User Association for Load Balancing in 5G Network Based on Min-Max Optimization

Ji-hong GUI, Sui-xiang GAO and Zhi-peng JIANG

University of Chinese Academy of Sciences, No.19 (A) Yuquan Road, Shijingshan District, Beijing, P. R. China

*Corresponding author

Keywords: 5G network, Load balancing, User association, Min-Max optimization, Bisection method.

Abstract. The rapid increase of user equipments has stressed the heterogeneous and enormous wireless network in recent years. As an effective way to promote resource utilization and improve network performance, load balancing has becoming more and more important in the upcoming fifth generation (5G) network. In this work, a load balancing problem with limited bandwidth constraints of different cells is investigated from base stations perspective and the way to balance these loads is to design an proper user association scheme. Consequently, A 0-1 min-max linear optimization model is established to describe the problem mathematically and robustly. Moreover, based on the upper and lower bound analysis of the model, a bisection method algorithm is designed to approach the optimal solution and its time complexity is $O(mn\log n)$, where $m$, $n$ are the amount of cells and users respectively. The numerical results of simulation experiment demonstrate the reliability and rapid computational efficiency of bisection method.

Introduction

As the amount of user equipments (UE) increases rapidly and significantly, the wireless networks has become more heterogeneous and enormous in recent years [1], [2]. At the same time, the huge demand for user mobile data cause massive usage of bandwidth sources and the expansion of channel capacity in the upcoming 5G network. Additionally, excessive load will stress the traffic pressure and communication congestion while the low load cannot meet the user quality of service (Qos) and the resources redundancy in the wireless access network. Nevertheless, it is inevitable for the unbalanced distribution of the load because of the user position and status without regularity [3].

User association substantially affects the system performance in a load-balancing network, involving the cell selection and handover decision. In the existing Long Term Evolution Advanced (LTE-A) systems, the maximum received signal strength (max-RSS) is the most prevalent evaluation criterion to decide which cell the UE should be connected to. The decision is to ensure the Qos requirements and maximizing the radio resource exploitation [4]. However, the channel nature and equilibrium targets in the emerging 5G network inevitably render such a rudimentary user association rule ineffective [3]. The emerging 5G network with various types of cells coexisting eagerly needs operators and managers to design an effective user association scheme for balancing the load of the dense heterogeneous small cell networks (HSCN) and meet the limited bandwidth.

Present works shown below have provided lots of excellent contributions on load balancing through user association. An user association strategy for balancing the network loads by maximizing the weighted sum of long-term rates is designed in [5]. Meanwhile, optimal user association for load balancing problem in cellular network with hybrid cognitive radio relays is investigated by [6]. Furthermore, an fair user association policy for encapsulating the reallocation cost of potential handover and capturing the erratic nature of mm Wave channel in [7].

In this work, we consider the trade-off of user connections amount between various cells which causes too much pressure of cells service. It is worth to mention that diversity of base station types and different user bit rate demands are comprehensively considered in this work owing to the
multiple types of data requests from numbers of users such as voice and video, which are reflected in the different bandwidth requirements ultimately. Furthermore, a min-max combinatorial optimization model with the limit bandwidth of different cells is established to measure load equilibrium quantitatively. As a consequence, an $O(mn \log n)$ algorithm like bisection method is provided based on the upper and lower bound from the rigorous mathematical analysis of optimization model.

The rest of this paper is organized as follows. The min-max optimization model for load balancing is established and the upper and lower bound are provided in Sec. II. A bisection method based on the bounds of optimization model is designed and the time complexity of it is provided in Sec. III. Ultimately, the simulation results and numerical analysis are presented in Sec. IV.

**Problem Formulation**

**Model Formulation**

Consider a 5G Macro-Femto-Pico cell mixed heterogeneous networks (HetNets) scenario illustrated in Fig. 1. There are $m$ cells in the region $\Omega$, which consists of $m_1$ macrocell base stations (MBS), $m_2$ femtocells and $m_3$ picocells. Naturally, $m = m_1 + m_2 + m_3$. Meanwhile, supposed that there are $n$ user equipments (UE) evenly distributed in $\Omega$. The $m$ cells and $n$ UEs make up the HetNets.

Let $P_i$ denote the transmitted power of cell $i$ and the $l_{ij}$ is the channel gain from cell $i$ to UE $j$. Then the received power of UE $j$ from cell $i$ is $P_{ij} = P_i l_{ij}$. The signals from different cells will interfere UE and the Signal to Interference plus Noise Ratio of UE $j$ from cell $i$ is denoted by $r_{ij} = \frac{P_i l_{ij}}{\sum_{h \neq i} P_h l_{hj} + \delta^2}$, where the numerator is received signal and the denominator is interference from other cells and system noise. Let $R_j$ denote the bit rate of service requested by user $j$ and according to Shannon’ formula, the bandwidth requested for cell $i$ from UE $j$ is denoted by $b_{ij} = \frac{R_j}{\log_2(1+r_{ij})}$.

Additionally, let binary variable $x_{ij}$ denote the association between cell $i$ and UE $j$. The UE $j$ will associate to cell $i$ if $x_{ij} = 1$, otherwise $x_{ij} = 0$ indicates no connection between them. Based on the assumption that UE will connected to only one cell at same point, the constraint condition $\sum_{j} x_{ij} = 1$ must be satisfied. Moreover, the total requested bandwidth for cell $i$ from all connected UEs is at most the limited maximum bandwidth $BW_i$, namely, $\sum_{j} b_{ij} x_{ij} \leq BW_i$.

Essentially, the objective of this work is to balance the load at each cell $i$. The associated UE amount at cell $i$ is $\sum_{j} x_{ij}$ and the higher associated user amount will stress service pressure at same cells. Therefore, the load balancing should minimize the maximum load cell and min-max optimization has robustness and reliability mathematically.

Ultimately, the min-max optimization model is established as model (I).
Model Transformation

The solution $X=[x_{ij}]_{mn}$ is the form of $m \times n$ matrix, whereas the decision variable should be vector form traditionally. For the purpose, stacking the columns of matrix $X$ and reforming it as vector by $x_{ij} = z_i + m(j-1)$. Then the objective function and constraints should be transformed as the function of vector $z$ entirely, which can be approached by structure matrix $A$, $C$ and $D$.

Firstly, the sum of matrix $X$'s columns, i.e., $\sum_i x_{ij}$ can be transformed by matrix

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{m \times n},$$

where $\sum_i d_{ik} z_k = \sum_i x_{ij}$ is rewritten for constraint (2) in model (I).

Analogously, the sum of matrix $X$'s rows, i.e., $\sum_j x_{ij}$ an transformed by matrix

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n \times m},$$

and

$$\sum_k c_{ik} z_k = \sum_j x_{ij}$$

is for the objective function in model (I).

Lastly, the structure matrix $D$ can be approached by imitating the forms of matrix $C$ for rewriting the constraint (1) in model (I). Similarly,

$$D = [D_1, D_2, \cdots, D_m]_{n \times m},$$

and

$$D_j = \begin{bmatrix} b_{ij} & 0 & \cdots & 0 \\ 0 & b_{j2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{jm} \end{bmatrix}_{m \times n},$$

where $\sum_j d_{ik} z_k = \sum_j b_{ij} x_{ij}$.

Therefore, the model (I) can be transformed as model (II). Then the objective function and the constraints in model (II) are all the function of vector $z$ and remain the property of linearity.

Algorithm

Based on the model (II) and the analysis of its bounds, Abisection method is designed for the load balance based on the model (II). For the completeness of the narrative, variable substitution will be adopted for the model (II). Let $t = \max_i \sum_k c_{ik} z_k$, then $t \geq \sum_k c_{ik} z_k$, $\forall i$. Therefore the model (II) can be written as
\[
\begin{align*}
\min \quad & t \\
\text{st.} \quad & t \geq \sum_{k}^{mn} c_{ik} z_{k}, \quad \forall i \quad (1) \\
& \sum_{k}^{mn} d_{ik} z_{k} \leq BW, \quad \forall i \quad (2) \quad \text{(III)} \\
& \sum_{i}^{m} a_{ik} z_{k} = 1, \quad \forall k \quad (3) \\
& z_{k} = 0/1, \quad \forall k
\end{align*}
\]

Algorithm: Bisection method for balance

Input: \( t = \frac{n}{m}, T = n, \) tolerance \( \epsilon > 0. \)

Output: The optimal solution \( \bar{z} \)

1: While \( T - 1 > \epsilon \) Do
2: \( t := \frac{1+T}{2} \).
3: Solve the linear feasibility problem (III) with fixed \( t \).
4: If (III) is feasible and \( z \) is the feasible solution of it then
5: \( T := t \).
6: Else
7: \( t := z \).
8: Return \( \bar{x} = z \).

The above problem is a linear feasibility problem about \( z \) with a fixed \( t \). There are two cases for the optimal solution \( \bar{z} \), \( t \geq \bar{z} \) if the problem is feasible with a fixed \( t \) and then try a smaller \( t \) to judge the feasibility. Otherwise, \( t \leq \bar{z} \) and a larger \( t \) should be selected properly to make the problem feasible.

The basic idea of bisection method is to adjust \( t \) according the feasibility of the problem. Meanwhile, the initial lower bound \( t \) is selected as \( \frac{n}{mn} \) and the initial upper bound \( T \) is set as \( n \).

In each iterations, the time complexity of judging the feasibility linear problem (III) is \( O(mn) \) for the \( mn \) decision variables in the model. Therefore the total time complexity of bisection method is \( O(mn \log n) \), where the default \( \epsilon \) is \( 1/2 \) due to the integer solution property.

Performance Simulation

In order to verify the validity and accuracy of the algorithm, numerical simulation experiments implemented by MATLAB are presented in this section. The first part is the construction of simulation environment including the hardware device and the network topology. The next part is the analysis of aforementioned algorithm validity and correctness. The last part provides an optimal access scheme and verifies the \( O(mn \log n) \) time complexity of the bisection method.

System Simulation Environment

The simulation parameters include the system parameters of hardware devices and the network parameter. The simulation network consists of 10 different type cells, i.e., 1 MBS, 3 Femto cells and 6 Pico cells. Meanwhile the max bandwidth of them are 200, 100 and 50 MHZ respectively because it’s realistically assumed that the greater the cells' capacity, the larger the bandwidth of them.

The Calculation Validity Analysis

For convenience, numbering the UEs from 1 to 100 and running the bisection method in the above simulation environment. The actual running time is just 0.4927s and the optimal access scheme between UEs and different cells illustrated in Fig. 1.

For the quantitative evaluation of the associated UEs' amount of different type cells, the statistics of these associated amount are illustrated in Fig. 2. Obviously, the associated amount to MBS is close to the amount to femtocell 1 and 2. Certainly, the ideal circumstance is absolutely equalization while it’s impossible to approach in order to meet the bandwidth demand because the different location and transmission power of these cells. It's effective for the small cells to absorb the excessive UEs originally associated to MBS and the network will reduce the traffic congestion on the MBS. In addition, low cost and easy deployment can make an advantage for building more small cells.
The Calculation Speed Analysis

In order to verify the rapid calculation of the algorithm, two group complementary simulation are executed. The first is to investigate the relationship of calculation time and total UE amount $n$ with different cell amount $m$, while the other is the relationship of calculation time and $m$ with different $n$. The default network parameters involving $m$ and $n$ are expanded in this part and $m$ range from 10 to 100 and $n$ is from 100 to even 1000, illustrated in Fig. 4 and Fig. 5.

Totally speaking, all of the simulation result about calculation time are not more than 60 seconds, especially for large scale problem with $m=100$ and $n=1000$. Meanwhile, the trend of line growth in these two picture is very moderate without spurting obviously, demonstrating the robustness and solidity of the algorithm. In additional, comparing Fig.4 and Fig.5, the trend of $n$ is approximately between linear and quadratic while the trend of $m$ is relatively more linear, which provides strong support for the $O(mn \log n)$ time complexity of the bisection method.

Conclusion

In this paper, a 0-1 min-max optimization model is established to balance the load and meet the bandwidth demand in 5G heterogeneous Macro-Femto-Pico mixed networks from the stations view. Based on the lower and upper bound analysis of this model, an optimal UE access scheme will be approached through the Bisection method and the time complexity of the algorithm is $O(mn \log n)$. The experimental results demonstrated the significant effectiveness and extremely fast solving speed. Moreover, it's indicated that the massive deployment of small cell will balance the load effectively, whose low cost and easy deployment are both very instructive for the establishment of 5G station.

Acknowledgments

Thanks to China Mobile Group Beijing Company Limited (CMBJ).
References

[1] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. Mccann, “Modulation and multiple access for 5g networks,” IEEE Communications Surveys and Tutorials, vol. 20, no. 1, pp. 629–646, 2018.

[2] D. N. K. Jayakody, S. K. Biswash, and D. Rui, “Recent advances and future research challenges in non-orthogonal multiple access for 5g networks,” in IEEE Vehicular Technology Conference, 2018.

[3] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. K. Wong, R. Schober, and L. Hanzo, “User association in 5g networks: A survey and an outlook,” IEEE Communications Surveys and Tutorials, vol. 18, no. 2, pp. 1018–1044, 2015.

[4] D. Kakadia, J. Yang, and A. Gilgur, Evolved Universal Terrestrial Radio Access Network (EUTRAN). New Delhi: Springer India, 2017, pp. 61–81.

[5] T. Zhou, Z. Liu, J. Zhao, C. Li, and L. Yang, “Joint user association and power control for load balancing in downlink heterogeneous cellular networks,” IEEE Transactions on Vehicular Technology, vol. PP, no. 99, pp. 1–1, 2016.

[6] H. Guo, F. Zhou, L. Feng, P. Yu, and W. Li, “User association for load balancing in cellular network with hybrid cognitive radio relays,” in International Conference on Network and Service Management, 2017, pp. 1–8.

[7] S. Goyal, M. Mezzavilla, S. Rangan, S. Panwar, and M. Zorzi, “User association in 5g mm wave networks,” in Wireless Communications and Networking Conference, 2017, pp. 1–6.