

# Backhaul-Aware and Context-Aware User-Cell Association Approach

Mona Jaber<sup>1</sup>, Oluwakayode Onireti<sup>2</sup>, Muhammad Ali Imran<sup>2</sup>

<sup>1</sup> Fujitsu Laboratories of Europe, Hayes, Middlesex, UB4 8FE, UK

Email: monamrhb@gmail.com

<sup>2</sup> School of Engineering, University of Glasgow, Glasgow, G12 8QQ, UK

Email: {Oluwakayode.Onireti, Muhammad.Imran}@glasgow.ac.uk

## *Abstract—*

The cell range extension (CRE) has been successfully implemented to bias the user to base station (BS) association policy in a way that achieves load balancing and increases the capacity of heterogeneous networks. The user-centric backhaul (UCB) scheme is a CRE evolution that is both backhaul-aware and user-context-aware —two constraints that are shaping the 5G network development. In this work, we formulate and solve the multi-objective optimisation problem of the UCB user-BS association. We derive analytical expressions of the ergodic throughput resulting from the UCB and, accordingly, identify the optimum association policy. The study demonstrates the gain margins that can be realised with pertinent user-cell association which is aware of the end-to-end network limitations and users requirements.

## I. INTRODUCTION

The network evolution towards 5G is expected to rely on ultra-dense heterogeneous networks with multiple radio access technologies, radio access network (RAN) architectures (e.g., Cloud-RAN [1] and Fog-RAN [2]) and a hybrid backhaul (BH) network [3]. The 5G usage and service portfolio is also characterised with the diversification of network users and applications. As such, the user-BS association process in 5G is more complex and challenging than any other incumbent cellular generation. Traditionally, in mono-layer (referred to as macro-cell) networks, users select cells with the highest received signal. Lower cell layers are slowly growing beneath the macro-cell layer with the spread of small cells (SCs). These have lower transmitting power than macro-cells and their antennae are often situated at lower heights. The purpose of inserting SCs is generally to absorb traffic in hotspots or to extend the coverage to a blackspot (e.g., indoor). As the received signal from macro-cells is likely to be higher than that from SCs (due to the difference in transmitting power), a bias factor is used to influence users to select SCs. This feature is called the cell range extension (CRE) or expansion and is a mere bias on the cell ranking that does not entail any change in transmitting powers. The CRE has been proved to address the load balancing issues in a heterogenous network and to increase its holistic capacity. Authors in [4], assume that cells within one tier have common radio characteristics, such as, total transmit power, allocated spectrum; consequently, the optimum CRE-based user-tier association is derived.

Although the CRE succeeds in pushing the traffic to low network layers, and as such, in enhancing the radio resource efficiency, it is blind to the status of the BH and to users' distinctive requirements. The BH may be a bottleneck in emerging networks which are connected through a multi-hop and hybrid BH. Since the fast spread of SCs dictates fast solutions for backhauling, this often results in sub-optimal BH links. Consequently, although cells may have similar radio access capabilities to offer to potential users, their BH characteristics are likely to differ greatly. To this end, BH-aware cell selection schemes have become a focal point in joint radio access and BH optimisation. Specifically, different approaches of BH-aware implementation of CRE have been proposed in recent literature, such as, [5] and [6] which use an altered CRE scheme that is BH-delay-aware. However, accounting for the BH delay in the user-BS association does not suffice in 5G, knowing that the network users (devices and applications) have very different quality needs. The key concept in unlocking the constrained BH lies in a judicious and pertinent approach for resource allocation that is context-aware and considers the users' quality needs. Indeed, authors in [7] say that: "5G needs to be a chameleon technology that can adapt to differing demands of wireless services whether to support high bandwidth, low latency, bursty traffic, ultra-reliable services, or a combination of these capabilities". The user-centric-backhaul (UCB) is the first scheme to address the disparate needs of users and to adapt the cell selection accordingly, as detailed in [8].

In this work, we propose an analytical solution to the UCB scheme which, based on the network topology and BH characteristics, would yield the optimum association scheme. We use elements of stochastic geometry to represent the network topology. Accordingly, analytical expressions are derived and employed to identify the multi-objective optimum association policy when two BH technologies are randomly allocated to base stations in a mono-layer network. A recent work [9] proposes an analytical formulation to address the optimal biasing for maximising the user throughput in a heterogeneous network with constrained BH. The analytical work based on stochastic geometry is important as it demonstrates the benefits of dynamic biasing in terms of capacity gain. Similarly, authors in [10] apply the analytical performance modelling of wireless BH to a use-case based on the UCB with a single constraint based on holistic network capacity. Nonetheless, in

this paper, we present the first work to jointly address the context-aware and BH-aware multi-constraints in an analytical formulation.

## II. SYSTEM MODEL

User association strategies can be categorised as either *channel borrowing* or *traffic transfer* [11]. The UCB adopts the latter type of strategy through the usage of BH-aware and context-aware biasing. The UCB scheme proposes a multiple-attribute CRE approach in which each cell identifies a set of bias factors, each representative of its capabilities or constraints with respect to a given attribute. At the same time, users associate different weights to different attributes, hence are attracted differently by the candidate cells. The objective of the UCB is to maximise the throughput on each BH link without exceeding its limit and while satisfying the users' quality needs.

In this paper, the wireless network consists of one tier of SCs. The following system model is adopted to replicate and test the UCB scheme. The location of SCs is modeled as a two-dimensional homogeneous Poisson Point Process (PPP)  $\Phi_b$  with intensity  $\lambda_b$ . The location of users in the network is modeled as another independent homogeneous PPP  $\Phi_u$  with intensity  $\lambda_u$ . The adoption of PPP in the representation of SCs' locations is justified since such randomness may be expected in SCs' deployment. All SCs have identical radio characteristics (transmit power  $P$ ) and have a purposely over-dimensioned radio access (bandwidth  $W$ ). The downlink desired and interference signals from any SC are assumed to experience propagation loss with a path loss exponent  $\alpha$ . A user receives a power  $P \cdot H \cdot x^{-\alpha}$  from a SC at a distance  $x$ , where  $H$  is the random channel power gain (assumed to be Rayleigh distributed with average unit power, i.e.,  $H \sim \exp(1)$ ). The noise is assumed additive with power  $\sigma^2$ .

In this paper, the adopted notation is as follows:  $\Pr\{x\}$  indicates the probability of an event  $x$ ,  $E_x\{h(x, y)\}$  indicates the expected value of  $h(x, y)$  conditioned on the value  $y$ ,  $f_X(x)$  is the probability distribution function of the variable  $X$ ,  $F_X(x)$  is the cumulative distribution function of the variable  $X$ , and  $\bar{F}_X(x)$  is the complementary cumulative distribution function of the variable  $X$ . The notation  $\hat{X}$  is used to denote the rate of difference, for example the load of a link with capacity  $X'$  Mbps that carries  $X$  Mbps is  $\hat{X} = (X - X')/X$ .

### A. Heterogeneous backhaul and disparate user types

In this work, the last mile connecting any of the SCs to the gateway could employ any of the two technologies in  $\mathbf{B} = \{B_1, B_2\}$ . For instance,  $B_1$  and  $B_2$  may denote copper-based VDSL2<sup>1</sup> and wireless technologies, respectively. Let  $\mathbf{K} = \{K_1, \dots, K_K\}$  be the set of quality attributes that are considered in the multi-objective UCB optimisation problem. Such attributes may be throughput, latency, resilience, etc. In this work, let  $K_1$  and  $K_2$  indicate the throughput-related

and the resilience-related attributes, respectively. As such, the constraints of the BH links are defined according to these quality attributes, where  $Q_{1,k}$  indicates the capability of a SC with a last mile of type  $B_1$  with respect to quality  $k$  and  $Q_{1,1}$  represents its BH available capacity. A SC with a last mile of type  $B_2$  is characterised by  $Q_{2,k}$ , where  $Q_{2,1}$  represents its available capacity. Such a network can be modelled, using a stationary mixed PPP  $\Phi_b$  with a randomised intensity function  $L$  having a two-point distribution such that:

$$\Pr(L = \lambda_1) = p_{b,1}, \quad \Pr(L = \lambda_2) = p_{b,2} = 1 - p_{b,1}, \quad (1)$$

where,  $0 \leq p_{b,1} \leq 1$  is the probability of having a SC with last mile of type  $B_1$ , and  $p_{b,2}$  is the probability of having a SC with last mile of type  $B_2$ . Hence, the intensity of  $\Phi_b$  is  $\lambda_b = p_{b,1}\lambda_1 + (1 - p_{b,1})\lambda_2$ , where  $\lambda_1 > 0$  is the intensity of SC with last mile link of type  $B_1$  and  $\lambda_2 > 0$  is the intensity of SC with last mile link of type  $B_2$ .

Let  $\mathbf{U}$  be the set of total users in the network. We distinguish  $K$  types of users based on their quality targets with respect to the attributes in  $\mathbf{K}$ . A user  $u$  is thus characterised by two sets. The first represents the quality targets of user  $u$  with respect to each of the attributes in  $\mathbf{K}$ ,  $\mathbf{M}_u = \{M_{u,1}, \dots, M_{u,K}\}$ . The second set represents the weight associated by user  $u$  to the attributes in  $\mathbf{K}$ ,  $\omega_u = \{\omega_{u,1}, \dots, \omega_{u,K}\}$ . In this work, we assume that a user prioritises a single attribute by associating a full weight with the defined attribute and null weights to all other attributes, i.e., the user-SC association is based on a single attribute. To this end, the set  $\mathbf{U}$  can be divided into  $K$  non-overlapping subsets  $\mathbf{U}_k$ , as shown below:

- Set  $\mathbf{U}_1$  with  $\omega_v = \{\omega_{v,1} = 1, \omega_{v,2} = 0, \dots, \omega_{v,K} = 0\}$ ,  $\forall v \in \{1, \dots, |\mathbf{U}|\}$
- Set  $\mathbf{U}_k$  with  $\omega_v = \{\omega_{v,1} = 0, \dots, \omega_{v,k} = 1, \dots, \omega_{v,K} = 0\}$ ,  $\forall v \in \{1, \dots, |\mathbf{U}|\}$
- Set  $\mathbf{U}_K$  with  $\omega_v = \{\omega_{v,1} = 0, \dots, \omega_{v,K} = 1\}$ ,  $\forall v \in \{1, \dots, |\mathbf{U}|\}$

Each of the sets  $\mathbf{U}_k$  is disjoint because a user in  $\mathbf{U}$  can belong to exactly one of them. Let  $p_{u,k}$  be the average fraction of users belonging to the sets  $\mathbf{U}_k$ . Since the actual locations of users with respect to each other do not matter for the purpose of this work, each of  $\Phi_{u_k}$  can be equivalently modelled as independent PPPs with densities  $p_{u,k} \cdot \lambda_u$ . In other words, they can be modeled as thinned versions of the original process  $\Phi_u$  with retention probabilities  $p_{u,k}$  independent of the locations of the users.

### B. User-centric metrics

The metric we propose to measure the users' satisfaction gauges the gaps between each of the target and delivered quality of service (QoS) values, and is affected by the corresponding weight. Let  $M'_{u,K}$  represent the measured  $k$ -related QoS value delivered by the network to user  $u$ . For instance,  $M'_{u,1} \triangleq W \cdot \log_2(1 + \gamma)$ , where  $W$  is the channel bandwidth and  $\gamma$  is the signal-over-interference-and-noise-ratio (SINR). For the resilience-related attribute, the achieved QoS is  $M'_{u,2} = Q_{1,2}$ , if the user  $u$  is served by a SC with a last mile of type  $B_1$  and  $M'_{u,2} = Q_{2,2}$ , otherwise. We define

<sup>1</sup>Very-high-bit-rate digital subscriber line 2; a digital subscriber line (DSL) technology.

the level of dissatisfaction ( $\hat{M}_{u,k}$ ) of a user  $u$  with respect to the quality attribute  $k$  as follows:

$$\hat{M}_{u,k} \triangleq \omega_{u,k} \left( \frac{M'_{u,k} - M_{u,k}}{M_{u,k}} \right) \quad (2)$$

Consequently, optimising the association process between users and candidate SCs is constrained by  $K$  conditions  $\hat{M}_{u,k} \leq \theta_k \forall k = \{1, \dots, K\}$ , where  $\theta_k$  is the maximum allowed QoS gap for attribute  $k$ .

### C. UCB optimisation problem

The UCB optimisation problem consists of finding the optimum association policy  $\mathbf{A}(\mathbf{O}_1, \mathbf{O}_2) = [A_1(\mathbf{O}_1, \mathbf{O}_2), A_2(\mathbf{O}_1, \mathbf{O}_2)]$  that would maximise the ergodic throughput without exceeding the capacity limits of the BH links which are  $Q_{1,1}$  and  $Q_{2,1}$  for  $B_1$  and  $B_2$ , respectively. The factors  $A_1$  and  $A_2$  represent the proportion of users associated to SCs with last mile of type  $B_1$  and  $B_2$ , respectively. The association policy is a function of the bias (or offset) settings in both SC types  $(\mathbf{O}_1, \mathbf{O}_2)$  such that  $\mathbf{O}_s = \{O_{s,1}, \dots, O_{s,K} | O_{s,k} \geq 1\}$ , for  $s = \{1, 2\}$  and where  $O_{s,k}$  reflects the capability/constraint with respect to attribute  $k$  of the SC with type  $B_s$  last mile. In other words, if a lightly loaded SC has also a lightly loaded last mile, it would have a high  $O_{s,1}$  value; if either the radio interface or the last mile are loaded, it would have a low value instead. On the other hand, if a SC has a robust last mile BH, it would have a high  $O_{s,2}$  value; whereas an unreliable last mile would lead to a low  $O_{s,2}$  value.

The optimisation problem can, thus, be formulated as follows:

$$\begin{aligned} \max_{\mathbf{A}} \quad & T(\mathbf{A}) = A_1 \cdot T_1(\mathbf{A}) + A_2 \cdot T_2(\mathbf{A}) \quad (3) \\ \text{subject to} \quad & \end{aligned}$$

$$Q_{1,1} - T_1 \times G_1 \geq 0 \quad (4)$$

$$Q_{2,1} - T_2 \times G_2 \geq 0 \quad (5)$$

$$\hat{M}_{u,k} \leq \theta_k, \quad \forall k = \{1, \dots, K\} \quad (6)$$

where  $T_1$  and  $T_2$  are the average throughput of SCs with last mile of type  $B_1$  and  $B_2$ , respectively. The BH capacity constraints are (4) and (5), which dictate that the average throughput ( $T_1$  and  $T_2$ ) affected by the overhead  $G_1$  and  $G_2$ , should not exceed the BH nominal capacities  $Q_{1,1}$  and  $Q_{2,1}$ , respectively. The user-centric constraints are expressed in (6), where  $\hat{M}_{u,k}$  is obtained using the expression (2). These conditions ensure that the user-centric QoS gaps are within a user defined maximum that correspond to each attribute.

## III. ANALYTICAL SOLUTION

In order to solve the optimisation problem, we need to first derive the analytical expression for the system throughput. The ergodic throughput pertaining to the traditional user-SC association policy, is derived in many works that employ elements of stochastic geometry to capture the network topology (e.g., [12]). Similarly, biased user association policy is addressed in [4]. However, the UCB association policy is

the only one that employs multiple bias factors per SC and that distinguishes between different users' priorities. To this end, we first formulate the association policy in Section III-A. Next, we determine the distribution of the distance between a typical user and corresponding serving SC (based on the association policy) in Section III-B. Finally, we derive the ergodic throughput in Section III-C.

We denote the variable  $x$  as the distance separating the typical user located at the origin, without loss of generality, from the closest SC in  $\Phi_b$ . First, we find the complementary cumulative distribution function (ccdf) of  $x$ ,  $\bar{F}_X(x) = \Pr[X > x] = \exp(-\lambda\pi x^2)$  which may be derived using the null probability of a 2-D Poisson process, where  $\pi x^2$  is the area of the circle drawn by  $x$ . The probability distribution function (pdf) of  $x$  can then be obtained as the derivative of  $F_X(x)$ :  $f_X(x) = 2\pi\lambda \cdot x \cdot \exp(-\pi\lambda x^2)$ .

### A. Basic UCB association policy

A typical user  $u$  located at the centre ranks the candidate SCs according to the maximum biased average received power from all SCs with type  $B_s$  last mile, where  $s \in \{1, 2\}$  (i.e., the fading component is averaged out). The rank  $R_{u,s}$  attributed by the user  $u$  to a SC with type  $B_s$  last mile is:

$$R_{u,s} = P \cdot x_s^{-\alpha} \times \sum_k O_{s,k} \cdot \omega_{u,k}, \quad (7)$$

and the UCB association policy identifies the highest rank:  $R_u \triangleq \max_s R_{u,s}$ . The probability of a typical user associating with a SC with type  $B_s$  last mile is thus equivalent to the probability of  $R_{u,s} > R_{u,n}$  for all SCs with type  $B_n$  last mile. Based on our system model, a typical user either connects to SC with a last mile of type  $B_1$  or  $B_2$ . Thus, we need to define the association policy  $\{A_{1,k}, A_{2,k}\}$  of a typical user in  $U_k$  with distances  $X_{k,1}$  and  $X_{k,2}$  with respect to SCs with last mile of type  $B_1$  and  $B_2$ , respectively.

$$\begin{aligned} A_{1,k} &\triangleq \Pr[s = 1] = \mathbb{E}_{X_{k,1}} [\Pr[R_{k,1}(X_{k,1}) > R_{k,2}]] \\ &\stackrel{(a)}{=} \mathbb{E}_{X_{k,1}} \left[ \Pr \left[ X_{k,2} > X_{k,1} \left( \frac{O_{2,k}}{O_{1,k}} \right)^{\frac{1}{\alpha}} \right] \right] \\ &= \int_0^\infty \Pr \left[ X_{k,2} > X_1 \left( \frac{O_{2,k}}{O_{1,k}} \right)^{\frac{1}{\alpha}} \right] f_{X_{k,1}}(x) dx \\ &\stackrel{(b)}{=} \int_0^\infty e^{-\pi\lambda \left( \frac{O_{2,k}}{O_{1,k}} \right)^{\frac{2}{\alpha}} x^2} \times 2\pi\lambda \cdot x \cdot e^{-\pi\lambda x^2} dx \\ &\stackrel{(c)}{=} \frac{1}{\left( \frac{O_{2,k}}{O_{1,k}} \right)^{\frac{2}{\alpha}} + 1}. \end{aligned} \quad (8)$$

The equality (a) is a direct result of replacing  $R$  by the expression in (7) and (b) follows from using the ccdf expression  $\bar{F}_X(x)$ . The final result (c) is obtained by replacing  $y = x^2$  and a straight forward integration. The association probability

of a user in  $\mathbf{U}_k$  with SC with last mile of type  $B_2$  can be obtained in a similar manner as:

$$A_{2,k} = \frac{1}{\left(\frac{O_{1,k}}{O_{2,k}}\right)^{2/\alpha} + 1}, \quad (9)$$

and it can be verified that  $A_{1,k} + A_{2,k} = 1$ , as expected. The probabilities of any user associating with SCs with last mile of type  $B_1$  and  $B_2$  can be expressed as follows:

$$A_1 = \sum_{k=1}^K p_{u,k} A_{1,k} \quad (10)$$

$$A_2 = \sum_{k=1}^K p_{u,k} A_{2,k}. \quad (11)$$

### B. Policy-based user-cell distance

The minimum distance between a typical user and the closest SC in  $\Phi_b$  is  $x$ , such as  $f_X(x) = 2\pi\lambda x \cdot \exp(-\pi\lambda x^2)$ , as explained in Section II. We consider a typical user in  $\mathbf{U}_1$  at the origin associated with a SC with a last mile of type  $B_1$  or  $B_2$ , according to the UCB. Denote  $X_{k,1}$  (or  $X_{k,2}$ ) as the distance between the user and its serving BS. Since SCs are deployed as a PPP,  $X_{k,1}$  and  $X_{k,2}$  are random variables described by their probability density function  $f_{X_{k,1}}(x)$  and  $f_{X_{k,2}}(x)$  that we derive in this section.

$$\bar{F}_{X_{k,1}}(x) = \Pr[X_{k,1} > x | s = 1] = \frac{\Pr[X_{k,1} > x, s = 1]}{\Pr[s = 1]} \quad (12)$$

The probability  $\Pr[s = 1]$  is equal to  $A_{1,k}$ , as shown in Section III-A. The joint probability  $\Pr[X_{k,1} > x, s = 1]$  can be found as follows:

$$\begin{aligned} \Pr[X_{k,1} > x, s = 1] &= \Pr[X_{k,1} > x, R_{k,1}(X_{k,1}) > R_{k,2}] \\ &= \int_x^\infty \Pr[R_{k,1}(x) > R_{k,2}] f_{X_{k,1}}(x) dx \\ &= 2\pi\lambda \int_x^\infty x \cdot e^{-\pi\lambda x^2 \left(\left(\frac{O_{2,k}}{O_{1,k}}\right)^{2/\alpha} + 1\right)} dx \\ &= A_{1,k} \cdot e^{-\pi\lambda x^2 / A_{1,k}} \end{aligned} \quad (13)$$

Plugging (13) into (12) leads to the expression of the ccdf as follows:

$$\bar{F}_{X_{k,1}}(x) = e^{-\pi\lambda x^2 / A_{1,k}} \quad (14)$$

Consequently, the pdf of the policy-based user-SC distance can be found as follows:

$$\begin{aligned} f_{X_{k,1}}(x) &= \frac{dF_{X_{k,1}}(x)}{dx} = \frac{d(1 - \bar{F}_{X_{k,1}}(x))}{dx} \\ &= \frac{2\pi\lambda}{A_{1,k}} \cdot x \cdot e^{-x^2 \left(\frac{\pi\lambda}{A_{1,k}}\right)} \end{aligned} \quad (15)$$

Similarly, the pdf of the distance between a user in  $\mathbf{U}_1$  and a SC with last mile of type  $B_2$  according to UCB can be obtained as follows:

$$f_{X_{k,2}}(x) = \frac{2\pi\lambda}{A_{2,k}} \cdot x \cdot e^{-x^2 \left(\frac{\pi\lambda}{A_{2,k}}\right)} \quad (16)$$

### C. Ergodic throughput

In this section, we derive the average ergodic throughput  $T$  of a typical user that has  $p_{u,k}$  probability of being in set  $\mathbf{U}_k$ , where  $\sum_{k=1}^K p_{u,k} = 1$ . Moreover, this user has  $A_1$  probability of connecting to a SC with last mile of type  $B_1$  and  $A_2$  otherwise.

$$T = \underbrace{\sum_{k=1}^K p_{u,k} A_{1,k} T_{1,k}}_{T_1} + \underbrace{\sum_{k=1}^K p_{u,k} A_{2,k} T_{2,k}}_{T_2} \quad (17)$$

where,  $T_{1,k}$  and  $T_{2,k}$  are the average throughput of a typical user in  $\mathbf{U}_k$  associated with a SC with last mile of type  $B_1$  and  $B_2$ , respectively. The average throughput  $T_{s,u}(x_{u,s}, \gamma) \triangleq \mathbb{E}_x[\mathbb{E}_\gamma[W \log_2(1 + \gamma)]]$  is a function of two variables: the distance between user and associated SC ( $X_{u,s}$ ) and  $\gamma$ , the received SINR. The SINR of a typical user at a random distance  $x$  from its associated SC  $c$  is defined as follows:

$$\gamma(x) = \frac{P \cdot x^{-\alpha} \cdot H_c}{\sigma^2 + \sum_{i \in \Phi_b \setminus c} (P \cdot x_i^{-\alpha} \cdot H_i)} \quad (18)$$

where,  $H_c$  is the exponentially distributed serving channel power with unit mean,  $x_i$  is the distance between interfering SC  $i$  and the typical user at the origin, and  $H_i$  is the exponentially distributed interfering channel power with unit mean from SC  $i$ . The expression  $\mathbb{E}_\gamma[W \log_2(1 + \gamma(x))]$  is derived in [12] as follows:

$$\begin{aligned} &\mathbb{E}_\gamma[W \log_2(1 + \gamma(x))] \\ &= W \ln 2 \int_0^\infty \exp\left\{-\frac{2^\tau - 1}{\gamma'}\right\} \mathcal{L}_{I_x}(x^\alpha(2^\tau - 1)) d\tau \end{aligned} \quad (19)$$

where,  $\tau$  is the target throughput,  $\gamma'$  is the signal-over-noise ratio (SNR), and  $\mathcal{L}$  is the Laplace function as defined below:

$$\begin{aligned} &\mathcal{L}_{I_x}(x^\alpha(2^\tau - 1)) \\ &= \exp\left(-\pi\lambda x^2(2^\tau - 1)^{\frac{2}{\alpha}} \int_{(2^\tau - 1)^{-\frac{2}{\alpha}}}^\infty \frac{1}{1 + z^{\frac{\alpha}{2}}} dz\right) \end{aligned} \quad (20)$$

The general average ergodic throughput of a user  $u$  at a variable location associated with a SC with  $B_s$  last mile can be expressed as in (21) shown at the top of the next page. This can be further expanded by substituting (19) and (15), as shown in (22).

1) *Special interference-limited case:* For a particular case in which the radio coverage is interference limited, the SINR can be approximated with the signal-to-interference ratio ( $\text{SIR} = \dot{\gamma}$ ) (i.e.,  $\sum_{i \in \Phi_b \setminus c} (P \cdot x_i^{-\alpha} \cdot H_i) \gg \sigma^2$ ). If, in addition, we assume that  $\alpha = 4$ , the expected throughput at a defined distance can be further simplified as follows [12]:

$$\begin{aligned} &\mathbb{E}_\gamma[W \log_2(1 + \dot{\gamma}(x))] \\ &= W \ln 2 \int_{\tau > 0} \mathcal{L}_{I_x}(x^4(2^\tau - 1)) d\tau \\ &= W \ln 2 \int_{\tau > 0} e^{-\pi\lambda x^2 \sqrt{2^\tau - 1} \left(\frac{\pi}{2} - \arctan\left(\frac{1}{\sqrt{2^\tau - 1}}\right)\right)} d\tau \end{aligned} \quad (24)$$

$$T_{s,u} = \int_0^\infty \mathbb{E}_\gamma [W \log_2(1 + \gamma(x))] \times f_{X_{s,u}}(x) dx \quad (21)$$

$$T_{s,u} \stackrel{(a)}{=} \frac{2\pi\lambda}{A_{s,u}} \times W \ln 2 \int_0^\infty \int_0^\infty e^{-\pi\lambda \frac{x^2}{A_{s,u}}} \exp\left\{-\frac{2^\tau - 1}{\gamma'}\right\} \times \mathcal{L}_{I_x}(x^\alpha(2^\tau - 1)) d\tau dx \quad (22)$$

$$\begin{aligned} T_{s,u}(\gamma \sim \dot{\gamma}, \alpha = 4) &= \frac{2\pi\lambda}{A_{s,u}} \times W \ln 2 \int_0^\infty \int_0^\infty e^{-\pi\lambda \frac{x^2}{A_{s,u}}} \times \exp\left\{-\pi\lambda x^2 \sqrt{2^\tau - 1} \left(\frac{\pi}{2} - \arctan\left(\frac{1}{\sqrt{2^\tau - 1}}\right)\right)\right\} x dx d\tau \\ &= W \ln 2 \underbrace{\int_0^\infty \frac{1}{1 + A_{s,u} \times \sqrt{2^\tau - 1} \left(\frac{\pi}{2} - \arctan\left(\frac{1}{\sqrt{2^\tau - 1}}\right)\right)} d\tau}_{\mathcal{G}(A_{s,u})} \end{aligned} \quad (23)$$

Consequently, the expression of the average ergodic throughput of a user  $u$  at a variable location associated with a SC with  $B_s$  last mile, (22), can be further simplified to (23), shown in the top of the next page. Plugging (23) in (17), we get the average ergodic throughput of a typical user in the network (for the interference limited basic deployment with  $\alpha = 4$ ):

$$T = W \ln 2 \sum_s \sum_k A_{s,k} \times \mathcal{G}(A_{s,k}), \quad (25)$$

where the function  $\mathcal{G}(\cdot)$  is defined as follows:

$$\mathcal{G}(g) \triangleq \int_0^\infty \frac{1}{1 + g \times \sqrt{2^\tau - 1} \left(\frac{\pi}{2} - \arctan\left(\frac{1}{\sqrt{2^\tau - 1}}\right)\right)} d\tau \quad (26)$$

The average number of users of type  $U_k$  associated with a SC with  $B_s$  last mile, can be obtained as the ratio  $\mathcal{N}_{s,k} = p_{u,k} \cdot A_{s,k} \cdot \lambda_u / \lambda_b$ , where  $\lambda_u$  is the intensity of the users' represented by  $\Phi_u$  and  $\lambda_b$  is the density of SCs in  $\Phi_b$ , as defined in Section II. For the given particular case, the average throughput of a SC with  $B_s$  type of last mile can thus be expressed as follows.:

$$\begin{aligned} T_s &= \sum_k \mathcal{N}_{s,k} \times T_{s,k} \\ &= \frac{\lambda_u}{\lambda_b} \times W \ln 2 \sum_k p_{u,k} \cdot A_{s,k} \times \mathcal{G}(A_{s,k}) \end{aligned} \quad (27)$$

#### IV. RESULTS AND ANALYSIS

In this section, we solve the UCB optimisation problem in which we consider two optimisation attributes (i.e.,  $K = 2$ ): Throughput and Resilience. Hence, there are two user types in the network,  $U_1$  and  $U_2$ . Moreover, we consider that users of both types share the same QoS targets but associate different weights, as shown below:

- Set  $\mathbf{U}_1$  with  $\omega_1 = \{\omega_{1,1} = 1, \omega_{1,2} = 0\}$ ,
- Set  $\mathbf{U}_2$  with  $\omega_2 = \{\omega_{2,1} = 0, \omega_{2,2} = 1\}$ ,
- $\mathbf{M}_1 = \mathbf{M}_2 = \{M_1, M_2\}$ .

Each SC has a single last mile BH employing either VDSL2 technology or V-band millimeter wave (71 – 76 GHz and 81 – 86 GHz). VDSL2 is highly resilient with limited throughput while the V-band is less robust with ample capacity. The network-centric objective is to maximise the cumulative throughput without exceeding the capacity of the BH links.

The user-centric constraints insure that the QoS gap for both types of users is limited to the defined threshold.

##### A. Exhaustive search results

The optimisation problem is solved by using an exhaustive search approach. There are two independent variables that define the association policy  $A_{1,1}$  and  $A_{1,2}$ , which indicate the percentage of users of type  $U_1$  that associate with SCs having a VDSL2 BH, and users of type  $U_2$  that choose the same SCs, respectively. A feasible solution would entail  $A_{2,1} = 1 - A_{1,1}$  and  $A_{2,2} = 1 - A_{1,2}$  percentage of users of each type associating with SCs having a V-band BH (i.e., any user is served by exactly one SC). Accordingly, we explore all possible values of  $A_{1,1}$  and  $A_{1,2}$  and identify the joint feasible solution space from a network-centric perspective, i.e., values that do not overload any type of last mile links. In parallel, the acceptable ranges of each of variables  $A_{1,1}$  and  $A_{2,2}$  are defined with respect to the user centric conditions  $\hat{M}_{1,1}$  and  $\hat{M}_{2,2}$ , respectively. The solution space is further refined according to the user-centric constraints. The association policy that yields the highest throughput is optimal.

The exhaustive approach is demonstrated next, where we adopt the default values in Table I with  $p_{u,1} = 0.8$ . In addition, we assume an interference-limited radio access scenario, thus, the expressions derived in Section III-C1 may be used. Figure 1 shows the solution space determined from a network-centric perspective. The x-axis shows the variation of  $A_{1,2}$  from 0 – 60% while the y-axis shows (a) the cumulative radio access load of the cells (b) the unused percentage of the VDSL2 BH link (for instance, a 0% value on the y-axes for VDSL2 indicates that there is no unused capacity and the BH is fully loaded). As expected, the maximum cumulative throughput is achieved when the users are equally distributed among the SCs, from a radio access dimensioning point of view. Nonetheless, the VDSL2 BH constraint limits the feasible solution space to  $A_{1,1} < 40\%$ . For  $A_{1,1} = 40\%$  and  $A_{1,2} = 0\%$ , the VDSL2 BH is fully loaded with 40% of users of type  $U_1$  associated to these SC and all users of type  $U_2$  served by SC with V-band last mile. Thus, the highest feasible throughput setting, from a network perspective, is  $A_{1,1} = 30\%$  and  $A_{1,2} = 40\%$  which also leads to full load on VDSL2 links.

The user-centric constraints are examined in Figure 2, as-

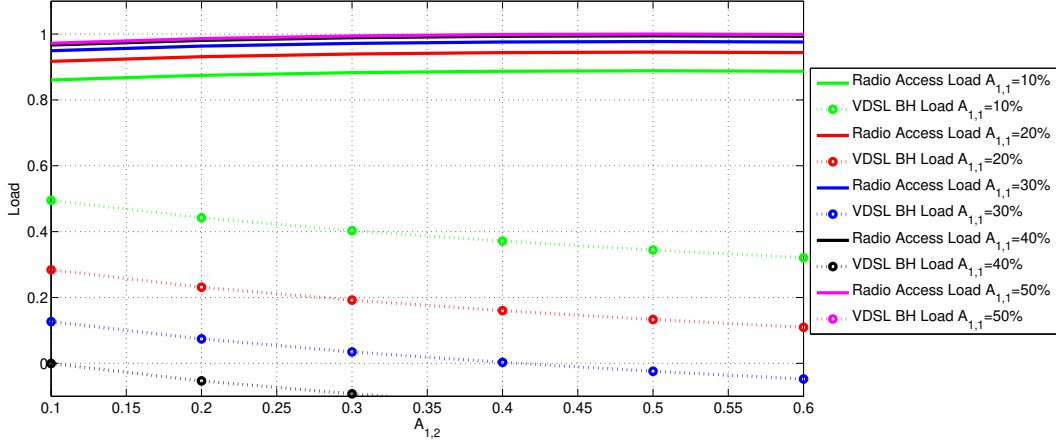


Fig. 1. Network-centric performance for different association policies. The figure shows the radio access load for various combinations of association policies in continuous lines with the highest value for equal partition, as expected. The dotted lines show the rate of unused capacity of the VDSL links; negative values indicate overload of the BH link.

suming a common threshold  $\theta_1 = \theta_2 = 15\%$ . The throughput-related  $\hat{M}_{u,1}$  is valid between values of  $30\% \leq A_{1,1} \leq 70\%$ , whereas the resilience-related  $\hat{M}_{u,2}$  is valid for values of  $A_{2,1} \geq 30\%$ . It can be verified that the highest-throughput BH-aware solution identified in Figure 1 complies with these conditions, and is hence selected as the optimum association policy resulting in  $\mathbf{A} = \{A_1 = 32\%, A_2 = 68\%\}$  (using (10)) and an air interface load  $\hat{T} = 97.59\%$ . The radio access load balancing user association policy would have resulted in equal load to all SCs. Looking at Figure 1, 50% load on each type of SC would entail 30% overload on VDSL2 based SCs. Consequently, the congested VDSL2 links would cause quality degradation to 50% of the users in both throughput and latency aspects.

TABLE I  
PARAMETERS EMPLOYED TO OBTAIN EXPERIMENTAL RESULTS.

Parameter	Range	Default
$\lambda_s$ (SCs/km <sup>2</sup> )	40	40
$p_{s,1}$	0.1 to 0.9	0.5
$p_{u,1}$	0.2 to 0.8	0.5
$\lambda_u$ (users/km <sup>2</sup> )	2000	2000
$G_1 = G_2$	1.3	1.3
MIMO rank	2	2
$Q_{1,1}$ (Mbps)	50 to 150	80
$Q_{2,1}$ (Mbps)	1000	1000
$Q_{1,2}$ (%)	100	100
$Q_{2,2}$ (%)	70 to 100	80
$M_1$ (Mbps)%	3.5	3.5
$M_2$ (%)	99	99
$\theta_1 = \theta_2$ (%)	15	15

### B. Q-learning implementation of UCB

The experiment conducted in the previous section is valuable as it demonstrates the potential of UCB is ameliorating the network and user-centric performance metrics and the respective margins that result from pertinent associations. However, from an implementation perspective, exhaustive search

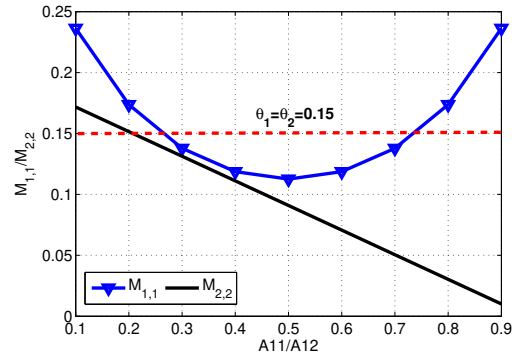


Fig. 2. User-centric performance for different association policies.

is not the best approach for two main reasons. Firstly, the computational power needed to find the optimum association policy becomes rapidly prohibitive in a realistic scenario with multiple BH types and quality attributes. Second, it is desirable to compute the optimum association policy in a distributed as opposed to centralised fashion to guarantee practical response times, agility, and scalability of the scheme. To this end, authors in [8] propose a UCB Q-learning implementation (a reinforcement learning technique) and demonstrate the gains in a two-attributes scenario: Throughput and latency. The results in [8] corroborate the analytical results derived in this paper as the improvements in user experience are 9.73% for throughput and 24.34% for latency. Nonetheless, it is difficult to compare quantitatively the gains resulting from Q-learning to the analytical output in the previous section. This is due to the fact that the Q-learning approach allows for individual cell bias setting as opposed to one bias per group type in the analytical modelling. The increased degrees of freedom unlock further the capacity of the network and provide the required agility to address the users' diversity.

Furthermore, the method is extended to three quality attributes in [13] in a heterogeneous network and it is shown to deliver significant improvement in users' satisfaction when compared to the state-of-the-art association schemes. Similar to [8], the network centric performance may suffer some degradation (3.3% recorded in [13]) while the user-centric performance is significantly and systematically improved. Authors in [14] use Fuzzy-Q-learning in their implementation of the UCB and show superior gains compared to the basic Q-learning approach at the cost of additional complexity. A memory-based implementation of the UCB that takes advantage of historical knowledge of previously identified optimum policies is presented in [15] and is shown to deliver better performance in terms of both convergence and gains. In both [14] and [15], the authors select two attributes: throughput and resilience. The memory-based scheme is the best implementation scheme as it outperforms the Fuzzy-Q-learning by 51% and 82%, with respect to throughput and resilience and it outperforms the basic Q-learning by 63% and 87%, respectively.

Although the merits of the UCB have been repeatedly demonstrated through network simulations in the cited papers, this work is the first to enable the generalisation of these results under variable conditions. On the other hand, the analytical solution allows fast and accurate analysis of the impact of different factors, whereas Q-learning-based simulations are onerous as they require numerous iterations and scenarios to yield representative results.

## V. CONCLUSION

In this work, we have presented the first complete analytical solution to the User-Centric-Backhaul (UCB) optimisation problem. In a mono-layer network with two possible types of BH technologies, we derive the throughput expressions as a function of a variable association policy. An exhaustive search approach is followed to find the optimum association that would yield the maximum throughput without overloading the constrained BH links nor breaching the users' quality of experience. The presented work is crucial as it demonstrates the merits of the UCB in accommodating different user types in a network with constrained BH links, whenever a feasible solution exists. From a different perspective, the UCB is also key in identifying the limiting constraint(s) in the network and in allowing for a pertinent improvement.

This paper is the first to propose a solution to the BH-aware and user-centric user-BS association problem. It is employed to demonstrate the unique ability of the UCB in capitalising on the diversity of 5G user-types in order to optimise the usage of the constrained realistic BH network.

## REFERENCES

- [1] P. Rost, C. Bernardos, A. Domenico, M. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wubben, "Cloud technologies for flexible 5G radio access networks," *IEEE Communications Magazine*, vol. 52, pp. 68–76, May 2014.
- [2] S. H. Park, O. Simeone, and S. Shamai, "Joint optimization of Cloud and Edge processing for Fog radio access networks." [Online], Available: <http://arxiv.org/abs/1601.02460>, Jan 2016.
- [3] X. Costa-Perez, A. Garcia-Saavedra, X. Li, T. Deiss, A. de la Oliva, A. di Giglio, P. Iovanna, and A. Moored, "5G-Crosshaul: An SDN/NFV integrated fronthaul/backhaul transport network architecture," *IEEE Wireless Communications*, vol. 24, pp. 38–45, Feb. 2017.
- [4] H.-S. Jo, Y. J. Sang, P. Xia, and J. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Transactions on Wireless Communications*, pp. 3484–3495, October 2012.
- [5] G. Zhang, T. Quek, M. Kountouris, A. Huang, and H. Shan, "Fundamentals of heterogeneous backhaul design - analysis and optimization," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 876 – 889, 2016.
- [6] G. Zhang, T. Quek, A. Huang, M. Kountouris, and H. Shan, "Backhaul-aware base station association in two-tier heterogeneous cellular networks," in *IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 390–394, Jun. 2015.
- [7] P. Marshall, "5G Operator survey," tech. rep., The Telecommunications Industry Association (TIA), Jan. 2017.
- [8] M. Jaber, M. Imran, R. Tafazolli, and A. Tukmanov, "A distributed SON-based user-centric backhaul provisioning scheme," *IEEE Access*, vol. 4, pp. 2314 – 2330, May 2016.
- [9] H. Zhang, Y. Chen, Z. Yang, and X. Zhang, "Flexible coverage for backhaul-limited ultra-dense heterogeneous networks: Throughput analysis and  $\epsilon$ -optimal biasing," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2018.
- [10] M. Jaber, F. J. Lopez-Martinez, M. A. Imran, A. Sutton, A. Tukmanov, and R. Tafazolli, "Wireless backhaul: Performance modelling and impact on user association for 5g," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2018.
- [11] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. Andrews, "User association for load balancing in heterogeneous cellular networks," *Wireless Communications, IEEE Transactions on*, vol. 12, pp. 2706–2716, June 2013.
- [12] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, pp. 3122–3134, November 2011.
- [13] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "A multiple attribute user-centric backhaul provisioning scheme using distributed SON," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec 2016.
- [14] F. Pervez, M. Jaber, J. Qadir, S. Younis, and M. A. Imran, "Fuzzy Q-learning-based user-centric backhaul-aware user cell association scheme," in *13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, (Valencia, Spain), Jun. 2017.
- [15] F. Pervez, M. Jaber, J. Qadir, S. Younis, and M. A. Imran, "Memory-based user-centric backhaul-aware user cell association scheme," *IEEE Access*, vol. 6, pp. 39595–39605, 2018.