Performance Study of 5G Indoor Small Cells for Industrial MEC

Amy Sokhna Sidibé
University of Helsinki
amy.sidibe@helsinki.fi

Saimanoj Katta
Cumucore Oy, Espoo, Finland
saimanoj.katta@cumucore.com

Jose Costa Requena
Aalto University
jose.costa@aalto.fi

***Abstract*—The Fifth-generation (5G) and future 6G are targeted for industrial environments where reliability and low latency processing are required for automation of factories. Multi-access Edge Computing (MEC) is the emerging paradigm expected to address those requirements. In this study we utilize an Open Radio Access Network (ORAN) equipment to conduct extensive performance measurements, primarily focusing on latency and bandwidth metrics. Industrial MEC use cases require the deployment of indoor small cells capable of providing low latency communications for small coverage. In this paper, we are interested in analyzing the effectiveness of a co-deployment of MEC with the access network as promised by the 3rd Generation Partnership Project (3GPP) [1]. Leveraging the iperf tool, we conduct measurements both within our custom MEC platform, collocated with the gNodeB (gNB), and externally against a public iperf server deployed in the cloud. By comparing these measurements, the paper provides valuable insights into the efficiency of ORAN technology in industrial MEC environments, shedding light on its potential advantages and limitations. This empirical evaluation serves to inform future deployments and optimizations, contributing to the advancement of efficient and reliable edge computing solutions for industrial use cases.**

***Index Terms*—5G, 6G, MEC, Open RAN, latency, bandwidth**

I. Introduction

With the revolution of Industry 4.0, smart factories are integrating new technologies such as Internet of Things (IoT) devices (e.g. advanced sensors, etc.) into their processes [2]. The widespread use of smart and mobile devices goes in parallel with an increase in data consumption. This increase requires network operators to find ways to fulfill the promise of better customer experience in the network especially for ultra-low latency and reliability. Moreover, 5G technology is the first generation of mobile networks that is designed with a set of features such as 5G Local Area Network (LAN) and Time Sensitive Networking (TSN) specifically to support industrial communications [3], and to open up new business opportunities for mobile operators. The demand for better consumer experience and support for industrial communications requires new types of deployments such as Non-Public Networks (NPN) (also called campus networks, dedicated or private networks) and technologies such as network slicing and Multi-access Edge Computing (MEC).

Edge computing is a new paradigm introduced with 5G to bring the data processing of applications closer to their physical location, either the end user or the data source. This reduces the traffic load that reaches the mobile operator infrastructure and out-sources the computing process to edge nodes closer to the user. Another area of consideration in edge computing is security and data privacy. At an industry level (or country level, e.g., for healthcare), data processing at the edge is foreseen as an advantage to keep sensitive information locally, nevertheless it also brings many challenges as presented in [4], [5]. MEC is an application scenario of edge computing used for mobile networks with an access agnostic approach. It is a network platform standardized by the European Telecommunications Standards Institute (ETSI) [6]. It aims to provision the edge nodes with computing and data processing resources required for applications. The goal is to achieve better performance (e.g., low latency, reduced network congestion, higher data rate, etc.) on services requested by customers or industrial machines. The 3rd Generation Partnership Project (3GPP), which is the forum defining mobile communication specifications, has recently completed the Release 17 of standards defining in its architecture the native support of edge computing services [7].

In this paper a prototype of 3GPP Release 17 MEC architecture was implemented and integrated with a 5G Core network (5GC). The goal is to measure the network performance when using different Radio Access Networks (RAN) to estimate overall MEC performance in different life deployments. Significant efforts have been done by vendors to produce compliant product for indoor coverage [8], [5]. However, under different scenarios and workload effects, estimations on serving latency do not comply with the reality check. The complexity of such system design combined with demanding metrics of computation, slow the launch process of end-to-end (e2e) ready edge native platforms. Our measurements show that the latency reached between 9 ms to 100 ms for MEC with an ORAN gNB. However, due to complexities in the RAN management, poor throughput was measured with our multiple simulation scenarios. In regards with the different components used to perform the measurements, our goal is to provide an overview of actual indoor small cells collocated with MEC infrastructure and give concrete insights on the bottlenecks for evaluation and adaptation to future research.

Contributions. We present a custom implementation of MEC platform adapted to 3GPP's specifications. The platform is collocated with ORAN equipment used for performance measurements. We present our results on latency and bandwidth metrics with different spectrum usage configurations for Time Division Duplexing (TDD).

II. Background

A. *Indoor Small Cells*

Small cells and network densification were introduced in 3GPP Release 12 with the requirements on 4G Long Term

Evolution (LTE) small cells. As transmissions are more challenging inside buildings, there is a need to improve signal strength to meet the rising demand in data rate. The transformation in data consumption enhanced by 4G LTE enabled fundamental changes in 5G context. 5G networks in regards with the stringent requirements in terms of speed, ultra-low latency and heterogeneity of devices, rely on indoor network architectures to enhance consumers' experience. It allows operators to provide a variety of services with use cases tailored for indoor communications e.g., massive Machine Type Communication (mMTC), ultra-reliable Low Latency Communication (urLLC), enhanced Mobile Broadband (eMBB).

Smaller scale networks provide localised coverage (i.e. reliable connectivity in specific zones), hence are designed to be integrated with portable equipments on which vendors are investing in their production. These are Software Defined Radio (SDR) systems that offer network operators the flexibility to tune their infrastructure based on user requirements. Another key aspect of such deployments is that it allows users to perform seamless handovers between cells during mobility. The prominence of such cells introduced technologies like Cloud RAN (C-RAN) and ORAN to overcome the lack of flexibility and cost problems of traditional networks. Several benefits are listed in [9] as improvements of RAN technologies in wireless systems.

C-RAN offers RAN functions over a generic compute platform instead of a purpose-built hardware platform like in traditional networks. Network Functions (NFs) are cloud-native software as microservices deployed in containers (isolation) to run on Commercial off-the-shelf (COTS) hardware platforms which simplifies network management.

ORAN is considered an evolution of C-RAN concepts through virtualized RAN (vRAN). The design of ORAN targets the following solutions: open interoperable interfaces, hardware-software disaggregation, enable AI-driven solutions in the RAN domain (e.g. RAN Intelligent Controller (RIC)) [9]. The principle of White Box Hardware for ORAN allows an optimal instantiation of Central Units (CUs) and Distributed Units (DUs) for scalable network deployments.

B. 5G Architecture

5G is an evolution from the monolithic architecture in 4G LTE into a more distributed approach called Service Based Architecture (SBA). The objective is to facilitate the integration of new services and functionalities using NFs without major changes and reconfigurations.

The SBA consists of several NFs all interconnected through well-defined REST API interfaces and service endpoints defined in 3GPP as shown in Fig. 1. 5G Core Network requires a set of NFs to provide default functionalities to register, authenticate and provide data sessions to devices [10], [11]. The SBA's flexibility allows to add new NFs with additional functionalities that would be defined in future releases of 3GPP specifications. Fig. 1 includes some of the NFs implemented in the 5GC to be used for testing.

C. Multi-Access Edge Computing

The 3GPP Release 17 has defined a set of new functions to enable edge computing and interactions with NFs in the 5G

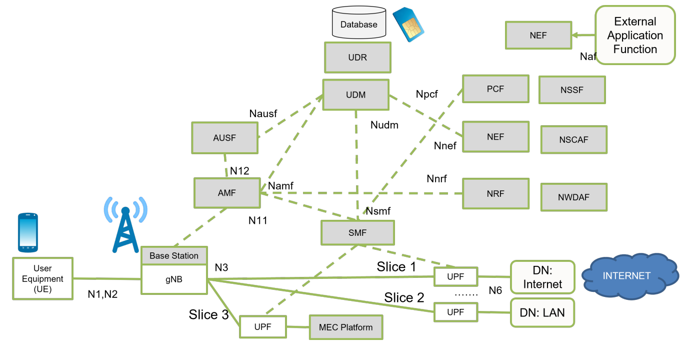


Fig. 1: Service Based Architecture with 5G network functions

System (5GS) [7]. In our study the original MEC platform is redefined to keep only the necessary components for a simple use case of performance measurements (bandwidth & latency). Therefore, a shortened list of MEC components are described in our implementation. Section III elaborates on the MEC platform and its integration with the 5GS.

3GPP's Edge Enabler architecture is complementing the ETSI MEC reference-point architecture [6]. However, we are implementing a tiny MEC architecture with a service-based approach. The aim is to minimize operational costs and complexities associated with orchestration of multiple servers. Particularly, the SBA grants the authority to enable other entities to access their services through vendor-agnostic API calls as recommended by the standards and shown in Fig. 2. We are using Transport Layer Security (TLS) as a security mechanism for authentication and authorization of applications and API requests. The redefined components are as follows:

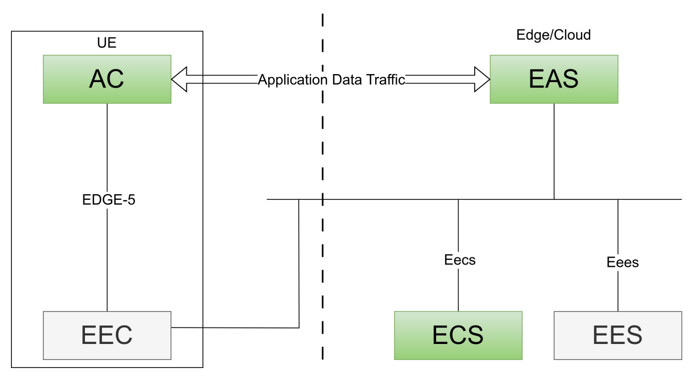


Fig. 2: 3GPP Service-based Edge Enabler

- **Application Client. (AC)** is an application in the UEs. For the simple use case of performance measurements, it is represented by an iperf client capable of sending requests to the server at the edge or in the cloud.
- **Edge Application Server. (EAS)** performs the server functions, in this context iperf server located in the Data Network (DN) with the User Plane Function (UPF). It is entitled to provide services to the AC. Moreover, it invokes the Core Network (CN) capabilities (e.g., UE location, UE identity, parameter provisioning) through the ECS instead of the original EES.

- **Edge Configuration Server. (ECS)** is implemented as an external Application Function (AF). It is redefined to support functionalities of registration of the EAS to enable its discovery and connection. It interacts with 5GC via the Network Exposure Function (NEF) to access CN capabilities and enable discovery, registration and connection to the EAS. The entity is the management level of the MEC platform to get requirements of the application server (e.g. computation, availability, slice configurations).
- **Edge Application Server Discovery Function. (EASDF)** acts as an internal DNS resolver to avoid additional latency of relying on a public Domain Name Server (DNS) e.g., Google. The EASDF handles DNS queries from the UE as a DNS proxy with a database of local domain information.

In our forthcoming research, we aim to incorporate supplementary components (for orchestration and management) to broaden the scope of our use cases and enhance the depth of our experiments. These studies will include AI services, and implementations tailored for smart factory applications. We intend to diversify the user's connection scenario to include edge relocation for mobility.

Distinctive vendors are enhancing the implementation and deployment of edge native platforms e.g., the Eclipse Foundation, GSMA Operator Platform Group (OPG), Linux Foundation, and other industry initiatives [2], [8], [12]. Within their working groups, they aim to build open-source and production-ready solutions for edge applications. The platforms gravitate around ETSI MEC framework design to include system level management for edge orchestrator and host level management. With such an approach, vendors aim for better interoperability through multi-operator schemes of edge integration while considering security aspects [5].

III. MEC Platform

This section describes the test setup for measuring the performance of MEC. The results include distributed ORAN technologies. The RAN is connected to the 5GC (Cumucore 5GC) that supports 5G LAN and TSN features to be used with network slicing and MEC [13]. The 5GC includes the Network Data Analytics Function (NWDAF) [14] from our custom implementation, used by the MEC platform to find the optimal edge location to run the edge host. The MEC platform was redesigned to have a lightweight component in regards with our RAN infrastructure. Hence, any orchestration service as proposed in ETSI's framework is not included [6], [13].

A. MEC setup

Our setup consists of the topology in Fig. 3. We attach a custom MEC platform to one of our UPFs which represents the data plane towards a DN. The MEC is a multi-components feature developed as an API with the aim to provide services to end users without relying on the public network. For our case study the service is an iperf application for network performance measurements. We are using two iperf servers, one residing in the MEC collocated with the gNB and another available in the public internet.

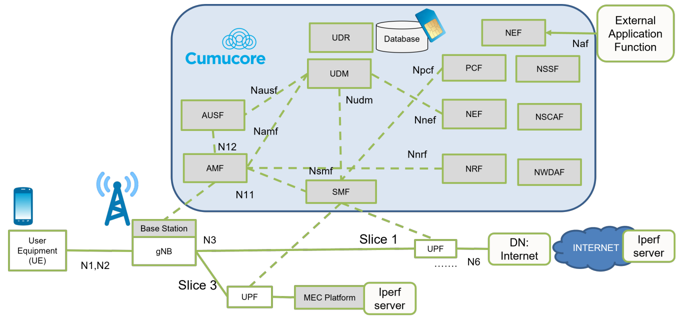


Fig. 3: Test setup for MEC access performance measurements

B. MEC workflow

The interactions between 5GC NFs and the MEC functions are shown in Fig. 4. The external application function mentioned in section II-C is represented by the ECS. Its task is to deploy the EAS's information to the core network. An interface between the two components (EAS & ECS) allows the exchange of information related to the application server's supported features. Prior to any deployment of the edge server information, the ECS requests some capabilities from the NEF. The NEF extracts analytics about slices' capacities from NWDAF.

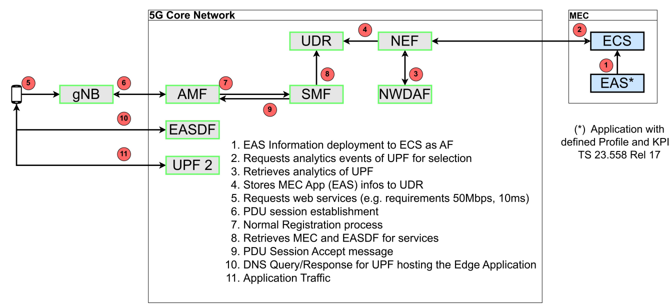


Fig. 4: UE interactions with MEC

The analytics are used to select the UPF to which the MEC platform is attached. After selection, the slice parameters are included in the set of data sent to the NEF. When the UE requests a service from the 5GC the Packet Data Unit (PDU) session establishment message includes MEC information extracted by the Session Management Function (SMF) from the Unified Data Repository (UDR) as shown in Fig. 4. The provisioning information given to the UE includes the DNS server IP address which belongs to the EASDF. The proxy is set up as a security layer to prevent attacks in the private network. When a DNS query is received by the EASDF the domain name field is extracted, we then proceed to a lookup from the database for its IP address. In case of a known service, a DNS response is sent to the UE with the IP address of the UPF to which the MEC platform is attached. If no IP address is found in the database, the DNS request is forwarded to a public DNS which answers back to the EASDF before replying to the UE. Further improvements are foreseen to consider the implementation of an AI solution for the UPF selection. The main idea is to collect network data such as slice congestion, communication delay per UE, mobility information of UEs, to find the optimal UPF based on those parameters.

IV. Measurements

In [15], 3GPP has defined e2e Key Performance Indicators (KPIs) in 5G networks. Interestingly, the calculation of latency and delay are separated at each level of the NG-RAN i.e. gNB-DU, gNB-CU-UP, gNB-CU-CP, etc. The average delay unit is 0.1ms, hence the aggregated value for an e2e latency measurement expected is 10ms as refers in most standards. We perform intensive latency and bandwidth measurements in our setup to evaluate the reliability of such metrics in regards with real-life deployment scenario. In following sections we describe RAN equipment used to perform the measurements with detailed explanations on the data collection and processing.

A. Radio Equipment settings

The ORAN radio equipment used in this study comprises a sophisticated two-part module. The hardware component is sourced from the ASKEY 5G Sub-6 Indoor Small Cell SCE2200 series for its performance in Sub-6 GHz frequency bands. This hardware is capable of supporting up to 100MHz bandwidth, suitable for high-speed data transmission in 5G networks. Complementing the hardware is a software layer facilitated by a Node-H management server API, which enables seamless configuration of the gNB through a REST API interface as shown in Fig. 5. This software architecture ensures flexibility and ease of management, allowing for customization of network parameters specially for the frame structure. Furthermore, the software is accessible and managed locally via WebGUI, hosted directly on the hardware. Table I presents a detailed description of the equipment's specifications.

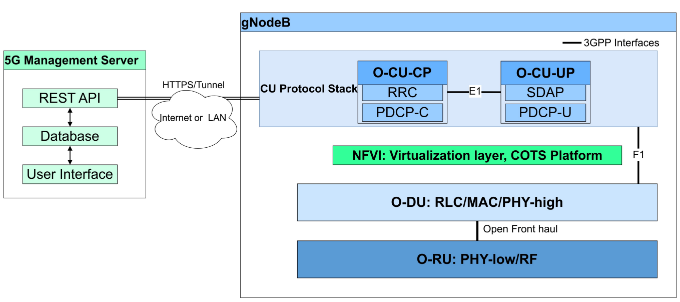


Fig. 5: ORAN Architecture

TABLE I: Radio Specifications

Specifications	ORAN
Band	n78/n79
Band Frequency	SKU1:N78(3.3-3.8GHz) SKU2:N79(4.4-5.0GHz)
Carrier Bandwidth	n78:20/30/40/50/60/70/80/90/100MHz; n79:100MHz
RF Output Power	24dBm per port
Modulation scheme	64QAM
Antenna	2x2 MIMO

B. Data Collection

The data collection process involved the use of two Nokia XR20 smartphones, each serving distinct roles in capturing relevant metrics. One phone was dedicated to conduct iperf measurements, assessing aspects such as bandwidth and throughput, while the other focused solely on monitoring latency. For the latter, we relied on the Round-Trip Time (RTT) of Internet Control Message Protocol (ICMP) packets, sending continuous ping requests to the target iperf to capture fluctuations in latency. Given iperf's limitation of allowing only one UE connection in TCP mode, we simulated parallel connections from the client side to mimic real-world scenarios, ranging from 10 to 50 threads. This simulated load was applied both in the Uplink (UL) and Downlink (DL) directions to comprehensively evaluate performance. The measurements were conducted with various TDD frame structures denoted as xDyU[1], as outlined in Table II. We aim by this to explore the impact of different configurations on network performance. To ensure robust data collection, we observed an interval of time between 30s and 1mn before initiating subsequent measurements, hoping for adequate stabilization of network conditions. This meticulous approach facilitated the generation of comprehensive and reliable datasets for analysis and interpretation.

TABLE II: Simulation Parameters

Parameters	Value
Iperf duration (per trial)	10s
Parallel connections (Thread)	10, 20, 30, 40, 50
Number of Reports in DL/UL	10 per thread
Number of UEs	2
Phone model	Nokia XR20 5G
Frame Structures	7D2U, 2D2U, 1D2U
Reporting interval	between 30s to 1mn
Slice capacity	1000Mbps UL/DL
5QI	9 (Non-GBR)
PDB	300ms

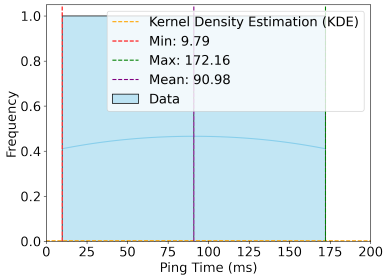
V. Results and Discussion

The data analysis focused on generating Gaussian distributions of latency measurements for each frame structure case, aiming to characterize the performance across various network configurations. Given the potential for unpredictable bottlenecks in the network, removal of outliers was done to ensure the reliability of the analysis. The resulting distributions enabled us to extract key statistical parameters, including the minimum, maximum, and mean latency values. These metrics were compared with the estimated latency in 3GPP specifications for MEC, providing insights into the network's adherence to industry standards. The UEs were allocated a 5G Quality of Service Identifier (5QI)[2] of 9, representing a non-Guaranteed Bit Rate (Non-GBR) according to 3GPP TS 23.501.

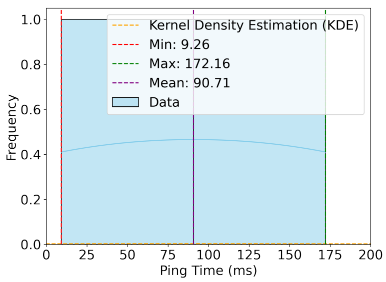
The measured latency values were juxtaposed with the Packet Delay Budget (PDB) of 300 ms corresponding to the QoS level. Interestingly, the analysis revealed mean latency values of 90.30 ms, 90.71 ms, and 90.98 ms for frame structures 1D2U, 2D2U, and 7D2U (considered most suitable for DL transmission in terms of bandwidth), respectively as shown in Fig. 6. These values represented at most 31% of the PDB, a marginally lower latency, indicating favorable latency

[1] xDyU: Represents the allocation of x time slots in downstream, y in upstream and 1 free time slot for synchronization.

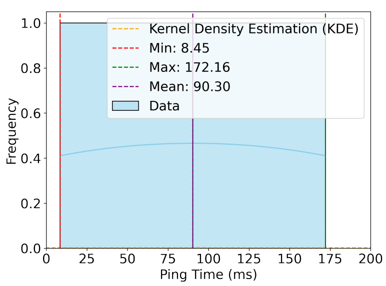
[2] In 5G, packets are classified into different QoS classes, each QoS with their characteristics (e.g. priority level, packet delay, etc.)



(a) 7D2U

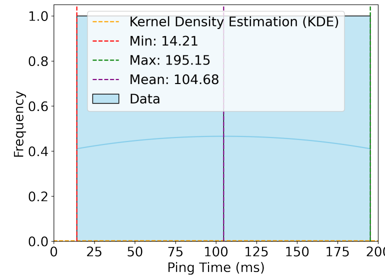


(b) 2D2U

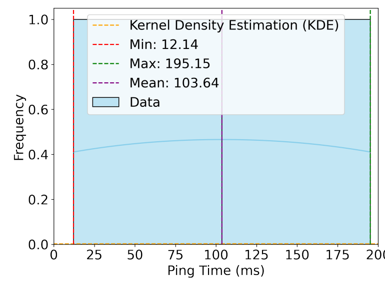


(c) 1D2U

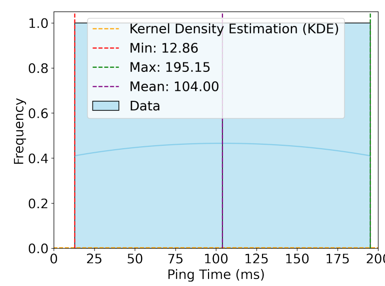
Fig. 6: Latency results for MEC



(a) 7D2U



(b) 2D2U

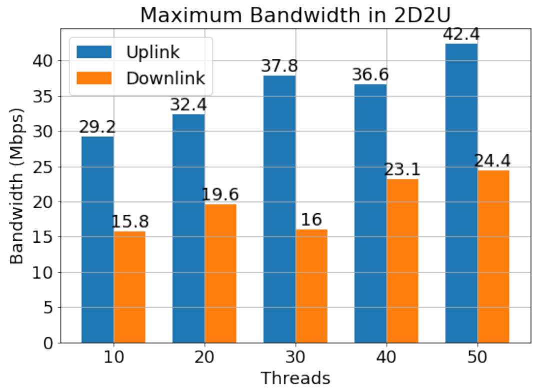


(c) 1D2U

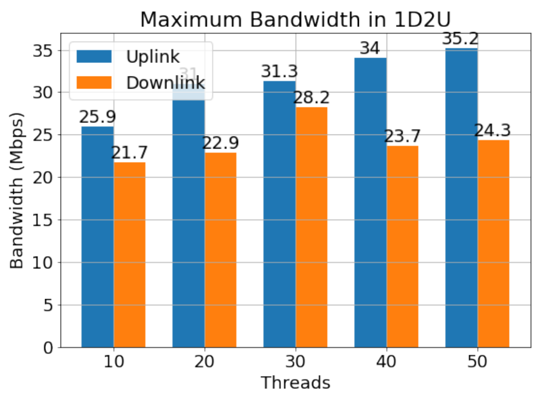
Fig. 7: Latency results for public Iperf

performance within the allocated QoS constraints. However, it was observed that communication management at the RAN level introduced additional overhead in terms of latency (due to https communication between the hardware and software), underscoring the importance of efficient network management strategies in minimizing latency impacts. As anticipated, the latency observed in the public iperf exceeds that of the MEC, aligning with established expectations. However, the disparity between the two is not that substantial. Fig. 7 demonstrates a tiny difference, typically ranging between 10 to 15 ms for the mean and merely 3 to 5 ms for the minimum latency when compared to the MEC across all considered frame structures. Such insights underscore the significance of MEC in providing competitive latency performance compared to traditional public networks, but still there is room of improvements.

In discussing the findings, it is essential to highlight certain limitations that affect the generalization and interpretation of the results. The study's reliance on a small number of equipment instances (both radio and UE) may restrict the extent to which findings can be applied to broader network contexts. Different vendors may implement technologies and optimizations differently, leading to variations in performance even with similar configurations. Future research efforts should aim to address these vendor-specific characteristics challenges and explore strategies for mitigating their impact on network performance assessment and optimization. The use of Non-GBR communication introduces complexities in optimizing network performance, as these communication modes prioritize best-effort delivery, traffic load and dynamic resource sharing over GBR. Consequently, achieving optimal performance requires finding a balance between slice capacity, QoS, frame structures, and end-user requirements (e.g. beamforming, MIMO). As shown in Fig. 8a, the throughput results are very low contrary to what was expected with asymmetric frame structure (2D2U), with the slice capacity set to 1000Mbps UL/DL. Similar low results are observed in Fig. 8b, yet the higher performance in UL are compliant with the frame structure. It is important to recognize that optimizing RAN and



(a) 2D2U



(b) 1D2U

Fig. 8: Bandwidth evolution with 2D2U and 1D2U

MEC performance involves a cautious consideration of various factors, rather than relying on a one-size-fits-all approach. The simplified MEC architecture, featuring a punctual server, was designed specifically to focus on bandwidth and latency measurements. However, the introduction of orchestration and management functions could significantly impact future results and dynamics in real-world deployments.

VI. Conclusion

This paper has provided a detailed account of our custom MEC implementation based on 3GPP Release 17. Our MEC architecture is seamlessly integrated with an ORAN base station, which served for our measurement campaign focused on latency and bandwidth assessment. Through experimentation of different frame structures while using iperf servers deployed both locally within the MEC platform and remotely in the cloud, we have gained valuable insights into the performance characteristics of our setup. Our findings regarding latency underscore the current gap between observed latency levels and the sub-ms latency targets advocated by industry standards. However, this study serves as a bridge for future research endeavors, providing clear guidance on areas of optimization and improvement. Key focal points include QoS (in software radio) enhancements, optimization of internal communication at the RAN management level, considerations for MIMO technology in the user side, among others.We emphasize the necessity of synchronized efforts between ORAN and Core Network levels to achieve substantial improvements in network performance. Our exploration into C-RAN solutions reveals promising insights, with initial results indicating enhanced stability and efficiency, particularly in achieving a balance between latency and bandwidth requirements. These findings pave the way for our future research initiatives aimed at harnessing the full potential of MEC to deliver superior network services and user experiences in industrial environments.

Acknowledgment

This work was supported by the European Union's Horizon Europe project DETERMINISTIC6G under grant agreement No. 101096504. This work was supported by 6GBridge 6G Test Network Finland project funded by Business Finland under grant agreement No. 10769/31/2022.

References

[1] 3GPP, "5G System Enhancements for Edge Computing," Technical Specification (TS) 23.548, Version 17.6.0.

[2] Gómez, D. la Fuente, García, Rosillo, and Puche, "A vision of industry 4.0 from an artificial intelligence point of view," 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:37885473

[3] Y. Seol, D. Hyeon, J. Min, M. Kim, and J. Paek, "Timely Survey of Time-Sensitive Networking: Past and Future Directions," *IEEE Access*, vol. 9, pp. 142 506–142 527, 2021.

[4] T. W. Nowak, M. Sepczuk, Z. Kotulski, W. Niewolski, R. Artych, K. Bocianiak, T. Osko, and J.-P. Wary, "Verticals in 5G MEC-Use Cases and Security Challenges," *IEEE Access*, vol. 9, pp. 87 251–87 298, 2021.

[5] B. Ali, M. A. Gregory, and S. Li, "Multi-Access Edge Computing Architecture, Data Security and Privacy: A Review," *IEEE Access*, vol. 9, pp. 18 706–18 721, 2021.

[6] ETSI, "MEC Deployments in 4G and Evolution Towards 5G," February 2018.

[7] 3GPP, "Architecture for Enabling Edge Applications," Technical Specification (TS) 23.558, Version 17.6.0.

[8] G. Nencioni, R. G. Garroppo, and R. F. Olimid, "5G Multi-Access Edge Computing: A Survey on Security, Dependability, and Performance," *IEEE Access*, vol. 11, pp. 63 496–63 533, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:259299979

[9] L. Gavrilovska, V. Raković, and D. Denkovski, "From cloud RAN to open RAN," *Wireless Personal Communications*, vol. 113, no. 3, pp. 1523–1539, 3 2020. [Online]. Available: https://doi.org/10.1007/s11277-020-07231-3

[10] 3GPP, "System architecture for the 5G System (5GS)," Technical Specification (TS) 23.501, Version 17.7.0.

[11] ——, "Procedures for the 5G System (5GS)," Technical Specification (TS) 23.502, Version 17.7.0.

[12] ETSI, "MEC support towards Edge Native Design," June 2023.

[13] ——, "Cloud RAN and MEC: A Perfect Pairing," February 2018.

[14] 3GPP, "Architecture Enhancements for 5G System (5GS) to Support Network Data Analytics Services," Technical Specification (TS) 23.288, Version 17.7.0.

[15] ——, "Management and Orchestration; 5G end to end Key Performance Indicators (KPI)," Technical Specification (TS) 28.554, Version 17.8.0.