# JOINT USER ASSOCIATION AND CONTENT PLACEMENT FOR CACHE-ENABLED WIRELESS ACCESS NETWORKS

*Binbin Dai and Wei Yu*

Department of Electrical and Computer Engineering
University of Toronto, Toronto, ON, Canada M5S 3G4
Emails: {bdai, weiyu}@comm.utoronto.ca

## ABSTRACT

This paper considers the optimal placement of content in cache-enabled base-stations (BSs) for reducing backhaul traffic in a densely deployed wireless access network. By caching popular files, users requesting these files can be served directly by their associated BSs without needing to fetch content from the core network. This paper makes an observation that a real network consists of distinct classes of users with different file preferences, so jointly optimizing cache placement and user-BS association can result in significant benefit. This paper considers such a joint optimization problem for achieving an optimized tradeoff between load balancing and backhaul saving, while accounting for both the physical layer wireless propagation characteristics and the finite cache size at the BSs. By proposing a numerical algorithm that iteratively optimizes the content placement policy for fixed user-association and optimizes the user association policy for fixed content placement, with a goal of maximizing a backhaul-aware proportional fairness network utility, this paper shows that placing similar content at strategically located BSs can result in significant backhaul saving without sacrificing as much in user access rates.

## 1. INTRODUCTION

Wireless access networks are increasingly deployed with dense set of base-stations (BSs) in order to support the ever-increasing demand for user data. As a result, backhaul congestion between the BSs and the core network often becomes the limiting factor, especially at peak hour and especially in networks with wireless backhaul. This paper examines the potential for reducing the backhaul traffic between the BSs and the core network by equipping the BSs with low-cost cache storage. Local storages at the BSs allow popular files to be cached for serving users without needing the backhaul.

In conventional content distribution networks, simply the most popular files are cached at the network edge. This paper goes one step further by making the following two key observations. First, there are distinct classes of users with very different content preferences. Second, a densely deployed wireless network affords an additional degree of freedom of allowing users the flexibility of associating with the BS depending on both channel condition and content availability. Thus, caching and user-BS association can be optimized jointly.

As a concrete example, the class of users desiring “sports” content may be quite distinct from the class of users desiring “music” content. By dedicating some BSs for caching “sports” videos and some other BSs for caching “music” videos, and by allowing the users to associate with the BSs of their choice, significant backhaul saving can be obtained.

The user-BS association problem has been considered extensively in the literature [1, 2]. The key issue here is load balancing, i.e., the BS association policy must account for both the wireless channel and interference characteristics, as well as the number of users already associated with each BS. A popular BS association policy is to offset the received signal-to-interference-and-noise ratio (SINR) by a bias term, so that users may be incentivized toward less crowded BSs even when the SINR from that BS is not the largest.

This paper brings the additional consideration of caching into BS association. Toward this end, [3] considers the downloading time as the objective function and uses a matching algorithm to optimize user association for given caching policy, while [4] assumes dynamic user association so that a user can connect with different BSs depending on which BS has its requested file, and designs the optimal file placement at each BS also to minimize the overall file downloading time.

In contrast to [3, 4], this paper considers the case that user association is static and that each user is only connected with one BS, but proposes the joint optimization of content placement and user association. We formulate a backhaul-aware network utility maximization problem, and use the pricing approach of [1, 2] to arrive at an optimized tradeoff between network utility and backhaul reduction.

## 2. PROBLEM FORMULATION

Consider a downlink wireless network with $L$ BSs serving $K$ users. Both the BSs and the users are equipped with a single antenna each. Each BS transmits at a fixed power spectrum density (PSD) $p_l$, $l \in \mathcal{L} = \{1, 2, \cdots, L\}$ over a total bandwidth of $W$. The channel from BS $l$ to user $k$ is assumed to be flat-fading with channel gain $h_{lk} \in \mathbb{C}$. We assume maximum frequency reuse factor of 1. It has been shown in [1] that at each BS, equal allocation of the available bandwidth among its associated users is optimal in terms of maximizing the log-utility of user data rates. In this case, if user $k$ is associated with BS $l$, then the data rate of user $k$ can be written as

$$R_{kl} = \frac{W}{K_l} \log\left(1 + \text{SINR}_{kl}\right), k \in \mathcal{K} = \{1, 2, \cdots, K\}, \quad (1)$$

where $K_l$ is the number of users associated with BS $l$, and

$$\text{SINR}_{kl} = \frac{|h_{lk}|^2 p_l}{\sum_{m \neq l} |h_{mk}|^2 p_m + \sigma_k^2}, \quad (2)$$

is the SINR of user $k$ if it is associated with BS $l$. Here, $\sigma_k^2$ is the background noise PSD. The proportional fair log-utility of the network can be written as

$$\sum_{kl} x_{kl} \log\left(R_{kl}\right), \quad (3)$$

where $x_{kl}$ is a binary variable defined as

$$x_{kl} = \begin{cases} 1, & \text{if user } k \text{ is associated with BS } l \\ 0, & \text{otherwise} \end{cases}. \quad (4)$$

The user-BS association problem as considered in [1, 2] is that of determining $\{x_{kl}\}$ to maximize (3).

We now describe the caching model. Suppose that there are a total of $F$ files, each of the same size $z$ for simplicity. In a wireless network without caching capabilities, each file request leads to a backhaul consumption of $z$ for the serving BS. However, if the request file has already been cached locally at the BS, then the BS can directly transmit the file to the user without costing backhaul. This paper considers a network where each BS $l$ can cache up to $C_l$ files, and poses the question of optimal content placement in the cache.

Let $y_{lf}$ be a binary variable indicating whether or not BS $l$ caches file $f \in \mathcal{F} = \{1, 2, \cdots, F\}$, i.e.,

$$y_{lf} = \begin{cases} 1, & \text{if file } f \text{ is cached in BS } l \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

then the cache size constraint at BS $l$ can be formulated as

$$\sum_f y_{lf} \leq C_l, \quad l \in \mathcal{L}. \quad (6)$$

In this paper, we also assume that different users may have different file preferences and denote the probability that user $k$ requests file $f$ as $P_{kf}$. It is not difficult to see that the total backhaul reduction of the network by associating users with the BS that caches the requested content is

$$\sum_{kl} x_{kl} \left( \sum_f P_{kf} y_{lf} \right). \quad (7)$$

There is a tradeoff between maximizing the network utility and maximizing backhaul saving. To maximize network utility, each user should associate with a BS of reasonably high SINR (while accounting for BS load), but such a BS may not have in its cache the content that the user desires. To maximize backhaul saving, the user should associate with the BS with the largest number of its required content, but such a BS may be far away.

This paper considers a joint optimization of the user-BS association policy $\{x_{kl}\}$ and content placement policy $\{y_{lf}\}$ at each BS that maximizes a weighted objective of network utility (3) and backhaul saving (7) as follows:

$$\begin{aligned} \underset{\{x_{kl}, y_{lf}, K_l\}}{\text{maximize}} \quad & \sum_{kl} x_{kl} \log\left(R_{kl}\right) + \lambda \sum_{klf} x_{kl} P_{kf} y_{lf} && (8) \\ \text{subject to} \quad & \sum_l x_{kl} = 1, \quad \sum_k x_{kl} = K_l, && (9) \\ & \sum_f y_{lf} \leq C_l, && (10) \\ & x_{kl} \in \{0, 1\}, \quad \forall k, l, && (11) \\ & y_{lf} \in \{0, 1\}, \quad \forall l, f. && (12) \end{aligned}$$

where (9) is the user-BS association constraint that ensures each user is only associated with one BS, and each BS is associated with $K_l$ users; (10) is the cache size constraint at each BS. By varying $\lambda$, we arrive at different tradeoff points between network utility and backhaul saving.

## 3. PROPOSED ALGORITHM

The user association and content placement problem is an integer programming problem, so finding its globally optimal solution is expected to be difficult. By taking advantage of the fact that the user-BS association constraint (9) and the caching constraint (10) are separable, this paper proposes an iterative algorithm that optimizes content placement at each BS assuming fixed user association and optimizes user association assuming fixed caching policy, as described in the following.

### 3.1. Optimal Content Placement with Fixed Association

With fixed user association $\{x_{kl}\}$, the caching problem is decoupled among the BSs. In particular, the content placement problem for BS $l$ can be written as

$$\begin{aligned} \underset{\{y_{lf}\}}{\text{maximize}} \quad & \sum_f \mu_{lf} y_{lf} && (13) \\ \text{subject to} \quad & \sum_f y_{lf} \leq C_l, \quad y_{lf} \in \{0, 1\} \;\; \forall f && (14) \end{aligned}$$

where $\mu_{lf} = \sum_k x_{kl} P_{kf}$ is the amount of backhaul reduction if BS $l$ caches file $f$. The above problem can be interpreted as follows: Given the benefit of caching file $f$ at BS $l$ as $\mu_{lf}$, which $C_l$ files should BS $l$ cache? The optimal solution to such a problem is simply to cache the $C_l$ files that have the largest benefits, i.e.,

$$y_{lf}^* = \begin{cases} 1, & \text{if } \mu_{lf} \in \{\mu_{l[1]}, \mu_{l[2]}, \cdots, \mu_{l[C_l]}\} \\ 0, & \text{otherwise} \end{cases}, \quad (15)$$

where $\mu_{l[j]}$ is the $j$'th largest item in the list of $\mu_{lf}$'s.

The above solution is obtained under the condition that each file is of the same size. In the case that files have different sizes, the cache placement problem for each BS becomes the well known *knapsack problem* [5], which can be solved exactly within pseudo-polynomial time $O(FC_l)$ using dynamic programming; more efficient approximation algorithms are also possible.

Note that the problem formulation in this paper assumes static user locations. In particular, we assume in (13) that user association is known at the content placement stage. In practical situations in which contents are placed during off-peak hours and accessed during peak hours, the optimization of content placement would require predicting user association.

### 3.2. Optimized User Association with Fixed Cache

With fixed content placement at each BS, the user association problem can be stated as

$$\underset{\{x_{kl}, K_l\}}{\text{maximize}} \quad \sum_{kl} (a_{kl} + \lambda b_{kl})\, x_{kl} - \sum_l K_l \log(K_l) \quad (16)$$

$$\text{subject to} \quad \sum_l x_{kl} = 1, \; x_{kl} \in \{0, 1\}, \quad \forall k, l \quad (17)$$

$$\sum_k x_{kl} = K_l, \quad \forall l, \quad (18)$$

where $a_{kl} = \log(W \log(1 + \text{SINR}_{kl}))$ is the log of spectrum efficiency and $b_{kl} = \sum_f P_{kf} y_{lf}$ is the amount of backhaul reduction if user $k$ is associated with BS $l$. Problem (16) is a mixed integer programming problem, for which obtaining the global optimum solution is challenging. This paper adopts a similar idea used in [1, 2] to solve a partially dualized version of the problem with respect to the constraint (18):

$$\max_{\{x_{kl}, K_l\}} \sum_{kl} (a_{kl} + \lambda b_{kl} - \eta_l)\, x_{kl} - \sum_l K_l \log(K_l) + \eta_l K_l$$

$$\text{s.t.} \quad \sum_l x_{kl} = 1, \; x_{kl} \in \{0, 1\}, \quad \forall k, l \quad (19)$$

where $\eta_l$ is the dual variable associated with constraint (18).

The optimization problem (19) is now decoupled among the users. It has explicit analytical solution:

$$x_{kl}^* = \begin{cases} 1, & \text{if } l = \arg\max_{l'} (a_{kl'} + \lambda b_{kl'} - \eta_{l'}) \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

for fixed dual variable $\eta_l$. The optimal $\eta_l^*$ can be found through the subgradient method [1] or the coordinate descent method in the dual domain [2]. Note that it is possible to have multiple BSs achieving the same maximum value in (20). In such a case, a tie-breaking procedure is needed [2].

The solution (20) is quite intuitive: each BS sets a price $\eta_l$; each user associates with the BS that maximizes its "profit" minus the price, where "profit" is defined as $(a_{kl} + \lambda b_{kl})$. In the user-BS association problem without caching [1, 2], the log spectral efficiency term $a_{kl}$ is considered as the user's "profit". With the possibility of caching content at the BSs, we take into account the backhaul reduction $b_{kl}$ as an additional motivation for users to associate with a BS. At small tradeoff constant $\lambda$, the log-utility term $a_{kl}$ is the dominating "profit"; as $\lambda$ increases the backhaul reduction $b_{kl}$ becomes the main concern in user-BS association.

Finally, we mention that due to the nonconvex nature of problem (16), the optimal solution (20) of the dual problem (19) may not be the optimal solution to (16). However, as verified in [1, 2], this pricing-based approach often produces excellent solutions to the overall problem. We summarize the proposed iterative algorithm for joint user association and caching design in Algorithm 1.

**Algorithm 1** Joint User Association and Content Placement

**Initialization**: Set the initial user association $\{x_{kl}\}$ using the algorithm in [1, 2] without considering caching;

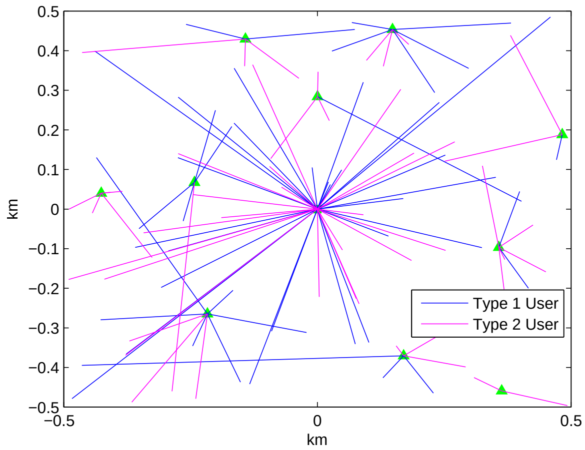
**Repeat**:

1. Fix user association $\{x_{kl}\}$, find the optimal content placement $\{y_{lf}\}$ according to (15);
2. Fix caching policy $\{y_{lf}\}$ at each BS, update the user association $\{x_{kl}\}$ by solving (19), i.e., according to (20) with optimal dual variable $\eta_l^*$.

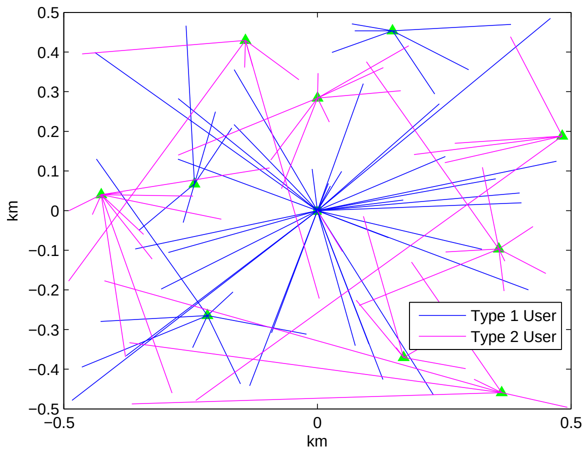
**Until** convergence

## 4. SIMULATION RESULTS

The performance of the proposed joint user association and caching placement strategy is evaluated for two topologies. First, consider a two-tier heterogeneous network with 1 macro-BS at the center of a cell and $L'$ pico-BSs randomly deployed in an 1 squared kilometer of geographic area. The transmit PSDs for the macro and pico BSs are set to be $-27$dBm/Hz and $-40$dBm/Hz respectively over a 10MHz bandwidth. The channels are modeled with distance-dependent path-loss $128.1 + 37.6 \log_{10}(d)$ dB and $140.7 + 36.7 \log_{10}(d)$ dB ($d$ in km) for the macro and pico BSs respectively with 8 dB log-normal shadowing. The PSD of the combined background noise and outer-cell interference is set at $-150$dBm/Hz. A total of 100 users are uniformly distributed in the area and a total of 1000 video files are assumed to be requested by the users. For simplicity, each BS is assumed to have the same cache size.




(a) User association without caching




(b) User association with caching

**Fig. 1**. User association without and with caching at the BSs

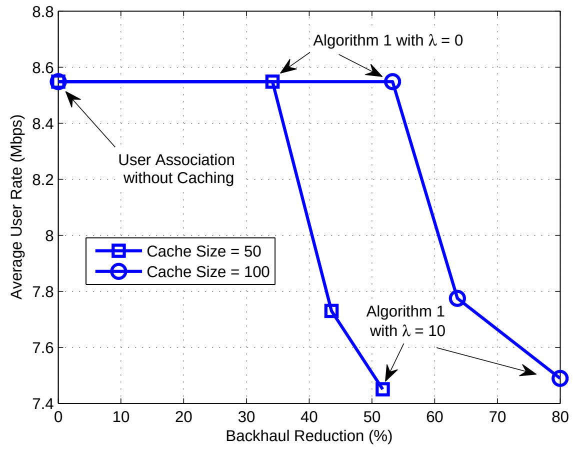



**Fig. 2**. Log-utility and backhaul reduction tradeoff

Fig. 1 shows the effect of cache-aware user association. Two types of files are considered: sports videos and music videos with 500 each. Both the sports and music videos are requested with the same Zipf distribution [6] with parameter 0.4. Two types of users with different file preferences are considered: Type 1 users prefer watching sports and request sports videos with probability 0.8 and music videos with probability 0.2; Type 2 users prefer music and request music videos with probability 0.8 and sports videos with probability 0.2. A total number of $L^{'} = 10$ pico-BSs are deployed in the network. In Fig. 1(a), the user association is done using the algorithm proposed in [1, 2] without caching at the BSs. In Fig. 1(b), each BS can cache up to 100 files and the user association is jointly optimized with cache placement using Algorithm 1 with the tradeoff constant $\lambda = 10$. As we can observe, the BSs in Fig. 1(a) are associated with similar number of Type 1 and Type 2 users, while BSs in Fig. 1(b) are almost always associated with one type of users or the other, but not both, in order to minimize backhaul consumption.

Fig. 2 plots the tradeoff curve between the average user access rate under log-utility and the total backhaul reduction for a second network topology with $L^{'} = 50$ pico-BSs and with different cache sizes at the BSs. We consider 10 different types of videos with 100 videos per type, which are requested with the same Zipf distribution with parameter 0.4. A total of 10 types of users are considered: each type of users prefers a different type of videos with probability 0.8 and all the rest of videos with probability 0.2 in total. Each point in the tradeoff curve is obtained by running Algorithm 1 with a particular tradeoff constant $\lambda$. The left end-point corresponds to the user association strategy of [1, 2] without caching, which achieves the maximum log-utility. If the BSs are cache-enabled, even without changing the user association (corresponding to $\lambda = 0$), the total backhaul consumption can already be reduced by $(34\%, 53\%)$ under the BS cache sizes of $(50, 100)$ files respectively. By increasing $\lambda$ and running the backhaul-aware user association iteratively with optimized content placement, the total backhaul consumption can be further reduced by $(52\%, 80\%)$ respectively, with only around $13\%$ sacrifice on average user access rate.

## 5. CONCLUSION

This paper considers the joint design of user association and content placement in cache-enabled wireless networks by formulating an optimal tradeoff problem between network utility and backhaul reduction. We propose an efficient algorithm iterating between cache-aware user association and association-aware content placement. Numerical results verify that a significant amount of backhaul reduction can be achieved by associating the BSs with the users of similar file preference.

## 6. REFERENCES


[1] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J.G. Andrews, “User association for load balancing in heterogeneous cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, 2013.

[2] K. Shen and W. Yu, “Distributed pricing-based user association for downlink heterogeneous cellular networks,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1100–1113, 2014.

[3] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, “Cache-aware user association in backhaul-constrained small cell networks,” in *12th Int. Symp. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Netw. (WiOpt)*, May 2014, pp. 37–42.

[4] K. Shanmugam, N. Golrezaei, A.G. Dimakis, A.F. Molisch, and G. Caire, “Femtocaching: Wireless content delivery through distributed caching helpers,” *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[5] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman & Co., New York, NY, USA, 1990.

[6] M. Zink, K. Suh, Y. Gu, and J. Kurose, “Characteristics of youtube network traffic at a campus network measurements, models, and implications,” *Comput. Netw.*, vol. 53, no. 4, pp. 501 – 514, 2009.