Network-Aided Intelligent Traffic Steering in 6G ORAN: A Multi-Layer Optimization Framework

Van-Dinh Nguyen, Thang X. Vu, Nhan Thanh Nguyen, Dinh C. Nguyen, Markku Juntti,

Nguyen Cong Luong, Dinh Thai Hoang, Diep N. Nguyen and Symeon Chatzinotas

Abstract

To enable an intelligent, programmable and multi-vendor radio access network (RAN) for 6G networks, considerable efforts have been made in standardization and development of open RAN (ORAN). So far, however, the applicability of ORAN in controlling and optimizing RAN functions has not been widely investigated. In this paper, we jointly optimize the flow-split distribution, congestion control and scheduling (JFCS) to enable an intelligent traffic steering application in ORAN. Combining tools from network utility maximization and stochastic optimization, we introduce a multi-layer optimization framework that provides fast convergence, long-term utility-optimality and significant delay reduction compared to the state-of-the-art and baseline RAN approaches. Our main contributions are three-fold: i) we propose the novel JFCS framework to efficiently and adaptively direct traffic to appropriate radio units; ii) we develop low-complexity algorithms based on the reinforcement learning, inner approximation and bisection search methods to effectively solve the JFCS problem in different time scales; and iii) the rigorous theoretical performance results are analyzed to show that there exists a scaling factor to improve the tradeoff between delay and utility-optimization. Collectively, the insights in this work will open the door towards fully automated networks with enhanced control and flexibility. Numerical results are provided to demonstrate the effectiveness of the proposed algorithms in terms of the convergence rate, long-term utility-optimality and delay reduction.

Index Terms

V.-D. Nguyen is with the College of Engineering and Computer Science, VinUniversity, Vinhomes Ocean Park, Hanoi 100000, Vietnam (e-mail: dinh.nv2@vinuni.edu.vn).

T. X. Vu and S. Chatzinotas are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, L-1855 Luxembourg City, Luxembourg (e-mail: {thang.vu, symeon.chatzinotas}@uni.lu).

N. T. Nguyen and M. Juntti are with Centre for Wireless Communications, University of Oulu, P.O.Box 4500, FI-90014, Finland, (email: {nhan.nguyen, markku.juntti}@oulu.fi).

N. C. Luong is with the Faculty of Computer Science, PHENIKAA University, Hanoi 12116, Vietnam (e-mail: luong.nguyencong@phenikaa-uni.edu.vn).

Dinh C. Nguyen is with the Elmore Family School of Electrical and Computer Engineering, Purdue University, USA (e-mail: nguye772@purdue.edu).

D. T. Hoang and D. N. Nguyen are with the School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: {hoang.dinh, diep.nguyen}@uts.edu.au)
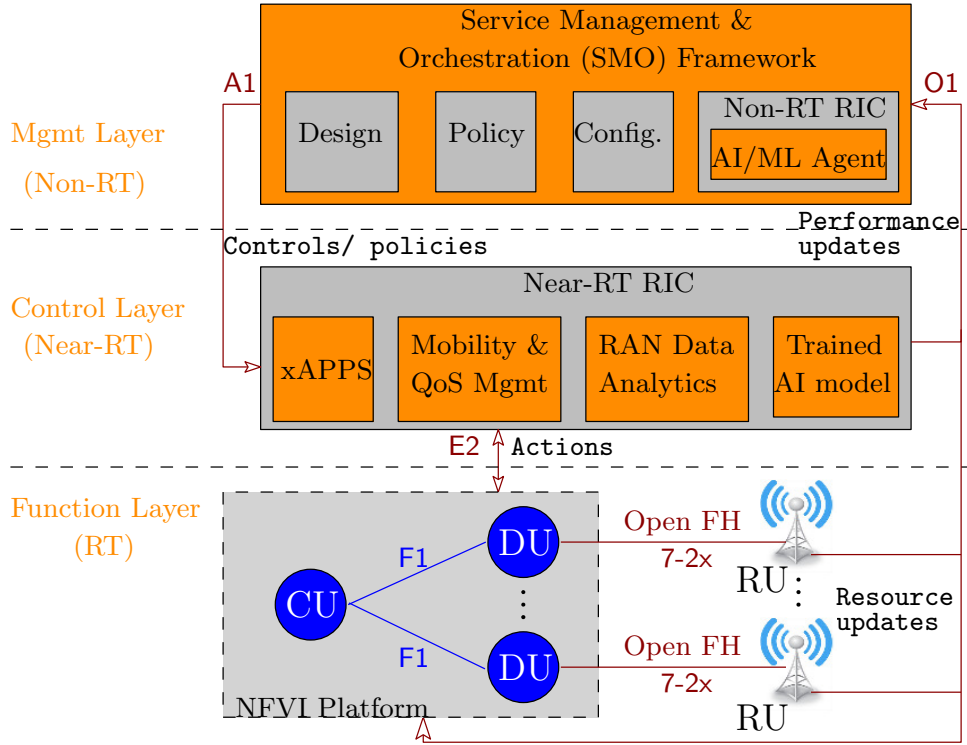
Fig. 1: ORAN Alliance reference architecture and workflow [3], with Non-RT and Near-RT RICs. A base station is disaggregated to CU, DU and RU.

Open radio access network, intelligent resource management, traffic steering, reinforcement learning, resource sharing.

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 [1]. As we move towards 6G, the latest frontier in this endeavor is open radio access network (ORAN) by disaggregating RAN components and opening up interfaces, which is considered today the most promising approach to revolutionize the wireless technology from "*connected things*" to "*connected intelligence*" [2]–[4]. ORAN is expected to fully enable programmable, intelligent, interoperable and multi-vendor RAN [5].

Fig. 1 illustrates the high-level ORAN Alliance reference architecture [3], where the "blue" parts and interfaces are defined by the 3rd Generation Partnership Project (3GPP), while the "orange" parts and interfaces are defined by ORAN Alliance. ORAN initiatives were developed to split the RAN into the radio unit (RU), distributed unit (DU) and centralized unit (CU), allowing for the interoperability

of open hardware (HW), software (SW) and interfaces (*e.g.* O1, A1 and E2) [3], [4]. The ORAN architecture typically has three main layers (or loops), including the management, control and function layers as illustrated in Fig. 1. In particular, the management layer takes place in non-real-time (Non-RT) over 1 s (second) with orchestration, automation functions and trained artificial intelligence (AI) and machine learning (ML) models. The control layer is executed in near real-time (Near-RT) between 10 ms and 1 s to provide functions like radio resource management (RRM), quality-of-service (QoS) management and interference management. Finally, the function layer provides the RAN optimization of a timescale below 10 ms, such as scheduling, power control and radio-frequency assignment, etc. Two important parts introduced in ORAN are Non-RT RAN intelligent controller (Non-RT RIC) and Near-RT RIC that allow to access RRM functions. The former enables AI/ML workflow for RAN components and RRM like traffic steering (TS) as well as policy-based guidance of applications in Near-RT RIC, while the latter is embedded with the control/optimization algorithms of RAN and radio resources [5]–[8].

A. Motivation

Yet the existing research efforts on ORAN in the academic community are isolated, providing only tailored solutions to problems at either the physical or higher layers [9]–[12]. The understanding of how ORAN could help improve network performance by controlling data traffic and optimizing RAN functions remains rather limited in the literature. In this paper, we aim to fill this gap by conducting an in-depth analysis of the multi-layer design between the physical and higher layers and developing low-complexity algorithms for network control, scheduling and resource allocation in different time scales as well as analyzing their impact on throughput and delay performances in the 6G ORAN context.

In light of the above discussions, this paper focuses on designing the TS control to intelligently direct the user traffic through a group of RUs, taking into account available resources and users' service requirements. To fully realize the potential performance of the TS scheme, ORAN allows customization of user-centric strategies, multi-path routing and multi-connectivity as well as proactive optimization of network parameters through RICs. However, the problem becomes more challenging in the ORAN setting due to several complicating factors: i) the traffic demand of user equipements (UEs) often varies over time, and the complete information of the RAN layer is indeterminate at the time of optimization algorithm execution. Hence, the policies and control decisions at the service management and orchestration (SMO) must be adapted to the variation of data traffic; ii) the total data traffic is distributed unevenly to RUs due to different downlink (DL) throughput capabilities, causing high queueing delay; and iii) the strong correlation between congestion control and scheduling optimization influences the optimal choice of flow-split distribution of data traffic across all RUs. In

addition, the deployment of fully automated networks is an intricate problem in ORAN that calls for intelligent, scalable and self-organizing strategies for a holistic multi-layer optimization framework. In this regard, reinforcement learning (RL) plays an important role in achieving long-term utility optimization. *To the best of our knowledge, the TS optimization problem for ORAN as outlined above has not been thoroughly addressed in the literature.*

B. Main Contributions

In this paper, we consider a practical scenario where 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 and address the challenges above, 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 three key contributions:

- We propose a novel JFCS framework to efficiently and adaptively direct traffic to appropriate RUs. Our framework not only generalizes the classical queue-length-based congestion control and scheduling (QCS) method [13], but also provides a synergy between RL, QCS and updated network state information, and thus enabling a closed-loop control of the TS in the ORAN context.

- To ensure the practicality and scalability, we identify inherent properties of the JFCS problem and propose an intelligent resource management algorithm to solve it effectively by leveraging the stochastic optimization framework [14]. In particular, by exploiting the historical system information accumulated from the previous time-slots, an RL process is developed to build the smoothed best response while maximizing the long-term utility for each data-flow under arbitrary changes in traffic demands. Given the updated queue-length vector and the optimal flow-split distribution, two low-complexity algorithms are developed to effectively solve the short-term power control optimization subproblem in an iterative fashion.

- Given a scaling factor φ to minimize the Lyapunov drift [15], the theoretical performance results are analyzed to show that the queueing network is stable. In addition, the expected divergence in queue-length and the optimality gap of congestion control rate still scale as $\mathcal{O}(\sqrt{\varphi})$ and $\mathcal{O}(1/\sqrt{\varphi})$, respectively. Thus, there always exists a scaling factor to balance utility-optimality and latency.

We numerically evaluate the performance of the proposed framework. Results show that the proposed

framework can improve network resource utilization significantly while achieving fast convergence and long-term utility-optimality, compared to state-of-the-art approaches.

C. Paper Organization and Mathematical Notation

The remainder of this paper is organized as follows. The related work is discussed in Section II. In Section III, we first introduce the network model and then present the problem formulation. The proposed JFCS framework and its solutions are provided in Sections IV and V, respectively. Section VI presents the key theoretical performance results of the JFCS framework. Numerical results are given in Section VII, while Section VIII concludes the paper.

Mathematical notation: Throughout this paper, matrices and vectors are written as bold uppercase and lowercase letters, respectively, while scalar number is denoted in lowercase. $\mathbf{h}^{\mathsf{H}}$ is the Hermitian transpose of vector $\mathbf{h}$. The notation $x \sim \mathcal{CN}(0, \sigma^2)$ implies that x is circularly-symmetric complex Gaussian random variable with zero mean and variance σ^2. $\| \cdot \|$ stands for the vector's Euclidean norm. $\mathbb{C}$ and $\mathbb{R}$ denote the sets of all complex and real numbers, respectively. Finally, $\mathbb{E}\{\cdot\}$ denotes the expectation of a random variable.

II. RELATED WORK

Multi-layer (a.k.a. cross-layer) optimization for traditional cellular RAN architectures has been extensively studied in the literature (see *e.g.*, [16] and references therein). For example, Tang *et al.* [17] studied a multi-layer resource allocation problem to minimize the overall system power consumption in a cloud-RAN (C-RAN), which jointly optimizes the service scaling, remote radio head selection, and beamforming. In [18], a joint design of virtual computing and radio resource allocation was proposed. It was shown that this approach can efficiently allocate the virtual computing of the baseband unit (BBU) pool to achieve load balancing among users with the significantly reduced power consumption. These problems are often solved by the difference of convex algorithm due to the combinatorial nature and strong coupling between optimization variables. To address this challenge, graph theory techniques were introduced in [19] and [20] to effectively solve the joint coordinated scheduling and power optimization problem in C-RAN. Recently, the multi-layer network coding was also investigated in [21]–[23], taking into account the rate heterogeneity of different users to remote radio heads. In general, these existing works only optimized radio resources, while other factors at higher layers (*e.g.* congestion control and routing) were overlooked, making guaranteed multi-layer QoS for ORAN 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 ORAN architecture. Kumar *et al.* [9] proposed an automatic neighbour relation (ANR) approach to manage neighbour cell relationships by leveraging ML techniques, hence improving gNodeB (gNB) handovers. The

work in [24] introduced an intelligent user access control algorithm based on deep reinforcement learning, aiming to maximize the overall throughput and avoid frequent handovers. The authors in [10] developed an RL-based dynamic function splitting which is shown to be able to effectively decide the ORAN's function splits and reduce operating costs. Based on the Working Group (WG)-2 AI/ML specifications of the O-RAN Alliance, Acumos framework and open network automation platform were introduced in [11] to generate AI/ML models to be deployed in RIC modules and monitor the designed workflow, respectively. However, all these studies do not reveal any observable information about the RAN layer to SMO via periodic feedback loops. Thus, RICs in these studies are unable to monitor RAN in a timely manner and allow for their management automation within ORAN.

In traditional RAN architectures, the TS solutions are typically determined by users' radio conditions of a serving cell while treating signals from neighboring cells as interference [25]. The authors in [26] 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.* [27] 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-RT RIC and assumes that all network information are available at Near-RT RIC to optimize radio resource allocation.

Different from all above works and others in the literature that focus on a single layer, 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. This holistic multi-layer optimization framework guarantees the long-term utility-optimality with far less latency than state-of-the-art approaches, opening the door towards fully automated networks with enhanced control and flexibility.

III. Network Model and Problem Formulation

A. Network Model

As shown in Fig. 2, we consider an ORAN architecture with one CU, I DUs and J RUs, where each DU connects to multiple RUs for cost-effective deployment. Let us denote by $\mathcal{I} \triangleq \{1, 2, \cdots, I\}$ the set of DUs. 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 total number of RUs' antennas is thus $M_\Sigma = \sum_{\forall(i,j)} M_{i,j}$. 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 assume that the midhaul (MH) link between the CU and DU and
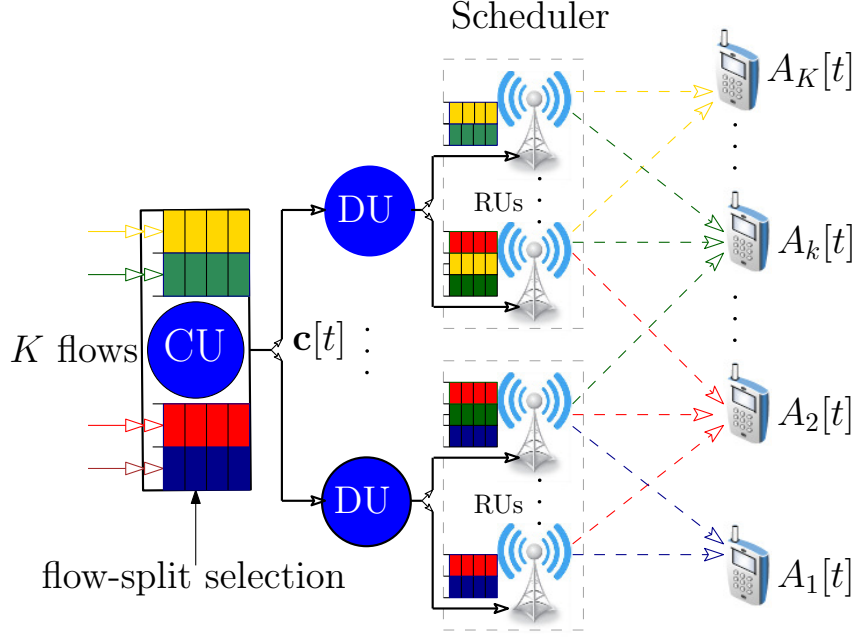
Fig. 2: Illustration of the ORAN-based system model enabling TS where each DU connects to multiple RUs towards cost-effective deployment.
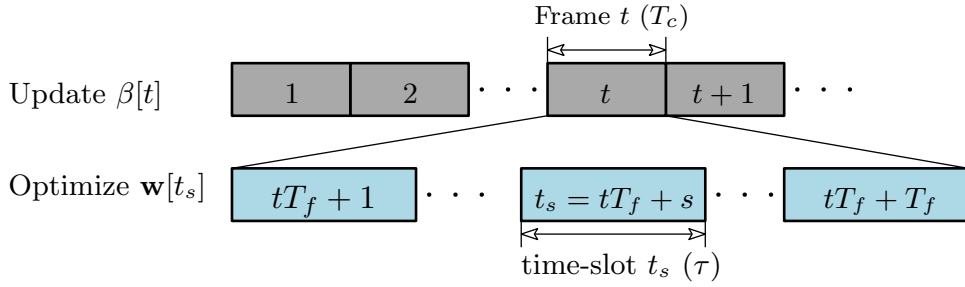

Fig. 3: Illustration of frame structure with each time-frame t (corresponding to one large-scale coherence time) consisting of T_f time-slots.

fronthaul link between the DU and RU have sufficient capacity (*i.e.*, high-speed optical ones), so that the transmission latency from CU to RUs and queueing latency at CU and DUs are negligible.

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 illustrated in Fig. 3. 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 $s \in \{1, 2, \cdots, T_f\}$. At CU, there exists K independent data-flows, each of which is intended for one UE. The CU splits the 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 [28], [29], so-called "traffic steering". For 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. To improve the system throughput, 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 $\mathscr{B}[t] \triangleq \{\boldsymbol{\beta}_k[t], \forall k | \sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t] = 1, \forall k\}$, where each column flow-split vector $\boldsymbol{\beta}_k[t] \triangleq \left[\beta_k^{i,j}[t]\right]_{(i,j)\in\mathcal{P}_k}^{\mathsf{T}} \in \mathbb{R}^J$ corresponds to the flow-split vector of data-flow k.

1) Wireless Channel Model and Downlink Throughput

The large-scale fading coefficients are assumed to be invariant within one frame T_c, while the small-scale fading components with a low degree of mobility are assumed to be unchanged during time-slot t_s with duration of τ and vary independently in the next time-slot. For example, the large-scale fading coefficients may stay invariant for a period of at least 40 small-scale fading coherence intervals for indoor scenarios [30]. 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 $\mathbf{h}_k^{i,j}[t_s] = \sqrt{\xi_k^{i,j}[t]}\left(\sqrt{\kappa_k^{i,j}[t]/(\kappa_k^{i,j}[t]+1)}\bar{\mathbf{h}}_k^{i,j}[t] + \sqrt{1/(\kappa_k^{i,j}[t]+1)}\tilde{\mathbf{h}}_k^{i,j}[t_s]\right)$ where $\xi_k^{i,j}[t]$ represents the large-scale fading; $\bar{\mathbf{h}}_k^{i,j}[t]$ and $\tilde{\mathbf{h}}_k^{i,j}[t_s] \sim \mathcal{CN}(0,\mathbf{I})$ are the line-of-sight (LoS) and non-LoS (NLoS) components, which follow a deterministic channel and Rayleigh fading models, respectively. We let $\mathbf{H}[t_s] \triangleq \left[\mathbf{h}_1[t_s]\cdots\mathbf{h}_K[t_s]\right] \in \mathbb{C}^{M\times K}$ denote the channel matrix between all RUs and UEs in time-slot t_s where $\mathbf{h}_k[t_s] \triangleq \left[(\mathbf{h}_k^{i,j}[t_s])^{\mathsf{H}}\right]_{\forall i,j}^{\mathsf{H}} \in \mathbb{C}^{M\times 1}$ corresponds to the channel vector between RUs and UE k.

Let us denote 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 in time-slot t_s, respectively. The received signal at UE k in time-slot t_s can be written as

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

where $\omega_k[t_s]$ is the additive white Gaussian background noise (AWGN) with power N_0. The downlink achievable rate (bits/s) of UE k from RU (i,j) in time-slot t_s can be written as $r_k^{i,j}(\mathbf{w}[t_s]) \triangleq W \log_2\left(1 + \gamma_k^{i,j}(\mathbf{w}[t_s])\right)$, where W is the system bandwidth and the signal-to-interference-plus-noise ratio (SINR) $\gamma_k^{i,j}(\mathbf{w}[t_s])$ is given by $\gamma_k^{i,j}(\mathbf{w}[t_s]) = |(\mathbf{h}_k^{i,j}[t_s])^{\mathsf{H}}\mathbf{w}_k^{i,j}[t_s]|^2/\Phi_k^{i,j}(\mathbf{w}[t_s])$ with

$$
\Phi_k^{i,j}(\mathbf{w}[t_s]) \triangleq \underbrace{\sum_{(i',j')\in\mathcal{P}_k\setminus\{(i,j)\}} |(\mathbf{h}_k^{i',j'}[t_s])^{\mathsf{H}}\mathbf{w}_k^{i',j'}[t_s]|^2}_{\text{Intra-user interference}}
$$

$$+ \underbrace{\sum_{k' \in \mathcal{K} \setminus \{k\}} \sum_{(i,j) \in \mathcal{P}_{k'}} |(\mathbf{h}_k^{i,j}[t_s])^{\mathsf{H}} \mathbf{w}_{k'}^{i,j}[t_s]|^2}_{\text{Inter-user interference}} + N_0 \tag{2}$$

and $\mathbf{w}[t_s] \triangleq \big[(\mathbf{w}_k^{i,j}[t_s])^{\mathsf{H}}\big]_{k \in \mathcal{K}, (i,j) \in \mathcal{P}_k}^{\mathsf{H}}$ being the vector embracing all the beamformers. The overall effective data rate of data-flow k (or UE k) can be computed as $r_k(\mathbf{w}[t_s]) = \sum_{(i,j) \in \mathcal{P}_k} r_k^{i,j}(\mathbf{w}[t_s])$. Then, for each $\mathbf{H}[t_s]$ and a given $\boldsymbol{\beta}_k[t]$, we define the *instantaneous achievable rate region* under beamformer $\mathbf{w}[t_s]$ as

$$\mathscr{C}_{\mathbf{H}[t_s]} \triangleq \left\{ r_k(\mathbf{w}[t_s]), \forall k \left| \begin{array}{l} r_k(\mathbf{w}[t_s]) = \sum\limits_{(i,j) \in \mathcal{P}_k} r_k^{i,j}(\mathbf{w}[t_s]) \\ \sum\limits_{k \in \mathcal{K}} \|\mathbf{w}_k^{i,j}[t_s]\|_2^2 \leq P_{\max}^{i,j}, \forall(i,j) \end{array} \right. \right\} \tag{3}$$

where $P_{\max}^{i,j}$ denotes the transmit power budget of RU (i,j). We note that the achievable rate $r_k^{i,j}(\mathbf{w}[t_s])$ is upper bounded by $r_k^{i,j}(\mathbf{w}[t_s]) \leq W \log_2\big(1 + P_{\max}^{i,j} \|\mathbf{h}_k^{i,j}[t_s]\|_2^2 / N_0\big)$ for a limited transmit power budget $P_{\max}^{i,j}$, leading to $r_k(\mathbf{w}[t_s]) < \infty, \forall k, t$.

2) Queueing Model

As illustrated in Fig. 2, each RU maintains a separate queue for each UE. Let $A_k[t]$ (bits/s) be the total rate of data arriving at RU 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 $A_k[t] \leq A^{\max} < \infty, \forall k, t$, 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: $q_k^{i,j}[t_{s+1}] = \Big[q_k^{i,j}[t_s] + \beta_k^{i,j}[t]A_k[t]\tau - r_k^{i,j}(\mathbf{w}[t_s])\tau\Big]^+$, where $[x]^+ \triangleq \max\{0, x\}$. By $\mathbf{q}[t_s] \triangleq \big[q_k^{i,j}[t_s]\big]_{k,(i,j)}^{\mathsf{T}}$ and following [15], a queueing network is *stable* if the steady-state total queue-length remains finite, such as

$$\limsup_{t_s \to \infty} \mathbb{E}\{\|\mathbf{q}[t_s]\|_1\} < \infty. \tag{4}$$

B. Problem Formulation

Let $\bar{r}_k \triangleq \lim\limits_{t_s \to \infty} \frac{1}{t_s} \sum_{\ell=1}^{t_s} r_k(\mathbf{w}[\ell])$ denote the long-term average rate of data-flow k. Each UE k is associated with a utility function, denoted by $U_k(\bar{r}_k)$. To facilitate the analysis presented later, we make the following assumption to the utility function [15], [28], [31], [32].

Assumption 1. *The utility function $U_k(\cdot)$ is assumed to satisfy the following conditions*

- $U_k(\cdot)$ *is twice continuously differentiable, increasing, and strictly concave.*
- *There exist positive constants $0 < \psi < \Psi < \infty$, such as $\psi \leq -U_k''(\bar{r}_k) \leq \Psi, \forall \bar{r}_k \in [0, \bar{r}^{\max}]$, with $\bar{r}^{\max}$ being the maximum long-term average rate of any data flow.*

Our goal is to maximize the network utility function $\sum_{k \in \mathcal{K}} U_k(\bar{r}_k)$, subject to the probabilistic

delay constraint, achievable rate region and queue-stability constraint. Based on the network utility maximization (NUM) framework, the joint flow-split distribution, congestion control and scheduling optimization problem (JFCS) can be mathematically formulated as

$$\textbf{JFCS}: \max_{\boldsymbol{\beta}, \bar{\mathbf{r}}, \mathbf{w}} \sum_{k \in \mathcal{K}} U_k(\bar{r}_k) \tag{5a}$$

$$\text{s.t.} \limsup_{t_s \to \infty} \mathbb{E}\{\|\mathbf{q}[t_s]\|_1\} < \infty \tag{5b}$$

$$r_k(\mathbf{w}[t_s]) \in \mathscr{C}_{\mathbf{H}[t_s]}, \forall t_s, k \in \mathcal{K} \tag{5c}$$

$$\boldsymbol{\beta}_k[t] \in \mathscr{B}[t], \forall t, k \in \mathcal{K} \tag{5d}$$

$$\mathsf{Prob}\Big(\frac{q_k^{i,j}[t_s]}{\bar{A}_k} \le \bar{d}_k\Big) \ge \epsilon_k, \ \forall t_s, k, (i,j) \tag{5e}$$

where $\boldsymbol{\beta} \triangleq \big[\boldsymbol{\beta}_k^\mathsf{T}\big]_{k \in \mathcal{K}}^\mathsf{T}$ and $\bar{\mathbf{r}} \triangleq \big[\bar{r}_k\big]_{k \in \mathcal{K}}^\mathsf{T}$. Constraint (5e) 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. It is stated that the probability of UEs' maximum allowable delay should be greater than or equal to a certain positive constant ϵ_k.

Remark 1. *It is clear that problem* (5) *needs to be executed in different time scales (i.e., over the long-term scale t at Non-RT RIC and the short-term scale t_s at Near-RT RIC), as shown in Fig. 3. In particular, the global flow-split vector $\boldsymbol{\beta}[t]$ is only updated once per time-frame t to reduce computational complexity, information exchange and to ensure a stable queueing system. On the other hand, the beamforming vector $\mathbf{w}[t_s]$ and the instantaneous achievable rate $\mathbf{r}[t_s]$ are optimized based on the real-time effective CSI $\mathbf{H}[t_s]$ in time-slot t_s, adapting to dynamic environments.*

IV. JFCS-BASED NETWORK UTILITY OPTIMIZATION

A. Tractable Form of the JFCS Problem (5)

Challenges of Solving JFCS Problem (5): We can observe that that constraint (5c) is nonconvex while (5e) is a nonconvex probabilistic constraint, generally making problem (5) 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) [33], 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. The flow-split decisions mainly rely on the previous states updated by the RAN layer. Towards practical applications, an efficient and adaptive solution to the long-term subproblem of (5) is necessary to achieve high QoE for all UEs in every TTI.

Let us start by transforming problem (5) into a more tractable form. Towards a safe design, we

consider the replacement of constraint (5e) by its deterministic constraint. From the basic property of probability, we can rewrite (5e) as $\mathsf{Prob}\big(q_k^{i,j}[t_s] \geq \bar{A}_k \bar{d}_k\big) \leq 1 - \epsilon_k$. It follows from the well-known Markov inequality [34] that $\mathsf{Prob}\big(q_k^{i,j}[t_s] \geq \bar{A}_k \bar{d}_k\big) \leq \mathbb{E}\{q_k^{i,j}[t_s]\}/\bar{A}_k \bar{d}_k$, yielding

$$\sum_{\ell=1}^{t} \beta_k^{i,j}[\ell]\bar{A}_k \tau - (1 - \epsilon_k)\bar{A}_k \bar{d}_k - \sum_{\ell=1}^{t_{s-1}} r_k^{i,j}(\mathbf{w}[\ell])\tau$$
$$\leq r_k^{i,j}(\mathbf{w}[t_s])\tau, \ \forall t_s, k \in \mathcal{K}, (i,j) \in \mathcal{P}_k \tag{6}$$

where each queue-length is always non-negative. We note that (6) is a relaxed constraint of (5e), which implies that any feasible of the former is also feasible for the latter but not vice versa due to the Markov upper bound on the outage probabilities.

To facilitate the following optimization, we introduce congestion control variables $\boldsymbol{a}[t_s] \triangleq \big[a_k[t_s]\big]_{k \in \mathcal{K}}^{\mathsf{T}}$, satisfying $\bar{a}_k - \bar{r}_k \leq 0, \forall k$, where $\bar{a}_k \triangleq \lim_{t_s \to \infty} \frac{1}{t_s} \sum_{\ell=1}^{t_s} a_k[\ell]$. Problem (5) is then rewritten as

$$\max_{\boldsymbol{\beta}, \bar{\boldsymbol{a}}, \bar{\mathbf{r}}, \mathbf{w}} \sum_{k \in \mathcal{K}} U_k(\bar{a}_k) \tag{7a}$$

$$\text{s.t.} \quad (5b), (5c), (5d), (6) \tag{7b}$$

$$\bar{a}_k - \bar{r}_k \leq 0, \forall k. \tag{7c}$$

We also introduce a new auxiliary queue-length vector $\hat{\mathbf{q}}[t_s] \triangleq \big[\hat{q}_k[t_s]\big]_{k \in \mathcal{K}}^{\mathsf{T}}$, where $\hat{q}_k[t_{s+1}] = \big[\hat{q}_k[t_s] + a_k[t_s]\tau - r_k(\mathbf{w}[t_s])\tau\big]^{+}$ to associate constraint (7c) 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 $L[t_s] = \frac{1}{2}\big(\sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{P}_k} \frac{q_k^{i,j}[t_s]^2}{\tau^2} + \sum_{k \in \mathcal{K}} \frac{\hat{q}_k[t_s]^2}{\tau^2}\big)$, which is the quadratic Lyapunov function [14], [35]. 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 (*i.e.*, (5b) and (7c) hold true), we adopt the drift-plus-penalty procedure [14] to minimize the drift of a quadratic Lyapunov function and rewrite (7) as

$$\max_{\boldsymbol{\beta}, \bar{\boldsymbol{a}}, \bar{\mathbf{r}}, \mathbf{w}} \quad \varphi \sum_{k \in \mathcal{K}} \mathbb{E}\{U_k(a_k[t_s])\} - \mathbb{E}\{\Delta L[t_s]\} \tag{8a}$$

$$\text{s.t.} \quad (5c), (5d), (6) \tag{8b}$$

where φ is a scaling factor to balance two objective functions. We now show that constraint (7c) holds with equality at optimum by introducing the following lemma.

Lemma 1. *For each data-flow of UE k, the optimal congestion control rate is equal to the optimal long-term average service rate, i.e., $\bar{a}_k^* - \bar{r}_k^* = 0, \forall k$.*

The proof Lemma 1 is straightforward by examining the Karush–Kuhn–Tucker (KKT) complementary slackness condition over the increasing and strictly concave objective function $U_k(\cdot), \forall k$.

Algorithm 1: Intelligent Resource Management Algorithm for Solving JFCS Problem (5), compliant with ORAN

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 the queue-length vector $\hat{\mathbf{q}}[t_s]$, solve S-SP1 (12) 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 the queue-length vector $\hat{\mathbf{q}}[t_s]$ and the flow-split distribution $\boldsymbol{\beta}^*[t]$, each RU $(i,j) \in \mathcal{P}_k$ schedules the service rate $r_k^{i,j}(\mathbf{w}[t_s])$ for UE $k \in \mathcal{K}$ by solving S-SP2:

$$\max_{\mathbf{r}[t_s], \mathbf{w}[t_s]} \sum_{k \in \mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k(\mathbf{w}[t_s]), \text{ s.t. } (5c), (6).$$

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

$$q_k^{i,j}[t_{s+1}] = \big[q_k^{i,j}[t_s] + \beta_k^{i,j}[t]A_k[t]\tau$$
$$- r_k^{i,j}(\mathbf{w}[t_s])\tau\big]^+, \ \forall k, (i,j)$$
$$\hat{q}_k[t_{s+1}] = \big[\hat{q}_k[t_s] + a_k[t_s]\tau - r_k(\mathbf{w}[t_s])\tau\big]^+, \ \forall k.$$

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**

B. Overall Intelligent Resource Management Algorithm

To solve problem (8) in different time scales, we now decompose it into three subproblems. To do so, we consider a worst-case design by developing an upper bound of $\Delta L[t_s]$ for given $(\mathbf{q}[t_s], \hat{\mathbf{q}}[t_s])$. From the inequality $([x]^+)^2 \leq x^2$ and $(x+y)^2 - x^2 = 2xy + y^2$, we have

$$\Delta L^{\text{UB}}[t_s] \triangleq \sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{P}_k} \frac{q_k^{i,j}[t_s]}{\tau}\big(\beta_k^{i,j}[t]A_k[t] - r_k^{i,j}(\mathbf{w}[t_s])\big)$$

$$+ \sum_{k \in \mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau}\big(a_k[t_s] - r_k(\mathbf{w}[t_s])\big) + B[t_s] \geq \Delta L[t_s] \qquad (9)$$

where $B[t_s] \triangleq \frac{1}{2}\sum_{k \in \mathcal{K}}\sum_{(i,j) \in \mathcal{P}_k}\big(\beta_k^{i,j}[t]A_k[t] - r_k^{i,j}(\mathbf{w}[t_s])\big)^2 + \frac{1}{2}\sum_{k \in \mathcal{K}}\big(a_k[t_s] - r_k(\mathbf{w}[t_s])\big)^2$ is the summation of the second moments of the arrival and service processes. Following [14] and [28], we

consider that $B[t_s]$ is finite and bounded by $\bar{B}$ for all t_s, *i.e.*, $\mathbb{E}\{B[t_s]\big|\mathbf{q}[t_s],\hat{\mathbf{q}}[t_s]\} \leq \bar{B}$. As a result, problem (8) is simplified to

$$\max_{\boldsymbol{\beta},\bar{\boldsymbol{a}},\bar{\mathbf{r}},\mathbf{w}} \quad \varphi \sum_{k\in\mathcal{K}} \mathbb{E}\{U_k(a_k[t_s])\} - \mathbb{E}\{\Delta L^{\text{UB}}[t_s]\} \tag{10a}$$

$$\text{s.t.} \quad (5\text{c}), (5\text{d}), (6). \tag{10b}$$

Long-term subproblem (L-SP): The flow-split distribution subproblem at time-frame t is given as

$$\textbf{L-SP}: \max_{\boldsymbol{\beta}_k[t]\in\mathscr{B}[t],\forall k} \quad \sum_{k\in\mathcal{K}} \mathscr{L}_k[t] \tag{11}$$

where $\mathscr{L}_k[t] = \sum_{(i,j)\in\mathcal{P}_k} \frac{q_k^{i,j}[t_s]}{\tau}\big(r_k^{i,j}(\mathbf{w}[t_s]) - \beta_k^{i,j}[t]A_k[t]\big)$. Although problem (11) 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

$$\textbf{S-SP1}: \max_{\boldsymbol{a}[t_s]\geq 0} \quad \sum_{k\in\mathcal{K}} \big(\varphi U_k(a_k[t_s]) - \frac{\hat{q}_k[t_s]}{\tau}a_k[t_s]\big) \tag{12}$$

which is an unconstrained convex problem. The optimal solution of (12) 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

$$\textbf{S-SP2}: \max_{\mathbf{r}[t_s],\mathbf{w}[t_s]} \quad \sum_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau}r_k(\mathbf{w}[t_s]), \text{ s.t. } (5\text{c}),(6). \tag{13}$$

The overall intelligent resource management algorithm for solving the JFCS problem (5) is summarized in Algorithm 1, where the solutions of subproblems will be provided next.

V. PROPOSED ALGORITHMS FOR SOLVING SUBPROBLEMS

We are now in position to solve L-SP (11) and S-SP2 (13) in different time scales. The optimality of the latter depends heavily on the optimal flow-split decisions, which often require a prior knowledge of the statistical information of all possible paths at Non-RT RIC. However, the assumption of complete information is unrealistic due to the dynamic environment and the data collected from the RAN layer being only updated to Non-RT RIC only on the long-term scale. In this work, at time-frame t we aim to exploit historical system information accumulated from the previous time-slot, which can be used to build the smoothed best response in maximizing the long-term utility for each data flow.

A. Reinforcement Learning Algorithm for Solving L-SP (11)

The flow-split decision $\boldsymbol{\beta}_k[t]$ in problem (11) can be estimated separably by minimizing $\mathscr{L}_k[t]$. This implies that the larger the queue-length $q_k^{i,j}[t_s]$, the lower the flow-split decision value $\beta_k^{i,j}[t]$ to guarantee fairness among all RUs $(i,j) \in \mathcal{P}_k$ (*i.e.*, to avoid large queue-lengths $q_k^{i,j}$ at some RUs in the next time-slot t_{s+1}). Let us denote by $u_k^{i,j}[t] \triangleq \frac{q_k^{i,j}[t_s]}{\tau}\big(r_k^{i,j}(\mathbf{w}[t_s]) - \beta_k^{i,j}[t]A_k[t]\big)$ 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

$$u_k[t] = \sum_{(i,j)\in\mathcal{P}_k} u_k^{i,j}[t]. \tag{14}$$

However, it is unable to build a smoothed best response based on $u_k^{i,j}[t]$ as it is not revealed at the beginning of time-frame t. Inspired by [36], 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:

$$\hat{u}_k^{i,j}[t] = \hat{u}_k^{i,j}[t-1] + \eta_u[t]\mathbb{1}_{\{\mathbf{c}_k[t]=c_k^{i,j}[t]\}}\big(\bar{u}_k[t] - \hat{u}_k^{i,j}[t-1]\big), \ \forall t > 1 \tag{15}$$

where $\eta_u > 0$ is the decreasing step size (learning rate), which is often decreased over time. Naturally, $\hat{u}_k^{i,j}[1]$ is initialized as $\hat{u}_k^{i,j}[1] = 0$ for $t = 1$. The indicator function $\mathbb{1}_{\{x=y\}} = 1$ (resp. 0) if the condition $x = y$ is true (resp. false).

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 for action $\mathbf{c}_k[t] = c_k^{i,j}[t]$ as

$$\hat{\theta}_k^{i,j}[t] = \hat{\theta}_k^{i,j}[t-1] + \eta_\theta[t]\mathbb{1}_{\{\mathbf{c}_k[t]=c_k^{i,j}[t]\}}\big(\bar{u}_k[t]\hat{u}_k^{i,j}[t] - \hat{\theta}_k^{i,j}[t-1]\big), \ \forall t > 1 \tag{16}$$

with $\hat{\theta}_k^{i,j}[1] = 0$ and $\eta_\theta[t]$ being the learning rate. In order to achieve high performance in the long term, the L-SP must balance exploration and exploitation processes. 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 (15) 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

$$f(\hat{\boldsymbol{\theta}}[t]) := \operatorname*{argmin}_{\boldsymbol{\beta}_k[t]\in\mathscr{B}[t]} \Big\{h\big(\boldsymbol{\beta}[t]\big) - \lambda \sum_{k\in\mathcal{K}} \sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t]\hat{\theta}_k^{i,j}[t]\Big\}. \tag{17}$$

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 (17) will dominate the best response function and then the actions associated with highest estimated regret will be selected [36].

Regularization function: The regularization function allows to learn the best paths that maximize its own performance and stabilize the flow-split decisions. The solutions of problem (11) 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 (17), we have

$$f(\hat{\boldsymbol{\theta}}[t]) := \operatorname*{argmin}_{\boldsymbol{\beta}_k[t]\in\mathscr{B}[t],\forall k} \bigg\{ \sum_{k\in\mathcal{K}} \sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t] \ln\big(\beta_k^{i,j}[t]\big) - \lambda \sum_{k\in\mathcal{K}} \sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t]\hat{\theta}_k^{i,j}[t] \bigg\}. \qquad (18)$$

The function $f(\hat{\boldsymbol{\theta}}[t])$ is convex and separable for each $\beta_k^{i,j}[t]$. By solving the following equation

$$\partial f(\hat{\boldsymbol{\theta}}[t])/\partial \beta_k^{i,j}[t] = \ln\big(\beta_k^{i,j}[t]\big) + 1 - \lambda\hat{\theta}_k^{i,j}[t] = 0$$

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 $\sum_{(i,j)\in\mathcal{P}_k} \beta_k^{i,j}[t] = 1, \forall k$ (*i.e.* the unit simplex for data-flow k), we normalize $f_k^{i,j}(\hat{\boldsymbol{\theta}}_k[t])$ through the exponentiated mirror function as

$$\begin{aligned} f_k^{i,j}(\hat{\boldsymbol{\theta}}_k[t]) &= \frac{\exp\big(\big[\lambda\hat{\theta}_k^{i,j}[t] - 1\big]^+\big)}{\sum_{(i',j')\in\mathcal{P}_k} \exp\big(\big[\lambda\hat{\theta}_k^{i',j'}[t] - 1\big]^+\big)} \\ &= \frac{\exp\big(\lambda\big[\hat{\theta}_k^{i,j}[t]\big]^+\big)}{\sum_{(i',j')\in\mathcal{P}_k} \exp\big(\lambda\big[\hat{\theta}_k^{i',j'}[t]\big]^+\big)}. \end{aligned} \qquad (19)$$

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

$$\beta_k^{i,j}[t] = \beta_k^{i,j}[t-1] + \eta_\beta[t]\big(f_k^{i,j}(\hat{\boldsymbol{\theta}}_k[t]) - \beta_k^{i,j}[t-1]\big) \qquad (20)$$

for $t > 1$, where $\boldsymbol{\beta}_k[1] = \frac{1}{|\mathcal{P}_k|}[1,\cdots,1]$ and $\eta_\beta[t]$ is the learning rate. The three-step reinforcement learning procedure includes (15), (16) and (20), which do not require expensive computations and projection to the feasible space.

Convergence properties: The convergence conditions for the three-step reinforcement learning procedure are given as follows:

$$\lim_{T\to\infty} \sum\nolimits_{t=1}^T \eta_u[t] = +\infty \;\&\; \lim_{T\to\infty} \sum\nolimits_{t=1}^T \eta_u^2[t] < +\infty$$

$$\lim_{T\to\infty} \sum\nolimits_{t=1}^T \eta_\theta[t] = +\infty \;\&\; \lim_{T\to\infty} \sum\nolimits_{t=1}^T \eta_\theta^2[t] < +\infty$$

$$\lim_{T\to\infty} \sum\nolimits_{t=1}^T \eta_\beta[t] = +\infty \;\&\; \lim_{T\to\infty} \sum\nolimits_{t=1}^T \eta_\beta^2[t] < +\infty$$

$$\lim_{t \to \infty} \frac{\eta_\theta[t]}{\eta_u[t]} = 0 \quad \& \quad \lim_{t \to \infty} \frac{\eta_\beta[t]}{\eta_\theta[t]} = 0. \tag{21}$$

This implies that the learning rates must be decreased over time to guarantee the convergence of the proposed three-step RL procedure. The detailed proof of a multiple-timescales RL algorithm can be found in [36], [37]. Following the same arguments as those in [36] and the conditions in (21), the three-step RL procedure in (15), (16) and (20) converges to the optimal solution with the positive trade-off factor $\lambda > 0$, satisfying $\lim_{t \to \infty} \beta_k[t] = \beta_k^*, \ \forall k \in \mathcal{K}$.

B. Proposed Algorithm for Solving S-SP2 (13)

Given the optimal flow-split distribution of data-flow k, $\beta_k^*[t]$, we denote by $\mathcal{P}_k^*[t]$ the set of selected path states in time-frame t, which only includes $c_k^{i,j}[t] = 1$ with $(i,j) \in \mathcal{P}_k$. In this section, we present two low-complexity transmission designs for $\mathbf{w}$, *namely* maximum ratio transmission (MRT) and zero-forcing beamforming (ZFBF), and then develop low-complexity iterative algorithms for their solution.

1) MRT-Based Transmission Design

Each RU (i,j) performs MRT beamforming (a.k.a. channel-mathched beamforming) using local CSI as $\mathbf{w}_k^{i,j}[t_s] = \frac{\sqrt{p_k^{i,j}[t_s]}}{\sqrt{\nu_k^{i,j}[t_s]}} \mathbf{h}_k^{i,j}[t_s], \ \forall (i,j) \in \mathcal{P}_k[t]$ and $k \in \mathcal{K}$, where $\nu_k^{i,j}[t_s] \triangleq \|\mathbf{h}_k^{i,j}[t_s]\|_2^2$ and $p_k^{i,j}[t_s]$ is the transmit power coefficient allocated to UE k from RU (i,j) in time-slot t_s. The corresponding SINR is rewritten as

$$\gamma_k^{i,j}(\mathbf{p}[t_s]) = \frac{p_k^{i,j}[t_s]\nu_k^{i,j}[t_s]}{\Phi_k^{i,j}(\mathbf{p}[t_s])} \tag{22}$$

where $\Phi_k^{i,j}(\mathbf{p}[t_s]) \triangleq \sum_{(i',j') \in \mathcal{P}_k[t] \setminus \{(i,j)\}} p_k^{i',j'}[t_s]\nu_k^{i',j'}[t_s] + \sum_{k' \in \mathcal{K} \setminus \{k\}} \sum_{(i,j) \in \mathcal{P}_{k'}[t]} p_{k'}^{i,j}[t_s] \frac{|(\mathbf{h}_k^{i,j}[t_s])^{\mathsf{H}} \mathbf{h}_{k'}^{i,j}[t_s]|^2}{\nu_{k'}^{i,j}[t_s]} + N_0$ is linear in $\mathbf{p}[t_s] \triangleq \left[p_k^{i,j}[t_s]\right]_{k \in \mathcal{K},(i,j) \in \mathcal{P}_k}$. As a result, the short-term power optimization problem (13) with MRT reduces to the following problem:

$$\max_{\mathbf{p}[t_s]} \quad \sum_{k \in \mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k(\mathbf{p}[t_s]) \tag{23a}$$

$$\text{s.t.} \quad \bar{R}_k^{i,j}[t_s] \leq r_k^{i,j}(\mathbf{p}[t_s])\tau, \ \forall k, (i,j) \tag{23b}$$

$$\sum_{k \in \mathcal{K}} p_k^{i,j}[t_s] \leq P_{\max}^{i,j}, \ \forall (i,j) \tag{23c}$$

where $r_k(\mathbf{p}[t_s]) = \sum_{(i,j) \in \mathcal{P}_k^*} r_k^{i,j}(\mathbf{p}[t_s])$ with $r_k^{i,j}(\mathbf{p}[t_s]) \triangleq W \log_2\left(1 + \gamma_k^{i,j}(\mathbf{p}[t_s])\right)$, and $\bar{R}_k^{i,j}[t_s] \triangleq \sum_{\ell=1}^{t} \beta_k^{i,j}[\ell]\bar{A}_k\tau - (1 - \epsilon_k)\bar{A}_k\bar{d}_k - \sum_{\ell=1}^{t_s-1} r_k^{i,j}(\mathbf{p}[\ell])\tau$.

Problem (23) is nonconvex due to the nonconcavity of $r_k^{i,j}(\mathbf{p}[t_s])$. We will now apply IA method to effectively solve (23) in an iterative manner. Following from inequality (A.2) in Appendix A with $v = p_k^{i,j}[t_s]\|\mathbf{h}_k^{i,j}[t_s]\|_2^2$ and $z = \Phi_k^{i,j}(\mathbf{p}[t_s])$, the global concave lower bound of $r_k^{i,j}(\mathbf{p}[t_s])$ at the

---

Algorithm 2: Proposed Iterative Algorithm for Solving (13) with MRT-Based Transmission Design

---

Initialization: Set $n := 1$ and generate an initial feasible value for $\mathbf{p}^{(0)}[t_s]$ to constraints in (25)
1: **repeat**
2:     Solve (25) to obtain the optimal transmission power $\mathbf{p}^*[t_s]$
3:     Update $\mathbf{p}^{(n)}[t_s] := \mathbf{p}^*[t_s]$
4:     Set $n := n + 1$
5: **until** Convergence
6: **Output:** $\mathbf{p}^*[t_s] = \mathbf{p}^{(n)}[t_s]$ and $\mathbf{w}_k^{i,j,*}[t_s] = \frac{\sqrt{p_k^{i,j*}[t_s]}}{\sqrt{\nu_k^{i,j}[t_s]}} \mathbf{h}_k^{i,j}[t_s], \ \forall k, (i,j)$.

---

updated feasible point $\mathbf{p}^{(n)}[t_s]$ found at iteration n, denoted by $r_k^{i,j(n)}(\mathbf{p}[t_s]; \mathbf{p}^{(n)}[t_s])$, is given as

$$
\begin{aligned}
r_k^{i,j}(\mathbf{p}[t_s]) \geq \ & r_k^{i,j}(\mathbf{p}^{(n)}[t_s]) - W \log_2 e \Bigg[\gamma_k^{i,j}(\mathbf{p}^{(n)}[t_s]) \\
& - 2 \frac{\nu_k^{i,j}[t_s] \sqrt{p_k^{i,j(n)}[t_s]} \sqrt{p_k^{i,j}[t_s]}}{\Phi_k^{i,j}(\mathbf{p}^{(n)}[t_s])} + \gamma_k^{i,j}(\mathbf{p}^{(n)}[t_s]) \\
& \times \frac{p_k^{i,j}[t_s] \nu_k^{i,j}[t_s] + \Phi_k^{i,j}(\mathbf{p}[t_s])}{p_k^{i,j(n)}[t_s] \nu_k^{i,j}[t_s] + \Phi_k^{i,j}(\mathbf{p}^{(n)}[t_s])} \Bigg] \\
:= \ & r_k^{i,j(n)}(\mathbf{p}[t_s]; \mathbf{p}^{(n)}[t_s])
\end{aligned}
\tag{24}
$$

with $r_k^{i,j(n)}(\mathbf{p}^{(n)}[t_s]; \mathbf{p}^{(n)}[t_s]) = W \log_2(1 + \gamma_k^{i,j}(\mathbf{p}^{(n)}[t_s]))$. As a result, we successively solve the following inner convex approximate program of (23) at iteration n:

$$
\max_{\mathbf{p}[t_s]} \sum_{k \in \mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k^{(n)}(\mathbf{p}[t_s])
\tag{25a}
$$

$$
\text{s.t. } \bar{R}_k^{i,j}[t_s] \leq r_k^{i,j(n)}(\mathbf{p}[t_s]; \mathbf{p}^{(n)}[t_s])\tau, \ \forall k \in \mathcal{K}, (i,j)
\tag{25b}
$$

$$
\sum_{k \in \mathcal{K}} p_k^{i,j}[t_s] \leq P_{\max}^{i,j}, \ \forall (i,j)
\tag{25c}
$$

and update the feasible point $\mathbf{p}^{(n)}[t_s]$ until convergence, where $r_k^{(n)}(\mathbf{p}[t_s]) = \sum_{(i,j) \in \mathcal{P}_k^*} r_k^{i,j(n)}(\mathbf{p}[t_s]; \mathbf{p}^{(n)}[t_s])$. The proposed iterative procedure to solve (13) is summarized in Algorithm 2. An initial feasible value for $\mathbf{p}^{(0)}[t_s]$ to start Algorithm 2 is easily found by successively solving the following simple convex program:

$$
\max_{\mathbf{p}[t_s]} \varrho \triangleq \min_{\forall k,(i,j)} \left\{ r_k^{i,j(n)}(\mathbf{p}[t_s]; \mathbf{p}^{(n)}[t_s])\tau - \bar{R}_k^{i,j}[t_s] \right\}
\tag{26a}
$$

$$
\text{s.t. } \sum_{k \in \mathcal{K}} p_k^{i,j}[t_s] \leq P_{\max}^{i,j}, \ \forall (i,j)
\tag{26b}
$$

until reaching $\varrho > 0$.

Convergence and complexity analysis: The convergence of an IA-based algorithm is already provided in [38]. In particular, Algorithm 2 generates an improved solution after each iteration, which

converges to at least a local optimal solution of (13) when $n \to \infty$. The worst-case of per-iteration complexity of Algorithm 2 is $\mathcal{O}\big(\sqrt{c}(v)^3\big)$ by the interior-point method [39, Chapter 6], where $c = KJ + J$ and $v = KJ$ are the numbers of linear constraints and scalar variables, respectively.

2) ZFBF-Based Transmission Design

To make ZFBF efficient and feasible, the number of antennas of each RU (i,j) is required to be larger than the number of UEs, i.e. $M_{i,j} > K$, $\forall (i,j) \in \mathcal{J}$, to cancel the inter-user interference transmitted 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 proposed ZFBF technique, beamformer $\mathbf{w}_k^{i,j}[t_s]$ at RU (i,j) is designed to satisfy $(\mathbf{h}_{k'}^{i,j}[t_s])^{\mathsf{H}}\mathbf{w}_k^{i,j}[t_s] = 0, \forall k' \in \mathcal{K}\backslash\{k\}$. We denote by $\mathbf{H}_{-k}^{i,j}[t_s] \triangleq \big[\mathbf{h}_1^{i,j}[t_s] \cdots \mathbf{h}_{k-1}^{i,j}[t_s]\, \mathbf{h}_{k+1}^{i,j}[t_s] \cdots \mathbf{h}_K^{i,j}[t_s]\big] \in \mathbb{C}^{M \times (K-1)}$ the channel matrix from RU (i,j) to UEs, except UE k. Let $\mathbf{V}_k^{i,j}[t_s] \in \mathbb{C}^{M_{i,j} \times (M_{i,j}-K+1)}$ be the null space of $(\mathbf{H}_{-k}^{i,j}[t_s])^{\mathsf{H}}$. 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 $\tilde{\mathbf{w}}_k^{i,j}[t_s] \in \mathbb{C}^{(M_{i,j}-K+1)\times 1}, \forall k, (i,j)$ 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 $\tilde{\mathbf{w}}_k^{i,j}[t_s] = \sqrt{\tilde{p}_k^{i,j}[t_s]}\frac{(\tilde{\mathbf{h}}_k^{i,j}[t_s])^{\mathsf{H}}}{\sqrt{\tilde{\nu}_k^{i,j}[t_s]}}$, where $\tilde{\mathbf{p}}[t_s] \triangleq \big[\tilde{p}_k^{i,j}[t_s]\big]_{k,(i,j)\in\mathcal{P}_k}$ are the solutions to the following problem:

$$\max_{\tilde{\mathbf{p}}[t_s]} \quad \sum_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k(\tilde{p}_k^{i,j}[t_s]) \tag{27a}$$

$$\text{s.t.} \quad \bar{R}_k^{i,j}[t_s] \le r_k^{i,j}(\tilde{p}_k^{i,j}[t_s])\tau, \ \forall k, (i,j) \tag{27b}$$

$$\sum_{k\in\mathcal{K}} \tilde{p}_k^{i,j}[t_s] \le P_{\max}^{i,j}, \ \forall (i,j) \tag{27c}$$

where $r_k^{i,j}(\tilde{p}_k^{i,j}[t_s]) \triangleq W^{i,j}\log_2\big(1 + \frac{\tilde{p}_k^{i,j}[t_s]\tilde{\nu}_k^{i,j}[t_s]}{N_0}\big)$. The function $r_k^{i,j}(\tilde{p}_k^{i,j}[t_s])$ is concave in $\tilde{p}_k^{i,j}[t_s]$, leading to the convexity of problem (27). From (27b), one can show that $\tilde{p}_k^{i,j}[t_s] \ge \tilde{p}_{k,\min}^{i,j}[t_s] := \frac{N_0}{\tilde{\nu}_k^{i,j}[t_s]}2^{\frac{\bar{R}_k^{i,j}[t_s]}{W^{i,j}\tau}}-1$. We now develop an efficient method to solve (27) by formulating the partial Lagrangian as

$$L(\tilde{\mathbf{p}}[t_s], \boldsymbol{\mu}) = \sum_{k\in\mathcal{K}} \frac{\hat{q}_k[t_s]}{\tau} r_k(\tilde{p}_k^{i,j}[t_s]) + \sum_{(i,j)\in\mathcal{J}} \mu_{i,j}\big(P_{\max}^{i,j} - \sum_{k\in\mathcal{K}} \tilde{p}_k^{i,j}[t_s]\big) \tag{28}$$

where $\boldsymbol{\mu} \triangleq \{\mu_{i,j} \ge 0\}_{(i,j)\in\mathcal{J}}$ are the Lagrange multipliers of constraint (27c). The dual function can be written as $g(\boldsymbol{\mu}) = \max_{\tilde{\mathbf{p}}[t_s]\ge 0}\{L(\tilde{\mathbf{p}}[t_s], \boldsymbol{\mu})|\tilde{p}_k^{i,j}[t_s] \ge \tilde{p}_{k,\min}^{i,j}[t_s], \forall k, (i,j)\}$. We note that $L(\tilde{\mathbf{p}}[t_s], \boldsymbol{\mu})$ is separable with respect to $\tilde{p}_k^{i,j}[t_s]$. Thus, by solving

$$\tilde{p}_k^{i,j*}[t_s] = \operatorname*{argmax}_{\tilde{p}_k^{i,j}[t_s]\ge\tilde{p}_{k,\min}^{i,j}[t_s]} \left\{\frac{\hat{q}_k[t_s]}{\tau}W\log_2\Big(1 + \frac{\tilde{p}_k^{i,j}[t_s]\tilde{\nu}_k^{i,j}[t_s]}{N_0}\Big) - \mu_{i,j}\tilde{p}_k^{i,j}[t_s]\right\} \tag{29}$$

Algorithm 3: Proposed Low-Complexity Algorithm for Solving (13) with ZFBF-Based Transmission Design

---

Initialization: Set $n := 1$ and generate initial values $\underline{\mu}_{i,j} = 0$ and $\overline{\mu}_{i,j} = +\infty, \forall (i,j) \in \mathcal{J}$

1: **for** each RU $(i,j) \in \mathcal{J}$ *in parallel* **do**
2:     **repeat**
3:         Compute $\mu_{i,j}^{(n)} = (\underline{\mu}_{i,j} + \overline{\mu}_{i,j})/2$ and $\tilde{p}_k^{i,j(n)}[t_s]$ as   in (30)
4:         **if** $\sum_{k \in \mathcal{K}} \tilde{p}_k^{i,j}{}'[t_s] - P_{\max}^{i,j} \leq 0$ **then**
5:             Compute $\mu_{i,j}' = (\underline{\mu}_{i,j} + \overline{\mu}_{i,j})/2$ and update $\overline{\mu}_{i,j} := \mu_{i,j}'$
6:         **else**
7:             Update $\mu_{i,j}' = (\underline{\mu}_{i,j} + \overline{\mu}_{i,j})/2$ and update $\underline{\mu}_{i,j} := \mu_{i,j}'$
8:         **end if**
9:         Set $n := n + 1$
10:     **until** $\overline{\mu}_{i,j} - \underline{\mu}_{i,j} \leq \delta$ {/*Satisfying a given accuracy level*/}
11: **end for**
12: **Output:** $\mu_{i,j}^* = \mu_{i,j}^{(n)}$, $\tilde{p}_k^{i,j*}[t_s] = \max\left\{ \tilde{p}_{k,\min}^{i,j}[t_s], \frac{\hat{q}_k[t_s]W^{i,j}}{\tau \mu_{i,j}^* \ln 2} - \frac{N_0}{\tilde{\nu}_k^{i,j}[t_s]} \right\}$ and

    $\mathbf{w}_k^{i,j,*}[t_s] = \frac{\sqrt{\tilde{p}_k^{i,j,*}[t_s]}}{\sqrt{\tilde{\nu}_k^{i,j}[t_s]}} \mathbf{V}_k^{i,j}[t_s] (\tilde{\mathbf{h}}_k^{i,j}[t_s])^{\mathsf{H}}, \; \forall k, (i,j)$.

---

for a given $\mu_{i,j}$, the optimal solution to $\tilde{p}_k^{i,j}[t_s]$ is given as

$$\tilde{p}_k^{i,j*}[t_s] = \max\left\{ \tilde{p}_{k,\min}^{i,j}[t_s], \frac{\hat{q}_k[t_s]W^{i,j}}{\tau \mu_{i,j} \ln 2} - \frac{N_0}{\tilde{\nu}_k^{i,j}[t_s]} \right\}. \tag{30}$$

The optimal Lagrange multiplier $\mu_{i,j}$ is efficiently found by applying a bisection search method between $\underline{\mu}_{i,j} = 0$ and a sufficiently large $\overline{\mu}_{i,j}$. An efficient algorithm for solving (13) with ZFBF is summarized in Algorithm 3, which does not rely on existing convex optimization solvers.

C. ORAN-based Implementation of Algorithm 1

Fig. 4 illustrates the key steps for implementing the proposed JFCS management scheme at time-frame t in the ORAN architecture.

(1) At the beginning of time-frame $t > 1$, the three-step RL procedure for solving L-SP is carried out at Non-RT RIC based on the collected data in SMO. The collected data include performance/observation and resource updates from CU, DU, RU and Near-RT RIC to SMO. For $t = 1$, the flow-split decisions are initialized as $\boldsymbol{\beta}_k[1] = \frac{1}{|\mathcal{P}_k^{(1)}|}[1, \cdots, 1], \forall k$ where $\mathcal{P}_k^{(1)}$ is the set of RUs in the feasible communication range of UE k.

(2) The optimal flow-split decisions $\boldsymbol{\beta}^*[t]$ are sent to Near-RT RIC via the A1 interface for real deployment.

(3) Given $\boldsymbol{\beta}^*[t]$, xAPPs deployed in Near-RT RIC control congestion and optimize RAN resources and functions in each time-slot t_s by solving S-SP1 and S-SP2 to obtain the optimal solutions of congestion control $\boldsymbol{a}^*[t_s]$ and beamformer $\mathbf{w}^*[t_s]$.

(4) Subsequently, the RAN Data Analytic component in Near-RT RIC updates queue-lengths as in Step 6 of Algorithm 1.

(5) Given $\boldsymbol{\beta}^*[t]$ and $\mathbf{w}^*[t_s]$, the optimal service rate $\mathbf{r}(\mathbf{w}^*[t_s])$ is scheduled and applied to CU and
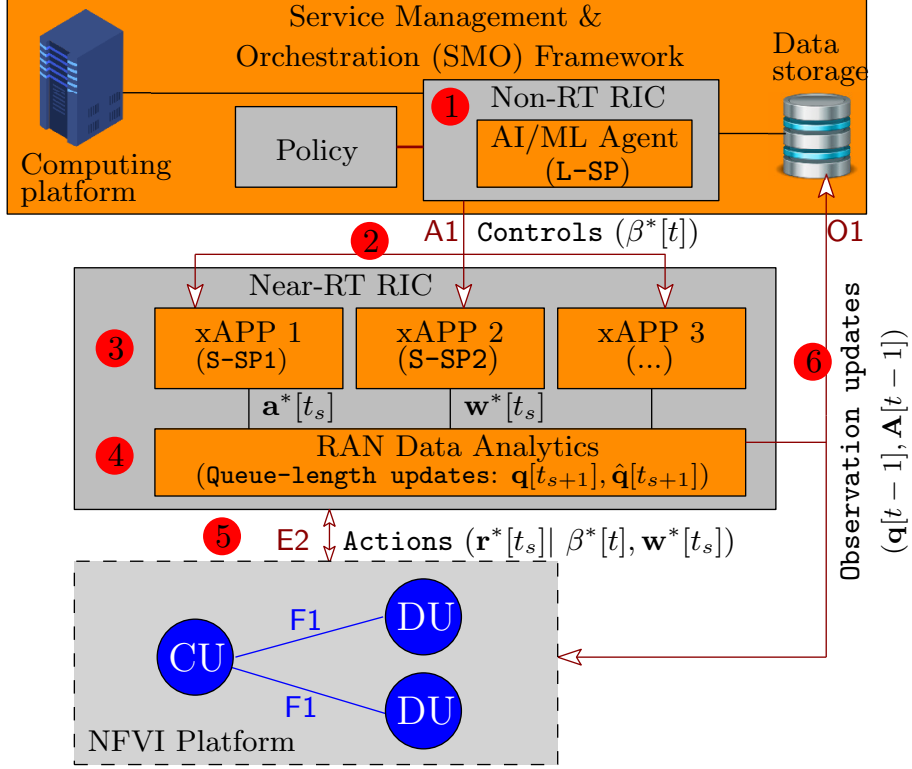
Fig. 4: ORAN Alliance reference architecture for implementing the proposed JFCS management scheme at time-frame t.

DUs through the E2 interface.

⑥ After T_f time-slots in the short-term scale t_s, performance and observations (e.g. $\mathbf{q}[t-1]$, $\mathbf{A}[t-1]$) are updated to SMO through the O1 interface to re-estimate the flow-split decision $\boldsymbol{\beta}^*[t+1]$.

VI. PERFORMANCE ANALYSIS OF THE JFCS FRAMEWORK

In this section, we analyze the main theoretical performance results of Algorithm 1 and discuss their key insights, followed by concrete proofs of the theorems.

Assumption 2. *To facilitate the analysis, we make the following additional assumptions.*

- *Under the limited transmit power budget at RUs, the achievable rate of UE k is upper bounded by $r^{\max} > 0$, i.e., $r_k(\mathbf{w}[t_s]) \leq r^{\max}$, $\forall k, t_s$.*
- *The congestion control variable $a_k[t_s]$ satisfies the condition $\mathbb{E}\{a_k^2[t_s]\} \leq A_1^{\max}$, where $A_1^{\max}$ is a sufficiently large positive constant [32].*

Theorem 1 (Bounding the mean divergence of the auxiliary queue-length). *For a given scaling factor φ, let $\hat{\mathbf{q}}_{(\varphi)}^{\infty}$ and $\hat{\mathbf{q}}_{(\varphi)}^*$ be the steady-state and optimal queue-lengths, respectively. From Assumptions*

1 and 2, the expected upper bound of the divergence of $\hat{\mathbf{q}}_{(\varphi)}^\infty$ from $\hat{\mathbf{q}}_{(\varphi)}^*$ is given as

$$\mathbb{E}\{\|\hat{\mathbf{q}}_{(\varphi)}^\infty - \hat{\mathbf{q}}_{(\varphi)}^*\|_2\} \leq \mathsf{C}_1\sqrt{\varphi} = \mathcal{O}(\sqrt{\varphi}) \tag{31}$$

where $\mathsf{C}_1 \triangleq \sqrt{\frac{K\tau^2\Psi}{2}\big(A_1^{\max} + (r^{\max})^2\big)}$ is a positive constant.

The proof is detailed in Appendix B. Theorem 1 implies that the divergence of the steady-state queue-length is bounded by $\mathcal{O}(\sqrt{\varphi})$. In particular, the smaller the value of φ, the less the divergence of $\hat{\mathbf{q}}_{(\varphi)}^\infty$. However a small φ will also result in a small congestion control rate and a faster convergence. When φ is large, a better congestion control rate is achieved but with the cost of larger steady-state queue-length divergence (i.e. larger delay and slower convergence). Hence, there exists an appropriate value of φ to make this tradeoff more efficient. Theorem 1 will immediately lead to the following result.

Corollary 1 (Queue-stability). *Given a scaling factor φ and C_1 in (31), the steady-state total queue-length remains finite and scales as $\mathcal{O}(\varphi) + \mathcal{O}(\sqrt{\varphi})$, i.e.*

$$\limsup_{t_s\to\infty} \mathbb{E}\{\|\hat{\mathbf{q}}_{(\varphi)}[t_s]\|_1\} \leq \tau\Psi K A^{\max}\varphi + \sqrt{K}\mathsf{C}_1\sqrt{\varphi}$$

$$= \mathcal{O}(\varphi) + \mathcal{O}(\sqrt{\varphi}). \tag{32}$$

Proof. The proof of (32) is straightforward by noticing the fact that $\limsup_{t_s\to\infty} \mathbb{E}\{\|\hat{\mathbf{q}}_{(\varphi)}[t_s]\|_1\} = \mathbb{E}\{\|\hat{\mathbf{q}}_{(\varphi)}^\infty\|_1\} = \mathbb{E}\{\|\hat{\mathbf{q}}_{(\varphi)}^\infty - \hat{\mathbf{q}}_{(\varphi)}^*\|_1 + \|\hat{\mathbf{q}}_{(\varphi)}^*\|_1\}$. Applying the inequality $\|\mathbf{x}\|_1 \leq \sqrt{K}\|\mathbf{x}\|_2$ for any $\mathbf{x} \in \mathbb{R}_+^K$ yields: $\|\hat{\mathbf{q}}_{(\varphi)}^\infty - \hat{\mathbf{q}}_{(\varphi)}^*\|_1 + \|\hat{\mathbf{q}}_{(\varphi)}^*\|_1 \leq \sqrt{K}\|\hat{\mathbf{q}}_{(\varphi)}^\infty - \hat{\mathbf{q}}_{(\varphi)}^*\|_2 + \|\hat{\mathbf{q}}_{(\varphi)}^*\|_1$. From (31) and Step 4 of Algorithm 1, it follows that

$$\limsup_{t_s\to\infty} \mathbb{E}\{\|\hat{\mathbf{q}}_{(\varphi)}[t_s]\|_1\}$$

$$\leq \sqrt{K}\mathbb{E}\{\|\hat{\mathbf{q}}_{(\varphi)}^\infty - \hat{\mathbf{q}}_{(\varphi)}^*\|_2\} + \|\hat{\mathbf{q}}_{(\varphi)}^*\|_1$$

$$\leq \sqrt{K}\mathsf{C}_1\sqrt{\varphi} + \tau\sum_{k\in\mathcal{K}} U_k'\big(a_k^*\big)\varphi \leq \sqrt{K}\mathsf{C}_1\sqrt{\varphi} + \tau\Psi\sum_{k\in\mathcal{K}} a_k^*\varphi$$

$$\leq \sqrt{K}\mathsf{C}_1\sqrt{\varphi} + \tau\Psi K A^{\max}\varphi \quad (\text{due to } a_k^* \leq A^{\max}, \forall k) \tag{33}$$

showing (32). $\qquad\square$

Let $\boldsymbol{a}_{(\varphi)}^\infty \triangleq [a_{(\varphi),k}^\infty]_{k\in\mathcal{K}}^\mathsf{T}$ with $a_{(\varphi),k}^\infty = \mathbb{E}\{\min\{A^{\max}, U_k'^{-1}\big(\frac{\hat{q}_{(\varphi),k}^\infty}{\varphi\tau}\big)\}\}$ be the mean steady-state congestion control rate vector. We also denote by $U(\boldsymbol{a}) \triangleq \sum_{k\in\mathcal{K}} U_k(a_k)$ the total utility function of problem (7). The utility-optimality of Algorithm 1 is stated by the following theorem, whose proof is given in Appendix C.

TABLE I: Simulation Parameters

Parameter	Value
System bandwidth, W	20 MHz
Number of RUs, J	8
Number of UEs, K	12
Number of antennas at RUs (i,j), $M_{i,j} \equiv M$	16
RUs' height	10 m
UEs' antenna altitude	1.5 m
Power budget at RU (i,j), $P_{\max}^{i,j} \equiv P_{\max}$	43 dBm
Noise figure, NF	9 dB
Maximum average delay, $\bar{d}_k \equiv d$	10 ms
Require reliable communication, $\epsilon_k \equiv \epsilon$	0.95
Number of frames, T	10000
Number of time-slots per frame, T_f	10
Duration of one frame, T_c	10 ms
Duration of one time-slot, τ	1 ms
Trade-off factor (Boltzmann temperature), λ	0.3

Theorem 2 (Optimality). *Given a scaling factor φ, Algorithm 1 produces the mean steady-state congestion control rate vector $\boldsymbol{a}_{(\varphi)}^{\infty}$, satisfying*

$$\|\boldsymbol{a}_{(\varphi)}^{\infty} - \boldsymbol{a}^*\|_2 \leq \mathsf{C}_2 \frac{1}{\sqrt{\varphi}} = \mathcal{O}(1/\sqrt{\varphi}) \tag{34}$$

where $\mathsf{C}_2 \triangleq \frac{\mathsf{C}_1}{\psi\tau} = \sqrt{\frac{K\Psi}{2\psi}\left(A_1^{\max} + (r^{\max})^2\right)}$. Therefore, the optimal network utility maximization is bounded as

$$U(\boldsymbol{a}^*) - \mathsf{C}_3 \frac{1}{\varphi} = U(\boldsymbol{a}^*) - \mathcal{O}(1/\varphi) \leq U(\boldsymbol{a}_{(\varphi)}^{\infty}) \tag{35}$$

where $\mathsf{C}_3 \triangleq \frac{\Psi\mathsf{C}_1^2}{2\psi^2\tau^2} = \frac{K\Psi^2}{4\psi}\left(A_1^{\max} + (r^{\max})^2\right)$.

The analytical results in Theorem 2 show that the divergence of the steady-state congestion control rate vector $\boldsymbol{a}_{(\varphi)}^{\infty}$ from $\boldsymbol{a}^*$ scales as $\mathcal{O}(1/\sqrt{\varphi})$, which is the same as in [32], [40]. The utility-optimality gap can be reduced by increasing φ, but this will also lead to a larger steady-state queue-length divergence.

VII. NUMERICAL RESULTS

In this section, we first present simulation setup and parameters in Section VII-A and then provide numerical results of Algorithm 1 in Section VII-B. The results and performance comparison over existing schemes will be provided in Section VII-C.

A. Simulation Setups and Parameters

We consider a system topology given in Fig. 5, including 8 RUs and 12 UEs located within a circle of 1-km radius. There are two DUs, each connected to 4 RUs. RUs are uniformly distributed
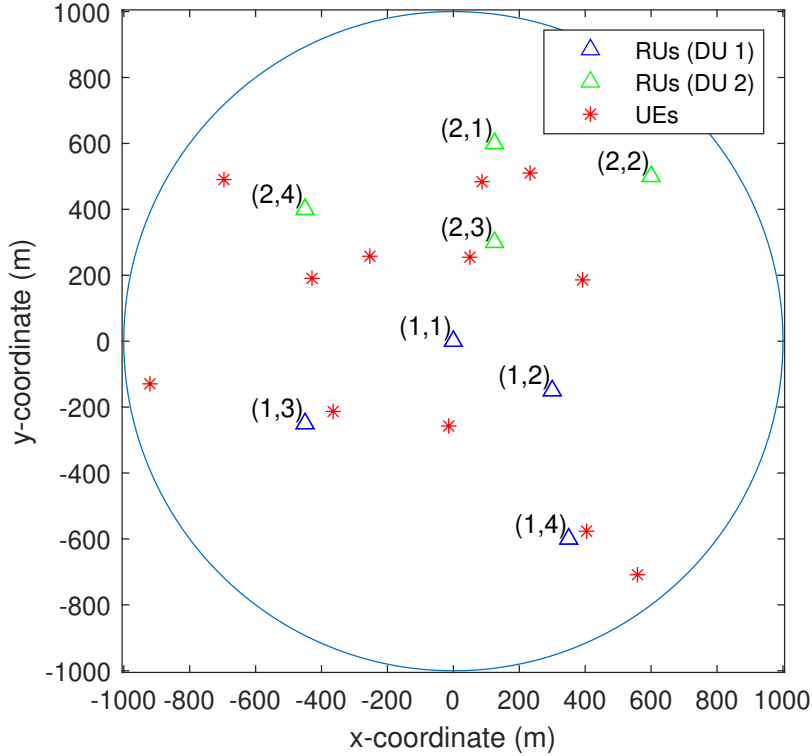
Fig. 5: A system topology with $J = 8$ RUs and $K = 12$ UEs.

in the area, while those of UEs are randomly located in each time-frame t. 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 [41], 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 $\kappa = P_{\mathsf{LoS}}(d[t]) / (1 - P_{\mathsf{LoS}}(d[t]))$, where the LoS probability follows the 3GPP–UMa model as $P_{\mathsf{LoS}}(d[t]) = \min\left(\frac{18}{d[t]}, 1\right)\left(1 - \exp(-\frac{d[t]}{36})\right) + \exp(-\frac{d[t]}{36})$ [42]. We consider uniform linear arrays with half-wavelength distance between array elements to model the LoS channels at RUs. 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 $\left[\boldsymbol{a}(\phi_k^{i,j}[t])\right]_m = \exp(j\pi(m-1)\sin\phi_k^{i,j}[t])$ 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, where $\mathsf{NF} = 9$ dB denotes the noise figure.

We run Algorithm 1 over $T = 10000$ frames, each consists of $T_f = 10$ time-slots (subframes) and has duration of $T_c = 10$ ms, followed by 5G NR Frame structure [43]. 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 $\eta_\beta[t] = 1/(t+1)^{0.6}$ [44]. We adopt the proportional fairness metric to model the utility function as: $U_k(r_k) = \log(0.001 + r_k), \forall k$ [45]. The key parameters are summarized in Table I for ease of cross-referencing, followed studies in [36], [41]–[44]. 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:

- "NUM with fixed resource allocation (NUM-FRA)" [46]: Under Algorithm 1, RUs allocate power equally to UEs.

- "NUM with equal flow-split distribution (NUM-EFSD):" CU splits data-flows of all UEs equally among the selected paths, *i.e.*, $\beta_k^{i,j}[t] = 1/|\mathcal{P}_k|, \forall (i,j) \in \mathcal{P}_k$.

- "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 of Algorithm 1

We first study the impacts of φ and λ on the convergence behavior of Algorithm 1 in Fig. 6. From Fig. 6(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 (see Theorem 2), 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 (8a) will prevail over the Lyapunov drift function $\Delta L[t_s]$, which requires more iterations to guarantee network stability. In Fig. 6(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 the estimated utility that can be achieved with the cost of lower convergence speed of the RL process. From (18), the paths associated with highest estimated regret $\hat{\theta}_k^{i,j}[t]$ will be selected to minimize the best response function $f(\hat{\boldsymbol{\theta}}[t])$. 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.

In Fig. 7, we evaluate the performance of Algorithm 1 with different transmission strategies, *namely* MRT and ZFBF. For a fixed $\varphi = 25$, we vary the number of antennas at RUs $M \equiv M_{i,j}, \forall (i,j)$ from 16 to 128. For each transmission design, we also plot the steady-state congestion control rate $\mathbb{E}\{\|\boldsymbol{a}_{(25)}^\infty\|\}$ with equal flow-split distribution. As seen from Fig. 7 that the steady-state congestion control rate of all schemes increases as M increases. Unsurprisingly, Algorithm 1 with ZFBF offers better performance in terms of congestion control rate than that of MRT when the number of antennas at RUs is sufficiently large to cancel the inter-user interference transmitted by the same RU. It is obvious that the higher the effective data rate of a data-flow in the downlink, the lower the total
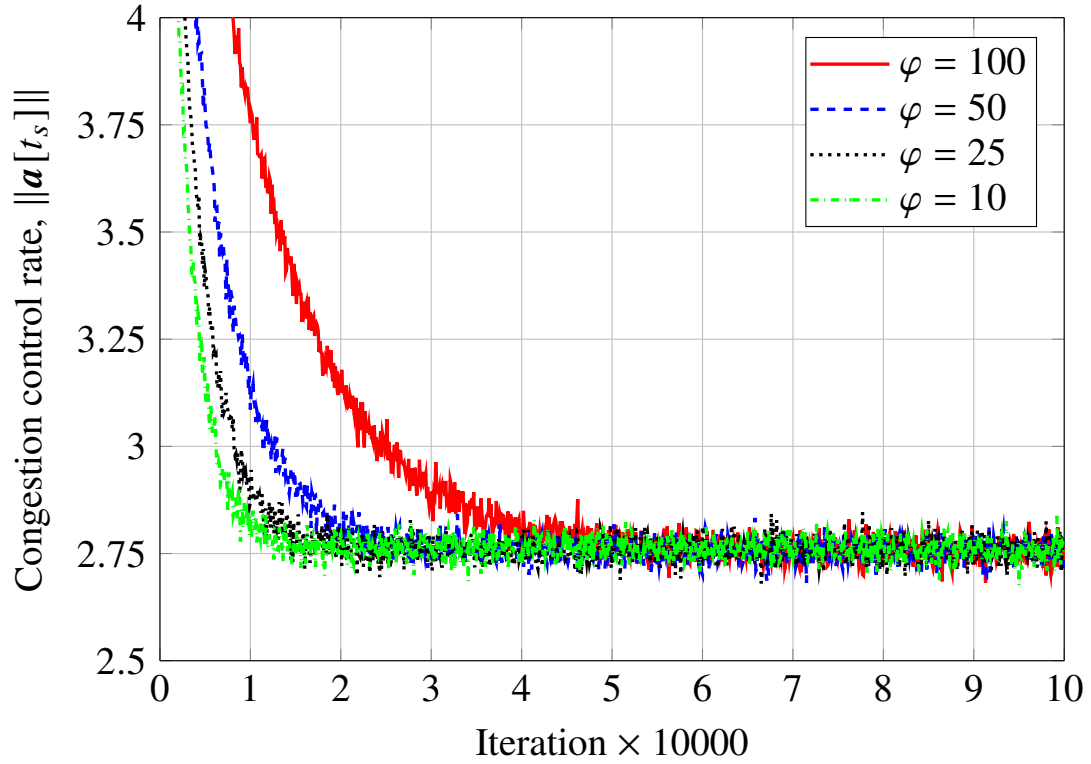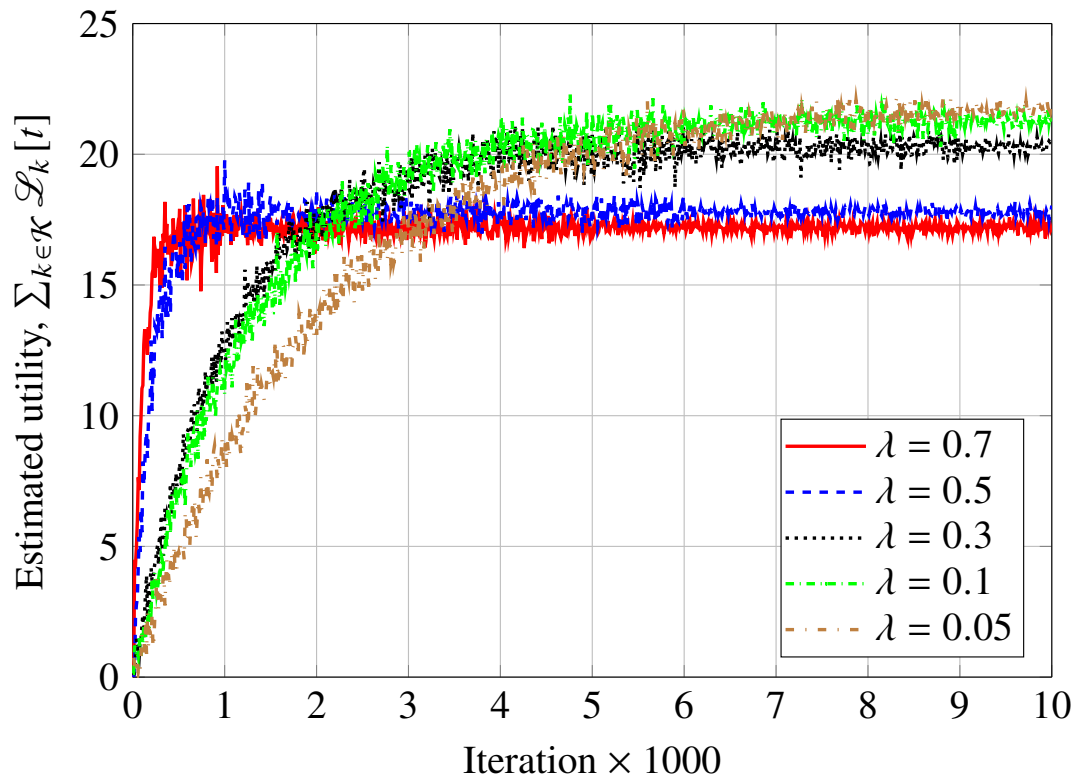
(a) Impact of φ on congestion control rate

(b) Impact of λ on estimated utility
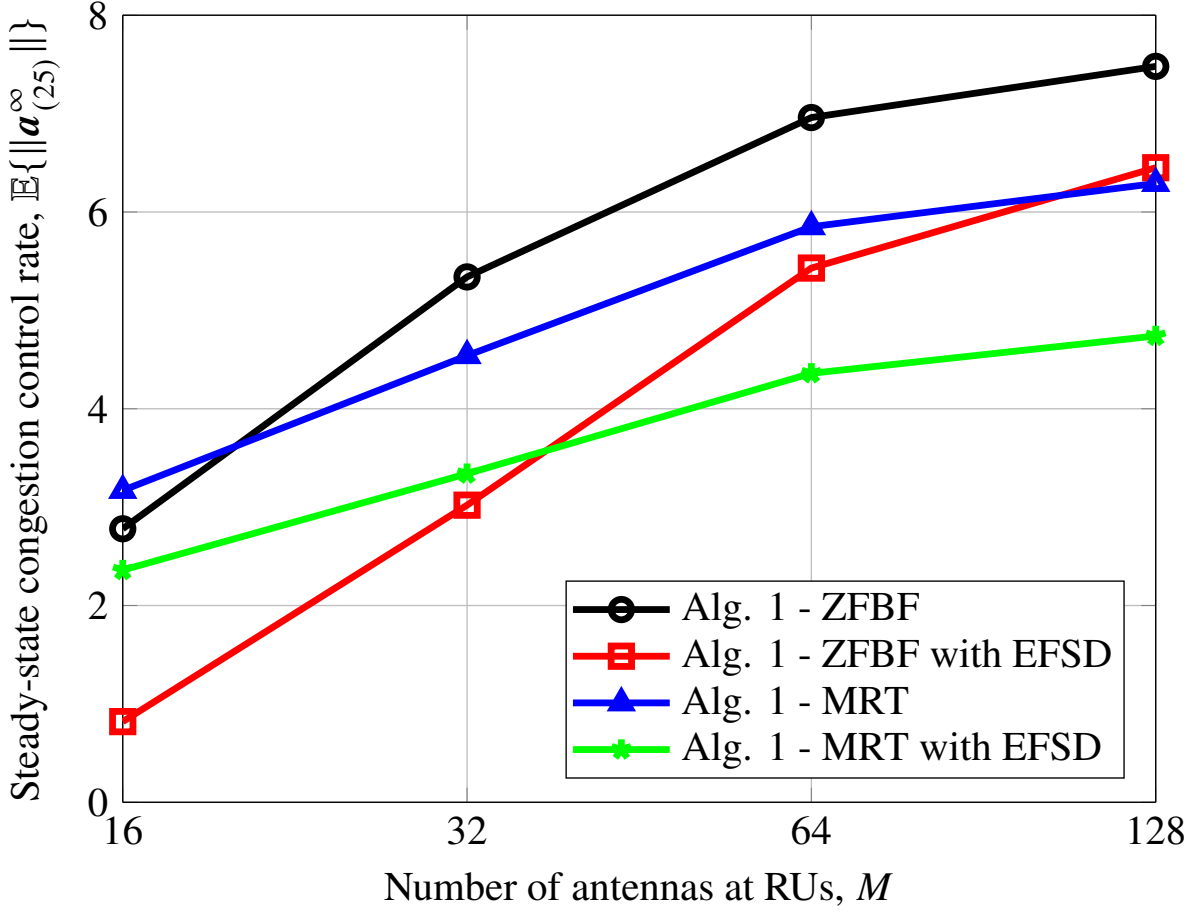
Fig. 6: Convergence behavior of Algorithm 1 with ZFBF.

Fig. 7: Performance of Algorithm 1 with different transmission strategies versus the number of antennas at RUs, $M \equiv M_{i,j}, \forall(i,j)$.

queue-length of that data-flow (or user), resulting in a higher congestion control rate.

Since Algorithm 1 with MRT is based on the IA method that requires high computation complexity and relies on existing convex optimization solvers, we provide only the performance of Algorithm 1 with ZFBF in the following section.

C. Performance Comparison

Next, we show the performance comparison in terms of the steady-state congestion control rate $\mathbb{E}\{\|\boldsymbol{a}_{(25)}^{\infty}\|\}$ among the considered schemes versus the number of antennas at RUs in Fig. 8. We fix $\varphi = 25$ and vary M 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. Clearly, Algorithm 1 outperforms the benchmark schemes in all ranges of M, and the gap is deeper when M is small. In addition, the NUM-FRA and NUM-NRU, which fairly allocate the
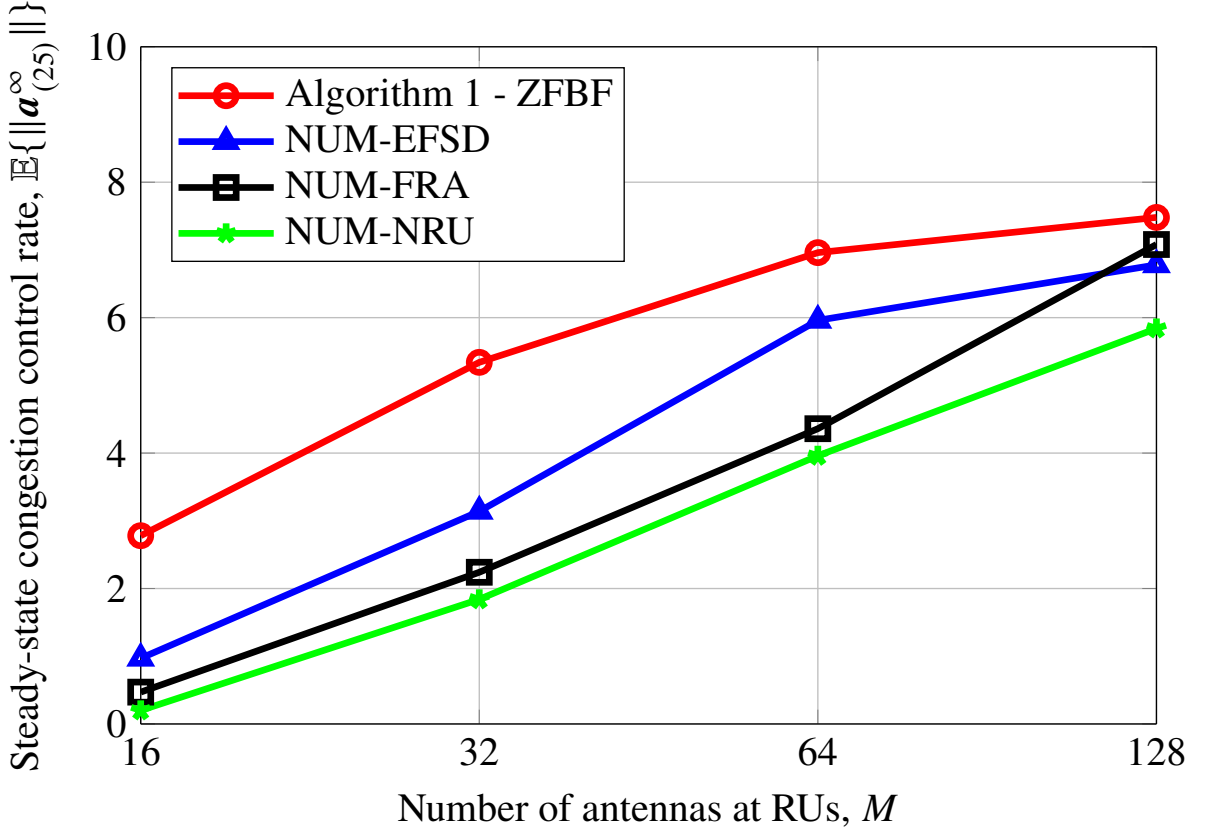
Fig. 8: The steady-state congestion control rate with respect to the number of antennas at RUs, $M \equiv M_{i,j}, \forall(i,j)$.

power budget and fix the path selection to UEs, respectively, provide the worst performance. These observations demonstrate the effectiveness of the proposed Algorithm 1 by jointly optimizing the flow-split distribution, congestion control, scheduling and radio resource allocation.

Lastly, the impacts of scaling factor φ on the steady-state total queue-length $\mathbb{E}\{\|\hat{\mathbf{q}}_{(\varphi)}^{\infty}\|_1\}$ and average worst-case delay (*i.e.*, the delay of slowest data-flow) are plotted in Figs. 9 and 10, respectively. It can be seen from Fig. 9 that the steady-state total queue-length of all schemes monotonically scales as $\mathcal{O}(\varphi) + \mathcal{O}(\sqrt{\varphi})$, which confirms our theoretical results in Corollary 1. We recall from Theorem 2 that the utility-optimality gap can be narrowed by increasing φ, but with the cost of higher delay, as shown in Fig. 10. When φ is larger than 25, all the considered schemes violate the maximum allowable average delay of $\bar{d} = 10$ ms. It implies that the data traffic cannot be completely transmitted to UEs in each time-frame. Nevertheless, Algorithm 1 still provides the best performance out of the schemes considered.
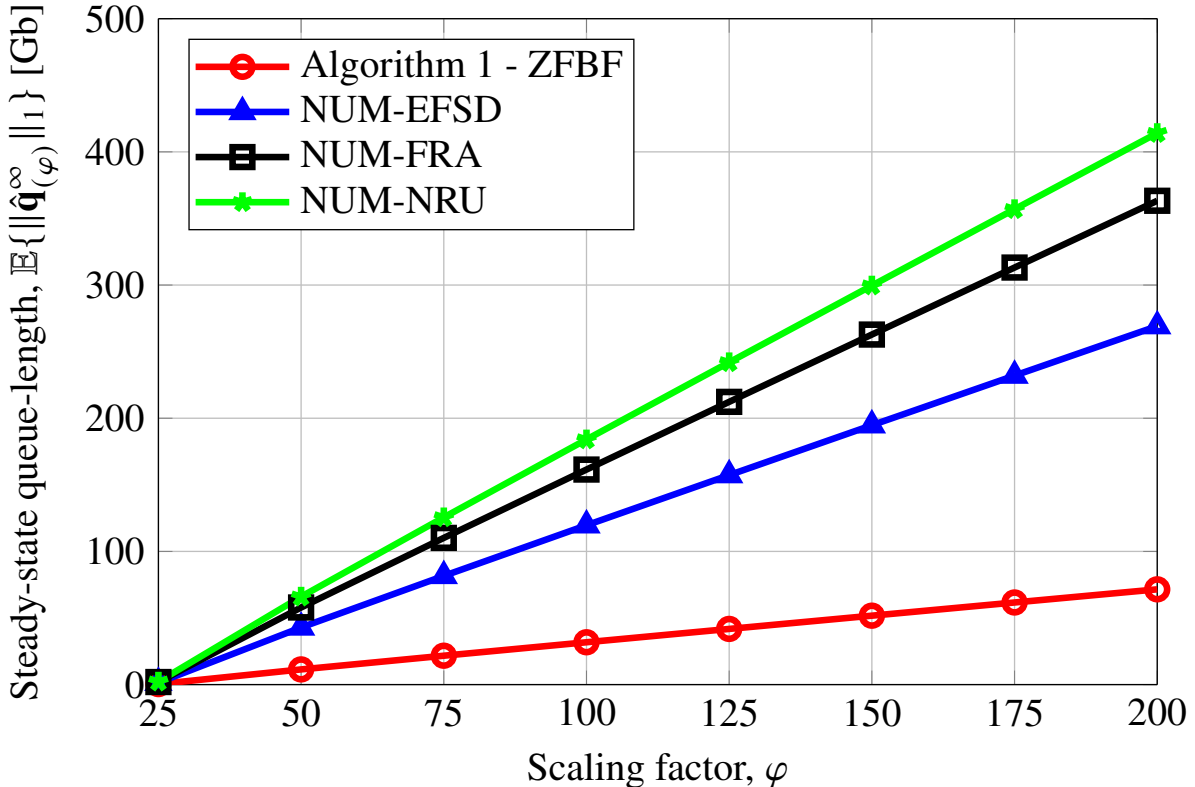
Fig. 9: The steady-state total queue-length with respect to φ.

VIII. CONCLUSION

We have proposed a new holistic multi-layer optimization framework, called JFCS, to enable intelligent traffic steering in a hierarchical ORAN architecture. In particular, we have developed an intelligent resource management algorithm based on network utility maximization and stochastic optimization to efficiently and adaptively direct traffic to appropriate RUs by jointly optimizing the flow-split distribution, congestion control and scheduling. JFCS is proved to achieve fast convergence, long-term utility-optimality and significant delay reduction compared to state-of-the-art approaches. To that end, the insights in this work will foster future studies in this area, especially in the design of more advanced AI/ML solutions to achieve enhanced control and flexibility in ORAN.

APPENDIX A: DERIVATION OF INEQUALITY

We will find the concave lower bound of $r_k^{i,j}[t_s]$. By [47, Appendix A], it is true that the function $r(x,y) = -\ln(1 - x^2/y)$ is convex in the domain $y > x^2$ with $x, y \in \mathbb{R}_+$. The global concave lower bound of $r(x,y)$ at the feasible point $(\bar{x}, \bar{y})$ is given as

$$r(x,y) \geq r(\bar{x}, \bar{y}) + \left\langle \left(\frac{\partial r(\bar{x}, \bar{y})}{\partial \bar{x}}, \frac{\partial r(\bar{x}, \bar{y})}{\partial \bar{y}} \right), (x - \bar{x}, y - \bar{y}) \right\rangle$$
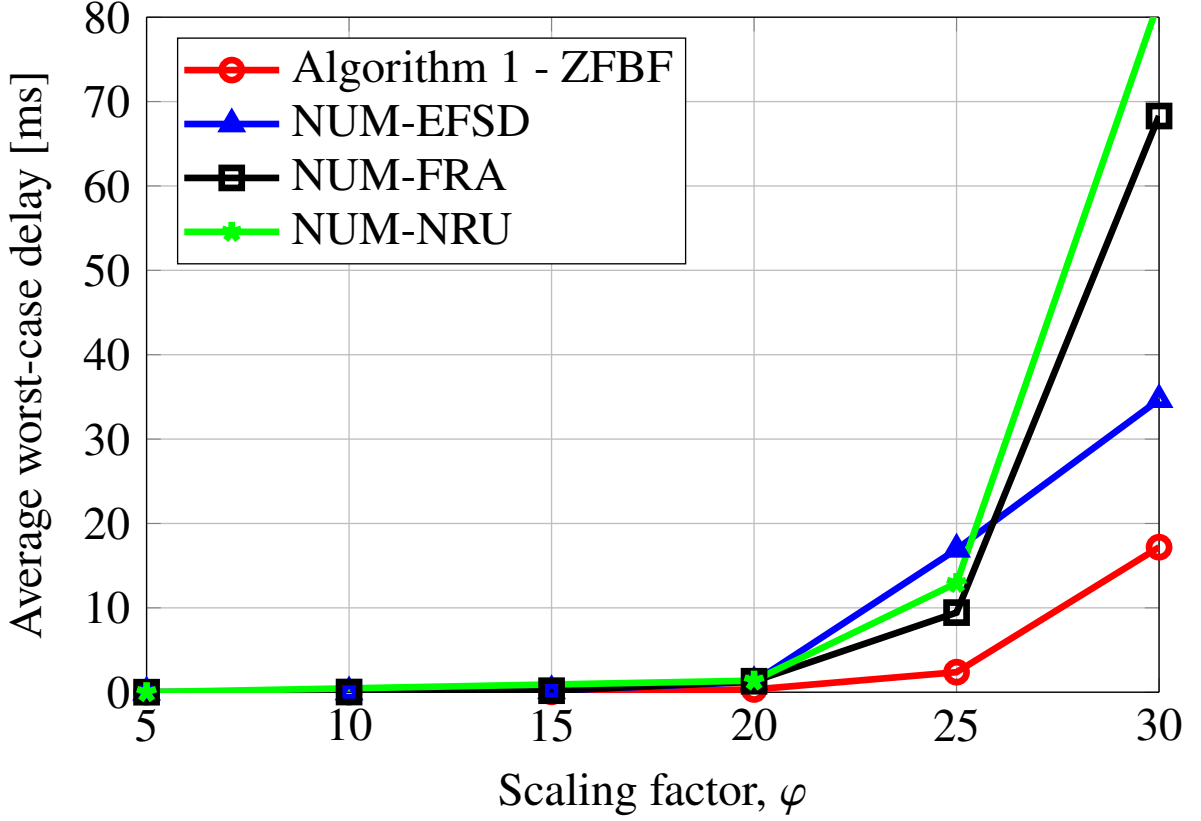
Fig. 10: Average worst-case delay with respect to φ.

$$= r(\bar{x}, \bar{y}) - \frac{\bar{x}^2}{\bar{y} - \bar{x}^2} + 2\frac{\bar{x}x}{\bar{y} - \bar{x}^2} - \frac{\bar{x}^2}{\bar{y} - \bar{x}^2}\frac{y}{\bar{y}} \tag{A.1}$$

by applying the first-order Taylor approximation. By the fact that $\ln\left(1 + \frac{x^2}{z}\right) = -\ln\left(1 - \frac{x^2}{z+x^2}\right)$ and substituting $y = z + x^2$, $\bar{y} = \bar{z} + \bar{x}^2$, $x = \sqrt{v}$ and $\bar{x} = \sqrt{\bar{v}}$ into (A.1), we obtain

$$r(v, z) \triangleq \ln\left(1 + \frac{v}{z}\right) \geq r(\bar{v}, \bar{z}) - \frac{\bar{v}}{\bar{z}} + 2\frac{\sqrt{\bar{v}}\sqrt{v}}{\bar{z}} - \frac{\bar{v}(z+v)}{\bar{z}(\bar{z} + \bar{v})}$$

$$:= \bar{r}(v, z; \bar{v}, \bar{z}) \tag{A.2}$$

where $\bar{r}(v, z; \bar{v}, \bar{z})$ is concave and $\bar{r}(\bar{v}, \bar{z}; \bar{v}, \bar{z}) = r(\bar{v}, \bar{z})$ whenever $(v, z) = (\bar{v}, \bar{z})$.

APPENDIX B: PROOF OF THEOREM 1

For a given φ, the quadratic Lyapunov function defined in Section IV-A is rewritten with respect to $\hat{\mathbf{q}}_{(\varphi)}[t_s]$ as: $L(\hat{\mathbf{q}}_{(\varphi)}[t_s]) = \frac{1}{2\tau^2}\|\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}^*_{(\varphi)}\|_2^2$. Following [32, Theorem 3], the mean Lyapunov drift from time-slot t_s to t_{s+1} is computed as

$$\Delta\bar{L}(\hat{\mathbf{q}}_{(\varphi)}[t_s]) = \mathbb{E}\{\Delta L(\hat{\mathbf{q}}_{(\varphi)}[t_s])\} = \mathbb{E}\{L(\hat{\mathbf{q}}_{(\varphi)}[t_{s+1}]) - L(\hat{\mathbf{q}}_{(\varphi)}[t_s])\}$$

$$= \frac{1}{2\tau^2}\mathbb{E}\left\{\left(\hat{\mathbf{q}}_{(\varphi)}[t_{s+1}] + \hat{\mathbf{q}}_{(\varphi)}[t_s] - 2\hat{\mathbf{q}}^*_{(\varphi)}\right)^\mathsf{T}\left(\hat{\mathbf{q}}_{(\varphi)}[t_{s+1}] - \hat{\mathbf{q}}_{(\varphi)}[t_s]\right)\right\}$$

$$\leq \frac{1}{2\tau}\mathbb{E}\Big\{\big(2\hat{\mathbf{q}}_{(\varphi)}[t_s] + \big(\boldsymbol{a}[t_s] - \mathbf{r}(\mathbf{w}[t_s])\big)\tau - 2\hat{\mathbf{q}}_{(\varphi)}^*\big)^\mathsf{T}\big(\boldsymbol{a}[t_s] - \mathbf{r}(\mathbf{w}[t_s])\big)\Big\}$$

$$= \underbrace{\frac{1}{2}\mathbb{E}\{\|\boldsymbol{a}[t_s] - \mathbf{r}(\mathbf{w}[t_s])\|_2^2\}}_{\triangleq \mathsf{B}_1} + \underbrace{\frac{1}{\tau}\mathbb{E}\{(\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^*)^\mathsf{T}\big(\boldsymbol{a}[t_s] - \mathbf{r}(\mathbf{w}[t_s])\big)\}}_{\triangleq \mathsf{B}_2} \quad \text{(B.1)}$$

by using the inequalities: $([x]^+)^2 \leq x^2$ and $x^2 - y^2 = (x+y)(x-y)$, and the fact that $\hat{\mathbf{q}}_{(\varphi)}[t_{s+1}] - \hat{\mathbf{q}}_{(\varphi)}^* = \hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^* + \big(\boldsymbol{a}[t_s] - \mathbf{r}(\mathbf{w}[t_s])\big)\tau$.

We first focus on providing the expected bound of B_1 as

$$\mathsf{B}_1 = \frac{1}{2}\mathbb{E}\{\|\boldsymbol{a}[t_s]\|_2^2 - 2\boldsymbol{a}[t_s]^\mathsf{T}\mathbf{r}(\mathbf{w}[t_s]) + \|\mathbf{r}(\mathbf{w}[t_s])\|_2^2\}$$

$$\leq \frac{1}{2}\mathbb{E}\{\|\boldsymbol{a}[t_s]\|_2^2 + \|\mathbf{r}(\mathbf{w}[t_s])\|_2^2\}$$

$$\leq \frac{K}{2}\big(A_1^{\max} + (r^{\max})^2\big) \triangleq \mathsf{B}_1^{\mathsf{UB}} \quad \text{(B.2)}$$

where the last inequality follows from Assumption 2. To bound B_2, we first rewrite it equivalently as

$$\mathsf{B}_2 = \frac{1}{\tau}(\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^*)^\mathsf{T}\big(\mathbb{E}\{\boldsymbol{a}[t_s]\} - \mathbf{r}^*\big) + \frac{1}{\tau}\mathbb{E}\{(\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^*)^\mathsf{T}\big(\mathbf{r}^* - \mathbf{r}(\mathbf{w}[t_s])\big)\}. \quad \text{(B.3)}$$

From (7), it follows that $(\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^*)^\mathsf{T}\big(\mathbb{E}\{\boldsymbol{a}[t_s]\} - \mathbf{r}^*\big) \leq 0$. By applying the Cauchy–Schwarz inequality, i.e. $|\mathbf{x}^\mathsf{T}\mathbf{y}| \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2$, to the first term in (B.3), we have

$$\frac{1}{\tau}(\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^*)^\mathsf{T}\big(\mathbb{E}\{\boldsymbol{a}[t_s]\} - \mathbf{r}^*\big) \leq -\frac{1}{\tau}\sum_{k\in\mathcal{K}}|\hat{q}_{(\varphi),k}[t_s] - \hat{q}_{(\varphi),k}^*||a_k[t_s] - r_k^*|. \quad \text{(B.4)}$$

By Assumption 1 on Ψ-smooth and Step 4 of Algorithm 1, it is true that $a_k[t_s] - r_k^* = U_k^{'-1}\big(\frac{\hat{q}_{(\varphi),k}[t_s]}{\varphi\tau}\big) - U_k^{'-1}\big(\frac{\hat{q}_{(\varphi),k}^*[t_s]}{\varphi\tau}\big) \leq 0$ and $\big|U_k'\big(\frac{\hat{q}_{(\varphi),k}[t_s]}{\varphi\tau}\big) - U_k'\big(\frac{\hat{q}_{(\varphi),k}^*[t_s]}{\varphi\tau}\big)\big| \leq \Psi\big|\frac{\hat{q}_{(\varphi),k}[t_s]}{\varphi\tau} - \frac{\hat{q}_{(\varphi),k}^*[t_s]}{\varphi\tau}\big|$. In addition, we have $\big|U_k^{'-1}\big(\frac{\hat{q}_{(\varphi),k}[t_s]}{\varphi\tau}\big) - U_k^{'-1}\big(\frac{\hat{q}_{(\varphi),k}^*[t_s]}{\varphi\tau}\big)\big| \geq \frac{1}{\Psi}\big|\frac{\hat{q}_{(\varphi),k}[t_s]}{\varphi\tau} - \frac{\hat{q}_{(\varphi),k}^*[t_s]}{\varphi\tau}\big|$ due to the inverse function lemma. From the fact that $(\hat{\mathbf{q}}_{(\varphi)}^*)^\mathsf{T}\mathbf{r}^* - (\hat{\mathbf{q}}_{(\varphi)}^*)^\mathsf{T}\mathbf{r}(\mathbf{w}[t_s]) \geq 0$, we can further bound B_2 as

$$\mathsf{B}_2 \leq -\frac{1}{\tau^2\Psi\varphi}\|\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^*\|_2^2 + \frac{1}{\tau}\mathbb{E}\{(\hat{\mathbf{q}}_{(\varphi)}[t_s])^\mathsf{T}\big(\mathbf{r}^* - \mathbf{r}(\mathbf{w}[t_s])\big)\} \quad \text{(B.5)}$$

where the term $\mathbb{E}\{(\mathbf{r}^* - \mathbf{r}(\mathbf{w}[t_s]))\}$ is a constant with respect to $\hat{\mathbf{q}}_{(\varphi)}[t_s]$. Substituting (B.2) and (B.5) into (B.1) yields

$$\Delta\bar{L}(\hat{\mathbf{q}}_{(\varphi)}[t_s]) \leq -\frac{1}{\tau^2\Psi\varphi}\|\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^*\|_2^2 + \mathsf{B}_1^{\mathsf{UB}} + \frac{1}{\tau}\mathbb{E}\{(\hat{\mathbf{q}}_{(\varphi)}[t_s])^\mathsf{T}\big(\mathbf{r}^* - \mathbf{r}(\mathbf{w}[t_s])\big)\}. \quad \text{(B.6)}$$

We now compute the mean Lyapunov drift over TT_f time-slots as

$$\Delta\bar{L} = \sum_{t=1}^{T}\sum_{s=1}^{T_f}\mathbb{E}\{L(\hat{\mathbf{q}}_{(\varphi)}[t_{s+1}]) - L(\hat{\mathbf{q}}_{(\varphi)}[t_s])|\hat{\mathbf{q}}_{(\varphi)}[1_1]\}$$

$$= \sum_{t=1}^{T}\sum_{s=1}^{T_f}\sum_{\hat{\mathbf{q}}_{(\varphi)}\geq 0}\Big(\mathsf{Prob}\big(\hat{\mathbf{q}}_{(\varphi)}[t_s] = \hat{\mathbf{q}}_{(\varphi)}|\hat{\mathbf{q}}_{(\varphi)}[1_1]\big)$$

$$\times \mathbb{E}\{L(\hat{\mathbf{q}}_{(\varphi)}[t_{s+1}]) - L(\hat{\mathbf{q}}_{(\varphi)}[t_s])|\hat{\mathbf{q}}_{(\varphi)}[t_s] = \hat{\mathbf{q}}_{(\varphi)}\}\Big). \tag{B.7}$$

Let us denote by $\rho_{\hat{\mathbf{q}}_{(\varphi)}}^{\infty}$ the stationary distribution of the Markov chain $\hat{\mathbf{q}}_{(\varphi)}[t_s] \geq 0$, i.e. $\rho_{\hat{\mathbf{q}}_{(\varphi)}}^{\infty} = \lim_{T\to\infty} \frac{1}{TT_f} \sum_{t=1}^{T} \sum_{s=1}^{T_f} \mathsf{Prob}\big(\hat{\mathbf{q}}_{(\varphi)}[t_s] = \hat{\mathbf{q}}_{(\varphi)}|\hat{\mathbf{q}}_{(\varphi)}[1_1]\big)$. By substituting (B.6) into (B.7) and dividing both side with TT_f, we have

$$\sum_{\hat{\mathbf{q}}_{(\varphi)} \geq 0} \rho_{\hat{\mathbf{q}}_{(\varphi)}}^{\infty} \Big(-\frac{1}{\tau^2 \Psi \varphi} \|\hat{\mathbf{q}}_{(\varphi)}[t_s] - \hat{\mathbf{q}}_{(\varphi)}^*\|_2^2 + \mathsf{B}_1^{\mathsf{UB}} + \frac{1}{\tau}(\hat{\mathbf{q}}_{(\varphi)}[t_s])^{\mathsf{T}} \mathbb{E}\big\{\big(\mathbf{r}^* - \mathbf{r}(\mathbf{w}[t_s])\big)\big\}\Big)$$

$$= -\frac{1}{\tau^2 \Psi \varphi} \mathbb{E}\big\{\|\hat{\mathbf{q}}_{(\varphi)}^{\infty} - \hat{\mathbf{q}}_{(\varphi)}^*\|_2^2\big\} + \mathsf{B}_1^{\mathsf{UB}} + \frac{1}{\tau}\mathbb{E}\big\{(\hat{\mathbf{q}}_{(\varphi)}^{\infty})^{\mathsf{T}}\big(\mathbf{r}^* - \mathbf{r}^{\infty}\big\} \geq 0 \tag{B.8}$$

where $\mathbf{r}^{\infty} = \operatorname*{argmax}_{r_k(\mathbf{w})\in\mathscr{C}_{\mathbf{H}[\infty]},\forall k\in\mathcal{K}} \sum_{k\in\mathcal{K}} \hat{q}_k^{\infty} r_k(\mathbf{w})$. We note here that $(\hat{\mathbf{q}}_{(\varphi)}^{\infty})^{\mathsf{T}}\mathbf{r}^{\infty} = \max_{r_k(\mathbf{w})\in\mathscr{C}_{\mathbf{H}[\infty]},\forall k\in\mathcal{K}} \sum_{k\in\mathcal{K}} \hat{q}_k^{\infty} r_k(\mathbf{w}) \geq (\hat{\mathbf{q}}_{(\varphi)}^{\infty})^{\mathsf{T}}\mathbf{r}^*$, yielding

$$\frac{1}{\tau^2 \Psi \varphi} \mathbb{E}\big\{\|\hat{\mathbf{q}}_{(\varphi)}^{\infty} - \hat{\mathbf{q}}_{(\varphi)}^*\|_2^2\big\} - \mathsf{B}_1^{\mathsf{UB}} \leq 0 \tag{B.9}$$

This implies that $\mathbb{E}\big\{\|\hat{\mathbf{q}}_{(\varphi)}^{\infty} - \hat{\mathbf{q}}_{(\varphi)}^*\|_2\big\} \leq \sqrt{\frac{K\tau^2\Psi}{2}\big(A_1^{\max} + (r^{\max})^2\big)}\sqrt{\varphi}$ where $\mathsf{B}_1^{\mathsf{UB}} = \frac{K}{2}\big(A_1^{\max} + (r^{\max})^2\big)$, showing the inequality (31) in Theorem 1.

APPENDIX C: PROOF OF THEOREM 2

To prove (34), we first recall that $a_{(\varphi),k}^{\infty} - a_k^* = U_k^{'-1}\big(\frac{\hat{q}_{(\varphi),k}^{\infty}}{\varphi\tau}\big) - U_k^{'-1}\big(\frac{\hat{q}_{(\varphi),k}^*}{\varphi\tau}\big)$ and $\big|U_k'\big(\frac{\hat{q}_{(\varphi),k}^{\infty}}{\varphi\tau}\big) - U_k'\big(\frac{\hat{q}_{(\varphi),k}^*}{\varphi\tau}\big)\big| \geq \psi\big|\frac{\hat{q}_{(\varphi),k}^{\infty}}{\varphi\tau} - \frac{\hat{q}_{(\varphi),k}^*}{\varphi\tau}\big|$ using Assumption 1. By the inverse function lemma, we have $\big|U_k^{'-1}\big(\frac{\hat{q}_{(\varphi),k}^{\infty}}{\varphi\tau}\big) - U_k^{'-1}\big(\frac{\hat{q}_{(\varphi),k}^*}{\varphi\tau}\big)\big| \leq \frac{1}{\psi}\big|\frac{\hat{q}_{(\varphi),k}^{\infty}}{\varphi\tau} - \frac{\hat{q}_{(\varphi),k}^*}{\varphi\tau}\big|$, which yields

$$\|\boldsymbol{a}_{(\varphi)}^{\infty} - \boldsymbol{a}^*\|_2 \leq \frac{1}{\psi\tau\varphi}\|\hat{\mathbf{q}}_{(\varphi)}^{\infty} - \hat{\mathbf{q}}_{(\varphi)}^*\|_2 \overset{(31)}{\leq} \frac{\mathsf{C}_1}{\psi\tau}\frac{1}{\sqrt{\varphi}}. \tag{C.1}$$

Next, it is assumed that $U_k(\cdot)$ is twice continuously differentiable, increasing, and strictly concave. If the utility function $U(\boldsymbol{a})$ has a maximizer $\boldsymbol{a}^*$, then

$$U(\boldsymbol{a}^*) - U(\boldsymbol{a}_{(\varphi)}^{\infty}) \leq \frac{\Psi}{2}\|\boldsymbol{a}^* - \boldsymbol{a}_{(\varphi)}^{\infty}\|_2^2 \leq \frac{\Psi\mathsf{C}_1^2}{2\psi^2\tau^2}\frac{1}{\varphi} \tag{C.2}$$

where the last inequality follows from (C.1). The proof is thus complete.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[2] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, 2019.

[3] O-RAN Alliance, "ORAN-WG1 O-RAN architecture description v01.00.00." *Technical Specification*, Feb. 2020.

[4] SAMSUNG, "ORAN - The Open Road to 5G," *White Paper*, July 2019. [Online]. Available: https://www.samsung.com/global/business/networks/insights/whitepapers/ORAN-the-open-road-to-5g/

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

[6] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in O-RAN for data-driven nextG cellular networks," *IEEE Commun. Mag.*, vol. 59, no. 10, pp. 21–27, 2021.

[7] L. Gavrilovska, V. Rakovic, and D. Denkovski, "From cloud RAN to ORAN," *Wireless Personal Communications*, Mar. 2020.

[8] J. Wang, H. Roy, and C. Kelly, "OpenRAN: The next generation of radio access networks," *Accesnture Startegy, Tech. Rep.*, Nov. 2019. [Online]. Available: https://telecominfraproject.com/openran/

[9] 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.

[10] 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.

[11] H. Lee, J. Cha, D. Kwon, M. Jeong, and I. Park, "Hosting AI/ML workflows on O-RAN RIC platform," in *IEEE Global Commun. Conf. Workshops (IEEE GLOBECOM Wkshps.)*, 2020, pp. 1–6.

[12] J. A. Ayala-Romero, A. Garcia-Saavedra, X. Costa-Perez, and G. Iosifidis, "Bayesian online learning for energy-aware resource orchestration in virtualized RANs," in *IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, 2021, pp. 1–10.

[13] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. on Netw.*, vol. 16, no. 2, pp. 396–409, 2008.

[14] 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.

[15] 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.

[16] 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.

[17] J. Tang, W. P. Tay, and T. Q. S. Quek, "Cross-layer resource allocation with elastic service scaling in cloud radio access network," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5068–5081, 2015.

[18] P. Luong, F. Gagnon, C. Despins, and L.-N. Tran, "Joint virtual computing and radio resource allocation in limited fronthaul green C-RANs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2602–2617, 2018.

[19] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Coordinated scheduling and power control in cloud-radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2523–2536, 2016.

[20] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Low-complexity scheduling and power adaptation for coordinated cloud-radio access networks," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2298–2301, 2017.

[21] M. S. Al-Abiad, A. Douik, and S. Sorour, "Rate aware network codes for cloud radio access networks," *IEEE Trans. Mobile Comput.*, vol. 18, no. 8, pp. 1898–1910, 2019.

[22] A. Douik, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Rate aware instantly decodable network codes," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 998–1011, 2017.

[23] M. S. Al-Abiad, A. Douik, S. Sorour, and M. J. Hossain, "Throughput maximization in cloud-radio access networks using cross-layer network coding," *IEEE Trans. Mobile Comput.*, vol. 21, no. 2, pp. 696–711, 2022.

[24] Y. Cao, S.-Y. Lien, Y.-C. Liang, K.-C. Chen, and X. Shen, "User access control in open radio access networks: A federated deep reinforcement learning approach," *IEEE Trans. Wire. Commun.*, vol. 21, no. 6, pp. 3721–3736, 2022.

[25] 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(accessedon10November2021)

[26] 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.

[27] 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.

[28] 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.

[29] S. Singh, M. Geraseminko, S.-p. Yeh, N. Himayat, and S. Talwar, "Proportional fair traffic splitting and aggregation in heterogeneous wireless networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1010–1013, 2016.

[30] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.

[31] P. Key, L. Massoulie, and D. Towsley, "Path selection and multipath congestion control," in *IEEE Conf. Comput. Commun. (IEEE INFO- COM)*, 2007, pp. 143–151.

[32] J. Liu, A. Eryilmaz, N. B. Shroff, and E. S. Bentley, "Understanding the impacts of limited channel state information on massive MIMO cellular network optimization," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1715–1727, 2017.

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

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

[35] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Automatic Control*, vol. 37, no. 12, pp. 1936–1948, 1992.

[36] 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.

[37] D. S. Leslie and E. Collins, "Convergent multiple-timescales reinforcement learning algorithms in normal form games," *Anna. Appl. Prob.*, vol. 13, no. 4, pp. 1231–1251, 2003.

[38] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.

[39] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization.*   Philadelphia: MPS-SIAM Series on Optimi., SIAM, 2001.

[40] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," *IEEE/ACM Trans. Net.*, vol. 15, no. 6, pp. 1333–1344, 2007.

[41] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.

[42] A. H. Jafari, D. López-Pérez, M. Ding, and J. Zhang, "Study on scheduling techniques for ultra dense small cell networks," in *IEEE Veh. Technol. Conf. (VTC-Fall)*, 2015, pp. 1–6.

[43] 3GPP, *NR, Physical channels and modulation (Release 15)*, document 3GPP TS 38.211 version 15.2.0 Release 15, 2017.

[44] S. Samarakoon, M. Bennis, W. Saad, and M. Latva-aho, "Backhaul-aware interference management in the uplink of wireless small cell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5813–5825, 2013.

[45] 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.

[46] 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.

[47] V.-D. Nguyen, T. Q. Duong, H. D. Tuan, O.-S. Shin, and H. V. Poor, "Spectral and energy efficiencies in full-duplex wireless information and power transfer," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2220–2233, May 2017.