Fairness Guaranteed and Auction-Based x-Haul and Cloud Resource Allocation in Multi-Tenant O-RANs

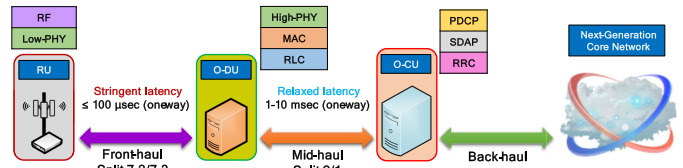Sourav Mondal, *Member, IEEE*, and Marco Ruffini, *Senior Member, IEEE*

Fig. 1.　A schematic diagram showing O-RAN architecture with functions of RU, O-DU, and O-CU and their corresponding interfaces.

Abstract— The open-radio access network (O-RAN) embraces cloudification and network function virtualization for base-band function processing by dis-aggregated radio units (RUs), distributed units (DUs), and centralized units (CUs). These enable the cloud-RAN vision in full, where multiple mobile network operators (MNOs) can install their proprietary or open RUs, but lease on-demand computational resources for DU-CU functions from commonly available open-clouds via open x-haul interfaces. In this paper, we propose and compare the performances of *min-max fairness* and *Vickrey-Clarke-Groves (VCG) auction*-based x-haul and DU-CU resource allocation mechanisms to create a multi-tenant O-RAN ecosystem that is sustainable for small, medium, and large MNOs. The min-max fair approach *minimizes the maximum OPEX of RUs* through cost-sharing proportional to their demands, whereas the VCG auction-based approach *minimizes the total OPEX for all resources utilized while extracting truthful demands from RUs*. We consider time-wavelength division multiplexed (TWDM) passive optical network (PON)-based x-haul interfaces where PON virtualization technique is used to flexibly provide optical connections among RUs and edge-clouds at macro-cell RU locations as well as open-clouds at the central office locations. Moreover, we design efficient heuristics that yield significantly better economic efficiency and network resource utilization than conventional greedy resource allocation algorithms and reinforcement learning-based algorithms.

Index Terms— Min-max fairness, multi-tenant Open-RAN, reinforcement learning, resource allocation, VCG auction.

I. Introduction

THE fifth-generation (5G) radio access networks (RANs) are standardized to meet a diverse set of QoS requirements to support broadband, low-latency, and machine-type communications. Applications like mixed reality, telesurgery, high-definition video streaming, and Industrial Internet-of-Things, to name a few, will be free from the spectrum crunch and network resource scarcity issues of the legacy RANs. However, the existing mobile networks with their "one size

fits all" architecture lack sufficient flexibility and intelligence for efficient catering of such requirements [1]. Therefore, the necessity for a major architectural revolution is envisaged for beyond 5G and sixth-generation (6G) RANs. Over the past few years, major mobile network operators (MNOs) across the globe are collaborating within the *Open-RAN (O-RAN) Alliance* to standardize an open and smart RAN architecture that can perform complex RAN management with the aid of software-defined networking (SDN), network function virtualization (NFV), and edge computing (EC) technologies [2]. This architecture typically follows 3GPP recommendations where the RUs perform low-PHY functions (typically split 7.2 and 7.3), while high-PHY, MAC, RLC, RRC, and PDCP functions are processed by the DU-CUs that can be hosted on OLT-Clouds with commercial off-the-shelf (COTS) hardware, as shown in Fig. 1. Recently, the IEEE P1914.1 standardization working group was created to specify the next-generation front-haul interface (NGFI). The RU-DU interface is known as the *NGFI-I*, or the *front-haul* (maximum one-way latency bound = 100 μsec), and the DU-CU interface is known as the *NGFI-II* or the *mid-haul* (maximum one-way latency bound = 1 msec) [3]. The interface beyond CU to the 5G core is known as the *back-haul*; hence, the general term *x-haul* is used.

The incorporation of open clouds for DU-CU function processing over the open front/mid-haul interfaces in the O-RAN architecture creates new business opportunities for small, medium, and large MNOs as well as network service providers (NSPs) [4]. In turn, this creates a *multi-tenant O-RAN ecosystem* where several MNOs deploy their RUs with macro and small-cell coverage over a certain geographic area but procure front/mid-haul and DU-CU function processing resources from the open and shared resource pool provided by various NSPs [5]. The primary benefit of this multi-tenant O-RAN architecture is minimization of the CAPEX and OPEX

for the MNOs. The techno-economic analysis in [6] shows that $\sim$40% CAPEX and $\sim$15% OPEX over 5 years can be reduced by adopting SDN-based architectures for mobile network virtualization. In practice, government, municipality, or an alliance of MNOs can be the NSP that owns the open x-haul and cloud resources and distribute the resources among the MNOs. On the other hand, a competitive market model can also be created where the MNOs compete against each other or form opportunistic coalitions for procuring their required x-haul and cloud resources. These observations motivate us to propose efficient resource allocation mechanisms that create a multi-tenant O-RAN ecosystem that is sustainable for small, medium, and large MNOs.

The cloud servers installed at a central office (CO) or optical line terminal (OLT) locations are referred to as OLT-Clouds, but their significant intermediate distance may become disadvantageous for supporting low-latency applications and fronthaul interfaces (typical PON length $\geq$ 10 km). This hurdle can be overcome by installing Edge-Clouds at macro-cell RU locations to host DU-CU and *local core* functions for some of the neighboring small-cell RUs [7]. Moreover, efficiently utilizing geographically distributed Edge-Clouds can lead to a better cost efficiency of a RAN than centralized OLT-Clouds. Nonetheless, the RUs supporting latency-tolerant broadband applications can be connected to OLT-Cloud and 5G core without such issues. Therefore, we consider the TWDM-PON architecture proposed in [8] as the x-haul interfaces to create a logical mesh topology that facilitates the small-cell RUs to be connected with OLT-Clouds at CO locations or Edge-Clouds at macro-cell RU locations in a flexible manner. This architecture *supports East-West communication* along with traditional North-South communication and its efficiency over similar architectures in literature was also proven in [8].

We critically observe that a large body of the existing literature mainly focuses on allocating computational resources only and ignores communication resources of the x-haul interfaces. Moreover, while connecting RUs from different MNOs to either Edge-Cloud or OLT-Cloud over the open front/midhaul interfaces, the OPEX of the RUs are calculated by either of the well-known methods like uniform sharing, utility maximization, min-max fairness, and proportional fairness [9]. Note that this resource allocation problem can be considered as an *assignment problem*, but the RUs can not demand any specific amount of resources as in the conventional setting. Each RU only knows its front/mid-haul datarate and RU-DU-CU processing requirements corresponding to its split option. After all the RUs inform their respective front/midhaul datarate to the NSP, sufficient resources are allocated by the NSP such that the front/mid-haul data generated by the RUs in each *slot duration* (5G slot duration can be 125, 250, 500, or 1000 μsec) are transmitted and processed within the maximum latency bounds. Therefore, in the uniform sharing approach, when the cost of total utilized resources is uniformly distributed among the RUs, inefficiency may arise if RUs with lower resource requirements pay higher prices. In the utility maximization approach, the profit of the NSP is maximized while RUs are connected to Edge/OLT-Clouds.

Hence, the RUs from wealthy MNOs will get priority and the RUs from poor MNOs may suffer from resource starvation at high-load conditions. The proportional fairness is a fair resource allocation method where fairness is achieved through maximization of a logarithmic utility function.

Nevertheless, in this paper, we embrace the *min-max fairness approach with proportional cost sharing* method, where we connect the RUs to Edge/OLT-Clouds such that the maximum OPEX of the RUs is minimized by allocating resources proportional to their demands and satisfy their latency requirements. Also, the RUs from different MNOs are fairly chosen for allocation such that poor MNOs do not suffer heavily during high-load conditions. We design this method for creating a multi-tenant O-RAN ecosystem where all the small, medium, and large MNOs get fair opportunities for OPEX minimization. However, the decisions made by this scheme strongly depend on the *revealed resource demands of the RUs to the NSPs* and the RUs may not be always truthful in revealing their resource demands if there exist opportunities to gain extra incentives from the market. This motivates us to design a *Vickrey-Clarke-Groves (VCG) auction-based mechanism* that allocates resources to RUs while minimizing the cost of total utilized resources but uses a special payment rule that enforces truthful revelation of resource requirements as a weakly dominant strategy equilibrium for the RUs [10]. Our contributions in this paper are:

(a) We propose a multi-tenant O-RAN architecture where RUs from small, medium, and large MNOs can be connected to Edge/OLT-Clouds for their DU-CU functions for lowlatency and broadband applications in a sustainable manner over TWDM-PON-based front/mid-haul interfaces via East-West and North-South links.

(b) We formulate an integer non-linear program (INLP) for the min-max fair resource allocation. In this formulation, we minimize the maximum cost for leasing front/midhaul and DU-CU resources of each RU (resource allocation is proportional to demand) while satisfying the latency requirements of the low-latency and broadband applications.

(c) We formulate a second INLP for the VCG auction-based resource allocation. In this formulation, we minimize the total cost for leasing front/mid-haul and DU-CU resources of all the RUs. Moreover, a payment rule is designed that ensures truthful revelation of resource demands of the RUs to prevent them from taking unfair advantages while paying for consumed resources.

(d) We design polynomial-time algorithms for efficient implementation of the min-max fair and VCG auction formulations. Furthermore, we compare the economic efficiency and network resource utilization achieved by our proposed algorithms against state-of-the-art nearest-first (greedy) and reinforcement learning (RL)-based (multi-arm bandit) algorithms through numerical evaluation to showcase their usefulness in practice.

The rest of this paper is organized as follows. Section II reviews some related works. Section III describes the multitenant O-RAN architecture. Section IV presents the system

model. Section V presents the min-max fairness, VCG auction, and baseline (greedy nearest-first and RL-based) methods. Section VI presents numerical evaluation results. Finally, Section VII provides the concluding remarks.

II. REVIEW OF RELATED WORKS

Resource allocation and management problems are fundamental research challenges in any networking environment and a large volume of literature exists on this area spanning across all types of network scenarios [11]. The O-RAN for beyond 5G/6G mobile communication systems is no exception to this as its flexibility in terms of bandwidth, latency, and QoS requirements introduces several interesting research challenges [12]. Before the O-RAN architecture was proposed, several resource allocation or radio resource head (RRH) to base-band unit (BBU) assignment problems were solved using mathematical optimization and game theoretic tools for the Cloud-RAN (C-RAN) architecture by the authors of [13], [14], [15], [16]. The authors of [17] designed a dynamic two-stage mechanism for downlink resource allocation and BBU-RRH assignment in C-RAN. The authors of [18] investigated a joint RRH-BBU association and energy sharing problem to minimize brown energy usage. Again, the authors of [19] investigated the RRH-BBU mapping problem to minimize the network power consumption by reducing the number of active BBUs. Moreover, the authors of [20] studied the joint RRH clustering and RRU activation problem with QoS constraints to minimize the energy consumption of RRHs. The authors of [21] demonstrated a multi-vendor multi-standard PON for 5G x-haul that performs the control and management operations by SDN/NFV technologies.

After the formation of the O-RAN Alliance, as the standardization of virtualized RAN started, researchers from academia and industry started to propose various interesting solutions to overcome O-RAN deployment and resource management challenges. Recently, the authors of [22] provided a very elaborate overview of the architecture and components of O-RAN, explored artificial intelligence (AI)-based use cases, and discussed various research opportunities across different engineering sectors. The authors of [23] provided detailed discussions on the ongoing O-RAN Alliance standardization activities with various analyses supported by a study of the traffic steering use case in a modular way following the open networking approach. We also shared several insights on optical transmission network (OTN) and optical distributed network (ODN)-based front/mid-haul network design for O-RANs from our observations in [24]. Alongside these, the authors of [25] formulated a two-step mixed-integer programming problem for finding the optimal power allocation, physical resource block (PRB) assignment, the number of virtual network functions (VNFs), and the number of RUs. The authors of [26] modeled the RU-DU assignment problem as a 2D bin packing problem and proposed a deep reinforcement learning-based self-play method to achieve efficient RU-DU resource management. Moreover, the authors of [27] designed a team learning algorithm for implementing a near-real-time (near-RT) radio intelligent controller (RIC) of O-RAN.

However, neither of the aforementioned works focused on the challenges of designing flexible front/mid-haul interfaces between RUs and Edge/OLT-Clouds. Moreover, no comparative analysis is available between the conventional greedy heuristics and learning-based resource allocation algorithms.

Another important aspect of the O-RAN architecture, which essentially evolves from the C-RAN architecture, is its natural ability to facilitate multi-tenancy, i.e., a pool of network resources can be shared among multiple MNOs [28]. The multi-operator RAN (MORAN) allows two or more MNOs to share every component of a RAN except the radio carriers, whereas the multi-operator core network (MOCN) allows two or more core networks to share the same RAN or the carriers [29]. In [30], the authors demonstrated a virtual network controller enabled multi-tenant virtual network on top of multi-technology OTNs. We also performed some initial studies on the resource allocation problem for multi-tenant O-RAN ecosystems in [31]. However, a more detailed investigation of system performance, economic analysis, and robust resource allocation mechanisms implementable in a practical competitive market scenario is required.

III. MULTI-TENANT O-RAN ARCHITECTURE

Fig. 2 shows the considered O-RAN architecture in a multi-tenant scenario where multiple MNOs install neighboring RUs with hexagonal macro-cell and circular small-cell coverage. Each MNO pays a fee for leasing networking (i.e., for x-haul) and computing resources (i.e., DU-CU processing at Edge/OLT-Clouds) according to a certain payment scheme. Furthermore, all the MNOs need to pay a default price to the mediator, acting as the open platform provider to cover the cost of the resources required for RAN management and control plane operations. Recently, ITU-T has drafted recommendations for using TWDM-PONs as an optical front/mid-haul solution as TWDM-PONs can support 100 Gbps or more aggregated datarate (i.e., in upcoming standardization) which can be scaled further by combining additional wavelengths [32]. Moreover, other recent work has addressed PON slicing isolation [33] and compliance with service level agreement (SLAs) [34]. Thus, TWDM-PON-based interfaces are used to connect both the macro and small-cell RUs to a CO with multi-level *reflective splitters*. These splitters are designed so that they can be dynamically reconfigured to pass through or reflect back (i.e., towards the end points) the desired set of wavelengths (the concept is taken from [8]). The wavelengths that are passed through, establish the North-South communication links (downlink: green, uplink: blue), whereas the reflected wavelengths establish the East-West communication links (downlink: purple, uplink: red). In terms of network hierarchy, each level-1 reflective splitter aggregates multiple RUs and each level-2 reflective splitter aggregates multiple level-1 reflective splitters and all their respective RUs for a cost-efficient deployment. A set of level-1 reflective splitters are used to connect RUs and Edge-Clouds directly, while multiple level-1 reflective splitters are connected to a level-2 reflective splitter to reach through other PON branches.

A local connectivity between small-cells and macro-cells via East-West communication links can be achieved by installing
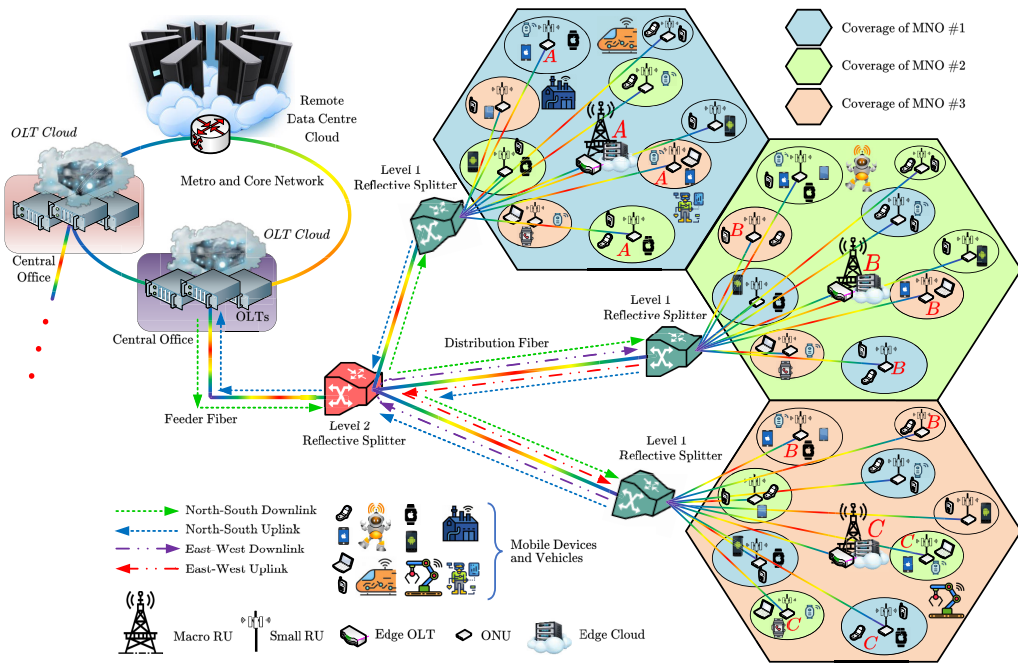
Fig. 2. The proposed TWDM-PON-based multi-tenant O-RAN architecture where RUs from multiple MNOs (hexagonal macro-cell and circular small-cell coverage area are shown in blue, green, and orange for three different MNOs) can be connected to Edge-Cloud or CO OLT-Cloud via the North-South or East-West virtual-PONs (indicated by red A, B, C) for the respective DUs and CUs.

an Edge-OLT at the macro-cell, while the small-cells can host a simple ONU. Control signals via the North-South communication links can be sent to these Edge-OLTs at macro-cells and ONUs at small-cells to create virtual-PON instances that communicate via the East-West communication links. For example, three virtual-PON instances are shown in Fig. 2 and the ONUs and Edge-OLTs belonging to the same virtual-PON are labeled as A, B, and C in red. This direct communication enables ultra-low latency and ultra-low jitter communications as the signals remain in the optical domain while reflected back at the splitter. Note that ONUs in virtual-PON instance A communicate only via level-1 reflective splitter. The same occurs for instance C. However, ONUs in virtual-PON instance B can communicate via both level-1 and level-2 reflective splitters (i.e., they extend across two PON branches). Both OLT and Edge-Clouds can host the DU-CU functions. Although the OLT-Clouds host the main 5G core, the Edge-Clouds can be used to host local 5G cores [7]. The back-haul traffic can be routed to the remote data centers via metro and core networks.

Fig. 3 shows the user plane and control plane interfaces of the proposed TWDM-PON-based multi-tenant O-RAN architecture. The RUs for ultra-reliable and low-latency (uRLLC) services are prioritized to be connected to Edge-Clouds, whereas RUs for enhanced mobile broadband (eMBB) services can be flexibly connected to Edge/OLT-Clouds. Although Fig. 1 shows the most general schematic to highlight the flexible RAN deployment options provided by the O-RAN architecture, we choose to place DU and CU functions at a common Edge/OLT-Cloud because this is the most efficient configuration for uRLLC and eMBB applications in our judgment. The near-RT RIC mainly interacts with DUs and CUs through the $E2$ interface, whose control loops operate with a periodicity between 10 msec and 1 sec. The near-RT RIC consists of multiple applications called *xApps* for per-UE controlled load-balancing, resource block management, interference detection and mitigation, QoS management, connectivity management, and seamless handover control [35]. Alongside this, the near-RT RIC is connected to the non-real time (non-RT) RIC by the $A1$ interface. This non-RT RIC is a component of the service management and orchestration (SMO) framework and consists of *rApps* to complement the near-RT RIC for intelligent RAN operation and optimization on a time scale larger than 1 sec. Therefore, our proposed resource allocation algorithms in this paper can be implemented as control mechanisms that involve the periodic exchange of information and decision between RUs, Near-RT RIC, and Non-RT RIC. We consider that the RUs report their incoming resource demands to SMO every 1 sec over the $O1$ interface. Observing the information over a few seconds interval (operators decide based on the dynamicity of traffic), the rApps execute our proposed decision-making algorithms and pass on the decision to COs over the $O1$ interface. Accordingly, RUs are connected to Edge/OLT-Clouds over North-South or East-West TWDM-PON links. All the intermediate UE connectivity and handover management functions are handled by the xApps in near-RT RIC.

IV. SYSTEM MODEL

In this section, we describe the TWDM-PON-based front/mid-haul communication and RU-DU-CU function processing models considered for our problem formulation. The datarate for the front/mid-haul interface mainly depends on the split option chosen between RU and DU [36]. With Split-7.2, all the radio frequency processing, fast Fourier
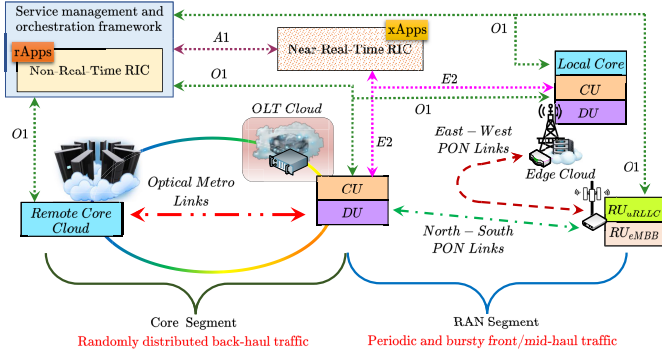
Fig. 3. A schematic diagram showing the user plane and control plane interfaces of the proposed TWDM-PON-based multi-tenant O-RAN architecture that supports uRLLC and eMBB services.

transform (FFT)/inverse FFT, cyclic prefix removal/addition, digital beamforming, and resource element mapping are done at the RU. The datarate can be calculated as follows [36]:

$$W_{7.2} = N_P \times N_{RB} \times N_{RB}^{SC} \times N_{sym}^{SF} \times T_{SF}^{-1} \\ \times \mu \times N_Q \times 2 \times \zeta, \quad (1)$$

where, N_P denotes the number of antenna ports, N_{RB} denotes the number of resource blocks (RB), N_{RB}^{SC} denotes the number of sub-carriers per RB, T_{SF}^{-1} denotes sub-frame duration, μ denotes the maximum RB utilization, N_Q denotes the quantizer bit resolution per I/Q dimension, and ζ denotes the front-haul overhead. With Split-7.3, precoding, layer mapping, and modulation are also done with the aforementioned tasks and the datarate is calculated as follows:

$$W_{7.3} = N_L \times N_{RB} \times N_{SC}^{RB} \times N_{sym}^{SF} \times T_{SF}^{-1} \\ \times \mu \times (1 - \eta) \times N_Q \times \log_2(M_{mod}) \times \zeta, \quad (2)$$

where, N_L denotes the number of spatial layers, η denotes resource overhead, and M_{mod} denotes the modulation order. Note that 3GPP recommends Split-7.3 mainly to be used for downlink transmission, but Split-7.2 can be used for both uplink and downlink [37]. As front/mid-haul data are transmitted as periodic bursts of Ethernet frames, the number of frames in a burst can be calculated as $\mathcal{B} = \lceil R_D \times \delta_t / \mathcal{P} \rceil$, where R_D denotes the front/mid-haul datarate, δ_t denotes the burst interval duration, and $\mathcal{P}$ denotes the payload size of an Ethernet frame (1500 Bytes). Hence, the actual throughput of a flow can be calculated by $(\mathcal{B} \times F/\delta_t)$, where F is the maximum Ethernet frame size (1542 Bytes). This data is transmitted over TWDM-PON and *cooperative dynamic bandwidth allocation (Co-DBA)* protocol [38] is used for coordinating RAN and PON capacity scheduling in the uplink transmission. Furthermore, it is crucial to note that sufficient communication resources should be available for each RU to transmit each burst of front/mid-haul data to DU-CU without failure to ensure a successful end-to-end communication.

The total RU-DU-CU function processing effort per slot in Giga operations per second (GOPS) is given by [39]:

$$\mathcal{C}_{RDC} = \left(3N_a + N_a^2 + \frac{1}{3} \times \mathcal{M} \times \Psi \times N_L \right) \times \frac{N_{RB}}{5}, \quad (3)$$

where, N_a denotes the number of MIMO antennas, $\mathcal{M}$ denotes the number of modulation bits, and Ψ denotes the coding rate. This total computational effort $\mathcal{C}_{RDC}$ is distributed among RU, DU, and CU based on the chosen intermediate split options. For example, 40% processing is done by RU with Split-7.2, but 50% processing is done by RU with Split-7.3. The remainder of the processing is done by the DU-CU and the total RU-DU-CU processing time can be computed by the polynomial expressions provided in [40].

V. ASSIGNMENT OF RUs TO EDGE/OLT-CLOUDS HOSTING DU/CU FUNCTIONS

In this section, we formulate a min-max fairness-based and a VCG auction-based problem for connecting the RUs to some Edge/OLT-Cloud over front/mid-haul interfaces that host both the corresponding DU and CU functions. We also design some efficient heuristics for each of these problem formulations that can be implemented in practice. We consider that both macro-cell and small-cell RUs can support either or both uRLLC and eMBB applications. Thus, for mathematical convenience, we define two different sets for denoting RUs supporting uRLLC and eMBB applications, but some members of both these sets can be co-located and we need to optimally connect them to Edge-Clouds via East-West communication links or OLT-Clouds via North-South communication links. The min-max fairness-based resource allocation creates a multi-tenant O-RAN ecosystem where all the small, medium, and large MNOs get fair opportunities for OPEX minimization. In this scheme, each RU pays the price for allocated resources in proportion to their demand. However, to prevent affluent MNOs from influencing the fairness of resource allocation by revealing a higher resource demand, we formulate the VCG auction-based resource allocation problem with a different allocation and payment rule that makes each RU pay a price that is independent of their respective resource demand but dependent on the resource demands of other RUs. Thus, the RUs cannot gain any incentive by revealing any false resource demand. Although this formulation ensures truthful resource demand revelation from all the RUs, it cannot guarantee a fair OPEX for all MNOs because it minimizes the OPEX of the overall network and does not consider the OPEX of RUs individually. Therefore, NSPs can choose the min-max fairness resource allocation mechanism where truthful demand revelation is possible through strict market regulations (e.g., huge economic penalty or market ban on detection of false information). For an open and competitive market scenario, the NSPs can choose the VCG auction-based mechanism. Furthermore, we describe a nearest-first (greedy) and RL-based resource allocation mechanism for performance comparison.

A. Min-Max Fairness Guaranteed Resource Allocation

Our primary objective here is to allocate front/mid-haul and DU-CU resources for RUs such that the OPEX of RUs with worst/high values are minimized. We denote the set of uRLLC RUs by $\mathcal{R}_u = \{1, 2, \ldots, R_u\}$ and the set of eMBB RUs by $\mathcal{R}_m = \mathcal{R} \setminus \mathcal{R}_u$, where $\mathcal{R} = \{1, 2, \ldots, R_u,$

$R_u + 1, \ldots, R_u + R_m\}$. Note that at each RU location, one RU for uRLLC services and one RU for eMBB services can coexist, whose data are scheduled to be transmitted at different PRBs within each slot. Also, we denote the set of Edge-Clouds by $\mathcal{E} = \{1, 2, \ldots, E\}$, and the set of OLT-Clouds by $\mathcal{Q} = \mathcal{Y} \setminus \mathcal{E}$, where $\mathcal{Y} = \{1, 2, \ldots, E, E+1, \ldots, E+Q\}$. The *binary variable* x_{ry} denotes if an RU $r \in \mathcal{R}$ is connected to an Edge/OLT-Cloud $y \in \mathcal{Y}$, i.e.,

$$x_{ry} = \begin{cases} 1; & \text{if RU } r \text{ and Edge/OLT-Cloud } y \text{ are connected} \\ 0; & \text{otherwise.} \end{cases}$$

The parameter z_{ry} indicates if RU $r \in \mathcal{R}$ and Edge/OLT-Cloud $y \in \mathcal{Y}$ can be connected over a virtual-PON (East-West or North-South) when its value is 1. The parameters C_r, C_λ, and C_P denote the default cost to the mediator (€), the cost for throughput used (€/Gbps), and the cost for cloud resources leased (€/GOPS) by each RU r, respectively. As the Edge-Clouds are located at some macro-cell RU location, they can be owned by the respective MNO, and the attached RUs from the same MNO do not need to pay the costs of computational resources. The neutral NSP can also own the Edge-Clouds, but it needs to provide some price discount to the respective MNOs. To incorporate these facts, we incorporate a discount factor $\gamma_{ry} \in [0, 1]$ where $\gamma_{ry} = 0$ indicates full discount and $\gamma_{ry} = 1$ indicates no discount. The parameters W_r^{UL} and W_r^{DL} denote the uplink and downlink front/mid-haul datarate of RU r. The parameters B_y^{UL}, B_y^{DL}, $\forall y \in \mathcal{E}$ denote the maximum uplink, and downlink throughput of the East-West TWDM-PON links and B_y^{UL} and B_y^{DL}, $\forall y \in \mathcal{Q}$ denote the maximum uplink and downlink throughput of the North-South TWDM-PON links. The maximum throughput of each PON link can vary according to the number of configured wavelengths. The parameters η_r^{UL} and η_r^{DL} denote the required uplink and downlink GOPS/slot, H_r^{UL} and H_r^{DL} denote the available uplink and downlink GOPS/slot for RU processing. The parameters Γ_r^{UL} and Γ_r^{DL} denote the required GOPS/slot for DU-CU processing of RU r and G_y^{UL}, G_y^{DL} denote maximum available GOPS/slot at Edge/OLT-Clouds y. The parameter θ_{ry} denotes the burst interval over which data are transmitted from ONUs connected to RU r in East-West or North-South TWDM-PONs. Finally, the parameters Δ_r^H and Δ_r^{RDC} denote the maximum one-way front/mid-haul latencies and total RU-DU-CU processing for RU r. Now, we formulate the min-max fair resource allocation problem for a multi-tenant O-RAN ecosystem as follows:

$$\mathcal{P}_1: \quad \min_{x_{ry}} \max_r \left\{ \sum_{y \in \mathcal{Y}} (C_r + C_\lambda B_{ry} + \gamma_{ry} C_P G_{ry}) \, x_{ry} \right\} \quad (4)$$

$$\text{subject to } x_{ry} \le z_{ry}, \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y}, \quad (5)$$

$$\sum_{y \in \mathcal{Y}} x_{ry} \le 1, \quad \forall r \in \mathcal{R}, \quad (6)$$

$$B_{ry} = \left(\frac{W_r^{UL} B_y^{UL}}{\varepsilon + \sum_r x_{ry} W_r^{UL}} \right) + \left(\frac{W_r^{DL} B_y^{DL}}{\varepsilon + \sum_r x_{ry} W_r^{DL}} \right), \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y}, \quad (7)$$

$$G_{ry} = \left(\frac{\Gamma_r^{UL} G_y^{UL}}{\varepsilon + \sum_r x_{ry} \Gamma_r^{UL}} \right) + \left(\frac{\Gamma_r^{DL} G_y^{DL}}{\varepsilon + \sum_r x_{ry} \Gamma_r^{DL}} \right), \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y}, \quad (8)$$

$$x_{ry} \left\{ \delta_{ry} + \frac{D_{ry}}{v_l} \right\} + \left\lceil \frac{\theta_{slot}}{\theta_{ry}} \right\rceil \left\{ \frac{\sum_r x_{ry} W_r^{UL} \theta_{ry}}{B_y^{UL}} \right\} \le \Delta_r^H, \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y}, \quad (9)$$

$$x_{ry} \left\{ \frac{D_{ry}}{v_l} \right\} + \left\lceil \frac{\theta_{slot}}{\theta_{ry}} \right\rceil \left\{ \frac{\sum_r x_{ry} W_r^{DL} \theta_{ry}}{B_y^{DL}} \right\} \le \Delta_r^H, \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y}, \quad (10)$$

$$\frac{\eta_r^{UL}}{H_r^{UL}} + \left\{ \frac{\sum_r x_{ry} \Gamma_r^{UL}}{G_y^{UL}} \right\} \le \frac{\Delta_r^{RDC}}{\theta_{slot}}, \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y}, \quad (11)$$

$$\frac{\eta_r^{DL}}{H_r^{DL}} + \left\{ \frac{\sum_r x_{ry} \Gamma_r^{DL}}{G_y^{DL}} \right\} \le \frac{\Delta_r^{RDC}}{\theta_{slot}}, \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y}, \quad (12)$$

$$x_{ry} \in \{0, 1\}, \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y}. \quad (13)$$

The objective function of the problem $\mathcal{P}_1$ is given by (4), which indicates the minimization of maximum OPEX of each RU r. The first term is the default cost, the second term is the front/mid-haul throughput leasing cost, and the third term is the DU-CU function processing resources leasing cost. Note that *the price for throughput and computational resources paid by each RU is proportional to their demands*. The constraint (5) ensures that RU r can be associated with Edge/OLT-Cloud y only when an East-West or North-South connection exists and the constraint (6) restricts RU r to be connected to one Edge-Cloud or OLT-Cloud y at most. The constraints (7) indicates the allocated share of throughput to RU r over front/mid-haul interface to Edge/OLT-Cloud y. Similarly, the constraint (8) indicates the allocated share of GOPS to RU r for DU-CU processing at Edge/OLT-Cloud y. Note that a very small constant $\varepsilon \approx 0$ is added to the denominator of each of the terms in (7)-(8) to avoid division by zero. Furthermore, the constraint (9) ensures that the uplink front/mid-haul latency from RU r to Edge/OLT-Cloud y is within Δ_r^H. The parameter δ_{ry} denotes the *average queuing latency* of uplink data due (considering the use of the Co-DBA mechanism). The second term with x_{ry} indicates *propagation latency* where the parameter D_{ry} denotes the distance from RU r to Edge/OLT-Cloud y and v_l denotes speed of light within fiber (2×10^5 km/s). The third term indicates *data transmission latency* where data is transmitted in multiple bursts of duration θ_{ry} within each TTI, θ_{slot}. Similarly, constraint (10) ensures that the downlink front/mid-haul latency from RU r to Edge/OLT-Cloud y is within Δ_r^H. Finally, the constraints (11)-(12) ensure the uplink and downlink RU-DU-CU processing latencies are within Δ_r^{RDC}, respectively. The first term indicates the RU processing latency and the second and third terms indicate the DU-CU processing latencies at Edge/OLT-Cloud y, respectively.

TABLE I
NETWORK PARAMETERS AND SETS

Symbol	Definition
$\mathcal{R}_u$	Set of RUs for uRLLC services
$\mathcal{R}_m$	Set of RUs for eMBB services
$\mathcal{R}$	Set of all RUs present in the system ($\mathcal{R} = \mathcal{R}_u \cup \mathcal{R}_m$)
$\mathcal{E}$	Set of Edge-Cloud locations
$\mathcal{Q}$	Set of OLT-Cloud locations
$\mathcal{Y}$	Set of all Edge/OLT-Cloud locations ($\mathcal{Y} = \mathcal{E} \cup \mathcal{Q}$)
D_{ry}	Distance between RU $r \in \mathcal{R}$ and Edge/OLT-Cloud $y \in \mathcal{Y}$
C_r	Default cost of participation to the mediator (€)
C_λ	Cost for throughput used (€/Gbps)
C_P	Cost for cloud resources leased (€/GOPS)
$W_r^{UL/DL}$	The uplink and downlink front/mid-haul datarate of RU r
$\Gamma_r^{UL/DL}$	Required GOPS/slot for DU-CU processing of RU r
$B_y^{UL/DL}$	Maximum throughput of the TWDM-PON links
$G_y^{UL/DL}$	Maximum available GOPS/slot at Edge/OLT-Clouds y
$\eta_r^{UL/DL}$	Required uplink and downlink GOPS/slot for RU processing
$H_r^{UL/DL}$	Available uplink and downlink GOPS/slot for RU processing
Δ_r^H	Maximum one-way front/mid-haul latency of RU r
Δ_r^{RDC}	Maximum RU-DU-CU processing latency of RU r
θ_{slot}	Transmit time slot of RU r
θ_{ry}	Transmission burst interval of TWDM-PON uplinks
δ_{ry}	The reduced waiting time for uplink data at ONUs
v_l	Speed of light in optical fiber (2×10^8 m/s)

B. Heuristic for the Min-Max Fair Resource Allocation

We observe that $\mathcal{P}_1$ is an NP-hard problem and the primary reason behind the NP-hardness is that the locations of active Edge/OLT-Clouds are not known when we start to connect RUs to Edge/OLT-Clouds. Note that the problem $\mathcal{P}_1$ has a unique structure that converts a multi-objective problem into a single-objective problem such that standard optimization methods can be employed. In this case, we convert the minimization problem of OPEXs of multiple RUs into a minimization problem of the maximum OPEX of RUs. However, the problem $\mathcal{P}_1$ is still very inconvenient to solve due to the presence of $\max\{.,.\}$ function in the objective. Therefore, we need to reformulate this problem into an equivalent *epigraph form* as follows:

$$\mathcal{P}_1^r : \min_{x_{ry},M} \quad M \tag{14}$$

$$\text{subject to } M \geq \sum_{y \in \mathcal{Y}} (C_r + C_\lambda B_{ry} + \gamma_{ry} C_P G_{ry})$$

$$x_{ry}, \quad \forall r \in \mathcal{R},$$

$$\text{constraints } (5) - (13). \tag{15}$$

It is straightforward to show that an optimal solution for $\mathcal{P}_1^r$ is also a solution for $\mathcal{P}_1$ [41]. Nonetheless, as both these problems are INLP, the evaluation of an optimal solution cannot be guaranteed in polynomial time. Hence, a heuristic algorithm is required. In general, we understand that OPEX of each RU in (4) can be minimized if each front/mid-haul link and Edge/OLT-Cloud resources are leased by a maximum number of RUs while satisfying constraints (9)-(12). In addition, we observe the following interesting property of optimal solutions of $\mathcal{P}_1^r$.

Proposition 1: An optimal solution of $\mathcal{P}_1^r$ can guarantee fairness if and only if the OPEX of an RU with lower resource

---

Algorithm 1 Algorithm for min-max fair resource allocation

Input: $\mathcal{R}, \mathcal{E}, \mathcal{Y}, D_{ry}, B_y^{U/DL}, W_r^{U/DL}, \Gamma_r^{U/DL}, G_r^{U/DL}$
Output: Near-optimal solution: x_{ry}^* and $\mathcal{C}_r^*$
Initialize: Sort the elements of $\mathcal{R}$ in the increasing order of $(\max\{W_r^{DL}, W_r^{UL}\})$ and/or $(\max\{\Gamma_r^{DL}, \Gamma_r^{UL}\})$ while maintaining an uniform distribution of the percentage of ownership of MNOs;

```
1: for r ← 1 to |R| do
2:     if r = 1 then                          ▷ choose best possible r = 1
3:         Set Ȳ ← Y;
4:         Set assign ← 0;
5:         while assign ≠ 1 and Ȳ ≠ ∅ do
6:             Find y' = arg min_y{D_ry}, y' ∈ Ȳ;
7:             if constraints (5), (9)-(12) are satisfied then
8:                 Set x_ry' ← 1;
9:                 Set assign ← 1;
10:                Calculate B_ry, G_ry, and C_r;
11:            else
12:                Set Ȳ ← Ȳ \ {y'};
13:            end if
14:        end while
15:        if assign ≠ 1 and Ȳ = ∅ then
16:            break;                          ▷ infeasibility condition
17:        end if
18:    else if 1 < r ≤ |R| then
19:        Find all y ∈ Y such that constraints (5), (9)-(12)
20:        are satisfied for the current RU r and create Ȳ;
21:        if |Ȳ| ≥ 1 then                     ▷ if at least one such y exists
22:            Calculate all dummy OPEX values for RU r,
23:            C_ry, if r was connected to each of y ∈ Ȳ;
24:            Find y' = arg min_y{C_ry}, y' ∈ Ȳ;
25:            Set x_ry' ← 1;
26:            Update B_ry, G_ry, and C_r, ∀r with Σ_y x_ry = 1;
27:        else
28:            Set B_ry = 0, G_ry = 0, and C_r = 0;
29:        end if
30:    end if
31: end for
32: return x_ry and C_r;
```

requirements does not exceed the OPEX of an RU with higher resource requirements.

Please refer to Appendix A for the proof. In general, we can achieve the best possible value of M if full-mesh connectivity is available among RU and Edge/OLT-Clouds. However, in practice, mostly partial-mesh connectivity can be observed, i.e., constraint (6) along with constraints (9)-(12) will have a strong influence on the solution. Nonetheless, in general, we are able to connect a higher number of RUs to Edge/OLT-Clouds if we start with lower resource requirements. Based on the above insights, we design a heuristic algorithm, summarized as Algorithm 1. At first, we sort the RUs in $\mathcal{R}$ in the increasing order of $(\max\{W_r^{DL}, W_r^{UL}\})$ and $(\max\{\Gamma_r^{DL}, \Gamma_r^{UL}\})$. This step is crucial to maintain consistency with Proposition 1. We also order the RUs such that the percentage of ownership of MNOs is uniformly maintained. Then we start to iteratively connect each RU r to an Edge/OLT-Cloud y. For $r = 1$, we initialize the flag $assign \leftarrow 0$, the dummy set $\bar{\mathcal{Y}} \leftarrow \mathcal{Y}$, and find the nearest $y' = \arg\min_y\{D_{ry}\}, y' \in \bar{\mathcal{Y}}$. If the constraints (5), (9)-(12) are satisfied for this y', we set $x_{ry'} \leftarrow 1$, $assign \leftarrow 1$,

and calculate the corresponding OPEX value $\mathcal{C}_r = \sum_y (C_r + C_\lambda B_{ry} + \gamma_{ry} C_P G_{ry}) x_{ry}$. If this is not successful, then we remove y' from $\bar{\mathcal{Y}}$ and continue this process until $\bar{\mathcal{Y}} = \emptyset$. In the subsequent iterations, i.e., for $1 < r \leq |\mathcal{R}|$, we find all $y \in \mathcal{Y}$ that satisfy constraints (5), (9)-(12) for the current RU r and reinitialize the dummy set $\bar{\mathcal{Y}}$. If at least one such y exists, i.e., $|\bar{\mathcal{Y}}| \geq 1$, then we calculate the dummy OPEX values $\mathcal{C}_{ry} = (C_r + C_\lambda B_{ry} + \gamma_{ry} C_P G_{ry})$ if RU r was connected to each of the Edge/OLT-Cloud y. Then we find $y' = \arg\min_y \{\mathcal{C}_{ry}\}, y' \in \bar{\mathcal{Y}}$, set $x_{ry'} \leftarrow 1$, and calculate the updated values of B_{ry}, G_{ry}, and $\mathcal{C}_r$ for all RUs. The first for loop iterates for $|\mathcal{R}|$ times to return x_{ry} and $\mathcal{C}_r$ while finding a suitable Edge/OLT-Cloud y from a set of maximum size $|\mathcal{Y}|$ at every iteration. Therefore, the worst-case time-complexity of this loop, as well as Algorithm 1, is $\mathcal{O}(|\mathcal{R}| \times |\mathcal{Y}|)$. Now, if we denote the optimal number of Edge/OLT-Cloud as Y^*, then the number of remaining RUs yet to be connected at every iteration t of Algorithm 1 is at most $R_t \leq R_{t-1}(1 - (1/Y^*))$. Thus, at the end of all $|\mathcal{R}|$ iterations, we must have $1 \leq |\mathcal{R}| \times (1 - (1/Y^*))^{Y_{|\mathcal{R}|}} = |\mathcal{R}| \times (1 - (1/Y^*))^{Y^* \times \frac{Y_{|\mathcal{R}|}}{Y^*}}$. By using the Taylor series approximation $(1 - (1/x))^x \approx (1/e)$, we get $1 \leq |\mathcal{R}| \times ((1/e))^{\frac{Y_{|\mathcal{R}|}}{Y^*}}$, or $Y_{|\mathcal{R}|} \leq Y^* \log_e(|\mathcal{R}|)$, where $Y_{|\mathcal{R}|}$ denotes the number of active Edge/OLT-Clouds after all $|\mathcal{R}|$ iterations. Hence, the solution produced by Algorithm 1 approximates the optimal solution by a factor of $\mathcal{O}(\log_e(|\mathcal{R}|))$.

C. VCG Auction-Based Resource Allocation

Although the solution obtained from the min-max fairness method described in previous sub-sections is very efficient, the solution is dependent on the *private information* like W_r^{UL}, W_r^{DL}, Γ_r^{UL}, and Γ_r^{DL} shared by the RUs to the NSP. To prevent the RUs from sharing false information and gaining unfair incentives from the market, we design a VCG auction-based resource allocation method in this sub-section. Note that this problem is very close to an auction of multiple divisible items [42] but with specific unique characteristics. Each RU wants to connect to an Edge/OLT-Cloud via some East-West or North-South TWDM-PON link to avail sufficient throughput and Edge/OLT-Cloud resources to successfully transmit its front/mid-haul data generated within each slot duration, satisfying the maximum latency bounds. Therefore, the private valuation function of each RU r is given as:

$$V_r(x_{ry}) = \begin{cases} (C_\lambda B_y + C_P G_y) x_{ry}; & \text{if } r \text{ is connected to} \\ & \text{some } y \text{ with successful} \\ & \text{front/mid-haul data} \\ & \text{transmission,} \\ 0; & \text{otherwise,} \end{cases}$$

where, $B_y = (B_y^{UL} + B_y^{DL})$ and $G_y = (G_y^{UL} + G_y^{DL})$. Note that if an RU gets connected to some Edge/OLT-Cloud but fails to transmit its front/mid-haul data, then also its valuation of resources is zero. We assume that the NSP wants a fair market competition and hence, keeps the cost parameters C_r, C_λ, and C_P the same for all the competing MNOs. At first, each RU r submits their front/mid-haul datarate and DU-CU processing requirements to the NSP as a message

$b_r = (\hat{W}_r^{UL}, \hat{W}_r^{DL}, \hat{\Gamma}_r^{UL}, \hat{\Gamma}_r^{DL})$. After receiving messages $\boldsymbol{b} = (b_1, b_2, \ldots, b_{|\mathcal{R}|})$ from all RUs, the NSP solves the following cost-minimization problem while connecting a maximum number of RUs to Edge/OLT-Clouds and the solution $\hat{x}_{ry}^*$ can be considered as the *allocation rule* [43].

$$\mathcal{P}_2 : \min_{x_{ry}, t_y} \left\{ \sum_{y \in \mathcal{Y}} (C_\lambda B_y + C_P G_y) t_y \right\} \qquad (16)$$

$$\text{subject to } t_y \leq x_{ry}, \quad \forall r \in \mathcal{R}, \ y \in \mathcal{Y},$$

$$\text{constraints } (5) - (6), \quad (9) - (13), \qquad (17)$$

where t_y is a binary decision variable and the constraint (17) implies that t_y is equal to 1 if at least one RU r is connected to the Edge/OLT-Cloud y.

Proposition 2: The allocation rule $\hat{x}_{ry}^*$ derived as a solution of $\mathcal{P}_2$ is allocatively efficient.

Please refer to Appendix B for the proof. Although several other dominant-strategy truthful mechanisms exist in the *quasi-linear setting*,[1] only with Groves mechanisms, we can implement an allocative efficiency in dominant strategies among agents with arbitrary quasi-linear utilities. Observe that the valuation for resources of each RU r is non-zero only if it is connected to some Edge/OLT-Cloud y and gets sufficient resources to transmit its front/mid-haul data generated in each slot within the maximum latency bounds. Otherwise, if an RU r is unallocated or fails to transmit its front/mid-haul data generated in each slot within the maximum latency bounds, its valuation is zero. Note that the same resources B_y and G_y are shared by multiple RUs that are allocated to Edge/OLT-Cloud y. Therefore, the net worth of resources allocated to each RU r, $\hat{\mathcal{C}}_r$ varies within $[0, (C_\lambda B_y + C_P G_y)]$. This implies that if only one RU is connected to an Edge/OLT-Cloud y, then it must bear the cost of the consumed resources $\hat{\mathcal{C}}_r = (C_\lambda B_y + C_P G_y)$ all alone. However, if multiple RUs are connected to an Edge/OLT-Cloud y, then the total cost is uniformly divided among them, i.e., $\hat{\mathcal{C}}_r = [(C_\lambda B_y + C_P G_y) / \sum_r \hat{x}_{ry}^*]$. Using this observation, we design a *payment rule* for the RUs similar to Clarke's mechanism [43] as follows:

$$P_r(x_{ry}, \boldsymbol{b}) = \left[\sum_{j \neq r} \tilde{\mathcal{C}}_j(x_{jy}^{-r*}(\boldsymbol{b}^{-r}), b_j) \right] - \left[\sum_{j \neq r} \tilde{\mathcal{C}}_j(x_{jy}^*(\boldsymbol{b}), b_j) \right]$$

$$= \sum_{j \neq r} \left(\frac{C_\lambda B_y + C_P G_y}{\sum_{k \neq r} x_{ky}^*} \right)$$

$$- \sum_{j \neq r} \left(\frac{C_\lambda B_y + C_P G_y}{\sum_k x_{ky}^*} \right), \qquad (18)$$

which can be interpreted as the total cost of all RUs other than r under an efficient allocation when RU r is absent in the system minus the total cost of all RUs other than r under an efficient allocation when RU r is present in the system. Note that $\hat{x}_{jy}^{-r*}$ or $\hat{x}_{ky}^{-r*}$ denote allocation rules when RU r is absent in the system.

---

[1] The utility of agent r with private valuation v_r from obtaining a fraction x of a divisible good at a price p is $u_r(x, p) = v_r(x) - p$.

Algorithm 2 Algorithm for VCG Auction-Based Allocation

Input: $\mathcal{R}, \mathcal{E}, \mathcal{Y}, D_{ry}, B_y^{U/DL}, W_r^{U/DL}, \Gamma_r^{U/DL}, G_r^{U/DL}$
Output: Near-optimal solution: x_{ry}^* and $\mathcal{C}_r^*$
Initialize: Sort the elements of $\mathcal{R}$ in the increasing order of $(\max\{W_r^{DL}, W_r^{UL}\})$ and $(\max\{\Gamma_r^{DL}, \Gamma_r^{UL}\})$ while maintaining an uniform distribution of the percentage of ownership of MNOs;

1: **for** $r \leftarrow 1$ **to** $|\mathcal{R}|$ **do**
2:    **if** $r = 1$ **then**   ▷ *choose best possible $r = 1$*
3:      Set $\bar{\mathcal{Y}} \leftarrow \mathcal{Y}$;
4:      Set $assign \leftarrow 0$;
5:      **while** $assign \neq 1$ **and** $\bar{\mathcal{Y}} \neq \emptyset$ **do**
6:        Find $y' = \arg\min_y\{D_{ry}\}, y' \in \bar{\mathcal{Y}}$;
7:        **if** constraints (5), (9)-(12) are satisfied **then**
8:          Set $x_{ry'} \leftarrow 1$ and $t_{y'} \leftarrow 1$;
9:          Set $assign \leftarrow 1$;
10:          Calculate $\mathcal{C}_{tot}$ and $\tilde{\mathcal{C}}_r$;
11:        **else**
12:          Set $\bar{\mathcal{Y}} \leftarrow \bar{\mathcal{Y}} \setminus \{y'\}$;
13:        **end if**
14:      **end while**
15:      **if** $assign \neq 1$ **and** $\bar{\mathcal{Y}} = \emptyset$ **then**
16:        break;   ▷ *infeasibility condition*
17:      **end if**
18:    **else if** $1 < r \leq |\mathcal{R}|$ **then**
19:      Find all $y \in \mathcal{Y}$ such that constraints (5), (9)-(12)
20:      are satisfied for the current RU r and create $\bar{\mathcal{Y}}$;
21:      **if** $|\bar{\mathcal{Y}}| \geq 1$ **then**   ▷ *if at least one such y exists*
22:        Calculate all dummy cost values for RU r,
23:        $\mathcal{C}_{tot,y}$, if r was connected to each of $y \in \bar{\mathcal{Y}}$;
24:        Find $y' = \arg\min_y\{\mathcal{C}_{tot,y}\}, y' \in \bar{\mathcal{Y}}$;
25:        Set $x_{ry'} \leftarrow 1$ and $t_{y'} \leftarrow 1$;
26:        Update $\mathcal{C}_{tot}$ and $\tilde{\mathcal{C}}_r, \forall r$ with $\sum_y x_{ry} = 1$;
27:      **else**
28:        Set $\tilde{\mathcal{C}}_r = 0$ and keep $\mathcal{C}_{tot}$ unchanged;
29:      **end if**
30:    **end if**
31: **end for**
32: **for** $r \leftarrow 1$ **to** $|\mathcal{R}|$ **do**
33:    Set $\mathcal{R} \leftarrow \mathcal{R} \setminus \{r\}$ and execute lines 1-31;
34:    Calculate $P_r(x_{ry}, \boldsymbol{b})$ and $\mathcal{C}_r = C_r + P_r(x_{ry}, \boldsymbol{b})$;
35:    Set $\mathcal{R} \leftarrow \mathcal{R} \cup \{r\}$;   ▷ *add r back to $\mathcal{R}$*
36: **end for**
37: **return** x_{ry} and $\mathcal{C}_r$;

Theorem 1: The payment rule (18) ensures that truthful private information sharing to NSPs is a weakly dominant strategy for all RUs.

Please refer to Appendix C for the proof. This also shows that the mechanism is *weakly budget balanced* as $\sum_r P_r(\hat{x}_{ry}^*, \boldsymbol{b}) \geq 0$. With the aforementioned allocation and payment rules, the utility of each RU r is defined as follows:

$$\mathcal{U}_r(\hat{x}_{ry}^*, \boldsymbol{b}) = V_r(\hat{x}_{ry}^*, \boldsymbol{b}) - P_r(\hat{x}_{ry}^*, \boldsymbol{b}). \qquad (19)$$

Therefore, the *individual rationality* of the RUs is always maintained in this mechanism as $\mathcal{U}_r(\hat{x}_{ry}^*, \boldsymbol{b}) \geq 0$. In addition to $P_r(\hat{x}_{ry}^*, \boldsymbol{b})$, each RU r also needs to pay the default cost C_r to the NSP if it is connected to some Edge/OLT-Cloud y.

D. Heuristic for the VCG Auction-Based Resource Allocation

As *computational efficiency* is an important property for a mechanism design, we design a polynomial-time heuristic to solve the problem $\mathcal{P}_2$ as outlined in Algorithm 2 (page 30). We start to iteratively connect each RU r to an Edge/OLT-Cloud y. For $r = 1$, we initialize the flag $assign \leftarrow 0$, the dummy set $\bar{\mathcal{Y}} \leftarrow \mathcal{Y}$, and find the nearest $y' = \arg\min_y\{D_{ry}\}, y' \in \bar{\mathcal{Y}}$. If the constraints (5), (9)-(12) are satisfied for this y', we set $\hat{x}_{ry'} \leftarrow 1$, $t_{y'} \leftarrow 1$, $assign \leftarrow 1$, and calculate the cost of total utilized resources $\mathcal{C}_{tot} = \sum_{y \in \mathcal{Y}} (C_\lambda B_y + C_P G_y) t_y$, which is also equal to $\hat{\mathcal{C}}_r$, the *shared cost of RU r*. If the assignment is not successful, then we remove y' from $\bar{\mathcal{Y}}$ and continue this process until $\bar{\mathcal{Y}} = \emptyset$. In the subsequent iterations, i.e., for $1 < r \leq |\mathcal{R}|$, we find all $y \in \mathcal{Y}$ that satisfy constraints (5), (9)-(12) are satisfied for the current RU r and reinitialize the dummy set $\bar{\mathcal{Y}}$. If at least one such y exists, i.e., $|\bar{\mathcal{Y}}| \geq 1$, then we calculate the *dummy costs of total utilized resources* $\mathcal{C}_{tot,y}$ if RU r was connected to each of the Edge/OLT-Cloud y. Then we find $y' = \arg\min_y\{\mathcal{C}_{tot,y}\}, y' \in \bar{\mathcal{Y}}$, set $\hat{x}_{ry'} \leftarrow 1$, $t_{y'} \leftarrow 1$, and calculate the updated value of $\mathcal{C}_{tot}$. We also calculate the shared cost of RU r as $[(C_\lambda B_y + C_P G_y)/\sum_r \hat{x}_{ry}]$ for $y = y'$.

Once the RU to Edge/OLT-Cloud allocation is complete, we calculate the payment rule. For this purpose, we solve $\mathcal{P}_2$ using steps 1-31 for another $|\mathcal{R}|$ times in the absence of each of the RUs from $\mathcal{R}$. Thus, we can calculate $P_r(\hat{x}_{ry}, \boldsymbol{b})$, the payment made by each RU r using (18) and the OPEX of the RU r can be calculated as $\hat{\mathcal{C}}_r = C_r + P_r(\hat{x}_{ry}, \boldsymbol{b})$. Note that the first for loop iterates for $|\mathcal{R}|$ times to return $\hat{x}_{ry}^*$ and tries to find a suitable Edge/OLT-Cloud y from a set of maximum size $|\mathcal{Y}|$ at every iteration. Therefore, the worst-case time-complexity of this loop is $\mathcal{O}(|\mathcal{R}| \times |\mathcal{Y}|)$. The second for loop iterates for $|\mathcal{R}|$ times to return $\hat{\mathcal{C}}_r$, but it executes the first for loop with complexity $\mathcal{O}(|\mathcal{R}| - 1)$. Therefore, the overall time-complexity of Algorithm 2 is $\mathcal{O}(|\mathcal{R}|^2 \times |\mathcal{Y}|)$. Also, similar to Algorithm 1, it can shown that the solution produced by Algorithm 2 approximates the optimal solution by a factor of $\mathcal{O}(\log_e(|\mathcal{R}|))$.

E. Nearest-First (Greedy) and Reinforcement Learning-Based Resource Allocation Mechanisms

To create a performance baseline, we design conventional nearest-first (greedy) and RL-based resource allocation mechanisms. In the *greedy method*, we select the RUs sequentially and attempt to associate them to their nearest Edge/OLT-Cloud subject to the network connectivity (5)-(6), communication latency (9)-(10), and processing latency (11)-(12) constraints, i.e., $y' = \arg\min\{D_{ry}|y' \in \mathcal{Y}, (5) - (6), (9) - (12)\}$. For each RU r, if their nearest Edge/OLT-Cloud y' is unavailable, then we remove y' from the list of Edge/OLT-Clouds and proceed to find the next nearest Edge/OLT-Cloud. In this way, we iteratively try to assign each RU r to an Edge/OLT-Cloud y and $\tilde{x}_{ry}^*$ provides the final allocation. Once the maximum possible RUs are assigned to all Edge/OLT-Clouds, we can calculate the OPEX of the assigned RUs as $\tilde{\mathcal{C}}_r = \sum_{y \in \mathcal{Y}} \left(C_r + C_\lambda \tilde{B}_{ry} + C_P \tilde{G}_{ry} \right) \tilde{x}_{ry}^*$, where the share of throughput and DU-CU function processing resources

Algorithm 3 Algorithm for Nearest-First Resource Allocation

Input: $\mathcal{R}, \mathcal{E}, \mathcal{Y}, D_{ry}, B_y^{U/DL}, W_r^{U/DL}, \Gamma_r^{U/DL}, G_r^{U/DL}$
Output: Near-optimal solution: x_{ry}^* and $\mathcal{C}_r^*$
Initialize: Sort the elements of $\mathcal{R}$ in the increasing order of $(\max\{W_r^{DL}, W_r^{UL}\})$ and $(\max\{\Gamma_r^{DL}, \Gamma_r^{UL}\})$ while maintaining an uniform distribution of the percentage of ownership of MNOs;

1: **for** $r \leftarrow 1$ **to** $|\mathcal{R}|$ **do**
2:     Set $\tilde{\mathcal{Y}} \leftarrow \mathcal{Y}$;
3:     Set $assign \leftarrow 0$;
4:     **while** $assign \neq 1$ **and** $\tilde{\mathcal{Y}} \neq \emptyset$ **do**
5:        Find $y' = \arg\min_y\{D_{ry}\}, y' \in \tilde{\mathcal{Y}}$;
6:        **if** constraints (5), (9)-(12) are satisfied **then**
7:           Set $x_{ry'} \leftarrow 1$;
8:           Set $assign \leftarrow 1$;
9:        **else**
10:          Set $\tilde{\mathcal{Y}} \leftarrow \tilde{\mathcal{Y}} \setminus \{y'\}$;
11:        **end if**
12:     **end while**
13: **end for**
14: **for** $r \leftarrow 1$ **to** $|\mathcal{R}|$ **do**
15:     **if** $\sum_y x_{ry} = 1$ **then**
16:        Calculate $\tilde{B}_{ry}, \tilde{G}_{ry}$, and $\mathcal{C}_r$;
17:     **else**
18:        Set $\tilde{B}_{ry} = 0, \tilde{G}_{ry} = 0$, and $\mathcal{C}_r = 0$;
19:     **end if**
20: **end for**
21: **return** x_{ry} and $\mathcal{C}_r$;

for each RU r are considered either as proportional sharing:

$$\tilde{B}_{ry} = \left(\frac{W_r^{UL} B_y^{UL}}{\sum_r x_{ry} W_r^{UL}}\right) + \left(\frac{W_r^{DL} B_y^{DL}}{\sum_r x_{ry} W_r^{DL}}\right), \quad (20)$$

$$\tilde{G}_{ry} = \left(\frac{\Gamma_r^{UL} G_y^{UL}}{\sum_r x_{ry} \Gamma_r^{UL}}\right) + \left(\frac{\Gamma_r^{DL} G_y^{DL}}{\sum_r x_{ry} \Gamma_r^{DL}}\right), \quad (21)$$

or as uniform sharing:

$$\tilde{B}_{ry} = \left(\frac{B_y^{UL}}{\sum_r x_{ry}}\right) + \left(\frac{B_y^{DL}}{\sum_r x_{ry}}\right), \quad (22)$$

$$\tilde{G}_{ry} = \left(\frac{G_y^{UL}}{\sum_r x_{ry}}\right) + \left(\frac{G_y^{DL}}{\sum_r x_{ry}}\right). \quad (23)$$

The OPEX of the unassigned RUs are considered equal to zero. All the steps of this algorithm are summarized in Algorithm 3. We observe that the worst-case time-complexity of the first for loop is $\mathcal{O}(|\mathcal{R}| \times |\mathcal{Y}|)$ and the time-complexity of the second for loop is $\mathcal{O}(|\mathcal{R}|)$. This implies that the overall time-complexity of Algorithm 3 is $\mathcal{O}(|\mathcal{R}| \times |\mathcal{Y}|)$.

In the *RL-based method*, each RU r randomly selects an Edge/OLT-Cloud y and attempts to establish a front/mid-haul connection by following a *multi-arm bandit algorithm* [44]. The reward of each RU ($\mathfrak{R}_r$) is calculated as the average of Δ_r^H to front/mid-haul latency ratio and Δ_r^{RDC} to RU-DU-CU processing latency ratio. The RUs perform both exploration and exploitation (ϵ-greedy with $\epsilon = 0.3$, i.e., explore strategies with probability 0.3, else exploit) to maximize its cumulative reward values and eventually find the best Edge/OLT-Cloud. After RU to Edge/OLT-Cloud allocation is done, the OPEX of the RUs are calculated as above.
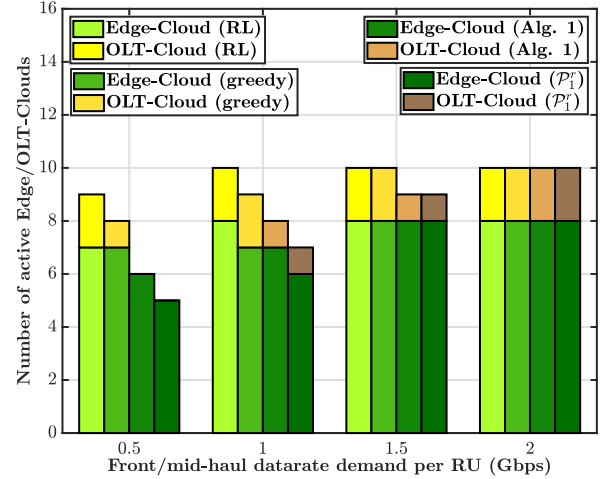
(a) Min-max fairness vs. baseline algorithms
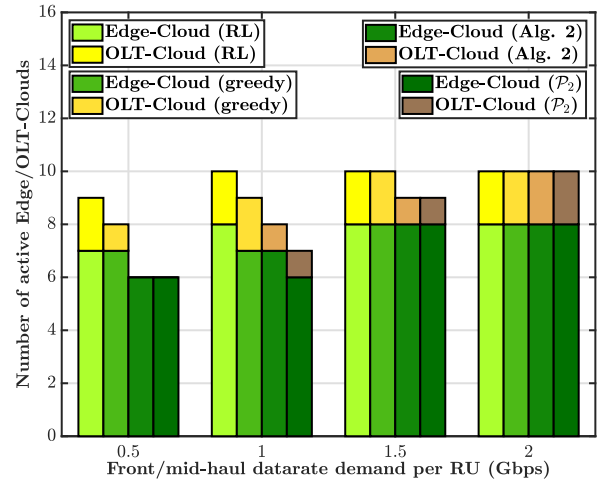
(b) VCG auction vs. baseline algorithms

Fig. 4. Comparison of the total number of active Edge/OLT-Clouds obtained by the optimal solution, heuristics, and baseline methods with more computational resources at Edge-Clouds than at OLT-Clouds.

VI. RESULTS AND DISCUSSIONS

To evaluate our proposed frameworks, we consider a multi-tenant O-RAN deployment area of dimension 5×5 km^2. In this area, 8 macro-cell RUs (coverage = 1 km) and 30 small-cell RUs (coverage = 0.5 km) from three different MNOs coexist (refer to Appendix D). The front/mid-haul datarate and RU-DU-CU processing efforts of the RUs vary over time depending on the RU configurations and the chosen split option. For example, if the RUs are configured with 2×2 MIMO, 2 layers, 50 MHz bandwidth, 15 kHz sub-carrier spacing, MCS index 16, and slot duration = 0.5 msec, the maximum uplink datarate with Split-7.2 is 2.304 Gbps, and the maximum downlink datarate with Split-7.3 is 0.432 Gbps. Each radio head's total RU-DU-CU processing requirement is nearly 550 GOPS/slot. We consider that in each radio head, 25% resources are used for uRLLC services and 75% resources are used for eMBB services. The one-way front-haul link latency bound is $\sim$100 μsec and we choose the RU-DU-CU processing latency bound 325 μsec for uRLLC services and 975 μsec for eMBB services [45]. The reduced
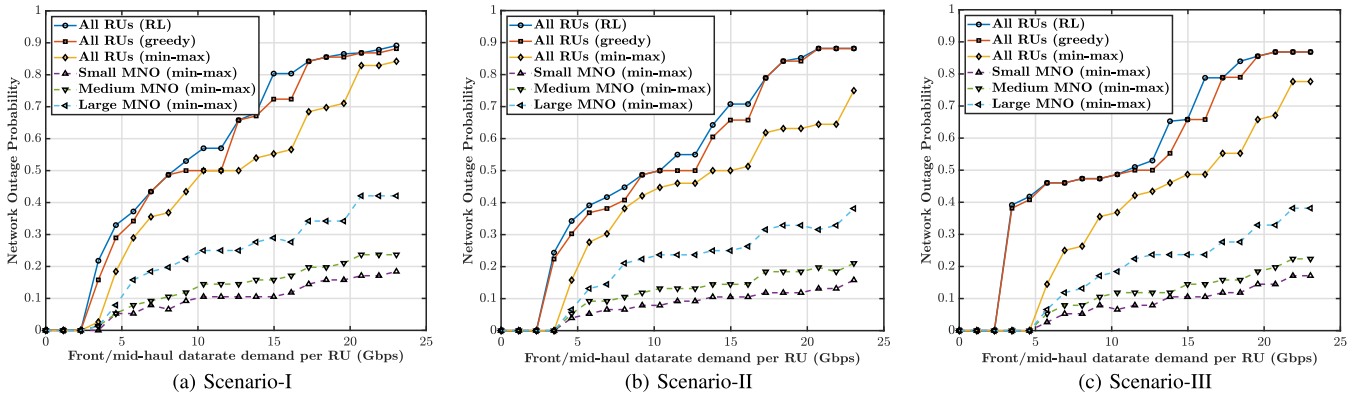
Fig. 5. Comparison of overall O-RAN outage probability and outage probability for small, medium, and large MNOs against increasing front/mid-haul datarate demand per RU with different Edge/OLT-Cloud resource distributions obtained by min-max fairness and baseline (greedy nearest-first and reinforcement learning-based) algorithms.
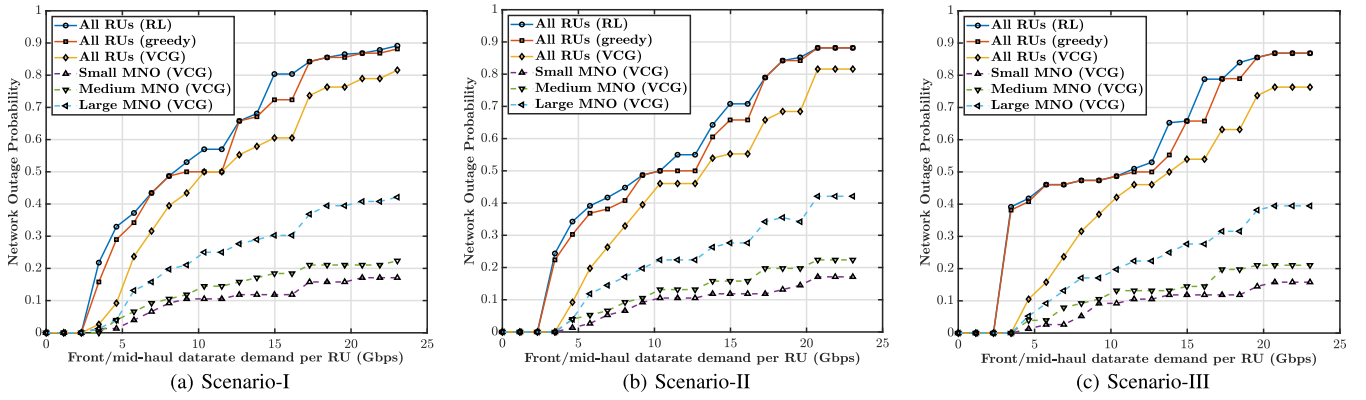

Fig. 6. Comparison of overall O-RAN outage probability and outage probability for small, medium, and large MNOs against increasing front/mid-haul data per RU with different Edge/OLT-Cloud resource distributions obtained by VCG auction-based and baseline (greedy nearest-first and reinforcement learning-based) algorithms.

queueing latency of the uplink data at the ONUs is 15 μsec as Co-DBA is used for uplink transmission [46] and the data are transmitted as periodic bursts of duration 31.25 μsec [7]. Each wavelength of the TWDM-PON can support a throughput of 25 Gbps and we can aggregate multiple such wavelengths to achieve a few hundred Gbps of total throughput. A group of RUs are connected to a level-1 reflective splitter (locations found through k-means clustering) and a few level-1 reflective splitters are connected to a level-2 reflective splitter located at the center of the area. We assume that the RUs can be connected to the OLT-Clouds via the North-South links and to the Edge-Clouds via creating East-West virtual-PON links. We arbitrarily assume that each RU pays a default cost of €100/day to the NSP (other network economic pricing schemes can also be employed but beyond the scope of this paper), although it has very little effect on the results as it is not associated with any decision variables. In addition, the cost for throughput used is €0.5/Gbps [47], and the cost for leasing DU/CU resources is €1.5/GOPS [48].

Firstly, we compare the optimal solutions of the min-max fairness and VCG auction-based resource allocation methods with their corresponding heuristics. In this evaluation, each Edge-Cloud has a maximum capacity of 4.5×10^4 GOPS/slot and each OLT-Cloud has a maximum capacity of 1.5×10^4 GOPS/slot. These values are chosen such that a feasible

TABLE II

NETWORK EVALUATION SCENARIOS

	Scenario - I		Scenario - II		Scenario - III	
	Edge	OLT	Edge	OLT	Edge	OLT
Maximum GOPS/slot	1.5×10^4	4.5×10^4	3×10^4	3×10^4	4.5×10^4	1.5×10^4
Maximum throughput	50 Gbps	600 Gbps	100 Gbps	400 Gbps	150 Gbps	200 Gbps

benchmark solution can be obtained by the greedy baseline algorithm with front/mid-haul datarate demand upto 2 Gbps per RU. We use OCTERACT, a global optimal mixed-integer nonlinear programming solver integrated with the AMPL platform to evaluate the optimal solutions of our formulated INLPs in a computer with Intel Core i7 processor and 32 GB RAM. In Fig. 4(a), we compare the optimal solution of $\mathcal{P}_1^r$ against the solutions obtained using Algorithm 1, the greedy baseline Algorithm 3, and the RL-based method. Again, in Fig. 4(b), we compare the optimal solution of $\mathcal{P}_2$ against the solutions obtained using Algorithm 2, the greedy baseline Algorithm 3, and the RL-based method with the same dataset. From these figures, we observe that the optimal solution of $\mathcal{P}_1^r$ is the highest performing during lower load conditions. The nearest-first method makes several sub-optimal RU to Edge/OLT-Cloud assignments as it connects each RU to its nearest (based on intermediate North-South
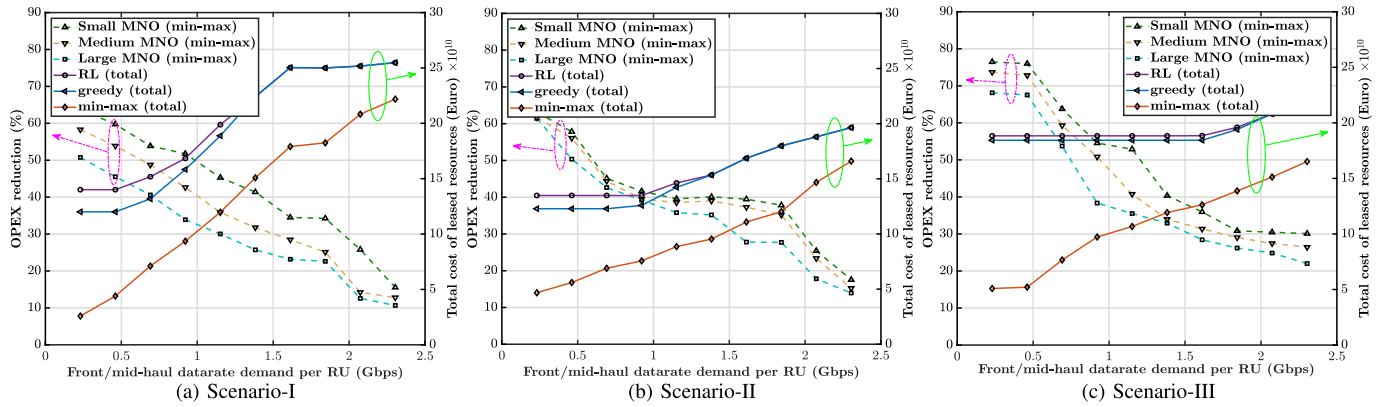
Fig. 7. Comparison of the total cost of leased resources (€) and OPEX reduction percentage for small, medium, and large MNOs against network load variation with different Edge/OLT-Cloud resource distributions obtained by min-max fairness and baseline (greedy nearest-first and reinforcement learning-based) algorithms.
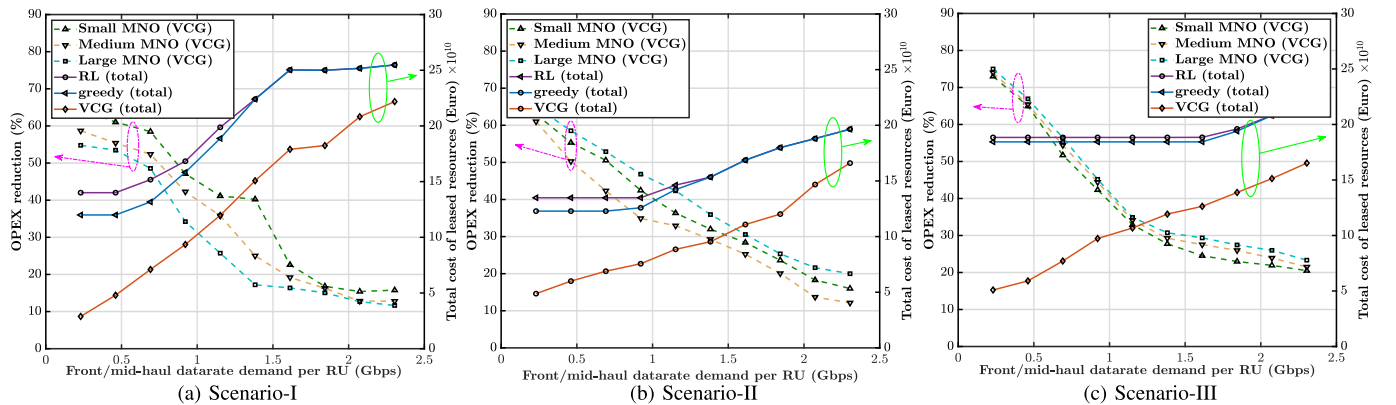

Fig. 8. Comparison of the total cost of leased resources (€) and OPEX reduction percentage for small, medium, and large MNOs against network load variation with different Edge/OLT-Cloud resource distributions obtained by VCG auction-based and baseline (greedy nearest-first and reinforcement learning-based) algorithms.

or East-West link distance) Edge/OLT-Cloud if the latency constraints are satisfied. The RL-based method requires less computation than other methods, but each RU independently attempts to connect to some Edge/OLT-Cloud that gives a higher reward, in turn lower latency values. Thus, the RU to Edge/OLT-Cloud assignment does not incorporate minimization of overall network resource utilization. However, all the solutions at medium and higher load conditions become similar as network load increases, because a large percentage of the potential Edge/OLT-Clouds gets activated and the scope of sub-optimal RU to Edge/OLT-Cloud assignments is reduced.

Next, we compare the *network outage probabilities* (defined as the ratio of the total number of RUs that could not be connected to some Edge/OLT-Cloud to the total number of RUs present in the network) achieved by our proposed methods. We consider that the RUs are owned by three different MNOs where a small MNO-1 has 20%, a medium MNO-2 has 30%, and a large MNO-3 has 50% ownership. Additionally, we consider three scenarios with different computational resource distributions among Edge-Clouds and OLT-Clouds as outlined in Table II. Figs. 5(a)-5(c) show the overall O-RAN outage probabilities obtained through the min-max fairness and baseline (greedy nearest-first and RL-based) algorithms. Similarly, Figs. 6(a)-6(c) show the

overall O-RAN outage probabilities obtained through the VCG auction-based and baseline (greedy nearest-first and RL-based) algorithms. We observe that all the methods show zero outage probability as long as there are sufficient front/mid-haul and Edge/OLT-Cloud resources to accommodate all the RUs. However, network outage probability increases fastest with the RL-based method, followed by the nearest-first baseline method due to their inefficient utilization of network resources, but the outage probability increases relatively slower with the VCG auction-based method and slowest with the min-max fairness method. We also plot the outage probabilities of the small, medium, and large MNOs, which show that the outage probabilities of the small, medium, and large MNOs vary according to their percentage of ownership, i.e., the highest for the large MNO and smallest for the small MNO, in all scenarios.

Finally, we compare the total cost of leased front/mid-haul and Edge/OLT-Cloud resources obtained by the min-max fairness and baseline (greedy nearest-first and RL-based) algorithms. In Figs. 7(a)-7(c), we observe that during the low-load conditions, the performance of the min-max fairness approach is significantly better than the baseline (greedy nearest-first and RL-based) methods. However, as the front/mid-haul datarate demand per RU increases, the performance of all

algorithms becomes similar due to the saturation of available resources. Note that the total cost of leased resources (right y-axis) increases as we place more resources at the Edge-Clouds. Also, the percentage of OPEX savings decreases (left y-axis) as resource utilization increases for all the MNOs. In all scenarios, *the OPEX saving percentage is highest for the small MNO and lowest for the large MNOs.* Similarly, Figs. 9(a)-9(c) compare the total cost of leased resources (right y-axis) obtained and OPEX reduction percentages (left y-axis) by the VCG auction-based and baseline (greedy nearest-first and RL-based) algorithms. We observe that the total cost of leased resources obtained by the VCG auction-based algorithm is very similar to the min-max fairness algorithm, but the OPEX saving percentages for small and medium MNOs are not always lower than the large MNO because the cost of the leased resources is uniformly shared among the RUs. Thus, often the small and medium MNOs pay a much higher price than their actual resource requirements. These results justify that our proposed min-max fair resource allocation framework creates a multi-tenant O-RAN ecosystem that is sustainable for small, medium, and large MNOs.

VII. CONCLUSION

In this paper, we have proposed a multi-tenant O-RAN architecture where RUs from multiple MNOs can lease resources for their DU-CU function processing from Edge/OLT-Clouds over TWDM-PON-based open front/mid-haul interfaces. These TWDM-PONs use reflective splitters to facilitate East-West communication links along with traditional North-South communication links for better mesh connectivity among the RUs and Edge/OLT-Clouds. We have proposed two methods for allocating front/mid-haul and DU-CU processing resources to RUs based on min-max fairness and VCG auction mechanisms. In turn, we have formulated the corresponding INLPs and have designed time-efficient heuristic algorithms. Through numerical evaluation, we have investigated the performance of these frameworks against different distributions of cloud resources at Edge and OLT locations. We have shown that both min-max fairness and VCG auction-based methods achieve a much lower network outage probability than the baseline (greedy nearest-first and RL-based) methods due to the efficient utilization of network resources. Note that the VCG auction-based method can ensure the extraction of truthful information from the RUs, but the min-max fairness method ensures that the OPEX of the RUs are proportional to their actual resource requirements. Moreover, we have shown that the min-max fairness method can reduce the OPEX of all MNOs by more than 20% (at high load) to nearly 75% (at low load). We believe that our proposed frameworks have successfully laid cornerstones for developing further O-RAN resource allocation strategies.

APPENDIX A
PROOF OF PROPOSITION 1

Proof: We observe that if our main problem formulation $\mathcal{P}_1$ has an optimal solution x_{ry}^*, then our reformulated problem $\mathcal{P}_1^r$ also has an optimal solution (x_{ry}^*, M^*). This implies that we have found the best possible M^* as the maximum value of $\sum_{y \in \mathcal{Y}} (C_r + C_\lambda B_{ry} + C_P G_{ry}) x_{ry}$ for some $r \in \mathcal{R}$ but it is the minimum among all feasible values of M. In this sense, *the solution x_{ry}^* guarantees fairness for all $r \in \mathcal{R}$.* However, sometimes the network connectivity, communication latency, and computation latency constraints may play a very strong role to produce an optimal solution (x_{ry}^*, M^*) without maintaining fairness. Now, we analyze all possible network scenarios to show the validity of this proposition.

Case 1: Constraints (6), (9)-(12) are relaxed.

This is the trivial case when there is full-mesh connectivity among the RUs and Edge/OLT-Clouds and no strict latency bound exists. Therefore, any RU can be connected to any Edge/OLT-Cloud and the optimal solution will be dictated by the objective (4) and constraints (5) and (13). Therefore, the optimum value of M will be achieved only when all the RUs are connected to a single Edge/OLT-Cloud with $\min\{B_y^{UL} + B_y^{DL}\}$ and/or $\min\{G_y^{UL} + G_y^{DL}\}$. It is also obvious from (4) that the OPEX of each RU will be directly proportional to their resource demands W_r^{UL}, W_r^{DL}, Γ_r^{UL}, and Γ_r^{DL}, which ensure fairness.

Case 2: Constraint (6) is relaxed only.

In this case, although there is full-mesh connectivity among the RUs and Edge/OLT-Clouds, strict communication and processing latency bounds are applied. Therefore, only a limited number of RUs can be connected to each Edge/OLT-Cloud to satisfy constraints (9)-(12). In this case, the best possible solution can be achieved only if all RUs can be connected to (one or multiple) Edge/OLT-Clouds with minimum total cost of resources. Without loss of generality, consider that there are total $|\mathcal{R}|$ number of RUs and 2 front/mid-haul links and Edge/OLT-Clouds. If the front/mid-haul and Edge/OLT-Cloud 1 have sufficient resources to host all $|\mathcal{R}|$ number of RUs, then there is no issue of fairness. However, if there are insufficient resources in Edge/OLT-Cloud 1 and some RUs need to be assigned to Edge/OLT-Cloud 2, then it is best to choose RUs with higher resource demands. By following this method, the fairness of the shared cost of each RU assigned to either of the Edge/OLT-Cloud can be guaranteed. If we assign RUs in any other randomized way, fairness cannot be guaranteed. For example, if we can assign only the RU with the lowest demand to Edge/OLT-Cloud 2 and the rest of the RUs to Edge/OLT-Cloud 1, then the optimal value M^* remains the same (total cost of resources of front/mid-haul and Edge/OLT-Cloud 2), but the fairness is not maintained. Note that this argument remains valid even if we generalize our scenario with more than 2 Edge/OLT-Clouds.

Case 3: Constraint (9)-(12) are relaxed only.

In this case, usually, there is partial-mesh connectivity among the RUs and Edge/OLT-Clouds but communication and processing latency bounds are relaxed. Therefore, in spite of sufficient resources being present, we cannot connect all the RUs to a single Edge/OLT-Cloud due to network connectivity constraints. This implies that the cost of shared resources of multiple RUs with the exact same resource demand may vary if they are connected to different Edge/OLT-Clouds by being forced by the network connectivity constraints. However, their

shared costs will be proportionally fair in comparison with the other RUs connected to each Edge/OLT-Clouds. For example, let us consider two RUs with the exact same resource demands but allocated to two different Edge/OLT-Clouds and a different number of other RUs connected to these Edge/OLT-Clouds. Therefore, the OPEX values may be different for these RUs but their OPEX values will be fair in comparison to other RUs connected to their corresponding Edge/OLT-Clouds.

Case 4: All constraints (6), (9)-(12) are active.

This is the most general case and characteristics of all the previous cases can be observed. Depending on the dataset, either constraint (6) or constraints (9)-(12) will dictate the optimal solution. Although sufficient resources may be present at some Edge/OLT-Cloud to serve a large number of RUs, constraint (6) might interfere. Therefore, in this case, also the cost of shared resources of multiple RUs with the exact same resource demand may vary if they are connected to different Edge/OLT-Clouds. However, the OPEX values of each RU will be proportionally fair to their resource demands in comparison with other RUs connected to each Edge/OLT-Clouds.

In general, we observe by testing a number of diverse datasets that to obtain the best possible solution M^* while maintaining the fairness of the OPEX values of the RUs, we must start to assign RUs to Edge/OLT-Clouds in the increasing order of their resource requirements. ∎

APPENDIX B
PROOF OF PROPOSITION 2

Proof: We consider *a resource allocation mechanism is allocatively efficient* if it optimizes the sum of the valuations of the agents for each given type profile of the agents [43]. Now, in our problem formulation $\mathcal{P}_2$, the total cost of front/mid-haul and Edge/OLT-Clouds are minimized. Although the binary variable t_y indicates if an Edge/OLT-Cloud y is activated or not, the constraint $t_y \leq x_{ry}, \forall r \in \mathcal{R}, y \in \mathcal{Y}$ makes this formulation equivalent to minimizing the sum of valuation functions $V_r(x_{ry})$ for all RUs. Therefore, our allocation rule $\hat{x}_{ry}^*$ derived as an optimal solution of $\mathcal{P}_2$ can be considered as allocatively efficient. ∎

APPENDIX C
PROOF OF THEOREM 1

Proof: The proposed payment rule can be interpreted as the total value of all RUs other than r under an efficient allocation when RU r is absent in the system minus the total value of all RUs other than r under an efficient allocation when RU r is present in the system. Observe that each RU r is connected to an Edge/OLT-Cloud y based on the shared information $b_r = (\hat{W}_r^{UL}, \hat{W}_r^{DL}, \hat{\Gamma}_r^{UL}, \hat{\Gamma}_r^{DL})$ from all RUs.

Case 1: If the shared information b_r is *higher than its actual requirement*, then there is a risk that it might remain unallocated when the network is in high-load condition. Then the valuation, payment, and utility of the RU r are given by:

$$V_r(\hat{x}_{ry}^*) = 0,$$
$$P_r(\hat{x}_{ry}^*, \boldsymbol{b}) = 0,$$
$$\mathcal{U}_r(\hat{x}_{ry}^*, \boldsymbol{b}) = 0.$$

Case 2: If the shared information b_r is *lower than its actual requirement*, then the RU might get allocated to some Edge/OLT-Cloud, but due to insufficient resources in the RU's share, the front/mid-haul data generated per slot may fail to get delivered within the maximum one-way front/mid-haul latency requirements. In this case, the valuation is zero but the payment is non-zero, yielding a negative utility. Thus, the valuation, payment, and utility of the RU r are:

$$V_r(\hat{x}_{ry}^*) = 0,$$
$$P_r(\hat{x}_{ry}^*, \boldsymbol{b}) = \sum_{j=1, j \neq r}^{|\mathcal{R}|} \left(\frac{C_\lambda B_y + C_P G_y}{\sum_{k \neq r} \hat{x}_{ky}^{-r*}} \right) - \sum_{j=1, j \neq r}^{|\mathcal{R}|} \left(\frac{C_\lambda B_y + C_P G_y}{\sum_k \hat{x}_{ky}^*} \right),$$
$$\mathcal{U}_r(\hat{x}_{ry}^*, \boldsymbol{b}) = 0 - P_r(\hat{x}_{ry}^*, \boldsymbol{b}) < 0.$$

Case 3: There may arise network scenarios where an RU r might get connected to some Edge/OLT-Cloud by sharing false information (both higher and lower) and its front/mid-haul data generated per slot is also successfully delivered within the maximum one-way front/mid-haul latency requirements. However, after the RU to Edge/OLT-Cloud allocation, when the payment amount is calculated for each RU r, only the total number of RUs connected to each Edge/OLT-Cloud y is used and the revealed front/mid-haul datarate does not have any role to play. Thus, *the utility of the RU with true revealed information as well as false revealed information remains the same*. In this case, the valuation, payment, and utility of the RU r are given below:

$$V_r(\hat{x}_{ry}^*) = \sum_y (C_\lambda B_y + C_P G_y)\hat{x}_{ry}^*,$$
$$P_r(\hat{x}_{ry}^*, \boldsymbol{b}) = \sum_{j=1, j \neq r}^{|\mathcal{R}|} \left(\frac{C_\lambda B_y + C_P G_y}{\sum_{k \neq r} \hat{x}_{ky}^{-r*}} \right) - \sum_{j=1, j \neq r}^{|\mathcal{R}|} \left(\frac{C_\lambda B_y + C_P G_y}{\sum_k \hat{x}_{ky}^*} \right),$$
$$\mathcal{U}_r(\hat{x}_{ry}^*, \boldsymbol{b}) = V_r(\hat{x}_{ry}^*) - P_r(\hat{x}_{ry}^*, \boldsymbol{b}) \geq 0.$$

Therefore, any RU r can not gain any advantages in its utility with the payment rule by sharing false information with the NSP. This ensures that sharing truthful information with the NSP is a weakly dominant strategy for all the RUs. ∎

Example 1: [One RU and One Edge/OLT-Cloud]

Let us consider the total x-haul throughput is B_1 and total Edge/OLT-Cloud resources is G_1. The RU 1 submits a message b_1 to the NSP. The results of Case 1 and Case 2 are straight-forward to show. If the scenario is like Case 3, then the RU 1 is connected to Edge/OLT-Cloud 1. Therefore the valuation of the RU is:

$$V_1 = (C_\lambda B_1 + C_P G_1).$$

Now, the cost of resources for the RU 1 is $(C_\lambda B_1 + C_P G_1)$ as there is no other RU to share the cost of resources. However, if RU 1 was not present, then the entire block of resources
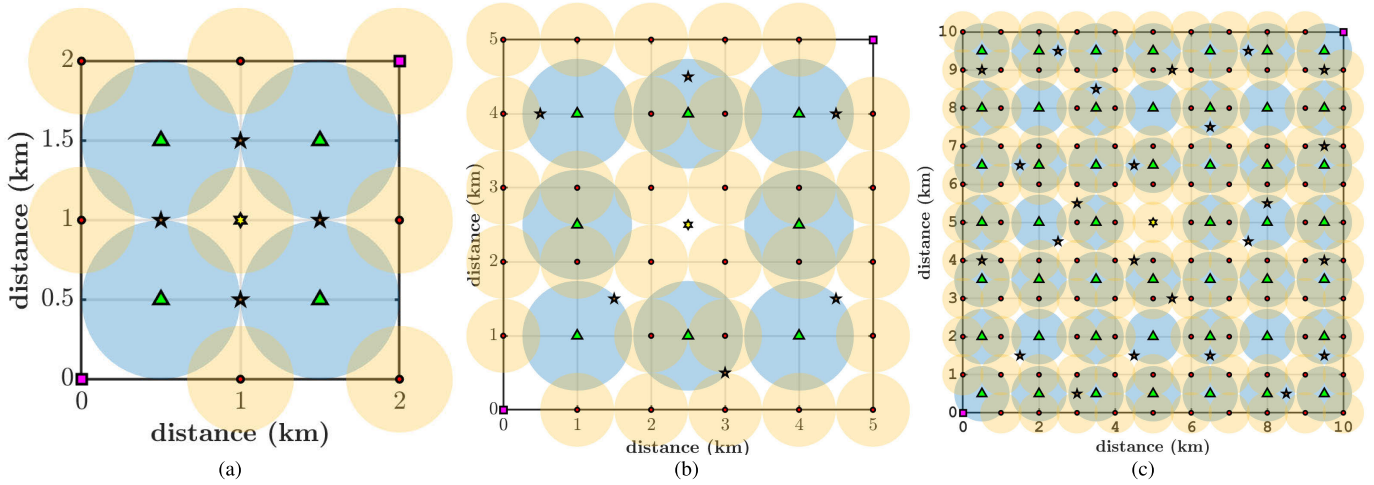
Fig. 9. Considered O-RAN deployment areas with two CO locations at diagonally opposite corner points with total dimension (a) 2 km × 2 km, (b) 5 km × 5 km, and (c) 10 km × 10 km (pink square: CO locations, orange pentagon: level-1 splitter, yellow hexagon: level-2 splitter, red circle: small-cell RU, green triangle: macro-cell RU).

would have remain unallocated with zero cost of resources. Hence, the actual payment made by RU 1 is given by:

$$P_1 = (C_\lambda B_1 + C_P G_1) - 0$$
$$= (C_\lambda B_1 + C_P G_1).$$

This shows that if an RU gets sufficient resources to transmit its radio data generated in each TTI, then its payment rule is independent of its shared private information. Therefore, sharing truthful information is a weakly dominant strategy in this scenario.

Example 2: [Two RUs and One Edge/OLT-Cloud]

Let us consider the total x-haul throughput is B_1 and total Edge/OLT-Cloud resources is G_1. The RU 1 submits a message b_1 and the RU 2 submits a message b_2 where $b_1 \leq b_2$. The results of Case 1 and Case 2 can be shown in a straight-forward manner. If the scenario is like Case 3, then both the RUs are connected to Edge/OLT-Cloud 1. Therefore the valuations of both the RUs are:

$$V_1 = V_2 = (C_\lambda B_1 + C_P G_1).$$

Now, the cost of resources for each of the RUs is $(C_\lambda B_1 + C_P G_1)/2$. However, if RU 1 was not present, then the entire block of resources would have been allocated to RU 2 with cost of resources $= (C_\lambda B_1 + C_P G_1)$. Therefore, the actual payment made by RU 1 is given by:

$$P_1 = (C_\lambda B_1 + C_P G_1) - \left(\frac{C_\lambda B_1 + C_P G_1}{2}\right)$$
$$= \left(\frac{C_\lambda B_1 + C_P G_1}{2}\right).$$

A similar analysis is applicable for RU 2. This shows that if an RU gets sufficient resources to transmit its radio data generated in each TTI, then its payment rule is independent of its shared private information. Therefore, sharing truthful information is a weakly dominant strategy in this scenario.

Example 3: [Three RUs and Two Edge/OLT-Clouds]

Let us consider the total x-haul throughput are B_1, B_2 and total Edge/OLT-Cloud resources are G_1, G_2. The RUs submit

messages (b_1, b_2, b_3) to the NSP where $b_1 \leq b_2 \leq b_3$. The results of Case 1 and Case 2 are straight-forward to show. If the scenario is like Case 3 and we assume that *RU 1 and RU 2 are allocated to Edge-Cloud 1 whereas RU 3 is allocated to Edge-Cloud 2*, then the valuations of all the RUs are:

$$V_1 = V_2 = (C_\lambda B_1 + C_P G_1),$$
$$V_3 = (C_\lambda B_2 + C_P G_2).$$

Now, the cost of shared resources for each of the RUs are given as follows:

$$\tilde{C}_1 = \tilde{C}_2 = (C_\lambda B_1 + C_P G_1)/2,$$
$$\tilde{C}_3 = (C_\lambda B_2 + C_P G_2).$$

However, if RU 1 was not present, then two cases can happen. Either RU 3 is allocated to Edge/OLT-Cloud 1 along with RU 2 while Edge/OLT-Cloud 2 remains inactive, or RU 2 is allocated to Edge/OLT-Cloud 1 while RU 2 is allocated to Edge/OLT-Cloud 2. Nonetheless, in both the cases, the total cost of resources to be paid by RUs allocated to Edge/OLT-Clouds 1 is $(C_\lambda B_1 + C_P G_1)$. Therefore, the actual payment made by RU 1 is given by:

$$P_1 = (C_\lambda B_1 + C_P G_1) - \left(\frac{C_\lambda B_1 + C_P G_1}{2}\right)$$
$$= \left(\frac{C_\lambda B_1 + C_P G_1}{2}\right).$$

We can analyze the payment rules for RU 2 and RU 3 by following the previous examples. Furthermore, *if all the three RUs were allocated to Edge/OLT-Cloud 1* in the first place, then we can show through a similar analysis as is Example 2 that the payment made by RU 1 is given by:

$$P_1 = \left[\left(\frac{C_\lambda B_1 + C_P G_1}{2}\right) + \left(\frac{C_\lambda B_1 + C_P G_1}{2}\right)\right]$$
$$- \left[\left(\frac{C_\lambda B_1 + C_P G_1}{3}\right) + \left(\frac{C_\lambda B_1 + C_P G_1}{3}\right)\right]$$
$$= (C_\lambda B_1 + C_P G_1) - \frac{2}{3}(C_\lambda B_1 + C_P G_1)$$

$$= \left(\frac{C_\lambda B_1 + C_P G_1}{3} \right).$$

This clearly shows that if an RU gets sufficient resources to transmit its radio data generated in each TTI, then its payment rule is independent of its shared private information. Therefore, sharing truthful information is a weakly dominant strategy in this scenario.

Appendix D
Considered O-RAN Deployment Areas

In this section, we graphically show some O-RAN deployment areas that are considered for the numerical evaluation of our proposed frameworks. Firstly, Fig. 9(a) shows an O-RAN deployment area of dimension 2 km × 2 km, secondly, Fig. 9(b) shows an O-RAN deployment area of dimension 5 km × 5 km, and thirdly, Fig. 9(c) shows an O-RAN deployment area of dimension 10 km × 10 km. In all these figures, each macro-cell RU has an approximate coverage area of 1.0 km (shown using light-blue circles) and each small-cell RU has an approximate coverage area of 0.5 km (shown using light-yellow circles). Also, a few symbols are used to denote key elements, i.e., pink square: CO locations, orange pentagon: level-1 splitter, yellow hexagon: level-2 splitter, red circle: small-cell RU, and green triangle: macro-cell RU. The actual fiber connections are not shown to avoid the figures from being over-crowded and stingy. The coordinates of the two CO locations are given by $(0,0)$ and (D,D) where $D \in \{2,5,10\}$. A group of TWDM-PONs branch out from each of the CO locations that are used to connect the RUs, e.g., the RUs located below the primary diagonal are all connected to the CO at $(0,0)$, but the RUs located above the primary diagonal are all connected to the CO at (D,D). Although a rectangular grid-like deployment of RUs are considered for our proposed framework evaluation, any practical RU deployment data can be used as an input to this framework.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, Oct. 2019.

[2] *O-RAN: Towards An Open and Smart RAN*, O-RAN Alliance, Alfter, Germany, Oct. 2018.

[3] I. Chih-Lin, H. Li, J. Korhonen, J. Huang, and L. Han, "RAN revolution with NGFI (xHaul) for 5G," *J. Lightw. Technol.*, vol. 36, no. 2, pp. 541–550, Jan. 15, 2018.

[4] L. Peterson et al., "Democratizing the network edge," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 49, no. 2, pp. 31–36, May 2019.

[5] (Jul. 2019). *The Open Road to 5G*. Samsung Networks. [Online]. Available: https://image-us.samsung.com/SamsungUS/samsungbusiness/pdfs/Open-RAN-The-Open-Road-to-5G.pdf

[6] B. Naudts, M. Kind, F.-J. Westphal, S. Verbrugge, D. Colle, and M. Pickavet, "Techno-economic analysis of software defined networking as architecture for the virtualization of a mobile network," in *Proc. Eur. Workshop Softw. Defined Netw.*, Oct. 2012, pp. 67–72.

[7] T. Pfeiffer, P. Dom, S. Bidkar, F. Fredricx, K. Christodoulopoulos, and R. Bonk, "PON going beyond FTTH [invited tutorial]," *J. Opt. Commun. Netw.*, vol. 14, no. 1, pp. A31–A40, 2022.

[8] S. Das, F. Slyne, A. Kaszubowska, and M. Ruffini, "Virtualized EAST–WEST PON architecture supporting low-latency communication for mobile functional split based on multiaccess edge computing," *J. Opt. Commun. Netw.*, vol. 12, no. 10, pp. D109–D119, 2020.

[9] L. Massoulie and J. Roberts, "Bandwidth sharing: Objectives and algorithms," *IEEE/ACM Trans. Netw.*, vol. 10, no. 3, pp. 320–328, Jun. 2002.

[10] D. Niyato, N. C. Luong, P. Wang, and Z. Han, "Second-price sealed-bid auction," in *Auction Theory for Computer Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2020, ch. 6, pp. 119–157.

[11] L. Tan, *Resource Allocation and Performance Optimization in Communication Networks and the Internet*. Boca Raton, FL, USA: CRC Press, 2017.

[12] D. Wypiór, M. Klinkowski, and I. Michalski, "Open RAN—Radio access network evolution, benefits and market trends," *Appl. Sci.*, vol. 12, no. 1, p. 408, Jan. 2022.

[13] J. Yao and N. Ansari, "QoS-aware joint BBU-RRH mapping and user association in cloud-RANs," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 881–889, Dec. 2018.

[14] M. Barahman, L. M. Correia, and L. S. Ferreira, "A fair computational resource management strategy in C-RAN," in *Proc. Int. Conf. Broadband Commun. Next Gener. Netw. Multimedia Appl. (CoBCom)*, Jul. 2018, pp. 1–6.

[15] N. Mharsi and M. Hadji, "Joint optimization of communication latency and resource allocation in cloud radio access networks," in *Proc. Int. Conf. Smart Commun. Netw. Technol. (SaCoNeT)*, Oct. 2018, pp. 13–18.

[16] M. Peng, Y. Yu, H. Xiang, and H. V. Poor, "Energy-efficient resource allocation optimization for multimedia heterogeneous cloud radio access networks," *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 879–892, May 2016.

[17] M. Y. Lyazidi, N. Aitsaadi, and R. Langar, "Dynamic resource allocation for cloud-RAN in LTE with real-time BBU/RRH assignment," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[18] S. Guo, D. Zeng, L. Gu, and J. Luo, "When green energy meets cloud radio access network: Joint optimization towards Brown energy minimization," *Mobile Netw. Appl.*, vol. 24, no. 3, pp. 962–970, Jun. 2019.

[19] K. Boulos, M. El Helou, and S. Lahoud, "RRH clustering in cloud radio access networks," in *Proc. Int. Conf. Appl. Res. Comput. Sci. Eng. (ICAR)*, Oct. 2015, pp. 1–6.

[20] H. M. Soliman and A. Leon-Garcia, "QoS-aware joint RRH activation and clustering in cloud-RANs," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–6.

[21] D. Liang, R. Gu, Q. Guo, and Y. Ji, "Demonstration of multi-vendor multi-standard PON networks for network slicing in 5G-oriented mobile network," in *Proc. Asia Commun. Photon. Conf.*, 2017.

[22] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, programmable, and virtualized 5G networks: State-of-the-art and the road ahead," *Comput. Netw.*, vol. 182, Dec. 2020, Art. no. 107516.

[23] M. Dryjański, u. Kułacz, and A. Kliks, "Toward modular and flexible open RAN implementations in 6G networks: Traffic steering use case and O-RAN xApps," *Sensors*, vol. 21, no. 24, p. 8173, Dec. 2021.

[24] S. Mondal and M. Ruffini, "Optical front/mid-haul with open access-edge server deployment framework for sliced O-RAN," *IEEE Trans. Netw. Service Manage.*, vol. 19, no. 3, pp. 1–18, Sep. 2022.

[25] M. Karbalaee Motalleb, V. Shah-Mansouri, S. Parsaeefard, and O. L. A. Lopez, "Resource allocation in an open RAN system using network slicing," *IEEE Trans. Netw. Service Manage.*, vol. 20, no. 1, pp. 471–485, Mar. 2022.

[26] X. Wang, J. D. Thomas, R. J. Piechocki, S. Kapoor, R. Santos-Rodríguez, and A. Parekh, "Self-play learning strategies for resource assignment in open-RAN networks," *Comput. Netw.*, vol. 206, Apr. 2022, Art. no. 108682.

[27] H. Zhang, H. Zhou, and M. Erol-Kantarci, "Team learning-based resource allocation for open radio access network (O-RAN)," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 4938–4943.

[28] M. Kassis, S. Costanzo, and M. Yassin, "Flexible multi-operator RAN sharing: Experimentation and validation using open source 4G/5G prototype," in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G Summit)*, Jun. 2021, pp. 205–210.

[29] M. Sauter, "Long term evolution (LTE) and LTE-advanced pro," in *From GSM to LTE-Advanced Pro and 5G*. Hoboken, NJ, USA: Wiley, 2017, ch. 4, pp. 211–333.

[30] R. Vilalta et al., "Network virtualization controller for abstraction and control of OpenFlow-enabled multi-tenant multi-technology transport networks," in *Proc. Opt. Fiber Commun. Conf.*, 2015.

[31] S. Mondal and M. Ruffini, "A min-max fair resource allocation framework for optical x-haul and DU/CU in multi-tenant O-RANs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 452–457.

[32] (Sep. 2020). *5G Wireless Fronthaul Requirements in a Passive Optical Network Context*, Telecommunication Standardization Sector of ITU (ITU-T). [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.Sup66-202009-I!!PDF-E&type=items

[33] M. Ruffini, A. Ahmad, S. Zeb, N. Afraz, and F. Slyne, "Virtual DBA: Virtualizing passive optical networks to enable multi-service operation in true multi-tenant environments," *J. Opt. Commun. Netw.*, vol. 12, no. 4, pp. B63–B73, 2020.

[34] F. Slyne, S. Zeb, and M. Ruffini, "Stateful DBA hypervisor supporting SLAs with low latency and high availability in shared PON," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2021, pp. 1–3.

[35] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "ColO-RAN: Developing machine learning-based xApps for open RAN closed-loop control on programmable experimental platforms," *IEEE Trans. Mobile Comput.*, early access, Jul. 4, 2022, doi: 10.1109/TMC.2022.3188013.

[36] (2015). *Small Cell Virtualization Functional Splits and Use Cases*. [Online]. Available: https://scf.io/en/documents/159_-_Small_Cell_Virtualization_Functional_Splits_and_Use_Cases.php

[37] (2017). *Study on New Radio Access Technology: Radio Access Architecture and Interfaces*. Accessed: Jul. 2, 2022. [Online]. Available: https://www.3gpp.org/DynaReport/38801.htm

[38] T. Tashiro et al., "A novel DBA scheme for TDM-PON based mobile fronthaul," in *Proc. Opt. Fiber Commun. Conf.*, 2014, pp. 1–3.

[39] M. Shehata, A. Elbanna, F. Musumeci, and M. Tornatore, "Multiplexing gain and processing savings of 5G radio-access-network functional splits," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 982–991, Dec. 2018.

[40] S. Khatibi, K. Shah, and M. Roshdi, "Modelling of computational resources for 5G RAN," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2018, pp. 1–5.

[41] E. K. P. Chong and S. H. Zak, "Multiobjective optimization," in *An Introduction to Optimization*. Hoboken, NJ, USA: Wiley, 2008, ch. 23, pp. 541–562.

[42] L. M. Ausubel and P. Cramton, "Auctioning many divisible goods," *J. Eur. Econ. Assoc.*, vol. 2, nos. 2–3, pp. 480–493, May 2004.

[43] Y. Narahari, "Vickrey-Clarke-Groves (VCG) Mechanisms," in *Game Theory and Mechanism Design*. Singapore: World Scientific Publishing Company, 2014, ch. 18, pp. 267–284.

[44] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.

[45] D. H. Hailu, B. G. Gebrehaweria, S. H. Kebede, G. G. Lema, and G. T. Tesfamariam, "Mobile fronthaul transport options in C-RAN and emerging research directions: A comprehensive study," *Opt. Switching Netw.*, vol. 30, pp. 40–52, Nov. 2018.

[46] Y. Nakayama et al., "Efficient DWBA algorithm for TWDM-PON with mobile fronthaul in 5G networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[47] *Preliminary Fiber Network Design and Business Plan Framework*, Columbia Telecommunications Corporation, Kensington, MD, USA, Jun. 2005.

[48] Y. Xiao, J. Zhang, and Y. Ji, "Can fine-grained functional split benefit to the converged optical-wireless access networks in 5G and beyond?" *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 3, pp. 1774–1787, Sep. 2020.

Sourav Mondal (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Kalyani Government Engineering College, affiliated with the West Bengal University of Technology, in 2012, the M.Tech. degree in telecommunication systems engineering from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, in 2014, and the Ph.D. degree from the Department of Electrical and Electronic Engineering, The University of Melbourne, in 2020. He was an Engineer with Qualcomm India Pvt. Ltd. from 2014 to 2016. He is currently working as an EDGE/Marie Skłodowska-Curie Post-Doctoral Fellow with the CONNECT Centre for Future Networks and Communication, Trinity College Dublin, Ireland.

Marco Ruffini (Senior Member, IEEE) received the M.Eng. degree in telecommunications from Marche Polytechnic University, Italy, in 2002, and the Ph.D. degree from Trinity College Dublin (TCD) in 2007. After working as a Research Scientist with Philips, Germany, he joined TCD in 2005, where he is currently an Associate Professor and a fellow. He is a Principal Investigator of both the IPIC Photonics Integration Centre and the CONNECT Telecommunications Research Centre, TCD. He is currently involved in several Science Foundation Ireland and H2020 projects, including a new research infrastructure to build a beyond 5G testbed in Dublin. He leads the Optical Network Architecture Group, TCD, and the new SFI funded Ireland's Open Networking testbed infrastructure, and has authored over 150 international publications, over ten patents, and contributed to standards at the broadband forum. He has raised research funding in excess of €7 million. His main research is in the area of 5G optical networks, where he carries out pioneering work on the convergence of fixed-mobile and access-metro networks, and on the virtualization of next generation networks. He has been invited to share his vision through several keynote and talks at major international conferences across the world.