QoE-aware user association and resource allocation for mixed services in heterogeneous networks

To cite this article: Shuxian Li *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **490** 072023

IOP Conf. Series: Materials Science and Engineering **490** (2019) 072023 doi:10.1088/1757-899X/490/7/072023

QoE-aware user association and resource allocation for mixed services in heterogeneous networks

Shuxian Li [1,*], Liangrui Tang[1], Jiangyu Yan[2]

[1]State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China
[2]School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

*Corresponding author e-mail: lishuxian798@163.com

Abstract. In heterogeneous networks (HetNets), user association and resource allocation (UARA) is a great challenge for radio resource management (RRM). In this paper, a novel QoE-aware UARA scheme is proposed. The delay factor is introduced to QoE metric to weight delays' effect on service transmitting. Then the UARA scheme based on QoE is formulated as an optimization problem which is a mixed integer nonlinear programming (MINLP) problem. To solve the problem, we adopt the genetic algorithm (GA) and associate the user association and the resource with population and fitness respectively. Through the evolutions of genetic algorithm, the optimal UARA scheme can be worked out. The simulation results indicate that the proposed scheme can reduce packet loss rate, and improve the system load balancing performance.

1. Introduction

HetNets are the integration of macrocells and small cells including microcells, picocells and femtocells, which can eliminate the coverage holes and relieve the hot spots[1]. Thus, HetNets can provide better service for users compared to conventional networks. Meanwhile, the coexistence of different types of networks introduces great challenges in terms of UARA. Researchers have studied UARA schemes from different aspects [3-8], and objective metrics such as load balancing, energy efficiency, interference, etc. are used to optimize system performance, thus resources may be used blindly without quality of experience (QoE) improvement.

Generally, the delay and rate of services has an obvious effect on users' QoE. Howbeit, most researches fail to consider delay's effect of different types of services. By properly making use of the difference, users' QoE and resource efficiency can be improved. Motivated by the above reason, we study a QoE-aware UARA scheme from a new prospective in this paper. Unlike the efforts in above literatures, we introduce a delay factor to measure QoE based on actual rates and rate requirements. And then we model the UARA as an optimization problem of the sum QoE of the system services' QoS constraints which is a MINLP proplem. To solve the problem, the genetic algorithm for UARA is designed and the scheme can guarantee delay requirements and provide better load balancing performance.

2. System Model and Problem Formulation

2.1 System Model

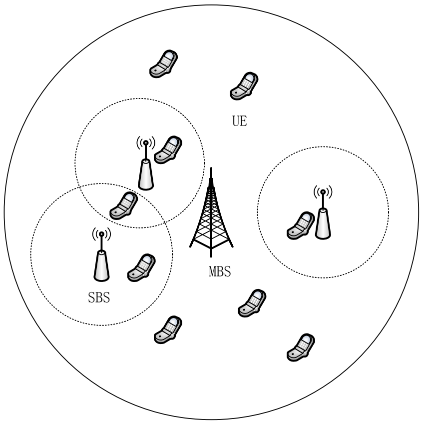


Figure 1. HetNet scene

Figure 1 shows the HetNet scene. In the scene, macro base stations (MBSs) and small base stations (SBSs) coexist and their covering areas overlap. Assuming there are N BSs and K users (one user only has one service) in the heterogeneous network. For presentation, denote the BS set as $\boldsymbol{N}=\{1,2,3,....,N\}$, and the service set as $\boldsymbol{K}=\{1,2,3...,K\}$, the system bandwidth is B per BS. When service k gets access to BS n, the achievable rate can be calculated by

$$r_{k,n}=B\log(1+\frac{P_n\left|h_{k,n}\right|^2}{\sum_{l=1,l\neq k}^{N}P_l\left|h_{l,n}\right|^2+\sigma^2}), \quad (1)$$

where P_n is the transmitting power of BS n, $h_{k,n}$ is the channel gain between service k and BS n, σ^2 is the thermal noise power.

2.2 Problem Formulation

In this paper, the mixed services in HetNets are divided into two types, important services (ISs) and standard services (SSs) and the former type has more stringent delay requirement than the latter.

Supposing τ_k is the waiting time of service k, DL_k is its deadline, $U_D(\tau_k)$ is the delay utility which is the function of waiting time. And in this paper, the delay utility is defined as

$$U_D(\tau_k)=1-\frac{1}{1+\exp(-m(\tau_k-DL_k))}, 0<\tau_k\leq DL_k, \quad (2)$$

where m is a tunable parameter. It is used to adjust the slope of delay utility curve around the deadline. As m increases, the starting point of obvious drop gets closer to DL_k and the decline rate grows faster and faster. Since ISs has more stringent delay requirement, m of ISs is larger than that of SSs.

And the derivative of the delay utility function can be used to characterize the urgency of services. The corresponding delay factors of ISs and SSs are defined as:

$$\omega_k^{IS}(\tau_k)=\beta_1(c_1+\beta_2\omega_k^0), \quad \omega_k^{IS}(\tau_k)\in[\varepsilon_2,1], \quad (3)$$

$$\omega_k^{SS}(\tau_k)=\beta_3(c_2+\beta_4\omega_k^0), \quad \omega_k^{SS}(\tau_k)\in[0,\varepsilon_1], \quad (4)$$

where $\omega_k^0=\left|\frac{dU_D(\tau_k)}{md\tau_k}\right|$, $\beta_1,c_1,\beta_2,\beta_3,c_2,\beta_4$ are variables that adapt ε_2 and ε_1.

A UARA scheme is designed to maximize a sum QoE based on the delay factor and rate utility, and it is formulated as

IOP Conf. Series: Materials Science and Engineering **490** (2019) 072023 doi:10.1088/1757-899X/490/7/072023

$$\begin{aligned} \max_{x_{k,n},y_{k,n}} \quad & \sum_{k=1}^{K} x_{k,n} w_k(\tau_k) U_R^k \left(\sum_{n=1}^{N} y_{k,n} r_{k,n}\right) \\ s.t. \quad & x_{k,n} = \{0,1\}, \forall k \in K, n \in N \\ & 0 \le y_{k,n} \le x_{k,n}, \forall k \in K, n \in N \; , \\ & \sum_{k=1}^{K} y_{k,n} \le 1, \forall n \in N \\ & \tau_k + \frac{L_k}{r_k} \le DL_k, \quad \forall k \in K \end{aligned} \tag{5}$$

where $U_R^k(\sum_{n=1}^{N} y_{k,n} r_{k,n})$ is the rate utility of service k, $U_R^{IS}(r_k) = 1 - \frac{1}{1+\exp(\gamma_k(r_k - r_{k\min}))}$, $\gamma_k > 0$ for ISs and $U_R^{SS}(r_k) = 1 - \exp(-\alpha_k r_k), \alpha_k > 0$ for SSs, L_k is the packet length in bit of service k, $r_k = \sum_{n=1}^{N} y_{k,n} r_{k,n}$, $x_{k,n}$ is a binary variable. That $x_{k,n} = 1$ means service k is associated with BS n, while $x_{k,n} = 0$ service k not associated with BS n, $w_k(\tau_k) = \omega_k^{IS}(\tau_k)$ for ISs and $w_k(\tau_k) = \omega_k^{SS}(\tau_k)$ for SSs.

3. Optimization and Solution

In this paper, we assume that a service accesses to one BS at most at once. Therefore, $r_k = \sum_{n=1}^{N} y_{k,n} r_{k,n} = y_{k,n} r_{k,n}$, and formula (7) can be rewritten as:

$$\begin{aligned} \max_{x_{k,n},y_{k,n}} \quad & \sum_{n=1}^{N}\sum_{k=1}^{K} x_{k,n} w_k(\tau_k) U_R^k (y_{k,n} r_{k,n}) \\ s.t. \quad & x_{k,n} = \{0,1\}, \forall k \in K, n \in N \\ & \sum_{n=1}^{N} x_{k,n} \le 1, \forall k \in K \\ & 0 \le y_{k,n} \le x_{k,n}, \forall k \in K, n \in N \, , \\ & \sum_{k=1}^{K} y_{k,n} \le 1, \forall n \in N \\ & \tau_k + \frac{L_k}{r_k} > DL_k, \quad \forall k \in K \end{aligned} \tag{6}$$

Let $X_n = [x_{1,n}, x_{2,n}, ..., x_{K,n}]^T, \forall n \in N$ which represents the user association of BS n. Denote $\boldsymbol{X}$ as the set of X_n, i.e. $\boldsymbol{X} = [X_1, X_2, ..., X_N]$, and it is the user association scheme of the overall system. Similarly, we define the resource allocation scheme as $\boldsymbol{Y}$, $\boldsymbol{Y} = [Y_1, Y_2, ..., Y_N]$ where $Y_n = [y_{1,n}, y_{2,n}, ..., y_{K,n}]^T, \forall n \in N$. Formula (6) is a MINLP problem which is difficult to solve directly. Genetic algorithm (GA) can solve complexed optimizing problem that conventional optimizing algorithm can hardly solve. Thus, we adopt genetic algorithm to solve our problem.

3.1 Determination and Solution of Fitness Function

In this paper, the fitness function is determined by $\boldsymbol{X}$ and $\boldsymbol{Y}$. Combined with formula (6), the fitness function is defined as

$$fit(\boldsymbol{X},\boldsymbol{Y}) = \sum_{n=1}^{N}\sum_{k=1}^{K} x_{k,n} w_k(\tau_k) U_R^k (y_{k,n} r_{k,n}) \, . \tag{7}$$

Among all schemes, the best scheme is the one with the largest fitness. Therefore, the optimal scheme can be obtained by solving the following optimizing problem:

$$\begin{aligned} &\max_{X,Y} \ fit(\boldsymbol{X},\boldsymbol{Y}) \\ s.t. \quad & x_{k,n} = \{0,1\}, \forall k \in K, n \in N \\ & \sum_{n=1}^{N} x_{k,n} \le 1, \forall k \in K \\ & 0 \le y_{k,n} \le 1, \forall n \in N, y_{k,n} \in Y_n \\ & \sum_{y_{k,n} \in Y_n} y_{k,n} \le 1, \forall n \in N \\ & \tau_k + \frac{L_k}{r_k} > DL_k, \quad \forall k \in K \end{aligned} \tag{8}$$

The fitness function can be calculated when the corresponding individual is determined, i.e. $fit(\boldsymbol{X},\boldsymbol{Y})$ can be calculated when $\boldsymbol{X}$ is given. In such situation $fit(\boldsymbol{X},\boldsymbol{Y}) = fit(\boldsymbol{Y})$.Determined by locations of service k and BS n, $r_{k,n}$ is a constant, thus and the following formula is equivalent to formula (8).

$$\begin{aligned} &\max_{Y_n} \sum_{k \in \{l | x_{l,n}=1\}} w_k(\tau_k) U_R^k(y_{k,n}) \\ s.t. \quad & 0 \le y_{k,n} \le 1, \quad \forall k \in \{l \mid x_{l,n} = 1\} \\ & \sum_{k \in \{l | x_{l,n}=1\}} y_{k,n} \le 1 \\ & \tau_k + \frac{L_k}{y_{k,n} r_{k,n}} \le DL_k, \quad \forall k \in \{l \mid x_{l,n} = 1\} \end{aligned} \tag{9}$$

By calculating formula (9), we can get the optimal power allocation at a single BS. To solve formula (9), we adopt interior-point method [9]. Based on the above analysis, we can get the fitness function value through formula (9).

3.2 Optimization and Solution

We adopt decimal numbers to code BSs. There are N BSs in the heterogeneous network, and they are coded as $1 \sim N$ respectively. The population size is s, and the population is denoted as $\boldsymbol{P} = [P_1, P_2, ..., P_s]^T$, where $P_m = [p_{m,1}, p_{m,2}, ..., p_{m,K}], 1 \le m \le s$ and $p_{m,k} \in [0, N]$. When $p_{m,k} = n \ (k \in K, n \in N)$, service k is associated to BS n, correspondingly $x_{k,n} = 1$, and $x_{k,n'} = 0, n' \ne n$. When $p_{m,k} = 0$, service k gets no access to any BS, i.e. $x_{k,n} = 0, \forall n \in \boldsymbol{N}$. Apparently, P_m is a user association scheme and equivalent to $\boldsymbol{X}$. The goal is to find the proper P_m to maximize the system utility. We design the following algorithm to optimize the UARA scheme, and the algorithm is shown in Table 1.

Table 1. GA for UARA scheme

GA for UARA scheme
Initialization: The iteration $t = 0$, maximum iteration number is set as $t_{\max}$, population $\boldsymbol{P}_t$, Set $\boldsymbol{X}=\boldsymbol{0}, \boldsymbol{Y}=\boldsymbol{0}$.
while $t \le t_{\max}$
for i=1 to s
for n=1 to N
$\boldsymbol{X} = P_i$;
Calculate formula (9),and get the solution Y_n;
end for
$\boldsymbol{Y} = (Y_n)_{1 \times N}$,and substitute $\boldsymbol{Y}$ into formula (7);
Calculate formula (7) and obtain the its value fit_i;
end for
perform crossover, mutation and selecting operations on $\boldsymbol{P}_t$ and get $\boldsymbol{P}_{t+1}$;
$\boldsymbol{P}_t = \boldsymbol{P}_{t+1}$;
$t = t + 1$;
end while
calculate the fitness functions of $\boldsymbol{P}_t$;
Find the individual with largest fitness, and get $\boldsymbol{X}$ and $\boldsymbol{Y}$.

4. Analysis of Simulation Results

The networks consists of 1 MBS (46 dBm transmission power) which is located in the center and 5 SBSs (30dBm transmission power) which are uniformly dropped in a square specified by $[0,1000]\times[0,1000]$. The coverage radiuses are 1km and 300m for MBS and SBSs respectively, and the system bandwidth B is $10MHz$ per BS. The path loss is modeled as $PL_{n.k}=30+40\log_{10}d_{n,k}$ where $d_{n.k}$ is the distance between service k and BS n in meter. Services to be transmitted include 40% ISs and 60% SSs which are distributed uniformly and geographically. More specially, the tolerable delay of ISs is 250ms, its required rate is 30kbps, the tolerable delay of SSs is 2000ms, and its required rate is 5kbps. The rest of simulation parameters are summarized in Table 2.

Table 2. Simulation parameters and values

Parameter	Value
σ^2	-95dBm
η	0.9
m (for ISs)	0.1
m (for SSs)	0.018
β_1, c_1, β_2	0.5,1,4
β_3, c_2, β_4	1,0,4
γ_k	1.5
α_k	5

In the following, we carry out performance comparisons among max SINR scheme, QoE scheme and the proposed scheme. Using max SINR scheme, services associates with the BS of highest SINR, and shares its bandwidth equally with other users/services of the BS. QoE scheme aims to optimize the aggregate rate utility which is the function of data rate. For the convenience, the proposed scheme is denoted as wQoE (short for QoE with ω).

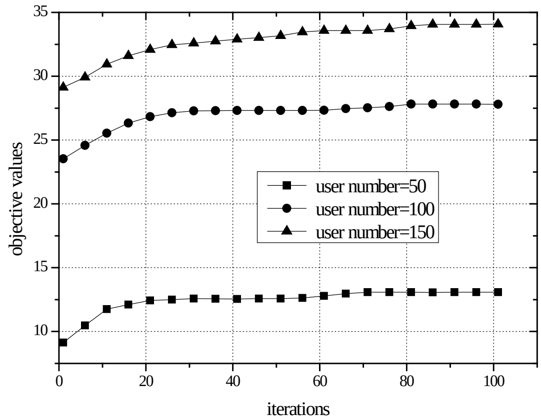
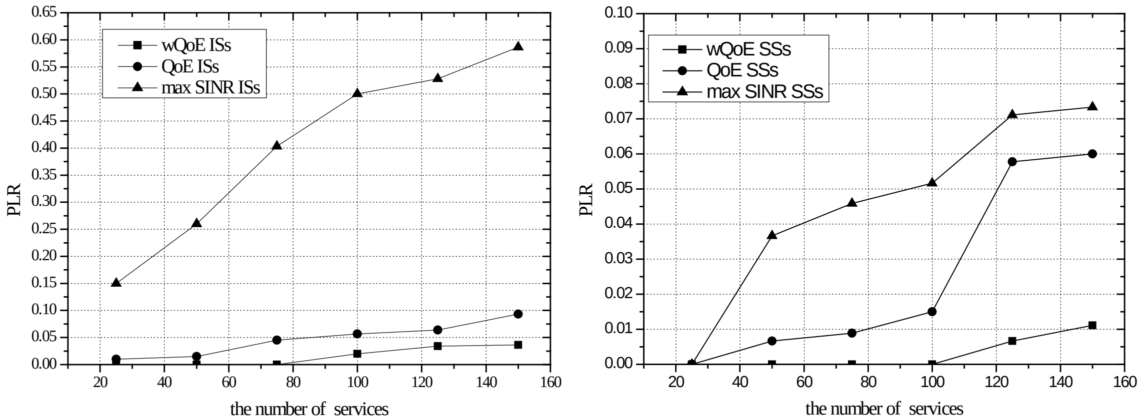


Figure 2. The convergence of GA for UARA

Figure 2 shows the convergence performance of GA for UARA. Through the iteration increases, the objective value increases gradually and gets convergent after 80 iterations regardless of user numbers. From the figure, we can see that the algorithm can quickly converge.

In this paper, the definition of packet loss is defined as following: for service k, if $\tau_k+\frac{L_k}{r_k}>DL_k$, then it is thought to be lost. Thus, the smaller the packet loss rate (PLR) is, the more services are transmitted and the better the delay requirement is satisfied.



(a)PLR of ISs （b） PLR of SSs

Figure 3. PLR of ISs and SSs

Figure 4 compares PLR of wQoE, QoE and max SINR schemes. Figure 3(a) indicates the PLR of ISs and Figure 3(b) the SSs. In the two figures as the number of services increases from 25 to 150, wQoE has the best PLR performance, followed by QoE, max SINR worst. The observations can be explained by the fact that max SINR and QoE schemes do not pay attention to the time natures. Services associate with the BSs that provide the max SINR among neighbor BSs when max SINR scheme is adopted and share its bandwidth equally. So the more services one BS has, the less resource its services can obtain. With the objective to maximize aggregate rate utility, PLR of QoE scheme is much better than that of max SINR scheme. By contrast with the two schemes, wQoE scheme introduces the delay factor taking packet waiting time and delay constraint into account. Since there is a positive correlation between the delay factor and the service urgency, packets approaching the deadline will get greater probability to be transmitted.

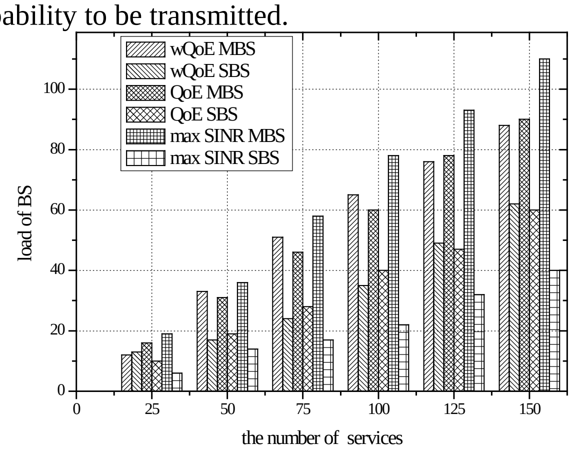


Figure 4. Load balancing condition

Figure 6 shows the load balancing performance of the three schemes. Max SINR associates many services with MBS while SBSs serve few services. Obviously, the load is unbalanced. In wQoE and QoE scheme, the load balancing performance is improved. More services get access to SBSs comparing with that of max SINR scheme. Max SINR scheme associates services with the BS providing the highest SINR. Since the transmitting power of MBS is much larger than that of SBSs, the service number of MBS is much larger than SBSs'. Consequently, MBS is overloaded and SBSs are lightly loaded. The wQoE and QoE scheme stand from the point of view of services, thus users transmit their services through BSs with light load.

5. Conclusion

To guarantee the QoE of mixed services in HetNets, the delay factor is introduced into conventional QoE based UARA scheme in this paper, and A novel UARA scheme is proposed. We model the scheme as a MINLP problem and design the GA for UARA to solve it. The simulation results show that the algorithm converges quickly to the optimal solutions and the proposed scheme can guarantee services' QoE, ensure delay requirements of services, and balance the network load when compared to the conventional UARA scheme.

Acknowledgements

This work is supported by the National High Technology Research and Development of China 863 Program (2014AA01A701), the Fundamental Research Funds for Central Universities of China (No.2015XS07), and Beijing Natural Science Foundation (4142049).

References

[1] Ghosh A , Mangalvedhe N , Ratasuk R , et al. Heterogeneous cellular networks: From theory to practice[J]. Communications Magazine IEEE, 2012, 50(6):54-64.
[2] Hongxiang S , Hangsheng Z , Youming S , et al. QoE-Aware Downlink User-Cell Association in Small Cell Networks: A Transfer-matching Game Theoretic Solution with Peer Effects[J]. IEEE Access, 2016, 4(99):10029-10041.
[3] Luo X . Delay-Oriented QoS-Aware User Association and Resource Allocation in Heterogeneous Cellular Networks [M]. IEEE Press, 2017.
[4] Schepper, T. D., Latré, S., & Famaey, J.. A transparent load balancing scheme for heterogeneous local area networks. Ifip/ieee International Symposium on Integrated Network Management. IEEE, 2017.
[5] Zhou, T. Q., Huang, Y. M., & Yang, L. X. Joint user association and resource partitioning with qos support for heterogeneous cellular networks. Wireless Personal Communications, 2015, 83(1), 383-397.
[6] Zhu, J., & Yang, H. C. Low-complexity qos-aware coordinated scheduling for heterogenous networks. IEEE Transactions on Vehicular Technology, 2016(99), 1-1.
[7] Adedoyin M A, Falowo O E. QoS-based radio resource management for 5G ultra-dense heterogeneous networks[C]// European Conference on Networks and Communications. IEEE, 2017:1-6.
[8] Tian H , Xie W , Gan X , et al. Hybrid user association for maximising energy efficiency in heterogeneous networks with human-to-human/machine-to-machine coexistence.[J]. Iet Communications, 2016, 10(9):1035-1043.
[9] Byrd R H , Hribar M E , Nocedal J . An Interior Point Algorithm for Large Scale Nonlinear Programming[M]. Society for Industrial and Applied Mathematics, 1999.