Enabling Intelligent Traffic Steering in A Hierarchical Open Radio Access Network

Van-Dinh Nguyen[1], Thang X. Vu[2], Nhan Thanh Nguyen[3], Dinh C. Nguyen[4], Markku Juntti[3], Nguyen Cong Luong[5], Dinh Thai Hoang[6], Diep N. Nguyen[6] and Symeon Chatzinotas[2]
[1]*CECS, VinUniversity, Vietnam (dinh.nv2@vinuni.edu.vn);* [2]*SnT, University of Luxembourg;* [3]*University of Oulu, Finland*
[4]*University of Alabama in Huntsville, USA;* [5]*PHENIKAA University, Vietnam;* [6]*University of Technology Sydney, Australia*

***Abstract*—In this paper, we aim to enable an intelligent traffic (TS) steering application in the open radio access network (O-RAN) by jointly optimizing the flow-split distribution, congestion control and scheduling (*i.e.* so-called JFCS). To do so, we develop a multi-layer optimization framework based on network utility maximization and stochastic optimization methods. The proposed algorithm provides fast convergence, long-term utility-optimality and significantly low latency compared to state-of-the-art RAN approaches. In particular, our main contributions are as follows: i) we propose the novel JFCS framework to efficiently and adaptively route traffic to indented users in appropriate radio units, and ii) we develop low-complexity algorithms to effectively solve the JFCS problem in different time scales, enabling a closed-loop control of the TS in the O-RAN context. The insights presented in this work will pave the way for O-RAN that are completely automated, offering improved control and flexibility.**

I. Introduction

With the great success of mobile Internet, fifth generation (5G) cellular networks have been standardized to meet competing demands (*e.g.* extremely high data rate, low-latency and massive connectivity) and proliferation of heterogeneous devices. However, the existing "one-size-fits-all" 5G architecture lacks sufficient intelligence and flexibility to enable the coexistence of these demands. As we move towards 6G, the forefront of this endeavor lies in open radio access network (O-RAN). This approach involves the separation of radio access network components and the opening of interfaces, which is currently regarded as the most promising approach to transform wireless technology from "*connected things*" to "*connected intelligence*" [1]

Multi-layer (a.k.a. cross-layer) optimization for traditional cellular RAN architectures has been extensively studied in the literature (see *e.g.*, [2] and references therein). In general, the existing works only optimize radio resources while other factors at higher layers (*e.g.* congestion control and routing) are overlooked, making guaranteed multi-layer quality-of-service (QoS) infeasible. In addition, the non-causal statistical knowledge of traffic demands is required to model queue states, which is again impractical. So far, there have been only few attempts to study the applicability of the O-RAN architecture. Kumar *et al.* [3] proposed an automatic neighbour relation (ANR) approach to manage neighbour cell relationships by leveraging machine learning (ML) techniques, hence improving gNodeB (gNB) handovers. The authors in [4] developed an RL-based dynamic function splitting which is shown to be able to effectively decide the O-RAN's function splits and reduce operating costs. In traditional RAN architectures, the traffic steering (TS) solutions are typically determined by users' radio conditions of a serving cell while treating signals from neighboring cells as interference [5]. The authors in [6] proposed a distributed TS scheme through edge servers, where the matrix-based shortest path selection and matrix-based multipath searching algorithms are developed to dynamically decide the best paths for traffic steering. Very recently, Kavehmadavani *et al.* [7] showed that a dynamic multi-connectivity (MC)-based TS scheme can help steer traffic flows towards the most suitable cells based on user-centric condition. However, this work does not embed AI/ML solutions in Non-real-time RAN intelligent controller (Non-RT RIC) and assumes that all network information are available at Near-RT RIC to optimize radio resource allocation. In this paper, different from the aforementioned works, we propose a fully multi-layer optimization framework that captures interplays between the physical and higher layers, enabling proactive optimization of network parameters through RICs with periodic feedback loops.

We consider the fact that the complete information of the RAN layer is not available at the beginning of each time-frame. Instead, we assume that only their expected values are available to approximately measure queueing delay. An interesting question naturally arises: *How does the incomplete information of user traffic demands affect the optimal choices of the TS scheme?* To answer this question, we introduce a holistic multi-layer optimization framework, which jointly optimizes the flow-split distribution, congestion control and scheduling (called JFCS). The proposed framework effectively characterizes the complex interactions between layers (*e.g.* flow-split selection, congestion control rate and power allocation). In summary, we make the following two key contributions:

- We propose a novel JFCS framework to efficiently and adaptively route data traffic to appropriate radio units. Our framework provides a synergy between reinforcement learning (RL), QCS and updated network state information, and thus enabling a closed-loop control of the TS in the O-RAN context.
- To ensure the practicality and scalability, we identify inherent properties of the JFCS problem and propose an intelligent resource management algorithm to build the

This work was supported in part by the VinUniversity Seed Grant Program.

smoothed best response while maximizing the long-term utility for each data-flow under arbitrary changes in traffic demands.

II. Network Model and Problem Formulation

A. Network Model

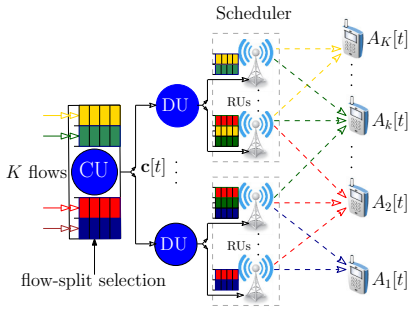


Fig. 1: Illustration of the system model enabling TS where each DU connects to multiples RUs towards cost-effective deployment.

As in Fig. 1, we consider the O-RAN architecture with one centralized unit (CU), I distributed units (DUs) and J radio units (RUs), where each DU connects to multiple RUs for cost-effective deployment. We consider a downlink multi-user multiple-input single-output (MU-MISO) system, where J RUs simultaneously serve the set $\mathcal{K} \triangleq \{1, 2, \cdots, K\}$ of $K = |\mathcal{K}|$ single-antenna UEs. The j-th RU served by the i-th DU is referred to as RU (i, j), which is equipped with $M_{i,j}$ antennas. The set of RUs served by DU i is denoted by $\mathcal{J}_i \triangleq \{(i, 1), \cdots, (i, J_i)\}$ with $|\mathcal{J}_i| = J_i$ and $\sum_{i\in\mathcal{I}} J_i = J$. The total set of RUs is denoted as $\mathcal{J} \triangleq \cup_{i\in\mathcal{I}}\mathcal{J}_i$.

We consider that the system operates in discrete time-frame indexed by $t \in [1, 2, \cdots, T]$, which corresponds to one large-scale coherence time with duration of T_c, as shown in Fig. 2. Each frame is divided into T_f time-slots of equal duration $\tau = T_c/T_f$, where the time-slot is indexed by $t_s = tT_f + s$ with $s \in \{1, \cdots, T_f\}$. At CU, there exists K independent data-flows where each of which is intended for one UE. The CU splits data-flow of UE k, say flow k, into multiple sub-flows which are possibly transmitted through the set of paths and then aggregated at this UE, so-called "traffic steering". For each data-flow k, we denote by $\mathcal{P}_k \triangleq \{(i, j)\}_{\forall(i,j)\in\mathcal{J}}$ the set of path states, including queue states and routing tables. A subset of separate paths in the set $\mathcal{P}_k$ (*i.e.*, via neighboring RUs indexed by (i, j)) should be appropriately selected. Let us denote by $\mathbf{c}_k[t] \triangleq \left[c_k^{i,j}[t]\right]_{(i,j)\in\mathcal{P}_k}$ the flow-split selection (action) vector for data-flow k in time-frame t, *i.e.*, $c_k^{i,j}[t] = 1$ if path $(i, j)\in \mathcal{P}_k$ (*i.e.*, via RU (i, j)) is selected to transmit data of flow k; otherwise, $c_k^{i,j}[t] = 0$. We let $\beta_k^{i,j}[t] \in [0, 1]$ be the fraction of data-flow k which is routed via path (i, j) in time-frame (state) t by selecting action $c_k^{i,j}[t]$, where $\sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t] = 1$. The global flow-split decision is denoted by $\mathcal{B}[t] \triangleq \{\boldsymbol{\beta}_k[t], \forall k \big| \sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t] = 1, \forall k\}$, where each column flow-split vector $\boldsymbol{\beta}_k[t] \triangleq \left[\beta_k^{i,j}[t]\right]^{\mathsf{T}}_{(i,j)\in\mathcal{P}_k} \in \mathbb{R}^J$ corresponds to the flow-split vector of data-flow k.

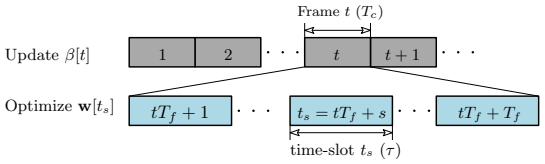


Fig. 2: Illustration of frame structure with each time-frame t.

1) Wireless Channel Model and Downlink Throughput

The channel vector between RU (i, j) and UE $k \in \mathcal{K}$ in time-slot t_s is denoted by $\mathbf{h}_k^{i,j}[t_s] \in \mathbb{C}^{M_{i,j}\times 1}$, which follows the Rician fading model with the Rician factor $\kappa_k^{i,j}[t]$. In particular, $\mathbf{h}_k^{i,j}[t_s]$ is modeled as $\mathbf{h}_k^{i,j}[t_s] = \sqrt{\xi_k^{i,j}[t]}\Big(\sqrt{\kappa_k^{i,j}[t]/(\kappa_k^{i,j}[t]+1)}\bar{\mathbf{h}}_k^{i,j}[t] + \sqrt{1/(\kappa_k^{i,j}[t]+1)}\tilde{\mathbf{h}}_k^{i,j}[t_s]\Big)$ where $\xi_k^{i,j}[t]$ represents the large-scale fading; $\bar{\mathbf{h}}_k^{i,j}[t]$ and $\tilde{\mathbf{h}}_k^{i,j}[t_s] \sim \mathcal{CN}(0, \mathbf{I})$. Denoting by $x_k^{i,j}[t_s]$ and $\mathbf{w}_k^{i,j}[t_s] \in \mathbb{C}^{M_{i,j}\times 1}$ a unit-power data symbol and a linear beamforming vector transmitted from RU (i, j) to UE k, respectively, the received signal at UE k can be written as

$$y_k[t_s] = \sum\nolimits_{(i,j)\in\mathcal{P}_k} (\mathbf{h}_k^{i,j}[t_s])^{\mathsf{H}}\mathbf{w}_k^{i,j}[t_s]x_k^{i,j}[t_s] + \sum_{k'\in\mathcal{K}\setminus\{k\}} \sum_{(i,j)\in\mathcal{P}_{k'}} (\mathbf{h}_k^{i,j}[t_s])^{\mathsf{H}}\mathbf{w}_{k'}^{i,j}[t_s]d_{k'}^{i,j}[t_s] + \omega_k[t_s] \tag{1}$$

where $\omega_k[t_s]$ is the AWGN with power N_0.

In this work, we consider the zero-forcing beamforming (ZFBF) for downlink transmisison. To make ZFBF efficient and feasible, we consider that $M_{i,j} > K_{i,j} \triangleq \sum_{k\in\mathcal{K}} c_k^{i,j}[t] \leq K$, $\forall(i, j) \in \mathcal{J}$, which helps cancel the inter-user interference caused by this RU. In addition, the system bandwidth is equally allocated to each RU (i, j), *i.e.* $W^{i,j} = W/J$, to completely remove the intra-user interference and interference caused by other RUs. Under the considered ZFBF technique, beamformer $\mathbf{w}_k^{i,j}[t_s]$ at RU (i, j) is designed to satisfy $(\mathbf{h}_{k'}^{i,j}[t_s])^{\mathsf{H}}\mathbf{w}_k^{i,j}[t_s] = 0, \forall k' \in \mathcal{K} \setminus \{k\}$. We let $\mathbf{V}_k^{i,j}[t_s] \in \mathbb{C}^{M_{i,j}\times(M_{i,j}-K_{i,j}+1)}$ be the null space of $\mathbf{H}_{-k}^{i,j}[t_s] \triangleq \left[\mathbf{h}_1^{i,j}[t_s] \cdots \mathbf{h}_{k-1}^{i,j}[t_s]\ \mathbf{h}_{k+1}^{i,j}[t_s] \cdots \mathbf{h}_K^{i,j}[t_s]\right] \in \mathbb{C}^{M\times(K-1)}$. We can then write $\mathbf{w}_k^{i,j}[t_s] = \mathbf{V}_k^{i,j}[t_s]\tilde{\mathbf{w}}_k^{i,j}[t_s]$, where $\tilde{\mathbf{w}}_k^{i,j}[t_s] \in \mathbb{C}^{(M_{i,j}-K_{i,j}+1)\times 1}, \forall k, (i, j)$ are the solutions to the ZFBF-based problem. By defining $\tilde{\nu}_k^{i,j}[t_s] \triangleq \|(\tilde{\mathbf{h}}_k^{i,j}[t_s])^{\mathsf{H}}\|_2^2$ with $\tilde{\mathbf{h}}_k^{i,j}[t_s] \triangleq (\mathbf{h}_k^{i,j}[t_s])^{\mathsf{H}}\mathbf{V}_k^{i,j}[t_s] \in \mathbb{C}^{1\times(M_{i,j}-K+1)}$, we can equivalently express $\tilde{\mathbf{w}}_k^{i,j}[t_s]$ as $\tilde{\mathbf{w}}_k^{i,j}[t_s] = \sqrt{p_k^{i,j}[t_s]}\frac{(\tilde{\mathbf{h}}_k^{i,j}[t_s])^{\mathsf{H}}}{\sqrt{\tilde{\nu}_k^{i,j}[t_s]}}$, where $p_k^{i,j}$ is the transmit power coefficient allocated to UE k by RU (i, j). The downlink achievable rate (bits/s) of UE k from RU (i, j) in time-slot t_s can be written as $r_k^{i,j}(p_k^{i,j}[t_s]) \triangleq W^{i,j}\log_2\big(1 + \frac{p_k^{i,j}[t_s]\tilde{\nu}_k^{i,j}[t_s]}{N_0}\big)$. By $\mathbf{p}_k[t_s] \triangleq \left[p_k^{i,j}[t_s]\right]_{(i,j)\in\mathcal{P}_k}$, the overall effective data rate of data-flow k (or UE k) can be computed as $r_k(\mathbf{p}_k[t_s]) = \sum_{(i,j)\in\mathcal{P}_k} r_k^{i,j}(p_k^{i,j}[t_s])$. Given $\mathbf{H}[t_s] \triangleq \left[\mathbf{h}_1[t_s] \cdots \mathbf{h}_K[t_s]\right] \in \mathbb{C}^{M\times K}$ with $\mathbf{h}_k[t_s] \triangleq \left[(\mathbf{h}_k^{i,j}[t_s])^{\mathsf{H}}\right]^{\mathsf{H}}_{\forall i,j} \in \mathbb{C}^{M\times 1}$ and $\boldsymbol{\beta}_k[t]$, we define the

instantaneous achievable rate region under $\mathbf{p}_k[t_s]$ as

$$\mathscr{C}_{\mathbf{H}[t_s]} \triangleq \left\{ r_k(\mathbf{p}_k[t_s]), \forall k \middle| \begin{matrix} r_k(\mathbf{p}_k[t_s]) = \sum\limits_{(i,j)\in\mathcal{P}_k} r_k^{i,j}(p_k^{i,j}[t_s]) \\ \sum\limits_{k\in\mathcal{K}} p_k^{i,j}[t_s] \le P_{\max}^{i,j}, \forall (i,j) \end{matrix} \right\}.$$

2) Queueing Model

Let $A_k[t]$ (bits/s) be the total rate of instantaneous arrived data destined for UE k in time-frame t with mean $\mathbb{E}\{A_k\} = \bar{A}_k$. We assume that $A_k[t]$ is upper bounded by a finite constant $A^{\max}$, such as $A_k[t] \le A^{\max} < \infty, \forall k, t$, and unknown at the beginning of time-frame t. As a result, the queue-length of data-flow k at RU (i,j) in time-slot t_s evolves as follows: $q_k^{i,j}[t_{s+1}] = \Big[q_k^{i,j}[t_s] + \beta_k^{i,j}[t]A_k[t]\tau - r_k^{i,j}(p_k^{i,j}[t_s])\tau\Big]^+$, where $[x]^+ \triangleq \max\{0, x\}$. By $\mathbf{q}[t_s] \triangleq \big[q_k^{i,j}[t_s]\big]^{\mathsf{T}}_{k,(i,j)}$ and following [8], a queueing network is *stable* if the steady-state total queue-length remains finite: $\limsup\limits_{t_s\to\infty} \mathbb{E}\{\|\mathbf{q}[t_s]\|_1\} < \infty$.

B. Problem Formulation

Let $\bar{r}_k \triangleq \lim\limits_{t_s\to\infty} \frac{1}{t_s}\sum_{\ell=1}^{t_s} r_k(\mathbf{p}_k[\ell])$ denote the long-term average rate of data-flow k. Each UE k is associated with a utility function, denoted by $U_k(\bar{r}_k)$.

Assumption 1. *The utility function $U_k(\cdot)$ is assumed to satisfy the following conditions: i) $U_k(\cdot)$ is twice continuously differentiable, increasing, and strictly concave; and ii) There exist positive constants $0 < \psi < \Psi < \infty$, such as $\psi \le -U_k^{''}(\bar{r}_k) \le \Psi, \forall \bar{r}_k \in [0,\ \bar{r}^{\max}]$, with $\bar{r}^{\max}$ being the maximum long-term average rate of any data flow.*

Based on the network utility maximization (NUM) framework, the joint flow-split distribution, congestion control and scheduling optimization problem (JFCS) is mathematically formulated as

$$\textbf{JFCS}: \max_{\boldsymbol{\beta},\bar{\mathbf{r}},\mathbf{p}} \sum_{k\in\mathcal{K}} U_k(\bar{r}_k) \tag{2a}$$

$$\text{s.t. } \limsup_{t_s\to\infty} \mathbb{E}\{\|\mathbf{q}[t_s]\|_1\} < \infty \tag{2b}$$

$$r_k(\mathbf{p}_k[t_s]) \in \mathscr{C}_{\mathbf{H}[t_s]}, \forall t_s, k \in \mathcal{K} \tag{2c}$$

$$\boldsymbol{\beta}_k[t] \in \mathscr{B}[t], \forall t, k \in \mathcal{K} \tag{2d}$$

$$\mathsf{Prob}\Big(\frac{q_k^{i,j}[t_s]}{\bar{A}_k} \le \bar{d}_k\Big) \ge \epsilon_k,\ \forall t_s, k, (i,j) \tag{2e}$$

where $\boldsymbol{\beta} \triangleq \big[\boldsymbol{\beta}_k^{\mathsf{T}}\big]^{\mathsf{T}}_{k\in\mathcal{K}}$, $\bar{\mathbf{r}} \triangleq \big[\bar{r}_k\big]^{\mathsf{T}}_{k\in\mathcal{K}}$ and $\mathbf{p} \triangleq \big[\mathbf{p}_k\big]_{k\in\mathcal{K}}$. Constraint (2e) ensures different minimum outage delay requirements for sub-flows, where $\bar{d}_k$ and ϵ_k ($0 \ll \epsilon_k \le 1$) are the maximum allowable average delay and the required reliable communication for each UE, respectively.

III. JFCS-based Network Utility Optimization

A. Tractable Form of the JFCS Problem (2)

Challenges of Solving JFCS Problem (2): We can observe that that constraint (2c) is nonconvex while (2e) is a nonconvex probabilistic constraint, generally making problem (2) NP-hard. In addition, the expectations in the constraints cause the stochastic nature of the problem, which cannot be solved directly. The classical optimization approaches, such as successive convex approximation (SCA) [9], are often applied to solve the optimization problems of nonconvex and deterministic constraints. However, the stochastic SCA-based algorithms can no longer guarantee a feasible and (sub)-optimal solution of all subsequent time intervals (TTIs) due to dynamics of physical layer at small timescales.

Towards a safe design, we consider the replacement of constraint (2e) by its deterministic constrain. From the well-known Markov inequality [10], we can show that $\mathsf{Prob}\big(q_k^{i,j}[t_s] \ge \bar{A}_k\bar{d}_k\big) \le \mathbb{E}\{q_k^{i,j}[t_s]\}/\bar{A}_k\bar{d}_k$, yielding

$$\sum\nolimits_{\ell=1}^{t} \beta_k^{i,j}[\ell]\bar{A}_k\tau - (1-\epsilon_k)\bar{A}_k\bar{d}_k - \sum\nolimits_{\ell=1}^{t_{s-1}} r_k^{i,j}(p_k^{i,j}[\ell])\tau \le r_k^{i,j}(p_k^{i,j}[t_s])\tau,\ \forall t_s, k \in \mathcal{K}, (i,j) \in \mathcal{P}_k \tag{3}$$

where each queue-length is always non-negative.

To facilitate the following optimization, we introduce congestion control variables $\boldsymbol{a}[t_s] \triangleq \big[a_k[t_s]\big]^{\mathsf{T}}_{k\in\mathcal{K}}$, satisfying $\bar{a}_k - \bar{r}_k \le 0, \forall k$, where $\bar{a}_k \triangleq \lim\limits_{t_s\to\infty} \frac{1}{t_s}\sum_{\ell=1}^{t_s} a_k[\ell]$. Problem (2) is then rewritten as

$$\max_{\boldsymbol{\beta},\bar{\boldsymbol{a}},\bar{\mathbf{r}},\mathbf{p}} \sum_{k\in\mathcal{K}} U_k(\bar{a}_k) \tag{4a}$$

$$\text{s.t.} \quad (2b), (2c), (2d), (3) \tag{4b}$$

$$\bar{a}_k - \bar{r}_k \le 0, \forall k. \tag{4c}$$

We also introduce a new auxiliary queue-length vector $\hat{\mathbf{q}}[t_s] \triangleq \big[\hat{q}_k[t_s]\big]^{\mathsf{T}}_{k\in\mathcal{K}}$, where $\hat{q}_k[t_{s+1}] = \big[\hat{q}_k[t_s] + a_k[t_s]\tau - r_k(\mathbf{p}_k[t_s])\tau\big]^+$ to associate constraint (4c) with a penalty function and $a_k[t_s] \in [0, A^{\max}]$. We define the total queue backlog of all UEs in time-slot t_s as $L[t_s] = \frac{1}{2}\big(\sum_{k\in\mathcal{K}} \sum_{(i,j)\in\mathcal{P}_k} \frac{q_k^{i,j}[t_s]^2}{\tau^2} + \sum_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]^2}{\tau^2}\big)$. For given $(\mathbf{q}[t_s], \hat{\mathbf{q}}[t_s])$, the Lyapunov drift from time-slot t_s to t_{s+1} is given as $\Delta L[t_s] = L[t_{s+1}] - L[t_s]$. To guarantee joint network stability and penalty minimization, we adopt the drift-plus-penalty procedure [11] to minimize the drift of a quadratic Lyapunov function and rewrite (4) as

$$\max_{\boldsymbol{\beta},\bar{\boldsymbol{a}},\bar{\mathbf{r}},\mathbf{p}} \varphi \sum_{k\in\mathcal{K}} \mathbb{E}\{U_k(a_k[t_s])\} - \mathbb{E}\{\Delta L[t_s]\}, \text{s.t. } (2c), (2d), \tag{5}$$

where φ is a scaling factor to balance two objective functions.

B. Overall Intelligent Resource Management Algorithm

From the inequality $([x]^+)^2 \le x^2$ and $(x+y)^2 - x^2 = 2xy + y^2$, we have

$$\Delta L^{\text{UB}}[t_s] \triangleq \sum\nolimits_{k\in\mathcal{K}} \sum\nolimits_{(i,j)\in\mathcal{P}_k} \frac{q_k^{i,j}[t_s]}{\tau}\big(\beta_k^{i,j}[t]A_k[t] - r_k^{i,j}(p_k^{i,j}[t_s])\big) + \sum\nolimits_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau}\big(a_k[t_s] - r_k(\mathbf{p}_k[t_s])\big) + B[t_s] \ge \Delta L[t_s] \tag{6}$$

where $B[t_s] \triangleq \frac{1}{2}\sum_{k\in\mathcal{K}}\sum_{(i,j)\in\mathcal{P}_k}\big(\beta_k^{i,j}[t]A_k[t] - r_k^{i,j}(p_k^{i,j}[t_s])\big)^2 + \frac{1}{2}\sum_{k\in\mathcal{K}}\big(a_k[t_s] - r_k(\mathbf{p}_k[t_s])\big)^2$. Following [12], we consider that $B[t_s]$ is finite and bounded by $\bar{B}$ for all t_s, *i.e.*, $\mathbb{E}\{B[t_s]\big|\mathbf{q}[t_s], \hat{\mathbf{q}}[t_s]\} \le \bar{B}$. As a result, problem (5) is simplified to

$$\max_{\boldsymbol{\beta},\bar{\boldsymbol{a}},\bar{\mathbf{r}},\mathbf{p}} \quad \varphi \sum_{k\in\mathcal{K}} \mathbb{E}\{U_k(a_k[t_s])\} - \mathbb{E}\{\Delta L^{\text{UB}}[t_s]\} \tag{7a}$$

$$\text{s.t.} \quad (2c), (2d), (3). \tag{7b}$$

Long-term subproblem (L-SP): Given $\mathcal{L}_k[t] = \sum_{(i,j)\in\mathcal{P}_k} q_k^{i,j}[t_s]/\tau(r_k^{i,j}(p_k^{i,j}[t_s]) - \beta_k^{i,j}[t]A_k[t])$, the flow-split

distribution subproblem at time-frame t is given as

$$\textbf{L-SP}: \max_{\boldsymbol{\beta}_k[t]\in\mathscr{B}[t],\forall k} \quad \sum_{k\in\mathcal{K}} \mathscr{L}_k[t]. \tag{8}$$

Although problem (8) is a linear program in $\boldsymbol{\beta}$, it cannot be solved directly by standard optimization techniques because $A_k[t], \forall k$ are *incompletely* known at the beginning of time-frame t.

Short-term subproblems (S-SPs): The congestion control subproblem at time-slot t_s is

$$\textbf{S-SP1}: \max_{\boldsymbol{a}[t_s]\geq 0} \quad \sum_{k\in\mathcal{K}} \big(\varphi U_k(a_k[t_s]) - \frac{\hat{q}_k[t_s]}{\tau} a_k[t_s]\big) \tag{9}$$

which is an unconstrained convex problem. The optimal solution of (9) exists and is unique that is $a_k^*[t_s] = U_k^{'-1}\big(\frac{\hat{q}_k[t_s]}{\varphi\tau}\big), \forall k$, where $U_k^{'-1}(\cdot)$ denotes the inverse function of the first derivation of $U_k(\cdot)$. Given the optimal solution $\boldsymbol{\beta}^*[t]$, the short-term power control optimization subproblem (*i.e.*, the weighted queue-length-based scheduling) at time-slot t_s is given as

$$\textbf{S-SP2}: \max_{\mathbf{r}[t_s],\mathbf{p}[t_s]} \quad \sum_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k(\mathbf{p}_k[t_s]), \text{ s.t. (2c), (3).} \tag{10}$$

The overall intelligent resource management algorithm for solving the JFCS problem (2) is summarized in Algorithm 1, where the solutions of subproblems will be provided next.

Algorithm 1: Intelligent Resource Management Algorithm for Solving JFCS Problem (2), compliant with O-RAN

Initialization: Set $t = 1$ and select a positive scaling factor φ. Initialize $\boldsymbol{\beta}_k[1] = \frac{1}{|\mathcal{P}_k|}[1, \cdots, 1]$ and all queues are set to be empty: $q_k^{i,j}[1_1] = 0$ and $\hat{q}_k[1_1] = 0, \forall (i,j), k$.

Main Loop:

1: **for** each frame $t = 1, 2, \cdots, T$ **do** {*/*Long-term scale t*/*}

2: **Flow-Split Distribution:** Given $\{\mathbf{q}[t-1], \mathbf{A}[t-1]\}$, CU splits data-flows of all UEs based on the optimal flow-split decisions $\boldsymbol{\beta}^*[t]$ by solving L-SP at Non-RT RIC:

$$\max_{\boldsymbol{\beta}_k[t]\in\mathscr{B}[t],\forall k} \sum_{k\in\mathcal{K}} \mathscr{L}_k[t].$$

3: **for** each time-slot $t_s = tT_f + s$ with $s \in \{1, \cdots, T_f\}$ **do** {*/*Short-term scale t_s*/*}

4: **Congestion Controller:** Given $\hat{\mathbf{q}}[t_s]$, solve S-SP1 (9) to obtain the optimal congestion control variables:

$$a_k^*[t_s] = \min\Big\{U_k^{'-1}\big(\frac{\hat{q}_k[t_s]}{\varphi\tau}\big), A^{\max}\Big\}, \forall k.$$

5: **Weighted Queue-Length-Based Scheduler:** Given $\hat{\mathbf{q}}[t_s]$ and $\boldsymbol{\beta}^*[t]$, each RU $(i,j) \in \mathcal{P}_k$ schedules the service rate $r_k^{i,j}(p_k^{i,j}[t_s])$ for UE $k \in \mathcal{K}$ by solving S-SP2:

$$\max_{\mathbf{r}[t_s],\mathbf{p}[t_s]} \quad \sum_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k(\mathbf{p}_k[t_s]), \text{ s.t. (2c), (3).}$$

6: **Queue-Length Updates:** Queue-Lengths are updated as

$$q_k^{i,j}[t_{s+1}] = [q_k^{i,j}[t_s] + \beta_k^{i,j}[t]A_k[t]\tau - r_k^{i,j}(p_k^{i,j}[t_s])\tau]^+, \ \forall k, (i,j)$$

$$\hat{q}_k[t_{s+1}] = [\hat{q}_k[t_s] + a_k[t_s]\tau - r_k(\mathbf{p}_k[t_s])\tau]^+, \ \forall k.$$

7: Set $s = s + 1$

8: **end for**

9: Update $\{\mathbf{q}[t], \mathbf{A}[t]\} := \{q_k^{i,j}[t], A_k[t]\}_{k,(i,j)}$ to Non-RT RIC.

10: Set $t = t + 1$

11: **end for**

IV. Proposed Algorithms for Solving Subproblems

A. Reinforcement Learning Algorithm for Solving L-SP (8)

Let us denote by $u_k^{i,j}[t] \triangleq \frac{q_k^{i,j}[t_s]}{\tau}\big(r_k^{i,j}(p_k^{i,j}[t_s]) - \beta_k^{i,j}[t]A_k[t]\big)$ the instantaneous utility observation of data-flow k at time-frame t when selecting path $(i,j) \in \mathcal{P}_k$. The total utility observation of data-flow k, denoted by $u_k[t]$, is thus $u_k[t] = \sum_{(i,j)\in\mathcal{P}_k} u_k^{i,j}[t]$. Inspired by [13], we denote $\hat{u}_k^{i,j}[t]$ as the estimated utility of data-flow k at time-frame t when selecting path (i,j). In addition, the actual utility observed by data-flow k at time-frame t, denoted by $\bar{u}_k[t]$, is given as $\bar{u}_k[t] = u_k[t-1]$, which is based on feedback from Near-RT RIC at time $t-1$. By initializing $\hat{u}_k^{i,j}[1] = 0$, the estimated utility of data-flow k is updated for action $\mathbf{c}_k[t] = c_k^{i,j}[t]$ as follows:

$$\hat{u}_k^{i,j}[t] = \hat{u}_k^{i,j}[t-1] + \eta_u[t]\mathbb{1}_{\{\mathbf{c}_k[t]=c_k^{i,j}[t]\}}\big(\bar{u}_k[t] - \hat{u}_k^{i,j}[t-1]\big) \tag{11}$$

for $\forall t > 1$ where $\eta_u > 0$ is the decreasing step size (learning rate).

Next, we denote $\hat{\boldsymbol{\theta}}_k[t] \triangleq [\hat{\theta}_k^{i,j}[t]]_{(i,j)\in\mathcal{P}_k}$ as the estimated regret vector of data-flow k, where each element is updated as

$$\hat{\theta}_k^{i,j}[t] = \hat{\theta}_k^{i,j}[t-1] + \eta_\theta[t]\mathbb{1}_{\{\mathbf{c}_k[t]=c_k^{i,j}[t]\}}\big(\bar{u}_k[t] - \hat{u}_k^{i,j}[t] - \hat{\theta}_k^{i,j}[t-1]\big), \ \forall t > 1 \tag{12}$$

with $\hat{\theta}_k^{i,j}[1] = 0$ and $\eta_\theta[t]$ being the learning rate. We note that trying all possible actions to choose the best paths (*e.g.* exploration) can offer the highest payoff, but with the cost of slow convergence and even computationally prohibitive. During the exploitation process, playing an action associated with the highest estimated utility in (11) will likely result in a very sub-optimal solution. To make this tradeoff more efficient, let us define the best response function $\hat{\boldsymbol{\beta}}[t] = f(\hat{\boldsymbol{\theta}}[t])$ as

$$f(\hat{\boldsymbol{\theta}}[t]) := \underset{\boldsymbol{\beta}_k[t]\in\mathscr{B}[t]}{\operatorname{argmin}} \Big\{h\big(\boldsymbol{\beta}[t]\big) - \lambda \sum_{k\in\mathcal{K}} \sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t]\hat{\theta}_k^{i,j}[t]\Big\}. \tag{13}$$

Here λ is the so-called trade-off factor (a.k.a. Boltzmann temperature) and $h\big(\boldsymbol{\beta}[t]\big)$ denotes the regularization function. We note that when $\lambda \to 0$, it leads to uniform probabilities of all actions, *i.e.*, $\beta_k^{i,j}[t] = 1/|\mathcal{P}_k|, \forall (i,j) \in \mathcal{P}_k$. For $\lambda \to \infty$, the second term in (13) will dominate the best response function and then the actions associated with highest estimated regret will be selected [13].

Regularization function: The solutions of problem (8) lie in the unit simplex for each data-flow. Therefore, we adopt the Gibbs-Shannon entropy as the regularization function, *i.e.* $h\big(\boldsymbol{\beta}[t]\big) = \sum_{k\in\mathcal{K}}\sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t]\ln\big(\beta_k^{i,j}[t]\big)$, which is K-strongly convex. Substituting $h\big(\boldsymbol{\beta}[t]\big)$ into (13), we have

$$f(\hat{\boldsymbol{\theta}}[t]) := \underset{\boldsymbol{\beta}_k[t]\in\mathscr{B}[t],\forall k}{\operatorname{argmin}} \Big\{\sum_{k\in\mathcal{K}}\sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t]\ln\big(\beta_k^{i,j}[t]\big) - \lambda\sum_{k\in\mathcal{K}}\sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t]\hat{\theta}_k^{i,j}[t]\Big\}. \tag{14}$$

The function $f(\hat{\boldsymbol{\theta}}[t])$ is convex and separable for each $\beta_k^{i,j}[t]$. By solving $\partial f(\hat{\boldsymbol{\theta}}[t])/\partial\beta_k^{i,j}[t] = \ln\big(\beta_k^{i,j}[t]\big) + 1 - \lambda\hat{\theta}_k^{i,j}[t] = 0$, we have $\beta_k^{i,j}[t] = f(\hat{\theta}_k^{i,j}[t]) = \exp\big(\lambda\hat{\theta}_k^{i,j}[t] - 1\big)$. To ensure $\sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t] = 1, \forall k$, we normalize $f_k^{i,j}(\hat{\boldsymbol{\theta}}_k[t])$ through

the exponentiated mirror function as

$$f_k^{i,j}(\hat{\boldsymbol{\theta}}_k[t]) = \frac{\exp\left(\lambda\left[\hat{\theta}_k^{i,j}[t]\right]^+\right)}{\sum_{(i',j')\in\mathcal{P}_k}\exp\left(\lambda\left[\hat{\theta}_k^{i',j'}[t]\right]^+\right)}. \quad (15)$$

As a result, the estimate value of each element of flow-split vector $\boldsymbol{\beta}_k[t]$ is updated for all actions with the regret as $\beta_k^{i,j}[t] = \beta_k^{i,j}[t-1] + \eta_\beta[t]\big(f_k^{i,j}(\hat{\boldsymbol{\theta}}_k[t]) - \beta_k^{i,j}[t-1]\big)$ for $t > 1$, where $\boldsymbol{\beta}_k[1] = \frac{1}{|\mathcal{P}_k|}[1,\cdots,1]$ and $\eta_\beta[t]$ is the learning rate.

B. Proposed Solution for Solving S-SP2 (10)

Given the optimal solution $\boldsymbol{\beta}_k^*[t], \forall k$, the short-term power optimization problem (10) with ZFBF can be reformulated as

$$\max_{\mathbf{p}[t_s]} \quad \sum\nolimits_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k(p_k^{i,j}[t_s]) \quad (16a)$$

$$\text{s.t.} \quad \bar{R}_k^{i,j}[t_s] \le r_k^{i,j}(p_k^{i,j}[t_s])\tau,\ \forall k,(i,j) \quad (16b)$$

$$\sum\nolimits_{k\in\mathcal{K}} p_k^{i,j}[t_s] \le P_{\max}^{i,j},\ \forall (i,j). \quad (16c)$$

The function $r_k^{i,j}(p_k^{i,j}[t_s])$ is concave in $p_k^{i,j}[t_s]$, leading to the convexity of problem (16). From (16b), one can show that $p_k^{i,j}[t_s] \ge p_{k,\min}^{i,j}[t_s] := \frac{N_0}{\tilde{\nu}_k^{i,j}[t_s]} 2^{\frac{\bar{R}_k^{i,j}[t_s]}{W^{i,j}\tau}} - 1$. We now formulate the partial Lagrangian as

$$L(\mathbf{p}[t_s],\boldsymbol{\mu}) = \sum\nolimits_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k(p_k^{i,j}[t_s]) + \sum\nolimits_{(i,j)\in\mathcal{J}} \mu_{i,j}\big(P_{\max}^{i,j} - \sum\nolimits_{k\in\mathcal{K}} p_k^{i,j}[t_s]\big) \quad (17)$$

where $\boldsymbol{\mu} \triangleq \{\mu_{i,j} \ge 0\}_{(i,j)\in\mathcal{J}}$ are the Lagrange multipliers of constraint (16c). The dual function can be written as $g(\boldsymbol{\mu}) = \max\{L(\mathbf{p}[t_s],\boldsymbol{\mu}) | p_k^{i,j}[t_s] \ge p_{k,\min}^{i,j}[t_s], \forall k,\ (i,j)\}$. We note that $L(\mathbf{p}[t_s],\boldsymbol{\mu})$ is separable with respect to $p_k^{i,j}[t_s]$. Thus, by solving

$$p_k^{i,j*}[t_s] = \underset{p_k^{i,j}[t_s]\ge p_{k,\min}^{i,j}[t_s]}{\operatorname{argmax}} \left\{ \frac{\hat{q}_k[t_s]}{\tau} W \log_2\Big(1 + \frac{p_k^{i,j}[t_s]\tilde{\nu}_k^{i,j}[t_s]}{N_0}\Big) - \mu_{i,j}^* p_k^{i,j}[t_s] \right\} \quad (18)$$

for a given optimal Lagrange multiplier $\mu_{i,j}^*$, the optimal solution of $p_k^{i,j*}[t_s]$ is determined as

$$p_k^{i,j*}[t_s] = \max\Big\{p_{k,\min}^{i,j}[t_s], \frac{\hat{q}_k[t_s]W^{i,j}}{\tau\mu_{i,j}^*\ln 2} - \frac{N_0}{\tilde{\nu}_k^{i,j}[t_s]}\Big\}. \quad (19)$$

The optimal Lagrange multiplier $\mu_{i,j}^*$ is efficiently found by applying the bisection search method between $\underline{\mu}_{i,j} = 0$ and a sufficiently large $\overline{\mu}_{i,j}$. At iteration n, each RU (i,j) computes $\mu_{i,j}^{(n)} = (\underline{\mu}_{i,j} + \overline{\mu}_{i,j})/2$ and $p_k^{i,j(n)}[t_s]$ as in (19). If $\sum_{k\in\mathcal{K}} p_k^{i,j(n)}[t_s] - P_{\max}^{i,j} \le 0$, then compute $\mu'_{i,j} = (\underline{\mu}_{i,j} + \overline{\mu}_{i,j})/2$ and update $\overline{\mu}_{i,j} := \mu'_{i,j}$; otherwise, compute $\mu'_{i,j} = (\underline{\mu}_{i,j} + \overline{\mu}_{i,j})/2$ and then update $\underline{\mu}_{i,j} := \mu'_{i,j}$. This procedure is repeated until convergence. As a result, the optimal ZFBF solution is recovered as $\mathbf{w}_k^{i,j*}[t_s] = \Big(\sqrt{\tilde{p}_k^{i,j*}[t_s]}/\sqrt{\tilde{\nu}_k^{i,j}[t_s]}\Big)\mathbf{V}_k^{i,j}[t_s](\tilde{\mathbf{h}}_k^{i,j}[t_s])^{\mathsf{H}},\ \forall k,(i,j)$.

V. Numerical Results

A. Simulation Setups and Parameters

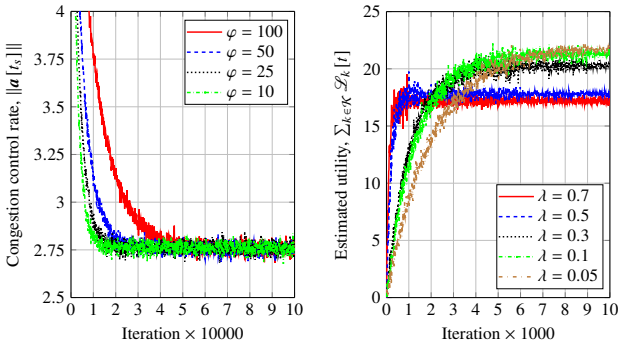
We consider a system topology which includes 8 RUs and 12 UEs located within a circle of 1-km radius. There are two DUs, each connected to 4 RUs. The large-scale fading coefficient $\xi[t] \in \{\xi_k^{i,j}[t]\}_{\forall(i,j),k}$, is modeled as the three-slope path loss model, such as $\xi[t] = \xi_0 - 35\log_{10}(d[t]) + 20c_0\log_{10}(d/d_0) + 15c_1\log_{10}(d/d_1)$ where $\xi_0 = -140.7 + \mathsf{SF}$ dB, $d_0 = 10$ m, $d_1 = 50$ m, and d is the distance between an RU and a UE; here $c_i = \max\{0, \frac{d_i - d}{|d_i - d|}\}$ with $i \in \{0,1\}$ and $\mathsf{SF} \sim \mathcal{CN}(0,\sigma_{\mathsf{SF}})$ denotes the shadowing factor with $\sigma_{\mathsf{SF}} = 8$ dB. The Rician factor $\kappa[t] \in \{\kappa_k^{i,j}[t]\}_{\forall(i,j),k}$ is given as $\kappa = P_{\mathsf{LoS}}(d[t])/\big(1 - P_{\mathsf{LoS}}(d[t])\big)$, where the LoS probability follows the 3GPP–UMa model as $P_{\mathsf{LoS}}(d[t]) = \min\Big(\frac{18}{d[t]}, 1\Big)\big(1 - \exp(-\frac{d[t]}{36})\big) + \exp(-\frac{d[t]}{36})$. The array response vector is generated as $\bar{\mathbf{h}}_k^{i,j}[t] = \boldsymbol{a}(\phi_k^{i,j}[t])$, where each element m is given as $\big[\boldsymbol{a}(\phi_k^{i,j}[t])\big]_m = \exp\big(j\pi(m-1)\sin\phi_k^{i,j}[t]\big)$ with $\phi_k^{i,j}[t] \in [-\pi/2, \pi/2)$ being the angle-of-departure (AoD) at RU (i,j). The noise power is modeled as $N_0 = -170 + 10\log_{10}(W) + \mathsf{NF}$ dBm with the noise figure $\mathsf{NF} = 9$ dB.

We run Algorithm 1 over $T = 10000$ frames, where each frame consists of $T_f = 10$ time-slots (subframes) and has duration of $T_c = 10$ ms, followed by 5G NR Frame structure. In each time-frame t, UE k is served by a subset of four RUs. To illustrate the heterogeneity of UEs, we assume that the arrival rate $A_k[t]$ is uniformly distributed in [1, 3] Gbps. The step sizes (learning rates) are set to decrease after each frame as $\eta_u[t] = 1/(t+1)^{0.51}$, $\eta_\theta[t] = 1/(t+1)^{0.55}$ and $\eta_\beta[t] = 1/(t+1)^{0.6}$ [14]. We adopt the proportional fairness metric to model the utility function as: $U_k(r_k) = \log(0.001 + r_k), \forall k$ [15]. The other parameters are given as $W = 20$ MHz, $M_{i,j} \equiv M = 16$, $P_{\max}^{i,j} \equiv P_{\max} = 43$ dBm, $\forall(i,j)$, $\bar{d}_k \equiv \bar{d} = 10$ ms and $\epsilon_k \equiv \epsilon = 0.95$, $\forall k$. In the following figures, results are averaged over the last 6000 frames.

Benchmark schemes: To demonstrate the benefits of the proposed JFCS algorithm, we consider the following three benchmark schemes: i) "NUM with fixed resource allocation (NUM-FRA)" [16]: Under Algorithm 1, RUs allocate power equally to UEs, ii) "NUM with equal flow-split distribution (NUM-EFSD):" CU splits data-flows of all UEs equally among the selected paths, *i.e.*, $\beta_k^{i,j}[t] = 1/|\mathcal{P}_k|, \forall(i,j) \in \mathcal{P}_k$, and iii) "NUM with the nearest RU selection (NUM-NRU):" Under Algorithm 1, each UE k selects only the nearest RU for the data transmission, *i.e.* $\beta_k^{i,j}[t] = 1$ if RU (i,j) is the nearest RU to UE k.

B. Numerical Results and Performance Comparison

From Fig. 3(a), it can be observed that the congestion control rates for different values of the scaling factor φ converge to the same optimal solution, and $\|\boldsymbol{a}[t_s]\|$ is almost independent of φ. In addition, increasing φ results in a smaller divergence of the steady-state congestion control rate, but also slows down the convergence rate of Algorithm 1. The reason is attributed to the fact that for a large φ, the network utility function $\sum_{k\in\mathcal{K}} U_k(a_k[t_s])$ in (5) will prevail over the Lyapunov drift function $\Delta L[t_s]$, which requires more iterations to guarantee network stability. In Fig. 3(b), we increase the trade-off factor λ (*i.e.* Boltzmann temperature) from 0.05 to 0.7. The result shows that the larger the value of λ, the better



(a) Impact of φ on congestion control rate

(b) Impact of λ on estimated utility

Fig. 3: Convergence behavior of Algorithm 1.

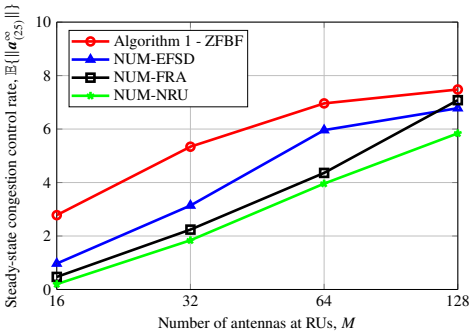


Fig. 4: The steady-state congestion control rate w.r.t. M.

the estimated utility that can be achieved with the cost of lower convergence speed of the RL process. Conversely, a low value of λ can speed up convergence by allocating traffic data uniformly to all paths, but leads to a very sub-optimal solution.

We show the performance comparison in terms of the steady-state congestion control rate $\mathbb{E}\{\|\boldsymbol{a}^{\infty}_{(25)}\|\}$ among the considered schemes versus the number of antennas at RUs in Fig. 4. We fix $\varphi = 25$ and vary $M \equiv M_{i,j}, \forall(i,j)$ from 16 to 128 to investigate the impact of the physical factor. As M increases, the downlink instantaneous achievable rates of all UEs also significantly increase since more degrees of freedom are added to leverage multi-user diversity, resulting in lower queue-lengths. For a fixed value of φ, the steady-state congestion control rate vector increases monotonically with M. Next, the impact of scaling factor φ on the steady-state total queue-length $\mathbb{E}\{\|\hat{\mathbf{q}}^{\infty}_{(\varphi)}\|_1\}$ is plotted in Fig. 5. It can be seen that the steady-state total queue-length of all schemes monotonically scales as $\mathcal{O}(\varphi)+\mathcal{O}(\sqrt{\varphi})$. Clearly, Algorithm 1 outperforms the benchmark schemes in all ranges of M, and the gap is deeper when M is small and φ is large.

VI. Conclusion

We have proposed a new holistic multi-layer optimization framework, called JFCS, to enable intelligent traffic steering in a hierarchical O-RAN architecture. By leveraging network utility maximization and stochastic optimization methods, we

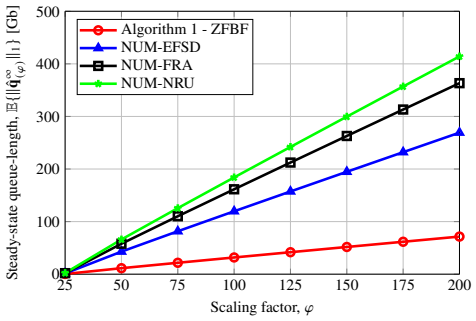


Fig. 5: The steady-state total queue-length with respect to φ.

have developed an intelligent resource management algorithm to efficiently and dynamically guide traffic to appropriate RUs by jointly optimizing the flow-split distribution, congestion control and scheduling. The proposed algorithm is proven to achieve rapid convergence, long-term utility-optimality and significantly low delay existing state-of-the-art methodologies.

References

[1] O-RAN Alliance, "ORAN Working Group 2: AI/ML workflow description and requirements," *Tech. Rep.*, Mar. 2019.

[2] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A comprehensive survey of RAN architectures toward 5G mobile communication system," *IEEE Access*, vol. 7, pp. 70 371–70 421, 2019.

[3] H. Kumar, V. Sapru, and S. K. Jaisawal, "O-RAN based proactive ANR optimization," in *IEEE Global Commun. Conf. Workshops (IEEE GLOBECOM Wkshps.)*, 2020, pp. 1–4.

[4] T. Pamuklu, M. Erol-Kantarci, and C. Ersoy, "Reinforcement learning based dynamic function splitting in disaggregated green open RANs," in *IEEE Inter. Conf. Commun. (IEEE ICC 2021)*, 2021, pp. 1–6.

[5] O-RAN.WG2.Use-Case-Requirements-v02.01, "Non-RT RIC & A1 interface: Use cases and requirements," *Technical Specification*, Nov. 2021. [Online]. Available: https://www.o-ran.org/specifications (accessed on 10 November 2021)

[6] M. R. Anwar, S. Wang, M. F. Akram, S. Raza, and S. Mahmood, "5G-enabled MEC: A distributed traffic steering for seamless service migration of internet of vehicles," *IEEE Internet of Things J.*, vol. 9, no. 1, pp. 648–661, 2022.

[7] F. Kavehmadavani, V.-D. Nguyen, T. X. Vu, and S. Chatzinotas, "Traffic steering for eMBB and uRLLC coexistence in open radio access networks," in *IEEE Inter. Conf. Commun. Workshops (IEEE ICC Wkshps.)*, 2022, pp. 242–247.

[8] A. Eryilmaz and R. Srikant, "Joint congestion control, routing, and MAC for stability and fairness in wireless networks," *IEEE J. Sel. Areas in Commun.*, vol. 24, no. 8, pp. 1514–1524, 2006.

[9] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, University of Minnesota, 2014.

[10] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.

[11] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures Commun. Netw.*, vol. 3, no. 1, pp. 1–211, 2010.

[12] T. K. Vu, M. Bennis, M. Debbah, and M. Latva-Aho, "Joint path selection and rate allocation framework for 5G self-backhauled mm-wave networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2431–2445, 2019.

[13] M. Bennis, S. M. Perlaza, P. Blasco, Z. Han, and H. V. Poor, "Self-organization in small cell networks: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3202–3212, 2013.

[14] S. Samarakoon *et al.*, "Backhaul-aware interference management in the uplink of wireless small cell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5813–5825, 2013.

[15] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, 2006.

[16] E. Stai and S. Papavassiliou, "User optimal throughput-delay trade-off in multihop networks under NUM framework," *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 1999–2002, 2014.