Joint Optimization of Frequency Allocation and User Association with Differentiated Service in Hyper-cellular Networks

Shan Zhang, Sheng Zhou, Zhisheng Niu
Tsinghua National Laboratory for Information Science and Technology,
Dept. of Electronic Engineering, Tsinghua Univ., Beijing, 100084, P.R. China
Email: zhangshan11@mails.tsinghua.edu.cn, {sheng.zhou, niuzhs}@tsinghua.edu.cn

***Abstract*—The existing architecture of heterogeneous networks is not energy and spectrum efficient as many lightly loaded base stations (BS) can not be turned off for coverage guarantee. To further improve energy and spectrum efficiency, a new architecture called "hyper-cellular" network has been proposed in our previous work. Under this architecture, the function of different types of BSs may not be the same, and the mechanism of user association should consider many factors, such as user mobility, traffic load distribution, and differentiated service demands, which are usually ignored in the existing studies. In addition, the spectrum allocation strategy also has great influence on the network performance. As a starting point, we explore the user association mechanism based on the differentiated service demands of the network users, and jointly optimize it with spectrum allocation, in order to maximize the network capacity with quality of service constraints. Although closed-form expression of the optimal solution can not be derived, numerical results are obtained. Our approach is shown to improve the network capacity more than four times over the baseline strategy, where the conventional user association method is adopted and all BSs use all available spectrum.**

I. Introduction

Heterogeneous Networks (HetNet), where macro base stations (BS), micro BSs, and pico BSs coexist, is expected to be the dominant scenario in 4G era [1]. However, the present architectures of HetNet are not energy and spectrum efficient enough, as many BSs being lightly loaded but still consuming great amount of spectrum and energy [9]. The reason why those lightly loaded BSs can not go into sleep is the coverage requirement. To investigate efficient mechanism for high spectrum and energy efficiency from the view of the whole system, we have proposed a framework of hyper-cellular system, where the control signaling and data service provided by the BSs are separated to realize soft coverage, soft resource matching, and service on-demand [10]. With this system architecture, more flexible network operation can be easily conducted, like BS sleeping and load balancing. Nevertheless, effective user association mechanism is necessary to realize these potential advantages, where many factors should be taken into consideration, such as traffic load distribution, service requirements, user mobility, performance indexes and so on. For example, users with high mobility should be associated with macro BSs to avoid from frequent hand-overs. Furthermore, users can associate with lightly loaded BSs for load balancing, while they can also be associated with heavily loaded ones so that there is more chance to turn off lightly loaded BSs. Besides user association mechanism, the frequency allocation strategies also have great influence on the network performance, whose design is usually coupled with user association.

Actually, the problems about user association and frequency allocation have been intensely studied in the existing works [3]- [8]. Range Expansion (RE), an improved association method based on the conventional one, has been adopted in many studies. According to the conventional user association method, users are associated with the BS which provides highest received signal to interference ratio (SINR), which can not make full use of the BSs with lower transmit power, like pico BSs. To increase the efficiency of pico BSs, a bias is added to the received SINR from them, which enlarges their coverage area size. Joint optimization of resource allocation and user association has also been studied [4] [5]. However, RE is far from enough as it only consider the received SINR.

As a starting point, we explore the capacity maximization of hyper-cellular networks by jointly optimizing user association and frequency allocation, where the users' association is also based on their differentiated services. The main difference between our work and the existing similar studies [4] [5] is the consideration of differentiated service demands during user association, which is to cater to the division of labor between different BSs in hyper-cellular networks. In addition, this association method based on differentiated service is also suitable in current HetNets for the unique properties of different BSs. The capacity maximization problem is formulated as an optimization problem with quality of service(QoS) constraints, where the

This work is sponsored in part by the National Basic Research Program of China (2012CB316001), the Nature Science Foundation of China (61201191, 60925002, 61021001), and Hitachi Ltd.

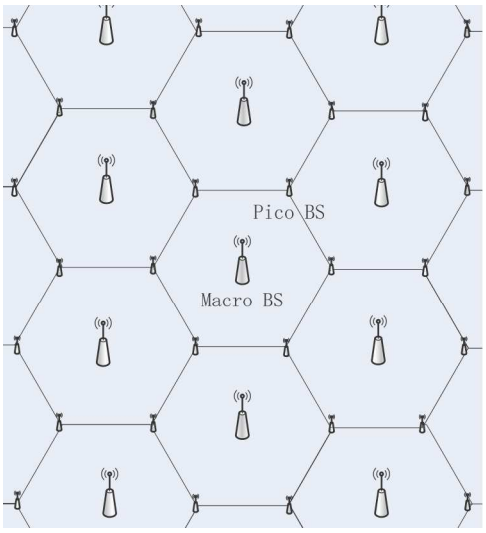


Fig. 1: A two tier HetNet topology

optimization variables are the ratio of frequency allocated to each tier and the association mechanism. Erlang approximation algorithm is applied for blocking probability analysis. Specially, it is found that the network capacity can be maximized if and only if load balancing is achieved, which helps to find the optimal solution. Numerical results show that the network capacity can be increased by more than four times compared with the baseline method, where all BSs use all available spectrum and users are associated according to the conventional mechanism.

The rest of the paper is organized as follows. Section II describes the system model adopted in this paper, which includes deployment scenario, link layer model and admission control. Problem formulation and solution are introduced in Section III. After that, numerical calculation is conducted in Section IV, where results and analysis are also presented. Finally, some conclusions are drawn in Section V.

II. System Model

A. Deployment Scenario

A two-tier HetNet (shown in Fig.1) is adopted in this paper, where users are assumed to be uniformly distributed. Macro tier is modeled as hexagonal network architecture, which is the most widely used model for traditional cellular networks. Pico BSs are deployed at the corner of every hexagon to improve user experience. Therefore, this network topology is the ideal case under our assumption when the density of the pico BSs is two times of the macro BSs. Nevertheless, our work can be easily extended to other network topologies where the position of the BSs are fixed. Orthogonal frequency allocation between macro and pico BSs is assumed for inter-tier interference avoidance, where the ratio is to be optimized. In addition, other frequency allocation schemes, like co-channel sharing [3], can also be analyzed by our model.

Services with constant bandwidth demands are considered in this paper, which include applications like voice calls, video conference, realtime TV and so on. For simplicity, these services are classified into two classes based on their data rate demands. The minimal data rate required by "*class-L*" users, r_{L}, is assumed to be relatively low, which mainly includes voice calls. "*Class-H*" services include other applications with high data rate demands r_{H}. Poisson process is used to model the random arrival of the users, where $(\lambda_{\mathrm{L}}, \lambda_{\mathrm{H}})$ denotes the arrival rate. Once associated with the BSs, the users will sojourn in the system for a random period of time, and the sojourn time of the two classes is assumed to be exponentially distributed with mean value of $(1/\mu_{\mathrm{L}}, 1/\mu_{\mathrm{H}})$. As the application of smart phones and 3G/LTE networks, the volume of class-H services is rapidly increasing and will still keep growing in the future, while the volume of voice calls is relatively steady. Therefore, we set λ_{L} as a parameter for simplicity, while λ_{H} is the target function which is to be maximized.

With the macro BSs are responsible for signaling and network coverage in hyper-cellular networks, the pico BSs only serve users with high data rate demands. In addition, this association method is also suitable for current HetNets for user expectations. For example, users at high moving speed (who are more suitable to be associated with macro BS) may bear a relatively low data rate service, while users associated with the pico BSs may expect higher data rate. Therefore, class-L users will be associated with their nearest macro BSs, and the service area partition for class-H users is shown in Fig.2. Class-H users in the pink area are still served by macro BSs, while other class-H users are associated with their nearest pico BSs. For example, the shadow area in Fig.2 shows the service area of the pico BS within that region. With this association model, class-L users are allowed higher mobility, while class-H users may enjoy better channel condition. One thing to mention is that the radius R_{M} is to be optimized, which corresponds to the association mechanism.

B. Link Model

Assume that the transmit power is denoted as P_{m} for macro BSs and P_{p} for pico BSs, and the bandwidth allocated to each user can be adjusted to satisfy its data rate demand. Then, the received SINR of user$u$ which is associated with BS$_i$ is given by

$$\gamma_u = \begin{cases} \dfrac{P_{\mathrm{m}} G_{iu}}{\sum\limits_{j=1, j\neq i}^{N_{\mathrm{m}}} P_{\mathrm{m}} G_{ju} + \sigma^2}, & \text{BS}_i \text{ is a macro BS} \\ \dfrac{P_{\mathrm{p}} G_{iu}}{\sum\limits_{j=1, j\neq i}^{N_{\mathrm{p}}} P_{\mathrm{p}} G_{ju} + \sigma^2}, & \text{BS}_i \text{ is a pico BS} \end{cases} \tag{1}$$

where G_{iu} is the channel gain between user$u$ and BS$_i$, σ^2 is the noise power, N_{m} denotes the number of macro BSs, and N_{p} denotes the number of pico BSs. Channel fading is not considered here as we focus on the average data rate of the users during a relatively long term. When adaptive

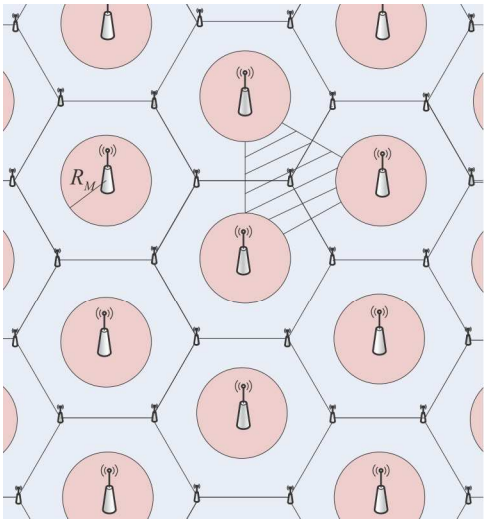


Fig. 2: Service area for high data rate service

modulation and coding are used, the spectrum efficiency function of received SINR γ_u can be approximated as [11]:

$$C(\gamma_u) = \log_2(1 + \beta\gamma_u), \tag{2}$$

where $\beta = -1.5\ln(5\varepsilon)$ is a constant related to the bit error rate requirement ε. Thus, the bandwidth demand of user$u$ is given by

$$W_u = \frac{r_u}{C(\gamma_u)}, \tag{3}$$

where r_u is the minimal data rate demand.

C. Admission Control

Admission control is needed to guarantee the performance of the admitted users. When user$u$ arrive at the cell of BS$_i$, it will be admitted if and only if following inequality holds:

$$\phi_u + \sum_{v \in \mathcal{A}_i} \phi_v \leq 1, \tag{4}$$

where $v \in \mathcal{A}_i$ means user$v$ is associated with BS$_i$, and ϕ_u is the normalized bandwidth demand of user$u$, which is given by

$$\phi_u = \frac{W_u}{W_i} = \frac{r_u}{C(\gamma_{iu})W_i}, \tag{5}$$

where W_i is the total bandwidth allocated to BS$_i$. If (4) is not satisfied, blocking happens. The blocking probability of BS$_i$ is given by

$$\mathbf{P} = \Pr\left\{\phi_u + \sum_{v \in \mathcal{A}_i} \phi_v > 1\right\}, \tag{6}$$

which should be guaranteed smaller than certain thresholds for QoS requirements.

III. Problem Formulation and Solution

A. Problem Formulation

Then our problem can be formulated as follows

$$\begin{aligned} \max_{\theta, R_{\mathrm{M}}} \quad & \lambda_{\mathrm{H}} \\ \text{s.t.} \quad & \mathbf{P}_{\mathrm{mL}}(\lambda_{\mathrm{H}}) \leq \eta_{\mathrm{L}} \\ & \mathbf{P}_{\mathrm{mH}}(\lambda_{\mathrm{H}}) \leq \eta_{\mathrm{H}} \\ & \mathbf{P}_{\mathrm{pH}}(\lambda_{\mathrm{H}}) \leq \eta_{\mathrm{H}}, \end{aligned} \tag{7}$$

where θ is the ratio of frequency allocated to macro BSs, η_{L} and η_{H} are the thresholds of blocking probability of class-L and class-H users respectively. $\mathbf{P}_{\mathrm{mL}}(\lambda_{\mathrm{H}})$ is the blocking probability of class-L users of the macro BS, which is given by (6). Similarly, $\mathbf{P}_{\mathrm{mH}}(\lambda_{\mathrm{H}})$ and $\mathbf{P}_{\mathrm{pH}}(\lambda_{\mathrm{H}})$ are the blocking probabilities of class-H users of the macro and pico BSs respectively.

B. Problem Simplification

The subproblem of (7) is to find the maximal λ_{H} under call blocking constraints when given θ and R_{M}, which is actually to find the maximal traffic arrival rate of a queueing network with multiple queues when given other network parameters. As the service process of each BS is independent, the analysis of the queuing network can be simplified as independent analysis of several queues. Therefore, the problem (7) is further decomposed into subproblems as (8) and (9):

$$\begin{aligned} \max \quad & \lambda_{\mathrm{mH}} \\ \text{s.t.} \quad & \mathbf{P}_{\mathrm{mL}}(\lambda_{\mathrm{mH}}) \leq \eta_{\mathrm{L}} \\ & \mathbf{P}_{\mathrm{mH}}(\lambda_{\mathrm{mH}}) \leq \eta_{\mathrm{H}}, \end{aligned} \tag{8}$$

$$\begin{aligned} \max \quad & \lambda_{\mathrm{pH}} \\ \text{s.t.} \quad & \mathbf{P}_{\mathrm{pH}}(\lambda_{\mathrm{pH}}) \leq \eta_{\mathrm{H}}, \end{aligned} \tag{9}$$

where parameters of the network are assumed to be given. Problem (8) describes the queue of the macro BSs, while (9) describes the queue of the pico BSs. Assume the optimal values of (8) and (9) are respectively λ^*_{mH} and λ^*_{pH}, then network capacity is given by

$$\lambda_{\mathrm{H}} = \min(\lambda^*_{\mathrm{mH}}, \lambda^*_{\mathrm{pH}}). \tag{10}$$

Specially, the optimal value of problem (7) can be achieved by $(\theta^*, R^*_{\mathrm{M}})$, if and only if the optimal values of problem (8) and (9) are equal under $(\theta^*, R^*_{\mathrm{M}})$. Denote the optimal value of problem (8) and (9) as λ_{mH} and λ_{pH} under (θ, R_{M}), and assume $\lambda_{\mathrm{mH}} < \lambda_{\mathrm{pH}}$ without losing generality. Then the network capacity is $\lambda_{\mathrm{H}} = \lambda_{\mathrm{mH}}$. Let $\theta' = \theta + \delta_\theta$, where $\delta_\theta \to 0, \delta_\theta > 0$. Then the new optimal values of subproblem (8) and (9) are respectively $\lambda'_{\mathrm{mH}} = \lambda'_{\mathrm{mH}} + \delta_{\lambda_{\mathrm{m}}}$ and $\lambda'_{\mathrm{pH}} = \lambda'_{\mathrm{pH}} - \delta_{\lambda_{\mathrm{p}}}$, where $\delta_{\lambda_{\mathrm{m}}} \to 0$, $\delta_{\lambda_{\mathrm{m}}} > 0$ and $\delta_{\lambda_{\mathrm{p}}} \to 0$, $\delta_{\lambda_{\mathrm{p}}} > 0$. Therefore, $\lambda'_{\mathrm{H}} = \min(\lambda'_{\mathrm{mH}}, \lambda'_{\mathrm{pH}}) = \lambda'_{\mathrm{mH}} > \lambda_{mH}$, which means (θ, R_{M}) can not be the optimal solution of problem (7). Note that the physical meaning of this conclusion is that the capacity of the

network can be maximized if of only if the traffic load of different BSs is balanced.

Then, problem (7) can be solved by following steps:

1) For any given frequency allocation ratio θ, find $R_{\mathrm{M}}^{*}(\theta)$ which makes the optimal values of problem (8) and (9) equal.
2) Compare the network capacity under $(\theta, R_{\mathrm{M}}^{*}(\theta))$, and find the optimal θ^{*} under which problem (7) can be maximized.

As the optimal values of problem (8) and (9) are continuous and monotonous functions of R_{M}, step 1) can be conducted by searching methods, like dichotomy.

C. Network Performance Analysis

Actually, accurate calculation of blocking probability given by equation (6) is impossible considering the complexity. The information of the bandwidth demand of every user is needed as it varies with user position. In addition, the dimension of system states is also uncertain as the number of users in the networks is random. Therefore, we apply an approximation method which is called "Erlang approximation" to analyze the network performance. The main idea is to approximate the random bandwidth demand by their mean values, then only the number of users of each service class is needed to describe the system state. Therefore, the complexity can be greatly reduced. By Erlang approximation, the service process of each BS can be modeled as an Erlang system, where the blocking probability can be easily derived by queuing theory [12]- [2]. Due to the symmetry of the network topology, we can simply focus on one macro cell and one pico BSs for the analysis of call blocking probability.

As the pico BS only offers class-H service, its service process can be modeled as an *M/M/s(0)* queuing system with Erlang approximation. And the number of servers s equals to $\lfloor 1/\bar{\phi}_{\mathrm{pH}} \rfloor$, where $\bar{\phi}_{\mathrm{pH}}$ is the average normalized bandwidth demand of class-H users associated with the pico BS:

$$\bar{\phi}_{\mathrm{pH}} = \frac{r_{\mathrm{H}}}{(1-\theta)W|\mathcal{A}_{\mathrm{p}}|} \int_{\mathcal{A}_{\mathrm{p}}} \frac{1}{C(\gamma(a))} \mathrm{d}a, \tag{11}$$

where $\mathcal{A}_{\mathrm{p}}$ is the coverage area of the pico BS, and $C(\gamma(a))$ is the corresponding spectrum efficiency at position a. According to the Erlang-B equation, the blocking probability of the pico BS is given by

$$\mathbf{P}_{\mathrm{pH}} \approx \frac{\rho_{\mathrm{pH}}^{s}}{s!} \Big(\sum_{j=0}^{s} \frac{\rho_{\mathrm{pH}}^{s}}{j!}\Big)^{-1}, \tag{12}$$

where ρ_{pH} is the traffic load of class-H for the pico BS:

$$\rho_{\mathrm{pH}} = \frac{\lambda_{\mathrm{H}}}{\mu_{\mathrm{H}}} |\mathcal{A}_{\mathrm{p}}|, \tag{13}$$

where the unit of μ_{H} is /s/m^2 and the unit of $|\mathcal{A}_{\mathrm{p}}|$ is m^2.

For the macro BS, the analysis of blocking probability is more complex. The average bandwidth demand is given by

$$\bar{\phi}_{\mathrm{mL}} = \frac{r_{\mathrm{L}}}{\theta W|\mathcal{A}_{\mathrm{mL}}|} \int_{\mathcal{A}_{\mathrm{mL}}} \frac{1}{C(\gamma(a))} \mathrm{d}a \tag{14}$$

and

$$\bar{\phi}_{\mathrm{mH}} = \frac{r_{\mathrm{H}}}{\theta W|\mathcal{A}_{\mathrm{mH}}|} \int_{\mathcal{A}_{\mathrm{mH}}} \frac{1}{C(\gamma(a))} \mathrm{d}a, \tag{15}$$

where $\mathcal{A}_{\mathrm{mH}}$ is the service area of the macro BS for class-H users, which is a round with radius being R_{M} , while $\mathcal{A}_{\mathrm{mL}}$ is the whole hexagonal cell. Then the access control condition in (4) can be approximated as:

$$n_{\mathrm{L}}\bar{\phi}_{\mathrm{mL}} + n_{\mathrm{H}}\bar{\phi}_{\mathrm{mH}} \leq 1, \tag{16}$$

where n_{L} is the number of class-L users associated with the macro BS and n_{H} is that of the class-H users. Denote the traffic load of each class as ρ_{mL} and ρ_{mH}, which can be similarly obtained by (13). According to queuing theory, the stationary probability distribution of state $\mathbf{n} = (n_{\mathrm{L}}, n_{\mathrm{H}})$ is given by:

$$\pi(\mathbf{n}) = \frac{\rho_{\mathrm{mL}}^{n_{\mathrm{L}}}}{n_{\mathrm{L}}!} \frac{\rho_{\mathrm{mH}}^{n_{\mathrm{H}}}}{n_{\mathrm{H}}!} \Big(\sum_{\mathbf{n}\in\mathcal{S}} \frac{\rho_{\mathrm{mL}}^{n_{\mathrm{L}}}}{n_{\mathrm{L}}!} \frac{\rho_{\mathrm{mH}}^{n_{\mathrm{H}}}}{n_{\mathrm{H}}!}\Big)^{-1}, \tag{17}$$

where $\mathcal{S}$ is the set of all feasible states which satisfies (16). The blocking probability of class-L users is given by:

$$\mathbf{P}_{\mathrm{mL}} \approx \sum_{(n_{\mathrm{L}}, n_{\mathrm{H}})\in\mathcal{S}_{\mathrm{L}}} \pi(n_{\mathrm{L}}, n_{\mathrm{H}}), \tag{18}$$

where $\mathcal{S}_{\mathrm{L}}$ denotes the set of states under which the new arrival class-L users will be blocked:

$$\mathcal{S}_{\mathrm{L}} = \left\{(n_{\mathrm{L}}, n_{\mathrm{H}}) | 1 - \bar{\phi}_{\mathrm{mL}} \leq n_{\mathrm{L}}\bar{\phi}_{\mathrm{mL}} + n_{\mathrm{H}}\bar{\phi}_{\mathrm{mH}} \leq 1\right\}. \tag{19}$$

For the class-H users, the blocking probability is similarly given by:

$$\mathbf{P}_{\mathrm{mH}} \approx \sum_{(n_{\mathrm{L}}, n_{\mathrm{H}})\in\mathcal{S}_{\mathrm{H}}} \pi(n_{\mathrm{L}}, n_{\mathrm{H}}), \tag{20}$$

where $\mathcal{S}_{\mathrm{H}}$ denotes the set of states under which the new arrival class-H users will be blocked by the macro BS:

$$\mathcal{S}_{\mathrm{H}} = \left\{(n_{\mathrm{L}}, n_{\mathrm{H}}) | 1 - \bar{\phi}_{\mathrm{mH}} \leq n_{\mathrm{L}}\bar{\phi}_{\mathrm{mL}} + n_{\mathrm{H}}\bar{\phi}_{\mathrm{mH}} \leq 1\right\}. \tag{21}$$

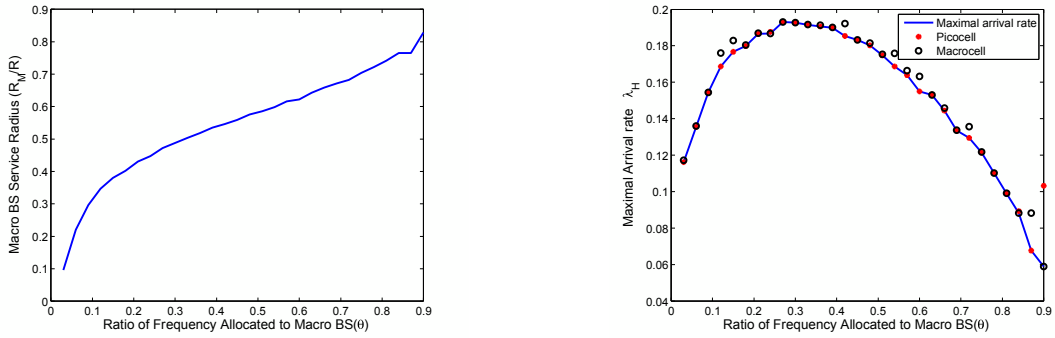
Although the explicit form of blocking probability can not be derived, numerical results can be calculated for any given network parameters. Generally, the computational complexity for the pico BS is much lower than the macro BS. For the macro BS, computational complexity is $O(N_{\mathrm{mL}}N_{\mathrm{mH}})$, where N_{mL} and N_{mH} denote the maximal number of users that can be accommodated by the macro BS for each class, i.e., $N_{\mathrm{mL}} = \lfloor 1/\bar{\phi}_{\mathrm{mL}} \rfloor$ and $N_{\mathrm{mH}} = \lfloor 1/\bar{\phi}_{\mathrm{mH}} \rfloor$.

IV. Numerical Results

A. Maximal Network Capacity

Following the two-step algorithm in SectionIII-B, we calculate the optimal solution of problem (7). First, we search for the optimal association strategy (corresponding to the service area radius of macro BS R_{M}) for any given frequency allocation (θ). Then, we obtain the optimal frequency allocation (θ^{*}) by comparing their network capacity $\lambda_{\mathrm{H}}^{*}(\theta)$.

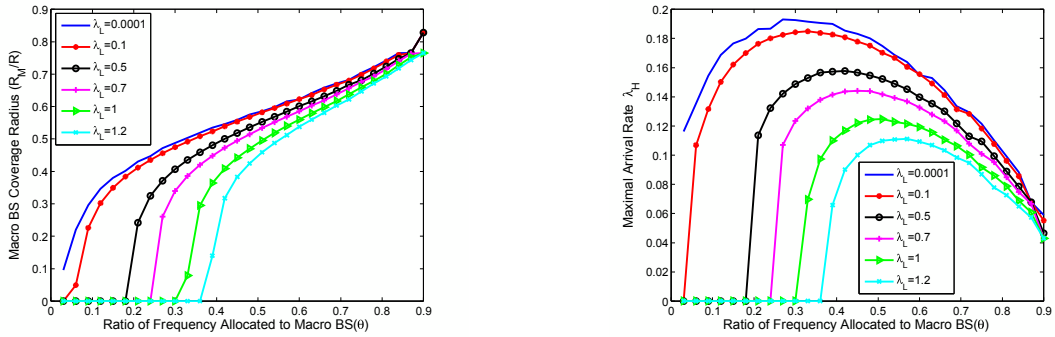
In the numerical calculation, the side length of the hexagonal macro cells R is 500m. The transmit power of the macro BS and the pico BS is respectively 10W and 0.5W,



(a) Optimal service radius for class-H users of macro BS

(b) Network capacity for class-H users

Fig. 3: Numerical results ($\lambda_L = 0.0001/s$)



(a) Optimal service radius for class-H users of macro BS

(b) Network capacity for class-H users

Fig. 4: Influence of λ_L

and the maximal received SINR is set to be 20dB. The channel gain only depends on the distance between the BS and the users: $G(s) = -130 - 35\log_{10} d$, where the unit of d is Km. The noise power is -104dBm, and the total bandwidth of the network is 10MHz. Spectrum efficiency is calculated by (2), where $\beta = 0.283$, which corresponds to the BER requirement of 10^{-3}. For the class-L users, the data rate demands are set to be 64kbps and their average sojourn time is 60s. For the class-H users, their data rate demands are 1Mbps while their average sojourn time is 300s. The thresholds of blocking probability of the two service classes are both 0.05.

Fig.3 shows the numerical results when the arrival rate of class-L users is set to be $\lambda_{\mathrm{L}} = 0.0001/\mathrm{s}$ per hexagon cell, which approximates the case with single service class. The optimal service area radius of macro tier R_{M} for the given frequency allocation is shown in Fig.3a, where the x-axis is the ratio of frequency allocated to macro tier and the y-axis is the normalized service radius R_{M}/R. As can be observed, the service area size of the macro tier increases with the resource allocated to it. Fig.3b shows the maximal arrival rate ($\lambda_H^*(\theta)$) for any the given resource allocation (θ). The blue solid line is the network capacity, the minimal value of the two dotted lines, which are the optimal solutions of (9) and (8) respectively. The chattering of the curves is caused by the errors of Erlang approximation, where the continuous random bandwidth demands are replaced by their mean value. But the accuracy is still acceptable shown as Fig.3b. As illustrated in Fig.3b, the network capacity can be maximized when the ratio of frequency allocated to the macro BSs is about 0.27. Then, the corresponding optimal normalized service radius of the macro BS is about 0.472, which can be found in Fig.3a. Therefore, the optimal solution of problem (7) is $(\theta^*, R_{\mathrm{M}}^*) \approx (0.27, 236m)$, and the corresponding network capacity is 0.193/s per hexagonal cell.

With optimal user association mechanisms for any given frequency allocation, the ratio of frequency allocated to the macro tiers indicates how many services are offered by the macro BSs. Obviously, these two kinds of BSs present their own advantages. Pico BSs have higher spatial spectrum reuse factor, while macro BSs have higher transmit power. As the network capacity depends on the spectrum efficiency, which increases with the spatial spectrum reuse factor and the received SINR at the end users, there exists a tradeoff between the service area of the two BSs. For example, larger service area of pico BSs means larger spatial spectrum reuse factor, but also means smaller received SINR at end users.

TABLE I: Network Capacity Comparison

λ_{L}	Baseline	Optimized
0.0001	0.0348	0.1930
0.01	0.0337	0.1843
0.5	0.0284	0.1577
0.7	0.0263	0.1442
1	0.0230	0.1248
1.2	0.0210	0.1111

This tradeoff relation explains the shape the curve in Fig.3b.

The optimal solutions and optimal values of problem (7) with different λ_{L} are compared in Fig.4. As can be observed, the network capacity is zero when θ is smaller than certain values (denoted as $\theta_{\min}(\lambda_{\mathrm{L}})$). The reason is that these solutions are not feasible. It can be explained as that the resource allocated to the macro BSs is so little that the QoS requirement of the class-L users can not be satisfied, even if all class-H users are associated with the pico BSs. In addition, this minimal ratio increases with the traffic load of the class-L users (for example, $\theta_{\min} = 0.03$ for $\lambda_{\mathrm{L}} = 0.1$ and $\theta_{\min} = 0.24$ for $\lambda_{\mathrm{L}} = 0.7$), which is quite apparent.

B. Efficiency Evaluation

To evaluate the efficiency of the our algorithm, we compared the network capacity of our optimal solution and the baseline strategy, where users are associated according to the conventional method and all BSs use all available spectrum. Under conventional method, users are associated with the BS which offers the highest received SINR. And the network capacity is shown in table I, where the units of all results are /s/macro-cell. As can be observed, the network capacity can be increased by more than 4 times with our optimal solution compared with that of the baseline strategy, which shows our algorithm is effective. Compared with the baseline strategy, our algorithm is beneficial for interference avoidance and load balancing, which helps to improve the network capacity.

V. Conclusion

In this paper, we have investigated the joint optimization of frequency allocation and user association in hyper-cellular networks, where the association strategy is based on users' service requirements besides their positions. In our system model, pico BSs only offer high data rate services. This association method is based on the division of labor between different tiers in the hyper-cellular networks, which is also suitable for current HetNets for the properties of different BSs and user expectations. An ideal deployment scenario has been studied. We formulated joint optimization of frequency allocation and user association as an optimization problem, whose target is to maximize the network capacity with call blocking probability constraints. Numerical results have been obtained by Erlang approximation methods. Specially, load balancing can be achieved with optimal solutions. The network capacity can be increased by more than four times with our algorithm compared with the baseline strategy, according to the numerical results. Other factors, like user mobility, different performance indexes, the nonuniformity of user distribution and so on, are left for our future work.

References

[1] A. Damnjanovic, et al., "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, issue 3, pp. 10-21, June 2011.

[2] Y. Wu, and Z. Niu, "Energy efficient base station deployment in green cellular neworks with traffic variations," in *IEEE ICCC'12*, Beijing, Chian, Aug. 2012.

[3] W. Lei, et al., "Heterogeneous network in LTE-Advanced system," in *IEEE ICCS'10*, Amsterdam, Netherlands, June, 2010.

[4] R. Madan, et al., "Cell association and interference coodination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, Dec. 2010.

[5] D. Fooladivanda, and C. Rosenberg, "Joint resource and user association for heterogeneous wireless cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, Jan. 2013.

[6] NTT DOCOMO, R1-103264, *Performance of eICIC with control channel coverage limitation*, 3GPP Std., Montreal Canada, May 2010.

[7] I. Guvenc, et al.,"Range expansion and inter-cell interference coodination (ICIC) for Picocell Networks", in *IEEE VTC'11*, San Francisco, US, Sep. 2011.

[8] A. Khandekar, et al., "LTE-advanced: Heterogeneous networks," in *IEEE European Wireless Conf.'10*, Lucca, Italy, Apr. 2010.

[9] Z. Niu, "TANGO: traffic-aware network planning and green operation", *IEEE Wireless Commun.*, vol. 18, pp. 22-29, Oct. 2011.

[10] Z. Niu, et al., "Energy efficiency and resource optimized hyper-cellular mobile comminication system architecture and its technical challenges," *Scientia Sinica(Informationis)*, vol. 42, no. 10, pp. 1191-1203, 2012.

[11] X. Qiu, and K. Chawla, On the performance of adaptive modulation in cellular systems *IEEE Trans. Commun.*, vol. 47. no. 6, pp. 884-895, Jun. 1999.

[12] T. Bonald, Wireless downlink data channels: user performance and cell dimensioning, in *ACM MobiCom'03*, San Diego, USA, Sept. 2003.

[13] M. Karray, Analytical evaluation of QoS in the downlink of OFDMA wireless cellular networks serving streaming and elastic traffic, *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1799-1807, May, 2010.