

Available online at <http://ijdea.srbiau.ac.ir>

Int. J. Data Envelopment Analysis (ISSN 2345-458X)

Vol. 12, No. 1, Year Article ID IJDEA-00422, Pages 1-9
Research Article



International Journal of Data Envelopment Analysis



Science and Research Branch (IAU)

Proposing a mathematical programming model for load balancing in mobile cellular networks

Shahram Saeidi^{*1}, Sahar Khoshfetrat²

¹ Department of Industrial Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran,

² Department of Mathematics, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

Received 23 July 2023, Accepted 18 January 2024

Abstract

A mobile cellular network consists of several base stations as antennas. Each antenna operates under a coverage radius and can serve users within its range. Based on the number of stations in the city and the overlapping of their service areas, a user may simultaneously be in the radio coverage radius of several antennas. However, he/she will be able to receive service from only one antenna. The number of users for whom each antenna can provide service is limited, and increasing the workload of the base station can lead to disturbances in the network performance. In this research, a linear programming model was presented to balance the workload among base stations in cellular mobile networks, and it was implemented in Lingo software on 12 real data sets in Tabriz city. The simulation results show that the proposed model can achieve the global optimal solution with the objective function equal to zero in eight examples. In the other examples, the local optimum solution with a minimal objective function value is obtained, and the workload is balanced throughout the network base stations.

Keywords: Linear Programming, Load Balancing, Cellular Mobile Networks, Lingo.

* Corresponding author: Email: sh_saeidi@iaut.ac.ir

1. Introduction

Each network consists of several connected nodes to perform a specific task. The mobile phone network consists of some fixed nodes as base stations, known as Base Transceiver Station (BTS) antennas, as service providers, and many mobile nodes that include client users. Each node has to identify and authenticate the users requesting the service and provide them with telecommunication services such as voice/video calls, SMS, or Internet. Each node (antenna), according to its type, communication protocol, and hardware capacity, can provide limited services, and more service requests will lead to the denial of the request as well as a decrease in the quality or speed of the service to users [1]. To maximize the efficiency and service quality of the entire network, it is recommended that the workload of the entire network is preferably distributed evenly throughout the entire network [2]. Load balancing is the process of dividing and distributing network traffic load between more than one server and provides the possibility of serving more activities on the entire network. In this procedure, the workload is appropriately distributed throughout the network, ensuring no network node is overloaded with traffic. Load imbalance is one of the significant problems of mobile networks. If there is no mechanism to maintain or balance the load in the network, it will lead to an increase in delay and a decrease in service quality [3]. Due to the unstable topology and the need for a centralized server in mobile networks, proper optimization of the load balancing between base stations has caused a challenge in these networks. Several efforts have been made to provide standard load-balancing algorithms to improve essential metrics related to these networks [3].

Load balancing using the resource allocation algorithm is based on distributing information packets to the

network's cells, where more users are gathered [2]. This is determined by using the distribution algorithm; the traffic load is forwarded to the direction where more resources are available, and the provision of services is determined according to the changes in the distribution of users. As there is an attempt to balance the workload on the entire network level between users, bandwidth limits, and quality of service workload must also be observed [4]. This article presents a mathematical programming model to balance the workload of base stations on the entire mobile network. The research aims to answer which antenna should serve which user at which moment so that the network workload always remains balanced. The proposed model is coded in Lingo software, and the calculation results on three experimental examples and twelve examples with real data indicate its accuracy and efficiency.

The rest of the paper is organized as follows: in the second section, the review of literature and previous works are discussed. The third section describes the proposed mathematical model, and the fourth section presents the numerical simulation. The general conclusion of the paper is given in section 5.

2. Literature Review

The cell is the smallest coverage area in the mobile network and is determined by the radio coverage of a BTS sector. The method of cell division and determining the radius of the cells depends on the geographical conditions of the covered area. It considers buildings and artificial obstacles, transmitter power, antenna gain, and type. Moreover, it has receiver sensitivity, and they usually use vectorization for the radio coverage of each cell [4]. In Figure (1), an example of a cellular network is given. A cell is an area covered by one or more BTSs. Figure (2) shows an example of BTS antenna placement in a small cell network, where

each cell is divided into three sectors. In this division, which is the most common type, each BTS tower includes three antennas with an angle of 120 degrees to

each other, and each antenna is responsible for covering the users of a sector [4].

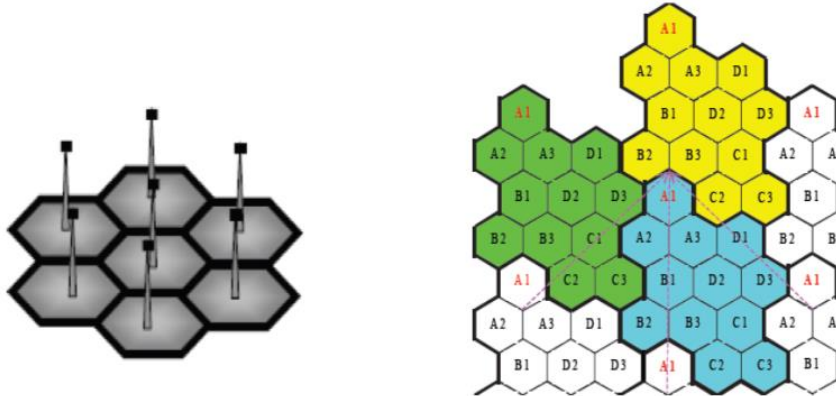


Fig. 1. A sample cellular network

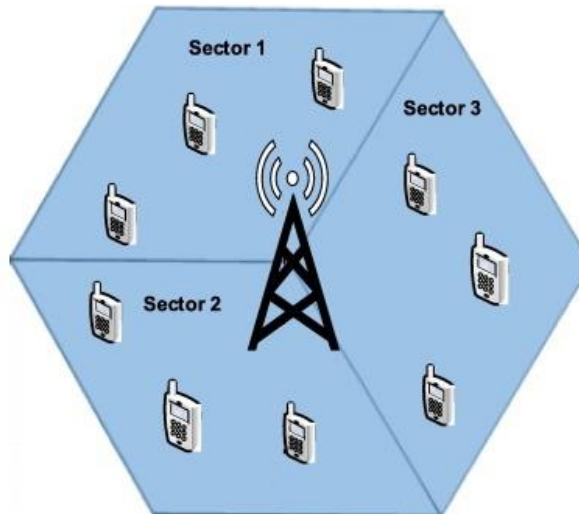


Fig. 2. BTS antenna placement in a cellular network (Saad et al., 2023)

Due to the mobility of users in the network, some of them always move away from the coverage radius of their base node and enter the range of new antennas. In a cellular network, radio frequencies or channels are not permanently assigned to a subscriber for the duration of a call. Suppose a mobile subscriber moves from one cell's coverage area to another during the conversation. In that case, it is necessary to direct the conversation to the new cell without interruption, and

sometimes, this channel transfer may be done for traffic management purposes. This process is called handover [1].

Load balancing can be defined as dividing and distributing tasks between more than one server, which provides the possibility of providing service to more users, and the efficiency of the whole system can increase. Load balancing is frequently performed in computer systems and telecommunications to share the load [5]. The essential components in the load-

balancing process are load definition, load measurement, and load-balancing mechanism [2]. The load criterion proposed by Nicholson et al. (2006) is based on the received signal strength and bandwidth indicator values that can be calculated when a new user is connected to an access point [6]. The authors considered the bandwidth as the primary measure as a percentage of the time the access point is busy transmitting. Considering that in mobile networks, base stations can have many services with different characteristics, the evaluation of the number of users or connections cannot be a correct measure to calculate the load of the station, as the delay, packet loss rate, and call blocking rate are only implicit information of the conditions and can only be helpful for decision-making [5].

Few studies have been done in this field with the possibility of handovers, which are primarily in systems based on the IEEE 802.16e standard, such as the algorithms presented by Lee et al. (2007) and Moiseev et al. (2006) [7,8]. Liu et al. (2008) proposed a handover with frequency reuse between frequency assignments to solve unbalanced load distribution in 802.16e systems. Moon et al. (2008) defined a load-balancing scheme in WiBro systems as a subset of IEEE 802.16 standards developed based on the mobile network. They proposed a combination algorithm of scheduling and dual handover. Casey et al. (2008) proposed an algorithm for mobile networks that can balance the system load with a directed handover mechanism. In this algorithm, the load balancing logic is located in the base station; in this way, the base stations force the users to be located in overlapping areas to change their connection from high-load antennas to low-load antennas [5].

Bazzo et al. (2010) presented a multi-layer handover algorithm for mobile WiMAX networks. The purpose of this algorithm is to balance resources between layers by transferring users from one layer to

another based on the channel status and their quality-of-service requirements [9]. Nazari (2010) presented a new distributed packet scheduling algorithm in mobile networks. This algorithm estimates the available resources for each connection in the uplink, and its purpose is to provide the requested resources for the connections according to the specifications and quality of service requirements of each connection [10]. In the load-balancing model presented by Zhou et al. (2011), one of the possible ways for the interaction between WiMAX and wireless local area network (WLAN), the load distribution between these two networks according to the services required by applications and network service characteristics is introduced. Askarian (2012) conducted a review study for load-balancing in WiMAX networks. Sharma et al. (2012) studied load-balancing techniques in mobile cellular networks and compared the efficiency of these methods. Afzal and Kavitha (2019) reviewed a detailed encyclopedia of load-balancing techniques. They discussed the advantages and limitations of existing methods with critical challenges that will be considered for developing efficient load-balancing algorithms in the future. The authors also proposed new insights regarding load balancing in cloud computing [11,12].

Lim et al. (2020) developed a load-balancing algorithm for mobile devices in an edge cloud computing environment based on graph coloring and genetic algorithms. They showed that the proposed load-balancing algorithm increases virtual machines' average CPU utilization, indicating high usage of edge servers [3]. Duan et al. (2022) presented a peer-to-peer three-layer cloud model for load balancing in mobile networks. Saad et al. (2023) simulated load balancing in 5G and 6G networks, showing that distance-based algorithms perform better [13].

Reviewing the literature and studying the previous work show that the mathematical

programming method has not been used for load-balancing in cellular networks so far, and the need to use operations research-based optimization methods is felt more.

3. The Proposed Model

In this section, the parameters and variables are introduced and the mathematical programming model is presented.

3.1 Parameters and Variables

I : The total number of users in the network at the moment

j : Total number of active stations in the network

K_{ij} : The volume of service received by the user i from antenna j

K_{\max} : The maximum service power of an antenna

x_i : Longitudinal coordinate of user i

y_i : Transverse coordinate if user j

B_{sx_j} : Longitudinal specification of base station j

B_{sy_j} : Transverse coordinates base station j

d_{ij} : Euclidean distance of user i from base station j

CR_j : Radio coverage radius of base station j

α_{ij} : Binary decision variable, whether user i is covered by antenna j . If yes, its value is 1; otherwise, zero.

β_{ij} : Binary decision variable, whether the antenna j should serve the user i or not. If yes, its value is 1; otherwise, zero.

3.2. The Objective Function

The objective function of the proposed model is to optimize the workload balance among network stations. The workload of

an antenna is the total of the service it provides to the users assigned to it at one moment. In the proposed model, the low value of the average workload (\overline{WL}) is

irrelevant as the reduction of (\overline{WL}) means reducing the workload of individual cells and thus reducing the amount of service to users. Therefore, the main goal of this model is to balance the workload provided by the network cells. To maintain this balance, we try to minimize each cell's workload deviation from this average. As a result, the objective function of the proposed model is equal to the total deviation of the workload of each antenna from the average value, which should be minimized.

3.3. The Mathematical Model

$$\min Z = \sum_{j=1}^J |WL_j - \overline{WL}| \quad (1)$$

$$s.t. WL_j = \sum_{i=1}^I \sum_{j=1}^J \beta_{ij} K_{ij}, \forall j \quad (2)$$

$$\overline{WL} = \frac{\sum_{j=1}^J WL_j}{J} \quad (3)$$

$$\sum_{j=1}^J Y_{ij} = 1, \forall i \quad (4)$$

$$\sum_{i=1}^I \sum_{j=1}^J \beta_{ij} K_{ij} \leq K_{\max}, \forall j \quad (6)$$

$$d_{ij} = \text{sqrt}\left(\left(x_i - Btsx_j\right)^2 + \left(y_i - Btsy_j\right)^2\right), \forall j, i \quad (7)$$

$$CR_j \geq d_{ij} + m(1 - \alpha_{ij}), \forall j, i \quad (8)$$

$$\begin{cases} \alpha_{ij}, \beta_{ij} = \{0, 1\}; \\ WL_j, d_{ij}, x_i, y_i, K_{ij}, Btsx_j, Btsy_j \geq 0 \end{cases}, \forall j, i \quad (9)$$

Equation (1) is the objective function of the model. Equation (2) calculates the working load of each antenna, and equation (3) calculates the average working load of each antenna. The constraint (4) states that each user can receive service from only one antenna at

any time. Constraint (5) states that if any antenna does not cover a user, then no antenna should provide service to that user. Equation (6) is the service limit and guarantees that the workload of each antenna does not exceed its maximum service capacity. Constraint (7) calculates the Euclidean distance of user i from base station j . Equation (8) determines whether the user is located in the coverage radius of the station or not. Equation (9) defines the variables of the model.

4. Computational Results

The proposed mathematical model was implemented in Lingo software using a personal computer on three experimental input data sets. These sets were randomly created and simulated. After verifying the model's accuracy in solving the test samples, the proposed model was implemented in 12 examples with medium and large-scale pseudo-real data sets adopted from the East Azarbaijan Province Telecommunication Center. For security reasons, these data are provided to the

researcher in a general and limited manner without detailed details. In all these examples, despite the complexity and non-linearity of the proposed model, the Lingo could achieve the global optimal solution in eight examples by calculating the zero value for the objective function and calculating the local optimal solution in the other four examples, the network workload is balanced or near to the optimum between the stations.

4.1 Evaluating the Model

At this stage, three sample examples on a small scale were randomly designed, and by examining the results, the validity of the proposed model was confirmed. For example, in the first example, a sample mobile phone network shown in Figure (3) was considered with 20 users and ten base stations (antennas). The coordinates of users and other values of the model parameters were entered as model inputs in Excel.

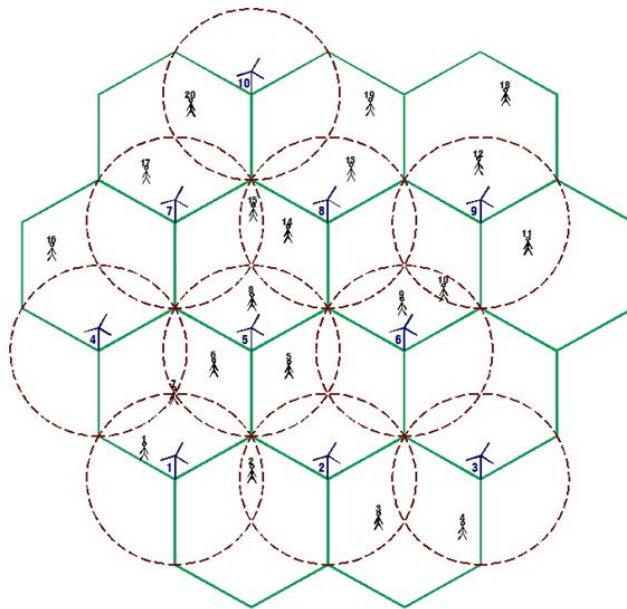


Fig. 3. The sample network of the first example

Examining the numerical results of the examples revealed that all the model's

constraints were satisfied. For example, users outside the coverage radius of any

station did not receive service. This procedure was repeated for examples 2 and 3 with different topologies and data, and the results were free of misallocating the value of the decision variables.

After validating the model, to evaluate the efficiency of the proposed model, 12

examples on a medium and large scale, whose data are similar to the actual environment conditions of segments of the mobile phone network of Tabriz city, are simulated. The details of the examples and calculation results are given in Table (1).

Table 1. Characteristics of medium and large-scale examples along with computation results

Example	I	J	K_{ij}	K_{max}	\bar{W}	Obj. function	Solution type	Exc. time (seconds)
1	20	10	50	500	100	0	Global Opt.	<1
2	50	10	50	500	250	0	Global Opt.	<1
3	80	10	50	500	400	0	Global Opt.	1
4	100	10	05	500	500	0	Global Opt.	<1
5	150	10	5	100	75	0	Global Opt.	1
6	180	20	5	500	332.5	5	Local Opt.	<1
7	250	10	10	500	250	0.113	Local Opt.	2
8	270	30	80	1000	6240	0	Global Opt.	<1
9	300	10	10	500	300	0.595	Local Opt.	1
10	400	10	5	5000	200	0.413	Local Opt.	3
11	1000	20	50	5000	5000	0	Global Opt.	2
12	1200	50	10	10000	600	0	Global Opt.	120

The calculation results show that, except for examples 6, 7, 9, and 10, the global optimum solution with a zero value for the objective function has been obtained, indicating a 100% workload balance. The resulting solution in the other four examples is of the local optimal type, and the value of the objective function, except for example 6, is minimal.

5. Conclusions

In this research, the load-balancing problem in the cellular mobile phone network was investigated, and a mathematical model was presented to optimize and balance the workload between the base stations (BTS antennas). The proposed model was non-linear and was simulated using Lingo ver. 18 software using a personal computer running Windows 10 operating system with an Intel Core i5-4.2 GHz processor. Three small-scale sample examples were

used to evaluate the model, whose numerical results could be checked and analyzed manually, and the model's validity was confirmed. In the next step, 12 examples on a medium and large scale were simulated with pseudo-real data Tabriz city network segments. In eight examples, the values obtained for the objective function were of the global optimal type with a value of zero (100% workload balance). In four other examples, local optimum solutions were obtained with minimal values close to zero. Despite the non-linearity and computational complexity of the proposed model, the time required to calculate the optimal solution has been reasonable and acceptable. As research problems, we can point out the difficulty/impossibility of obtaining the actual values of the parameters due to security reasons in providing the exact information required by the telecommunications company and

needing the necessary factors to compare with the previous approaches. To continue the research and carry out future work, the following are suggested:

- Solving the proposed model using meta-heuristic algorithms in very large-scale examples
- Use actual parameters in solving the model
- Considering more variety for the type of service requested by users

References

- [1] Sharma, A., Roy, A., Ghosal, S., Chaki R., and Bhattacharya, U. (2012). Load balancing in Cellular Network: A review, 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), Coimbatore, India, 1-5.
- [2] Wu, K. (2005). Load balancing of elastic data streams in cellular networks, Helsinki University of Technology.
- [3] Lim, JongBeom, and DaeWon Lee. (2020). A Load Balancing Algorithm for Mobile Devices in Edge Cloud Computing Environments, *Electronics*, 9(4): 686.
- [4] Saad, W., Shayea, I., Alhammadi, A., Sheikh, M., El-Saleh, A. (2023). Handover and load balancing self-optimization models in 5G mobile networks. *Engineering Science and Technology, and International Journal*. 42.
- [5] Casey, T., Veselinovic, N., Jantti. R. (2008). Base station-controlled load balancing with handovers in mobile WiMAX. in *Personal, Indoor, and Mobile Radio Communications, PIMRC 2008. IEEE 19th International Symposium on. 2008. IEEE*.
- [6] Nicholson, A.J., (2006). Improved access point selection. in *Proceedings of the 4th International Conference on Mobile Systems, applications and services. ACM*.
- [7] Lee, S., Han. Y. (2007). A novel inter-FA handover scheme for load balancing. IEEE 802.16 e system. in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*.
- [8] Moiseev, S., Fillin, S., Kondakov, M., Garmonov, A., Savinkov, A. (2006). Load-Balancing QoS-Guaranteed Handover in the IEEE 802.16e OFDMA Network. *Global Telecommunications Conference, GLOBECOM'06. IEEE*.
- [9] Bazzo, J., Cavalcante, A.M., de Sousa, M.J., Kuru, L., Moilanen, J. (2010). Load Balance for Multi-Layer Reuse Scenarios on Mobile WiMAX System. *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*.
- [10] Nazari, S., Beigy. H. (2010). A new distributed uplink packet scheduling algorithm in WiMAX networks. *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*.
- [11] Askarian, C., Beigy, H. A Survey for Load Balancing in Mobile WiMAX Networks. *Advanced Computing: An International Journal*, 2012. 3(2): 119-137.
- [12] Afzal, S., Kavitha, G. (2019). Load balancing in cloud computing – A hierarchical taxonomical classification. *Journal of Cloud Computing*, 22(8).
- [13] Duan, Z., Tian, C., Chang, N., Zhou, M., Yu, B., Wang, X., Guo, J., Wu, Y. (2022). A novel load balancing scheme for mobile edge computing. *Journal of Systems and Software*, 186, <https://doi.org/10.1016/j.jss.2021.111195>.