DOI:
10.1109/LSP.2019.2932193

Online Reinforcement Learning of X-Haul Content Delivery Mode in Fog Radio Access Networks

Jihwan Moon, Osvaldo Simeone, *Fellow*, *IEEE*, Seok-Hwan Park, *Member*, *IEEE*, and Inkyu Lee, *Fellow*, *IEEE*

Abstract

We consider a Fog Radio Access Network (F-RAN) with a Base Band Unit (BBU) in the cloud and multiple cache-enabled enhanced Remote Radio Heads (eRRHs). The system aims at delivering contents on demand with minimal average latency from a time-varying library of popular contents. Information about uncached requested files can be transferred from the cloud to the eRRHs by following either backhaul or fronthaul modes. The backhaul mode transfers fractions of the requested files, while the fronthaul mode transmits quantized baseband samples as in Cloud-RAN (C-RAN). The backhaul mode allows the caches of the eRRHs to be updated, which may lower future delivery latencies. In contrast, the fronthaul mode enables cooperative C-RAN transmissions that may reduce the current delivery latency. Taking into account the trade-off between current and future delivery performance, this paper proposes an adaptive selection method between the two delivery modes to minimize the long-term delivery latency. Assuming an unknown and time-varying popularity model, the method is based on model-free Reinforcement Learning (RL). Numerical results confirm the effectiveness of the proposed RL scheme.

I. Introduction

The architecture of the recently launched fifth generation (5G) mobile system can leverage cloud processing at Base Band Units (BBUs), as well as edge processing, including edge caching, at enhanced Remote Radio Heads (eRRHs) [1]. In order to enable a flexible functional split in this architecture, referred to as Fog-Radio Access Network (F-RAN) [2], the concept of *X-haul* has been introduced to integrate the traditionally distinct backhaul and fronthaul connectivity

J. Moon and I. Lee are with the School of Electrical Engineering, Korea University, Seoul 02841, South Korea (e-mail: {anschino, inkyu}@korea.ac.kr). O. Simeone is with the Department of Informatics, King's College London, London WC2R 2LS, U.K. (e-mail: osvaldo.simeone@kcl.ac.uk). S.-H. Park is with the Division of Electronic Engineering, Chonbuk National University, Jeonju 54896, South Korea (e-mail: seokhwan@jbnu.ac.kr).

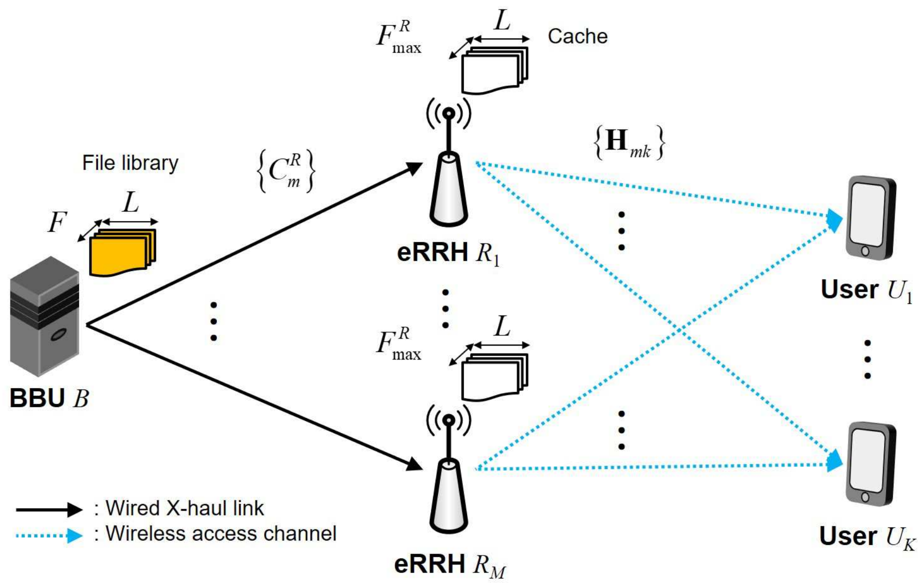

Fig. 1. Illustration of the F-RAN system under study

modes for the interface between the BBU and the eRRH into a unified framework [3]–[5]. The backhaul mode enables the transfer of data packets from the BBU in the cloud to the eRRHs. In contrast, the fronthaul mode allows the BBU to carry out joint baseband processing and deliver quantized baseband samples to the eRRHs as in Cloud-RAN (C-RAN) [6]–[8].

In this work, we study the adaptive selection of backhaul and fronthaul transfer modes with the aim of optimizing the performance of content delivery. The content delivery in F-RANs has been widely studied in recent years [9]–[15]. Most studies assume offline caching with a static popularity model. Under these assumptions, references [9] and [10] investigated the problem of instantaneous delivery latency minimization and minimum data rate maximization, respectively, while keeping the contents of the caches fixed. In contrast, in [11] and [12], information-theoretic performance bounds were provided on the optimal high Signal-to-Noise-Ratio (SNR) performance by considering also the optimization of uncoded caching strategies. An extension of this work that accounts for time-varying and possibly unknown file popularity with online caching was described in [13]. Under an unknown dynamic popularity model, the works [14] and [15] presented a Reinforcement Learning (RL) based optimization of online caching by assuming a backhaul mode.

In this paper, we investigate for the first time the online minimization of the long-term delivery latency over X-haul links in an F-RAN with time-varying and unknown file popularity. We focus on the joint optimization of linear precoding strategies and the choice between fronthaul and backhaul modes. The backhaul mode enables cache updates at the eRRHs, hence potentially reducing future latencies. In contrast, the fronthaul mode allows cooperative C-RAN transmissions which decrease the current delivery latency [9]–[11]. We propose a new model-free RL approach based on a linear value function approximation with properly selected features, and numerical results confirm the effectiveness of the proposed RL scheme.

Notations: $\mathbb{E}[\cdot]$ and $\Pr(\cdot)$ stand for expectation and probability, respectively. $|\mathcal{A}|$ represents the cardinality of set $\mathcal{A}$, and $\mathbb{C}^{m\times n}$ denotes an $m \times n$ complex matrix. $\mathbb{I}\{c\}$ outputs one if condition c is true and zero otherwise. For a matrix $\mathbf{X}$, $|\mathbf{X}|$, $\mathbf{X}^T$, $\mathbf{X}^H$, $\mathbf{X}^{-1}$ and $\mathrm{tr}(\mathbf{X})$ are defined as determinant, transpose, Hermitian, inverse and trace, respectively. $\mathbf{I}_m$ means an $m\times m$ identity matrix while $\otimes$ equals a Kronecker product operation. Also, $\mathrm{diag}\left(\mathbf{X}_1, ..., \mathbf{X}_N\right)$ represents block-wise diagonalization of matrices $\mathbf{X}_1, ..., \mathbf{X}_N$. Lastly, $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Omega})$ indicates a circularly symmetric complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Omega}$.

II. System Model

We study the F-RAN system illustrated in Fig. 1, which consists of a BBU in the cloud, connected to M cache-enabled eRRHs and K users. Each X-haul link between the BBU and the m-th eRRH has capacity C_m^R bits per symbols and can be operated in both backhaul and fronthaul modes [4] [5]. The k-th user and the m-th eRRH are equipped with N_k^U and N_m^R antennas, respectively. We assume a time-slotted operation [15], and the wireless channel matrix $\mathbf{H}_{mk}$ between the m-th eRRH and the k-th user is assumed to be constant for T_{B} time slots. We also define $\mathcal{F} \triangleq \{f_1, ..., f_F\}$ as the library of F L-bit files, which may be requested by the users. Finally, we denote $\mathcal{F}^R(t) \subseteq \mathcal{F}$ as the subset of files cached at time slot t at the eRRHs. This subset has a cardinality bounded by $F_{\max}^R$ files due to storage capacity constraints. Note that in this letter, we make a simplifying assumption that all the eRRHs store the same files in their respective caches. Generalization of the framework is possible but at the cost of a more cumbersome notation. Detailed request, online caching and delivery models are described in the following.

A. Request Model and Online Caching

In each time slot t, a subset $\mathcal{F}_{\mathrm{pop}}(t) \subseteq \mathcal{F}$ of files is popular in the sense that all users request files from $\mathcal{F}_{\mathrm{pop}}(t)$. Specifically, the k-th user requests a uniformly selected file $f_k^U(t)$ from subset $\mathcal{F}_{\mathrm{pop}}(t)$ without replacement [13]. The assumption of no replacement ensures that all requested files are distinct, yielding a worst-case performance analysis [11]. Let $\mathcal{K}_{\text{req,C}}(t)$ and $\mathcal{K}_{\text{req,NC}}(t)$ denote the indices of the users whose requested files $\mathcal{F}_{\text{req,C}}(t) \triangleq \left\{f_k^U(t)\right\}_{k\in\mathcal{K}_{\text{req,C}}(t)}$ are cached and the indices of users whose requested files $\mathcal{F}_{\text{req,NC}}(t) \triangleq \left\{f_k^U(t)\right\}_{k\in\mathcal{K}_{\text{req,NC}}(t)}$ are not cached at time t, respectively. In case the backhaul mode is selected at time slot t, the requested but uncached files in $\mathcal{F}_{\text{req,NC}}(t)$ are sent on all the X-haul links and cached. In order to make space for a new file, a previously cached file is evicted by following the standard Least Recently Used (LRU) rule [16].

B. Delivery Operation

At each slot t, the X-haul link is used in either fronthaul or backhaul mode for $\Delta^R(t, \mathrm{a}(t))$ symbols, where $\mathrm{a}(t) = 0$ and 1 indicate the selection of fronthaul and backhaul modes, respectively. Subsequently, the eRRHs deliver the requested files in set $\mathcal{F}_{\text{req}}(t) \triangleq \mathcal{F}_{\text{req,C}}(t) \cup \mathcal{F}_{\text{req,NC}}(t)$ over the wireless channel for $\Delta^U(t, \mathrm{a}(t))$ symbols, based on the signals received on the X-haul links and on the cached contents. This results in a total latency of $\Delta(t, \mathrm{a}(t)) = \Delta^R(t, \mathrm{a}(t)) + \Delta^U(t, \mathrm{a}(t))$ symbols for time slot t. Note that the eRRHs' caches are updated according to the caching mechanism described in Section II-A only if the backhaul mode is selected, i.e., if $\mathrm{a}(t) = 1$.

C. Problem Formulation

In this work, we aim at minimizing the average long-term delivery latency of the proposed F-RAN system over the selection of the delivery mode $\mathrm{a}(t)$. Given a forgetting factor $\gamma \leq 1$, the problem can be formulated as

$$\text{(P):} \quad \min_{\{\mathrm{a}(t)\}} \; \mathbb{E}_{\{\mathbf{H}_{mk}\}}\left[\sum_{t=1}^{\infty} \gamma^t \Delta(t, \mathrm{a}(t))\right] \tag{1a}$$

$$\text{s.t.} \quad \mathrm{a}(t) \in \{0, 1\}, \forall t, \tag{1b}$$

where the total latency $\Delta(t, \mathrm{a}(t))$ depends on the choice of the delivery mode and the optimized linear precoding as studied in [9] [10], which will be reviewed in Section III. In Section IV, we propose an RL-based approach to tackle problem (P) under the assumption that the popularity dynamics, defined by the evolution of the set $\mathcal{F}_{\text{pop}}(t)$, are unknown.

III. Minimum Instantaneous Latency

In this section, we review the instantaneous latency minimization at each time slot t given a X-haul delivery mode $\mathrm{a}(t)$ by following [9]. The time index t is omitted for simplicity.

A. Backhaul Mode

In the backhaul mode ($\mathrm{a}=1$), the BBU first fetches the requested but uncached files $\mathcal{F}_{\text{req,NC}}$ and transmits them to the eRRHs. The backhaul transmission to the m-th eRRH takes $\Delta_m^R = \left|\mathcal{F}_{\text{req,NC}}\right| L/C_m^R$ symbols, and the total backhaul latency is $\Delta^R = \max_m \Delta_m^R$, since all the eRRHs need to receive the files in $\mathcal{F}_{\text{req,NC}}$. As a result, all the requested files in $\mathcal{F}_{\text{req}}$ are available at the eRRHs and cooperative transmission across all eRRHs is feasible. Each file $f_k^U \in \mathcal{F}_{\text{req}}$ for the k-th user is encoded by each eRRH as the signal $\mathbf{s}_k \in \mathbb{C}^{n_k \times 1} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_{n_k}\right)$, where $n_k \leq N_k^U$ denotes the number of data streams allocated to the k-th user, which is assumed to be a fixed parameter. The transmit signal from the m-th eRRH is then given as $\mathbf{x}_m = \sum_{k\in\mathcal{K}_{\text{req}}} \mathbf{G}_{mk}\mathbf{s}_k$ where $\mathcal{K}_{\text{req}} \triangleq \mathcal{K}_{\text{req,C}} \cup \mathcal{K}_{\text{req,NC}}$, and $\mathbf{G}_{mk} \in \mathbb{C}^{N_m^R \times n_k}$ is the precoding matrix for $\mathbf{s}_k$ at the m-th eRRH. Accordingly, the achievable rate for the k-th user on the wireless channel can be written as [9]

$$R_{\text{back},k}^U\left(\{\mathbf{G}_k\}\right) = \log_2\left|\mathbf{I}_{N_k^U} + \mathbf{\Phi}_{\text{back},k}^U\right| \text{ [bits/symbol]}, \tag{2}$$

where we have $\mathbf{\Phi}_{\text{back},k}^U \triangleq \left(\sum_{\ell\in\mathcal{K}_{\text{req}}\setminus k} \mathbf{H}_k\mathbf{G}_\ell\mathbf{G}_\ell^H\mathbf{H}_k^H + \sigma_k^2\mathbf{I}_{N_k^U}\right)^{-1}\mathbf{H}_k\mathbf{G}_k\mathbf{G}_k^H\mathbf{H}_k^H$ with $\mathbf{H}_k \triangleq \left[\mathbf{H}_{1k}\cdots\mathbf{H}_{Mk}\right]$ and $\mathbf{G}_k \triangleq \left[\mathbf{G}_{1k}^T\cdots\mathbf{G}_{Mk}^T\right]^T$, and σ_k^2 represents the additive white Gaussian noise variance at the k-th user.

The latency Δ_k^U for delivering file f_k^U for the k-th user is obtained as $\Delta_k^U = L/R_{\text{back},k}^U\left(\{\mathbf{G}_k\}\right)$, and the overall wireless channel latency equals $\Delta^U = \max_k \Delta_k^U$, since every requesting user needs to receive the requested file. The minimum instantaneous latency Δ for $\mathrm{a}=1$ can hence be found as a solution of the problem

$$\text{(P1):} \quad \min_{\Delta^U, \{\mathbf{G}_k\}} \quad \Delta^R + \Delta^U \tag{3a}$$

$$\text{s.t.} \quad \Delta^U \geq \frac{L}{R_{\text{back},k}^U\left(\{\mathbf{G}_k\}\right)}, \ \forall k \in \mathcal{K}_{\text{req}}, \tag{3b}$$

$$\operatorname{tr}\left(\sum\nolimits_{k\in\mathcal{K}_{\text{req}}} \mathbf{E}_m\mathbf{G}_k\mathbf{G}_k^H\mathbf{E}_m^H\right) \leq P_m^R, m=1,...,M, \tag{3c}$$

where P_m^R denotes the maximum transmit power of the m-th eRRH, and we define $\mathbf{E}_m \triangleq \left[\mathbf{0}\cdots\mathbf{I}_{N_m^R}\cdots\mathbf{0}\right]$ in which an identity matrix $\mathbf{I}_{N_m^R}$ spans columns from $\sum_{\ell=1}^{m-1} N_\ell^R + 1$ to $\sum_{\ell=1}^{m} N_\ell^R$. Although problem (P1) is jointly non-convex, a local optimum can be attained by leveraging Successive Convex Approximation (SCA) as detailed in [9].

B. Fronthaul Mode

Under the fronthaul mode, any requested but uncached file $f_k^U \in \mathcal{F}_{\text{req,NC}}$ for the k-th user is jointly encoded and precoded at the BBU. The resulting signal dedicated for the m-th eRRH is written as $\hat{\mathbf{x}}_m = \sum_{k\in\mathcal{K}_{\text{req,NC}}} \mathbf{W}_{mk}\mathbf{s}_k$, where $\mathbf{s}_k \in \mathbb{C}^{n_k\times 1} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_{n_k}\right)$ encodes file f_k^U, and $\mathbf{W}_{mk} \in \mathbb{C}^{N_m^R\times n_k}$ represents the corresponding precoding matrix for the m-th eRRH. The BBU then performs compression on $\hat{\mathbf{x}}_m$ prior to transferring to the eRRHs. As a result, the decompressed signal at the m-th eRRH can be written by $\tilde{\mathbf{x}}_m = \hat{\mathbf{x}}_m + \mathbf{q}_m$ with quantization noise $\mathbf{q}_m \in \mathbb{C}^{N_m^R\times 1} \in \mathcal{CN}\left(\mathbf{0}, \mathbf{\Omega}_m\right)$ for a given covariance matrix $\mathbf{\Omega}_m$ [9] [10].

The rest of the requested cached files $\mathcal{F}_{\text{req,C}}$ are locally precoded with precoding matrices $\{\mathbf{G}_{mk}\}$ at the eRRHs in the same manner as in the backhaul mode. The final transmit signal at the m-th eRRH is then given as $\mathbf{x}_m = \sum_{k\in\mathcal{K}_{\text{req,C}}} \mathbf{G}_{mk}\mathbf{s}_k + \tilde{\mathbf{x}}_m$, and the achievable rate for the k-th user can be obtained as [9]

$$R_{\text{front},k}^U\left(\left\{\tilde{\mathbf{G}}_k\right\}, \mathbf{\Omega}_R\right) = \log_2\left|\mathbf{I}_{N_k^U} + \mathbf{\Phi}_{\text{front},k}^U\right| \text{ [bits/symbol]}, \tag{4}$$

where we have $\mathbf{\Phi}_{\text{front},k}^U \triangleq \left(\sum_{\ell\in\mathcal{K}_{\text{req}}\setminus k} \mathbf{H}_k\tilde{\mathbf{G}}_\ell\tilde{\mathbf{G}}_\ell^H\mathbf{H}_k^H + \mathbf{H}_k\mathbf{\Omega}_R\mathbf{H}_k^H + \sigma_k^2\mathbf{I}_{N_k^U}\right)^{-1}\mathbf{H}_k\tilde{\mathbf{G}}_k\tilde{\mathbf{G}}_k^H\mathbf{H}_k^H$, $\mathbf{\Omega}_R \triangleq \text{diag}\left(\mathbf{\Omega}_1, ..., \mathbf{\Omega}_M\right)$, $\tilde{\mathbf{G}}_k \triangleq \left[\tilde{\mathbf{G}}_{1k}^T \cdots \tilde{\mathbf{G}}_{Mk}^T\right]^T$ with $\tilde{\mathbf{G}}_{mk} \triangleq b_k^U\mathbf{G}_{mk} + \left(1 - b_k^U\right)\mathbf{W}_{mk}$, and $b_k^U = 1$ if $f_k^U \in \mathcal{K}_{\text{req,C}}$ and $b_k^U = 0$ otherwise for the k-th user.

The wireless channel latency Δ^U is defined in the same way as in the backhaul mode. For the fronthaul latency, by the rate-distortion theory, sending quantized signals to the m-th eRRH consumes

$$g_m\left(\left\{\tilde{\mathbf{G}}_k\right\}, \mathbf{\Omega}_R\right) = \log_2\left|\mathbf{I}_{N_m^R} + \mathbf{\Phi}_m^R\right| \text{ [bits/symbol]}, \tag{5}$$

with $\mathbf{\Phi}_m^R \triangleq \left(\mathbf{E}_m\mathbf{\Omega}_R\mathbf{E}_m^H\right)^{-1}\sum_{k\in\mathcal{K}_{\text{req,NC}}} \mathbf{E}_m\tilde{\mathbf{G}}_k\tilde{\mathbf{G}}_k^H\mathbf{E}_m^H$ [9]. Compressing Δ^U symbols produces $\Delta^U g_m \left(\left\{\tilde{\mathbf{G}}_k\right\}, \mathbf{\Omega}_R\right)$ bits, which need to be transferred from the BBU to the m-th eRRH. Therefore, the fronthaul latency is given by $\Delta^R = \max_m \Delta_m^R$ where $\Delta_m^R = \Delta^U g_m\left(\left\{\tilde{\mathbf{G}}_k\right\}, \mathbf{\Omega}_R\right)/C_m^R$, and the minimum instantaneous latency Δ for $\mathrm{a} = 0$ is calculated as a solution of the problem

$$\text{(P2):} \quad \min_{\Delta^R, \Delta^U, \{\tilde{\mathbf{G}}_k\}, \boldsymbol{\Omega}_R} \quad \Delta^R + \Delta^U \tag{6a}$$

$$\text{s.t.} \quad \Delta^R \geq \frac{\Delta^U g_m\left(\{\tilde{\mathbf{G}}_k\}, \boldsymbol{\Omega}_R\right)}{C_m^R}, m = 1, ..., M, \tag{6b}$$

$$\Delta^U \geq \frac{L}{R^U_{\text{front},k}\left(\{\tilde{\mathbf{G}}_k\}, \boldsymbol{\Omega}_R\right)}, \ \forall k \in \mathcal{K}_{\text{req}}, \tag{6c}$$

$$\text{tr}\Big(\sum\nolimits_{k \in \mathcal{K}_{\text{req}}} \mathbf{E}_m \tilde{\mathbf{G}}_k \tilde{\mathbf{G}}_k^H \mathbf{E}_m^H + \mathbf{E}_m \boldsymbol{\Omega}_R \mathbf{E}_m^H\Big)$$

$$\leq P_m^R, m = 1, ..., M, \tag{6d}$$

which can be tackled via the SCA approach detailed in [9].

IV. RL-Based X-Haul Online Optimization

In this section, we solve problem (P) by proposing an online on-policy RL-based optimization strategy [17].

A. Problem (P) as a Partially Observable Decision Process

Under the assumption that the popularity set $\mathcal{F}_{\text{pop}}(t)$ evolves as a Markov process along the time slot index t, problem (P) is a Markov Decision Process (MDP) [17] with state space $\mathrm{s}(t) = \{\mathcal{F}_{\text{pop}}(t), \, \mathcal{F}^R(t), \, \{\tau_{\text{req},f}(t)\}_{f \in \mathcal{F}^R(t)}\}$, where $\tau_{\text{req},f}(t)$ is the most recent time slot at which cached file f was requested at time slot t, the action space is obtained by $\mathrm{a}(t) = \{0, 1\}$, and the instantaneous reward is defined by the negative latency $\mathrm{r}(t+1) = -\Delta(t, \mathrm{a}(t))$.

The state $\mathrm{s}(t)$ is partially observable since the set $\mathcal{F}_{\text{pop}}(t)$ is unknown, and it is only observed indirectly via the file set $\mathcal{F}_{\text{req}}(t)$. In particular, at time t, only the history of observations $\mathrm{o}(1), ..., \mathrm{o}(t)$ with $\mathrm{o}(t) = \{\mathcal{F}_{\text{req}}(t), \, \mathcal{F}^R(t), \, \{\tau_{\text{req},f}(t)\}_{f \in \mathcal{F}^R(t)}\}$ is available to the system. Thus, a general policy can map the observations $\mathrm{o}(1), ..., \mathrm{o}(t)$ to a selected action $\mathrm{a}(t)$. In order to reduce the complexity of the policy, we optimize here over memoryless policies that select an action $\mathrm{a}(t)$ based only on the latest observation $\mathrm{o}(t)$ at time slot t. These policies have been successfully applied in many applications [18] [19].

B. SARSA with Linear Value Function Approximation

To optimize over memoryless policies, we adopt the online on-policy value-based strategy State-Action-Reward-State-Action (SARSA) with a carefully designed linear approximation [17].

The SARSA updates an action-value function, or Q-function, $q\left(o,a\right)$ that estimates the expected return $\mathbb{E}[G\left(t\right)|\mathrm{o}=o,\mathrm{a}=a]$ with $G\left(t\right)\triangleq\sum_{\tau=0}^{\infty}\gamma^{\tau}\mathrm{r}\left(t+\tau+1\right)$. Since the total size of the observation space $\mathrm{o}\left(t\right)$ in (P) grows exponentially with F, we propose a linear value function approximation $\hat{q}\left(o,a,\mathbf{w}\right)\triangleq\mathbf{w}^{T}\mathbf{x}\left(o,a\right)$, where $\mathbf{w}$ is a parameter vector to be learned, and $\mathbf{x}\left(o,a\right)$ denotes a feature vector representing the observation-action pair $\left(o,a\right)$ [17].

In order to determine a suitable feature vector, we first note that vector $\mathbf{x}\left(o,a\right)$ should contain sufficient information to quantify the value of caching for currently cached and requested files. Frequently requested files typically yield lower future latencies when cached, but an optimal choice should account not only for their popularity but also for their remaining *life time*, which is a duration that a file remains popular (see Sec. II of [20] for further discussion).

Based on these considerations, we introduce a variable $\phi_{\ell}\left(t\right)$ for every file $f_{\ell}\in\mathcal{F}$ as a function of the current observation $\mathrm{o}\left(t\right)$ at time slot t. We set it as $\phi_{\ell}\left(t\right)=1$ if $f_{\ell}\in\mathcal{F}_{\text{req,NC}}\left(t\right)$, $\phi_{\ell}\left(t\right)=2$ if $f_{\ell}\in\mathcal{F}^{R}\left(t\right)$ and $\phi_{\ell}\left(t\right)=0$ otherwise. Furthermore, we also include a variable $\theta\left(t\right)\triangleq t-\max_{f\in\mathcal{F}^{R}(t)}\tau_{\text{req},f}$ that measures the "age" of the currently cached files, that is, the maximum time elapsed since the last request of the cached files. We can quantize this variable by N_{Θ} ranges $\Theta_{1},...,\Theta_{N_{\Theta}}\subseteq\mathbb{R}^{+}$ with $\Theta_{i}\cap\Theta_{j}=\emptyset$ for all $i\neq j$ and $\bigcup\Theta_{i}=\mathbb{R}^{+}$. If the caches are up to date, the quantity $t-\tau_{\text{req},f}$ is small for all $f\in\mathcal{F}^{R}\left(t\right)$, and hence $\theta\left(t\right)$ is also small. Otherwise, if there exists any file $f\in\mathcal{F}^{R}\left(t\right)$ with large $t-\tau_{\text{req},f}$, a refresh of the caches may be required.

Using the variables introduced above, we define the feature vector $\mathbf{x}\left(o\left(t\right),a\left(t\right)\right)$ as

$$\mathbf{x}\left(o\left(t\right),a\left(t\right)\right)=\left[\boldsymbol{\phi}_{1}^{T}\left(t\right)\ \cdots\ \boldsymbol{\phi}_{F}^{T}\left(t\right)\ \boldsymbol{\theta}^{T}\left(t\right)\right]^{T}\otimes\mathbf{a}\left(t\right),\tag{7}$$

where we have used the one-hot encoded vectors $\boldsymbol{\phi}_{\ell}\left(t\right)\triangleq[\mathbb{I}\{\phi_{\ell}\left(t\right)=1\}\ \mathbb{I}\{\phi_{\ell}\left(t\right)=2\}\ \mathbb{I}\{\phi_{\ell}\left(t\right)=0\}]^{T}$, $\boldsymbol{\theta}\left(t\right)\triangleq[\mathbb{I}\{\theta\left(t\right)\in\Theta_{1}\}\ \cdots\ \mathbb{I}\{\theta\left(t\right)\in\Theta_{N_{\Theta}}\}]^{T}$ and $\mathbf{a}\left(t\right)\triangleq[\mathbb{I}\{\mathrm{a}\left(t\right)=0\}\ \mathbb{I}\{\mathrm{a}\left(t\right)=1\}]^{T}$. The feature vector $\mathbf{x}\left(o\left(t\right),a\left(t\right)\right)$ in (7) has dimension $2(N_{\Theta}+3F)$, which increases linearly in F and is hence significantly smaller than the size of the conventional look-up table-based SARSA. The effectiveness of the proposed feature vector $\mathbf{x}\left(\mathrm{o}\left(t\right),a\left(t\right)\right)$ will be verified in Section V.

The overall proposed procedure for solving (P) is summarized in Algorithm 1 where $\delta\left(t,\mathbf{w}\right)\triangleq\mathrm{r}\left(t+1\right)+\gamma\hat{q}\left(\mathrm{o}\left(t+1\right),\mathrm{a}\left(t+1\right),\mathbf{w}\right)-\hat{q}\left(o,a,\mathbf{w}\right)$ denotes the temporal difference error, $\mathbf{E}$ is the eligibility trace, which assigns credits to the most frequently visited states and selected actions (see [17] for details), and an ϵ-greedy exploration strategy with decreasing ϵ along the episodes is adopted.

Algorithm 1: Proposed RL-based solution for problem (P)
Initialize the total number of episodes N_{epi}, weight vector $\mathbf{w} = \mathbf{0}$, eligibility trace $\mathbf{E} = \mathbf{0}$, and parameter $\gamma, \lambda, \beta \in (0, 1]$
For $n_{\text{epi}} = 1 : N_{\text{epi}}$
Randomly initialize cached contents $\mathcal{F}^R(0)$ and generate channels $\{\mathbf{H}_{mk}\}$
For $t = 1 : T_{\text{B}}$
Collect observation $\mathrm{o}(t)$=$\{\mathcal{F}_{\text{req}}(t),\ \mathcal{F}^R(t), \{\tau_{\text{req},f}(t)\}_{f\in\mathcal{F}^R(t)}\}$
Choose the delivery mode greedily with probability $1–1/n_{\text{epi}}$ as $\mathrm{a}(t) = \arg\max_{a'} \mathbf{w}^T\mathbf{x}(\mathrm{o}(t), a')$, and uniformly with probability $1/n_{\text{epi}}$
If $\mathrm{a}(t) = 1$, update $\mathcal{F}_{\text{cache},R}(t)$ according to LRU
Set $\mathrm{r}(t+1) = -\Delta(t, \mathrm{a}(t))$
Update $\mathbf{E} \leftarrow \gamma\lambda\mathbf{E} + \mathbf{x}(o, a)$
Update $\mathbf{w} \leftarrow \mathbf{w} + \beta\delta(t, \mathbf{w})\mathbf{E}$
End
End

V. Numerical Results

In this section, the performance of the proposed RL-based algorithm is evaluated via numerical examples. We adopt the channel model $\mathbf{H}_{mk} = \sqrt{\rho_{mk}}\hat{\mathbf{H}}_{mk}$, where $\rho_{mk} \triangleq \rho_0\left(\frac{d_{mk}}{d_0}\right)^{-\eta}$ equals the distance-dependent path loss between eRRH R_m and user U_k, ρ_0 indicates the path loss at reference distance d_0, η is the path loss exponent, and d_{mk} represents the distance between the m-th eRRH and the k-th user. Each element of $\hat{\mathbf{H}}_{mk}$ follows an independent complex Gaussian distribution with zero mean and unit variance. The eRRHs and the users are circularly placed from the BBU at the center with uniformly distributed angles and distance $d_{BR} = 200$ m and $d_{BU} = 400$ m, respectively. The bandwidth is 20 MHz and the thermal noise is -170 dBm/Hz. We set $\rho_0 = 10^{-3}$, $d_0 = 1$ m, $\eta = 3.5$, $T_{\text{B}} = 100$ time slots, $F^R_{\max} = 4$ files, $P^R_m = 30$ dBm, $N^R_m = N^U_k = 1$ and $C^R_m = 0.1$ bits per symbol. For RL, we use the hyperparameters $\beta = 0.01$, $\gamma = 1$, $\lambda = 0.5$, and $\Theta_\ell = [2(\ell-1), \min(2(\ell-1)+1, \theta_{\max})]$ with $N_\Theta = 11$ where $\theta_{\max} = 20$ limits the maximum value of $\theta(t)$.

Reference [20] demonstrated that the popularity of files often exhibits temporal locality in

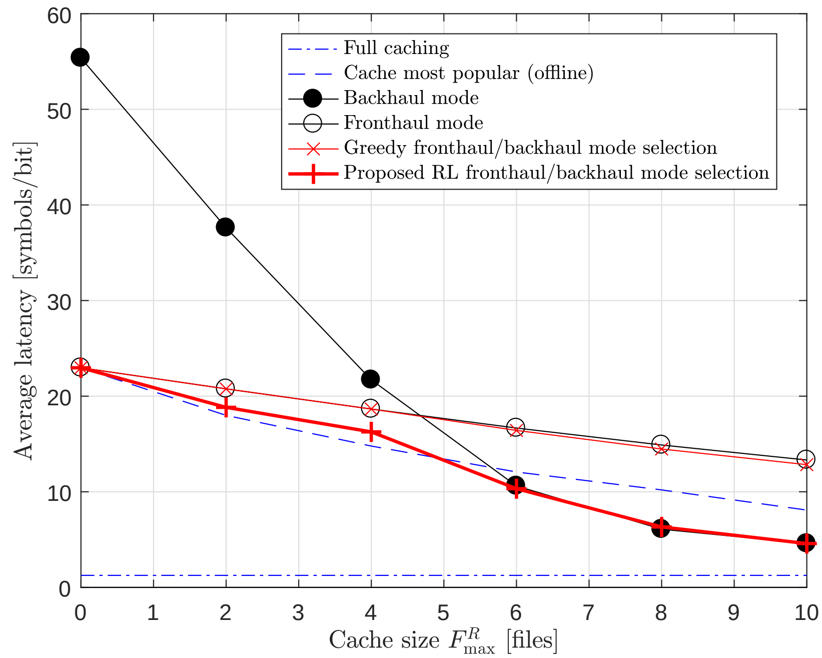

Fig. 2. Average latency with respect to the maximum cache size $F^R_{\max}$

the sense that the content is frequently requested in a bursty fashion for a certain life time. Motivated by these findings, we model the evolution of the subset $\mathcal{F}_{\text{pop}}(t)$ of popular files such that a currently unpopular file f has a probability of $\text{P}_{\text{pop},f}$ to become popular, and file f remains popular for $T_{\text{life},f}$ time slots. We assume Zipf's distribution [21] for $\text{P}_{\text{pop},f_\ell} = \ell^{-\xi}/\sum_{\nu=1}^{F}\nu^{-\xi}$ with $\xi = 1$. The proposed RL scheme is compared with a greedy fronthaul/backhaul mode selection that minimizes the current delivery latency at each time slot as well as with an offline scheme that keeps the $F^R_{\max}$ most popular files with the largest $\text{P}_{\text{pop},f}$ under the idealized assumption that this is known in prior.

Fig. 2 compares the average long-term latency performance as a function of the eRRHs' cache size $F^R_{\max}$ for $P^R_m = 30$ dBm, $T_{\text{life},f} = 10$ and $F = 20$. With $F^R_{\max} \leq 4$, the fronthaul mode is seen to yield a lower latency than the backhaul mode given the limited advantage of caching in this regime. The opposite is true when the eRRHs have larger caches, such as $F^R_{\max} > 4$, in which the backhaul mode outperforms the fronthaul mode. In agreement with the results in [9]–[11] and [13], the greedy scheme almost always selects the fronthaul mode and is hence strongly suboptimal for large enough $F^R_{\max}$. The proposed RL method exhibits the lowest latency among all schemes that do not assume the knowledge of the popularity probability. It can be checked that the gain is not obtained by statically selecting the best mode at each time instant, but rather by carrying out an optimized dynamic selection. It is also observed that in a large $F^R_{\max}$ regime, the proposed strategy can outperform the static offline scheme which assumes popularity dynamics to be known in advance.

VI. Conclusions

In this paper, we have demonstrated the advantage of adaptively selecting between the backhaul and fronthaul transfer modes as a function of the current cache contents and the history of past requests in an F-RAN system. The proposed RL-based strategy has been shown via numerical results to outperform baseline schemes, confirming the potential advantages of an X-haul implementation over static fronthaul or backhaul deployments.

References

[1] Y.-J. Ku, D.-Y. Lin, C.-F. Lee, P.-J. Hsieh, H.-Y. Wei, C.-T. Chou, and A.-C. Pang, "5G radio access network design with the fog paradigm: confluence of communications and computing," *IEEE Commun. Mag.*, vol. 55, pp. 46–52, Apr. 2017.

[2] Y.-Y. Shih, W.-H. Chung, A.-C. Pang, T.-C. Chiu, and H.-Y. Wei, "Enabling low-latency applications in fog-radio access networks," *IEEE Netw.*, vol. 31, pp. 52–58, Jan. 2017.

[3] A. D. L. Oliva, X. C. Prez, A. Azcorra, A. D. Giglio, F. Cavaliere, D. Tiegelbekkers, J. Lessmann, T. Haustein, A. Mourad, and P. Iovanna, "Xhaul: toward an integrated fronthaul/backhaul architecture in 5G networks," *IEEE Wireless Commun.*, vol. 22, pp. 32–40, Oct. 2015.

[4] T. Pfeiffer, "Next generation mobile fronthaul and midhaul architectures," *J. Opt. Commun. Netw.*, vol. 7, pp. 38–45, Nov. 2015.

[5] N. J. Gomes, P. Chanclou, P. Turnbull, A. Magee, and V. Jungnickel, "Fronthaul evolution: From CPRI to Ethernet," *Opt. Fiber Technol.*, vol. 26, pp. 50–58, Dec. 2015.

[6] H. Ren, N. Liu, C. Pan, M. Elkashlan, A. Nallanathan, X. You, and L. Hanzo, "Low-latency C-RAN: an next-generation wireless approach," *IEEE Veh. Technol. Mag.*, vol. 13, pp. 48–56, Jun. 2018.

[7] J. Kim, H. Lee, S.-H. Park, and I. Lee, "Minimum rate maximization for wireless powered cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 68, pp. 1045–1049, Jan. 2019.

[8] J. Kim, S.-H. Park, O. Simeone, I. Lee, and S. S. (Shitz), "Joint design of fronthauling and hybrid beamforming for downlink C-RAN systems," *accepted for IEEE Trans. Commun.*

[9] S.-H. Park, O. Simeone, and S. Shamai, "Joint cloud and edge processing for latency minimization in fog radio access networks," *in Proc. IEEE Int. Workshop Signal Adv. Wireless Comm. (SPAWC)*, Jul. 2016.

[10] S.-H. Park, O. Simeone, and S. Shamai, "Joint optimization of cloud and edge processing for fog radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 7621–7632, Nov. 2016.

[11] A. Sengupta, R. Tandon, and O. Simeone, "Fog-aided wireless networks for content delivery: fundamental latency Tradeoffs," *IEEE Trans. Inf. Theory*, vol. 63, pp. 6650–6678, Oct. 2017.

[12] J. Zhang and O. Simeone, "Fundamental limits of cloud and cache-aided interference management with multi-antenna edge nodes," [Online]. https://arxiv.org/abs/1712.04266.

[13] S. M. Azimi, O. Simeone, A. Sengupta, and R. Tandon, "Online edge caching and wireless delivery in fog-aided networks with dynamic content popularity," *IEEE J. Sel. Areas Commun.*, vol. 36, pp. 1189–1202, Jun. 2018.

[14] A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, "Optimal and scalable caching for 5G using reinforcement learning of space-time popularities," *IEEE J. Sel. Topics Signal Process.*, vol. 12, pp. 180–190, Feb. 2018.

[15] S. O. Somuyiwa, A. Gyrgy, and D. Gndz, "A reinforcement-learning approach to proactive caching in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, pp. 1331–1344, Jun. 2018.

[16] T. Johnson and D. Shasha, "2Q: A low overhead high performance buffer management replacement algorithm," *in Proc. 20th Int. Conf. Very Large Data Bases*, pp. 439–450, Sep. 1994.

[17] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction second edition*. The MIT Press, Oct. 2018.

[18] M. L. Littman, "Memoryless policies: theoretical limitations and practical results," *in Proc. International Conf. Simulation Adaptive Behavior*, pp. 238–245, Aug. 1994.

[19] Y. Li, B. Yin, and H. Xi, "Finding optimal memoryless policies of POMDPs under the expected average reward criterion," *Eur. J. Oper. Res.*, vol. 211, pp. 556–567, Jun. 2011.

[20] S. Traverso, M. Ahmed, M. Garetto, P. Giaccone, E. Leonardi, and S. Niccolini, "Temporal locality in today's content caching: why it matters and how to model it," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, pp. 5–12, Oct. 2013.

[21] D. M. W. Powers, "Applications and explanations of Zipf's law," *in Proc. Joint Conf. New Methods Lang. Process. Comput. Natural Lang. Learn.*, pp. 151–160, Jan. 1998.