* agent is rewarded every time it performs an adequate action, as follows:

$$r_m^t = e^{1-\varrho_m^t} \cdot e^{1-\upsilon_m^t}. \quad \textbf{(4-20)}$$

The first component, $e^{1-\varrho_m^t}$, drive each tenant agent to take better actions that drive the state-action values to converge towards the satisfaction of the SLA of tenant *m*.

$$\varrho_m^t = 1 - \log\left(\frac{\sum_{n \in \mathcal{N}} \eta_{n,m}}{\sum_{n \in \mathcal{N}} \chi_{n,m}}\right), \quad \text{if } \sum_{n \in \mathcal{N}} \chi_{n,m} < N O_n \quad \textbf{(4-21)}$$

$$\varrho_m^t = 1 - \log\left(\min\left(\frac{\sum_{n \in \mathcal{N}} \eta_{n,m}}{\sum_{n \in \mathcal{N}} \min\{\zeta_{n,m} + \delta_{n,m}, \chi_{n,m}\}}, 1\right)\right), \quad O.W \quad \textbf{(4-22)}$$

where $\delta_m^t$ denotes the guaranteed capacity that is not required by other tenants at time *t*.

$$\delta_m^t = \sum_{m' \in \mathcal{M}-m} \max(Y_{n,m'}^{max} - \chi_{m'}^t, 0) \quad \textbf{(4-23)}$$

which reaches $\delta_{m'}^t = 0$ when the offered load $\chi_m^t$ of all the tenants is higher or equal than their $Y_{n,m'}^{max}$.

The second the reward component, $e^{1-\upsilon_m^t}$, aims at measuring the degree of capacity over-provisioning. It is defined by the ratio between the system throughput provided to the tenant *m*, $\eta_m^t$ and its provided capacity during the last time step, that is:

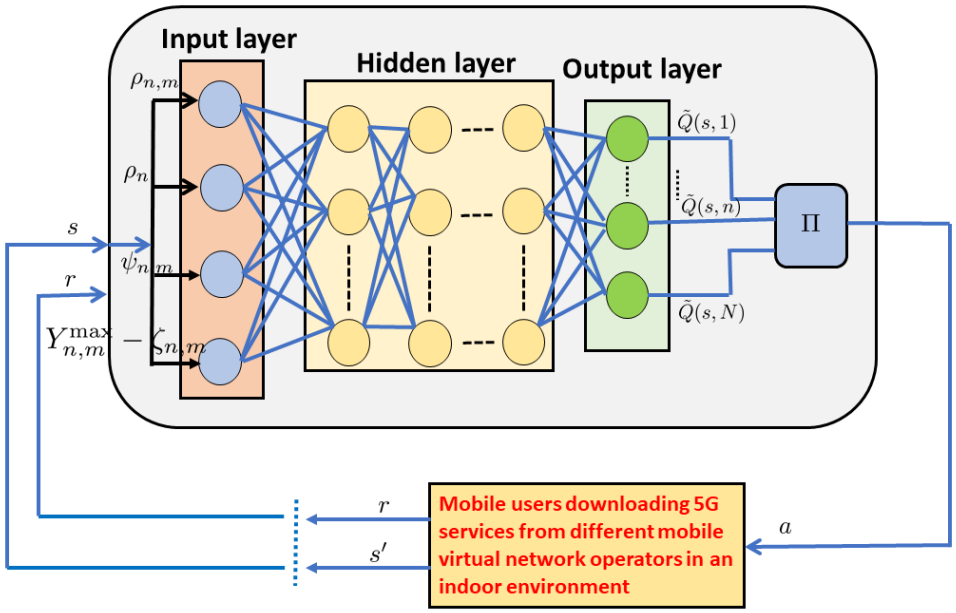



**Figure 4-2 Interaction of the DQN-based SDN controller with the sliced LiFi attocellular network environment**

$$\boldsymbol{v_m = 1 - \log\left(\frac{\eta_m^t}{\sum_{n\in\mathcal{N}} Y_{n,m}^{max}}\right)} \qquad \textbf{(4-24)}$$

The action-value function Q (s, a) tells how valuable it is to take a particular action *a* from the state *s*. It represents the expected total reward which the DQN agents at the SDN controller may get after taking the particular action *a* from the state *s* onward. The action-value functions are given by the Bellman expectation Equation as follows [49]

$$\boldsymbol{Q(s,a) = E(G^t|s,a) = E[r^t + \gamma Q(s',a')|s,a]}, \qquad \textbf{(4-25)}$$

where the action value Q(*s, a*) are recursively presented in terms of the immediate reward $r^t$ and the discounted value of the successor state-action Q(s' , a'), respectively. *a'* denotes the next action at the next state *s'*. A policy $\pi$ is a way of selecting actions. It can be viewed as a mapping from states to actions as it describes the set of probabilities for all possible actions to select from a given state. To solve the considered MDP problem the DRL agent needs to find the optimal policy through finding the optimal action-value function $Q^\star(s,a) = \max_\pi Q_\pi(s,a)$ for all state-action pairs. The Bellman optimality for $Q^\star(s,a)$ can be written as:

$$\boldsymbol{Q^\star(s,a) = E\left[r^t + \gamma \max_{a'\in\mathcal{A}} Q^\star(s',a')\middle|s,a\right].} \qquad \textbf{(4-26)}$$

The optimal policy is simply taking the best local actions from each state considering the expected rewards. The optimal policy can be learnt by solving the Bellman optimality $Q^\star(s,a)$ for $a^\star$, which can be challenging. DQN approaches address this problem by approximating the state-action value functions, $Q^\star(s,a)$ , by using deep neural networks (DNNs). The DQN agent is basically a DNN that consists of an input layer, hidden layers, and an output layer. The number of neurons in the input and output layers is equal to the state and action dimensions, respectively. Whereas the number of hidden layers and the number of neurons in each hidden layer are design parameters to be chosen. The system state *s* is fed to the DNN as an input, which then regularly updates its parameters, mainly the weights of all connections between neurons. This enable the DNN to predict the Q-values at the output for a given input state, as explained in Algorithm depicted in Figure 4-3. This enables the DQN agents to learn the optimal policy from the multi-tenant SDN-enabled sliced LiFi network shown in Figure 4-1, which supports URLLC and eMBB network slices. The DQN model design parameters are summarized in Table 4-3 .

**Initialize**:$\gamma = 0.9$, $\epsilon = 0.5$, $\forall \gamma, \epsilon \in [0, 1]$
Set the exploration probability by initially setting $\epsilon$ to 0, while sampling a probability $p$ from $[0, 1)$ at each time step for $\epsilon$-greedy.
Create DNN model and target model with weight $\omega$ and $\hat{\omega}$, respectively
**Initialize**: $\omega$ and $\hat{\omega}$, replay memory $\digamma$, total time-steps $TR$, $s$
**for** $t \leftarrow 1$ *to* $T_t$ **do**
  **for** $m \leftarrow 1$ *to* $M$ **do**
    **for** $n \leftarrow 1$ *to* $N$ **do**
      - Obtain the current observation $s^t_{n,m}$;
      - Take action $a^t_{n,m}$ according to $\pi = \epsilon-$greedy.
      - Observe at the end of the scheduling period $t$, get the next observation $s^{t+1}_{n,m}$ and reward $r^t_m$ from environment;
      - Append the observation $(s^t_{n,m}, a^t_{n,m}, r^t_m, s^{t+1}_{n,m})$ to $\digamma$;
      - $s \leftarrow s^{t+1}_{n,m}$;
      - Sample a random mini-batch of $z$ observations from $\digamma$;
      **for** $i \leftarrow 1$ *to* $z$ **do**
        predict $\hat{Q}_i(s_i|\hat{\omega})$ ;

$$z_i = \begin{cases} r_i, & \text{if t+1=}T_t. \\ r_i + \gamma \max_{a'} \hat{Q}_i(s', a'|\hat{\omega}), & \text{O.W.} \end{cases}$$

        Fit the DNN model for $(s_i, z_i)$ by applying gradient descent step on $(z_i - \hat{Q}_i)^2$ with respect to $\hat{\omega}$.
      **if** $(t\%\tau == 0)$ **then**
        $\hat{\omega} \leftarrow \nu\omega + (1 - \nu)\hat{\omega}$
      **if** $\omega$ *converges* **then**
        $\omega^\star \leftarrow \omega$
        break;
  Compute $\psi_m(t)$, $\forall m \in \mathcal{M}$ by applying the optimal bandwidth allocation procedure for the network slices in the cells.
$\hat{\omega}$ is used to estimate $\hat{Q}(s, a)$ required for $\hat{\pi}$ based on BAP (Best Action Policy)

**Figure 4-3 DQN Training for Multi-agent Learning Optimum Policy**

The algorithm procedure follows these steps:

- Initialize DQN parameters
- Maintain current and previous observations in replay memory, and training samples collection in mini batch of size **z**
- Use the DNN target model to stabilize the DNN model by reducing the correlation between the action-values Q(s, a) and the targets
- Take an action, in each iteration, according to a policy π based on the ϵ-greedy policy, where the collected reward and the successor state are observed

- Estimate the output vector $\hat{Q}$ of the target model for a given input state *s* in each experience sample by using the target model weights $\hat{w}$
- Update the model weights ω by fitting the model for the input states and the corresponding targets
- Update periodically the target model weights every τ time steps
- Stop the algorithm when the DNN model weights ω converge.

**Table 4-1 LiFi AP Channel Simulation Parameters**

| Parameter | Description | Value |
|---|---|---|
| $\phi_{1/2}$ | LED half power semi angle | $60^{\circ}$ |
| $\eta_w$ | Reflectivity factor of wall | 0.8 |
| $g_f$ | Gain of optical filter | 1 |
| $A_R$ | Physical area of photo detector (PD) | $10\ mm^2$ |
| $\eta_f$ | Reflectivity factor of floor | 0.8 |
| $N_0$ | Noise power spectral density | $10^{-21}\ A^2/Hz$ |
| B | Modulated bandwidth | 20 MHz |
| $\eta_c$ | Reflectivity factor of ceiling | 0.8 |
| $\Psi_c$ | Receiver field of view (FOV) | $90^{\circ}$ |
| $P\_t$ | Transmission power | 9 W |
| $\zeta$ | Refractive index | 1.5 |
| $n_r$ | PD's orientation vector | [0, 0, 1] |
| $R_{PD}$ | PD responsivity | 0.5 A/W |
| $n_t$ | AP's orientation vector | [0, 0, -1] |

**Table 4-2 Services and Traffic Simulation Parameters**

| Parameter | URLLC | eMBB |
|---|---|---|
| $d_m$ (delay) (ms) | [0.5, 3] | [10, 40] |
| $[\eta^n_{m,min}, \eta^n_{m,max}]$(Mbps) | [7, 13] | [20, 35] |
| $\lambda_m$ (packets/ms) | 3 | 2 |
| $L_m$ (bytes) | [85, 150] | [1250, 1450] |
| $\lvert\mathcal{K}^n_m\rvert$ per attocell | 4 | 6 |

**Table 4-3 DNN Training and Simulation Setup**

| Parameter | Description | Value |
|---|---|---|
| $\mathcal{N}$ | Number of LiFi APs | 4 |
| $\mathcal{M}$ | Number of Tenants | 2, 3 |
| $\mathcal{O}_n$ | half of sub-carriers per AP | 128 |
| $B_n$ | AP downlink BW | 20 MHz, 48 Mbps |
| $B^a_n$ | Aggregate APs downlink BW | 80 MHz, 192 Mbps |

| | | |
|---|---|---|
| Γ | capacity of DNN replay memory | 2000 |
| γ | discount factor | 0.9 |
| α | learning rate | 0.01 |
| ϵ | Probability of random action | 0.05 |
| $z$ | batch size | 64 |
| τ | $\widehat{\omega}$ update interval | 1000 |
| ν | update rate | 0.3 |
| $TR$ | Maximum number of training steps | $2 \times 10^5$ |
| $\Delta t$ | Time step duration | 180 s |
| Δ | Action step | 0.05 s |
| $r_1$ | reward weight for tenant 1 | 0.5 |
| $r_2$ | reward weight for tenant 2 | 0.4 |
| DNN Configuration | input/hidden/output layers | $16 \times (64 \times 64) \times 8$ |

### 4.1.6 Performance evaluation

The performance of the proposed DRL-based autonomic LiFi attocellular network slicing (DQN) is evaluated in Python and the result graphs are produced in MATLAB. The network model is shown in Figure 4-1. The DQN training model and the traffic and services simulation parameters are summarized in Table 4-3 and Table 4-2. The fixed network slicing (FNS) and the utility scheduler-based network slicing (UBNS) [45] approaches are served as baselines to provide a performance comparison to the proposed DQN and UBNS approach. The fixed network slicing approach equally shares the downlink bandwidth of LiFi APs between the supported URLLC and eMBB network slices.

The first scenario evaluates the capability of the proposed DQN model to autonomically match the offered traffic load with their allocated bandwidth subject to the total network capacity [48]. The allocated bandwidth to tenant, $m$, at time step t, $Y_m^a$ , is defined as the total resource quota (data rate) allocated to match the offered load received by their users associated to the different network APs. As a result, the trained DQN model reacts efficiently to the offered traffic load of the eMBB and URLLC network slices, according to their aggregate data rate across the network, as shown in Figure 4-3. The training phase results in DQN agent learning a policy which they use to achieve a particular target data rate of their tenant in the evaluation phase. The model follows the DQN training process in [48]. It trains the different tenant DQN models (agents), considering various SLA parameters in terms of the maximum and minimum quota guarantees of the different slices under different traffic loads. After training the DQN models, the different agents may use their policy or inherent each other's policy. This allows new tenants to share with them the spectrum without the need to pass through the same training process. The allocated bandwidth of both URLLC and eMBB are adapted to the offered traffic load, as shown in Figure 4-4.

The second scenario evaluates the total average reward of eMBB and URLLC network slices (tenants) in terms of the average throughput and delay service satisfaction across the network. The DQN+UBNS approach achieves a noticeable performance improvement in system utility compared with the only UBNS and FNS approaches. The proposed pre-computation performance monitoring module and information sharing between the different agents of network slices improve the actions in terms of applying the increment to the bandwidth of network slices at the different APs. Obviously the UBNS schedule traffic packets with aim to maximize their utility. It satisfies the SLA for URLLC subscribers while not decreasing the SLA of other subscribers. The pre-computation module enables the DQN model to predict the next actions to apply to the shares of network slices, which contributes to increasing their rewards in terms of the service satisfaction.

It pursues higher throughput performance as shown in **Figure** 4-7. The DQN+UBNS approach improves the network utility while efficiently accommodating the offered traffic offered load.

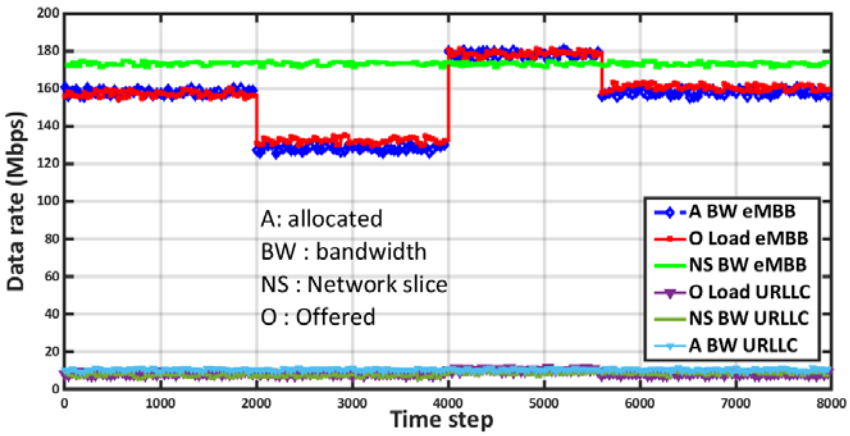



**Figure 4-4 DQN enabled-autonomic matching of the Offered load with their allocated bandwidth for the eMBB and URLLC network slices in Mbps**

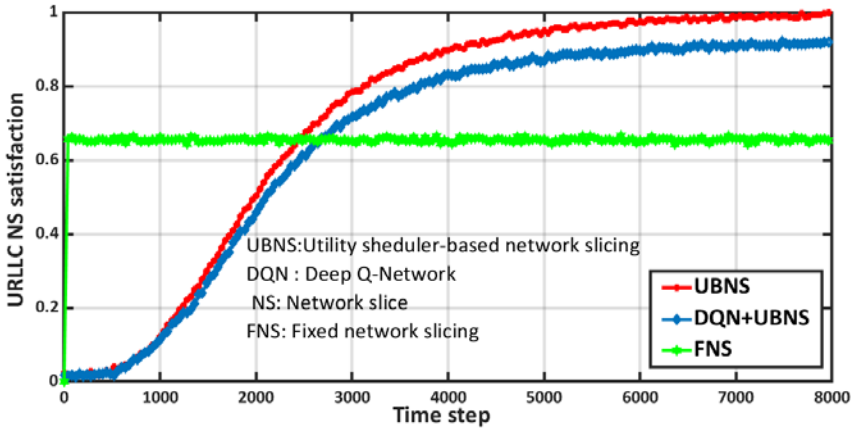



**Figure 4-5 URLLC network slice service satisfaction ratio**

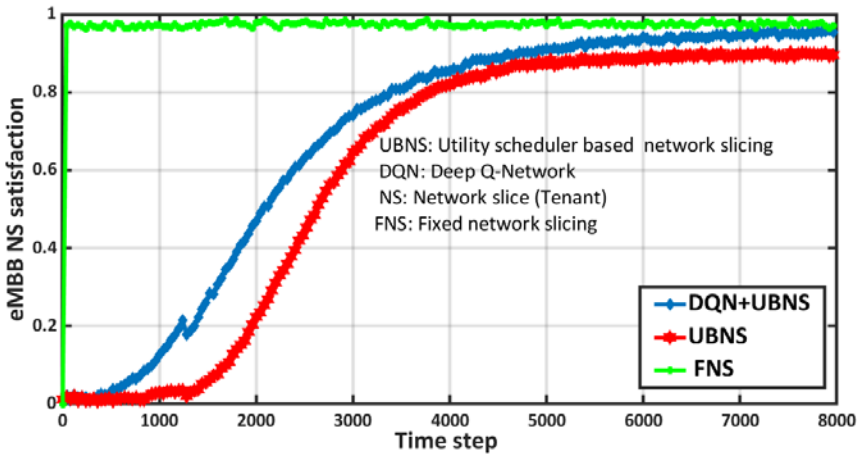



**Figure 4-6 eMBB network slice service satisfaction ratio**

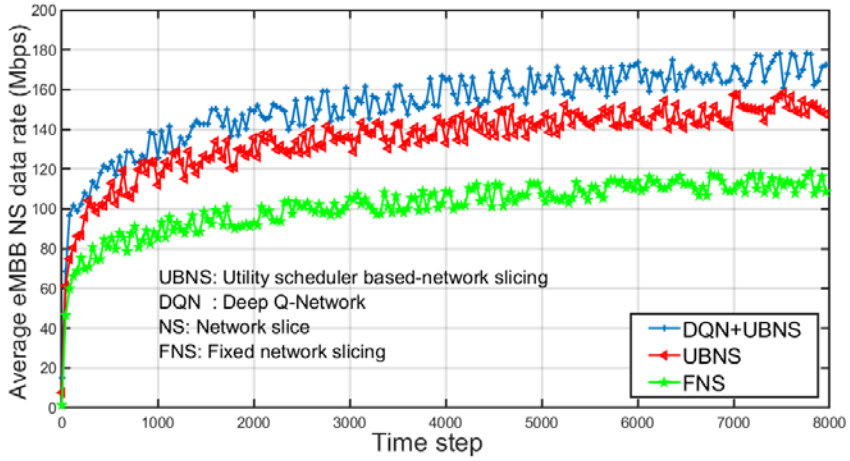



**Figure 4-7 Average eMBB network slice data rate in Mbps**

In conclusion, the proposed DRL algorithm uses the information collected from the LiFi network APs to identify the patterns of varying service and traffic demands in temporal and spatial space across the network. This information is pre-processed by the pre-computation module and then fed to the DRL algorithm. This automatically allocates the resource quota for each network slice across the network, which results in an intelligent resource management strategy for the network resource.

## 4.2 Network-wide slicing of Wi-Fi networks with variable load in space and time

### 4.2.1 Motivation

In 5G-CLARITY D3.2 [2] , we presented a mechanism to support slicing on Wi-Fi networks, which works by provisioning slices as different SSIDs radiated from the same access point (AP), and allocating an airtime percentage to each SSID in each AP. In this section, we propose a centralized scheduling policy that leverages the mechanism described in 5G-CLARITY D3.2 [2] to tune the airtime weights of a set of APs distributed over a geographical area. The intuition behind our approach is that the load offered to the slice over the geographic area will vary in space and time, and therefore the optimal airtime weights in each AP should be updated accordingly. We call our scheduling policy *Global-Airtime Deficit Round Robin (G-ADRR)*.

The work presented here relates to the ML driven multi-tenant slice RRM algorithm presented in 5G-CLARITY D4.2 [36]. The difference with that work is that our approach is focused on Wi-Fi networks, instead of 5G, and is not based on ML, but rather on a quadratic programming framework.

Finally, the content of this section is a summarized version of the full work published in [50].

### 4.2.2 G-ADRR design

#### 4.2.2.1 System model

Figure 4-8(a) depicts the considered system architecture, where we introduce the following terminology:

- Let A = {1, ..., |A|} be the overall set of APs managed by a centralized Wi-Fi controller, where the GADRR policy is executed. APs are assumed to be configured in orthogonal channels to avoid cross-AP interference.
- Let S = {1, ..., |S|} be the set of slices active in the system, where each slice represents a tenant
- Let As ∈ A be a sub-set of APs where slice s ∈ S is configured.

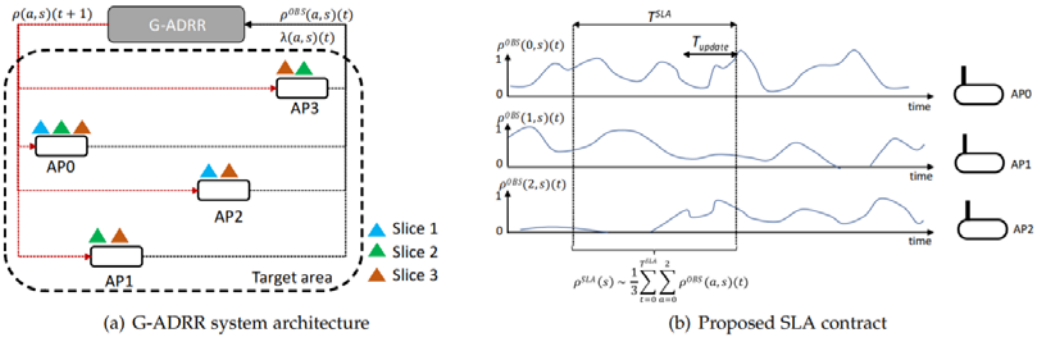



**Figure 4-8 G-ADRR system model (a) and SLA contract (b)**

For example, we can see in Figure 4-8(a) a deployment with 4 APs and 3 slices, where slice 1 is configured in AP0 and AP2, slice 2 in AP0, AP1 and AP3, and slice 3 is configured in all APs. Each AP $a \in A$ implements a local work-conserving airtime based slicing policy, where the airtime resource is allocated to a slice $s \in S$ using a weight parameter $0 \leq \rho a,s(t) \leq 1$, which can be updated periodically. We refer to this local policy as Airtime DRR (ADRR), although alternative work-conserving queuing disciplines, not necessarily based on DRR, could be also considered at the AP level.

In addition to scheduling traffic from the different slices according to the configured weight parameters, each AP $a \in A$ continuously monitors two additional parameters, which are periodically reported to the centralized Wi-Fi controller:

- $\lambda_{a,s}(t)$: The normalized offered load defined as the airtime that would have been required in the last reporting period to serve the overall traffic (bytes) received from slice $s \in S$ in AP $a \in A$, including both uplink and downlink. Notice that $\lambda a,s(t) \geq 0$ and it can grow bigger than 1.
- $\rho^{OBS}_{a,s}(t)$: The actual observed airtime utilization incurred by slice $s \in S$ in AP $a \in A$ during the last reporting period, including both uplink and downlink, where $0 \leq \rho^{OBS}_{a,s}(t) \leq 1$.

Individual APs report periodically the collected ($\lambda_{a,s}(t)$, $\rho^{OBS}_{a,s}(t)$) measurements to the centralized Wi-Fi controller. Based on these measurements, every $T_{update}$, a global scheduling policy, which we refer to as *Global Airtime Deficit Round Robin (G-ADRR)* computes the airtime weights of each slice in each AP, i.e. $\rho_{a,s}(t + 1)$, which are then communicated to the individual APs and applied during the next period. G-ADRR computes these weights in order to adhere to an SLA contract, which is defined next.

### 4.2.2.2 SLA contract definition

An SLA contract is defined for each slice $s \in S$ in the system as a pair ($\rho^{SLA}_s$, $T^{SLA}$), where:

- $\rho_{min} \leq \rho^{SLA}_s \leq \rho_{max}$ represents a fraction of the total system resources (airtime) that are reserved for this slice over the set of APs where this slice is configured, i.e., $A_s$. $\rho_{min}$ represents a minimum share to avoid that a slice is locked out from the system, and $\rho_{max}$ represents the maximum channel capacity.
- $T^{SLA}$ defines a sliding window in time over which $\rho^{SLA}_s$ needs to be delivered. $T^{SLA}$ is common for all slices.

Figure 4-8(b) illustrates the proposed SLA contract definition, where we can see the actual airtime $\rho_{a,s}(t)$ consumed by an exemplary slice across three different APs. To assess the level of SLA compliance of a given airtime allocation, the following steps are necessary:

- Compute the average allocation for this slice across all APs, i.e. $\hat{\rho}^{OBS}_s(t) = 1/|As| \sum_{a \in As} \rho^{OBS}_{a,s}(t)$.

- Compute a running average of the airtime allocation for slice s within a window of duration $T^{SLA}$ as $\hat{\rho}_s^{window}(t) = \hat{\rho}_s^{OBS}(t) * 1/T^{SLA}\, w(t)$, where w(t) is a square signal of duration $T^{SLA}$ and $*$ denotes the convolution operator.
- Compare $\hat{\rho}_s^{window}(t)$ to the SLA level defined for that slice, i.e. $\rho^{SLA}_s$.

In the case where a slice is configured in a subset of APs, $\rho^{SLA}_s$ refers to the allocation that the slice should receive across the set of configured APs. Thus, a normalized SLA can be defined as $\hat{\rho}^{SLA}_s = \rho^{SLA}_s |As|/|A|$ to compare SLA allocations across slices that are configured (or active) in different subsets of APs. Finally, we assume in our system model that slices are provisioned to use the full capacity of the system, i.e. $\sum_{s\in S} \hat{\rho}^{SLA}_s = \rho_{max}$.

#### 4.2.2.3 G-ADRR Cost Function

The goal of G-ADRR is to drive the actual allocation of airtime resources across the network for each slice s ∈ S over the time window $T^{SLA}$, as close as possible to the guaranteed SLA shares, i.e. $\rho^{SLA}_s$.

Let us define the average resource consumption of a slice across the network during period t as:

$$\hat{\rho}_s^{OBS}(t) = \frac{1}{|\boldsymbol{A_s}|} \sum_{a\in \boldsymbol{A_s}} \rho_{a,s}^{OBS}(t)\,, \quad \textbf{(4-27)}$$

which represents an average across the subset of active APs (As) of the actual resource allocations for slice s ∈ S during the last reporting period t. Given $\hat{\rho}^{OBS}_s(t)$, GADRR maintains a moving average of the resource consumption for each slice using an Exponential Weighted Moving Average (EWMA) defined as:

$$\rho_s^{ewma}(t) = \epsilon \hat{\rho}_s^{OBS}(t) + (1-\epsilon)\rho_s^{ewma}(t-1) \quad \textbf{(4-28)}$$

The parameter ϵ in the EWMA filter is linked to T SLA in the following way. In an EWMA filter, two samples separated by k observations have a weight ratio of r(t, t − k) = 1 (1− ϵ) k . Hence, we can define an EWMA effective time window, i.e. $T_{ewma}$, as $T_{update}$ multiplied by the number of samples that significantly contribute to the current value of the filter, i.e. samples such that r(t, t − k) ≤ $r_{thr}$. Thus, $T_{ewma} = T_{update} - \log r_{thr} \log(1-\epsilon)$ . By Taylor expansion and assuming that ϵ takes a small value (since $T^{SLA} >> T_{update}$), $T_{ewma} \approx T_{update} \log r_{thr}/\epsilon$ , which for $r_{thr} = 10$ results in $T^{SLA} \approx T_{update}/\epsilon$ , meaning that the parameter ϵ will allow to tune G-ADRR to the $T^{SLA}$ window, with a small ϵ resulting in G-ADRR averaging slice utilization over large time windows and vice-versa.

Armed with the moving average $\rho^{ewma}_s(t)$ and with $\rho^{SLA}_s$ ∀s ∈ S, G-ADRR decides how to allocate the per-slice airtime shares at each AP in the next period, i.e. $\rho_{a,s}(t+1)$ ∀a ∈ A, ∀s ∈ S, by attempting to minimize the mean square error (MSE) between the ideal resource share and the actual average resource allocation for each slice. Particularly, G-ADRR uses the following cost function:

$$\begin{aligned} \underset{\rho_{a,s}(t+1)}{\operatorname{argmin}} \; J_1(\rho_{a,s}(t+1)) = \\ = \frac{1}{|\boldsymbol{S}|} \sum_{s\in \boldsymbol{S}} (\rho_s^{ewma}(t+1) - \rho_s^{SLA})^2 = \\ = \frac{1}{|\boldsymbol{S}|} \sum_{s\in \boldsymbol{S}} (\epsilon \hat{\rho}_s^{OBS}(t+1) + (1-\epsilon)\rho_s^{ewma}(t) - \rho_s^{SLA})^2\,, \end{aligned} \quad \textbf{(4-29)}$$

where the term 1/|S| is constant and can be omitted, and $(1-\epsilon)\rho^{ewma}_s(t) - \rho^{SLA}_s$ is also constant at the end of the reporting period t, where our target is to determine the airtime weights for period t + 1.

In order to proceed, we need to express the cost function $J_1$ as a function of $\rho_{a,s}(t+1)$. The key assumption

in GADRR is to assume that $\rho^{OBS}_{a,s}(t+1)$, used to compute $\hat{\rho}^{OBS}_{s}(t+1)$ in Eq.(4-29) , equals $\rho_{a,s}(t+1)$, which implies that in period t + 1 the local airtime schedulers in each AP will schedule airtime exactly as instructed by G-ADRR through the $\rho_{a,s}(t+1)$ weights. Under this assumption we can proceed to express the cost function as:

$$\underset{\rho_{a,s}(t+1)}{\operatorname{argmin}} \; J_1(\rho_{a,s}(t+1)) = \sum_{s \in \boldsymbol{S}} \Big( \frac{\epsilon}{|\boldsymbol{A_s}|} \sum_{a \in \boldsymbol{A_s}} \rho_{a,s}(t+1) + \\ + (1-\epsilon)\rho_s^{ewma}(t) - \rho_s^{SLA} \Big)^2 \qquad \textbf{(4-30)}$$

Regarding the previous assumption, notice that even though the local slicing policy in each AP is controlled by G-ADRR, the local scheduler in the AP may need to redistribute airtime among slices to be work conserving. This is the case if some slice does not offer enough load to fill its reserved share, which would result in $\rho^{OBS}_{a,s}(t+1)$ not matching $\rho_{a,s}(t+1)$. This uncertainty is managed by G-ADRR in two ways: i) the objective function described in Eq.(4-30) is constrained by a forecast of the load offered by each slice in period t + 1, and ii) the mismatches that occur between the shares decided by G-ADRR and the actual allocation performed by each AP will be reported at the end of the period through $\rho^{OBS}_{a,s}(t+1)$, and hence will be automatically accounted for in the next period.

The cost function $J_1$ described in Eq. (4-30) is constrained in the following way:

$$\sum_{s \in \boldsymbol{S}} \rho_{a,s}(t+1) = \rho_{max}$$
$$\rho_{min} \leq \rho_{a,s}(t+1) \leq \hat{\lambda}_{a,s}(t+1)\,, \qquad \textbf{(4-31)}$$

where Eq.(4-31) ensures that all the resources available at each AP are fully allocated, and (4-31) (below) limits the allocated resources for slice $s \in S$ in AP $a \in A$ to a forecast of the load offered by this slice in this AP during the next period t + 1. Constraining the airtime allocations to the actual traffic demands allows G-ADRR to free available resources that can be used to balance under-served slices, and reduces the mismatch between $\rho^{OBS}_{a,s}(t+1)$ and $\rho_{a,s}(t+1)$. G-ADRR uses the periodic $\lambda_{a,s}(t)$ reports to forecast the offered load demand in the next period, i.e. $\hat{\lambda}_{a,s}(t+1)$. To avoid that a slice is locked out from the system, we set a minimum value for the expected load as $\hat{\lambda}_{a,s}(t+1) \geq \hat{\lambda}_{min}$. Estimators available in the literature such as [51] can be used to forecast the per-slice load in the next period.

There is an additional aspect to consider related to the constraints of cost function $J_1$. If the offered load from all slices in one AP is limited, i.e. $\hat{\lambda}_{a,s}(t+1)$ is small in Eq.(4-31) (below), then it may not be possible to fulfil the constraint of full resource utilization in that AP (Eq. (4-31) (above)). Thus, G-ADRR incorporates a heuristic that normalizes the per-slice load forecasts in each AP to ensure that Eq.(4-31) (above) can always be fulfilled. This heuristic is described in Algorithm 1, where it can be seen how the deficit load required to saturate an AP is added to the expected load of each slice active in that AP, weighted by $\rho^{SLA}/\sum_{s \in S}\{\rho^{SLA}_s \,|\, a \in A_s\}$. Again, it is worth noting that the local slicing policy at each AP will redistribute airtime among slices in order to be work conserving if $\hat{\lambda}_{a,s}(t+1)$ is overestimated, and any mismatch will be accounted for in subsequent iterations of G-ADRR if $\hat{\lambda}_{a,s}(t+1)$ is underestimated.

**Algorithm 1:** Load normalization heuristic: Redistribute deficit load according to slice share

**Input:** $\hat{\lambda}_{a,s}(t+1)$ given by forecast, $\rho_s^{SLA}$
**Output:** $\hat{\lambda}_{a,s}(t+1)$ to be used in Equation 6

**for** $a \in \boldsymbol{A}$ **do**
  $forecast\_load = \sum_{s \in \boldsymbol{S}} \{\hat{\lambda}_{a,s}(t+1) | a \in \boldsymbol{A_s}\}$;
  **if** $forecast\_load < 1$ **then**
    $deficit\_load = \rho_{max} - forecast\_load$;
    **for** $\{s \in \boldsymbol{S} \mid a \in \boldsymbol{A_s}\}$ **do**
      $share = deficit\_load \times \frac{\rho_s^{SLA}}{\sum_{s \in \boldsymbol{S}} \{\rho_s^{SLA} | a \in \boldsymbol{A_s}\}}$;
      $\hat{\lambda}_{a,s}(t+1) = \hat{\lambda}_{a,s}(t+1) + share$;
    **end**
  **end**
**end**

**Algorithm 1. Load normalization heuristic procedure**

The cost function $J_1$ described in Eq.(4-29) drives the average airtime allocation across all APs for a given slice close to its SLA allocation, thus compensating load variations in space and time, but it does not consider the internal balance, or fairness, between the allocations in different APs for a given slice. For example, if in a system with |A| = 2 APs, a slice needs to be allocated 30% of the resources, the cost function $J_1$ may regard as equally valid a solution that i) allocates 50% of the resources of AP 1 and 10% of the resources of AP 2 to this slice, or ii) allocates 30% of the resources in each AP. However, if enough offered load is available in each AP, we regard the second solution, which is more balanced, as preferable. Hence, we add a second term to the G-ADRR cost function, $J_2$, which takes care of balancing the resource allocation for a given slice across the APs in the system. A basic cost function to reduce intra-slice variance across all slices can be formulated as follows:

$$\underset{\rho_{a,s}(t+1)}{\operatorname{argmin}} \; J_2(\rho_{a,s}(t+1)) = \sum_{s \in \boldsymbol{S}} \sum_{a \in \boldsymbol{A_s}} (\rho_{a,s}(t+1) - \\ - \frac{1}{|\boldsymbol{A_s}|} \sum_{i \in \boldsymbol{A_s}} \rho_{i,s}(t+1))^2 \qquad (4\text{-}32)$$

The cost function described in Eq. (4-32) is proportional to the variance of the weights of a given slice across the APs where this slice is active, thus favouring similar weights for a slice across the APs. This cost function though does not consider the fact that a given slice may experience a different amount of offered load in each AP. Hence, we propose a normalized version of this cost function that reduces intra-slice variance while considering asymmetric loads:

$$\underset{\rho_{a,s}(t+1)}{\operatorname{argmin}} \; J_2^{norm}(\rho_{a,s}(t+1)) = \sum_{s \in \boldsymbol{S}} \sum_{a \in \boldsymbol{A_s}} (\frac{\rho_{a,s}(t+1)}{\hat{\lambda}_{a,s}(t+1)} - \\ - \frac{1}{|\boldsymbol{A_s}|} \sum_{i \in \boldsymbol{A_s}} \frac{\rho_{i,s}(t)}{\hat{\lambda}_{i,s}(t+1)})^2, \qquad (4\text{-}33)$$

where $\hat{\lambda}_{a,s}(t+1)$ is the forecasted offered load for slice $s \in S$ in AP $a \in A_s$ for the next period $t+1$, output by Algorithm 1.

Therefore, the final cost function that is periodically optimized by G-ADRR to decide on the per-slice and per AP airtime share allocations in the next period is obtained combining Eq.(4-30) and Eq. (4-33) as:

$$\underset{\rho_{a,s}(t+1)}{\operatorname{argmin}} J_{G-ADRR}(\rho_{a,s}(t+1)) = \\ = J_1(\rho_{a,s}(t+1)) + \gamma J_2^{norm}(\rho_{a,s}(t+1)), \quad (4\text{-}34)$$

constrained by Eq. (4-31) , where γ is a scalar factor used to scale $J^{norm}_2$ to a similar range than $J_1$.

In [50] it is shown that the G-ADRR cost function can be solved using Quadratic Programming in polynomial time.

### 4.2.3 G-ADRR Performance evaluation

#### 4.2.3.1 G-ADRR testbed

Figure 4-9 depicts the architecture of our G-ADRR prototype that consists of two main components: i) the physical access points (APs) including a local airtime-based scheduling function, and ii) a custom Radio Access Network (RAN) Controller that implements G-ADRR and allows us to remotely manage the per-slice weights allocated in each AP.

Physical APs in our testbed are implemented using Single Board Computers (SBC) from PC Engines (apu4d4model) running Linux kernel 5.5 and featuring IEEE 802.11n mini PCIe wireless cards from Qualcomm (Compex WLE200NX). Within the physical APs we implemented a Local Scheduler Agent (LSA) and a Telemetry exporter. The LSA includes the airtime based DRR scheduling system presented in [52], which modifies the ath9k and mac80211 kernel modules in [52] required to control Qualcomm Atheros 802.11n modems, and hostapd [53], which is an open access point implementation for Linux. hostapd allows to instantiate multiple virtual APs (vaps) over a physical AP, which is the feature we use to slice the Wi-Fi network, having a separate hostapd instance running per slice.

To control the airtime allocated to a slice, the LSA keeps track of the aggregated airtime assigned to all the active stations attached to the vap representing that slice through the mac80211 API (using the nl80211 driver). In order to enable global control of the slicing policies, we implement a python program that interacts with the different vap instances in a given AP and allows to dynamically modify the airtime of each slice exposing the scheduling weights through a REST API. As in [54], within a slice, airtime is distributed equally among the active stations by monitoring periodically (each 200 ms) the status of mac80211 queues. The default airtime quantum per station is 256μs.

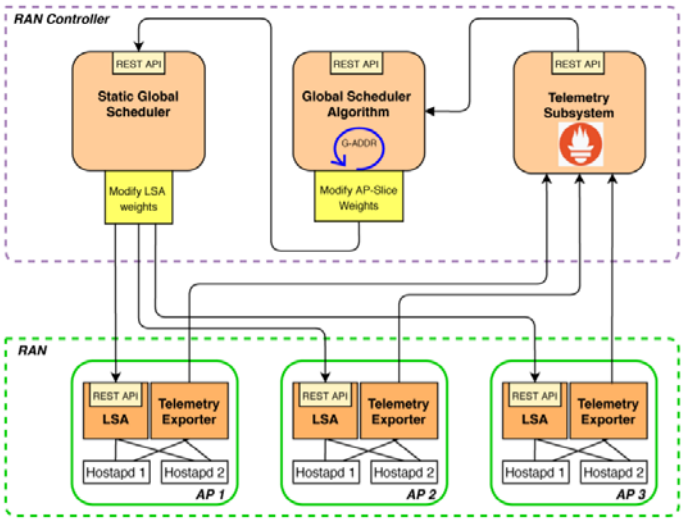



**Figure 4-9 G-ADRR experimental testbed**

The LSA maintains the relationship between the weights of the slices defined by G-ADRR, and the individual weight delivered to the active stations in each slice. Notice that given that the LSA implements a DRR policy, when one slice does not present enough load to consume its allocated resources, these resources are automatically distributed among the other slices.

The Telemetry exporter within a physical AP is implemented using a custom Prometheus [55] hostpad exporter that gathers the telemetry that G-ADRR requires to take scheduling decisions. In particular, the exporter delivers periodically the offered load, i.e. $\lambda_{a,s}(t)$, and the consumed airtime, i.e. $\rho^{OBS}_{a,s}(t)$, for each station in each slice (vap). To compute $\lambda_{a,s}(t)$ we consider the average Modulation and Coding Scheme (MCS) observed across the transmissions to/from users of slice $s \in S$ in AP $a \in A$. In our prototype implementation APs report these measurements every second. Our Prometheus hostpad exporter is available in [56].

To control the per-slice weights in each physical AP the RAN Controller depicted in Figure 4-9 is composed of: i) the Static Global Scheduler (SGS), ii) the Global Scheduler Algorithm (GSA) and iii) the Telemetry subsystem (TS). The RAN Controller is in a remote Virtual Machine (VM) and the physical APs are connected to it through Virtual Private Network (VPN) tunnels.

The SGS interacts with the API exposed by the LSA in each physical AP, implementing two main features. First, it allows to assign a global static weight for a slice, thus enforcing that weight into all APs where that slice is active. This is the feature we have used to implement a static global scheduling policy that we use as benchmark against GADRR. Second, the SGS module exposes a REST endpoint that allows an external module to modify the airtime allocated to a slice in a specific AP. This feature is used by the Global Scheduler Algorithm to enforce weights for each slice in each AP.

The Telemetry Subsystem in the controller is based on a Prometheus server that retrieves the metrics exposed by the hostapd exporter in each physical AP. These metrics are stored in a time-series database and can be queried by the GSA module. By means of Prometheus labels and operations, the monitored data can be aggregated per AP and per slice, as required by G-ADRR. Since the airtime metric values provided by hostapd exporter in each AP are cumulative, we use the rate operation from Prometheus to compute the percentage of airtime consumption during the last period. We use a period of Tupdate = 4 seconds, i.e., considering four reporting samples from the hostapd exporters, which we validate to deliver stable airtime estimates. Given that our target SLA is $T^{SLA}$ = 100 seconds and that in GADRR $T^{SLA} \approx$ Tupdate/$\epsilon$, we set $\epsilon$ = 0.04.

Finally, the GSA module implements G-ADRR using the python quadprog solver based on the Goldfarb/Idnani dual algorithm [57], which is the algorithm presenting the best performance among the solvers available in the python qpsolvers package [58]. The GSA module periodically retrieves the monitoring data from the Telemetry Subsystem, feeds them to the G-ADRR algorithm, obtains the per-slice and per-AP weights for the new period and updates the LSA schedulers in the physical APs via the SGS module. The GSA module implements a REST API used to provide the initial configuration and expects a set of slices and airtime weights to start its execution.

#### 4.2.3.2 G-ADRR experimental validation against a static scheduling policy

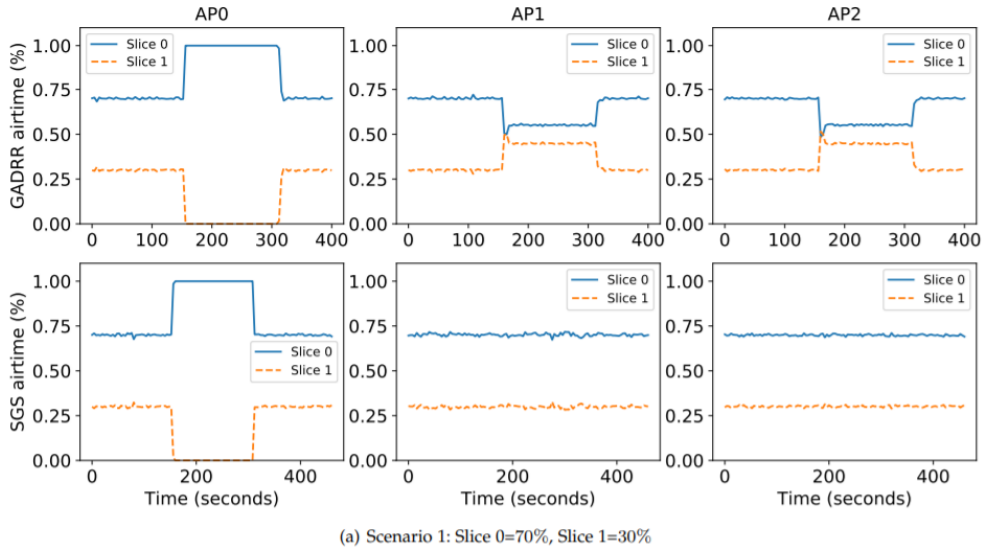
In this section we compare G-ADRR against a Static Global Scheduler (SGS) that allocates the same weight for a slice in all APs, i.e. $\rho_{a,s} = \rho^{SLA}_{s}$ $\forall a \in A$. For this purpose, we setup an indoor testbed consisting of 3 APs each hosting 2 slices, and we have one client per slice, i.e. 6 clients in total, implemented using two laptops, each with 3 Wi-Fi USB interfaces. Each AP is configured in 802.11a mode using 20 MHz orthogonal channels (i.e 36, 40 and 44). The traffic model consists of a constant bit rate downlink UDP stream of 30 Mbps using iperf, which is enough to saturate the wireless channel and allows us to understand the dynamics of G-ADRR. Using this setup, we evaluate the dynamic airtime allocation (DAA) of G-ADRR and SGS in two different

scenarios. In Scenario 1 the two slices are configured with $\rho^{SLA}_0 = 0.7$ and $\rho^{SLA}_1 = 0.3$, while in Scenario 2 the two slices are configured with $\rho^{SLA}_0 = 0.2$ and $\rho^{SLA}_1 = 0.8$, respectively. For both scenarios all Wi-Fi clients download the constant UDP stream, except for the slice 1 client in AP0, which stops receiving traffic during ≈ 150 s.

#### 4.2.3.2.1 **Scenario 1:** Dynamic airtime allocation ($\rho^{SLA}_0 = 0.7$, $\rho^{SLA}_1 = 0.3$)

We can see in the left column of Figure 4-10 how for AP0, after 150 s, the client of slice 1 stops receiving traffic for a period of ≈ 150 s. Given that the local scheduler in AP0 is work-conserving, both G-ADRR (upper row) and SGS (lower row) deliver all the bandwidth of AP0 to the client of slice 0.

However, G-ADRR and SGS differ on how they treat the clients of slice 0 and slice 1 in the other APs. Starting with SGS, we see how in AP1 and AP2 the default airtime allocation remains unaltered, and so the clients of slice 0 and slice 1 obtain 70% and 30% of the resources, respectively. However, if we look at the aggregate of the three APs, during the time where the slice 1 client in AP0 stops receiving traffic, SGS does not fulfil the aggregate SLA, since slice 0 receives 1+0.7+0.7 3 = 0.8 and slice 1 0.3+0.3 3 = 0.2. Unlike SGS, G-ADRR reacts to the traffic asymmetry in AP0 by reducing the allocation of slice 0 to 0.55 in AP1 and AP2 and increasing the one of slice 1 to 0.45 in AP1 and AP2, Thus, G-ADRR maintains the aggregate SLA for both slices, i.e., slice 0 receives 1+0.55+0.55 3 = 0.7 and slice 1 0.45+0.45 3 = 0.3.

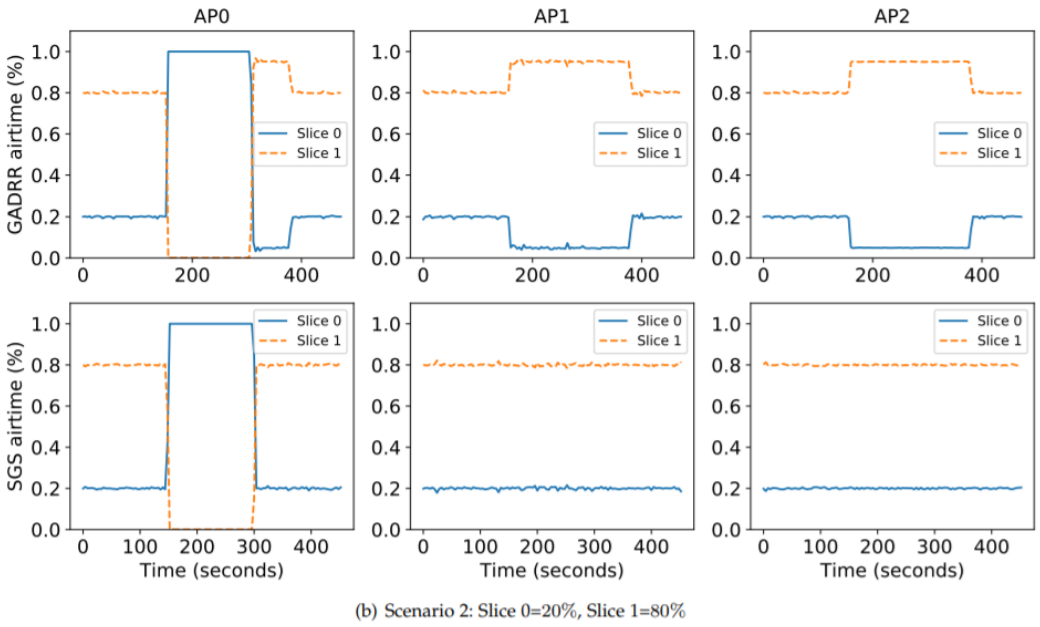



(a) Scenario 1: Slice 0=70%, Slice 1=30%

**Figure 4-10 Scenario 1: Network slices DAA ($\rho^{SLA}_0 = 0.7$ and $\rho^{SLA}_1 = 0.3$)**

#### 4.2.3.2.2 Scenario 2: Dynamic airtime allocation ($\rho^{SLA}_0 = 0.2$, $\rho^{SLA}_1 = 0.8$)

Note, however, that it may not always be possible for GADRR to maintain the overall SLA. One example is Scenario 2, where the G-ADRR (upper row) and SGS (lower row) dynamics are depicted in Figure 4-11. As in Scenario 1, when the slice 1 client stops receiving traffic in AP0, after 150 seconds, the local scheduler in AP0 delivers all the airtime resources to the slice 0 client to avoid wasting resources. In the case of SGS, the airtime allocation in the other APs remains unaltered, and so SGS delivers during this period an average resource share of 1+0.2+0.2 3 = 0.46 to slice 0 and of 0.8+0.8 3 = 0.53 to slice 1, which violates the SLA. Like in Scenario 1, G-ADRR tries to avoid the SLA violation by decreasing the airtime allocations for slice 0 in AP1 and AP2 and increasing the one for slice 1. This time though, G-ADRR can only decrease the slice 0 allocation in AP1 and AP2 to $\rho_{min} = 0.05$, which due to work conserving scheduling results in an allocation for slice 1 in these Aps of 0.95.




(b) Scenario 2: Slice 0=20%, Slice 1=80%

**Figure 4-11 Scenario 2: Network slices DAA ($\rho^{SLA}_0$ = 0.2 and $\rho^{SLA}_1$ = 0.8)**

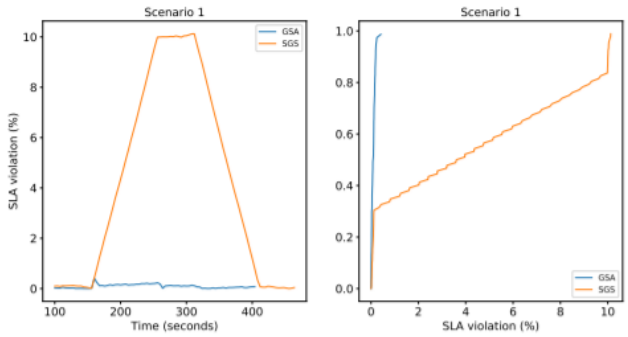
Therefore, G-ADRR results in an average airtime allocation for slice 0 during the off period of slice 1 in AP0 of 1+0.05+0.05 3 = 0.36 and for slice 1 of 0.95+0.95 3 = 0.63, which is a better allocation than SGS, but it is still an SLA violation. Recall though that G-ADRR always attempts to maintain an allocation that fulfils ρ SLA s during a window of duration T SLA, which is linked to the ϵ parameter in GADRR. Hence, when the slice 1 client becomes active again in AP0 at around 300 seconds we see how SGS (lower row) returns immediately to the default allocation, but G-ADRR does not. Instead, G-ADRR increases during ≈ 60 seconds the airtime allocated to slice 1 in AP0, and correspondingly reduces the one for slice 0, to compensate within the $T^{SLA}$ window the previous period where slice 1 received an allocation below its SLA.

#### 4.2.3.2.3 SLA analysis

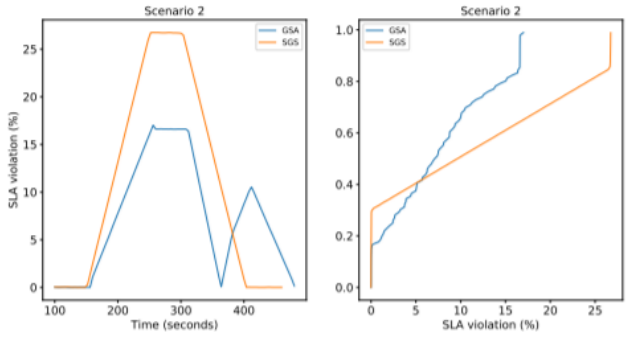
Let us now analyse how G-ADRR and SGS compare with respect to the SLA definition introduced in Section 4.2.2.1. Recall that the SLA contract for slice s consists of a guaranteed average resource allocation $\rho^{SLA}_s$ across all APs where this slice is active, which is evaluated over a sliding window of duration $T^{SLA}$. Therefore, given the observed average airtime allocation for slice s across all APs. i.e. $\hat{\rho}_s^{OBS}(t) = 1/|As| \sum_{a \in As} \rho^{OBS}_{a,s}(t)$, we can quantify the degree of SLA violation $\Delta^{SLA}(t)$ in the following way:

- Compute a running average of the airtime allocation for slice s within a window of duration $T^{SLA}$ as $\hat{\rho}_s^{window}(t) = \hat{\rho}_s^{OBS}(t) * 1/T^{SLA}\ w(t)$, where w(t) is a square signal of duration $T^{SLA}$.
- Compute the deviation with respect to the ideal allocation for that slice as, $\Delta^{SLA}_s(t) = |\hat{\rho}_s^{window}(t) - \rho^{SLA}_s|$
- The overall SLA violation can be computed as the average deviation across all slices as $\Delta^{SLA}(t) = 1/|S| \sum_{s \in S} \Delta^{SLA}_s(t)$

Figure 4-12 depicts $\Delta^{SLA}(t)$ for G-ADRR and SGS in Scenario 1 and Scenario 2, both as an instantaneous signal (left) and as a cumulative distribution function (CDF) (right). We can see in Figure 4-12(a) how G-ADRR avoids any SLA violation in Scenario 1, whereas SGS may grow up to an SLA violation of ΔSLA ≈ 10% during the period when slice 1 traffic is disabled in AP0 (c.f. Figure 4-10).




(a) Scenario 1: Slice 0=70%, Slice 1=30%




(b) Scenario 2: Slice 0=20%, Slice 1=80%

**Figure 4-12 Guaranteed network slices quota-based SLA violation**

Figure 4-12(b) depicts the same results for Scenario 2, where in this case G-ADRR reduces the worst-case SLA violation from $\Delta^{SLA} \approx 25\%$ down to $\Delta^{SLA} \approx 15\%$. It is also worth highlighting what happens in this case at around 360 seconds when $\Delta^{SLA}(t)$ starts to grow for G-ADRR but stays low for SGS. The reason is the overallocation that G-ADRR delivers to slice 1 in AP0 after it becomes active again in this AP (c.f. Figure 4-11 around 300 seconds). While this overallocation enhances the SLA when the off period of slice 1 falls within T SLA window, it penalizes the SLA once the slice 1 off period falls outside the $T^{SLA}$ window. This is inherent to the design of G-ADRR which always attempts to optimize the SLA while looking at the past window.

In [50], the interested reader can find an extensive simulation-based evaluation of G-ADRR, where the advantages of G-ADRR in terms of SLA compliance are demonstrated in scenarios with up to 5 APs and more than 100 users, while comparing G-ADRR with SGS and with two additional dynamic global scheduling policies.

## 4.3 Section Summary

Advanced intelligent LiFi and WiFi networks slicing, and resource management schemes are developed in this section. LiFi and WiFi networks slicing developments mainly include (i) deep multiagent reinforcement learning-based autonomic LiFi attocellular network slicing (ii) a utility-based scheduler at LiFi APs and a Global-Airtime Deficit Round Robin (G-ADRR) at WiFi APs to enforce the quota guarantees of network slices, and (iii) a centralized scheduling policy which tunes the airtime weights of WiFi APs distributed over a geographical area according to their traffic load and network slices requirements. The proposed network sharing schemes for the interface airtime of Wi-Fi APs and the downlink channel capacity of network LiFi APs are designed as self-organizing network (SON) functions. These use the information collected from the LiFi and Wi-Fi network APs to identify the patterns of varying service and traffic demands in temporal and spatial space across the network. This allowed to dynamically, in an autonomic manner, allocate the resource quota guarantees for each network slice across the network

# 5 Multi-WAT Based Localization Optimization and Implementations

5G-CLARITY proposes the use of four different WATs for localization purposes, as introduced in 5G-CLARITY D3.2 [2]. These WATs are a sub-6 GHz based indoor localization system, 60 GHz mmWave, LiFi and OCC/VLC. In this section we provide additional details about the main developments and updated, and improved results compared to those presented in 5G-CLARITY D3.2 [2].

## 5.1 Sub-6 GHz based indoor localization

The sub-6 GHz localization system was initially introduced in 5G-CLARITY D3.2, Section 4.2, [2]. It uses downlink time difference of arrival (DL-TDoA) for position estimation of the UEs. The anchor nodes (AN) or positioning APs, are synchronized and they transmit timestamped reference frames. These frames are received by the UE devices and the time of arrival (ToA) is estimated. Having the precisely estimated ToA and being these frames transmitted in an exact time instances, the position can be easily estimated. The developed system uses a channel bandwidth of 160 MHz in the 5 GHz ISM band. It transmits frames every 10 milliseconds. These frames are quite short and contain a pseudo noise (PN) Gold sequence [59] The system is developed and implemented using software defined radios (SDR). The SDRs used are Ettus N321 radios. The main advantage of using these radios is that they support large instantaneous bandwidth and have the possibility for synchronization using 10 MHz/1 PPS signals as well as synchronization using WR.

The architecture of the positioning system using 10 MHz/1 PPS is shown in Figure 5-1. A total of 3 cables are connected to each of the SDRs. One cable is used for 1 Gbps Ethernet data connection, needed to transmit the reference frames and the other 2 cables are coaxial cables needed for the 10 MHz/1 PPS signals. This approach was tested in 5G-CLARITY D3.2 [2] and the initial results were presented therein. The main disadvantage is the use of multiple cables and their limited length. The initial tests were performed in small areas of approx. 200-300 $m^2$. This requires coaxial cables of approx. 10-12 meters. For 5G-CLARITY UC2.2, larger areas need to be covered, requiring long coaxial cables for synchronization. The first problem derived from this kind of deployment would be the cable attenuation of the signal. If the synchronization signal is strongly attenuated, additional jitter would be present and, in the worst-case scenario, the overall system might not work. Additionally, long coaxial cables are susceptible to interference due to their imperfections, which might also affect the quality of synchronization. Finally, installing three copper cables for a demonstration or deployment purposes can be a tedious task.

To avoid the above-mentioned issues, we decided to use another type of synchronization. The one that suits us best is based on White Rabbit (WR), as it offers sub-nanosecond precision synchronization, which is particularly suitable if high precision positioning is required. The architecture of the positioning system with WR synchronization is given in Figure 5-3. For the WR synchronization, a single monomode fiber is required, where two wavelengths are used for transmission and reception. For the data communication with the SDRs, the same fiber can be used but other two wavelengths must be chosen. For this approach, a wavelength division multiplexer should be used. Another approach is to use duplex monomode fibers, i.e., 2 fibers attached in a single cable. With this approach, a single cable can be used for synchronization and data communication with the SDR. In this case, the data communication with the SDR can be performed at a data rate of 10 Gbps (not possible with copper cables over longer distances), providing greater flexibility.

The initial testing was conducted to estimate the timing synchronization precision. The ANs were set up in a single room and their 1 PPS outputs were connected to an oscilloscope. The synchronization was performed using a single WR switch and monomode fibers with different lengths. The PPS waveforms are given in Figure 5-2. The synchronization of the ANs is better than +/-100 ps. Additionally, the timing offset between the SDRs was constant during the tests, meaning that it can be compensated, achieving much higher synchronization precision.

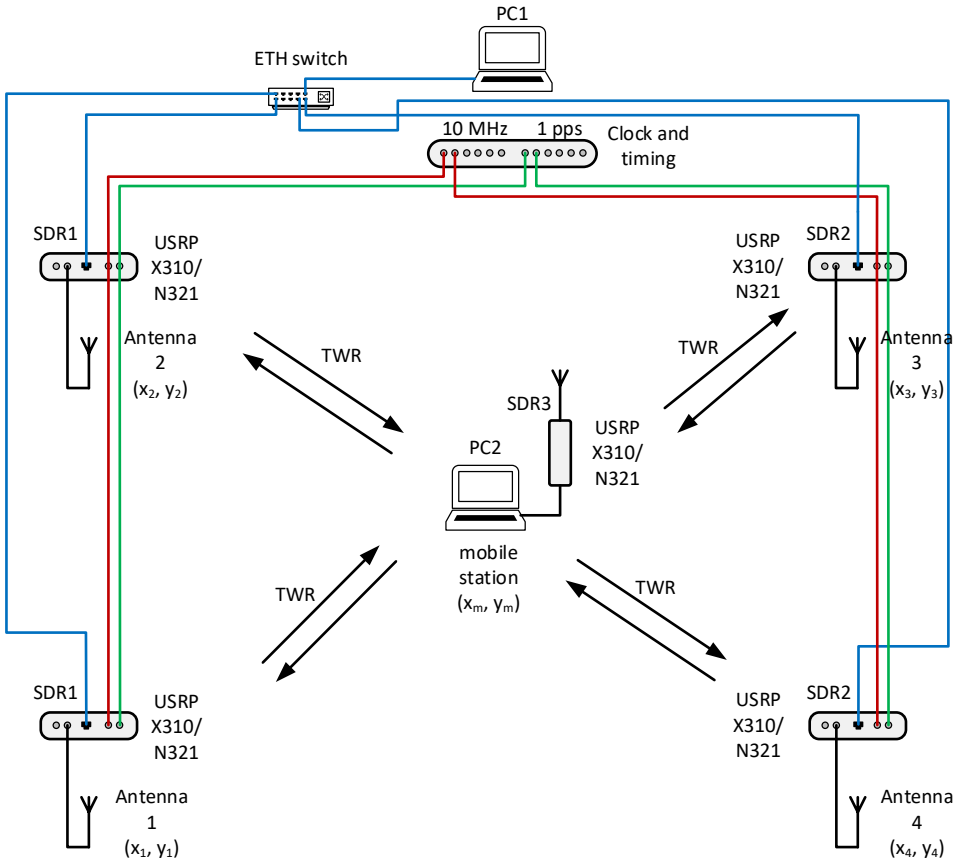
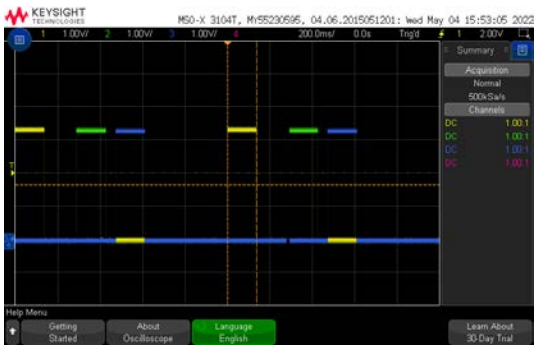
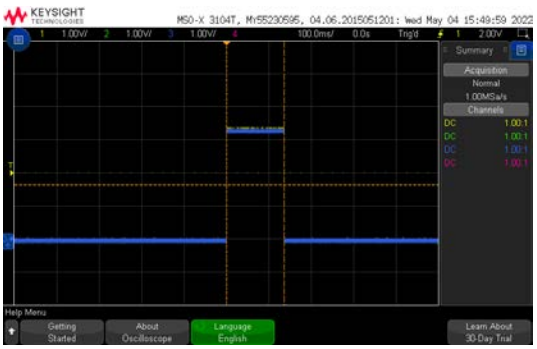


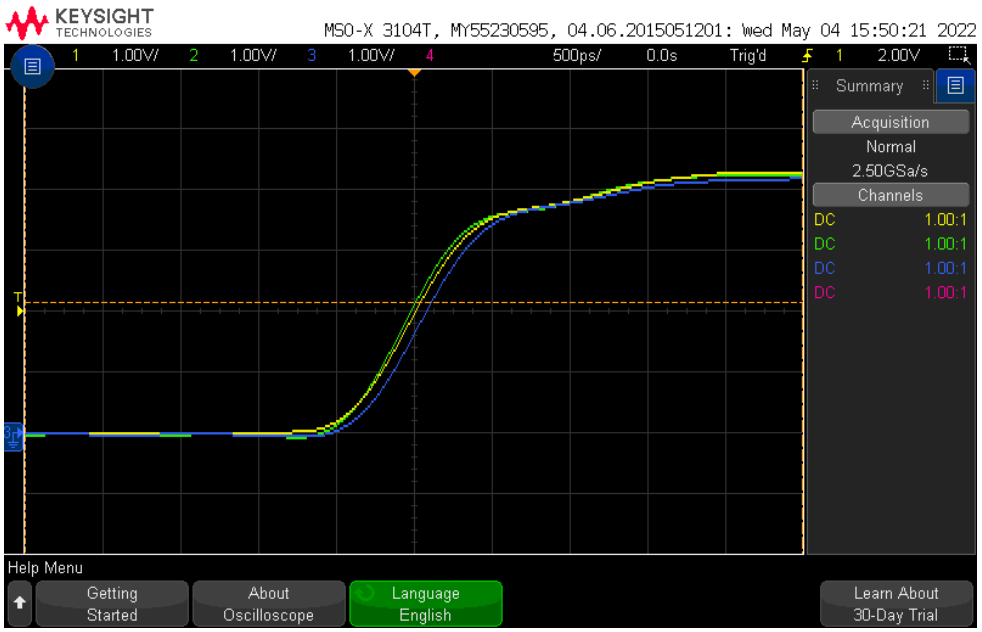

**Figure 5-1 Sub-6 GHz positioning system using 10 MHz/1 PPS synchronization**



**a)**



**b)**




c)

**Figure 5-2 PPS output of the ANs; a) unsynchronized; b) synchronized; c) leading edge of PPS pulse - synchronized ANs**

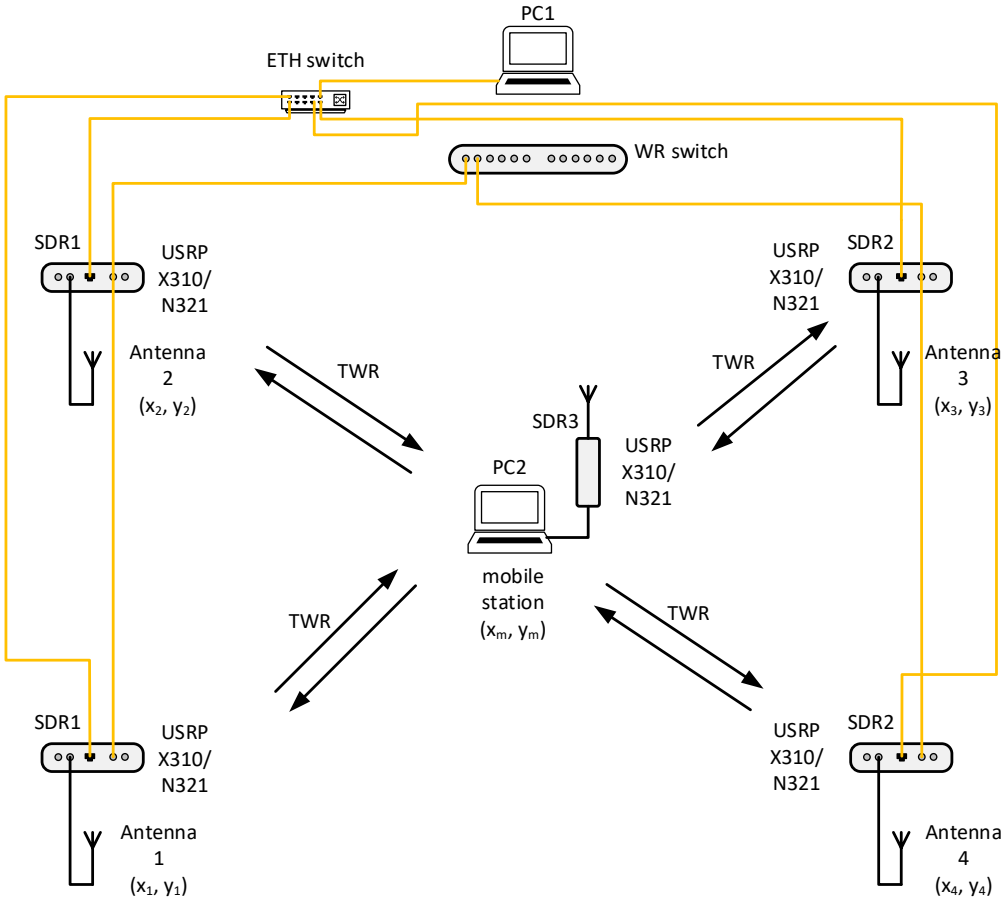



**Figure 5-3 Sub-6 GHz positioning system using WR synchronization**

Additionally, we deployed in a lab environment a positioning system like the one shown in Figure 5-3. Using this setup, a TDOA of the frames transmitted from each of the ANs was estimated at the UE.

Then the ToA at the UE was estimated, whose distribution is depicted in Figure 5-4. In this figure, the distribution of the TDOA between the ToA of the $n^{th}$ AN, and the $0^{th}$ AN is given. The TDOA is given in samples, where the sampling rate is 245.76 MHz, which corresponds to a period of 4.069 ns, or a distance of 1.2199 meters. As can be seen in Figure 5-4, the samples are distributed around a few points. This can be noticed in Figure 5-4 a), where the TDOAs are distributed around -1, 0 and 1. This is due to a discovered bug, i.e., due to transferring the timestamps to the SDRs in time units instead of samples. The conversion of the time units will be fractional and, therefore, the SDR driver rounds them. This causes the samples to be grouped around different samples as can be noticed. By correcting this bug, the TDOAs would be grouped around a single sample, leading to a positioning precision better than ±0.25 samples, i.e., better than ±0.3 meters.

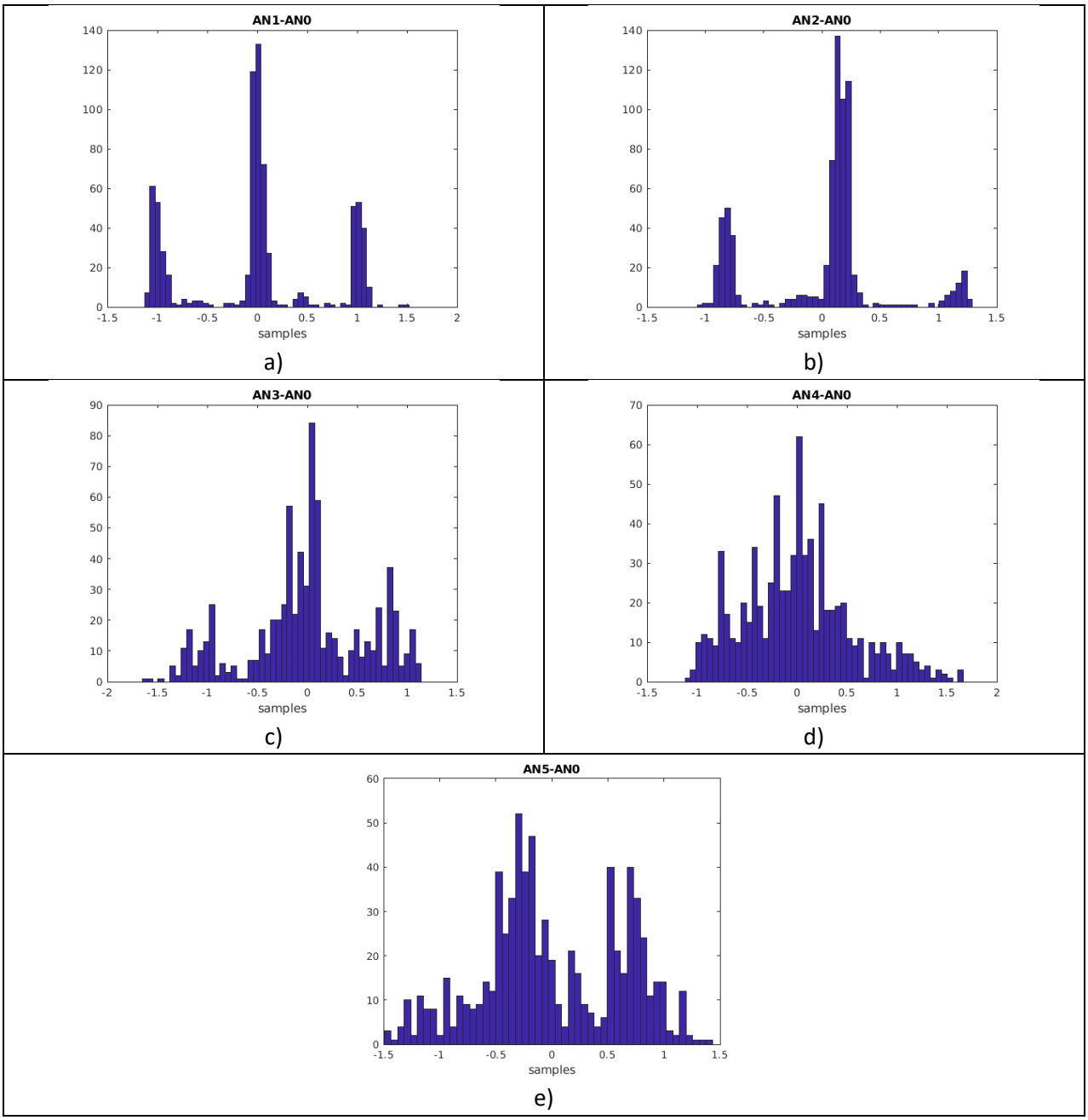



**Figure 5-4 Distribution of TDOA from different APs**

## 5.2 mmWave based localization

In this section we present the mmWave positioning solution developed within this project. The initial ranging results using two radios were shown in D3.2 [2]. In this deliverable we present the deployment and results of a mmWave system used to perform positioning. Two mmWave nodes are deployed as ANs and one mmWave node as a UE. Positioning is performed at the UE mmWave node by estimating the distance to the ANs using two-way ranging (TWR). The main issue with this approach is the synchronization of the nodes, i.e., detection of the frames arriving at the UE and the ANs. For the lab tests, a cabled synchronization between the ANs and the UE was used.

To test the positioning approach, three mmWave nodes were used. The system architecture is given in Figure 5-5. The system consists of two ANs and a single UE. The UE can move around. The distances between each of the ANs (AN1 and AN2) to the UE, i.e., $r_1$ and $r_2$ respectively, are measured using TWR. These two distances describe 2 circles that would intersect in two points. Introducing a constraint that the UE can be only on one side, the position of the UE can be unambiguously found.

To test the positioning precision, the setup from Figure 5-5 was deployed. The UE was moved on a few different positions and its position was estimated multiple times. In Figure 5-6 the positions of the ANs are marked with black dots and the estimated positions of the UE are marked in red, green, and blue. These represent three different positions of the UE. It can be noted that these estimates are tightly grouped since the estimation precision is quite high. The CDFs of these position estimates are given in Figure 5-7. It can be noticed that the positioning precision is better than 3 centimeters in all three cases. Of course, it must be stated that this is the case for a lab experiment and distances of up to a few meters. Further deterioration of the positioning precision is expected in real scenarios with longer distances between UE and ANs.

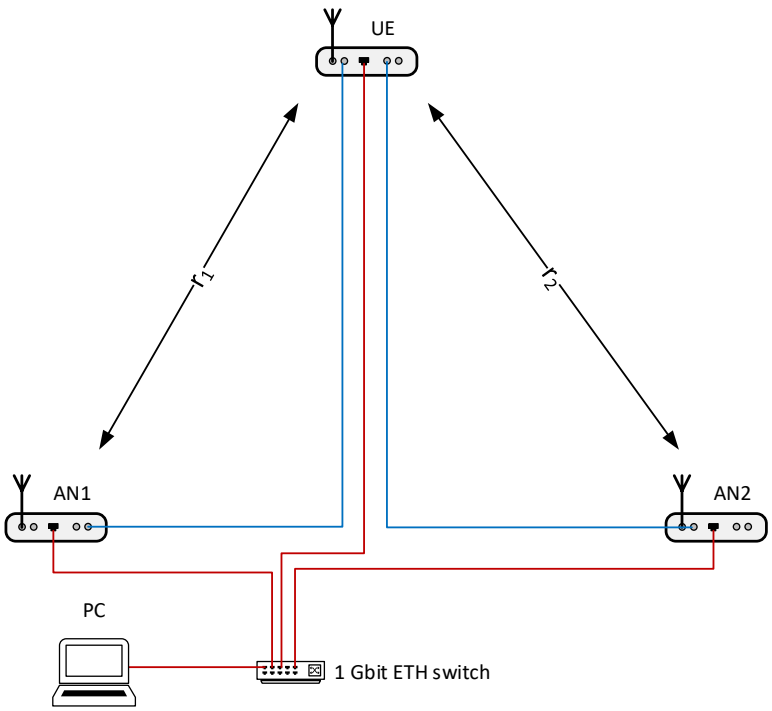



**Figure 5-5 mmWave positioning system architecture**

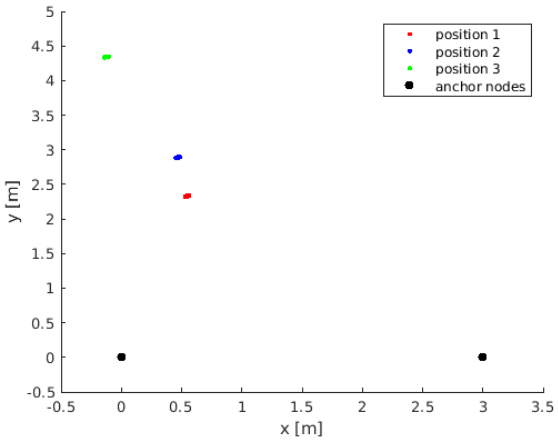



**Figure 5-6 mmWave localization showing the anchor nodes and the estimated position of the UE**

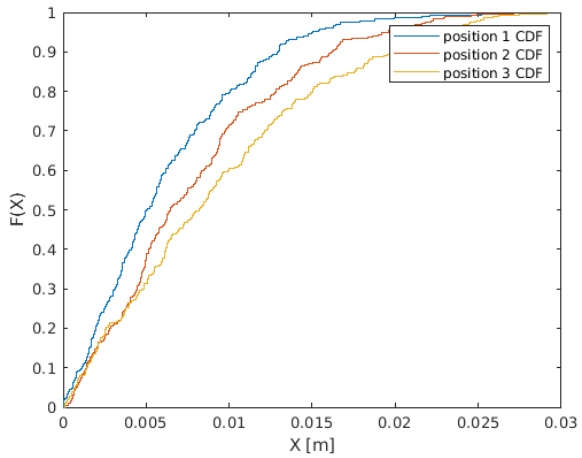



**Figure 5-7 CFDs of the estimated position of the UE**

## 5.3 LiFi based localization

Within 5G-CLARITY, the LiFi technology is part of the overall multi-WAT localisation system. LiFi based localization is implemented by using the received signal strength (RSS)-based positioning approach. The work presented in 5G-CLARITY D3.2 [2] provided a preliminary study to develop a platform for LiFi-based localization system. In this D3.3, the LiFi localization system has been improved and an attempt has been made to integrate the LiFi with the overall localization based 5G-CLARITY multi-WAT system.

The LiFi-based localisation is performed while maintaining illumination and wireless data communications to user devices (UDs). A typical use case scenario is illustrated in Figure 5-8, which shows the communication with the localisation server. Two LiFi APs are installed in the ceiling. Each of them is connected to a lamp with known location, $(x_1, y_1)$ and $(x_2, y_2)$. A switch connects the two LiFi APs and a PC that acts as a localisation server with known fixed IP address. The user device (mobile station) is LiFi-enabled, as it is connected to a LiFi USB dongle. The RSSI values detected by the mobile station are sent to the localisation server over UDP connections. The server estimates the position of the mobile station by considering inputs from other WATs.

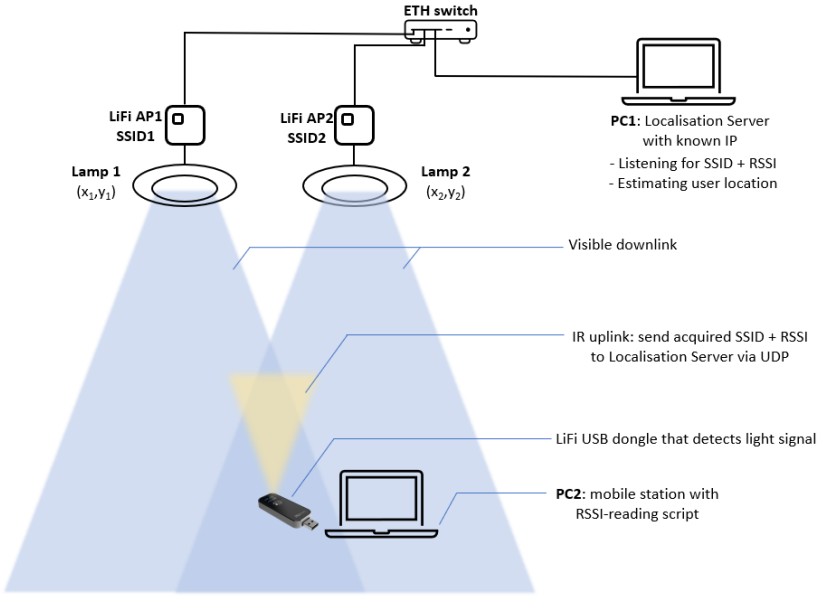



**Figure 5-8 LiFi based localisation system**

The results which were presented in 5G-CLARITY D3.2 [2] show the acquired RSSI values from a single LiFi AP. Also, as discussed in 5G-CLARITY D3.2 [2], there were some limitations, such as

- Low data resolution,
- Only support reading from single AP,
- Communication with the localisation server was not implemented.

In this D3.3, these aspects have been addressed accordingly, with the details are given below.

**Increased data resolution:** In 5G-CLARITY D3.2 [2], the RSS value was only read using four levels by the user device, which adds some inaccuracy in the calculation of UD's position. Now, the signal strength is readable in a range from -110 dBm to -40 dBm. To simplify the values presentation, a quality value is derived by adding 110 to the signal level, resulting in a value from 0 to 70. It is also possible to convert the received signal strength indicator (RSSI) to a percentage as follows:

$$RSSI = [(RSSI + 110) / 70 ] * 100$$

In the new implementation, the RSSI values are updated within each second, which can be accurately mapped to the UD's position along their movement.

**Support for multiple APs:** Increasing the number of APs improves the reading of RSSIs values, which noticeably improves the localisation accuracy. If the LiFi technology is used as a standalone localisation technology, at least 3 APs would be required for accurate estimation. In this study, the LiFi is part of the multi-WAT localisation system, and therefore using two APs would still contribute to the overall system. In current implementation, the different APs have been assigned different SSIDs (Service Set IDentifier), which make it possible to read the RSSIs from them. The RSSI values are recorded for each SSID continually. Compared to the reading from a single connected AP, this approach provides higher accuracy. However, it takes 5-6 seconds for each reading to be completed and saved. Thus, according to the current use case requirements, both approaches could be combined to provide timely and accurate information.

**Integration with the localization server:** The collected RSSI values and corresponding SSIDs are sent to the localisation server with known IP over UDP at a pre-defined rate. Position estimation is done at the server, given the location of the lamps and some inputs from other WATs.

**Implementation:** The proposed system is implemented and tested with a stand-alone LiFi operation first. It is tested with the server.

The script for the operation is implemented in Python and tested in a Linux device.

- *'iwlist.py'*: This reads the RSSI values from the connected AP. Results could be updated within each second.
- '*rssi_scan_all.py'*: This creates '*RSSI-Local.txt*' file and records RSSI values with corresponding SSIDs. It takes several seconds for each reading to be updated.
- '*UDP-Server.py'*: This sets up the UDP server and start listening.
- '*UDP-Client.py'*: This sends the collected RSSI with each SSID to the server. Server will record data in file '*RSSI-Server.txt*'

An example output is as follows:

[{'signal': -54, 'quality': '56/70', 'ssid': 'LIFI2'}, {'signal': -46, 'quality': '64/70', 'ssid': 'LIFI'}]

'signal' gives the received signal strength in dBm; 'quality' gives the numeric data by adding 'signal' with 110 as introduced earlier – maximum is 70; 'ssid' gives the corresponding SSID.

**Test results:** A test has been done to validate the implemented script localization procedure, where the result is shown in Figure 5-9. The results are either sent, printed on a screen, or saved in a file. The implemented script could correctly record the RSS with its corresponding SSID. More tests are planned to be carried, considering an actual link to the localisation server.

As part of the multi-WAT localisation system, the LiFi technology contributes to the main server by providing RSSI values measurement from APs with known locations. Even it is not offering individual location estimation, it still provides additional accuracy to the overall system.

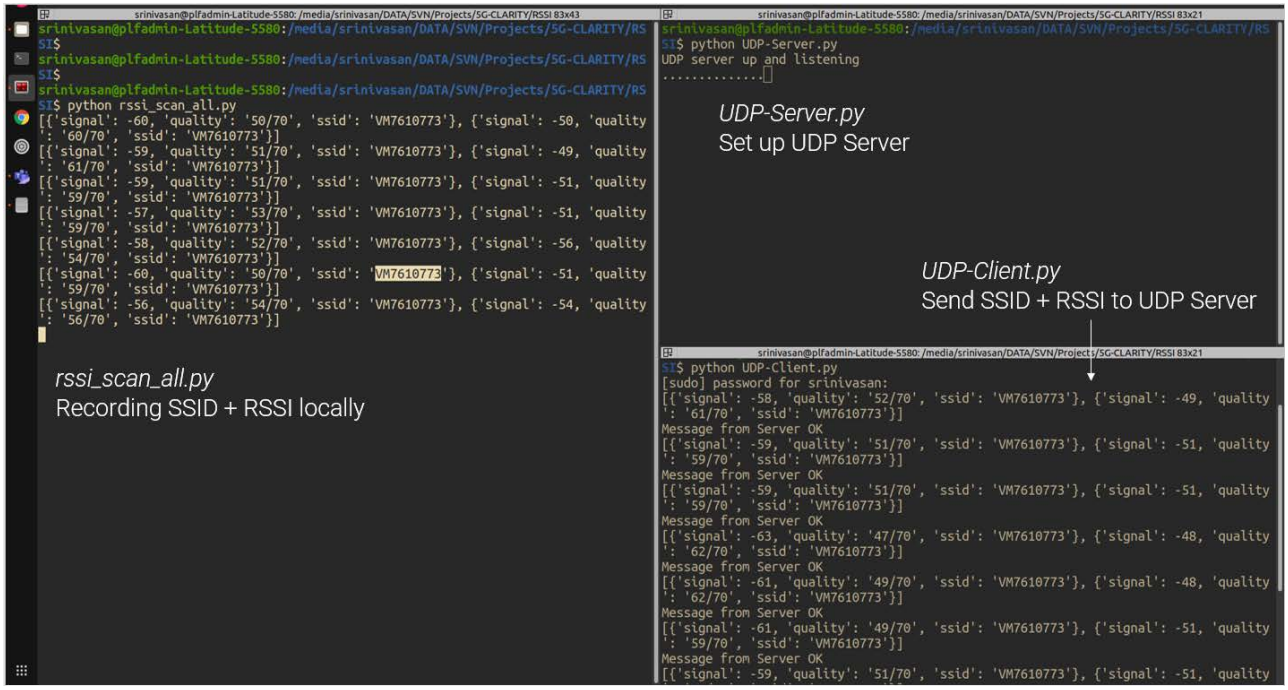



**Figure 5-9 Test result showing RSSI reading and communication with server**

## 5.4 Section summary

The main motivation for having a multi-WAT positioning system in 5G-CLARITY is that the different technologies considered in the project are complementary, and they have different features as well as different signal propagation characteristics, making them applicable in different scenarios.

For example, the sub-6 GHz technology enables signal penetration through thin walls as well as dielectric obstacles, allowing to cover large areas with obstacles, using small number of anchor nodes. This technology, on the other hand, cannot offer high precision, e.g., centimetre range, due to the limited available bandwidth in the sub-6 GHz band. The other technologies developed in this project, i.e., LiFi, OCC/VLP and mmWave, allow for high precision positioning but require a line of sight (LoS) scenarios to work. This means that any obstacle blocking the LoS path between the UE and the anchor node, would affect the localization capabilities of the system and, in the worst case, completely disable them.

Therefore, to achieve high availability of the indoor localization service, 5G-CLARITY proposes a dense deployment of high precision localization technologies (LiFi, OCC/VLP and mmWave) in the areas of interest and, additionally, covering the wider area with the lower precision sub-6 GHz technology. With this approach, in the areas which are mostly used, a high precision localization would be offered. Additionally, these areas and the areas not covered by these high precision technologies, would be covered by sub-6 GHz technology. This would enable a complete coverage (i.e., above 99%) with a precision better than 1 meter.

# 6 Enhanced System Level Simulations and Resource Management

One of the technical objectives, OBJ-TECH-4, of the 5G-CLARITY project is to achieve 500 Mbps/m$^2$ system area capacity via user and control plane functionalities, which has been envisaged in the 5G-CLARITY D3.1 [1]. Accordingly, a real-time traffic simulator, which has been developed to showcase 5G-CLARITY multi-connectivity framework implementation, is presented in section 5, 5G-CLARITY D3.2 [2]. 5GNR, Wi-Fi, LiFi radio and optical access technologies are simultaneously employed to enhance the achievable capacity results. On top of the multi-connectivity structure, both the cell densification and various bandwidth utilization factors are also tested to obtain deep insight into the real-world achievable data transmission limits of the given system structure.

The computer simulation results in 5G-CLARITY D3.2 [2] were provided for two cellular deployment scenarios; “Generic” and “Dense”, where the “Generic” deployment has been obtained by following the rule of thumb deployment rules for 5GNR, Wi-Fi and LiFi technologies. Unlike the “Generic” deployment scenario, the cell densification is employed on the “Generic” deployment to obtain the “Dense” deployment scenario, where the inter-AP distances are reduced significantly. It is important to note that the cell densification yielded a better average area traffic capacity value for LiFi optical access technology; however, this was not the case for the rest of WATs. This is because of the fundamental difference between the optical and RF spectrum, where the higher operation frequencies, in the THz region, are employed in optical wireless communications. Therefore, a higher pathloss (PL) coefficient brings some enhanced flexibility for the frequency reuse and cell densification in the optical spectrum. In addition to the various WAT and deployment scenarios, their conservative and opportunistic bandwidth allocation scenarios are also investigated in the simulation results presented in 5G-CLARITY D3.2 [2]. Accordingly, the conservative bandwidth values are adopted from the actual bandwidth values that are used by the current wireless communication systems. On the contrary, the opportunistic bandwidth values represent the “True Potential” for each technology, where the whole bandwidth for each specific WAT was envisaged to be utilized completely, which is not the case due to the technological immaturities as well as the regulations coming from the governing bodies. This was a crucial aspect of LiFi technology, where the current off-the-shelf light emitting diodes (LEDs) are the main limiting factor of the achievable capacity with around 20 MHz bandwidth for a blue LED with phosphor coating. On the contrary, it has been reported via laboratory measurements that the bandwidth of 6.8 GHz is possible by using a near-ultraviolet laser diode (LD) with red, green, and blue-emitting phosphors [60]. As a result, LiFi was able to deliver 7.42 Gbps/m$^2$ area traffic capacity value by itself in our link level simulations, which are reported in 5G-CLARITY D3.2 [2] . Overall, 8.88 Gbps/m$^2$ aggregate area traffic capacity was reported in 5G-CLARITY D3.2 [2] for a scenario with generic deployment and opportunistic bandwidth region. Although we were able to capture both realistic and futuristic traffic area capacity values for the 5G-CLARITY multi-connectivity architecture, the obtained results represent the absolute maximum that the system could reach. However, in real-world applications there are plenty of imperfections, such as redundancy in signalling, scheduling of resources, lack of full information about the instantaneous radio conditions, which could reduce the area capacity achieved in real environment. In the following subsections, we explain the practical enhancements and the design refinements that have been done on the real-time traffic simulator to investigate the effect of realistic considerations and limitations on the maximum achievable area capacity values.

More realistic user movement profile, resource management structure and modulation and coding schemes are considered. Furthermore, the simulator is enhanced to capture not only link level performance but also the system level key performance indicators (KPIs).

## 6.1 Enhanced simulation platform functionalities

The developed real-time traffic simulator platform in 5G-CLARITY D3.2 [2] could satisfy the real-time data delivery by realizing the L1 and L2 of the protocol stack. The simulator environment could mimic a real-time link level 5G-CLARITY network performance, for multiple UEs served by 5GNR, Wi-Fi and LiFi WATs. However, the developed simulator was lacking the functionalities such as resource and interference management, rate adaptation and realistic transmission frame structures. Moreover, the initial simulator has been developed in MATLAB by following the combination of procedural and functional programming principles. Procedural programming paradigm is known as the routines, subroutines, where data and operations are separated. This technique requires sending data to the procedure/functions. Similarly, functional programming is another programming paradigm that treats the computation as the evaluation of a set of mathematical functions, which avoids the concept of shared state and mutable data. The initial simulator has been developed in MATLAB environment by following both the procedural and functional programming principles, where each step is executed for each element in the simulation environment in a linear top-down fashion. However, this approach is not sustainable for a fully-fledged network simulation with $\mathcal{N}$ APs and $\mathcal{K}$ mobile terminals, specifically when $N$ and $K$ are large. In the initial simulator the total number of APs was taken to be $244$ and $527$ in generic and dense deployment scenarios, respectively. Furthermore, the number of functions, parameters and files become unmanageable both for the developer and the reader since very complex system elements and relationships among entities tried to be captured by the realistic and real-time simulator. In 5G-CLARITY RAN architecture, each of the APs and UEs require three WATs and dedicated PHY layer components and functions as well as MAC/Data link layer scheduling and resource management control functionalities. Mobility and interference at the mobile terminal will also affect the dynamics of all the functions and services. Therefore, a more scalable approach is needed, where the simulator is re-designed to be based on the object-oriented programming (OOP) paradigm. This is based on creating classes and objects. These contain data and procedures. Data in OOP is known as attributes, where the procedures/functions are known as methods.

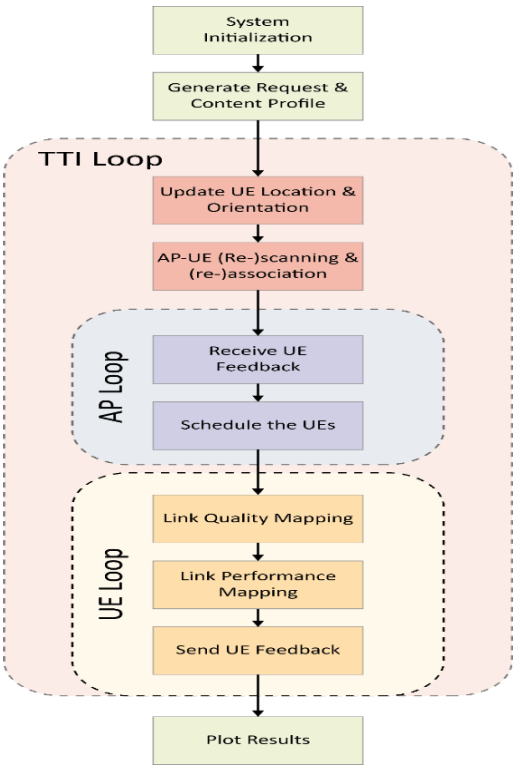



**Figure 6-1 Flow diagram of the OOP based system level simulator**

The main flow diagram of the new OOP based simulator is given in Figure 6-1. Accordingly, the first step is the system initialization, where the three-dimensional (3D) model of the indoor implementation scenario is generated. Then, the LiFi (337 THz), Wi-Fi (5 GHz), 5GNR-I (3.5 GHz) and 5GNR-II (26 GHz) WATs based AP objects are generated and located within the 3D environment. Finally, the UE objects are created with the assumption that all the UEs are equipped with the LiFi/Wi-Fi/5GNR-I/5GNR-II transceiver hardware. It is important to note that the supported WAT list of each UE could be individually defined in our simulation environment. However, they are assumed to support the exact same set of WATs in the simulation for the sake of simplicity. During the system initialization phase, a random location and an orientation angle are also assigned to each UEs. Like the analysis in 5G-CLARITY D3.2 [2], the UE is assumed to be located anywhere within a surface, which is elevated 1.44 m from the ground. The range of the $\boldsymbol{x}$ and $\boldsymbol{y}$ axis values that a UE could take is in the interval of $[5 \quad 135]$ m and $[5 \quad 65]$ m, respectively.

After the system initialization, the content and request profiles are generated. The content request and file sizes follow Zipf and exponential distributions, respectively. It is assumed that there are 10 files in the library that the UE could request. Further details of this process are provided in the following subsections. It is important to note that certain PHY and MAC layer functionalities, such as scheduling, require a transmit time interval (TTI) granularity level. Thus, the time that passes in each loop is adopted to be the basic TTI (temporal resolution of 1 ms) in this case. Within the TTI loop, the first action is to update the UE location and orientation with the predefined periodicity. Then, the UE-AP association will take place. The UE-AP association handshake diagram is provided in Figure 6-2. The association starts with the probe request broadcasted from the UE to the entire network to identify the UE. Then, the APs that receive the probe request return a probe response to identify the compatibility of the WATs in UE with the WATs in the APs. Once the technical foundations are laid, the UE picks $\boldsymbol{N}_{\text{cand}}$ candidate APs based on sorting the path loss (PL) value in ascending order. Once the APs with the best signal conditions are obtained, association request to each of those are sent from the UE. The association becomes complete among the UE and the APs that respond positively to the association request.

After the exchange of signalling between the UE and AP(s), the UE becomes officially attached to the network. The next item within the TTI loop is another loop, namely the AP loop. The UE feedback regarding the service provision is taken in this loop. Then, at each AP, the UEs in the queue are scheduled, depending on the available resources and active sessions within the AP. Once the AP loop is completed, another loop for the UEs is reached as the next thing in the flow. In the UE loop, link quality mapping calculations are conducted to obtain a PHY layer abstraction for the given network.

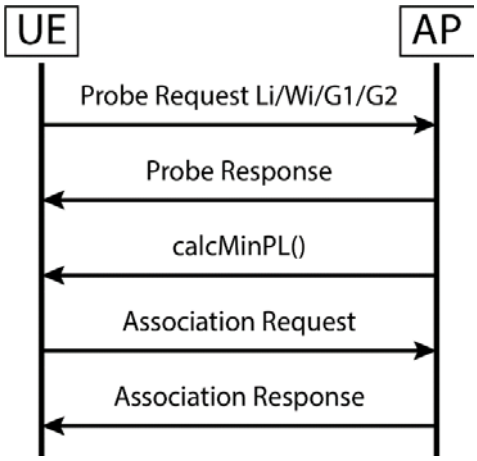



**Figure 6-2 Re-scanning and re-association handshake diagram**

In link performance modelling, the link level telemetry obtained via link quality mapping is mapped on more complex metrics, which can capture the higher layer overheads, operations, and imperfections. Further details on the link quality and link performance mappings are provided in the following subsections.

### 6.1.1 Content and request generation

Internet traffic follows a Zipf-like distribution at multiple aggregation levels, where the amount of traffic per flow is consistent with Zipf's law [61]. A traffic flow could be defined as coarse as all the traffic in a single network or as fine as the traffic between source and destination pair. Moreover, several research items have reported that the relative frequency of web content traffic also follows Zipf's law [62]. Since the goal is to model the potential content request profile, a library is created with 10 files. The relative content request probability vs. the content number is depicted in Figure 6-3.

As can also be seen from Figure 6-3 that the $n^{\text{th}}$ most frequent file has a normalized frequency, which follows Zipf's law, scales as follows:

$$\boldsymbol{p(n)} = \frac{\boldsymbol{n^{-\alpha}}}{\sum_{\forall n} \boldsymbol{n^{-\alpha}}} \qquad \textbf{(6-1)}$$

where the popularity/frequency rank of the content $n \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. The parameter $\alpha$ is the exponent/skewness parameter, which varies from trace to trace. It is reported in [63] that $\alpha$ takes values between $0.64$ and $0.83$ for modelling the web traffic, where it is becomes $0.56$ for YouTube traffic [62]. It is also reported in [64] that the popularity pattern of cloud mobile video streaming service follows Zipf-like distribution with the exponent $\alpha = 0.955$, where the data is collected for 4 months of period. Thus, in the simulations, the adopted values $\alpha = 0.955$ , since the focus is on modelling the internet request and streaming profile of mobile devices. Note that the content sizes are randomly chosen in the simulator, where the file/content sizes follow exponential distribution [65]. As it is also reported in [65], the mean value of the desktop and mobile web pages is 2.23 and 1.25 MB, respectively. Therefore, the content size is chosen to follow exponential distribution with the mean value of $\lambda_{\text{file}} = 3.48$ MB in our simulations. The probability density function (pdf) of the file/content size is expressed as follows:

$$\boldsymbol{f(x, \lambda_{\text{file}})} = \begin{cases} \boldsymbol{\lambda_{\text{file}} e^{-\lambda_{\text{file}} x}}, \text{ if } \boldsymbol{x \geq 0} \\ \boldsymbol{0}, \text{ if } \boldsymbol{x < 0} \end{cases} \qquad \textbf{(6-2)}$$

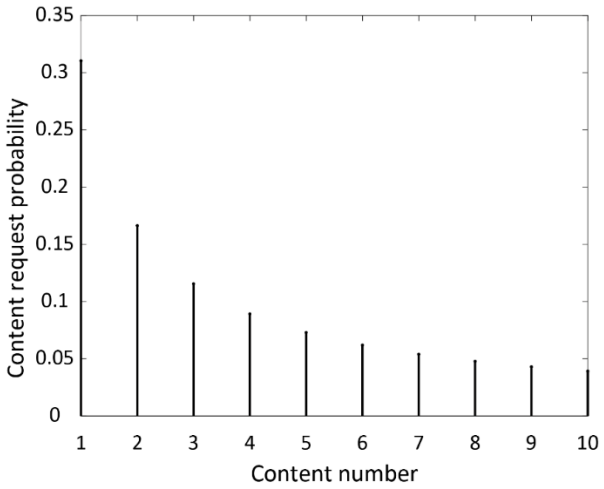



**Figure 6-3 Zipf content popularity distribution**

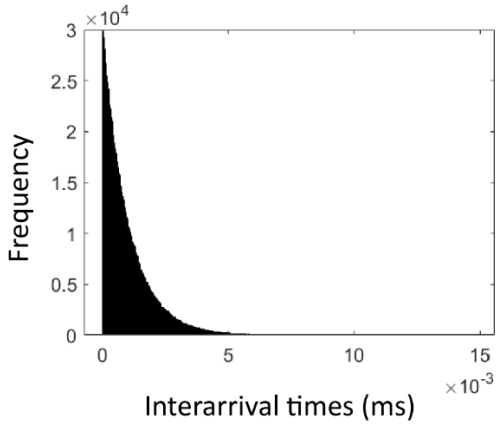
After the content type, size and request modelling, the content request distribution in the temporal domain is also needed. Accordingly, the time between successive request arrivals in the system becomes exponentially distributed, since the number of requests in a fixed interval follows a Poisson distribution, where the pdf is given as follows:

$$f(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, \text{ if } x \geq 0 \\ 0, \text{ otherwise} \end{cases} \quad \textbf{(6-3)}$$

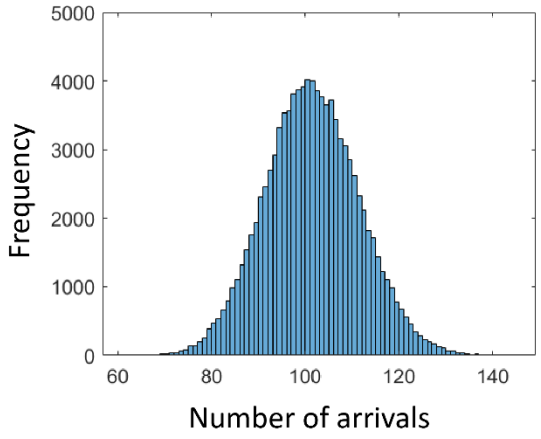
where the parameter $\lambda > 0$ also denotes the rate parameter. The empirical distribution of the interarrival times between the UE requests, which is obtained in our simulator, is depicted in Figure 6-4(a). As depicted in the figure, the interarrival requests follow an exponential distribution with mean and variance values of $10^{-3}$ and $1.003 \times 10^{-6}$. As mentioned previously, like the successive content request arrivals, the number of arrived requests at the network in a fixed interval of time follows a Poisson distribution, which is depicted in Figure 6-4(b). The pmf of the number of request arrivals is expressed as follows:

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \textbf{(6-4)}$$

where $k$ is the number of occurrences of a request at the APs. Note that the average number of events is set to be $\lambda = 100$ request arrivals/sec in the simulations. It can also be seen from the from Figure 6-4(b), the mean and variance values of the distribution for number of arrivals of UE requests become $100.94$ and $100.67$, respectively.




**(a) Distribution of the time between successive content requests, $Exponential(0.001, 1.003 \times 10^{-6})$**




**(b) Distribution of the number of content request arrivals, $Poisson(100.94, 100.67)$**

**Figure 6-4 Probability density functions of the content requests for $\lambda = 10^2$ events/sec**

### 6.1.2 Random mobility and orientation models for the UEs

The mobility model of the UEs in the simulation environment is chosen to be trapped random walk. The random walk models have been investigated and applied in many fields such as solid-state physics, polymer chemistry, photosynthesis, economics, and social sciences. The mathematical basis of the random walk has been established to model the basic principles of statistical mechanics e.g., Brownian motion, diffusion etc. It is considered all the UEs with a random walk in a confined geometry, where they start from a random location within the simulation environment. Starting from the randomly assigned origin point, each UE takes random steps in four possible directions (forward, back, left, and right) at each TTI, as shown in Figure 6-5.

It is important to note from Figure 6-5 that the UE can also choose not to take any action by staying at the same location, which is depicted by half red half orange coloured centre location point. Therefore, the probability values of the symmetric random walk can be expressed as follows:

$$\boldsymbol{p}_{\text{forward}} = \boldsymbol{p}_{\text{back}} = \boldsymbol{p}_{\text{left}} = \boldsymbol{p}_{\text{right}} = \boldsymbol{p}_{\text{no-action}} = \frac{\mathbf{1}}{\mathbf{5}} = \mathbf{0.2} \qquad \textbf{(6-5)}$$

where the probability of UE taking forward, back, left, and right step are denoted by $p_{\text{forward}}$, $p_{\text{back}}$, $p_{\text{left}}$ and $p_{\text{right}}$, respectively. Similarly, the probability of UE taking no action and staying at the same location is also given by $p_{\text{no action}}$. Unlike unrestricted random walks, the UE has restrictions in terms of the locations where it can travel. Since in 5G-CLARITY project, an indoor application of multi-connectivity framework is investigated, the UEs are limited in terms of the locations where they can be mobile. In the simulations, a random walk with a reflecting barrier is considered. In other words, when a UE hits any one of the reflecting barriers e.g., forward, back, left, and right, it would travel in the opposite direction. If the forward, back, left, and right reflecting barriers are denoted as $b_{\text{forward}}$, $b_{\text{back}}$, $b_{\text{left}}$ and $b_{\text{right}}$, respectively, and the element of its position vector for a given time $t$ is $p(t) = b_i$ where $i \in \{\text{forward}, \text{back}, \text{left}, \text{right}\}$, the same element in its position vector in time $t + 1$ becomes,

$$\boldsymbol{p}(\boldsymbol{t}+\mathbf{1}) = \begin{cases} \boldsymbol{b}_{\text{forward}} - \mathbf{1} \\ \boldsymbol{b}_{\text{back}} + \mathbf{1} \\ \boldsymbol{b}_{\text{left}} + \mathbf{1} \\ \boldsymbol{b}_{\text{right}} \end{cases} \qquad \textbf{(6-6)}$$

As can be seen from the above expression that the UE will bounce back from the barrier. Reflecting barrier random walks are not weakly or strongly stationary. However, they are still a Markov process since the most recent value of the process gives us as much information as the entire process even when it is at the reflecting barrier.

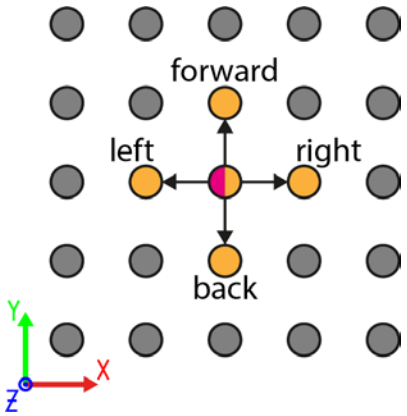



**Figure 6-5 Trapped random walk-based UE mobility diagram**

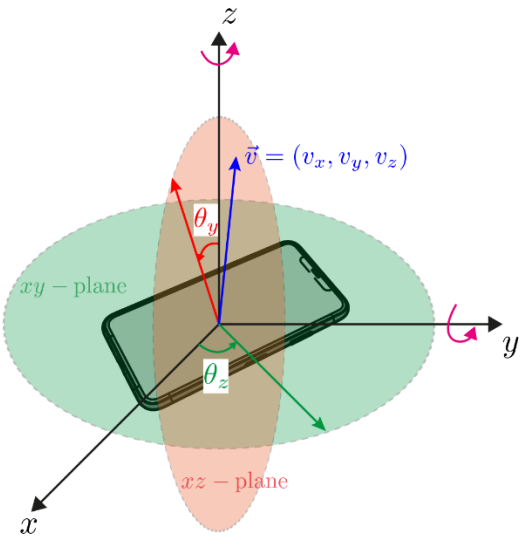



**Figure 6-6 Random orientation model for the mobile terminals**

Note that this type of random walks are not Martingales, as we know the next location value will be different from the current one if the UE is at the reflecting barrier.

The random orientation model of the mobile terminals is also depicted in Figure 6-6. Like D3.2 [2], the empirical data based random orientation model with approximately Gaussian distributed polar angle ($\theta_y$) with mean and variance values of $29.67°$ and $8.91$ dB [66], Moreover, the range of the polar angle in the real-world measurements is determined to be $\theta_y \in [0, \pi/2]$ degrees. The azimuth angle ($\theta_z$), follows a uniform distribution within the interval of $[-\pi, \pi]$ degrees.

### 6.1.3 Handover protocol

To provide seamless data connectivity under the random mobility effects of the UEs in cellular networks, the active session/ongoing connection might be needed to be transferred from one AP to another. Handover is required to ensure the following,

- The UE stays within the coverage area of the APs
- The APs that the UE is associated with are not fully loaded and available to serve if needed
- The received signal quality is kept within the interval to provide sufficient quality of service (QoS), which essentially means satisfying the service level agreement (SLA).

It is also important to note that since there are three WATs utilized in the 5G-CLARITY project, two types of handovers; (i) horizontal and (ii) vertical could be executed. In horizontal handovers, the handover occurs between the APs of a same WAT. However, in vertical handover structure, the connectivity is migrated/provided by a different WAT compared to the initial WAT that the UE was connected. In the current cellular networks, generally the horizontal handovers are executed to keep the received signal strength (RSS) within a certain limit at the UE. However, vertical handovers are used to support larger scale user mobility, such as user being travelling from indoors to outdoors. In the 5G-CLARITY project, the full exploitation of the available resources and WATs are envisaged to enhance the connectivity and capacity as much as possible. Therefore, both the horizontal and vertical handover schemes are utilized in our simulation environment to validate the potential limits of the proposed network structure in 5G-CLARITY project.

From the link establishment point of view, handovers could be divided into two categories: (i) hard handover and (ii) soft handover. While the hard make before break handover appears seamless to the UE, there is always a possibility that a call drop or a service interruption could occur due to potential connection re-establishment process. In soft handover, the UE connects to the new AP while keeping the old connection. Hence, the UE keeps at least one connection link, which is referred to as "make-before-break". Soft handovers propose a potential alleviation of the ping-pong effect as well as the fluctuation in the achievable rate. However, since there is no central node and no need for the power control due to the orthogonal modulation, soft handover scheme is not included in long term evaluation (LTE). In the simulations, the hard handover scheme is also implemented to comply with 3GPP reference architecture. Accordingly, when a new AP offers better RSS to the UE, the UE requests handover from the old AP. The handover becomes successful unless the new AP is not fully loaded and/or the maximum number of active connections are not reached.

### 6.1.4 Single point and coordinated multi-point transmission techniques

It is well known in the wireless network designing process that the simultaneous usage of the same resources between neighbouring cells introduces a significant performance degradation to the overall network performance as well as the individual mobile user quality of service (QoS). This phenomena of interfering signal among cells are referred to as co-channel interference (CCI). In the wireless communications literature, various CCI mitigation techniques are proposed. More specifically, the CCI could be mitigated by special design of the transmit/receive units and the propagation environment. However, a very confined radiation/emission pattern at the APs could introduce a patchy signal reception to the mobile users in the spatial domain. This "coverage islands" could potentially mean that the QoS and throughput degradation at the UE, which could eventually cause a user outage. Furthermore, non-uniformities between the signal reception pattern in certain locations could cause a rapid handover trigger, which is also known as "ping-pong effect". The most straightforward solution for the mentioned problems is the partitioning of the available resources in temporal, frequency, polarization, spatial, angle and power domains [67].

The coverage islands and cell-edge CCI problems are simultaneously solved by also fractional frequency reuse (FFR) schemes [68]. However, the FFR requires a compromise between the available bandwidth and CCI. Therefore, as an alternative transmission method called coordinated multi-point (CoMP) transmission technique is proposed in the literature [69]. Unlike the traditional single point (SP) transmission, the CoMP offers merged cell borders, where a coordinated transmission among the merged cell APs could alleviate the CCI. To alleviate the CCI related area traffic capacity reduction, both SP and CoMP transmission methods are adopted in our simulator. Accordingly, the best $N_{\text{assoc}}$ AP within the candidate APs list has been chosen in CoMP transmission method, where $N_{\text{assoc}} = 1$ for SP. Please note that neither Wi-Fi nor LiFi supports CoMP currently. However, the whole idea behind this investigation is to capture the current and future possibilities while reflecting the WAT specific capacity potential. Thus, the CoMP transmission method will be assumed to applied to all the WATs in our network without loss of generality.

## 6.2 Obtaining the system performance

The system-level simulations are important elements of the actual system design and implementation process due to the obtained insights for the deployment of the cells, frequency planning as well as the overall system performance. To be able to predict the performance of the multi-WAT 5G-CLARITY network, the *link-to-system-mapping is added,* which gives the opportunity to parameterize the input output relationship of the whole network in a simple manner. The physical layer of a 5GNR network on a system level is depicted in Figure 6-7.

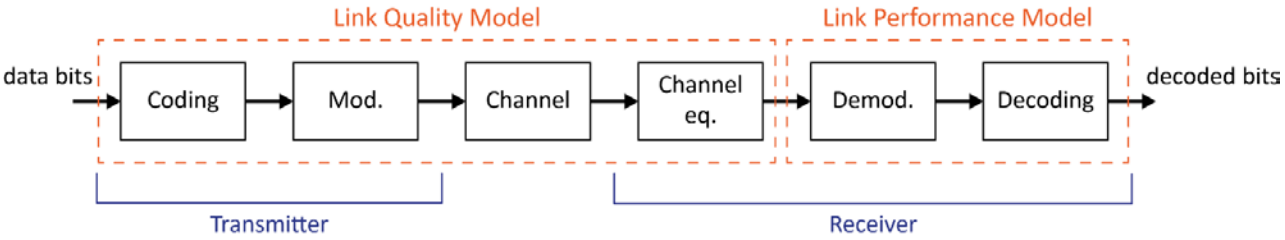



**Figure 6-7 5G-CLARITY PHY layer link abstraction model**

As can be seen from Figure 6-7, the PHY layer system model could be represented by two blocks for the sake of simplicity. The link quality model measures the quality of the received signal after the channel equalization. The link performance model translates the post-equalization SINR into block error ratio (BLER) and further into area spectral efficiency and effective throughput, based on the MCS mapping.

### 6.2.1 Link quality model

The link quality model could be represented in two parts.

1. Large-scale fading, which consists of path loss (PL) and shadowing
2. Small scale fading, which consists of fast, slow, and multi-path fading

On the one hand, the large-scale fading has the ability of modelling both propagation losses caused by the AP-UE distance and antenna gains jointly. Moreover, the fading that emerges due to the obstacles and irregularities between the propagation path between the AP and UE are also captured as log-normal shadowing with the respective parameters, which will be detailed in the following section. On the other hand, small scale fading is a time-dependent process, which captures the rapid changes in the signal amplitude and phase over a short period of time.

For the sake of simplicity in our system level simulations, the link quality model with only large-scale fading effects is considered in 5GNR, Wi-Fi and LiFi WATs without loss of generality. In case more accurate channel models are needed, complex theoretical, ray tracing based, or empirical channel models could easily be implemented within our simulator environment.

### 6.2.2 Link performance model

In cellular wireless communication networks, the quality of the received signal at the mobile terminal depends on the link quality from the associated/serving AP(s) and interference from the other AP(s), and the noise. To optimize the achievable area traffic capacity of the simulated 5G-CLARITY network, the transmit data rate for each user should be adaptively altered at the AP to match the varying channel conditions under a given signal power constraint. This is commonly referred as adaptive modulation and coding (AMC) in 5GNR networks.

In 5GNR, UE can report channel quality index (CQI) to assist the APs in selecting a suitable MCS for downlink (DL) transmission. Note that the reported CQI is not a direct indication of SNIR in 5GNR, where the UE reports the highest MCS that it can decode with transport BLER not exceeding 10% for an out-of-sync condition. It is also important to note that a high CQI feedback reporting frequency could yield an extensive signalling overhead for the network. Thus, a wideband aperiodic CQI reporting (Mode 1-0), which means that a UE report one wideband CQI value for the whole system bandwidth is adopted in our simulations.

The link performance model is used to determine the block error rate (BLER) under a given resource allocation modulation and coding scheme (MCS). In 5GNR, 15 different MCS values are associated with 15 channel quality indicator (CQI) values. The data rate calculation in LTE is based on the number of allocated

RBs, reported CQI, respective MCS and transport block size (TBS). Therefore, the reported CQI is mapped into an MCS index at the AP first. Then, the MCS index is mapped into TBS index for PDSCH based on [70] and [71]. The TBS is also obtained from the TBS index for a given number of resource blocks value, which is determined by the available bandwidth by [71]. Hence, the obtained TBS value will represent the number of bits that can be transmitted per transmission time interval (TTI). It is important to note that the mapping between CQI and MCS indexes is vendor specific. Therefore, like [72] and [73], the achievable rates in our simulation environment are obtained by using the UE reported SINR values. Accordingly, the UE reported SINR value is mapped into an MCS index of the resource block group (RBG) by using a one-to-one AMC mapping function. The mapping flow can be summarized by

$$\mathrm{SINR}_r^u \xrightarrow{F(\cdot)} \overline{\mathrm{SINR}}_{\mathrm{r}}^u \xrightarrow{G(\cdot)} \overline{\mathrm{SINR}}_g^u \xrightarrow{H(\cdot)} (\eta_g^u, i_{\mathrm{MCS},g}^u) \qquad \textbf{(6-7)}$$

where the SINR for the $u^{\mathrm{th}}$ UE and $r^{\mathrm{th}}$ resource block (RB) is denoted by $\mathrm{SINR}_r^u$. Similarly, the effective SNR of the $r^{\mathrm{th}}$ RB and $u^{\mathrm{th}}$ UE is given by $\overline{\mathrm{SINR}}_{\mathrm{r}}^u$. The effective SINR of sub-bands, which is mapped to $g^{\mathrm{th}}$ resource block group (RBG) for $u^{\mathrm{th}}$ UE is given by $\overline{\mathrm{SINR}}_g^u$. Lastly, the achievable rate, $\eta_g^u$, and MCS index, $i_{\mathrm{MCS},g}^u$, for $g^{\mathrm{th}}$ RBG and $u^{\mathrm{th}}$ UE are obtained because of the successive mappings. A straightforward method to determine the relationship between the SINR and appropriate CQI indexes at the UE, which satisfies the BLER $\leq 10^{-1}$, is to construct BLER thresholds look-up table for various MCS. Hence, the minimum SINR value required for that CQI index value with a maximum BLER of 0.1 could be obtained by inversing the BLER vs. SNR plot. The SINR-to-MCS index mapping look-up table could be represented mathematically as a one-to-one effective SINR mappings as follows [72]:

$$(\eta_{a,r}^u, i_{a,r}^u) = h\left(g\left(f(\gamma_{a,r}^u)\right)\right) \qquad \textbf{(6-8)}$$

where the effective SINR mapping function, which converts the SINR into effective SINR of a given RB for the UE, is denoted by $f(\cdot)$. Similarly, the RBG mapping function, which converts the effective SINR for a given RB to the effective SINR of a RBG for a UE is given by $g(\cdot)$. Lastly, the AMC mapping function, which maps the RBG of a given UE into the achievable data rate ($R_{u,r}^a$) and the MCS index ($i_{u,r}^a$) is denoted by $h(\cdot)$. The one-to-one AMC mapping function, $h(\cdot)$, which unifies the CQI index, MCS index, SINR, the modulation type, approximate code rate and spectral efficiency per symbol is given in Table x in a look-up table format. Then, the achievable rate for the system becomes,

$$R_{a,r}^u = \eta_{a,r}^u \frac{N_{\mathrm{RE}}^{\mathrm{sym}}\left(N_{\mathrm{sym}}^{\mathrm{sf}} - N_{\mathrm{sym}}^{\mathrm{PDCCH}}\right)}{2T_{\mathrm{slot}}} \qquad \textbf{(6-9)}$$

where the number of resource elements (REs) per symbol is denoted by $N_{\mathrm{RE}}^{\mathrm{sym}}$. Furthermore, the number of symbols in a subframe is given by $N_{\mathrm{sym}}^{\mathrm{sf}}$. Similarly, the number of symbols that are used for the physical downlink control channel (PDCCH) is also denoted by $N_{\mathrm{sym}}^{\mathrm{PDCCH}}$. Lastly, $T_{\mathrm{slot}}$ represents the slot duration. In our simulation environment, the methodology explained above is simplified further to map the SINR to the achievable throughput values. Accordingly, the physical resource block (PRB) allocation is managed by evolved NodeBs (eNBs) scheduling algorithms. Furthermore, although there is a direct mapping between the available bandwidth and number of PRBs, this choice is dependent on the sub-carrier spacing (SCS) and bandwidth in 5GNR. Based on the 3GPP Release 15 specifications [74], the number of orthogonal frequency division multiplexing (OFDM) symbols per sub frame is 14, with length of 1 ms. Moreover, the SCS of 60 kHz a bandwidth of 100 MHz is also considered. By using the number of PRBs and channel conditions the physical throughput of the system is calculated as follows [73]:

$$\mathbf{R^u} = \frac{\xi \mathbf{N}_{\mathrm{RE}}^{\mathrm{sym}}}{\mathbf{T}_{\mathrm{slot}}^{\boldsymbol{\mu}}} \sum_{\mathbf{i=1}}^{\mathbf{C}} \boldsymbol{\nu}_{\mathrm{l}}^{(\mathbf{i})} \mathbf{M}^{(\mathbf{i})} \boldsymbol{\rho}^{(\mathbf{i})} \mathbf{N}_{\mathrm{PRB}}^{\mathrm{sym},(\mathbf{i}),\boldsymbol{\mu}} (\mathbf{1} - \boldsymbol{\kappa}^{(\mathbf{i})}) \qquad \textbf{6-10}$$

where the maximum number of layers and modulation order are denoted by $\boldsymbol{\nu}_{\mathrm{l}}^{(i)}$ and $\boldsymbol{M}^{(i)}$, respectively. The number of PRBs per OFDM symbol, physical overhead factor that emerges due to the higher layer

processing, and code rate are given by $N_{\mathrm{PRB}}^{\mathrm{sym},(i)}$, $\boldsymbol{\kappa}^{(i)}$, and $\boldsymbol{\xi}$, respectively. The scaling factor, which takes the values between 0.75 and 1 is also given by $\boldsymbol{\rho}^{(i)}$. Note that the parameters $\boldsymbol{\mu}$ and $\boldsymbol{C}$ are the 5GNR numerology that represents SCS and the number of aggregated channels, respectively. Hence, the one-to-one mapping lookup table/function, $\mathcal{M}(\cdot)$, which maps the MCS indexes, SINR values and achievable throughput values is presented in Table 6-1.

**Table 6-1 MCS index, SINR, and Throughput Mapping Function, $\mathcal{M}(\cdot)$ [73]**

| MCS Index | SINR (dB) | Throughput (Mbps) |
|---|---|---|
| 1 | −4.63 | 73.00 |
| 2 | −3.615 | 95.69 |
| 3 | −2.6 | 117.64 |
| 4 | −1.36 | 152.99 |
| 5 | −0.12 | 187.73 |
| 6 | 1.17 | 231.01 |
| 7 | 2.26 | 273.68 |
| 8 | 3.595 | 320.61 |
| 9 | 4.73 | 366.93 |
| 10 | 6.13 | 413.87 |
| 11 | 7.53 | 414.48 |
| 12 | 8.1 | 460.80 |
| 13 | 8.67 | 529.07 |
| 14 | 9.995 | 597.33 |
| 15 | 11.32 | 674.13 |
| 16 | 12.78 | 750.93 |
| 17 | 14.24 | 802.13 |
| 18 | 14.725 | 800.92 |
| 19 | 15.21 | 852.12 |
| 20 | 16.92 | 945.37 |
| 21 | 18.63 | 1036.80 |
| 22 | 19.975 | 1126.40 |
| 23 | 21.32 | 1217.83 |
| 24 | 22.395 | 1314.75 |
| 25 | 23.47 | 1411.66 |
| 26 | 25.98 | 1503.09 |
| 27 | 28.49 | 1596.35 |

| MCS Index | SINR (dB) | Throughput (Mbps) |
|---|---|---|
| 28 | 31.545 | 1664.00 |
| 29 | 34.6 | 1733.49 |

## 6.3 Simulation results

In this section the computer simulation results that are obtained by using enhanced simulator structure is presented. Accordingly, the dimensions of BOSCH factory building, in Aranjuez, Spain, are considered to model an indoor application scenario, which is depicted in Figure 6-8. As can be seen from figure that the dimension of the considered scenario is $140 \times 70 \times 4$ m$^3$. Each UE object is randomly placed in a location within the considered scenario with a random orientation angle while they are generated. Then, depending adopted deployment scenario, (i) Generic deployment with moderate inter-AP distances, or (ii) Dense deployment, where the inter-AP separation distances are significantly smaller, the AP objects are created and located within the 3D environment. Then each UE changes their orientation and location autonomously, where the TTI loop initiates the object interactions in TTI basis. The details of the deployment types and the details of the parameters used in our simulator are given in Table 6-2 and

Table 6-3. The computer simulations will consist of two stages:

1. Obtaining SINR/SIR distribution maps for each WAT independently.
2. Obtaining the achievable throughput values for individual WATs as well as the aggregated values.

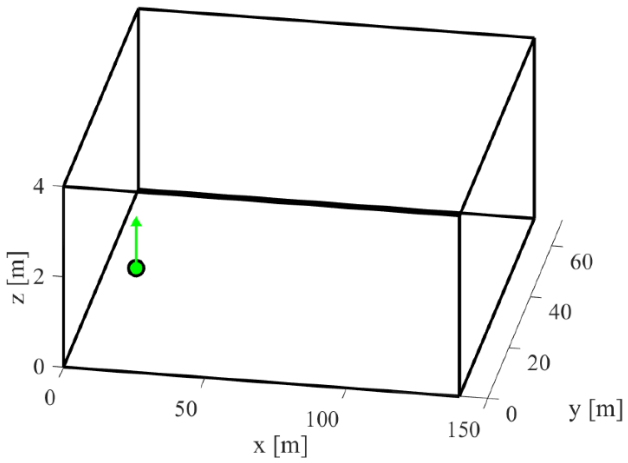



**Figure 6-8 Scenario description and random UE orientation**

**Table 6-2 Simulation Parameters**

| Parameter | Description | Value |
|---|---|---|
| $n_{\text{LAP}}$ | Number of LiFi APs | 136 (generic)<br>351 (dense) |
| $n_{\text{WAP}}$ | Number of Wi-Fi APs | 66 (generic)<br>102 (dense) |
| $n_{\text{GAP1}}$ | Number of 5GNR-I APs | 4 (generic)<br>8 (dense) |
| $n_{\text{GAP2}}$ | Number of 5GNR-II APs | 38 (generic)<br>66 (dense) |
| $x_{\text{LAP}}$ | Locations of LiFi APs in $x$-axis | 6:8:134 m (generic)<br>5:5:135 m (dense) |
| $y_{\text{LAP}}$ | Locations of LiFi APs in $y$-axis | 7:8:63 m (generic)<br>5:5:65 m (dense) |

| Parameter | Description | Value |
|---|---|---|
| $z_{\text{LAP}}$ | Locations of LiFi APs in $z$-axis | 4 m |
| $z_{\text{WAP}}$ | Locations of Wi-Fi APs in $z$-axis | 4 m |
| $z_{\text{GAP1}}$ | Locations of 5GNR-I APs in $z$-axis | 2 m |
| $z_{\text{GAP2}}$ | Locations of 5GNR-II APs in $z$-axis | 2 m |
| $\alpha$ | Zipf-like distribution exponent/skewness parameter | 0.955 |
| $n$ | Number of contents in the library | 10 |
| $N_{\text{assoc}}$ | Number of associated APs in CoMP transmission | 3 |
| $N_{\text{sym}}^{\text{sf}}$ | Number of OFDM symbols per sub frame | 14 |
| $N_{\text{RE}}^{\text{sym}}$ | Number of resource elements per OFDM symbol | 12 |
| $\lambda_c$ | Rate parameter of the content size distribution | 0.287 |
| $\lambda$ | Rate parameter for the number of content requests distribution | 100 events/sec |
| $T$ | Duration of the simulation | 1000 sec |
| $A$ | Area of the PD | 1 cm$^2$ |
| $\Phi_{1/2}$ | Semi-angle of half power of the LED | $60^\circ$ |
| $\Psi_{1/2}$ | FoV of the PD | $85^\circ$ |
| | Height of the LiFi AP | 4 m |
| $H_{\text{UE}}$ | Height of the mobile terminal | 1.44 m |

**Table 6-3 Adopted WAT Operation Frequencies [2]**

| Parameter | Description | Values | Conservative Values | Opportunistic Values |
|---|---|---|---|---|
| $\lambda_{\text{LAP}}$ | Operation wavelength region of LiFi APs | 337 THz (299.79-374.74 THz), where the visible band is optional | | |
| $f_{\text{WAP}}$ | Operation frequency of Wi-Fi APs | 5 GHz (5.15-5.725 GHz) | | |
| $f_{\text{GAP1}}$ | Operation frequency of type-I 5GNR APs | 3.5 GHz (3.4-3.8 GHz) | | |
| $f_{\text{GAP2}}$ | Operation frequency of type-II 5GNR APs | 26 GHz (24.5-26.5 GHz) | | |
| $B_{\text{LAP}}$ | Bandwidth of the LiFi APs | | 20 MHz | 6.8 GHz [75] |
| $B_{\text{WAP}}$ | Bandwidth of the Wi-Fi APs | | 80 MHz | 320 MHz |
| $B_{\text{GAP1}}$ | Bandwidth of the type-I 5GNR APs | | 80 MHz | 400 MHz |
| $B_{\text{GAP2}}$ | Bandwidth of the type-II 5GNR APs | | 800 MHz | 1.792 GHz [76] |

### 6.3.1 SIR distribution results

As the 5G-CLARITY project multi-connectivity framework proposes the utilization of three wireless access technologies (WATs); (i) 5GNR, (i) Wi-Fi and (iii) LiFi, our simulations are conducted to obtain a path gain and a wideband SINR value. The path gain in our simulations is defined as the average signal gain between the UE and the AP that the mobile device is associated with, where the signal gain includes the attenuation, shadowing, and antenna gains. The wideband SINR per UE is calculated as the ratio of the average power received from the associated cells and the average interference power received from the other cells plus noise as follows:

$$\mathrm{SINR}(\boldsymbol{u}) = \frac{\boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{sig}}}{\boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{int}}+\boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{noise}}} = \frac{\sum_{\forall i\in\Im} \boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{R},i}}{\sum_{(\forall j, i\neq j)} \boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{R},j}+\boldsymbol{N}^{\boldsymbol{u}}_{\boldsymbol{0}}\boldsymbol{B}^{\boldsymbol{u}}} \qquad \textbf{(6-11)}$$

where the received signal power that comes from the associated APs are given by $P^{u}_{\mathrm{sig}}$ for the $u^{\mathrm{th}}$ UE. Similarly, the received interference power from the other APs and the noise power for $u^{\mathrm{th}}$ UE are also denoted by $P^{u}_{\mathrm{int}}$ and $P^{u}_{\mathrm{noise}}$, respectively. It is important to note from the above expression that our system level evaluation simulations are designed to have a compatibility to support downlink (DL) CoMP transmission, where the received power from the multiple APs that are associated with the $u^{\mathrm{th}}$ UE are represented by $\sum_{\forall i\in\Im} P^{u}_{\mathrm{R},i}$. It is also important to note that the set $\Im$ contains the indexes of the APs that belong to association cluster. The wideband SINR metric sometimes called the "geometry factor" as the received signal strengths are highly dependent on the position and orientation of the UEs and APs.

It is important to emphasize that in a typical outdoor wireless network application, the overall system performance is generally limited by both the effective interference and noise powers. However, since we are considering an indoor application scenario, where the AP density in "Generic" and "Dense" deployment scenarios are significantly higher than the outdoor applications. Therefore, the main performance limiter in such scenarios becomes the interference, which is referred to as an "interference limited" network in the literature. Consequently, in our simulations, the metric to evaluate the system performance is chosen to be the SIR without loss of generality, where the SIR of the $u^{\mathrm{th}}$ UE is expressed as follows:

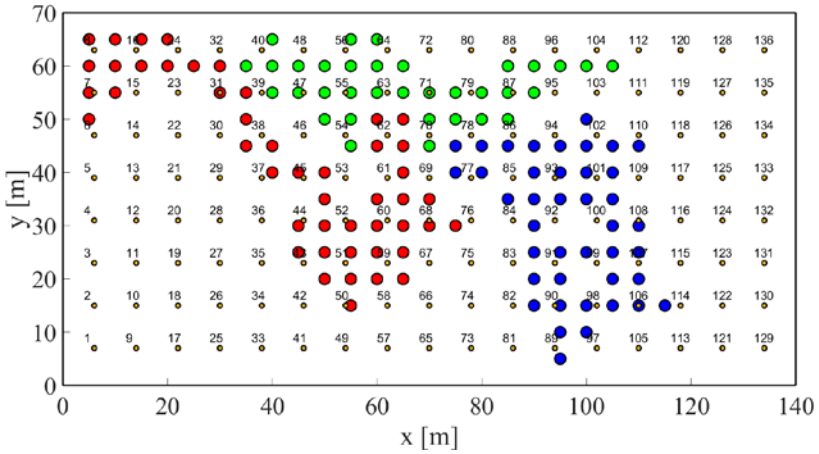
$$\mathrm{SIR}(\boldsymbol{u}) = \frac{\boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{sig}}}{\boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{int}}} = \frac{\sum_{\forall i\in\Im} \boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{R},i}}{\sum_{(\forall j, i\neq j)} \boldsymbol{P}^{\boldsymbol{u}}_{\mathrm{R},j}} \qquad \textbf{(6-12)}$$

In our Monte-Carlo trial-based computer simulations, the SIR distribution results for 20 000 mobile UEs, which take 100 random walk steps, are obtained. Please note that the number of UEs 20 000 is chosen to be large enough that law of large number applies without compromising very long simulation runtime. The results will be presented and investigated for each WAT when both the SP and CoMP transmission methods are adopted.

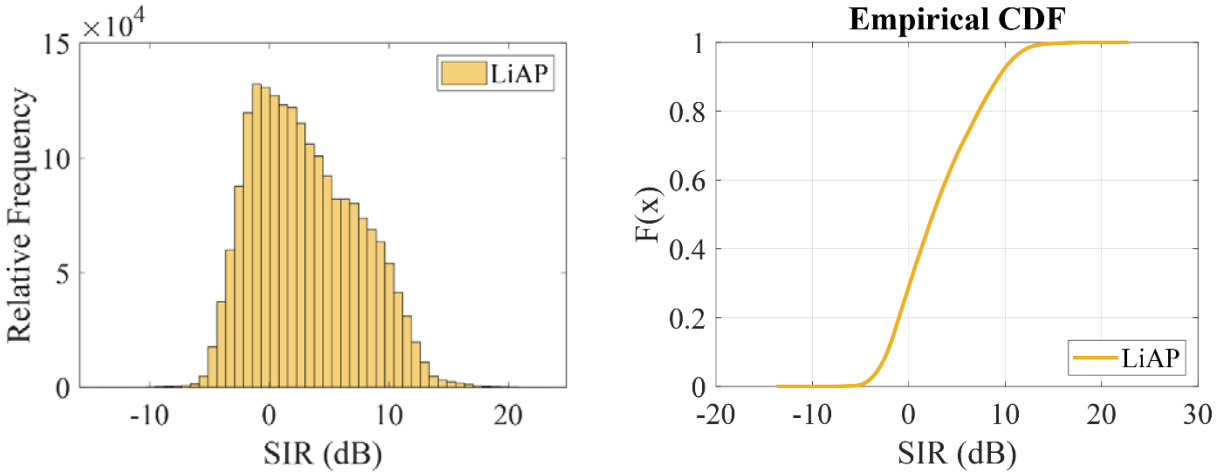
#### 6.3.1.1 Generic deployment with SP transmission

The following subsections present the SIR distribution results obtained by Monte-Carlo trial for 20 000 UEs with 100 step random walk and LiFi/Wi-Fi (5 GHz)/5GNR-I (3.5 GHz)/5GNR-II (26 GHz) WATs. Both the SP and CoMP transmission methods are also utilized in our simulations for "Generic" and "Dense" deployment scenarios to present the effect of deployment and AP utilization on the achievable area traffic capacity results.

The "Generic" deployment scenario for the LiFi APs and the labels of each APs are depicted in Figure 6-9(a). The red, green, and blue colours represent three realizations of random starting point 100 steps random walk for the UEs. Furthermore, the PDF distribution and empirical CDF plots for the SIR values at the UEs are also depicted when only LiFi APs are considered in Figure 6-9(b).




(a)




(b)

**Figure 6-9 (a) mobility map and AP deployment/labelling, (b) PDF and CDF distributions of the SIR values at the UEs**

For the sake readability and space issues, the deployment details as well as the PDF and CDF plots for each WAT, deployment scenario and transmission method is presented in Appendix.

The mean, variance, and median values of SIR values of "Generic" deployment with SP transmission method, that recorded at the UE side with respect to each WAT, is depicted in Figure 6-10 (a). Accordingly, the mean SIR value (in dB) for LiFi, Wi-Fi, 5GNR-I and 5GNR-II technologies become $3.06$, $-3.41$, $1.24$ and $-3.43$ dB, respectively. It can be seen from the results that LiFi outperforms the other three WATs by more than 1.8 dB in terms of the average SIR value. The main reason behind this is the favourable propagation characteristics of LiFi, where the PL is relatively high. Furthermore, higher density of the LiFi APs since the same APs will be used for illumination purposes simultaneously is another reason for this advantage. Similarly, the variance values of SIR for each LiFi, Wi-Fi, 5GNR-I and 5GNR-II technologies become $19.22$, $6.79$, $15.37$, $28.8$ dB, respectively. It is important to note that the main factor behind this variance is the random location updates, which alters the relative location of the UE with respect to its associated AP(s) for Wi-Fi, 5GNR-I and 5GNR-II technologies. However, in LiFi, the mobile UE experiences both the random orientation and orientation effects. The median value of the captured SIR for LiFi, Wi-Fi, 5GNR-I and 5GNR-II technologies could be given by $2.51$, $-3.8$, $0.31$, and $-3.91$ dB, respectively.

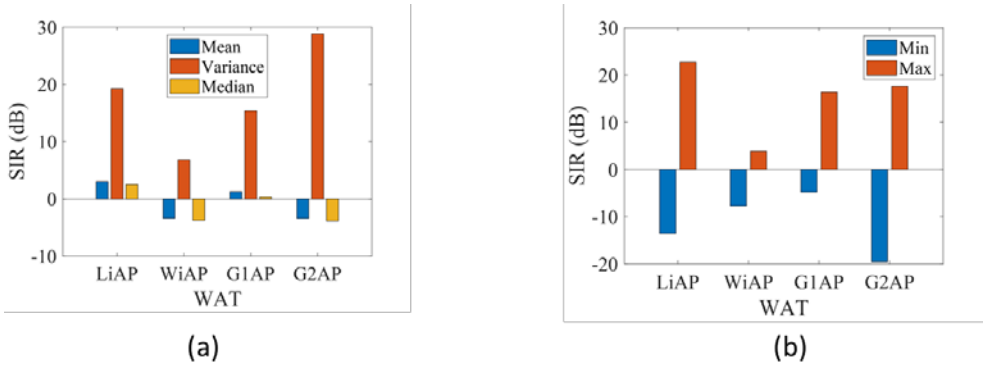



**Figure 6-10 (a) Mean, variance and median values, (b) Max and min values of the SIR distribution in dB**

Like Figure 6-11(a), the maximum and minimum values of the achieved SIR is depicted in Figure 6-11 (b) for the "Generic" deployment with SP transmission method. Accordingly, the maximum SIR values of 22.75, 3.83, 16.35, and 17.6 dB are obtained for LiFi, Wi-Fi, 5GNR-I and 5GNR-II technologies, respectively. As can be inferred from the figure that the AP density and higher PL advantages help LiFi to achieve the highest SIR values among all the WATs. In terms of the minimum values, the minimum achievable SIR values of $-13.62$, $-7{,}76$, $-4{,}77$, and $-19{,}57$ dB are obtained for LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies, respectively. This time the AP density worked against both LiFi and 5GNR-II technologies, where depending on the user location, the received interference could be significantly higher than the signal power.

### 6.3.1.2 Dense deployment with SP transmission

The results devised from the SIR distribution plots obtained for the "Dense" deployment scenario with the SP transmission are presented. In Figure 6-11 (a), the mean, variance, and median values of the SIR distribution plots obtained at the UE side are presented for a mobile UE that takes 100 random walk steps. Moreover, the random orientation model is also considered, which is important for LiFi transmission. As can be seen from Figure 6-11 (a), the achieved mean SIR values for LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs become $1.88$, $-4.83$, $-0.4$, and $-6.25$ dB, respectively. Like the "Generic" deployment case, LiFi outperforms other systems at least 2 dB in the "Dense" deployment scenario. It is important to note that mean SIR values are dropped significantly for each WAT, which is caused by the cell densification that increases the interference power significantly compared to the marginal increase in signal power.

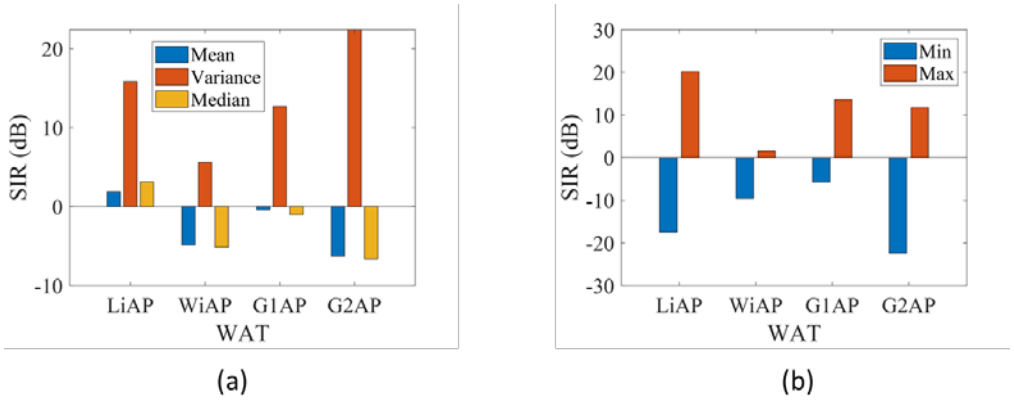



**Figure 6-11 The UE SIR distribution statistical properties when a random walk for 100 steps with LiFi, Wi-Fi, 5GNR-I and 5GNR-II APs are considered in the dense deployment scenario; (a) mean, variance and median values of SIR in dB, (b) max and min values of SIR in dB**

The variance values for LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs become $15.82$, $5.58$, $12.69$, and $22.4$ dB, respectively. Again, the high variance values caused by spatial distribution of the signal and interference powers are observed for also the "Dense" deployment scenario. Lastly, the median SIR distribution values of $3.12$, $-5.13$, $-1.04$, $-6.66$ dB, are obtained for LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies, respectively. The maximum and minimum values, in other words the range, of the SIR distributions are also plotted in Figure 6-11 (b). Accordingly, the maximum SIR values of $20.11$, $1.6$, $13.63$, and $11.69$ dB, are obtained for LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies, respectively. It is important to note that Wi-Fi suffers from the interference in terms of the peak SIR value the most.

The main reason for this is the deployment structure of the Wi-Fi APs as well as the propagation characteristics. However, other three technologies offer more favourable channel conditions for the given "Dense" deployment scenario. The minimum SIR values that are obtained by employing the LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies become $-17.49$, $-9.59$, $-5.72$, and $-22.44$ dB, respectively.

#### 6.3.1.3 Generic deployment with 3-CoMP transmission

In this subsection, the results obtained from the SIR maps will be summarized for "Generic" deployment scenario. Unlike the previously presented SP results, CoMP transmission technique is investigated in the following results. Accordingly, three highest signal power yielding APs are used in the signal power calculations, where all the other APs are considered as interferers. The mean, variance, and median values of the SIR distribution obtained at the UE for each WAT is depicted in Figure 6-12(a). The mean SIR values for the LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs become $8.72$, $0.32$, $11.08$, and $4.6$ dB, respectively. This time 5GNR-I outperforms all the other WATs by at least $2.3$ dB. The main reason behind this is the favourable PL characteristics in the 3.5 GHz band as well as the number of 5GNR-I APs. Since there is a total of four 5GNR-I APs, when three of them are contributing to the signal power there becomes only one interferer, which leads towards very high SIR values. However, this is not the case in LiFi, where there are 133 interferers rather the clear line-of-sight (LoS) signal power path along with the spatial location is the reason. If we investigate the variance values of the SIR distributions for the LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs we obtain $19.34$, $2.3$, $15.14$, and $45.45$ dB, respectively. The significant convergence of the SIR values from the average values in 5GNR-II, LiFi and 5GNR-I systems are stemmed from the relative location of the UE compared to the interfering and associated APs, which yield SIR to take extreme values near the corners of the environment. The median values obtained from the SIR distributions also become $7.84$, $-0.11$, $10.53$, and $3.89$ dB, respectively.

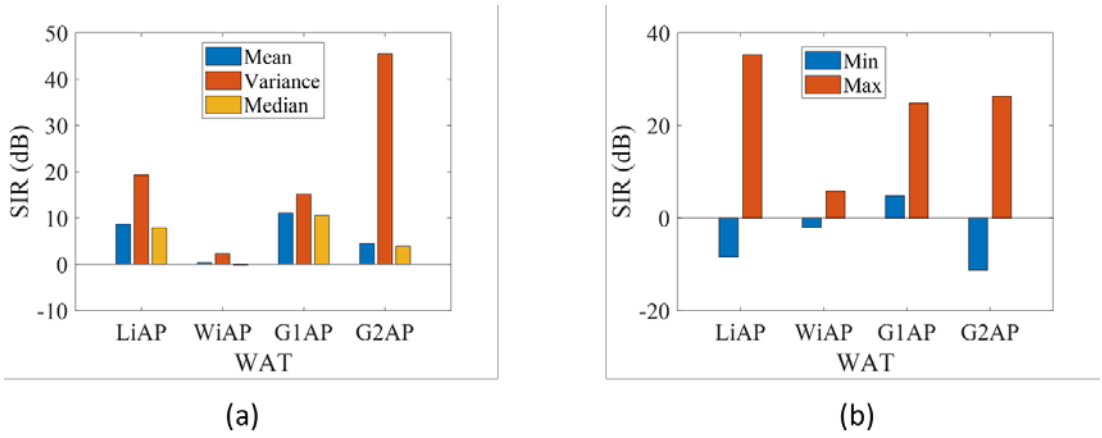



**Figure 6-12 The UE SIR distribution statistical properties when a random walk for 100 steps with LiFi, Wi-Fi, 5GNR-I and 5GNR-II APs are considered in the dense deployment scenario; (a) mean, variance and median values of SIR in dB, (b) max and min values of SIR in dB.**

The maximum and minimum achievable SIR value for each WAT for 3-CoMP transmission in "Generic" deployment scenario is given in Figure 6-12. Accordingly, the maximum SIR values of $35.28$, $5.87$, $24.86$, and $26.31$ dB are obtained for the LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies, respectively. It is interesting to note that LiFi achieves the highest SIR values among all the WATs even though the highest average SIR value is achieved by 5GNR-I technology.

Again, the main reason for this is the random orientation and mobility models, where the orientation impacts the link quality in a LoS dependent LiFi system severely. Therefore, the average value becomes significantly less compared to the max value in LiFi technology. The minimum SIR values for the LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies is given by $-8.38$, $-2.07$, $4.77$, and $-11.21$ dB, respectively. As can be seen from the values 5GNR-I achieves the maximum minimum SIR value. Like the average SIR value, picking three out of four APs as the associated APs and the last one as the interferer brings a significant advantage to the 5GNR-I system.

#### 6.3.1.4 Dense deployment with 3-CoMP transmission

Like the 3-CoMP transmission in "Generic" deployment, the "Dense" deployment scenario is investigated with the 3-CoMP transmission technique in this subsection. Note that the "Generic" and "Dense" deployment parameters are not different in SP and CoMP transmissions. The difference that deployment scenario brings to the CoMP is the higher degree of freedom of choosing $N_{\text{assoc}}$ out of all the APs per WAT compared to the SP transmission. The mean, variance and median values that are obtained from WAT specific SIR measuring simulations for the "Dense" deployment scenario with 3-CoMP transmission for a mobile UE is presented in Figure 6-13(a). Accordingly, the mean SIR values obtained by the LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies become $6.75$, $-1.1$, $5.78$, and $0.81$ dB, respectively. The inherent advantage of the LiFi technology persists in "Dense" deployment with 3-CoMP transmission, which outperforms the other WATs more than a dB. Similarly, the variance values of 13.22, 2.73, 8.9, and 33.97 dB are achieved by LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies, respectively. It is important to emphasize that the variance of 5GNR-II technology is significantly higher than the all the other WATs, due to the geo-spatial distribution of the 5GNR-II APs.

In Figure 6-13(b), the maximum and minimum achieved SIR values per WAT is depicted. Accordingly, the maximum value of $33.48$, $4.01$, $17.36$, and 19.34 dB, are obtained by utilizing LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs. Again, LiFi yields the highest achievable SIR values among all the WATs. Moreover, the minimum values of $-12.65$, $-4.1$, $1.63$, and $-13.45$ dB, are obtained by employing LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs, respectively.

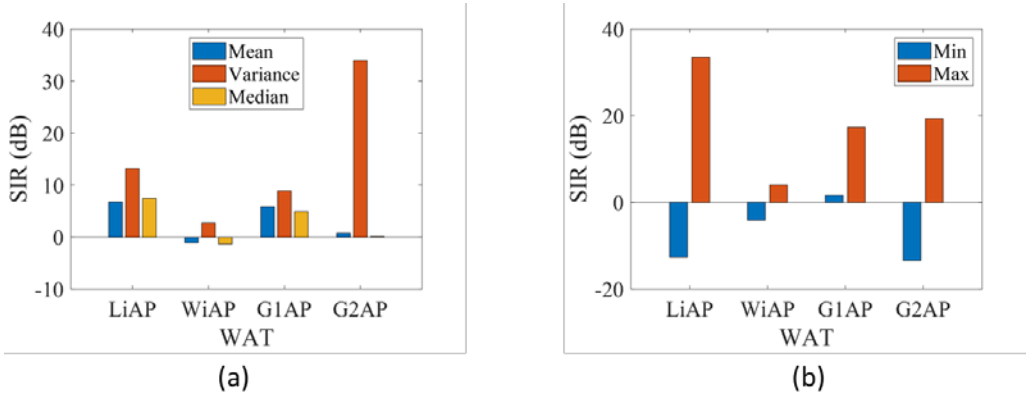



**Figure 6-13 The UE SIR distribution statistical properties when a random walk for 100 steps with LiFi, Wi-Fi, 5GNR-I and 5GNR-II APs are considered in the dense deployment scenario; (a) mean, variance and median values of SIR in dB, (b) max and min values of SIR in dB**

### 6.3.2 Achievable area traffic capacity results

This section reports the WAT-specific and aggregated achievable area traffic capacity results obtained from the SIR value distributions measured at the mobile UE. The result will be provided for two deployment scenarios; (i) "Generic" and (ii) "Dense" as well as two transmission techniques; (i) SP and (ii) CoMP. Note that OBJ-TECH-4 in the 5G-CLARITY project envisages more than 500 Mbps per meter square aggregate system area capacity in indoor scenarios through smart RRM algorithms. To achieve this goal, in 5G-CLARTY project the simulation scenarios considered various WATs, bandwidth values, deployment scenarios and transmission modes/techniques.

In Figure 6-14, the achievable area traffic capacity values for the given scenario parameters throughout this section is depicted. Specifically, in Figure 6-14(a), area traffic capacity values for "Generic" deployment with the SP transmission is given. The average achievable traffic capacity values of $290.60, 79.55, 225.87, 109.94$ Mbps/$\mathrm{m}^2$ are achieved by using LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs, respectively. It is important to emphasize that LiFi achieves the highest area traffic capacity results by outperforming the other WATs at least 64 Mbps/$\mathrm{m}^2$.

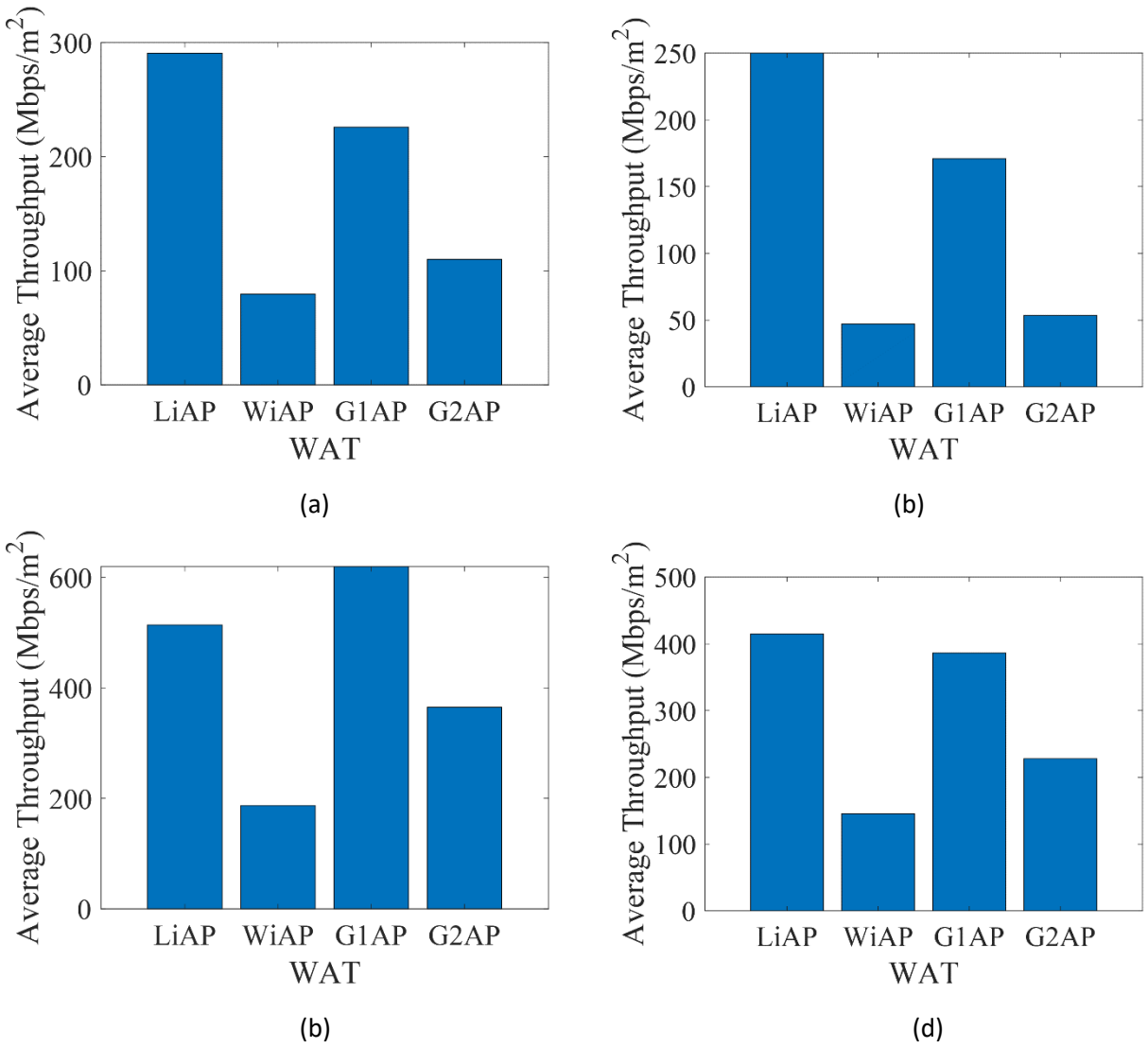



**Figure 6-14 Achievable throughput results for UEs with 100 steps random walk; (a) "Generic" deployment with SP transmission, (b) "Dense" deployment with SP transmission, (c) "Generic" deployment with CoMP-3 transmission, and (d) "Dense" deployment with CoMP-3 transmission**

**Table 6-4 Average Achievable Area Traffic Capacity Values for the Considered Scenarios**

| Transmission | Single Point (SP) Transmission | | Coordinated Multi-point (CoMP) Transmission | |
|---|---|---|---|---|
| Deployment | Generic | Dense | Generic | Dense |
| Aggregate | $C = 705.96$ Mbps/m$^2$ | $C = 521.53$ Mbps/m$^2$ | $C = 1685.93$ Mbps/m$^2$ | $C = 1175.01$ Mbps/m$^2$ |

For the given scenario, the aggregated area traffic capacity value is also given in Table 6-4. Accordingly, the aggregate area traffic capacity of the "Generic" deployment with the SP transmission becomes $705.96$ Mbps/m$^2$.

Similarly, the WAT-specific achievable area traffic capacity values are depicted in Figure 6-14(b) for the "Dense" deployment scenario with the SP transmission method. Accordingly, the area spectral efficiency values of $249.91$, $47.06$, $170.96$, and $53.6$ Mbps/m$^2$ are obtained by employing LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs, respectively. Again, the LiFi technology outperforms the other three WATs by at least $78$ Mbps/m$^2$ with its inherent high AP density and PL coefficient advantages. For this scenario, the aggregate area traffic capacity of $521.53$ Mbps/m$^2$ is achieved as given in Table 6-4. The reason behind this slight drop compared to the "Generic" deployment is the enhanced interference power, which increases with decreased inter AP distance.

In Figure 6-14(c), the WAT-specific achievable area traffic capacity is given for the "Generic" deployment when 3-CoMP transmission is adopted. Thus, the area traffic capacity values of $514.16$, $187.33$, $619.55$, and $364.89$ Mbps/m$^2$, respectively, are achieved by utilizing the LiFi, Wi-Fi, 5GNR-I, and 5GNR-II WATs, respectively. Note that the achievable area spectral efficiency values are enhanced with the 3-CoMP for all the WATs. This time the 5GNR-I technology outperforms the other three WATs more than $105$ Mbps/m$^2$. Table 6-4 explains that the aggregate area traffic capacity becomes $1.686$ Gbps/m$^2$, which corresponds to more than 2-folds area traffic capacity enhancement when CoMP is implemented.

Lastly, in Figure 6-14(d), the WAT-specific achievable area traffic capacity is given for the "Dense" deployment scenario when 3-CoMP transmission is used. Accordingly, the achievable area traffic capacity values become $414.85$, $145.65$, $386.45$, and $228.05$ Mbps/m$^2$ is achieved by utilizing LiFi, Wi-Fi, 5GNR-I, and 5GNR-II technologies, respectively. It is important to emphasize that the LiFi achieves at least $28$ Mbps/m$^2$ more area traffic capacity value compared to other three WATs. Table 6-4 indicates that the aggregate area traffic capacity of $1.175$ Gbps/m$^2$ is achieved when "Dense" deployment is employed with the 3-CoMP transmission method. Like the SP transmission, the overall area traffic capacity reduces in "Dense" configuration compared to the "Generic" deployment due to the enhanced interference power. Therefore, the highest area traffic capacity is achieved when the "Generic" deployment with CoMP transmission method is adopted.

## 6.4 Section summary

This section investigated the performance of the propped 5G-CLARITY PHY layer architecture by our system level simulations. The proposed RAN in 5G-CLARITY project consists of simultaneous operation of 3GPP and non-3GPP access technologies. Specifically, a new technology, namely LiFi, which uses the optical part of the electromagnetic spectrum is a promising solution for enhancing the achievable link quality in indoor private network applications. Since LiFi uses the light fixtures in an indoor environment for both the communication and illumination purposes it brings enhanced power efficiency and security merits as well as reduced deployment costs. On top of those, LiFi brings enhanced link quality and area spectral efficiency

potential to the 5G-CLARITY system architecture. Furthermore, it is also considered 3.5 5GNR, 5 GHz Wi-Fi and 26 GHz (mmWave) 5GNR WATs concurrently to achieve the maximum capacity that 5G-CLARITY architecture could offer.

# 7 Summary and Concluding Remarks

This deliverable, 5G-CLARITY D3.3, reported in detail the last tier of the refinements and potential diversions from the proposed initial design in D3.1 and those further developed in 5G-CLARITY D3.2. Particularly, the final 5G-CLARITY system architecture is based on underlying theoretical analysis and fundamentals, as well as practical design considerations. The technical objectives, OBJ-TECH-2 to OBJ-TECH-5, are achieved by the development of CBRS based spectrum sharing framework, MPTCP based AT3S and eAT3S multi-connectivity, advanced resource management and network slicing, multi-WAT coexistence and localization frameworks, and 5G-CLARITY modular multi-WAT simulator. These development activities have significantly enhanced the previous frameworks introduced in 5G-CLARITY D3.1 and 5G-CLARITY D3.2, which are demonstrated through simulation and experimental results for the final designed 5G-CLARITY multi-WAT architecture. It is important to note that the findings and final design details developed in 5G-CLARITY D3.3 will be reflected in the 5G-CLARITY D2.4, D2.5, D4.3 and D5.4.

Regarding D3.3 OBJ-1, D3.3 OBJ-2 and D3.3 OBJ-3, Section 2 introduced an E2E CBRS based spectrum sharing framework, which is supported by an E2E cloud native, and O-RAN Alliance disaggregated and open RAN intelligence. The framework testbed has been developed to enable the co-existence of CBRS and 5G-CLARITY multi-wireless technology integration framework. In Section 3, the 5G-CLARITY advanced multi-connectivity and multi-RAT Aggregation framework has been extended from 5G-CLARITY D3.2 and evaluated. The extensions and architecture enhancement mainly include (i) functional testbed with an API to manage the steering policies in MPTCP based AT3S user plan function in real-time, (ii) integration a multi-weight factor assignment in the eAT3S algorithm presented in D3.2 to support adaptive 5G URLLC and eMBB traffic load-balancing steering mode according to their QoS requirements, (iii) adaptive and probing based available bandwidth estimation on the multi-WAT interfaces to maximize their capacity utilization, (iv) implementation of the 5G-CLARITY multi-connectivity framework architecture with real 5GNR, Wi-Fi and LiFi access networks.

The enhanced 5G-CLARITY multi-connectivity framework architecture allowed to achieve an aggregated throughput of 1.45 Gbps, which mainly referred to the integration of MPTCP and AT3S operations, the designed reliability requirements based the air interface of 5GNR. The air interface latency was near the KPI objective, though new measurements are needed with mature 5GNR equipment. The updated eAT3S algorithm enables the inclusion of multiple network parameters in the steering decisioning. This would enable private/public network operators to derive user or environment specific network policy configurations to efficiently utilize 3GPP and non-3GPP networks.

Regarding D3.3 OBJ-4, Section 4 significantly enhanced the utility scheduler-based LiFi attocellular networks slicing and the interface airtime scheduler-based Wi-Fi AP slicing approaches, introduced in D3.2, by developing advanced intelligent self-network slicing and resource management schemes. The LiFi and Wi-Fi networks slicing developments mainly include (i) deep multiagent reinforcement learning-based autonomic LiFi attocellular network slicing (ii) a utility scheduler at LiFi APs and a Global-Airtime Deficit Round Robin (G-ADRR) at Wi-Fi APs to enforce the quota guarantees of network slices, and (iii) a centralized scheduling policy which tunes the airtime weights of Wi-Fi APs distributed over a geographical area according to their traffic load and network slices requirements.

The proposed network sharing schemes for the interface airtime of Wi-Fi APs and the downlink channel capacity of network LiFi APs are designed as self-organizing network (SON) functions. These use the information collected from the LiFi and Wi-Fi network APs to identify the patterns of varying service and traffic demands in temporal and spatial space across the network. This allowed to dynamically, in an autonomic manner, allocate the resource quota guarantees for each network slice across the network.

Regarding D3.3 OBJ-5, Section 5 explained the extensions that were developed to improve the accuracy of the 5G-CLARITY multi-WAT localization framework introduced in 5G-CLARITY D3.2. The extensions include the development of positioning technologies, namely, (i) novel positioning technologies using radio frequency in the sub-6 GHz and 60 GHz mmWave bands and (ii) enhanced light-based positioning technologies using more LiFi APs.

The LiFi technology is primarily used for data transmission, but within the 5G-CLARITY it was extended to be used also for positioning and localization. The initial results show that the achieved positioning precision is in the range of up to a couple of meters. This has been improved by combining VLC (visible light communication) using photodetectors and OCC (optical camera communication) based positioning technology. The developed localization-based VLC and OCC integration allowed to achieve a positioning precision of few centimetres with minimal hardware requirements.

Regarding D3.3 OBJ-4, Section 6 explains the extensions development to the 5G-CLARITY multi-WAT simulator introduced in D3.2, by providing an optimistic upper bound on the network deployment area capacity, considering more realistic assumptions and simulation environment. The simulator extensions mainly include (i) deployment of 5GNR, 5 GHz Wi-Fi and 26 GHz (mmWave) 5GNR WATs link quality and link performance models to conduct link level and higher-level system performance evaluations, (ii) deployment of network dense mode with CoMP transmission.

It is important to emphasize that the concurrent deployment of 5GNR, 5 GHz Wi-Fi and 26 GHz (mmWave) 5GNR WATs allowed to achieve the maximum capacity the 5G-CLARITY architecture could offer. The achieved area traffic capacity is still higher than the minimum value of 500 Mbps/$m^2$, which is one of the significant achievements in line with the technical objectives of the 5G-CLARITY project. Besides, other results have shown the potential of LiFi access network to provide enhanced link quality and area spectral efficiency to the 5G-CLARITY system architecture.

# 8 Bibliography

[1] 5G-CLARITY D3.1, "State-of-the-art review and initial design of the integrated 5GNR/Wi-Fi/LiFi network frameworks on coexistence, multi-connectivity, resource management and positioning," September 2020. Available: https://www.5gclarity.com/wp-content/uploads/2020/09/5G-CLARITY_D3.1.pdf .

[2] 5G-CLARITY D3.2, "Design Refinements and Initial Evaluation," 2021. Available: https://www.5gclarity.com/wp-content/uploads/2021/06/5GC-CLARITY_D32.pdf.

[3] 5G-CLARITY D2.2, "Primary System Architecture October," 2020. Available: https://www.5gclarity.com/wp-content/uploads/2020/12/5G-CLARITY_D41.pdf.

[4] 5G-CLARITY D2.1, "Use cases and requirements," 2020. Available: https://www.5gclarity.com/wp-content/uploads/2020/06/5G-CLARITY_D2.1.pdf.

[5] 5G-CLARITY D2.3, "Primary System Architecture," 2021. Available: https://www.5gclarity.com/wp-content/uploads/2021/10/5G-CLARITY_D23.pdf .

[6] 5G-CLARITY D2.5, "New services and applications," to be submitted soon. .

[7] FCC, "Report and Order and Second Further Notice of Proposed Rulemaking," April 21, 2015.

[8] FCC, "Order on Reconsideration and Second Report and Order," May 2, 2016.

[9] Winnforum, [Online]. Available: https://cbrs.wirelessinnovation.org/standards.

[10] CBRS Alliance, [Online]. Available: https://www.cbrsalliance.org/.

[11] "FCC - Auctions - Auction 105 results," [Online]. Available: https://www.fcc.gov/auction/105/round-results.

[12] DSA, "Automated Frequency Coordination - An established tool for modern spectrum management," March 2019.

[13] ATIS IOC, "International Mobile Subscriber Identity (IMSI) Assignment and Management Guidelines for Shared HNI for CBRS Range- Version 1.0," May 2018.

[14] OnGo/CBRS Alliance, "CBRS Network Services Use Cases and Requirements, CBRSA-TS-1001, V3.0.0," 18th Feb 2020.

[15] OnGo/CBRS Alliance, "CBRS Network Services Stage 2 and 3 Specification, CBRSA-TS-1002, V3.0.0," 18th Feb 2020.

[16] OnGo/CBRS Alliance, "CBRS Coexistence Technical Specifications, CBRSA-TS-2001, V3.0.0," 18th Feb 2020.

[17] OnGo/CBRS Alliance, "Extended Subscribers Authentication Technical Specification, OnGo-TS-1003, V4.0.0," December 15, 2020.

[18] OnGo/CBRS Alliance, "CBRS Network Services Use Cases and Requirements, OnGo-TS-1001, V4.1.0," November 10, 2021.

[19] OnGo/CBRS Alliance, “CBRS Network Services Stage 2 and 3 Specification, OnGo-TS-1002, V4.1.0,” November 10, 2021.

[20] OnGo/CBRS Alliance, “CBRS Coexistence Technical Specifications, OnGo-TS-2001, V4.0.0,” July 29, 2021.

[21] OnGo/CBRS Alliance, “CBRS Coexistence CSAS Profile, OnGo-TS-2002, V4.0.0,” March 16, 2021.

[22] JuanjoRepo., “https://github.com/jorgenavarroortiz/multitechnology_testbed_v0”.

[23] J. Navarro-Ortiz, “Multiconnectivity Testbed: REST API and Grafana Demo,” 2021. [Online]. Available: https://user-images.githubusercontent.com/17797704/125672786-4ee5dcec-b28b-4b11-885f-fd782e0a948f.mp4.

[24] Swagger, “API Development for Everyone,” [Online]. Available: https://swagger.io/.

[25] Uvicorn, “An ASGI Web Server for Python,” [Online]. Available: https://www.uvicorn.org/.

[26] VideoMPTCP, “https://www.youtube.com/watch?v=TxSEYxavJxI”.

[27] MPTCP, “MultiPath TCP - Linux Kernel implementation,” [Online] Available: https://www.multipath-tcp.org [Accessed 17 05 2022] , 2022.

[28] Lagscope, “[Online], Available at: https://github.com/microsoft/lagscope”.

[29] B. Kimura, D. Lima and A. Loureiro, “Packet scheduling in multipath TCP: Fundamentals, lessons, and opportunities,” *IEEE Systems Journal,* pp. 1445-1457, 2020.

[30] M. Martineau and O. O. O., “sing Upstream MPTCP in Linux Systems,” Netdev 0x14 conference, 2020.

[31] B. May, “Shuttle: how it works,” 2022.

[32] C. Windy, M. Lv and L. Yang, “ShadowSocks,” Available: https://github.com/shadowsocks/shadowsocks-libev (last accessed on 4 May 2022)., 2022.

[33] A. Bizjak, “tun2socks BadVPN,” Available: https://github.com/ambrop72/badvpn (last accessed on 4 May 2022), 2022.

[34] IP2Socks, “IP2Socks,” vailable: https://github.com/FlowerWrong/ip2socks (lasta accessed on4 May 2022), 2022.

[35] d. V. R. Berg, “TCPPing,” Available: http://www.vdberg.org/~richard/tcpping(last accessed on 4 May 2022), 2022.

[36] 5G-CLARITY D4.2, “Validation of 5G-CLARITY SDN/NFV Platform , Interface Design with 5G Service Platform, and Initial Evaluation of ML Algorithms,” , 2021. Available: https://www.5gclarity.com/wp-content/uploads/2021/11/5G-CLARITY_D42.pdf.

[37] 5G-CLARITY D4.3, “Evaluation of E2E 5G Infrastructure and Service Slices, and of the Developed Self-Learning ML Algorithms,” 2022. [Online].

[38] J. Iyengar and M. Thomson, “RFC 9000,QUIC: A UDP-Based Multiplexed and Secure Transport,” RFC Series, Internet Engineering Task Force (IETF), 2021.

[39] Y. Ma, Q. D. Coninck, O. Bonaventure and C. Huitema, "Internet-Draft: Multipath Extension for QUIC - draft-ietf-quic-multipath-01," Internet Engineering Task Force (IETF), 2022.

[40] L. Breslau, P. Cao, L. Fan and G. Phillips, "Web caching and zipf-like distributions: evidence and implications," in *IEEE INFOCOM*, 1999.

[41] M. Cha, H. Kwak, P. Rodriguez and Y. yeol, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Internet Measurement Conference (IMC)*, San Diego, CA, 2007.

[42] 5G-CLARITY D5.1, "Specification of Use Cases and Demonstration Plan," 2021. Available: https://www.5gclarity.com/wp-content/uploads/2021/02/5G-CLARITY_D51.pdf .

[43] 3GPP TR 38.901, *Study on channel model for frequencies from 0.5 to 100 GHz.*

[44] L. Jorguseski, "Self-Organizing Networks in 3GPP: Standardization and Future Trends," *IEEE Communication Magazine,* vol. 52, no. 12, pp. 28-34, 2014.

[45] H. Alshaer and H. Haas, "Dynamic LiFi Attocellular Networks Slicing for 5G," in *ICC*, 2022.

[46] H. Alshaer, H. Haas and O. Y. Kolawole, "An Optimal Networked LiFi Access Point Slicing Scheme for Internet-of-Things," in *Proc of ICC Workshop*, Montreal, 2021.

[47] H. Alshaer and H. Haas, "Bidirectional LiFi attocell access point slicing scheme," *IEEE Transactions on Network and Service Management,* vol. 15, no. 3, pp. 909-922, 2018.

[48] I. Vila, J. P. Romero, O. Sallent and A. Umbert, "A Multi-Agent Reinforcement Learning Approach for Capacity Sharing in Multi-Tenant Scenarios," *IEEE Transactions on vehicular Technology,* vol. 70, no. 9, pp. 9450-9464, 2021.

[49] R. Sutton and A. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[50] A. Betzler, D. C. Mur and M. Catalan, "G-ADRR: Network-wide slicing of Wi-Fi networks with variable loads in space and time," *IEEE Transactions on Mobile Computing,* pp. 1-14, 2021.

[51] V. Sciancalepore, K. Samdanis and X. Costa-Perez, "Mobile traffic forecasting for maximizing 5g network slicing resource utilization," in *IEEE INFOCOM* , Atlanta, GA, USA, 2017 .

[52] T. Hiland-Jrgensen, H. P., B. A. and others, "Polifi: Airtime policy enforcement for wifi," in *IEEE WCNC*, Marrakesh, Morocco, 2019.

[53] J. Malinen, "Hostapd: IEEE 802.11 AP, IEEE 802.1 X/WPA/WPA2/EAP/RADIUS Authenticator," https: //w1.fi/hostapd/ [Online; accessed 05-Feb-2021], 2013,.

[54] T. Hiland-Jrgensen, P. Hurtig and A. Brunstrom, "Polifi: Airtime policy enforcement for wifi," IEEE WCNC, 2019 .

[55] J. Turnbull, "Monitoring with Prometheus," 2018.

[56] M. Catalan, "Prometheus hostapd exporter," https://prometheus. io/docs/instrumenting/exporters/, 2020, [Online; accessed 05- Feb-2021]., 2020.

[57] D. Goldfarb and A. Idnani, "A numerically stable dual method for solving strictly convex quadratic

programs," *Mathematical Programming ,* pp. 1-33, September 1983.

[58] S. Caron, "QP Solvers for Python," https://prometheus.io/docs/ instrumenting/exporters/ , [Online; accessed 05-Feb-2021], 2020.

[59] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (Corresp.)," *IEEE Transactions on Information Theory,* vol. 13, pp. 619-621,, October 1967.

[60] C. Lee, C. Shen, C. Cozzan, R. M. Farrell and J. S. Speck, "Gigabit-per-second white light-based visible light communication using near-ultraviolet laser diode and red-, green-, and blue-emitting phosphors," *Optics Express,* vol. 25, no. 15, pp. 17480-17487, 2017.

[61] T. Yamakami, "A Zipf-Like Distribution of Popularity and Hits in the Mobile Web Pages with Short Life Time," in *Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'06)*, Taipei, Taiwan, 2006.

[62] P. Gill, M. Arlitt, Z. Li and A. Mahanti, "YouTube Traffic Characterization: A View From the Edge," HP Laboratories, Palo Alto, 2007.

[63] L. Breslau, P. Cao, L. Fan, G. Phillips and S. Shenker, "Web caching and Zipf-like distributions: evidence and implications," in *IEEE INFOCOM '99. Conference on Computer Communications. Proceedings*, New York, NY, USA, 1999.

[64] Y. Liu, F. Li, L. Guo, B. Shen, S. Chen and Y. Lan, "Measurement and Analysis of an Internet Streaming Service to Mobile Devices," *IEEE Transactions on Parallel and Distributed Systems,* vol. 24, no. 11, pp. 2240-2250, 2013.

[65] I. Grigorik, "The "Average Page" is a myth," [Online]. Available: https://www.igvita.com/2016/01/12/the-average-page-is-a-myth/. [Accessed 18 04 2022].

[66] M. D. Soltani, A. A. Purwita, Z. Zeng, H. Haas and M. Safari, "Modeling the Random Orientation of Mobile Devices: Measurement, Analysis and LiFi Use Case," *IEEE Transactions on Communications,* pp. 2157-2172, 2019.

[67] B. Ghimire and H. Haas, "Self-organising interference coordination in optical wireless," *EURASIP J Wireless Com Network,* vol. 2012, no. 131, pp. 1-15, 2012.

[68] H. Elfadil, M. Ibrahim Ali and M. Abas, "Fractional frequency reuse in LTE networks," in *2nd World Symposium on Web Applications and Networking (WSWAN)*, Sousse, Tunisia, 2015.

[69] S. Bassoy, H. Farooq, M. a. Imran and A. Imran, "Coordinated Multi-Point Clustering Schemes: A Survey," *IEEE Communications Surveys & Tutorials,* vol. 19, no. 2, pp. 743-764, 2017.

[70] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 12.3.0 Release 12)," 3GPP, 2014.

[71] 3GPP, "5G NR Physical layer procedures for data (3GPP TS 38.214 version 16.2.0 Release 16)," 3GPP, 2020.

[72] T. Cogalan, "Dense wireless network design and evaluation – an aircraft cabin use case," The University of Edinburgh, Edinburgh, 2018.

[73] A. R. Ramos, B. C. Silva, M. C. Lourenço, E. B. Teixeira and F. J. Velez, "Mapping between Average SINR

and Supported Throughput in 5G New Radio Small Cell Networks," in *22nd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Lisbon, Portugal, 2019.

[74] 3GPP, "5G Procedures for the 5G System (3GPP TS 23.502 version 15.2.0 Release 15)," 3GPP, 2018.

[75] C. Lee, C. Shen, C. Cozzan, R. Farrell, J. Speck, S. Nakamura, B. Ooi and S. DenBaars, "Gigabit-per-second white light-based visible light communication using near-ultraviolet laser diode and red-, green-, and blue-emitting phosphors," *Opt. Express,* pp. 17480-17487, 2017.

[76] OFCOM, "OfW48 UK Frequency Allocations for Fixed (Point-to-Point) Wireless Services and Scanning Telemetry," 2020.

[77] C. Lee, C. Shen, C. Cozzan and R. Farrell, "Gigabit-per-second white light-based visible light communication using near-ultraviolet laser diode and red-, green-, and blue-emitting phosphors,," *Opt. Express,* pp. 17480-17487, 2017.

[78] H. Chun, P. Manousiadis, S. Rajbhandari, D. Vithanage and G. Faulkner, "Visible Light Communication using a Blue GaN µLED and Fluorescent Polymer Colour Converter," *IEEE Photonics Technology Letters,* vol. 26, no. 20, pp. 2035-2038, 2014.