Computer Communications 138 (2019) 32–44

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Optimal backhauling for dense small-cell deployments using mmWave links☆

Po-Han Huang *, Konstantinos Psounis

Ming Hsieh Department of Electrical Engineering, University of Southern California, 3740 McClintock Ave, Los Angeles, CA, 90089, USA

ARTICLE INFO

Keywords:
Wireless backhaul networks
mmWave communication
Hybrid beamforming
Dense small-cell deployments
Approximation algorithm

ABSTRACT

Dense small-cell deployments of 5G networks require a wireless backhaul to efficiently connect the small cells to the macro base station (BS). We envision a wireless backhaul architecture where cells are grouped into clusters. One small cell per cluster plays the role of "cluster head" connecting to the BS via a millimeter wave (mmWave) multiple-input and multiple-output (MIMO) link; the rest of the small cells in the cluster will then connect to the "cluster head" to reach the BS. We formulate the problem of jointly selecting the cluster heads and the number of BS antennas dedicated to each mmWave MIMO link between the BS and each cluster head to maximize system throughput as an integer linear program (ILP) and prove its NP-hardness. We then propose an $O(ln|S|)$ approximation algorithm, where S is the number of small cells. We show via extensive simulations that the algorithm performs very close to the optimal in practice. Last, we show that the algorithm has, on average, more than 40% performance gain compared to previous work.

1. Introduction

To meet the high demand in 5G cellular networks, several research directions are explored to increase the capacity of both the access networks and the backhaul networks [2]. For wireless access networks, one promising solution is to use a highly dense base station deployment, which could enhance the whole system throughput by frequency reuse [3]. For example, recent industry reports [4,5] envision that the number of small cells in a dense deployment will be more than 200 per square km. To support such a network deployment, the backhaul network should be re-designed because it would be too expensive to connect such a large number of access points via fibers [6].

Millimeter Wave (mmWave) communication has recently matured thanks to hardware design advancements, and has been proposed to support the high bandwidth demand in 5G cellular networks [7]. The rationale behind using mmWave communication is to take advantage of higher frequency bands, e.g. from 30 GHz to 300 GHz, which could provide higher capacity with larger bandwidth than todays microwave bands. For this reason, mmWave communication is considered suitable for high-bandwidth backhaul connection of ultra-dense small cells [8].

However, there are some fundamental challenges with mmWave communication, such as directivity challenges, high pathloss, low penetration and more [9]. These challenges make mmWave communication only useful for short-range transmissions. To make mmWave links handle long-range transmissions, it has been recently proposed to use multiple-input and multiple-output (MIMO) beamforming [10]. This MIMO beamforming is not expected to incur significant inter-antenna interference thanks to the intrinsic directional nature of the transmissions which has led researchers to model mmWave backhaul links as pseudo-wires without interference [11].

Motivated by this, [12–30] discuss how to use mmWave communication in wireless backhaul networks. Some of this work focuses on a distributed architecture (e.g., [12–15]), where network operators deploy cluster heads connected with fibers to the core network and provide mmWave backhaul connection for the rest of the small cells. Another approach is to use a centralized architecture (e.g., [16–20]), where a central node like a macro cell controls every small cell via a mmWave backhaul. However, neither architecture is particularly appealing for a dense network deployment. For example, the centralized architecture has too high of a signaling complexity, and the distributed architecture cannot handle well the fluctuation of traffic demand as the number of cluster heads is fixed.

To address those problems a hybrid architecture has been proposed, see, for example, [21–30], where a centralized node (e.g., macro cell) controls several cluster heads which are also small cells via mmWave communication, and these cluster heads provide wireless backhaul for the rest of the small cells. In this fully wireless backhaul network architecture, the macro cell only controls the cluster heads rather than all small cells, and there is great flexibility for changing cluster heads if the traffic demand fluctuates. Still, due to the short range of mmWave communication, multi-hopping might be required, and, to

☆ This work was presented in part at 13th Wireless On-demand Network Systems and Services Conference, Jackson Hole, WY, USA, Feb. 2017 Huang and Psounis (2017) [1].

* Corresponding author.
E-mail addresses: pohanh@usc.edu (P.-H. Huang), kpsounis@usc.edu (K. Psounis).

https://doi.org/10.1016/j.comcom.2019.03.002
Received 17 September 2018; Received in revised form 5 February 2019; Accepted 2 March 2019
Available online 7 March 2019

avoid the performance issues of multi-hopping, e.g. latency [31], MIMO beamforming has been suggested to support long-range transmissions between the macro cell and the cluster heads [10].

In the context of such a hybrid architecture with MIMO-enabled mmWave backhaul links, in this work we study how to optimally select cluster heads among the small cells and how to optimally select the link capacity of the backhaul links between the macro cell and the cluster cells, that is, how many antennas of the macro cell to dedicate at each backhaul link, to maximize the achieved system throughput. For this purpose, we formulate this problem into an integer linear program (ILP) and prove its NP-hardness by reducing it to a k-set cover problem. To solve the problem in polynomial time, we first transform it into a simpler problem by allowing some cells to be out of range of all selected cluster heads and propose an optimal solution. We then require that all cells should be assigned to a cluster head and propose an approximation algorithm which uses the aforementioned optimal solution to solve the original problem, and show its near-optimality by simulation.

The main contributions of this paper are as follows:

1. To the best of our knowledge, this paper is the first work which jointly considers antenna partitioning and cluster heads selection in mmWave wireless backhauling under a hybrid architecture.
2. An optimal algorithm is proposed for the special case where there are no coverage constraints, i.e. some remote cells may not be covered by any cluster head.
3. The problem with coverage constraints is shown to be NP hard and an approximation algorithm is proposed and proved to be an $O(ln|S|)$-approximation where S is the number of small cells.
4. Extensive simulations show that the proposed algorithm is, in practice, within 5% off the optimal, and at least 40% superior than prior work [16–20,23–29].

The rest of the paper is organized as follows. Section 2 summarizes related work. Section 3 presents the system architecture that we consider. The objective of this work and a formal problem description is presented in Section 4. In Section 5, we solve the problem without coverage constraints by an optimal solution. Then, an approximation algorithm for the original NP hard problem is discussed in Section 6. In Section 7 we evaluate the performance of the approximation algorithm using simulations and show that it outperforms previous works under a variety of realistic scenarios. Practical considerations are discussed in Sections 8 and 9 concludes the paper.

2. Related work

We summarize prior work on backhaul design focusing on distributed, centralized, and hybrid architectures.

Distributed Architecture. A distributed architecture was presented in [12–15] with two types of nodes: anchored nodes and demand nodes, where anchored nodes were used for backhaul relaying and demand nodes were used for serving users. Note that anchored nodes connect to core networks via a wired backhaul and demand nodes connect to core networks with the help of anchored nodes via a wireless backhaul. [12,13] designed algorithms which minimize the deployment cost under a bandwidth resource constraint for a wireless small-cell backhaul network. The authors in [14,15] investigated how to select anchored nodes and where to place them while connecting them with wires for achieving high throughput. This line of prior work does not consider wireless backhauling between the anchored nodes and the macro cell. In contrast, we consider wireless backhauling between both small cells and cluster heads, and between cluster heads and the macro cell.

Centralized Architecture. A centralized architecture was investigated in [16–20], where a central node like a macro cell controls every small cell via a mmWave backhaul. Some work in this line of research (e.g., [16,17]) focused mainly on how to efficiently allocate antennas of the macro cell to provide a wireless backhauling solution towards all small cells. Other works (e.g., [18–20]) discussed the role of MIMO beamforming methods for providing better performance under different scenarios. More specifically, the authors in [18] took wind-induced beam misalignment into consideration, the authors in [19] addressed full-duplex small cells, and the authors in [20] considered both cell association and bandwidth allocation in their model. However, this prior work does not consider the option to select a subset of small cells to act as relays (cluster heads) for the rest of the small cells.

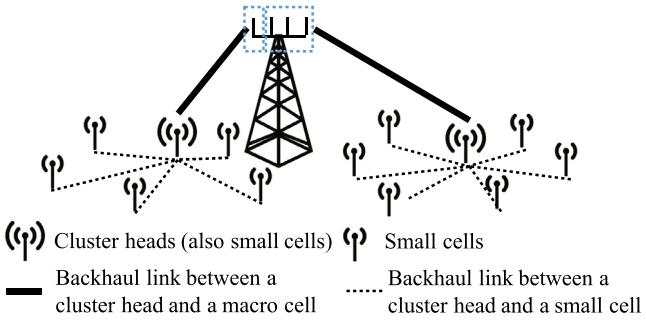

Fig. 1. Hybrid architecture.

Hybrid Architecture. A hybrid architecture has been recently introduced, see [21,22] and references therein, according to which a centralized node (e.g. macro cell) controls several cluster heads which are also small cells via mmWave communication, and these cluster heads provide wireless backhauling for the rest of the small cells. A number of associated problems have been studied, see [21–30]. Specifically, [23,24] study how to associate small cells with cluster heads subject to a predetermined number and location of cluster heads. In contrast, we study the optimal number and location of cluster heads. [25] studied dynamic radio resource management and transmission schemes to maximize the system throughput while maintaining low packet delay in the context of a fixed wireless backhaul, whereas we consider dynamic wireless backhauling for performance reasons. [26] studied packet delay in wired and wireless backhauling technologies and proposed a backhaul-aware BS association policy to minimize the delay, whereas we focus on throughput. [27–29] studied via heuristics how to fit more network flows (video streams) in a hybrid architecture with bandwidth resource constraints, without providing any formal performance guarantees. Last, [30] studied inter-operator resource management for maximizing the average sum rate. Specifically, the authors optimized the average sum rate and provided a cooperative and non-cooperative scheme to obtain a near-optimal performance. Note that all prior work on hybrid architectures does not take antenna partitioning into consideration, which has an impact on the system throughput.

In summary, different from prior work, this work takes advantage of both a fully-wireless backhaul and a dynamic selection of cluster heads to achieve higher throughput while providing optimal and approximation algorithms.

3. System architecture

3.1. Hybrid architecture for wireless backhaul

According to the hybrid architecture of wireless backhaul design for small cell networks (see Fig. 1), packets travel between the core network and a macro base station (BS) via fiber, then travel between the macro cell and a cluster head via mmWave communication, and last between a cluster head and the other small cell again via mmWave communication.

The main assumptions that we make are as follows. First, similar to prior work [11], we assume that the communication between cluster

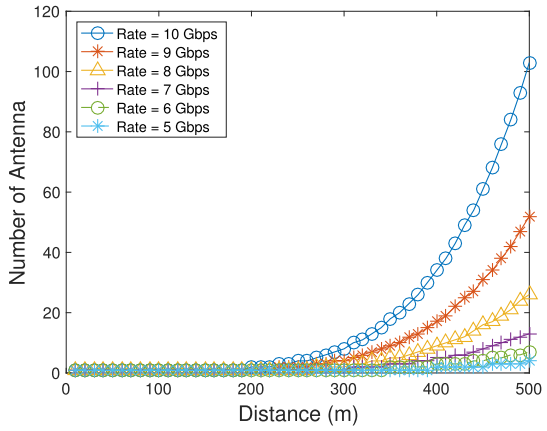

Fig. 2. Number of Required Antennas with Different Rate Requirement.

heads and the macro cell, as well as the communication between small cells and cluster heads have small interference, because main lobes are narrow and only side lobes contribute to interference. That said, we do take this small interference into consideration in our analysis, see Eq. (1).

Second, note that beamforming will suffer from alignment issues especially in a mmWave environment [18] whenever the antennas are reconfigured. For this reason, rapidly updating cluster heads is impractical and, instead, we assume that cluster heads change at much lower time scales than the duration of a transmission, such that the alignment issues can be addressed by standard schemes like the one presented in [18]. Another practical reason to change cluster heads at low time scales is the significantly higher cost of advanced antennas that can support fast reconfiguration, and that the signaling overhead increases as the frequency of antenna reconfigurations increases, see Section 8 for a discussion on signaling overhead.

Third, we assume that our network operates in the context of a half duplex communication system, and the scheduling conflicts from the half duplex system can be avoided using conflict-free routing tables and dual egress configurations, see [32] and references therein for more details. Therefore, we can fully utilize the resources in both frequency and time domain.

Last, we assume that every small cell always has packets to send, i.e. we operate in the saturated throughput regime. With this assumption in mind, the amount of traffic between small cells and cluster heads is not the main focus because it is always confined by the link capacity between the cluster heads and the macro cell. The main goal is to select the optimal set of cluster heads among all small cells such that the capacity of the backhaul network is maximized, subject to the constraints of coverage and maximum number of available antennas. Note that a small cell which becomes a cluster head should not only deal with the data from its users, but also with the traffic from the other small cells which use it as a relay.

3.2. Channel model in mmWave communication

Let d_i be the distance between the macro cell and small cell i. Then, the signal-to-noise ratio (SNR) without beamforming for small cell i can be obtained as follows,

$$D_i = \frac{P_t H_i d_i^{-\alpha}}{\mathcal{N}_0 W + \epsilon}, \forall i \in S, \tag{1}$$

where P_t is the transmission power, H_i is the small scale fading, α is the path-loss coefficient, W is the bandwidth of a frequency slot, and $\mathcal{N}_0$ is the noise power spectral density. In addition, ϵ captures the small

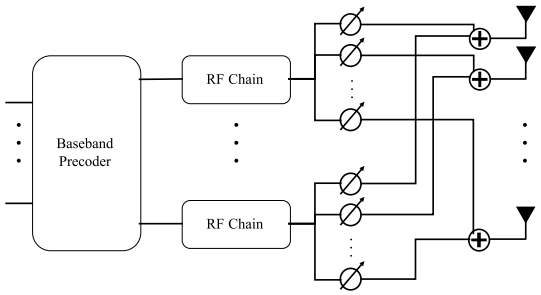

Fig. 3. Fully-Connected Hybrid Beamforming Architecture [37].

interference from the side lobes of other links, which, following [33], equals:

$$\epsilon = \sum_{\forall j \in S, j \neq i} p_t H_j d_{j,i}^{-\alpha},$$

where p_t is the transmission power of a side-lobe beam, H_j is the small-scale fading, and $d_{j,i}$ is the distance between small cell j and small cell i. Note that, as shown in [34], small-scale fading has a minor impact on the received power from LOS base stations when directional antennas are used in mmWave systems, but we include it for completeness nevertheless.

Then, the backhaul capacity C_i, achieved with N_i transmitter antennas and one receiver antenna, can been shown to equal [35]

$$C_i(N_i, D_i) = W \log_2 (1 + N_i D_i). \tag{2}$$

This equation shows the relationship between three parameters: the distance between a small cell and the macro cell, the number of antennas for a mmWave link, and the capacity of the link. We illustrate this relationship in Fig. 2, where it is evident that to maintain the same capacity as the distance grows linearly we need an exponentially larger number of antennas.

Note that the capacity model between the macro cell and a small cell only uses beamforming gain instead of multiplexing gain for MIMO beamforming. This is because multiplexing gains require a well-conditioned channel matrix from rich scattering. However, since rich scattering is typically unavailable in mmWave systems, the channel matrix is typically ill-conditioned, and beamforming gain is usually adopted in the model of mmWave MIMO beamforming, see [11,35–37].

3.3. Hybrid beamforming architecture

We assume a fully-connected hybrid beamforming architecture, see Fig. 3 for a schematic representation. Note that a fully-connected hybrid beamforming architecture allows arbitrary partitioning of the antenna elements to RF chains. Since narrow beams are usually found in mmWave setups, it is beamforming rather than multiplexing gains that are used in our context. Therefore, each cluster head should be associated with one or more RF chains/antenna elements (depending on the number of antennas and clusters heads) and each RF chain should transmit useful signal to one cluster head only. Besides fully-connected hybrid beamforming architectures, there is a plethora of partially-connected hybrid beamforming architectures aiming to reduce the cost at the expense of less flexibility. We discuss how to apply our work to partially-connected hybrid beamforming architectures in Section 8.2.

4. Problem formulation

In this section, we formulate the problem of **mm**Wave Wireless Backhauling for **H**ybrid **A**rchitecture **UL**tra-Dense Networks (mmHAUL) using integer linear programming (ILP). Specifically, we want to determine which small cells become cluster heads, which small cells connect

Table 1
System model notation.

Description	Notation
Decision variable for establishing connection between cluster head i and small cell j	$y_{i,j}$
The set of small cells	$\mathcal{S}$
Distance between macro cell and cluster head i	d_i
Transmission power of macro cell	P_t
Small-scale fading from macro cell to small cell i	H_i
Bandwidth of a frequency slot	W
Noise power spectral density	$\mathcal{N}_0$
Path loss exponent	α
Signal-to-noise ratio between macro cell and cluster head i	D_i
Number of antennas for backhauling of cluster head i	N_i
Backhaul capacity for cluster head i	$C_i(N_i, D_i)$
Maximum number of available antennas in macro cell	N_{MAX}
Adjacency matrix and its elements	$\mathcal{A} : a_{i,j}$

to these cluster heads while maximizing the system throughput and ensuring connectivity for every small cell, and how many antennas of the macro cell are allocated to the connection between the macro cell and each cluster head. Table 1 lists useful notation.

4.1. Modeling: Integer linear programming (ILP)

Given the set of small cells $\mathcal{S}$, the adjacency matrix

$$\mathcal{A} = \begin{bmatrix} a_{1,1} \cdots a_{1,|\mathcal{S}|} \\ \vdots \quad \ddots \quad \vdots \\ a_{|\mathcal{S}|,1} \cdots a_{|\mathcal{S}|,|\mathcal{S}|} \end{bmatrix},$$

which indicates whether small cell i and j are within range or not, and the number of available antennas at the macro cell N_{MAX}, we want to determine the values of the following decision variables: $y_{i,j}$ which indicates whether small cell j connects to cluster head i, and, N_i which indicates the number of antennas used for the link between the macro cell and the small cell i. Note that $N_i > 0$ indicates small cell i is selected to act as a cluster head. Under the saturation regime, the amount of traffic between the small cells and the cluster heads is eventually confined by the link capacity between the cluster heads and the macro cell. This is so because, by design, the link capacity between a cluster head and the macro cell would be less than the sum of the link capacities between the cluster head and the small cells it is serving, in order to take advantage of statistical multiplexing. Hence, under a saturation regime where all small cells are fully loaded, the bottleneck is the link between the cluster head and the macro cell, and not the links between a small cell and its cluster head. Therefore, the main goal of this problem is to select the right set of cluster heads and allocate to each of them the right number of antennas which would dictate the capacity of the backhaul links between cluster heads and the macro cell. We formulate this problem as follows.

$$Q1 : \max_{y_{i,j}, N_i} \sum_{\forall i \in \mathcal{S}} C_i(N_i, D_i) \tag{3}$$

s.t.

$$y_{i,j} \leq N_i \cdot a_{i,j}, \ \forall i, j \in \mathcal{S} \tag{4}$$

$$\sum_{\forall i \in \mathcal{S}} y_{i,j} = 1, \ \forall j \in \mathcal{S} \tag{5}$$

$$\sum_{\forall i \in \mathcal{S}} N_i \leq N_{MAX}, \tag{6}$$

$$y_{i,j} = \{0, 1\}, \ \forall i, j \in \mathcal{S}$$

$$N_i \in \mathbb{Z}+, \ \forall i \in \mathcal{S}$$

The rationale behind this model is as follows. Objective (3) aims to maximize the system throughput by deciding the allocation of antennas and the associations between the cluster heads and the rest of small cells. Note that even though the objective function does not depend on the allocation variable $y_{i,j}$, we need the allocation variable $y_{i,j}$ in the constraints to ensure coverage (see below). (4) represents that the connection $y_{i,j}$ can be chosen if small cells i and j are adjacent to each other (i.e., $a_{i,j} = 1$) and small cell i has been chosen as a cluster head (i.e., $N_i > 0$). (5) guarantees that every small cell connects to one cluster head. Last, (6) guarantees that the total number of used antennas for the links between cluster heads and the macro cell cannot exceed the number of available antennas at the macro cell.

4.2. Complexity analysis

In this section, we show that mmHAUL is NP-hard by reducing it to a k-set cover problem. We start by defining the k-set cover problem and then we prove NP-hardness.

Definition 1 (*k-Set Cover Problem*)**.** Given a universe $\mathcal{U}$, an integer k, and a family $\mathcal{T}$ of subsets of $\mathcal{U}$, a cover is a subfamily $\mathcal{C} \subset \mathcal{T}$ of sets whose union is $\mathcal{U}$, and $|\mathcal{C}| \leq k$.

Theorem 1. *mmHAUL is NP-hard*

Proof. Consider that $N_i = \{0, 1\}, \forall i \in \mathcal{S}$. We can transform our problem to the standard form of the k-set cover problem by the following steps. (i) Since $N_i = \{0, 1\}, \forall i \in \mathcal{S}$, we can get $C_{MAX} = C_i(1, D_{max})$, where D_{max} denotes the maximum SNR value. With C_{MAX}, we can render (3) into a minimization problem: $\min_{N_i, y_{i,j}} \sum_{\forall i \in \mathcal{S}} (C - C_{MAX}) \cdot N_i$ (ii) The constraints (4) and (5) are the typical coverage constraints in the k-set cover problem. (Note that (5) is usually expressed as a $\geq$ inequality rather than an equality but the solution is the same in our case.) (iii) Let $k = N_{MAX}$, (6) becomes $\sum_{\forall i \in \mathcal{S}} N_i \leq k$, which is the standard form of the k-set cover problem. From this, it is evident that the k-set cover problem is a special case of our problem. It is obvious that our problem is more complicated than the k-set cover problem because our $N_i, \forall i \in \mathcal{S}$ can be any positive integers. Since the k-set cover problem is known to be NP-complete, our problem is NP-hard. □

5. MmHAUL without coverage constraints

Since this problem is NP-hard, our first attempt to solve it efficiently assumes that every small cell is within range of any other small cell. Therefore, we reformulate the original problem ($Q1$) by assuming that the aforementioned adjacency matrix is full of ones. Note that in the next section we solve the original problem ($Q1$) using as part of the solution the methods developed in this current section.

Given the full connectivity assumption, (4) and (5) are always satisfied. Thus, these two constraints and the decision variable $y_{i,j}$ can be eliminated. Still, we have to decide which small cells should become cluster heads for maximizing the system throughput, and, the rest of the small cells will associate with a cluster head near them. With the above in mine we reformulate the problem as follows.

$$Q2 : \max_{N_i} \sum_{\forall i \in \mathcal{S}} C_i(N_i, D_i) \tag{7}$$

s.t.

$$\sum_{\forall i \in \mathcal{S}} N_i \leq N_{MAX}. \tag{8}$$

$$N_i \in \mathbb{Z}+, \forall i \in \mathcal{S}$$

In the following subsections, we propose a polynomial time algorithm to obtain the optimal solution of $Q2$. The algorithm consists of three steps which we discuss separately in the following subsections.

5.1. Upper bound

The first step to solve this problem is to reformulate it from ILP to linear programming (LP). In particular, the relaxed problem is the same as $Q2$, but now N_i is a real number instead of an integer.

Since it is an LP with one objective function and one constraint, we can use Lagrangian relaxation to solve it as follows:

$$\max_{N_i} \sum_{\forall i \in \mathcal{S}} [C_i(N_i, D_i) - \lambda N_i] + \lambda N_{MAX} \tag{9}$$

s.t.

$$N_i \in \mathbb{R}+, \forall i \in \mathcal{S}$$

where λ is the Lagrangian multiplier. Using the Karush–Kuhn–Tucker (KTT) conditions, we get the following, which, interestingly, resembles the KTT conditions for the well know waterfilling problem [38].

$$\left.\begin{array}{r} \frac{\mathrm{d}C_i(N_i,D_i)-\lambda N_i}{\mathrm{d}N_i} = 0 \\ \lambda(\sum_{\forall i \in \mathcal{S}} N_i - N_{MAX}) = 0 \\ \lambda \geq 0 \end{array}\right\} \text{KKT conditions.}$$

Then, we have the allocation policy as follows,

$$(\frac{W}{\lambda} - \frac{1}{D_i})^+ = N_i, \forall i \in \mathcal{S} \tag{10}$$

$$\sum_{\forall i \in \mathcal{S}} N_i = N_{MAX}, \tag{11}$$

Once this algorithm is performed, the optimal solution of the relaxed problem can be obtained, which is an upper bound of $Q2$, since N_i is not an integer value in $Q2$.

5.2. Rounding

In this step we simply round the real valued N_i to the largest integer less than or equal to N_i using the floor function, i.e.,

$$\tilde{N}_i = \lfloor N_i \rfloor, \forall i \in \mathcal{S}. \tag{12}$$

5.3. Greedy knapsack algorithm for leftover antennas

After rounding, there will be some left over antennas, which allows us to assign more antennas to some links. Note that we do not assign more than one antenna to a small cell before going over all small cells because of the logarithmic objective function (see (2) and (7)). Therefore, starting with $\tilde{N}_i$, we define a binary decision variable b_i to decide whether small cell i can have one more antenna or not. We solve the following problem:

$$Q3: \max_{b_i} \sum_{\forall i \in \mathcal{S}} C_i(\tilde{N}_i + b_i, D_i) \tag{13}$$

subject to

$$\sum_{\forall i \in \mathcal{S}} b_i \leq \tilde{N}_{MAX}, \tag{14}$$

$$b_i = \{0, 1\}, \forall i \in \mathcal{S}$$

where $\tilde{N}_{MAX} = N_{MAX} - \sum_{\forall i \in \mathcal{S}} \tilde{N}_i$.

$Q3$ is a standard binary Knapsack problem, and we can use a greedy algorithm [39] to get a polynomial-time solution. Specifically, at each step we select the small cell i with the largest value of D_i (i.e., the best SNR or the closest one) to give one antenna, and subtract one antenna from $\tilde{N}_{MAX}$ until $\tilde{N}_{MAX} \leq 0$. Please see Algorithm 1 for more details.

Notice that it is straightforward to show that the Greedy Knapsack Algorithm is the optimal solution for $Q3$:

Theorem 2. *Greedy Knapsack Algorithm is the optimal solution for $Q3$.*

Proof. Greedy Knapsack Algorithm is the exact optimal solution for $Q3$ since all variables have the same weight (i.e., 1) according to the results in [39]. □

Algorithm 1 Greedy Knapsack Algorithm for $Q3$.

Input: $\tilde{N}_i, \forall i \in \mathcal{S}$, $D_i, \forall i \in \mathcal{S}$, and $\tilde{N}_{MAX}$.
Output: $b_i, \forall i \in \mathcal{S}$.

```
Initialize b_i, ∀i ∈ S = {0, ..., 0}
Sort ∀i ∈ S by D_i in descending order with new index î.
for ∀î ∈ S do
    if Ñ_MAX ≥ 0 then
        b_î = 1
        Ñ_MAX = Ñ_MAX − b_î
    end if
end for
```

5.4. ClosURK: An optimal solution for mmHAUL without coverage constraints

We put everything together and propose an algorithm we referred to as ClosURK (**Clos**est-selected algorithm with **U**pper bound, **R**ounding, and greedy **K**napsack) which optimally solves $Q2$. The algorithm obtains the upper bound first, then performs rounding, and finally uses the Greedy Knapsack Algorithm. The details of ClosURK are shown in Algorithm 2.

Algorithm 2 ClosURK for mmHAUL without Coverage Constraints.

Input: $D_i, \forall i \in \mathcal{S}$ and N_{MAX}.
Output: $N_i^*, \forall i \in \mathcal{S}$.

```
N_i ← Upper Bound (D_i, N_MAX);
Ñ_i ← Rounding (N_i, D_i, N_MAX);
b_i ← Greedy Knapsack Algorithm (Ñ_i, D_i, Ñ_MAX);
N_i^* = Ñ_i + b_i;
```

The following theorem proves that ClosURK finds an optimal solution to $Q2$.

Theorem 3. *Given $D_i \geq 1, \forall i \in \mathcal{S}$, ClosURK is the optimal solution of $Q2$.*

Proof. Please see Appendix A. □

Remark. The intuition of this algorithm is quite straightforward. The first two steps (i.e., upper bound and rounding) are used to obtain the initial allocation of antennas and access the number of leftover antennas. Because of the logarithmic objective function, allocating additional antennas one by one to each small cell is better than allocating all antennas to one small cell. Therefore, for the leftover antennas, we allocate one more antenna to each small cell starting from the closest one to the macro cell with the highest SNR value and proceeding in decreasing order of SNR.

5.5. Complexity analysis

What remains is to prove that the algorithm can run in polynomial time.

Theorem 4. *Given $D_i \geq 1, \forall i \in \mathcal{S}$, ClosURK runs in polynomial time.*

Proof. Since the upper bound of $Q2$ can be obtained in polynomial time based on [38], the first part of ClosURK runs in polynomial time. Rounding can run in linear time as we can examine every small cell once. Greedy Knapsack Algorithm runs in $O(|\mathcal{S}| \log(|\mathcal{S}|))$, which is indeed polynomial time. In conclusion, ClosURK runs in polynomial time. □

6. MmHAUL with coverage constraints

In this section, we propose an approximation algorithm to solve mmHAUL with coverage constraints. We develop this algorithm based on ClosURK while taking coverage constraints into account simultaneously. In addition, we provide a formal proof for its performance guarantee.

Following the remark in Section 5, as long as the coverage constraints are satisfied, we can perform ClosURK to get the optimal solution. Therefore, our approach to solve mmHAUL is to satisfy the coverage constraints first with the minimum number of antennas, and then use ClosURK to get the maximum throughput with the remaining antennas. (Notice that in the following discussion, we assume that the coverage constraints can always be satisfied. Otherwise, no algorithm can solve this problem. We discuss how to check the feasibility of the problem in Section 6.4.)

6.1. Greedy set-cover based algorithm for coverage guarantee

To take care of the coverage constraints with the minimum number of antennas, we develop an algorithm based on the solution for the k-Set Cover Problem. Thus, we propose a Greedy Set-Cover based Algorithm, which can efficiently satisfy the coverage constraints by greedily choosing cluster heads. At each step of a greedy procedure we select the small cell i with the smallest $K_i = \frac{1}{|S_i|}$ to become cluster head, where S_i is the set of small cells that can be covered by cluster head i, and $|\cdot|$ denotes the cardinality of that set. In other words, we select the small cell that can cover the most number of small cells at each step, see Algorithm 3 for more details.

Algorithm 3 Greedy Set-Cover based Algorithm.

Input: $D_i, \forall i \in \mathcal{S}$ and N_{MAX}.
Output: $x_i, \forall i \in \mathcal{S}$.
1. Initialize $x_i, \forall i \in \mathcal{S} = \{0, \dots, 0\}$.
2. Calculate $|S_i|, \forall i \in \mathcal{S}$.
3. **while** $|\mathcal{S}|! = 0$ **do**
4. $\quad \tilde{i} = \arg\max_{\forall i \in \mathcal{S}} K_i = \frac{N_i}{|S_i|}$
5. $\quad x_{\tilde{i}} = 1$
6. $\quad \mathcal{S} \leftarrow \mathcal{S} - \tilde{i}$
7. $\quad$ Update $|S_i|, \forall i \in \mathcal{S}$.
8. **end while**

6.2. CovURK — An approximation algorithm

We propose an approximation algorithm which we refer to as CovURK (**Cov**erage aware algorithm with **U**pper bound, **R**ounding, and greedy **K**napsack algorithm) to solve the original mmHAUL problem. CovURK first checks whether $|\mathcal{S}| \leq N_{MAX}$ or $|\mathcal{S}| > N_{MAX}$, and then operates differently depending on the condition. In the first case, all small cells will become cluster heads (though they will be a cluster on their own) and there are no coverage constraints. Thus, we can just run ClosURK. In the second case, the basic idea of CovURK is to first run the Greedy Set-Cover based Algorithm to satisfy the coverage constraints and then run ClosURK to achieve the maximal throughput by allocating antennas appropriately, see Algorithm 4 for more details. In the following section, we prove that this algorithm provides an $O(ln|\mathcal{S}|)$ performance guarantee.

6.3. CovURK performance analysis

We first analyze the optimality of CovURK when $|\mathcal{S}| \leq N_{MAX}$ and the when $|\mathcal{S}| > N_{MAX}$.

Theorem 5. *Given $D_i \geq 1, \forall i \in \mathcal{S}$ and $|\mathcal{S}| \leq N_{MAX}$, ClosURK is the optimal solution of $Q1$.*

Algorithm 4 CovURK for mmHAUL.

Input: $D_i, \forall i \in \mathcal{S}$ and N_{MAX}.
Output: $N_i^*, \forall i \in \mathcal{S}$.
1. **if** $|\mathcal{S}| \leq N_{MAX}$ **then**
2. $\quad N_i^* \leftarrow$ ClosURK (D_i, N_{MAX});
3. **else**
4. $\quad N_i \leftarrow$ Greedy Set-Cover (D_i, N_{MAX});
5. $\quad N_i^* \leftarrow$ ClosURK (N_i, D_i, N_{MAX});
6. **end if**

Proof. Since $|\mathcal{S}| \leq N_{MAX}$, every small cell will have $\frac{N_{MAX}}{|\mathcal{S}|}$ antennas. Also, because $\frac{N_{MAX}}{|\mathcal{S}|} \geq 1$, this implies every small cell is covered, i.e., (4) and (5) are satisfied. Then, from the conclusion of Theorem 3, ClosURK is the optimal solution of $Q1$ in this case. □

Definition 2. For a given topology, let $L = C_i(N_i, D_{min})$ denote the lowest throughput and $U = C_i(N_i, D_{max})$ denote the highest throughput between the macro cell and a small cell, where D_{min} is the minimum SNR value in the given topology, and D_{max} is the maximum SNR value in the given topology.

Theorem 6. *For a given topology and $D_i \geq 1, \forall i \in \mathcal{S}$, CovURK is an $1 + (ln|\mathcal{S}| - 1)\frac{U}{L}$ approximation algorithm when $|\mathcal{S}| > N_{MAX}$.*

Proof. Please see Appendix B. □

The approximation ratio depends on how we (approximately) solve the set cover problem and there are several ways to solve it, see [40] and references therein. For example, we could get a tighter bound if we use the analysis in [40] or we could use a more complicated approximation algorithm than the greedy one we use, to get an $O(lnf)$-approximation algorithm, where f is the maximum number of sets in which any element appears.

Furthermore, the approximation ratio depends on the lowest throughput L and the highest throughput U in the given topology. If we have a hotspot network topology, where all small cells are distributed in a small area, the values of L and U will be closer and the bound will become tighter. Otherwise, if we have a uniform network topology, where small cells are evenly distributed within a macro cell, the values of L and U will diverge more and the bound will become looser.

6.4. Feasibility check

We use results from [41,42] to check whether the coverage constraints are satisfiable. We start by stating two useful facts:

1. If any row of $\mathcal{A}$ has all 0's, obviously the coverage constraint cannot be satisfied.
2. Define a dominated set as follows: a row of $\mathcal{A}$ where there is only one 1 among its entries. If the number of dominated sets are larger than $\ln|\mathcal{S}|N_{MAX}$, the coverage constraints might not be satisfied by the Greedy Set-Cover based Algorithm.

It takes a polynomial time to check both facts above. Clearly, if the first fact holds, there is no solution to the problem. Every time we choose a small cell to be the cluster head we check whether the second fact holds, and, if yes, we conservatively do not select the small cell as a cluster head as it would not be guaranteed that the Greedy Set-Cover based Algorithm could satisfy the coverage constraints.

6.5. Complexity analysis

What remains is to prove that the algorithm can run in polynomial time.

Table 2
Simulation parameters.

Parameters	Values	
	Small-scale	Large-scale
Radius of the macro cell	200 m	500 m
Radius of a small cell	100 m	200 m
Transmission power of the macro cell	30 dBm	46 dBm
Transmission power of a small cell	27 dBm	30 dBm
Carrier frequency	60 GHz	
Available bandwidth	1 GHz	
Bandwidth for a frequency slot	100 MHz	
Antenna gain	5 dBi	
Noise power spectral density	−174 dBm/Hz	
Path loss exponent	5	

Theorem 7. *Given topology and $D_i \geq 1, \forall i \in S$, CovURK runs in polynomial time.*

Proof. The approximation algorithm requires two algorithms: one is Greedy Set-Cover based Algorithm and the other is ClosURK. Since the algorithm for mmHAUL without coverage constraints has been proved to be run in polynomial time in Theorem 4, the remaining part of this proof is that Greedy Set-Cover based Algorithm runs in polynomial time. Based on the results in [40], we know that the time complexity of Greedy Set-Cover based Algorithm is $O(|S|^2 \log |S|)$, which is indeed polynomial time. Therefore, the approximation algorithm runs in polynomial time. □

7. Performance evaluation

We study the performance of various schemes via simulations under two network scales: A) In a small-scale scenario (i.e., when the network consists of a small number of small cells and antennas), mmHAUL without and with coverage constraints are investigated to study the optimality of ClosURK and the performance guarantee of CovURK, respectively. B) In a real-life large-scale scenario (i.e., on a network with a large number of small cells and antennas), we use the same process to test how efficient both algorithms are when the size of the network scales up.

We use the following legends to label various algorithms. For mmHAUL without coverage constraints ($Q2$), we use "Upper_Bound" to refer to the upper bound of this problem (see Section 5.1). We use "Rounding" to refer to the algorithm in Section 5.2. Last, we use "Optimal_woSC" to refer to the optimal values for $Q2$. For mmHAUL with coverage constraints ($Q1$), we use "Optimal" to refer to the optimal values for $Q1$. Note that the optimal values of "Optimal_woSC" and "Optimal" are obtained by CVX in MATLAB with Gurobi as the optimal solver.

7.1. Small-scale simulations

First, we study the system throughput under a varying number of small cells and a varying number of antennas on the macro cell for mmHAUL without coverage constraints. Specifically, we vary the number of small cells from 1 to 20, and we vary the total number of antennas from 1 to 10. Second, we study the system throughput under the same simulation settings and varying parameters for mmHAUL. Finally, we plot the cluster head distribution for ClosURK, Optimal, and CovURK to discuss how the selection of clusters affects performance.

7.1.1. Simulation settings

The parameters used for the simulation are based on [15,43,44], and are listed in Table 2. The backhaul system is operated in 60 GHz. The transmission power of the macro cell is 30 dBm, the transmission power of a small cell is 27 dBm, the bandwidth of a frequency slot is 100 MHz, and the total available bandwidth is 1 GHz. The path-loss coefficient is 5, and the noise power equals −174 dBm/Hz. Note that the path-loss model is based on the UMi path loss model for LOS with a BS height of 10 m, see [44]. In this scenario, small cells in the network are uniformly distributed within the range of a macro cell. The radius of the macro cell and the small cells is 200 m and 100 m respectively.

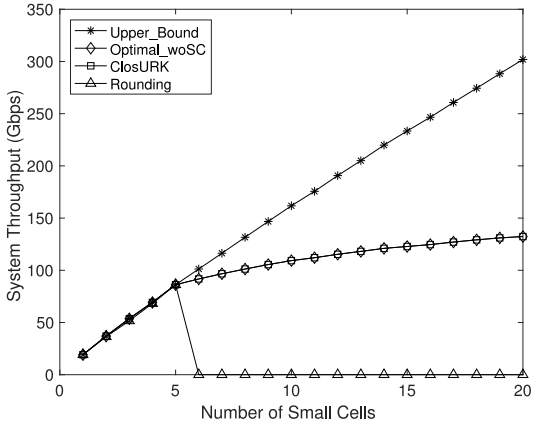

Fig. 4. System throughput of ClosURK v.s. number of small cells.

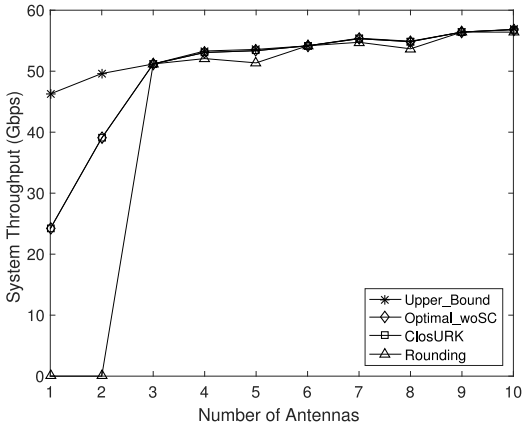

Fig. 5. System throughput of ClosURK v.s. number of antennas.

7.1.2. Simulation results — mmHAUL without coverage constraints

We first examine the performance for mmHAUL without coverage constraints ($Q2$). In Fig. 4, we show the results of ClosURK, Upper_Bound, Rounding, and Optimal_woSC in terms of the total system throughput under a varying number of small cells. For now we assume that there are 5 antennas on the macro cell.

First, when $|S| \leq N_{MAX}$, all of the algorithms perform similarly: Rounding has the smallest throughput, and ClosURK and Optimal_woSC have the same throughput in the middle of Upper_Bound and Rounding. The throughput difference between Upper_Bound and Rounding is a insignificant 0.2%. This is because the difference of the allocated number of antennas between before and after taking the floor function is less than 1, considered to be small in terms of the total system throughput.

Second, we analyze the performance of the algorithms when $|S| > N_{MAX}$. a) Upper_Bound keeps increasing when the number of small cells increases since it allows fractional number of antennas allocated in each cluster head. b) The performance of Rounding drops drastically: $N_i, \forall i \in S$ obtained from Upper_Bound is less than 1, which becomes 0 after Rounding. c) ClosURK and Optimal_woSC have the same performance no matter how many small cells are in the system.

Last, the system throughput of ClosURK and Optimal_woSC increases more slowly than that of Upper_Bound when the number of small cells increases. This shows the impact of linear programming

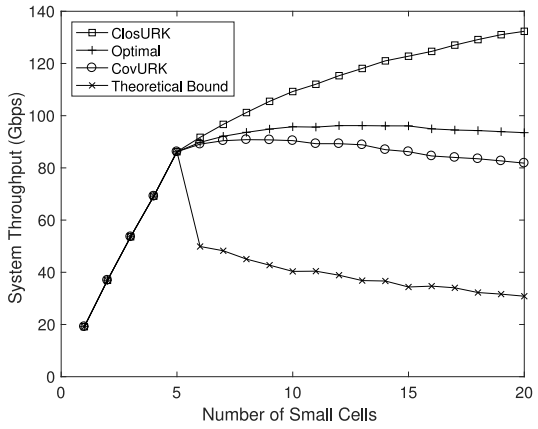

Fig. 6. System throughput of CovURK v.s. number of small cells. (Note that ClosURK does not satisfy the coverage constraints and this is why it is above the Optimal.).

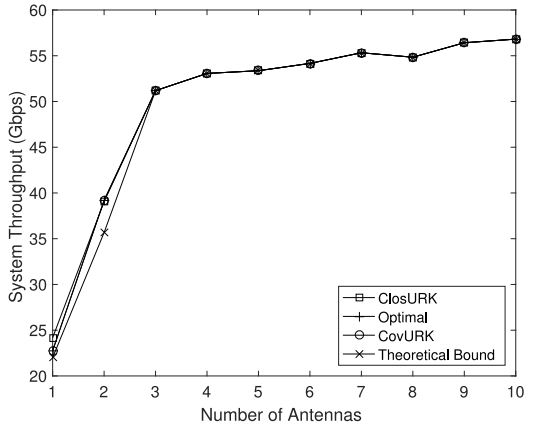

Fig. 7. System throughput of CovURK v.s. number of antennas.

relaxation in the algorithm. In the figure, the integrality gap in the algorithm is proportional to the number of small cells. In other words, the more small cells in the system, the harder it is to achieve the upper bound.

Fig. 5 shows the results of ClosURK, Upper_Bound, Rounding, and Optimal_woSC in terms of the system throughput under a varying number of antennas. We assume there are 3 small cells in the setting. In this figure, the performance of Rounding differs from that of the rest of the algorithms. When the number of antennas is a multiple of the number of small cells, the results of Rounding are the same as ClosURK and Optimal_woSC. Otherwise, there are gaps between Rounding and Optiaml_woSC. Notice that Greedy Knapsack Algorithm in ClosURK is the key step to fill those gaps, and to make ClosURK be the optimal solution of $Q2$.

7.1.3. Simulation results — mmHAUL with coverage constraints

In this section, we turn our focus on the original problem: mmHAUL ($Q1$). Fig. 6 shows the system throughput of ClosURK, Optimal, CovURK, and the theoretical lower bound under a varying number of small cells. Note that ClosURK does not satisfy the coverage constraint and thus is not a solution to $Q1$. We include it in this plot for reference only. Initially, we assume there are 5 antennas in the macro cell. Some observations can be summarized as follows.

First of all, when $|S| \leq N_{MAX}$, then CovURK, ClosURK, and Optimal have the same values of the system throughput. In other words, CovURK reaches the optimal values in this case. This result is expected since CovURK performs like ClosURK when $|S| \leq N_{MAX}$, and ClosURK is the optimal solution from Theorem 5.

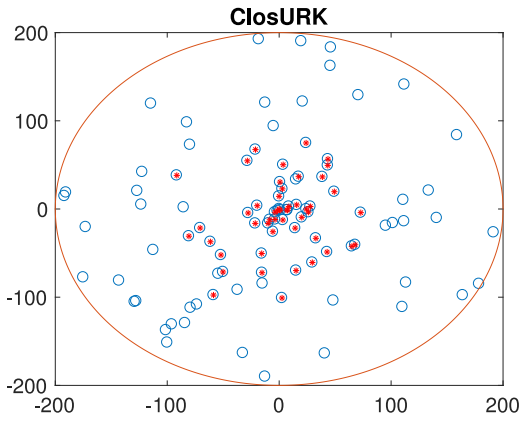

Fig. 8. Topology of cluster heads of ClosURK . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

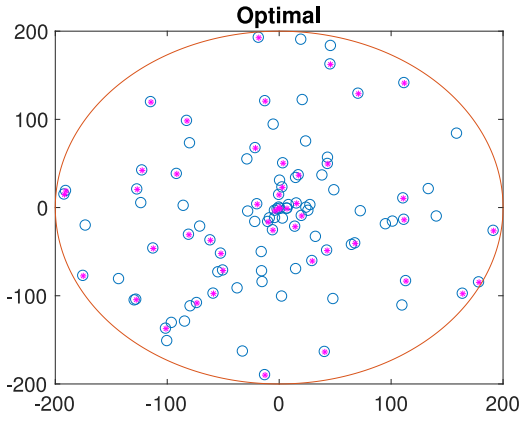

Fig. 9. Topology of cluster heads of Optimal . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

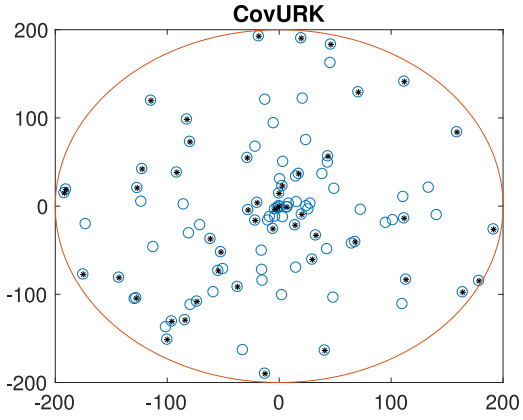

Fig. 10. Topology of cluster heads of CovURK . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

When $|S| > N_{MAX}$, the performance difference between ClosURK and Optimal is between 2.58% and 30.80%. Since ClosURK is the optimal solution of mmHAUL without coverage constraints and Optimal is the optimal value of mmHAUL with coverage constraints, this difference clearly shows the impact of the coverage constraints in the problem. The figure also shows that this difference is proportional to the number of small cells: more small cells require to be covered when the number of small cells increases, the performance therefore

is sacrificed when selecting more cluster heads with poor SINR. This result was expected.

Third, the performance difference between CovURK and Optimal is between 1.06% and 11.4% as the number of small cell increases. This shows that the approximation algorithm is near-optimal in most cases.

Interestingly, in this figure, the total system throughput of Optimal converges to 100 Gbps when the number of small cells increases. This is because the coverage constraints force the system to select the cluster heads that are relatively far from the macro cell such that traffic from remote small cells is relayed, resulting in a smaller total system throughput from these cluster heads as the number of small cells increases. Although there are still some leftover antennas to be allocated, because most of the antennas have already been used to satisfy the coverage constraints, the system cannot get a significant gain from these leftover antennas.

Moreover, the total system throughput of CovURK slightly decreases when the number of small cells increases. This is consistent with the fact that the performance bound of CovURK is proportional to the number of small cells: as the number of small cells increases, the Optimal converges to a fixed value, the distance between the optimal and CovURK increases and, thus, CovURK's performance decreases.

Lastly, based on our proof, the theoretical bound is inversely proportional to the number of small cells. In Fig. 6, the difference between Optimal and the theoretical bound is between 44.18% and 66.61% in the worst scenario when the number of small cells increases. Note that we can improve these results by implementing other algorithms for the Set Cover Problem as discussed in Section 6.3.

In Fig. 7, we show the results of ClosURK, Optimal, CovURK, and the theoretical bound in terms of the system throughput under a varying number of antennas. We assume 3 small cells in the simulation. In this figure, as expected, when the number of antennas increases, the performance also increases.

7.1.4. Topology of cluster heads

Figs. 8–10 illustrate the topology of cluster heads of ClosURK, Optimal, and CovURK respectively, to present their choices of the cluster heads. Note that in this simulation there are 5 antennas and 10 small cells in the range of the macro cell. The orange circle is the range of the macro cell, the blue small circles are the positions of small cells, and the red, pink, black stars are the positions of the cluster heads of ClosURK, Optimal, and CovURK, respectively. We run the same setting for 10 times. At each iteration, there are 5 small cells selected to be the cluster heads in ClosURK, Optimal, and CovURK. In Fig. 8, only the small cells in the middle, i.e., around the macro cell, are selected to be the cluster heads. As expected, since the small cells "close" to the macro cell have higher SNR values, we end up selecting them as cluster heads to maximize the system throughput. In Figs. 9 and 10, the results show that although both Optimal and CovURK have the same number of cluster heads in their networks, they select different sets of small cells to be the cluster heads. Specifically, Optimal selects the cluster heads closer to the macro cell whereas CovURK selects the cluster heads farther from the center. The main reason leading to this is that Greedy Set-Cover based Algorithm in CovURK is an approximation algorithm, not the optimal solution for satisfying the coverage constraints. Moreover, both figures produce different results compared to ClosURK since both algorithms need to take care of the coverage constraints.

7.1.5. Statistical significance

Fig. 11 shows confidence intervals for ClosURK, Optimal, and CovURK. Note that the simulation settings are the same as Fig. 6. As shown in the figure, the 95% confidence intervals of ClosURK, Optimal, and CovURK are $\sim$ 1 Gbps which corresponds to about 1% of variation off the mean. Thus, the average values of the performance for all the algorithms are representative, and, in the following sections we only present the average values.

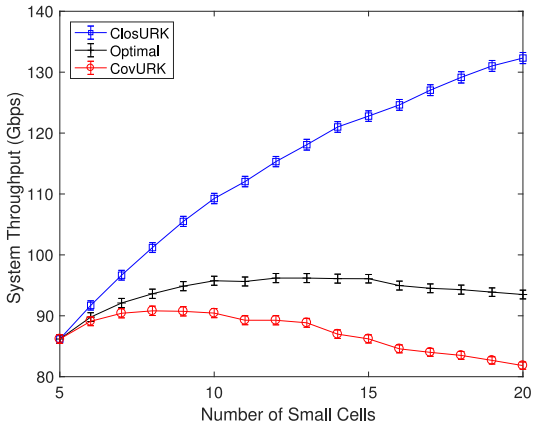

Fig. 11. System throughput with 95% confidence intervals v.s. number of small cells.

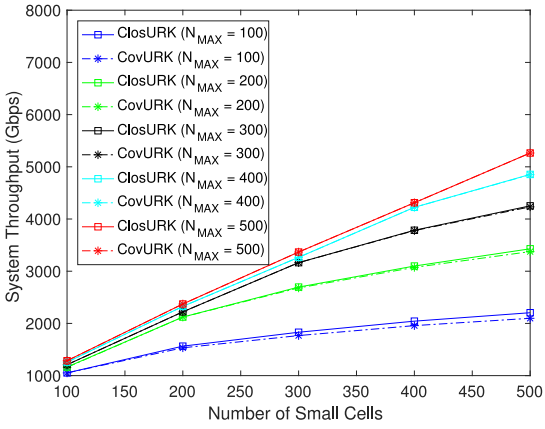

Fig. 12. System throughput of ClosURK and CovURK v.s. number of small cells.

7.2. Large-scale simulations

In this section, first, we study the system throughput under a varying number of small cells and a varying number of antennas on the macro cell for both ClosURK and CovURK. Specifically, we vary the number of small cells from 100 to 500, and we vary the total number of antennas from 100 to 500. Then, we compare ClosURK and CovURK with the algorithms of prior work to show how efficient our algorithms are.

7.2.1. Simulation settings

The same settings as in the small scale simulations are used, except that now the macro cell has a radius of 500 m and transmits with power 46 dBm, and each small cell has a radius of 200 m and transmits with power 30 dBm, see TABLE 2.

7.2.2. Varying number of small cells and antennas

In Fig. 12, we plot the performance of ClosURK and CovURK for a varying number of small cells and antennas together. As expected, the more antennas in the system, the higher system throughput we can obtain. Also, the most important observation from this figure is that CovURK has near-optimal performance in large-scale simulation. Note that the optimal values are in the middle of ClosURK and CovURK. Since the performance of CovURK is close to that of ClosURK as well as the optimal values, we conclude ConURK is near optimal.

In addition to the above-mentioned results, some observations are summarized as follows. First, for every pair of ClosURK and CovURK with the same number of antennas, there is a performance difference between these two algorithms when the number of small cells is larger

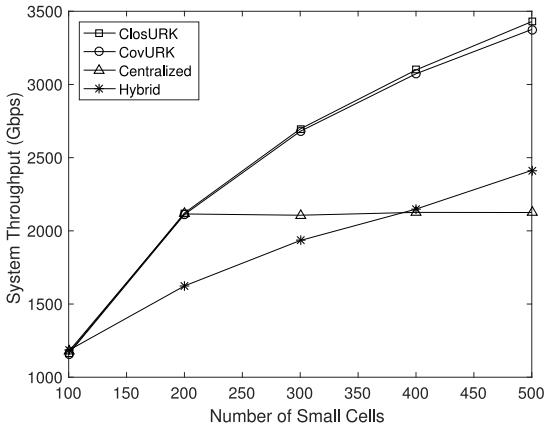

Fig. 13. System throughput of different architectures v.s. number of small cells.

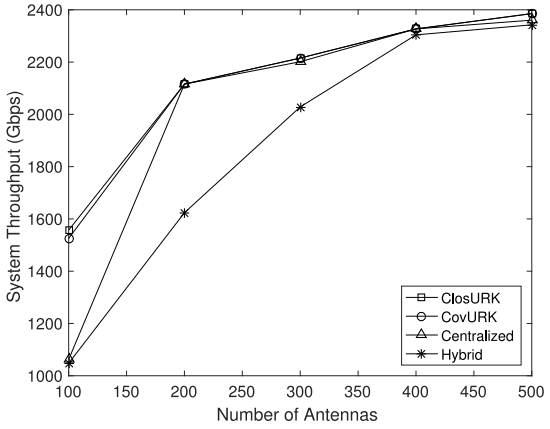

Fig. 14. System throughput of different architectures v.s. number of antennas.

than the number of antennas. Moreover, in the more constrained setting (i.e., a network with less antennas and more small cells), the performance difference becomes larger since the cluster heads selection by Greedy Set-Cover based Algorithm is more different from ClosURK and CovURK. Second, for both ClosURK and CovURK with the same number of antennas, the throughput increases slowly as long as the number of small cells is larger than the number of antennas. This is because every small cell can only get one antenna in that case. Note that in this case the system throughput will converge to the upper limit, $U \cdot N_{MAX}$, where U is the maximum SINR in Definition 2.

7.2.3. Comparison to prior work

Last, we study the performance of ClosURK and CovURK compared with the following prior studies under a varying number of small cells and a varying number of antennas in Figs. 13 and 14, respectively.

- **Centralized architectures ("Centralized", [16–20]):** In the case of $|S| \leq N_{MAX}$, Centralized evenly distributes its antennas to each cluster head. Moreover, in the case of $|S| > N_{MAX}$, Centralized makes its small cells use one antenna in turns.
- **Hybrid architectures ("Hybrid", [23,29]):** Hybrid has the same steps of CovURK: Greedy Set-Cover based Algorithm and the throughput enhancement. Instead of using Greedy Knapsack Algorithm for the throughput enhancement like CovURK, Hybrid uses random antenna allocation.

Fig. 13 shows the performance of ClosURK, CovURK, Centralized, and Hybrid under a varying number of small cells. We assume that the macro cell has 200 antennas in the simulation. As shown in the figure,

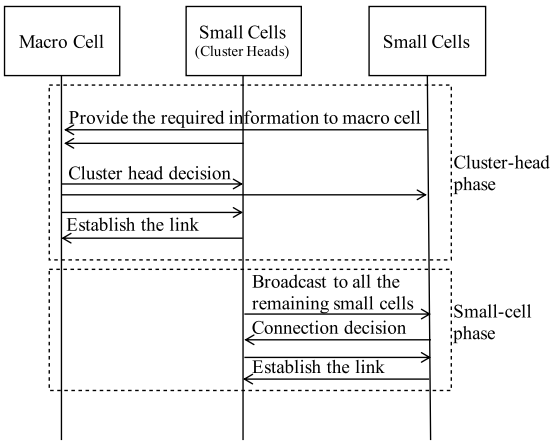

Fig. 15. Signaling flow for mmHAUL.

CovURK outperforms the others in most cases. As expected, Hybrid's performance is lower than that of ClosURK and CovURK since we choose the best way to allocate the antennas instead of randomly allocating them. In particular, CovURK can have a 30%–43% performance gain compared with Hybrid. When the number of small cells increases, the gap between Hybrid and ClosURK/CovURK becomes larger. This is because the probability that Hybrid has the same set of cluster heads as ClosURK and CovURK is smaller when there are more small cells in the network. In addition, Centralized has the same performance as ClosURK and CovURK when $|S| \leq N_{MAX}$ because in this case all the algorithms distribute the antennas evenly. However, when the number of small cells is larger than the number of antennas, small cells under the Centralized approach use the antennas in turns. When there are more small cells in the network, the performance cannot be more improved.

Fig. 14 shows the performance of ClosURK, CovURK, Centralized, and Hybrid under a varying number of antennas. We assume there are 200 small cells in the simulation. Similarly to before and as shown in the figure, CovURK outperforms the other algorithms in most cases. As expected, the system throughput of Hybrid will be closer to ClosURK and CovURK when the number of antennas increases since the probability that Hybrid has a similar trend as our schemes becomes larger. The performance gain of CovURK is 2%–46%. On the other hand, Centralized has the similar trend like ClosURK and CovURK when $|S| \leq N_{MAX}$ because all of these algorithms evenly distribute the antennas to the cluster heads. The performance gain of CovURK is 43.61% compared with Centralized when $|S| > N_{MAX}$.

The results establish that CovURK outperforms prior suggested algorithms by more than 40% depending on various scenarios.

8. Practical considerations

In this section we discuss practical issues related to the signaling between BSs and the computational requirements to run our algorithms in a real world context that would implement our schemes, and to the cost and availability of various hybrid beam-forming architectures.

8.1. Signaling and computational overhead

We adopt the structure of the existing 5G signaling flow [2,3].

Fig. 15 provides the detail procedure of the information exchange between the macro cell and small cells/cluster heads. There are two parts in the signaling flow: the cluster-head phase, which connects the macro cell to cluster heads, and the small-cell phase, which connects the remaining small cells to cluster heads. The cluster-head phase executes either the CovURK or the ClosURK algorithms by proceeding

as follows: (1) The beacons which are broadcasted by small cells are used by the macro cell to estimate the channel between small cells and itself. (2) The macro cell jointly decides the cluster head selection and antenna allocation on the links between itself and the cluster heads. (3) The macro cell broadcasts the decision to every small cell under its coverage and the links between the cluster heads and the macro cell are established. Then, the small-cell phase proceeds as follows: (1) The cluster heads broadcast the capacity of their link to the macro cell, and the number of already connected small cells, to all the remaining small cells. (2) The unconnected small cells make a connection to the cluster head that they prefer.

Note that the small-cell phase is very similar to the well know user-cell association problem in the context of a multi-cell network, see, for example [45,46], and thus we do not investigate it in this paper. It is worth to mention however that to tackle this problem one may use a centralized approach, where the macro cell optimally allocates unconnected small cells to cluster heads, or a distributed approach, where unconnected small cells pick the cluster head of their choice. In the later case, which we adopt here for scalability reasons [45], the ordering by which unconnected small cells select cluster heads affects the final outcome. To see this, when a new/unconnected small cell picks a cluster head, its decision will depend on the number of already connected small cells to this cluster head, as it would have to share the capacity of the cluster head with the rest of the connected small cells. As a last point, recall from Section 3 that we keep the same cluster heads for a period of time given that changing cluster heads takes time due to beam alignment challenges. Hence, the cluster-head phase described above will not be repeated every time a new small cell joins the network, whereas the small-cell phase will obviously be executed.

We finish this section with a discussion on the computational overhead to run our algorithms. We have recorded the time that it takes to run ClosURK and CovURK on a MacBook Pro and it takes less than 1 s to run either one of them for large scale scenarios involving hundreds of antennas and hundreds of small cells. Specifically, it takes about 0.6 s to run ClosURK and 0.9 to run CovURK. Today's macro BSs have far larger larger computational and storage capacity than a MacBook Pro, and the same will be true for upcoming cloud based BSs [47]. What is more, the antenna partitioning and cluster head selection decisions are expected to be updated at hourly timescales [48]. Thus, the computational overhead of our schemes is negligible.

8.2. Partially-connected hybrid beamforming architectures

Fully connected beamforming architectures (see Fig. 3) are becoming available but are rather costly and 5G providers may choose to avoid them in the context of small cells, at least till they become more affordable. Motivated by this, we discuss partially-connected hybrid beamforming architectures and how the problem under study would modified in this case. There are clearly many ways one can design a partially-connected hybrid beamforming architecture. For the shake of discussion, we consider the simplest and thus least expensive one, which is offered by many antenna manufacturers today: each RF chain is connected to a disjoint subset of the complete set of antenna elements.

Specifically, let B_{MAX} be the total number of RF chains and recall that N_{MAX} denotes the total number of antenna elements. As shown in Fig. 16, under the partially-connected architecture we consider, each RF chain connects to k antenna elements with $k \times B_{MAX} = N_{MAX}$. Now, let B_i be a decision variable of how many RF chains are used in small cell i. We can formulate the problem under study as follows:

$$Q4: \max_{y_{i,j},N_i,B_i} \sum_{\forall i \in S} C_i(N_i, D_i) \tag{15}$$

s.t.

$$y_{i,j} \leq N_i \cdot a_{i,j}, \ \forall i,j \in S \tag{16}$$

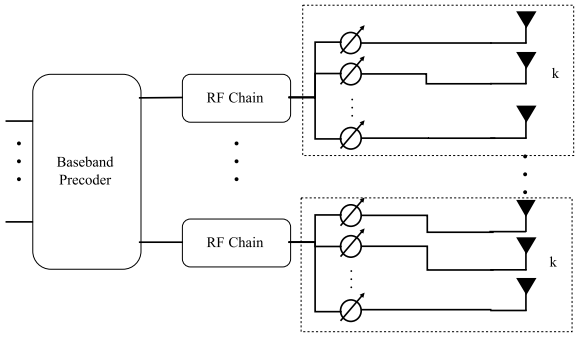

Fig. 16. Partially-Connected Hybrid Beamforming Architecture.

$$\sum_{\forall i \in S} y_{i,j} = 1, \ \forall j \in S \tag{17}$$

$$\sum_{\forall i \in S} N_i \leq N_{MAX}, \tag{18}$$

$$\sum_{\forall i \in S} B_i \leq B_{MAX}, \tag{19}$$

$$kB_i \leq N_i \leq k(B_i + 1), \tag{20}$$

$$y_{i,j} = \{0,1\}, \ \forall i,j \in S$$

$$N_i \in \mathbb{Z}+, \ \forall i \in S$$

$$B_i \in \mathbb{Z}+, \ \forall i \in S$$

where (16)–(18) are exactly the same as (4)–(6), (19) guarantees that the total number of used RF chains for the links between cluster heads and the macro cell cannot exceed the number of available RF chains at the macro cell, and (20) assures that there are sufficient number of RF chains to support the number of required antennas for small cell i.

Note that $Q4$ is, as expected, again an NP-hard problem. In the following discussion, we provide a polynomial-time solution which reuses the CovURK algorithm and a modified version of the ClosURK algorithm introduced before. The modified ClosURK algorithm has three steps like the original: Upper Bound, Rounding, and Greedy Knapsack Problem, and tries to solve the following problem:

$$Q5: \max_{B_i} \sum_{\forall i \in S} C_i(k \times B_i, D_i) \tag{21}$$

s.t.

$$\sum_{\forall i \in S} B_i \leq B_{MAX}, \tag{22}$$

$$B_i \in \mathbb{Z}+. \forall i \in S$$

The idea is to allocate k antenna elements when applying the waterfilling algorithm in the Upper Bound step and to leverage the rest of the steps in ClosURK as before (see Section 5.1). Then, the clustering part is still solved by CovURK. Note that there is no performance guarantee for the solution since the modified ClosURK does not necessarily find the optimal antenna partition in this case.

9. Conclusion

Connecting densely-deployed small cells with a macro cell in the context of future 5G networks is a challenging problem. We use a mmWave wireless backhauling network to do so. Specifically, we create an intermediate relay of cluster heads for the small cells to connect to the macro cell via these cluster heads. We jointly optimize the selection of cluster heads and the allocation of antennas of the macro cell to the mmWave links between the macro cell and the cluster heads to maximize the throughput. We introduce an approximation algorithm, CovURK, which can solve this problem in polynomial time. Compared to previous works, CovURK demonstrates an overall performance gain of 40% .

Acknowledgments

This work was supported by the National Science Foundation, USA [grant numbers ECCS-1444060 and NeTS-1618450], by Futurewei Technologies, USA, and by CISCO Systems, USA via a Cisco Research Center grant.

Appendix A. Proof of Theorem 3

We first analyze some properties of the logarithmic objective function in this problem. Then, based on these properties, we prove the optimality of ClosURK. The following proofs are based on the fact that $D_i \geq 1, \forall i \in \mathcal{S}$ for mmWave communication.

Lemma 1. *Given $D_i \geq 1, \forall i \in \mathcal{S}$, and $D_i \geq D_j, \forall i, j \in \mathcal{S}$, and $i \neq j$, $C(N, D_i) - C(N-1, D_i) > C(N+1, D_j) - C(N, D_j)$.*

Proof. We prove this lemma by contradiction. Assume given $D_i \geq 1, \forall i \in \mathcal{S}$, and $D_i \geq D_j, \forall i, j \in \mathcal{S}$, and $i \neq j$, we examine the relation below.

$$
\begin{aligned}
& C(N, D_i) - C(N-1, D_i) \leq C(N+1, D_j) - C(N, D_j) \\
\Longrightarrow\ & W \log_2(1 + N D_i) - W \log_2(1 + (N-1) D_i) \\
& \quad \leq W \log_2(1 + (N+1) D_j) - W \log_2(1 + N D_j) \\
\Longrightarrow\ & \log_2\left(\frac{1 + N D_i}{1 + (N-1) D_i}\right) \leq \log_2\left(\frac{1 + (N+1) D_j}{1 + N D_j}\right) \\
\Longrightarrow\ & \frac{1 + N D_i}{1 + (N-1) D_i} \leq \frac{1 + (N+1) D_j}{1 + N D_j} \\
\Longrightarrow\ & D_i D_j \leq D_j - D_i
\end{aligned}
$$

Since D_i and D_j are larger than or equal to 1, the left-hand-side is always larger than or equal to 1. Also, since $D_i \geq D_j$, the right-hand-side is always less than or equal to 0. The above relation is impossible to be true. Therefore, it reaches to the contradiction, and the lemma holds. □

Lemma 2. *Given $D_i \geq 1, \forall i \in \mathcal{S}$, and $D_i < D_j, \forall i, j \in \mathcal{S}$, and $i \neq j$, $C(N, D_i) - C(N-1, D_i) > C(N+1, D_j) - C(N, D_j)$.*

Proof. We prove this lemma by contradiction. Assume given $D_i \geq 1, \forall i \in \mathcal{S}$, and $D_i < D_j, \forall i, j \in \mathcal{S}$, and $i \neq j$, we examine the relation below.

$$
\begin{aligned}
& C(N, D_i) - C(N-1, D_i) \leq C(N+1, D_j) - C(N, D_j) \\
\Longrightarrow\ & W \log_2(1 + N D_i) - W \log_2(1 + (N-1) D_i) \\
& \quad \leq W \log_2(1 + (N+1) D_j) - W \log_2(1 + N D_j) \\
\Longrightarrow\ & \log_2\left(\frac{1 + N D_i}{1 + (N-1) D_i}\right) \leq \log_2\left(\frac{1 + (N+1) D_j}{1 + N D_j}\right) \\
\Longrightarrow\ & \frac{1 + N D_i}{1 + (N-1) D_i} \leq \frac{1 + (N+1) D_j}{1 + N D_j} \\
\Longrightarrow\ & D_i D_j \leq D_j - D_i \\
\Longrightarrow\ & D_i \leq 1 - \frac{D_i}{D_j} \quad (\because D_j \geq 1)
\end{aligned}
$$

Since D_i are larger than or equal to 1, the left-hand-side is always larger than or equal to 1. Also, since $D_i < D_j$, the right-hand-side is always less than or equal to 1. The above relation is impossible to be true. Therefore, it reaches to the contradiction, and the lemma holds. □

Based on the lemmas, given $D_i \geq 1, \forall i \in \mathcal{S}$, the following statement is always true.

Corollary 1. *Suppose all small cells are assigned with N antennas each and there are n leftover antennas, where $n \leq |S|$. Then, the policy that allocates $N+1$ antennas to exactly n small cells is better than a policy which, among these n small cells, allocates $N+2$ antennas in one cell, N antennas in another, and $N+1$ in the rest of them.*

Proof. Assume we arbitrarily choose n small cells to be with $N+1$ antennas and the rest of small cells with N antennas. Based on the results of Lemmas 1 and 2, we know that given $D_i \geq 1, \forall i \in \mathcal{S}$, $C(N+1, D_i) - C(N, D_i) > C(N+2, D_j) - C(N+1, D_j)$. This tells us that we cannot improve the performance by moving one antenna from one of these n small cells to another small cell of these n small cells. More specifically, the performance reduction by removing one antenna from small cell i is $C(N+1, D_i) - C(N, D_i)$, and the performance gain by allocating one more antenna to small cell j is $C(N+2, D_j) - C(N+1, D_j)$. Since $C(N+1, D_i) - C(N, D_i) > C(N+2, D_j) - C(N+1, D_j)$, the performance gain will not be larger than or equal to the performance reduction. Therefore, the policy that exactly n small cells have $N+1$ antenna is better than the policy that, among these n small cells, one of them has $N+2$ antennas, one of them has N antenna, and the rest of them have $N+1$ antenna. □

Theorem 3. *Given $D_i \geq 1, \forall i \in \mathcal{S}$, ClosURK is the optimal solution of $Q2$.*

Proof. Based on the result of Corollary 1, we know that before every small cell has the same number of antennas, we should not give any small cells more antennas than the maximum allocated number of antennas for each small cell. This implies that every small cell has at least $\lfloor \frac{N_{MAX}}{|S|} \rfloor$ antennas in the optimal solution, which is also the same as the results obtained from Rounding. Because $0 \leq \frac{1}{D_i} \leq 1$ in (10) and (11) ($D_i \geq 1$), the difference of $\frac{1}{D_i}$ does not exceed 1, which makes all $\tilde{N}_i = \lfloor \frac{N_{MAX}}{|S|} \rfloor$ after we execute Rounding. Therefore, $Q2$ can be reduced to $Q3$ after Step 2 in ClosURK. The only unsolved problem is how to allocate the leftover antennas, i.e., $N_{MAX} - \lfloor \frac{N_{MAX}}{|S|} \rfloor |S|$. Based on Theorem 2, Greedy Knapsack Algorithm is the optimal solution of $Q3$, which is also the optimal solution of $Q2$. In conclusion, ClosURK is the optimal solution of $Q2$. □

Appendix B. Proof of Theorem 6

Definition 3. Let OPT denote the maximum sum rate of mmHAUL ($Q1$), and $CovURK$ denote the sum rate obtained by CovURK.

Definition 4. Let n_{OPT} denote the number of small cells allocated one antenna by the optimal solution when $|S| > N_{MAX}$, and $\tilde{n}_{OPT}$ denote the number of small cells allocated one antenna for covering the other small cells when $|S| > N_{MAX}$.

Note that when $|S| > N_{MAX}$, every small cell can only get one antenna. In other words, $n_{OPT} = N_{MAX}$.

Lemma 3. *For a given topology and $D_i \geq 1, \forall i \in \mathcal{S}$, $L \cdot n_{OPT} \leq OPT \leq U \cdot n_{OPT}$.*

Proof. We can directly get this result by Definitions 2–4. □

Lemma 4. *For a given topology and $D_i \geq 1, \forall i \in \mathcal{S}$, $L \cdot n_{OPT} \leq CovURK$.*

Proof. When $|S| > N_{MAX}$, every small cell in either OPT or CovURK can only get one antenna. Therefore, $n_{CovURK} = n_{OPT} = N_{MAX}$. With the results of Definitions 2–4, $L \cdot n_{OPT} \leq CovURK$ can be easily stated. □

Theorem 6. *For a given topology and $D_i \geq 1, \forall i \in \mathcal{S}$, CovURK is $1 + (ln|S| - 1)\frac{U}{L}$ approximation algorithm when $|S| > N_{MAX}$.*

Proof. We prove this by construction. We use the general result of set cover problem, and then based on this result, provide the performance guarantee for CovURK. In the worst case scenario, i.e., none of the cluster heads selected by CovURK are the cluster heads for OPT. Recall that the number of cluster heads selected by the optimal solution $\tilde{n}_{OPT}$. Based on [40], the number of cluster heads selected by CovURK is at most $\ln|S|\tilde{n}_{OPT}$ since CovURK uses Greedy Set-Cover based Algorithm. Then, we can give the proof as follows.

$$\begin{aligned}
OPT - CovURK &\leq ClosURK(n_{OPT} - \tilde{n}_{OPT}) \\
&\quad - ClosURK(n_{OPT} - \ln|S|\tilde{n}_{OPT}) \quad (a) \\
&\leq (n_{OPT} - \tilde{n}_{OPT})U \\
&\quad - (n_{OPT} - \ln|S|\tilde{n}_{OPT})U \quad (b) \\
&= (\ln|S|\tilde{n}_{OPT} - \tilde{n}_{OPT})U \\
&= (\ln|S| - 1)\tilde{n}_{OPT}U \\
&\leq (\ln|S| - 1)n_{OPT}U \\
&= (\ln|S| - 1)\frac{U}{L}n_{OPT}L \\
&\leq (\ln|S| - 1)\frac{U}{L}CovURK \quad (c)
\end{aligned}$$

$$\implies OPT \leq \left(1 + (ln|S| - 1)\frac{U}{L}\right)CovURK$$

(a) follows the results of Theorem 3, (b) follows the results of Lemma 3, and (c) follows the results of Lemma 4. □

References

[1] P. Huang, K. Psounis, Efficient mmWave wireless backhauling for dense small-cell deployments, in: 2017 IEEE WONS, 2017.
[2] J. Rodriguez, Fundamentals of 5G Mobile Networks, John Wiley and Sons, 2015.
[3] P. Huang, H. Kao, W. Liao, Cross-tier cooperation for optimal resource utilization in ultra-dense heterogeneous networks, IEEE Trans. Veh. Technol. 66 (12) (2017) 11193–11207.
[4] Ultra dense network (udn) white paper, Technical Report, Nokia, 2016.
[5] 1000x: more small cells hyper-dense small cell deployments, Technical Report, Qualcomm, 2014.
[6] Backhaul technologies for small cells: use cases, requirements, and solutions, Technical Report, Small Cell Forum, 2013.
[7] U. Siddique, H. Tabassum, E. Hossain, D.I. Kim, Wireless backhauling of 5g small cells: challenges and solution approaches, IEEE Wirel. Commun. 22 (5) (2015) 22–31.
[8] R. Baldemair, T. Irnich, K. Balachandran, et al., Ultra-dense networks in millimeter-wave frequencies, IEEE Commun. Mag. 53 (1) (2015) 202–208.
[9] S. Sun, T.S. Rappaport, R.W. Heath Jr., et al., Mimo for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both? IEEE Commun. Mag. 52 (12) (2014) 110–121.
[10] A. Adhikary, E. Al Safadi, M.K. Samimi, R. Wang, G. Caire, T.S. Rappaport, A.F. Molisch, Joint spatial division and multiplexing for mmwave channels, IEEE J. Sel. Areas Commun. 32 (6) (2014) 1239–1255.
[11] S. Singh, R. Mudumdai, U. Madhow, Interference analysis for highly directional 60-GHz mesh networks, IEEE/ACM Trans. Netw. 19 (5) (2011) 1513–1527.
[12] O. Semiari, W. Saad, Z. Dawy, M. Bennis, Matching theory for backhaul management in small cell networks with mmwave capabilities, in: 2015 IEEE ICC, 2015.
[13] X. Xu, W. Saad, X. Zhang, et al., Joint deployment of small cells and wireless backhaul links in next-generation networks, IEEE Commun. Lett. 19 (12) (2015) 2250–2253.
[14] E. Karamad, R.S. Adve, Y. Lostanlen, et al., Optimizing placements of backhaul hubs and orientations of antennas in small cell networks, in: 2015 IEEE ICC Workshop on BackNets, 2015.
[15] S. Singh, M.N. Kulkarni, A. Ghosh, J.G. Andrews, Tractable model for rate in self-backhauled millimeter wave cellular networks, IEEE J. Sel. Area Commun. 33 (10) (2015) 2196–2211.
[16] R.J. Weiler, M. Peter, W. Keusgen, et al., Enabling 5g backhaul and access with millimeter-waves, in: 2014 IEEE EuCNC, 2014.
[17] J. Zhao, T.Q.S. Quek, Z. Lei, Heterogeneous cellular networks using wireless backhaul: fast admission control and large system analysis, IEEE J. Sel. Areas Commun. 33 (10) (2015) 2128–2143.
[18] S. Hur, T. Kim, D.J. Love, J.V. Krogmeier, T.A. Thomas, A. Ghosh, Millimeter wave beamforming for wireless backhaul and access in small cell networks, IEEE Trans. Commun. 61 (10) (2013) 4391–4403.
[19] H. Tabassum, A. Hamdi, E. Hossain, Analysis of massive mimo-enabled downlink wireless backhauling for full-duplex small cells, IEEE Trans. Commun. 64 (6) (2016) 2354–2369.
[20] N. Wang, E. Hossain, V.K. Bhargava, Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier hetnets with large-scale antenna arrays, IEEE Trans. Wirel. Commun. 15 (5) (2016) 3251–3268.
[21] X. Ge, H. Cheng, M. Guizani, T. Han, 5g wireless backhaul networks: challenges and research advances, IEEE Netw. Mag. 28 (6) (2014) 6–11.
[22] K. Zheng, L. Zhao, J. Mei, et al., 10 Gb/s HetSNets with millimeter-wave communications: access and networking - challenges and protocols, IEEE Commun. Mag. 53 (1) (2015) 222–231.
[23] M.Z. Islam, A. Sampath, A. Maharshi, et al., Wireless backhaul node placement for small cell networks, in: 2014 IEEE CISS, 2014.
[24] P. Hao, X. Yan, J. Li, et al., Flexible resource allocation in 5g ultra dense with self-backhaul, in: 2015 IEEE GLOBECOM, 2015.
[25] J. Garcia Rois, B. Lorenzo, F.J. Gonzalez-Castano, et al., Heterogeneous millimeter-wave/micro-wave architecture for 5g wireless access and backhauling, in: 2016 IEEE EuCNC, 2016.
[26] G. Zhang, T.Q.S. Quek, M. Kountouris, et al., Fundamentals of heterogeneous backhaul design - analysis and optimization, IEEE Trans. Commun. 64 (2) (2016) 876–889.
[27] Y. Zhu, Y. Niu, J. Li, et al., QoS-aware scheduling for small cell millimeter wave mesh backhaul, in: 2016 IEEE ICC, 2016.
[28] R. Li, P. Patras, WiHaul: Max-min fair wireless backhauling over multi-hop millimeter-wave links, in: 2016 ACM HotWireless, 2016.
[29] Y. Chiang, W. Liao, Mw-hierback: a cost-effective and robust millimeter wave hierarchical backhaul solution for hetnets, IEEE Trans. Mob. Comput. 16 (12) (2017) 3445–3458.
[30] O. Semiari, W. Saad, M. Bennis, Z. Dawy, Inter-operator resource management for millimeter wave multi-hop backhaul networks, IEEE Trans. Wirel. Commun. 16 (8) (2017) 5258–5272.
[31] G. Yang, M. Xiao, M. Alam, Y. Huang, Low-latency heterogeneous networks with millimeter-wave communications, IEEE Commun. Mag. 56 (6) (2018) 124–129.
[32] B. Panzer, Scheduling conflicts in wireless in-band backhaul for 5g millimeter wave communications, in: 2016 IEEE INFOCOM, 2016.
[33] T. Bai, R.W. Heath Jr., Coverage in dense millimeter wave cellular networks, in: 2013 IEEE Asilomar Conference on Signals, Systems and Computers, 2013.
[34] T.S. Rappaport, S. Sun, R. Mayzus, et al., Millimeter wave mobile communications for 5g cellular: it will work!, IEEE Access 1 (2013) 335–359.
[35] M.R. Akdeniz, Y. Liu, M.K. Samimi, et al., Millimeter wave channel modeling and cellular capacity evaluation, IEEE J. Sel. Areas Commun. 32 (6) (2014) 1164–1179.
[36] T. Halsig, D. Cvetkovski, E. Grass, B. Lankl, Measurement results for millimeter wave pure los mimo channels, in: 2017 IEEE WCNC, 2017.
[37] A. Alkhateeb M.N. Kulkarni, J.G. Andrews, A tractable model for per user rate in multiuser millimeter wave cellular networks, in: 2015 IEEE Asilomar Conference on Signals, Systems and Computers, 2015.
[38] J.G. Proakis, Digital Communication Systems, McGraw Hill, Boston, USA, 2001.
[39] S. Martello, P. Toth, Knapsack Problem: Algorithms and Computer Implementations, John Wiley and Sons, New York, NY, USA, 1990.
[40] D.-Z. Du, K.-I. Ko, X. Hu, Design and Analysis of Approximation Algorithms, Springer, 2012.
[41] P. Huang, S. Sun, W. Liao, Greencomp: energy-aware cooperation for green cellular networks, IEEE Trans. Mob. Comput. 16 (1) (2017) 143–157.
[42] H.M. Salkin, K. Mathur, Foundations of Integer Programming, North-Holland, Amsterdam, The Netherlands, 1989.
[43] H.S. Dhillon, G. Caire, Wireless backhaul networks: capacity bound, scalability analysis and design guidelines, IEEE Trans. Wireless Commun. 14 (11) (2015) 6043–6056.
[44] Tr 38.901 v14.0.0, study on channel model for frequencies from 0.5 to 100 GHz, Technical Report, 3GPP, 2017.
[45] W.C. Ao, K. Psounis, Approximation algorithms for online user association in multi-tier multi-cell mobile networks, IEEE/ACM Trans. Netw. 25 (4) (2017) 2361–2374.
[46] G. Athanasiou, P.C. Weeraddana, C. Fischione, L. Tassiulas, Optimizing client association for load balancing and fairness in millimeter-wave wireless networks, IEEE/ACM Trans. Netw. 23 (3) (2015) 846–850.
[47] N. Saxena, A. Roy, H. Kim, Traffic-aware cloud ran: a key for green 5g networks, IEEE J. Sel. Areas Commun. 34 (4) (2016) 1010–1021.
[48] Fengli Xu, Yong Li, Huandong Wang, Pengyu Zhang, Depeng Jin, Understanding mobile traffic patterns of large scale cellular towers in urban environment, IEEE/ACM Trans. Netw. 25 (2) (2017) 1147–1161.