Wireless Multicast for Cloud Radio Access Network with Heterogeneous Backhaul

Ya-Feng Liu
LSEC, ICMSEC, Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing, China
Email: yafliu@lsec.cc.ac.cn

Wei Yu
Electrical and Computer Engineering Department
University of Toronto, Canada
Email: weiyu@ece.utoronto.ca

***Abstract*—Cooperative communication in which multiple base-stations (BSs) jointly transmit to mobile users is a major advantage of the cloud radio-access network (C-RAN) architecture. This paper considers the use of wireless multicast for sharing user messages at multiple BSs in the C-RAN backhaul for cooperation purpose. To combat fading and path-loss in the wireless channels to the different BSs, this paper proposes the optimal provisioning of secondary backhaul to smooth out the channel disparity. The paper analyzes the problem structure and proposes an efficient algorithm for the optimization of secondary backhaul rate provisioning.**

I. Introduction

Cloud radio-access network (C-RAN) [1] is an emerging network architecture for wireless cellular networks in which the base-stations (BSs) are connected to and coordinated from a centralized cloud data centre, and are capable of jointly transmitting user messages in a synchronized fashion from across multiple BSs in the downlink, thereby benefiting the overall transmission by mitigating inter-cell interference. Crucial to the deployment of C-RAN is the provisioning of the high-speed links between the BSs and the cloud. These data links are referred to as *fronthaul*, if the BSs reduce their roles to analog front-end processing only and the cloud performs all the digital functionality, or as *backhaul*, if the user data messages are directly sent to the multiple BSs for joint encoding/decoding. This latter scenario, where the links between the cloud and the BSs carry the data messages directly, is the focus of this paper. The fronthaul/backhaul links of C-RAN are likely to require at least several times more capacity than that of traditional cellular networks. In fact, fronthaul/backhaul provisioning can often be a bottleneck in the overall network planning, especially for small-cell deployments in which reliable and high-speed connections may not always be available to every small-cell BSs.

This paper considers the use of wireless backhaul for C-RAN, not only because it is much easier to deploy when fixed wireline infrastructure is not available, but also because of the crucial *wireless multicast* advantage that allows the efficient delivery of user messages to multiple BSs at the same time. Wireless multicast is ideally suited for enabling the cooperative transmission benefit of C-RAN; but it also brings in the challenge of pathloss, fading, and shadowing effect of the wireless medium. In particular, because of the different locations of the BSs, there may be considerable disparity in the quality of their respective channels.

To deal with this issue of channel disparity in wireless multicast, this paper proposes the use of secondary backhaul to smooth out the difference in channel quality across the BSs. A secondary backhaul can be a capacity constrained digital subscribe line (DSL) connection that can be used to augment the wireless backhaul. (As an alternative, user content may also be pre-fetched in capacity-limited local storage to aid the BSs with weak channels.) The main contribution of this paper is to show that the optimization of the capacity allocation of the secondary backhaul can be done efficiently by taking advantage of the problem structure. Although the use of network coding with multicast [2] and the use of cache for cooperation [3] have been considered in the C-RAN literature, the possibility to smooth out channel disparity for multicast operation is a novel concept that has not been explored previously.

II. Problem Formulation

Consider a C-RAN consisting of K BSs connected to the central cloud processing center in a wireless backhaul as illustrated in Fig. 1. The transmit signal of the central processor is denoted as X; the received signals at the BSs are denoted as $Y_1, Y_2, \ldots, Y_K$. We assume a block fading model with channel state information at both the transmitter and the receivers, for which the achievable multicast rate for sending a common user message to all the BSs in each fading block (assuming a fixed input distribution) can be characterized as:

$$\min_{k \in \mathcal{K}} I(X; Y_k), \tag{1}$$

where $\mathcal{K} := \{1, \ldots, K\}$. The transmission of common message subsequently allows the BSs to cooperate in relaying the message to the users. Due to the path-loss, shadowing and channel fading, there is considerable disparity in $I(X; Y_k)$'s. The common message rate is constrained by the weakest channel in each block.

This paper considers the use of secondary backhaul links (or caching storage) at the BSs to boost the above common message rate. We model the secondary backhaul as having capacities C_k for $k \in \mathcal{K}$, so that the achievable common

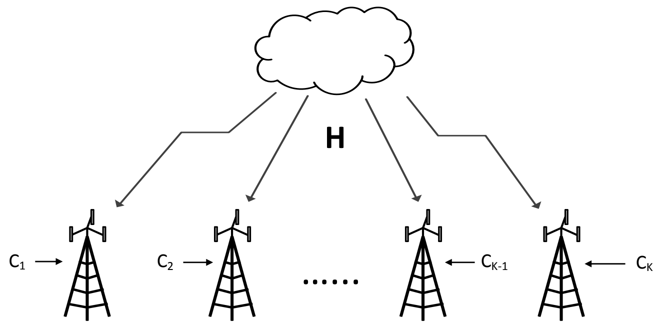

Fig. 1. Wireless multicast for C-RAN with secondary backhaul

information transmission rate in each fading block is improved to

$$R_0 < \min_{k \in \mathcal{K}} \{I(X; Y_k) + C_k\}. \tag{2}$$

(The effect of caching for providing side information can potentially be equivalently modeled in the same way.) The achievability of the above rate from a channel coding perspective can be understood as follows. To encode a common message, a random codeword with the above rate is generated and transmitted through the channels. The secondary backhaul links can be used to transmit additional parity bits of the common message codeword. Information theoretical consideration reveals that based on the received signal Y_k and the parity bits of rate C_k, successful decoding can be guaranteed at the rate R_0 in (2).

This paper aims to solve the following problem. Suppose that we have a fixed total secondary backhaul rate (or total cache storage size) C, how should it be allocated across the K BSs so as to maximize the overall common information transmission rate R_0? Secondary backhaul (or cache allocation) needs to take place at large time scale, so it can only be adapted to the statistics of the channels but not the instantaneous channel realization in each fading block. Let

$$R_k = I(X; Y_k) \tag{3}$$

be a random variable that depends on the channel realization $\mathbf{H}$ in each fading block. Define

$$r(C_1, \ldots, C_K) = \mathbb{E}_{\mathbf{H}} \left[\min_{k \in \mathcal{K}} \{R_k + C_k\} \right]. \tag{4}$$

The secondary backhaul rate allocation problem is

$$\begin{aligned} \text{maximize} \quad & r(C_1, \ldots, C_K) \\ \text{subject to} \quad & \sum_{k \in \mathcal{K}} C_k \le C, \\ & C_k \ge 0, \ \forall k \in \mathcal{K}. \end{aligned} \tag{5}$$

The optimization of C_k's accounts for the statistics of R_k. We assume that the probability density function (PDF) and the cumulative distribution function (CDF) of R_k, which can be derived from the channel distribution, are known, and are denoted by $\{f_k(x)\}_{k \in \mathcal{K}}$ and $\{F_k(x)\}_{k \in \mathcal{K}}$, respectively. Without loss of generality, we assume that all $\{f_k(x)\}_{k \in \mathcal{K}}$ are continuous in this paper. It is simple to verify that the PDF of the random variable $\min_{k \in \mathcal{K}} \{R_k + C_k\}$ is

$$\sum_{k=1}^{K} f_k(x - C_k) \prod_{j \neq k} (1 - F_j(x - C_j)).$$

The objective function of (5) can then be explicitly written as

$$r(C_1, \ldots, C_K) = \int \cdots \int_{-\infty}^{+\infty} \min\{x_1 + C_1, \ldots, x_K + C_K\} \, f_1(x_1) \cdots f_K(x_K) dx_1 \cdots dx_K \tag{6}$$

$$\begin{aligned} &= \sum_{k=1}^{K} \int_{-\infty}^{+\infty} x f_k(x - C_k) \prod_{j \neq k} (1 - F_j(x - C_j)) \, dx \\ &= \sum_{k=1}^{K} \int_{-\infty}^{+\infty} (x + C_k) \, f_k(x) \prod_{j \neq k} (1 - F_j(x + C_k - C_j)) \, dx. \end{aligned} \tag{7}$$

For any given $\{C_k\}_{k \in \mathcal{K}}$, calculating the value of $r(C_1, \ldots, C_K)$ (and also its gradient and Hessian) involves numerical integration. Throughout the paper, we use Simpson's rule [4, Section 5.1] to compute numerical integrals. The rest of the paper is devoted to solving the optimization problem (5).

III. Proposed Approach

A. Intuition

The idea is to aid the BS with the worst channel so as to maximize the minimal $R_k + C_k$ as in (4). In each particular fading state, the extra C_k is useful to BS k only if the BS k happens to hit the minimum. Consider the following question: Given an existing rate allocation $(C_1, \ldots, C_K)$, suppose that we have some small additional ΔC that can be allocated, which BS should the ΔC be given to? A moment of thought reveals that ΔC should be given to the BS with the largest probability of hitting the minimum. If there are more than one BSs with the same largest probability of hitting the minimum, then ΔC should be equally allocated among these BSs. This guiding principle can be made rigorous in the following analysis, and it leads to an efficient algorithm for allocating the C_k's.

B. Gradient Structure

We begin by computing the gradient of the objective function in (5). The k-th component of the gradient $\nabla r(C_1, C_2, \ldots, C_K)$ can be obtained from (7) as:

$$\begin{aligned} &\nabla_k r(C_1, C_2, \ldots, C_K) \\ &\quad = \int_{-\infty}^{+\infty} f_k(x) \prod_{j \neq k} (1 - F_j(x + C_k - C_j)) \, dx \\ &\quad = \int_{-\infty}^{+\infty} f_k(x - C_k) \prod_{j \neq k} (1 - F_j(x - C_j)) \, dx. \end{aligned} \tag{8}$$

Theorem 1. *For any $\{C_k \ge 0\}$, $\nabla_k r(C_1, C_2, \ldots, C_K)$ is the probability that $R_k + C_k$ is the minimum over all $k \in \mathcal{K}$.*

While the proof of the above theorem can be obtained from a close examination of the expression (8), we can also understand why the gradient $\nabla_k r(C_1, C_2, \ldots, C_K)$ is exactly the said probability by the following reasoning. Suppose a small additional ΔC is to be allocated to C_k. The improvement to the minimum (4) should be exactly ΔC times the probability that BS k hits the minimum. Since by definition $\nabla_k r(C_1, C_2, \ldots, C_K)$ is the proportional improvement of the objective function over small change of $C_k + \Delta C$, we must have that $\nabla_k r(C_1, C_2, \ldots, C_K)$ is the said probability. For convenience, the rest of the paper denotes the probability that the value of $R_k + C_k$ achieves the minimum of all K BSs, at given $\{C_k\}_{k=1}^K$, as

$$p_k(C_1, C_2, \ldots, C_K) = \nabla_k r(C_1, C_2, \ldots, C_K), \ \forall k \in \mathcal{K}. \quad (9)$$

C. Optimality Condition

We are now ready to state the following optimality condition for the optimization problem (5).

Theorem 2. *The function $r(C_1, C_2, \ldots, C_K)$ is concave in $\{C_k\}_{k\in\mathcal{K}}$. The optimal $\{C_k^*\}_{k\in\mathcal{K}}$ of (5) should be such that if $\mathcal{S} = \{k \mid C_k^* > 0\}$ is the set of BSs for which C_k^* is strictly greater than zero, i.e., $C_{k'}^* = 0$ for all $k' \notin \mathcal{S}$, then the probability that $R_k + C_k^*$ achieves the minimum must be the same for all $k \in \mathcal{S}$; further, it must be greater than the probability that $R_{k'}$ achieves the minimum for all $k' \notin \mathcal{S}$.*

Proof. The concavity of $r(C_1, C_2, \ldots, C_K)$ can be seen directly from (6). It is an integral of the concave function $\min\{x_1 + C_1, \ldots, x_K + C_K\}$ thus is also concave. The optimality condition of the convex optimization problem (5) is that there exists a nonnegative λ^* such that

$$\nabla_k r(C_1^*, C_2^*, \ldots, C_K^*) = \lambda^*, \quad \forall k \in \mathcal{S}, \quad (10)$$

$$\nabla_{k'} r(C_1^*, C_2^*, \ldots, C_K^*) < \lambda^*, \quad \forall k' \notin \mathcal{S}. \quad (11)$$

By (9), we obtain the optimality condition as stated in the theorem. ☐

As an interesting example, consider the two-BS case and further assume that the distributions of R_1 and R_2 are both symmetric. Then, it is easy to see that if as long as $R_1 + C_1$ and $R_2 + C_2$ have the same mean, we always have

$$\Pr(R_1+C_1 \geq R_2+C_2) = \Pr(R_1+C_1 \leq R_2+C_2) = \frac{1}{2}. \quad (12)$$

Thus, the optimal rate allocation strategy must be to allocate rate to the two BSs so as to equalize the mean of $R_k + C_k$. This may appear surprising at a first glance, as the optimal rate allocation strategy in this two-BS symmetric case does not depend at all on the exact distributions, except through the means. For example, the respective variances of R_k do not influence the rate allocation, (although they do influence the achievable multicast rate). We emphasize that this is a feature of the particular two-BS symmetric example. When the distributions are not symmetric or when there are more than two BSs, the optimal rate allocation does depend on the exact distributions.

D. Proposed Algorithm

It is now clear, based on the optimality condition, that an optimization strategy should begin by allocating rate to the BS with the largest probability that its $R_k + C_k$ achieves the minimum, until two or more BSs have the same such probability. Then, the rates should be further allocated to all such BSs to keep their probabilities of hitting the minimum the same, until all the rates are exhausted. Since the allocation of rates to different BSs affects their probabilities differently (depending on the probability distribution of R_k) and in some nonlinear fashion, a numerical algorithm is needed to perform each step of the rate allocation. This section proposes an efficient way to do so.

First, we give the following two lemmas that are used later in the algorithm design. For ease of presentation, we use $\mathcal{S}^c$ to denote the complement of the set $\mathcal{S}$ with respect to $\mathcal{K}$.

Lemma 1. *The following results hold true:*

(i) *For any $\mathcal{S} \subset \mathcal{K}$,*

$$p_\ell\left(\{C_k + c_k\}_{k\in\mathcal{S}}, \{C_{k'}\}_{k'\in\mathcal{S}^c}\right), \ \ell \in \mathcal{S}^c$$

is a nondecreasing function with respect to all $c_k \geq 0$.

(ii) *For any $c > 0$,*

$$p_\ell\left(\{C_k + c\}_{k\in\mathcal{K}}\right) = p_\ell\left(\{C_k\}_{k\in\mathcal{K}}\right), \ \ell \in \mathcal{K}.$$

Proof. Recall that p_ℓ is the probability that the rate of the ℓ-th BS achieves the minimum. Therefore, allocating $\{c_k \geq 0\}_{k\in\mathcal{S}}$ to the BSs in set $\mathcal{S}$ does not decrease the probability that the rate of the BSs not in the set hits the minimum. Moreover, allocating the same $c \geq 0$ rate to all BSs does not change the probability that the rate of all BSs hits the minimum. ☐

Lemma 2. *Suppose that all PDFs $f_k(x)$ of rate R_k are positive for all x, then the following results hold true:*

(i) *Any partial Hessian of $r(C_1, \ldots, C_K)$, the objective function of (5), is negative definite (thus invertible).*

(ii) *The full Hessian of $r(C_1, \ldots, C_K)$ is negative semidefinite and has only one zero eigenvalue.*

Proof. Let us compute the Hessian of $r(C_1, C_2, \ldots, C_K)$ in (5). The (k, ℓ)-th off-diagonal entry is

$$\nabla^2_{k\ell} r(C_1, C_2, \ldots, C_K) = \int_{-\infty}^{+\infty} f_k(x - C_k) f_\ell(x - C_\ell) \prod_{j\neq k,\, j\neq \ell} (1 - F_j(x - C_j))\, dx$$

and the k-th diagonal entry of the Hessian is

$$\nabla^2_{kk} r(C_1, C_2, \ldots, C_K) = -\sum_{\ell\neq k} \nabla^2_{k\ell} r(C_1, C_2, \ldots, C_K), \quad \forall k \in \mathcal{K}. \quad (13)$$

If all PDFs $f_k(x)$ of rate R_k are positive for all x, it follows that

$$\nabla^2_{k\ell} r(C_1, C_2, \ldots, C_K) > 0, \ \forall \ell \neq k. \quad (14)$$

Combining (13) and (14) shows that (i) any partial Hessian of $r(C_1, C_2, \ldots, C_K)$ is strictly diagonally dominant and thus negative definite (and invertible) [5, Theorem 6.1.10] and (ii) the full Hessian of $r(C_1, C_2, \ldots, C_K)$ is negative semidefinite. Moreover, it can be checked that the all-one vector is the eigenvector of the full Hessian of $r(C_1, C_2, \ldots, C_K)$ corresponding to the zero eigenvalue. By this and the fact that any partial Hessian of $r(C_1, C_2, \ldots, C_K)$ is negative definite, we obtain that the full Hessian has only one zero eigenvalue. $\square$

The proposed algorithm for solving problem (5) iteratively allocates rates to the BSs as follows. Fix $\epsilon > 0$ to be some solution tolerance. Let $\{C_k^t\}_{k\in\mathcal{K}}$ be the rate allocation at the t-th iteration that satisfies the optimality condition of Theorem 2 within tolerance ϵ so that a subset of $\mathcal{S}^t \subset \mathcal{K}$ BSs will be allocated some positive rates, where $\mathcal{S}^t$ is the ϵ-maximum set among all $\left\{p_\ell\left(\{C_k^t\}_{k\in\mathcal{K}}\right)\right\}_{\ell\in\mathcal{K}}$, i.e.,

$$\mathcal{S}^t = \left\{\ell \;\middle|\; p_\ell\left(\left\{C_k^t\right\}_{k\in\mathcal{K}}\right) \ge \max_{\ell\in\mathcal{K}}\left\{p_\ell\left(\left\{C_k^t\right\}_{k\in\mathcal{K}}\right)\right\} - \epsilon\right\}.$$

At the $(t+1)$-th iteration, we allocate some additional rates while keeping the optimality condition of Theorem 2 satisfied (within tolerance ϵ) by finding an appropriate level $\lambda > 0$ together with rate allocation $\{c_k \ge 0\}_{k\in\mathcal{S}^t}$ such that there exists some nonempty $\mathcal{S}' \subseteq (\mathcal{S}^t)^c$, we have, for all $\ell \in \mathcal{S}^t$,

$$\left|p_\ell\left(\left\{C_k^t + c_k\right\}_{k\in\mathcal{S}^t}, \left\{C_{k'}^t\right\}_{k'\in(\mathcal{S}^t)^c}\right) - \lambda\right| \le \epsilon/2; \quad (15)$$

for all $\ell \in \mathcal{S}'$,

$$\left|p_\ell\left(\left\{C_k^t + c_k\right\}_{k\in\mathcal{S}^t}, \left\{C_{k'}^t\right\}_{k'\in(\mathcal{S}^t)^c}\right) - \lambda\right| \le \epsilon/2; \quad (16)$$

and further for all $\ell \in \left(\mathcal{S}^t \bigcup \mathcal{S}'\right)^c$,

$$p_\ell\left(\left\{C_k^t + c_k\right\}_{k\in\mathcal{S}^t}, \left\{C_{k'}^t\right\}_{k'\in(\mathcal{S}^t)^c}\right) < \lambda - \epsilon/2. \quad (17)$$

Next, we argue that the level λ and the rate allocation $\{c_k \ge 0\}_{k\in\mathcal{S}^t}$ which jointly satisfy (15), (16), and (17) can be efficiently found by exploiting the special structure of the problem.

Let us first argue that $\{c_k\}_{k\in\mathcal{S}}$ satisfying (15) (with given λ and $\mathcal{S}$) can be found efficiently by applying Newton's algorithm to solve the set of nonlinear equations

$$p_\ell\left(\left\{C_k^t + c_k\right\}_{k\in\mathcal{S}}, \left\{C_{k'}^t\right\}_{k'\in\mathcal{S}^c}\right) = \lambda,\ \ell \in \mathcal{S}. \quad (18)$$

Let $\mathbf{c}^{i+1}$ be the vector of c_k's at the i-th iteration of Newton's algorithm. More specifically, let $\mathbf{c}^1 = \mathbf{0}$, then, for all $i \ge 1$, do

$$\mathbf{c}^{i+1} = \mathbf{c}^i + \left[\nabla^2 \mathbf{p}_{\mathcal{S}}\left(\left\{C_k^t + c_k^i\right\}_{k\in\mathcal{S}}, \left\{C_{k'}^t\right\}_{k'\in\mathcal{S}^c}\right)\right]^{-1} \left[\lambda\mathbf{e} - \mathbf{p}_{\mathcal{S}}\left(\left\{C_k^t + c_k^i\right\}_{k\in\mathcal{S}}, \left\{C_{k'}^t\right\}_{k'\in\mathcal{S}^c}\right)\right] \quad (19)$$

until (15) is satisfied, where $\mathbf{e}$ is the all-one vector of dimension $|\mathcal{S}|$ and $\mathbf{p}_{\mathcal{S}}\left(\left\{C_k^t + c_k^i\right\}_{k\in\mathcal{S}}, \{C_{k'}^t\}_{k'\in\mathcal{S}^c}\right)$ is a collection of $\left\{p_\ell\left(\{C_k^t + c_k\}_{k\in\mathcal{S}}, \{C_{k'}^t\}_{k'\in\mathcal{S}^c}\right)\right\}_{\ell\in\mathcal{S}}$. According to (i) of Lemma 2, if the cardinality of the set $\mathcal{S}$ is less than K, the above Newton iteration (i.e., the inverse) is well defined; otherwise the remaining rate should be equally allocated to all BSs by (ii) of Lemma 2. We remark that the above problem (18) can be solved without the need to enforce the positivity constraints on c_k^i for all $k \in \mathcal{S}$. However, to guarantee convergence, appropriate step size needs to be chosen for Newton's algorithm.

Algorithm 1 Iterative Rate Allocation for Solving (5)

Input: Parameters of the optimization problem (5), i.e., $C > 0$, $\{F_k, f_k\}_{k\in\mathcal{K}}$, and tolerance $\epsilon > 0$.
Initialization: Set $t = 0$ and $C_k^t = 0,\ k \in \mathcal{K}$.
Compute $\{p_k\}_{k\in\mathcal{K}}$ based on $\{F_k, f_k, C_k^t\}_{k\in\mathcal{K}}$ and obtain the ϵ-maximum set $\mathcal{S}^t$ based on $\{p_k\}_{k\in\mathcal{K}}$.
if $|\mathcal{S}^t| = K$ **then**
  Set $C_k^t = C/K,\ k \in \mathcal{K}$, return $\{C_k^t\}_{k\in\mathcal{K}}$, and terminate.
end if
loop
  Use binary search and Newton's algorithm to compute $\{c_k > 0\}_{k\in\mathcal{S}^t}$ such that (15), (16), and (17) are satisfied for some $\mathcal{S}'$.
  if $\sum_{k\in\mathcal{K}} C_k^t + \sum_{k\in\mathcal{S}^t} c_k \le C$ **then**
    Set $C_k^{t+1} = C_k^t + c_k,\ k \in \mathcal{S}^t$.
  end if
  if $\sum_{k\in\mathcal{K}} C_k^t + \sum_{k\in\mathcal{S}^t} c_k > C$ **then**
    Solve (18) with some level such that $\sum_{k\in\mathcal{S}^t} c_k = C - \sum_{k\in\mathcal{K}} C_k^t$.
    Set $C_k^{t+1} = C_k^t + c_k,\ k \in \mathcal{S}^t$.
    Return $\left\{C_k^{t+1}\right\}_{k\in\mathcal{K}}$ and terminate.
  end if
  Obtain the ϵ-maximum set $\mathcal{S}^{t+1} = \mathcal{S}^t \cup \mathcal{S}'$.
  if $|\mathcal{S}^{t+1}| = K$ **then**
    Return $\left\{C_k^{t+1} + \dfrac{C - \sum_{k\in\mathcal{K}} C_k^{t+1}}{K}\right\}$ and terminate.
  end if
  Set $t = t + 1$.
end loop

Next we argue that the level $\lambda > 0$ that satisfies (15), (16), and (17) can be efficiently found by binary search. Suppose that at the t-th iteration, the lower and upper bounds of the desired level are λ^l and λ^u, respectively. For fixed

$$\lambda^{\text{mid}} = \frac{\lambda^l + \lambda^u}{2},$$

we first use Newton's algorithm to solve problem (18) (with $\mathcal{S}$ there being replaced with $\mathcal{S}^t$) and find $\{c_k \ge 0\}_{k\in\mathcal{S}}$ satisfying (15); we then compute

$$p_\ell\left(\left\{C_k^t + c_k\right\}_{k\in\mathcal{S}}, \left\{C_{k'}^t\right\}_{k'\in\mathcal{S}^c}\right),\ \ell \in \mathcal{S}^c.$$

If both (16) and (17) are satisfied for some $\mathcal{S}'$, then the desired λ is found; if there exists $\ell \in \mathcal{S}^c$ such that

$$p_\ell\left(\left\{C_k^t + c_k\right\}_{k\in\mathcal{S}}, \left\{C_{k'}^t\right\}_{k'\in\mathcal{S}^c}\right) > \lambda + \epsilon/2,$$

then set $\lambda^l = \lambda^{\text{mid}}$; if for all $k \in \mathcal{S}^c$,

$$p_\ell\left(\left\{C_k^t + c_k\right\}_{k\in\mathcal{S}}, \left\{C_{k'}^t\right\}_{k'\in\mathcal{S}^c}\right) < \lambda - \epsilon/2,$$

then set $\lambda^u = \lambda^{\text{mid}}$. The above update rule of the upper and lower bounds of the level λ is due to the monotonicity of $p_\ell\left(\{C_k^t + c_k\}_{k\in\mathcal{S}}, \{C_k^t\}_{k\in\mathcal{S}^c}\right)$, $\ell \in \mathcal{S}^c$, as shown in (i) of Lemma 1. The initial λ^l and λ^u at the t-th iteration can be set as

$$\lambda^l = \min_{\ell\in\mathcal{S}^c}\left\{p_\ell\left(\left\{C_k^t\right\}_{k\in\mathcal{K}}\right)\right\}, \ \lambda^u = \max_{\ell\in\mathcal{S}}\left\{p_\ell\left(\left\{C_k^t\right\}_{k\in\mathcal{K}}\right)\right\}.$$

Clearly, we have $0 \le \lambda^l \le \lambda^u \le 1$.

E. Theoretical Properties

In this subsection, we list some interesting theoretical properties of the proposed Algorithm 1.

Finite termination. The proposed algorithm terminates after at most K (outer) iterations. Clearly, the algorithm terminates either when the remaining rate is not sufficient (see line 12) or when $|\mathcal{S}^t| = K$ for some t (see lines 4 and 18). If the first case happens, the algorithm terminates within K iterations, as the cardinality of the set $\mathcal{S}^t$ strictly increases as t increases. If the second case happens, the algorithm allocates the remaining rate equally to all BSs and again terminates within K iterations.

Low complexity per iteration. The most computationally expensive step (see line 8) at each (outer) iteration is to find the level λ and rate allocation $\{c_k \ge 0\}_{k\in\mathcal{S}^t}$ such that (15), (16), and (17) are jointly satisfied. The step further involves a double-loop computation. The outer loop aims to find the appropriate level λ via bisection search and the worst-case complexity is $\lceil \log_2(1/\epsilon)\rceil$. The inner loop solves the nonlinear equation (18) and finds $\{c_k\}_{k\in\mathcal{S}}$ satisfying (15) (with given λ and $\mathcal{S}$). This can be done very efficiently by using Newton's algorithm and the computational cost is negligible due to the quadratic convergence rate of Newton's algorithm. In summary, the worst-case complexity of our proposed iterative rate allocation Algorithm 1 is $K\lceil \log_2(1/\epsilon)\rceil$.

Global optimality at each iteration. The rate allocation of our proposed algorithm (with $\epsilon = 0$) is globally optimal at each iteration. More specially, for any $t \ge 0$, $\left\{C_k^{t+1}\right\}$ is the optimal solution to problem (5) with $C = \sum_{k\in\mathcal{K}} C_k^{t+1}$.

IV. Simulation Results

To illustrate the numerical performance of the proposed algorithm, this section considers a C-RAN topology with $K = 9$ BSs at distances 50m to 200m from the centralized cloud. We assume a single-antenna setup with Rayleigh fading. The transmit power is -27dBm/Hz from the wireless multicast backhaul transmitter; the background noise level is set to be -120dBm/Hz (to account for interference). We assume a distance dependent pathloss model according to $128.1 + 37.6\log_{10}(d)$ in dB, where d is expressed in km.

The capacity of the block-fading channel model follows certain distribution, which over multiple fading blocks can be approximated as a Gaussian distribution in aggregate [6]. Although on a block-by-block basis the channel distributions are not Gaussian, for the convenience and for the purpose of numerical illustration, we approximate them as Gaussian in our simulation. In particular, we approximate the distributions of the capacities of the channels to the 9 BSs as $\mathcal{N}(\mu_k, \sigma_k^2)$, $k = 1, \ldots, 9$, where $[\mu_k] = [8.5, 7, 5.2, 4, 3, 2, 2, 1, 1]$, and $[\sigma_k] = [1.8, 1.6, 1.6, 1.5, 1.3, 1.2, 0.5, 0.5, 0.1]$ in bps/Hz.

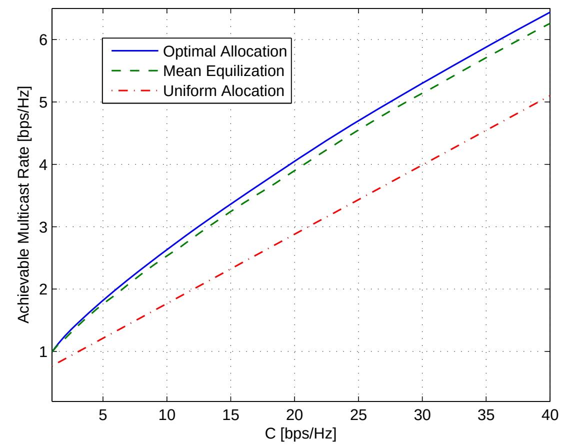

Fig. 2. Achievable multicast rate versus total secondary backhaul rate

We compare the proposed optimal algorithm with the following two suboptimal benchmarks. The first one is a mean equalization strategy, which allocates larger C_k to the BSs with smaller mean rates to make $\mathbb{E}[R_k] + C_k$ equal to each other (if possible); the second one is the uniform allocation strategy, which allocates the same C/K to all BSs.

Fig. 2 plots the multicast rates R_0 achieved by the three algorithms versus the total secondary backhaul C. It can be observed from Fig. 2 that the proposed algorithm outperforms the two benchmarks, although the margin over the mean equalization strategy is small, indicating that allocating rates to equalize the means is near optimal in this example. Moreover, the slope of the optimal achievable rate, which is λ^* and is also the largest probability of hitting the minimum among all BSs, gradually decreases as the total secondary backhaul rate increases. The slope eventually becomes $1/K$ if C is larger than a threshold, which is 36.5 in this example.

References

[1] T. Q. S. Quek, M. Peng, O. Simeone, and W. Yu, Eds., *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge University Press, 2017.

[2] L. Liu and W. Yu, "Cross-layer design for downlink multi-hop cloud radio access networks with network coding," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1728–1740, Apr. 2017.

[3] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.

[4] K. E. Atkinson, *An Introduction to Numerical Analysis*, 2nd ed. John Wiley & Sons, 1989.

[5] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[6] P. Smith and M. Shafi, "An approximate capacity distribution for MIMO systems," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 887–890, Jun. 2004.