Adapting Downlink Power in Fronthaul-Constrained Hierarchical Software-Defined RANs

Xianfu Chen, Zhu Han, Zheng Chang, Guoliang Xue, Honggang Zhang, and Mehdi Bennis

Abstract—The proof-of-concept software-defined radio access network (RAN) is not flexible enough due to the inherent delay and the necessity of high-capacity fronthaul links. We are hence motivated to propose a hierarchical software-defined RAN architecture, over which the base stations (BSs) are abstracted into multiple virtual local controllers while these local controllers are administered by a high-level controller. Under such a hierarchical network architecture, we particularly investigate in this paper how to adapt the BS transmit power over a long term according to the network dynamics under the constraints of mobile user queue stability and limited fronthaul capacity. We first formulate an off-line stochastic power adaptation problem. Through developing the Lyapunov method, we transform the problem into an approximate on-line optimization task. However, the challenge arises from the introduced per-cluster fronthaul capacity constraint. To solve the task efficiently and avoid extensive information exchange between the high-level controller and the local controllers, we put forward a novel low-complexity algorithm by designing a non-cooperative power adaptation game among the local controllers. Simulations are provided to evaluate the efficacy of the proposed studies.

I. INTRODUCTION

The exponentially growing mobile data traffic leads to the need of ever increasing capacity density in radio access networks (RANs) [1]. To keep pace with such demands, one of the promising solutions is to make the network infrastructure heterogeneous and dense. In a dense environment, the base stations (BSs) have to be operated over a common spectrum band, making the network operations extremely complex due to the tight coupling of control plane decisions at the neighbouring BSs. Moreover, the traditional RANs are dimensioned to cope with the peak traffic demands. Such designs are not flexible enough to match the radio resources with the spatially and temporally fluctuating traffic, resulting in low spectral efficiency and inferior energy efficiency as well [2].

Applying the idea of software-defined networking to RANs brings the immediate advantages of simplifying the management of a dense network. In a software-defined RAN, the control plane is decoupled from the data plane via virtualizing all independent BSs as a controller which makes centralized control plane decisions [3]. The centralized controller can thus optimize the network performance with the global view of the network and adapt radio resources to the network dynamics, i.e., the channel quality variations and the mobile traffic fluctuations from mobile users (MUs). However, the problems with the software-defined RAN concept lie in the inherent latency due to the totally centralized control plane and the need of large-capacity fronthaul links to connect the BSs and the controller [4], [5]. It thus becomes necessary to design a more flexible software-defined RAN architecture.

In this paper, we first consider a new hierarchical architecture of the software-defined RAN, which is shown in Fig. 1. Instead of virtualizing all BSs as a single centralized controller, multiple clusters are formed with regards to the BS geographic locations, with each being assigned a virtual local controller. The connections between the clusters and their associated local controllers are established via the capacity-limited fronthaul links [6]. A virtual high-level controller is responsible for coordinating control plane decisions among the local controllers. With the hierarchical design, the global network view at the high-level controller is aggregated across the local clusters. The local controllers make local decisions (e.g., intra-cluster interference mitigation), which avoids distant control from the high-level controller and thus alleviates control latency.

In spite of the benefits, challenges remain in facilitating a hierarchical software-defined RAN (HSDRAN) in practice. Particularly, mechanisms that efficiently utilize the decoupling of control plane and data plane in a hierarchical network architecture and the capacity-limited fronthaul links must be
developed to achieve stable data transmissions for MUs. There exists related work addressing part of the issues. In [7], the authors proposed a distributive stochastic learning algorithm for dynamic clustering to minimize the long-term average delay for MUs. The authors of [8], [9] developed resource allocation techniques to optimize the network performance considering the limited-capacity fronthaul constraints. The Lyapunov optimization framework has been adopted to achieve a suboptimal solution to a stochastic optimization constrained complexity algorithm is proposed. We elaborate on details of the system model that is considered in this paper. In Section III, we introduce the long-term average transmit power minimization in a HSDRAN as an off-line stochastic optimization problem and present the approximate solution to an on-line optimization problem, for which a low-complexity algorithm is proposed. The rest of the paper is organized as follows. In Section II, we elaborate on details of the system model that is considered in this paper. In Section III, we introduce the long-term average transmit power minimization in a HSDRAN as an off-line stochastic optimization problem and present the approximate on-line version. For the formulated on-line optimization, we propose a low-complexity algorithm in Section IV. In Section V, we provide preliminary simulation results to validate the proposed work. We conclude this paper in Section VI.

II. SYSTEM DESCRIPTIONS

As depicted in Fig. 1, we consider a HSDRAN, the heart of which constitutes a hierarchical structure of virtual controllers. The local controllers process baseband signals for MUs and optimize cluster-wide radio resource utilization. The analysis throughout this paper focuses on downlink communications. In the downlink, BSs communicate with MUs using RF signals, which carry the baseband signals and the precoding vectors that are determined by the local controllers. All BSs in the network share a common spectrum of bandwidth $W$ (Hz). The time dimension is partitioned into discrete time slots indexed by $t \in \mathbb{N}^+$ with equal duration $\delta$ (seconds). Let $\mathcal{K}$ be the set of local controllers/clusters. In each cluster $k \in \mathcal{K}$, a set $\mathcal{B}_k$ of BSs serve a set $\mathcal{I}_k$ of single-antenna MUs. For notational convenience, we assume that every BS is equipped with $N$ antennas.

We denote $\rho_{bk}^{b_k}(t) \in \mathbb{R}_+$ and $u_{bk}^b(t) \in \mathbb{C}^{N \times 1}$, respectively, as the allocated transmit power and the precoding vector of BS $b_k \in \mathcal{B}_k$ for MU $i_k \in \mathcal{I}_k$ at each time slot $t$, where the precoding vector $u_{bk}^b(t)$ has unit norm, i.e., $\|u_{bk}^b(t)\| = 1$. The baseband signal $y_{ik}(t)$ received by MU $i_k$ at time $t$ can thus be expressed by (1), where $h_{bk}^b(t) \in \mathbb{C}^{N \times 1}$ denotes the vector of channels between antennas of BS $b_k$ and MU $i_k$, $x_{ik}(t) \in C^{N(0,1)}$ denotes the transmitted signal for MU $i_k$, $z_{ik}(t) \in C^{N(0,\sigma^2)}$ denotes the additive white noise, and the superscript $^H$ denotes the Hermitian transpose operation. The channels hold their states for the duration of one time slot, and potentially change states at the time slot boundaries in an independent and identically distributed (i.i.d.) manner. Based on the instantaneous local cluster channel state information, i.e., $\mathbf{H}_{ik}(t) = \{h_{bk}^b(t) : b_k \in \mathcal{B}_k, i_k \in \mathcal{I}_k\}$, local controller $k$ adopts zero-forcing (ZF) precoding to cancel the intra-cluster interference. Specifically, the collection of local cluster ZF precoding vectors $\mathbf{u}_{ik}(t) = \{u_{bk}^b(t) : b_k \in \mathcal{B}_k, i_k \in \mathcal{I}_k\}$ should satisfy: $(h_{bk}^b(t))^H u_{bk}^b(t) = 0$, if $b'_k \neq i_k$. Assuming that the transmitted signals and the noise are mutually independent, we derive the signal-to-interference-plus-noise ratio (SINR) at MU $i_k$ during the time slot as follows

$$\text{SINR}_{ik}(t) = \frac{\sum_{b_k \in \mathcal{B}_k} \rho_{bk}^{b_k}(t) \left| (h_{bk}^b(t))^H u_{bk}^b(t) \right|^2}{M_{ik}(t) + \sigma^2},$$

where

$$M_{ik}(t) = \sum_{k' \in \mathcal{K}, k' \neq k} \sum_{b_{k'} \in \mathcal{B}_{k'}} \left| \sum_{b_k \in \mathcal{B}_k} \rho_{bk}^{b_k}(t) (h_{bk}^b(t))^H u_{bk}^b(t) \right|^2,$$

is the power of received inter-cluster interference. The corresponding data rate is given by

$$R_{ik}(t) = W \log_2(1 + \text{SINR}_{ik}(t)).$$

The transmit power consumed by a BS $b_k$ during time slot $t$ can thus be computed as

$$P_{bk}(t) = \sum_{i_k \in \mathcal{I}_k} \rho_{bk}^{b_k}(t).$$

At each local controller $k \in \mathcal{K}$, there are $|\mathcal{I}_k|$ separate downlink queues which buffer data bits for MUs in the cluster, where $|\mathcal{I}|$ is the cardinality of the set $\mathcal{I}$. Let $Q_{ik}(t) \in \{0\} \cup \mathbb{N}^+$ be the queue length (number of bits) of a MU $i_k$ at the beginning of each time $t$, its queue evolution can be expressed as in the form below,

$$Q_{ik}(t + 1) = [Q_{ik}(t) - \delta R_{ik}(t), 0]^+ + A_{ik}(t),$$

where $[q]^+$ is the projection of $q$ into a non-negative area and $A_{ik}(t)$ is the number of data bits arriving at the end of time slot $t$. The data arrivals get queued until they are transmitted. Like $h_{bk}^b(t)$, we assume that $A_{ik}(t)$ is i.i.d. across time slots according to a general distribution $\text{Pr}\{A_{ik}(t)\}$.

Next, we shall address the data transmissions from local controllers to BSs over the fronthaul links. It’s worth noting that in a cluster $k \in \mathcal{K}$, $\rho_{bk}^{b_k}(t) = 0$ implies that BS $b_k \in \mathcal{B}_k$ does not serve MU $i_k \in \mathcal{I}_k$ during time slot $t$. In contrast, if $\rho_{bk}^{b_k}(t) \neq 0$, the fronthaul link between local controller $k$ and BS $b_k$ needs to carry the baseband signal $x_{ik}(t)$ for MU $i_k$. 

...
\[
\begin{align*}
y_{ik}(t) &= \left( \sum_{b_k \in B_k} \sqrt{P_{i_k}^{b_k}(t)} \left( h_{i_k}^{b_k}(t) \right)^H u_{i_k}^{b_k}(t) \right) x_{ik}(t) + \sum_{i'_{ik} \in I_k \setminus \{ik\}} \left( \sum_{b_k \in B_k} \sqrt{P_{i'_{ik}}^{b_k}(t)} \left( h_{i'_{ik}}^{b_k}(t) \right)^H u_{i'_{ik}}^{b_k}(t) \right) x_{i'_{ik}}(t) + z_{ik}(t) \\
&= \text{the desired signal} + \text{the intra-cluster interference} + \text{the inter-cluster interference plus the noise}
\end{align*}
\]

As a result, the allocated fronthaul capacity for BS \( b_k \) can be mathematically written as

\[
C_{b_k}(t) = \sum_{i_k \in I_k} \theta_{i_k}^{b_k} \left( \rho_{i_k}^{b_k}(t) \right) R_{i_k}(t),
\]

where

\[
\theta_{i_k}^{b_k} \left( \rho_{i_k}^{b_k}(t) \right) = \begin{cases} 1, & \text{if } \rho_{i_k}^{b_k}(t) > 0; \\ 0, & \text{otherwise}, \end{cases}
\]

is defined as a BS-MU association indicator function.

### III. Problem Formulation

With the aforementioned concerns for a HSDRAN, we are particularly interested in investigating how much total BS transmit power is needed in the long term to maintain the queue stability for all MUs, while taking into account the limited fronthaul capacity constraint. Formally, we formulate the following optimization problem:

\[
\begin{align}
& \min_{\{ \rho(t) : t \in \mathbb{N}^+ \}} \sum_{k \in K} \sum_{b_k \in B_k} \mathcal{P}_{b_k} \\
& \quad \text{s.t. } \forall k \in K, \\
& \quad \quad \overline{A}_{i_k} \leq \delta \mathcal{P}_{i_k}, \forall i_k \in I_k; \tag{9b}
& \quad \quad \sum_{b_k \in B_k} C_{b_k}(t) \leq C_{\max}; \forall t; \tag{9c}
& \quad \quad P_{b_k}(t) \leq P_{\max}, \forall b_k \in B_k, \forall t; \tag{9d}
& \quad \quad \rho_{i_k}^{b_k}(t) \geq 0, \forall b_k \in B_k, \forall i_k \in I_k, \forall t. \tag{9e}
\end{align}
\]

In (9), \( \rho(t) = \{ \rho_{i_k}^{b_k} : k \in K, b_k \in B_k, i_k \in I_k \} \) is the power adaptation profile for all MUs in the network at time slot \( t \), \( \mathcal{P}_{b_k} = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^\tau E[P_{b_k}(t)] \), \( \overline{A}_{i_k} = \limsup_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^\tau E[A_{i_k}(t)] \), \( R_{i_k} = \liminf_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^\tau E[R_{i_k}(t)] \), and \( C_{\max} \) and \( P_{\max} \) are the capacity limit for the fronthaul links and the maximum transit power for the BSs. We refer to \( \lambda_{i_k} = \overline{A}_{i_k} \) as the mean data arrival rate of MU \( i_k \). Note that (9b) is the condition of maintaining queue stability for all MUs in the network.

Previous approaches usually solve the problem in (9) based on the Markov decision process framework \([12]\) with complete network statistics \([14]\) and suffer from the curse of dimensionality \([7]\). The introduced constraints of queue stability (9b) and limited fronthaul capacity (9c) make the problem solving even more challenging. Inspired by the penalty-plus-drift method \([13]\), we transform the off-line problem in (9) into an on-line optimization problem by minimizing the bound of one-step Lyapunov drift. Let us define the Lyapunov function as

\[
L(t) = \frac{1}{2} \sum_{k \in K} \sum_{i_k \in I_k} (Q_{i_k}(t))^2, \tag{10}
\]

and the one-step conditional Lyapunov drift as

\[
\Delta(t) = E[L(t+1) - L(t)|Q(t)], \tag{11}
\]

where \( Q(t) = \{ Q_{i_k}(t) : k \in K, i_k \in I_k \} \) is the global queue state information at each time slot \( t \). Moreover, the drift-plus-penalty function can be defined by

\[
\Delta_V(t) = \Delta(t) + V E \left[ \sum_{k \in K} \sum_{b_k \in B_k} P_{b_k}(t) | Q(t) \right], \tag{12}
\]

where \( V \) is a non-negative weight that trades off the performance. Suppose that for each MU, \( A_{i_k}(t) \) and \( R_{i_k}(t) \) are upper bounded by \( A_{\max} \) and \( R_{\max} \), respectively. By squaring both sides of the queue evolution equation (6) and taking the expectation with respect to \( Q(t) \), we deduce

\[
\Delta_V(t) \leq B + V E \left[ \sum_{k \in K} \sum_{b_k \in B_k} P_{b_k}(t) | Q(t) \right] - \sum_{k \in K} \sum_{i_k \in I_k} Q_{i_k}(t) E[\delta R_{i_k}(t) - A_{i_k}(t)|Q(t)], \tag{13}
\]

where \( B = (1/2) \sum_{k \in K} \sum_{i_k \in I_k} ((R_{\max}^0)^2 + (A_{\max})^2) \). Accordingly, the original transmit power minimization problem in (9) is transformed into solving the following optimization problem at each time slot \( t \),

\[
\begin{align}
& \min_{\rho(t)} V \sum_{k \in K} \sum_{b_k \in B_k} P_{b_k}(t) - \sum_{k \in K} \sum_{i_k \in I_k} \delta Q_{i_k}(t) R_{i_k}(t) \\
& \quad \text{s.t. constraints (9c), (9d) and (9e), } \forall k \in K. \tag{14}
\end{align}
\]

The key challenge involved in solving above problem arises from the limited fronthaul capacity constraint (9c) for each
cluster. In next section, we aim at designing efficient algorithms for (14) since the convergence speed is a main concern for the per-time slot optimization.

IV. LOW-COMPLEXITY ALGORITHM

It can be noticed that the formulated per-time slot optimization problem in (14) is with a non-convex structure due to the introduction of a BS-MU association indicator variable (8), and is considerably different from the problem studied in [15] where the BS-MU association and the radio resource allocation are independent. Without the fronthaul capacity constraint (9c), the optimal solution can be efficiently obtained [16], [17]. For (14), a brute force approach can be applied to achieve the optimal power adaptation. Yet such a method is computationally exhaustive. This motivates us to develop a low-complexity algorithm, which is the focus of this section. In brief, our idea is first to adapt the transmit powers of BSs for a specific set of active BS-MU associations, where all MUs are initially assumed to be served by all BSs in the same cluster (i.e., \( \theta_{ik}^b = 1 \), \( \forall b \in B_k \), \( \forall i \in I_k \)). If the fronthaul capacity constraint (9c) is violated, some of active BS-MU associations are then removed. This process iterates until convergence. In the following, we drop the time slot index \( t \) to avoid redundancy.

Specifically, let \( B_{ik}^{(\ell)} \subseteq B_k \) be the set of BSs that serve a MU \( i_k \in I_k \) in cluster \( k \in K \) at the \( \ell \)-th iteration with \( \ell \in \mathbb{N}_+ \) and \( B_{ik}^{(1)} = B_k \). Whilst to decouple the power adaptation from the BS-MU association, we make another fundamental assumption that the transmit powers allocated to MU \( i_k \) by BSs from \( B_{ik}^{(\ell)} \) are identical, i.e., \( \rho_{ik}^b = \rho_{i_k} \in \mathbb{R}_+ \), \( \forall b \in B_{ik}^{(\ell)} \). We then solve

\[
\min_{\rho_k} \sum_{k \in K} \sum_{b_k \in B_k} P_{bk} - \sum_{k \in K} \sum_{i_k \in I_k} \delta Q_{ik}^b(t) R_{ik} \quad (15a)
\]

s.t. \( \forall k \in K \),

\[
\rho_{ik}^b = 0, \quad \forall b_k \notin B_{ik}^{(\ell)}, \forall i_k \in I_k; \quad (15b)
\]

\[
\rho_{ik}^b = \rho_{i_k}, \quad \forall b_k \in B_{ik}^{(\ell)}, \forall i_k \in I_k; \quad (15c)
\]

constraints (9d) and (9e),

through the dual decomposition method for each iteration. The fronthaul capacity constraint (9c) can be rewritten as

\[
\sum_{b_k \in B_k} \sum_{i_k \in I_k} R_{ik} \leq C_{\text{max}}, \quad (16)
\]

for cluster \( k \in K \), where \( T_{b_k}^{(\ell)} \) denotes the set of MUs that are served by BS \( b_k \) at \( \ell \)-th iteration. Towards this end, we define the Lagrangian function of the primal problem (15) as

\[
J(\rho; \alpha) = V \sum_{k \in K} \sum_{b_k \in B_k} P_{bk} - \sum_{k \in K} \sum_{i_k \in I_k} \delta Q_{ik}^b R_{ik} + \sum_{k \in K} \sum_{b_k \in B_k} \alpha_{bk} (P_{bk} - P_{max}), \quad (17)
\]

where \( \alpha = \{ \alpha_{bk} : k \in K, b_k \in B_k \} \) is a set of the non-negative Lagrangian multipliers (LMs) associated with the maximum transmit power constraint (9d) and the boundary constraint (9e) is absorbed into the Karush-Kuhn-Tucker (KKT) conditions. The dual problem of the primal problem in (15) is given by

\[
\max \gamma \text{G}(\alpha), \quad (18)
\]

where

\[
\text{G}(\alpha) = \min \rho J(\rho; \alpha). \quad (19)
\]

Solving (18) directly leads to the optimal solution to (15).

For a given \( \alpha \), the KKT conditions are necessary and sufficient for (19). By rearranging (17), we have an aggregated version of the Lagrangian function as

\[
J(\rho; \alpha) = \sum_{k \in K} \left( \sum_{i_k \in I_k} J_{ik}(\rho; \alpha_k) + \sum_{b_k \in B_k} \alpha_{bk} P_{max} \right), \quad (20)
\]

which incorporates constraints (15b) and (15c). This enables us to solve (15) in a distributed way, hence alleviating the information exchange overheads between high-level controller and local controllers. We formulate the power adaptation among the clusters as a non-cooperative game, in which the local controllers are the players, the action of each local controller \( k \in K \) is the BS power adaptations \( \rho_k = \{ \rho_{bk} : b_k \in B_k, i_k \in I_k \} \) and the payoff function is defined as \( J_k(\rho_k, \rho_{-k}; \alpha_k) = \sum_{i_k \in I_k} J_{ik}(\rho_k, \rho_{-k}; \alpha_k) + \sum_{b_k \in B_k} \alpha_{bk} P_{max} \), where \( \rho_{-k} = \{ \rho_{bk}^* : k' \in K \setminus \{ k \} \} \). The non-cooperative power adaptation game can be characterized by

\[
\min_{\rho_k} J_k(\rho_k, \rho_{-k}; \alpha_k), \quad \forall k \in K. \quad (22)
\]

In cluster \( k \), the power adapted by a BS \( b_k \in B_{ik}^{(\ell)} \) for serving a MU \( i_k \in I_k \) can be expressed as a function of \( \rho_{-k} \) and \( \alpha_k \), namely,

\[
\rho_{ik} = \varphi_{ik}(\rho_{-k}; \alpha_k) = \left| \frac{\{b_k \in B_{ik}^{(\ell)}\}}{\ln(2)} \left( 1 + \alpha_{bk} \right) \right| \exp \left( \frac{\Omega_{ik}(\rho_k)}{\sum_{b_k \in B_{ik}^{(\ell)}}^\Omega(\rho_{bk})} \right), \quad (23)
\]

where \( \Omega_{ik}(\rho_{-k}) = \sum_{b_k \in B_{ik}^{(\ell)}} (\rho_{bk}^2 + \rho_{bk}^2)/(2\rho_{bk}^2) \). The Nash equilibrium solution to (22) can be expressed as \( \rho_{bk} = \varphi_{ik}(\rho_{-k}; \alpha_k), \forall k \in K, \forall i_k \in I_k \). The optimal LMs can be found using an incremental-update based sub-gradient method [18], which in general converges slowly. From (23), the power adaptation is a strictly decreasing function of the LMs. The bisection approach [10] can be accordingly explored to find the LMs that ensure (9d). The power adaptation at each \( \ell \)-th iteration is summarized in Algorithm 1.

The outcomes from Algorithm 1 do not take into account the fronthaul capacity constraint (16). If the power adaptation at the \( \ell \)-th iteration satisfies the fronthaul capacity constraint (16) for all clusters, the process terminates. Otherwise, if the
Algorithm 1 Power adaptation at each local controller $k \in \mathcal{K}$ during each $\ell$-th iteration

1: initialize $\alpha_{k,\text{min}} = \{\alpha_{b_k,\text{min}} : b_k \in B_k\}$ and $\alpha_{k,\text{max}} = \{\alpha_{b_k,\text{max}} : b_k \in B_k\}$ for local controller $k$.
2: repeat
3: Calculate the transmit power $\rho_{ik}$ allocated by BSs in $B_{ik}^{(\ell)}$ to each MU $i_k \in I_k$ according to (23).
4: Calculate the total transmit power $P_{b_k}$ of each BS $b_k \in B_k$ across MUs in $I_{b_k}^{(\ell)}$ according to (5).
5: For a BS $b_k \in B_k$, if $P_{b_k} \leq P_{\text{max}}$, set $\alpha_{b_k} = \alpha_{b_k,\text{min}}$; otherwise, update $\alpha_{b_k}$ using the bisection approach.
6: until Convergence

The problem left is how to select a BS-MU association to be shut down. In this paper, based on the power adaptation results at $\ell$-th iteration, we propose a criterion in the following form

$$(b_{ik}^{(\ell)}, i_k^{(\ell)}) = \arg \min_{(b_{ik}, i_k) \in B_{ik}^{(\ell)} \times I_k} \left\{ J_{ik} - J_{ik}(\rho_{ik}; \alpha_k) \right\},$$

where

$$J_{ik} = \sum_{b_k' \in B_{ik}^{(\ell)} \setminus \{b_k\}} \left( V + \alpha_{b_k'} \rho_{ik} - \delta Q_{ik} \tilde{R}_{ik} - b_k \right).$$

with

$$\tilde{R}_{ik} = W \log_2 \left( 1 + \frac{\rho_{ik} \left( \sum_{b_k' \in B_{ik}^{(\ell)} \setminus \{b_k\}} (h_{ik}'^{b_k'})^H u_{ik}'^{b_k'} \right)^2}{M_{ik} + \sigma^2} \right).$$

Algorithm 2 Fronthaul capacity constrained power adaptation at each time slot $t$

1: initialize $B_{ik}^{(1)} = B_k$ for all $i_k \in I_k$ in each cluster $k \in \mathcal{K}$.
2: repeat
3: Solve Problem (15) using Algorithm 1, obtain power adaptation $\rho_{ik}$ for each MU $i_k \in I_k$.
4: If fronthaul capacity constraint (16) does not hold, update $B_{ik}^{(\ell+1)}$ according to (24) and set $\ell = \ell + 1$.
5: until Convergence
6: Set the optimal power adaptation at time $t$ to be $\rho^*(t) = \rho$.

We use the Pareto distribution to simulate the traffic arrivals at the MU side [20]. More specifically, the probability that $\alpha$ data bits arrive during a time slot can be expressed as

$$f(\alpha) = \begin{cases} \frac{\zeta \nu^\zeta}{\alpha^{\zeta+1}}, & \text{if } 0 < \alpha < q; \\ \nu, & \text{if } \alpha \geq q, \end{cases}$$

where $\zeta$ is the shape factor, $\nu$ is the mode, $q$ is the cutoff threshold, and $\nu$ is calculated as

$$\nu = \left( \frac{q}{\zeta} \right)^{\frac{1}{\zeta}}.$$
problem. To solve the non-convex per time slot optimization and avoid extensive information exchange between the high-level controller and the local controllers, we propose a low-complexity algorithm by formulating a non-cooperative power adaptation game among the local controllers. Preliminary simulation experiments illustrate that our proposed algorithm can achieve optimal required transmit power while the MU queue stability being ensured.

VI. CONCLUSIONS

In this paper, we investigate a hierarchical architecture design of the software-define RANs and the problem of long-term downlink power adaptation in an HDRAN. The introduced queue stability constraint for MUs and limited fronthaul capacity constraint for each cluster make the problem solving extremely challenging. We show that the drift-plus-penalty method can achieve the queue stability for all MUs and the sum transmit power minimization of BSs across different clusters and thereby transforms the off-line stochastic transmit power adaptation into an on-line per time slot optimization problem. To solve the non-convex per time slot optimization and avoid extensive information exchange between the high-level controller and the local controllers, we propose a low-complexity algorithm by formulating a non-cooperative power adaptation game among the local controllers. Preliminary simulation experiments illustrate that our proposed algorithm can achieve optimal required transmit power while the MU queue stability being ensured.

REFERENCES