Research on Resource Allocation and Management of Mobile Edge Computing Network

Rui Zhang and Wenyu Shi
Anhui Xinhua University, Hefei, Anhui 230087, China
E-mail: nw2094@163.com

The popularity of mobile Internet makes the application of mobile terminals need more computing resources, and cloud computing enables mobile terminals to handle application tasks that need high computing resources under the premise of maintaining small specifications. However, it is difficult to obtain high-quality low latency services as the mobile Internet edge is far away from the cloud computing center; hence mobile edge computing (MEC) is proposed. This study introduced computing resource allocation methods based on power iteration and system utility, applied them to the mobile edge computing network, and carried out simulation experiments in MATLAB software. The experimental results showed that the network throughput and system utility under two resource allocation methods increased and the average transfer rate decreased with the increase of users in the mobile edge network; under the same number of access users, the edge network based on the system utility allocation method had higher throughput, average transfer rate and system utility.

Povzetek: Robno računalništvo (edge computing) omogoča boljše delovanje mrež, ker podatke v oblaku prestavi na rob mreže. Prispevek se ukvarja z analizo in izdelavo tovrstnih metod.

1 Introduction

The development of mobile Internet technology has further facilitated people's life, and the popularity of mobile terminals such as mobile phones and IPADS has greatly promoted mobile Internet technology [1]. The emergence of cloud computing [2] greatly reduces the requirements of the mobile terminal for computing and storage performance, thus reducing the manufacturing cost. However, even if the mobile Internet has a large coverage, there is only one data center that can provide cloud computing services. When data are transmitted in the nodes of the mobile Internet, there will inevitably be a delay. The further the distance with the data center is, the more serious the delay is. High delay will seriously affect the service reliability of various mobile applications. The emergence of mobile edge computing [3] solves the above problems. Compared with the cloud server in the data center, edge devices are closer to users and have shorter time delay. The combination of cloud server and edge device can provide more reliable services to users on the edge. Liu et al. [4] proposed a deep learning architecture based on the close connection network, transplanted it into the mobile edge algorithm, and found through the simulation that the algorithm had obvious overall efficiency advantage. Zhang et al. [5] proposed the weight based algorithm and mobile prediction based heuristic algorithm for users with certain and uncertain mobile tracks to reduce the network overhead caused by task migration and found through experiments that the two algorithms could effectively reduce the network overhead caused by task migration. In order to solve the joint optimization problem of task caching and offloading in

edge cloud computing, Hao et al. [6] proposed an efficient task caching and offloading algorithm based on the alternative iterative algorithm and found through the simulation that the algorithm had lower energy consumption. This study introduced a computing resource allocation method based on power iteration and a computer resource allocation method based on system utility, applied them to the mobile edge computing network, and carried out a simulation experiment on the two computing resource allocation methods in MATLAB software.

2 Mobile edge computing (MEC)

In recent years, the configuration performance of mobile terminals has been greatly improved with the progress of science and technology, mainly reflected in small volume and large amount of computing. However, in order to maintain its mobile convenience, the size of mobile terminal itself must be portable. Unless there is a breakthrough in the existing materials and technologies, the amount of computing must be limited and lower than that of large-scale computing equipment. In the face of today's huge mobile network applications, mobile terminals with limited computing and energy are gradually difficult to provide good services. The development of cloud computing technology has greatly liberated the computing burden of mobile terminals, but the increase of terminals which are accessed to mobile network for cloud computing has increased the network burden and delay. Cloud computing services usually give priority to meet the service request of mobile terminals near the data center;

therefore terminals on the edge of the network will have network delay.

In order to solve the problem that the service quality of edge network cloud computing reduces due to the excessive access of mobile terminals, mobile edge computing is proposed. The basic structure of MEC system includes the bottom structure, functional
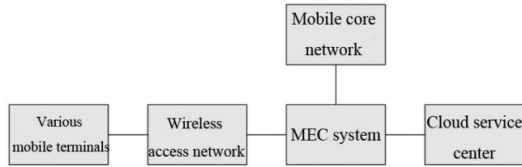

Figure 1: The basic architecture of MEC system.

components and application layer. The bottom structure mainly includes the virtual layer structure and hardware facilities which are used for generating virtual computing resources [7]; the functional components in MEC system play the role of interaction interface between internal and external data, assisting the system to access the mobile network without obstacles; the application layer is the application interface of the system, which is used for interacting information with users. After connecting MEC system to cloud computing service, its basic architecture is shown in Figure 1. The access location of MEC system is between the wireless access network and mobile core network. The wireless access network contains MEC servers, and they constitute the edge cloud and connect with various kinds of mobile terminals through the base station. The mobile core network is the center of the whole mobile Internet, which is used for realizing the large area transmission of communication data. The cloud service center is an important part of providing cloud computing.

The principle of MEC system improving the service quality of cloud computing can be generally described as lowering the task that needs to be executed in the cloud computing center to the edge server. In the traditional cloud computing data interaction, the request is first sent to the base station of wireless access network, and then the base station transmits the request to the core network for data protection. The request of any user follows the above process even though the request is the same. Once the number of access users increases, not only channel resources will be wasted, but also network congestion will cause delay [8]. After adding MEC, cloud computing resources are distributed to MEC server. When users repeat the same request, they can directly obtain resources from MEC server, which greatly improves the speed of service response.

3 Resource management based on power iteration

For the convenience of explanation, it is assumed that the mobile edge network has one base station (BS) and several mobile terminals (MT). MEC server is set in BS, which can directly carry out data interaction. Then all MTs

in the edge network can be expressed as $N = \{1,2,3,...,n\}$, all communication channel resources can be expressed as $M = \{1,2,3,...,m\}$, the tasks to be performed by the i-th MT can be expressed as $X_i = (D_i, C_i, T_i^{\max})$, where D_i, C_i and $T_i^{\max}$ are the size of calculation data, the required calculation power and maximum time delay respectively.

If the transmission power is used to control resource allocation, the resource allocation problem of the edge computing network can be expressed as:
objective function is:

$$\max\left(\sum_{i=1}^{N}\sum_{m=1}^{M} a_{i,m} B_m \log_2 \frac{\sigma_m^2 + I_{i,m} + p_{i,m} g_{i,m}}{\sigma_m^2 + I_{i,m}} - R_i^{\min}\right) \quad (1)$$

condition is:

$$\begin{cases} \sum_{m=1}^{M} B_m \log_2 \frac{\sigma_m^2 + I_{i,m} + p_{i,m} g_{i,m}}{\sigma_m^2 + I_{i,m}} \geq R_i^{\min} \\ \sum_{m=1}^{M} a_{i,m} = 1 \\ \sum_{m \in M} a_{i,m} p_{i,m} \in [0, p_i^{\max}] \\ SNR = \frac{\sum_{m \in M} a_{i,m} p_{i,m} g_{i,m}}{\sigma_m^2 + I_{i,m}} \geq \lambda_D \end{cases} \quad (2)$$

where $a_{i,m}$ indicates whether the i-th MT passes the m-th channel migration tasks or not, 1 for yes, and 0 for no; B_m indicates the bandwidth of the m-th channel, σ_m^2 stands for the Gaussian white noise on channel m [9], $I_{i,m}$ indicates the interference of other MT in channel m to the i-th MT, $p_{i,m}$ indicates the transmitting power of the i-th MT in channel m, $g_{i,m}$ indicates the channel gain of the i-th MT in channel m, $R_i^{\min}$ indicates the minimum transmission rate of the i-th MT in channel m to ensure the transmission quality, SNR indicates the signal to noise ratio [10], and λ_D indicates the threshold of the signal to noise ratio.

It is seen from equation (1) and (2) that the corresponding allocation scheme is optimal when $p_{i,m}$ is optimal, then the solution step of optimal $p_{i,m}$ is:

① Related iteration update parameters including number of iterations t and Lagrange multiplier λ and μ are initialized.

② Let $t = t+1$, indicating one time of iteration, and then the Lagrange multiplier is updated according to the following formula:

$$\begin{cases} \lambda_m(t+1) = \max\left[\lambda_m(t) + \alpha\left(\sum_{m \in M} a_{i,m} p_{i,m} - p_i^{\max}\right), 0\right] \\ \mu_m(t+1) = \max\left[\mu_m(t) + \beta\left(\lambda_D - \dfrac{\sum_{m \in M} a_{i,m} p_{i,m} g_{i,m}}{\sigma_m^2 + I_{i,m}}\right), 0\right] \end{cases} \quad (3)$$

③ Transmission power $p_{i,m}$ is calculated according to the Lagrange multiplier [11] obtained from the previous update, and the formula is:

$$p_{i,m} = \frac{B_m}{[\mu g_{i,m} - \lambda(I_{i,m} + \sigma_m^2)]\ln 2} - \frac{I_{i,m} + \sigma_m^2}{g_{i,m}} \quad (4)$$

④ $p_{i,m}$ which is calculated after updating and before updating is compared. If the difference between them is smaller than the set threshold value, iteration updating stops, otherwise step ② and ③ repeat. After several times of iterations, optimal $p_{i,m}$ is obtained.

4 Resource management based on system utility

First of all, the model structure of MEC system is, same as above, set as multi-user MT-single BS. The resource allocation of the edge computing network is designed based on the utility function of the system. Then the allocation can be expressed as follows:
objective function is:

$$\max\left(\sum_{i=1}^{N} \rho_i s_i \left(\gamma_i^T \frac{T_i^l - T_i^r}{T_i^l} + \gamma_i^E \frac{E_i^l - E_i^r}{E_i^l}\right)\right) ; (5)$$

condition is:

$$\begin{cases} s_i = \{0,1\} \\ p_i \in (0, p_{\max}] \\ f_i > 0 \\ \sum_{i \in S} f_i \le f_{\max} \\ \sum_{i \in N} s_i \le M \end{cases} \quad (6)$$

where ρ_i indicates the priority of the i-th MT receiving MEC service, [0,1] , s_i indicates whether terminal equipment I selects task immigration or not, 1 for migrating to MEC server and 0 for processing tasks locally, γ_i^T and γ_i^E indicate the preference of equipment I for improving efficiency and reducing energy consumption respectively [12], i.e., the intention of users to efficiently solve problems and save energy, [0,1], T_i^l and T_i^r indicate the time delay of processing tasks locally and processing tasks by migrating to MEC server respectively, the former depends on the CPU performance of equipment and the latter depends on computation resource f_i allocated by MEC server and the time delay of information transmission, E_i^l and E_i^r indicate the

energy consumption of processing tasks locally and processing tasks by migrating to MEC server respectively, the former depends on the power and processing time of equipment and the latter depends on the energy during data transmission, p_i indicates the transmitting power of equipment i, which cannot exceed its maximum transmitting power, S indicates a set of equipment participating in task migration.

Steps to solve objective function (5) are as follows.

① First, the optimal transmission power of the equipment in each mobile edge network was calculated using the dichotomy method [13].

② Then whether task migration is required for each equipment is determined. If necessary, a task migration request is issued; if not, a NULL message is issued.

③ After receiving the message from the device, the equipment is classified according to the following formula:

$$\begin{cases} S_l = \left\{ i \middle| \rho_i(\gamma_i^T + \gamma_i^E) - \Delta(i) \le 0 \right\} \\ S_r = \left\{ i \middle| \rho_i(\gamma_i^T + \gamma_i^E) - (\Delta(i) + \Delta(i|N - S_l - \{i\})) \ge 0 \right\} \\ S_s = N - S_l - S_r \\ \Delta(i) = \dfrac{\eta_i + \gamma_i p_i}{\log_2(1 + a_i p_i)} + \dfrac{\tau_i F_i^l}{f_{\max}} \\ \Delta(i|N - S_l - \{i\}) = \dfrac{2\sqrt{\tau_i F_i^l} \sum_{j \in N - S_l - \{i\}} \sqrt{\tau_j F_j^l}}{f_{\max}} \\ \eta_i = \dfrac{\rho_i \gamma_i^T D_i}{B_m T_i^l} \\ \gamma_i = \dfrac{\rho_i \gamma_i^E D_i}{B_m T_i^l \varsigma} \\ \tau_i = \rho_i \gamma_i^T \end{cases} \quad , (7)$$

where S_l , S_r and S_s are the equipment set of locally processed tasks, the equipment set of task migration and the equipment set to be allocated respectively, $\Delta(i), \Delta(i|N - S_l - \{i\})$ are the intermediate quantity for calculating the system marginal utility value in the original migration equipment set after adding equipment i, η_i, γ_i and τ_i are intermediate variables for calculating $\Delta(i), \Delta(i|N - S_l - \{i\})$, F_i^l is the working frequency of mobile equipment CPU, a_i is the ratio of channel gain to channel noise power during the transmission of equipment i, and ς is the working efficiency of transmission power amplifier.

④ Whether the number of equipment in S_r exceeds the total number of channels (K) in the edge network is determined. If it exceeds, then a equipment with the smallest system utility is selected from S_r and moved to S_l . The cycle stops until the number of equipment in S_r does not exceed K . Then S_r is output, and task migration and allocation of computation resources were performed according to S_r .

⑤ If the number of equipment in S_r is smaller than the total number of channels in the edge network, then the equipment with the largest system utility is selected from S_s and added to S_r. S_r does not shrink after adding the new equipment, i.e., the system utility of the set does not decrease. The step repeats until the number of equipment in S_r reaches the largest number of channels in the network or S_r does not extend. After the cycle stops, S_r is output, and the task migration and allocation of computation resources were performed according to S_r.

5   Simulation experiment

5.1   Experimental environment

In this study, the two edge computing network resource allocation methods mentioned above were simulated and analyzed by MATLAB software [14]. The experiment was carried out in a laboratory service. The server configuration included Windows7 system, i7 processor and 16 G memory.

5.2   Experiment setup

In this study, a mobile edge computing network area was established using MATLAB, and the basic parameters are shown in Table 1. In the simulated mobile edge network, the effective coverage of the network was 500 m. In the network, there was a base station and a MEC server. The base station and MEC server were connected. The total channel bandwidth provided by the base station of the edge network was 30 MHz, and its subchannel bandwidth was 1.5 MHz. The maximum number of subchannel that could be provided was 20. The maximum transmission power of MT held by users in the edge network was set as 25 dbm. The computing performance was 1.3 GHz. The user's preference for improving task processing efficiency and equipment energy saving was randomly distributed between 0.3 and 0.8. For the convenience of simulation calculation, the task data size of the MT which was needed to be processed by user was set as 450 kB, and the computing power required to process the task was 1200 Mcycles. The MEC server used for remote processing of tasks had a computing performance (computing resources) of 25 GHz.

The simulation network was was set as described above. Then grouping experiments were carried out according to the number of users in the edge network. There were seven groups in total, 5 users in the 1st group, 10 users in the 2nd group, 15 users in the 3rd group, 20 users in the 4th group, 25 users in the 5th group, 30 users in the 6th group and 35 users in the 7th group. The above two resource management methods are used in each group of simulation experiments. The indicators of resource management method are network throughput, average transmission rate and system utility. Network throughput [15] refers to the number of successfully transmitted data

Parameter	Radius of edge network area	Total channel bandwidth	Subchannel bandwidth
Numerical value	500 m	30 MHz	1.5 MHz
Parameter	Channel gain	Channel interference noise	Maximum transmitting power of MT
Numerical value	$128.1+37.5\lg(r)$	-175 dBm/Hz	25 dBm
Parameter	Task size	Resources required to process tasks	CPU performance of MT
Numerical value	450 kB	1200 MCycles	1.3 GHz
Parameter	User preferences (γ_i^T, γ_i^E)	MEC server performance	
Numerical value	0.3~0.8	25 GHz	

Table 1: Basic parameters of simulated mobile edge network.

in unit time, while System utility is the effective utilization of network computing resources.

5.3   Experimental results

Changes of throughput in the edge computing network with the increase of the number of access network users under the two resource management methods are shown in Figure 2. It was seen intuitively from Figure 2 that the network throughput under the two resource management methods was on the rise with the increase of the number of access users in the edge network, and the rise amplitude reduced when the number of users was larger than 20. Generally speaking, the throughput of the edge network based on power iteration was lower than that of the system utility based network under the same number of access
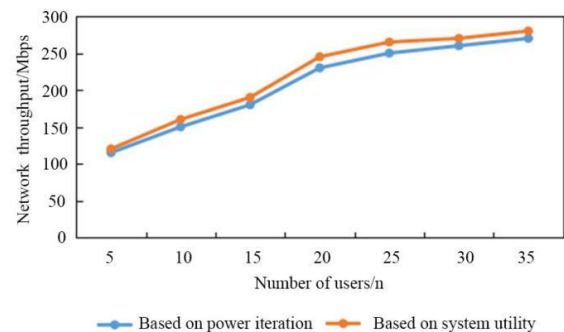
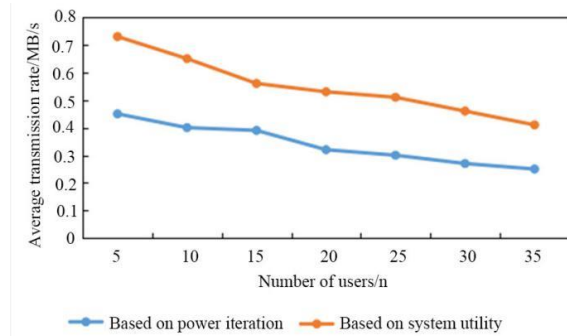Figure 2: Changes of network throughput with the number of users under the two resource management methods.

Figure 3: Changes of average transmission rate with number of users under two resource management methods.

Figure 4: The system utility of the edge network changes with the number of users under two resource management methods.

users, and the network throughput performance of the system utility based network was better.

Changes of the average transmission rate of the edge network under the two resource allocation methods with the increase of the number of access users are as shown in Figure 3. In the edge network using the power iterative based allocation method, the average transmission rate was 0.45 MB/s when the number of users was 5, 0.4 MB/s when the number of users was 10, 0.39 mb/s when the number of users was 15, 0.32 MB/s when the number of users was 20, 0.30 MB/s when the number of users was 25, 0.27 MB/s when the number of users was 30, and 0.25 MB/s when the number of users was 35. In the edge network using the system utility based allocation method, the average transmission rate was 0.73 mb/s for 5 users, 0.65 mb/s for 10 users, 0.56 mb/s for 15 users, 0.53 mb/s for 20 users, 0.51 mb/s for 25 users, 0.46 mb/s for 30 users, and 0.41 mb/s for 35 users. It was seen from Figure 3 that the average transmission rate of the edge network under the two resource allocation methods decreased with the increase of the access users in the edge network. The reason for the decrease was that the increase of users occupied more subchannels, and moreover the interference between the transmitted signals strengthened. In addition, under the same number of access users, the average transmission rate of the edge network under the system utility based allocation method was higher.

Changes of the system utility of the edge network under the power iteration and system utility with the increase of the number of access users are shown in Figure 4. As shown in Figure 4, when the number of access users was 5, the network system utility of the former was 1.9, and the system utility of the latter was 2.0; when the number of access users was 10, the system utility of the former was 2.8, and the system utility of the latter was 3.0; when the number of access users was 15, the system utility of the former was 3.7, and the system utility of the latter was 4.0; when the number of access users was 20, the system utility of the former was 4.5, and the system utility of the latter was 4.9; when the number of access users was 25, the system utility of the former was 5.0, and the system utility of the latter was 5.5; when the number of access users was 30, the system utility of the former was 5.6, and the system utility of the latter was 6.0; when the number of access users was 35, the system utility of the former was 5.0, and the system utility of the latter was 6.3. It was seen
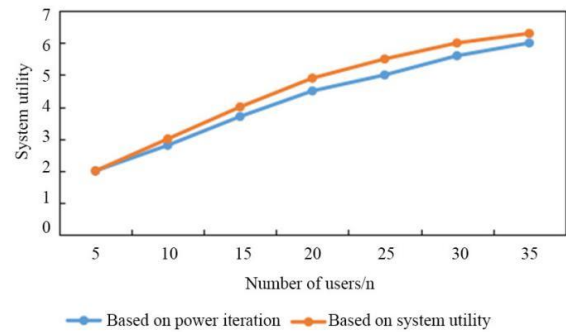
from Figure 4 that the system utility of the edge network under the two resource allocation methods increased with the increase of the number of access users in the edge network, and the increase amplitude decreased gradually; under the same number of access users, the edge network under the system utility based allocation method had higher system utility.

6 Conclusion

This study introduced computing resource allocation methods based on power iteration and system utility, applied them to the mobile edge computing network, and carried out the simulation experiment on the two methods in MATLAB software. The experimental results are as follows: (1) with the increase of users in the edge network, the network throughput under the two computing resource allocation methods showed an increasing tendency, and the edge network under the system utility based allocation method had higher throughput; (2) the average transmission rate of the edge network decreased with the increase of the number of access users, and the edge network under the system utility based allocation method had a higher average transmission rate; (3) with the increase of the number of users in the edge network, the system utility of the edge network under the two methods of computing resource allocation was on the rise, and the edge network under the system utility based allocation method had higher system utility.

7 References

[1] Al-Shuwaili A, Simeone O (2016). Energy-Efficient Resource Allocation for Mobile Edge Computing-Based Augmented Reality Applications. *IEEE Wireless Communication Letters*, PP(99). https://doi.org/10.1109/LWC.2017.2696539

[2] Ahmed E, Rehmani MH (2016). Mobile Edge Computing: Opportunities, solutions, and challenges. *Future Generation Computer Systems*, 70. https://doi.org/10.1016/j.future.2016.09.015

[3] Paymard P, Mokari N (2019). Resource allocation in PD-NOMA–based mobile edge computing system: Multiuser and multitask priority. *Transactions on Emerging Telecommunications Technologies*, (1), pp. e3631. https://doi.org/10.1002/ett.3631

[4]   Liu ZK, Yang XQ, Shen JX (2019). Optimization of multitask parallel mobile edge computing strategy based on deep learning architecture. *Design Automation for Embedded Systems*, (4). https://doi.org/10.1007/s10617-019-09222-5

[5]   Zhang F, Liu G, Zhao B, Fu X, Yahyapour R (2018). Reducing the network overhead of user mobility-induced virtual machine migration in mobile edge computing. *Software Practice and Experience*, (3). https://doi.org/10.1002/spe.2642

[6]   Hao Y, Chen M, Hu L, Hossain MS, Ghoneim A (2018). Energy Efficient Task Caching and Offloading for Mobile Edge Computing. *IEEE Access*, 6(99), pp. 11365-11373. https://doi.org/10.1109/ACCESS.2018.2805798

[7]   Pham QV, Le LB, Chung SH (2019). Mobile Edge Computing with Wireless Backhaul: Joint Task Offloading and Resource Allocation. *IEEE Access*, PP(99), pp. 1-1. https://doi.org/10.1109/access.2018.2883692

[8]   Ma LL, Yi SH, Carter N, Li Q (2018). Efficient Live Migration of Edge Services Leveraging Container Layered Storage. *IEEE Transactions on Mobile Computing*, PP(99), pp. 1-1. https://doi.org/10.1109/TMC.2018.2871842

[9]   Farris I, Taleb T, Flinck H (2018). Providing ultra-short latency to user-centric 5G applications at the mobile network edge. *Transactions on Emerging Telecommunications Technologies*, 29. https://doi.org/10.1002/ett.3169

[10] Shahzadi S, Iqbal M, Dagiuklas T, Qayyum ZU (2017). Multi-Access Edge Computing: Open issues, Challenges and Future Perspective. *Journal of Cloud Computing Advances Systems & Applications*, 6(1), pp. 30. https://doi.org/10.1186/s13677-017-0097-9

[11] An N, Yoon S, Ha T, Kim Y, Lim H (2018). Seamless Virtualized Controller Migration for Drone Applications. *IEEE Internet Computing*, PP(99), pp. 1-1. https://doi.org/10.1109/MIC.2018.2884670

[12] Zeng DZ, Gu L, Pan SL, Cai JJ, Guo S (2019). Resource Management at the Network Edge: A Deep Reinforcement Learning Approach. *IEEE Network*, 33(3), pp. 26-33. https://doi.org/10.1109/MNET.2019.1800386

[13] Wang Z, Zhao ZW, Min GY (2018). User mobility aware task assignment for Mobile Edge Computing. *Future Generation Computer Systems*, 85. https://doi.org/10.1016/j.future.2018.02.014

[14] Fang WW, Ding S, Li YY (2019). OKRA: optimal task and resource allocation for energy minimization in mobile edge computing systems. *Wireless Networks*, 25(5). https://doi.org/10.1007/s11276-019-02000-y

[15] Yang X, Chen ZY, Li KK (2018). Communication-Constrained Mobile Edge Computing Systems for Wireless Virtual Reality: Scheduling and Tradeoff. *IEEE Access*, 6, pp. 16665-16677. https://doi.org/10.1109/ACCESS.2018.2817288