A Service Migration Strategy Based on Multiple Attribute Decision in Mobile Edge Computing

Dan Zhao
State Key Laboratory of Network and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
e-mail: zhaodandan@bupt.edu.cn

Yuehui Jin
State Key Laboratory of Network and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
e-mail: yhjin@bupt.edu.cn

Tan Yang
State Key Laboratory of Network and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
e-mail: tyang@bupt.edu.cn

Yue Xu
State Key Laboratory of Network and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
e-mail: xuyuejiay@163.com

***Abstract*—With the shortcoming of long network delay and long response time, traditional mobile cloud computing can't meet the need of delay sensitive business, such as Virtual Reality and Augmented Reality. Mobile edge computing has emerged and provided a way to solve the problems. During the movement of the user, if an optimal data center is detected, the virtual machine migrates from the current data center to the optimal one. It is important to note that virtual machine migration increases costs while enhancing the user's experience. How to balance costs and benefits is what we need to focus on. The core issue of the migration strategy is whether, when, and where to migrate. In this paper, we propose an edge computing migration strategy based on multiple attribute decision making to deal with the issue. Furthermore, we demonstrate the effectiveness of the migration strategy by simulation.**

Keywords-mobile edge computing; multiple attribute decision making; virtual machine migration; mobility

I. INTRODUCTION

In recent years, with the rapid development of mobile Internet, mobile data traffic is at an alarming rate of growth. According to the Cisco VNI [1] report, the total amount of mobile data will grow sevenfold between 2016 and 2021. It is clear that super data volume will increase the network load, especially the core network. In response to growing mobile traffic, mobile operators realized the advantage of decentralization of their core network and begun to design solution based on traffic offloading. As a result, mobile edge computing (MEC) came into being.

Mobile edge cloud in MEC deploys calculation, storage, processing and other functions on the side of base station without uploading task to the cloud data center. It has a number of benefits, such as reducing bandwidth consumption, reducing latency by locating resources at nearby users and improving transmission efficiency. Due to the mobility of users, the data center that is providing service may not be the optimal one. In order to keep the quality of experience, virtual machine that containing context information needs to be migrated from one data center to another.

In this paper, we design a migration strategy based on MADA (multi-attribute decision making) in MEC. In our approach, the virtual machine can be migrated among multiple servers according to the network status, user location, task attributes and server load. At the same time, we use MATLAB to simulate the network and user's movement to the greatest extent. The simulation results show that our migration strategy is beneficial.

The remainder of this paper is organized as follows: Section II introduces related researches. The migration strategy based on multiple attribute decision making is conducted in Section III Sections IV presents simulation experiments and results discussion. The paper concludes in Section V.

II. RELATED WORK

There are some research work focusing on MEC, such as concepts, scenarios, benefits, and challenges. In [2], Hu *et al.* introduce the concept of mobile edge computing and the related market drivers, depicting the advanced blueprint of edge calculation. Some real-time computing scenarios and challenges for edge computing discussed in [3]. Cao *et al.* apply fog calculation to medical health monitoring, which helps to reduce latency and improve stability in [4]. Moreover, Tarik provides a framework of the edge calculation designed to migrate users' ongoing business from the original data center to another in [5]. In [6], Salman *et al.* combine fog, SDN and NFV three concepts, showing how to achieve MEC in mobile networks, and how to promote the development of Internet of Things. The author introduces the details of the virtual machine migration technology, providing a heuristic idea of migration to reduce power consumption and achieve load balancing in [7]. The author presented an architecture to validate and evaluate fog calculation and demonstrated the likelihood of user telemedicine equipment in [8].

In addition, several papers published in recent years describe computational offloading associated with MEC. In [9], Ksentini *et al.* describe a one-dimensional MDP with a specific cost function based on distance. Wang *et al.* proposed a new effective solution to speed up the calculation speed in [10]. These papers only consider the users location, didn't consider the load of edge server, bandwidth and other context conditions.

III. MADA-BASED MIGRATION STRATEGY

A. MEC Network Architecture

MEC refers to providing services near mobile terminals. Intuitively, those monstrous data center compressed and installed to the wireless base station. It is estimated that deploying application servers at the edge of wireless networks can save up to 35% of wired bandwidth usage and reduce network latency by 50% for users.

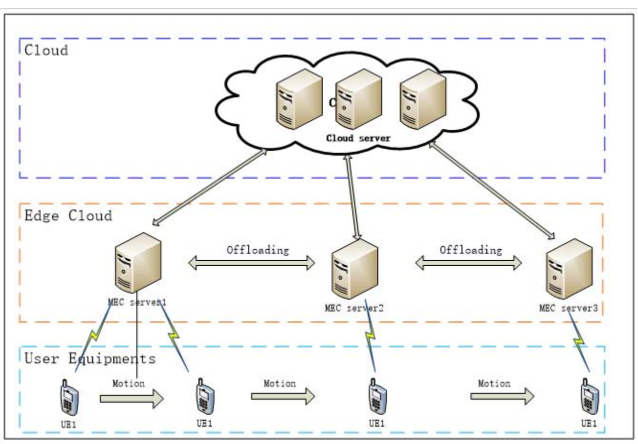


Figure 1. MEC network architecture.

The whole network is divided into three levels, namely, user equipment, edge data center and cloud data center. As shown in Fig. 1, the user obtains the service from the current location-optimal edge data center, and when the user moves, it may be necessary to migrate the VM to another edge data center. The purpose of the migration is to enable the user to always get the service from the optical data center (Geographically the nearest or the fastest processing).

B. Key Algorithms in Modeling

Definition 1 (Energy consumption model). This refers to the energy consumption of the server during the experiment. Eb is baseline energy consumption of server, and the server runs n virtual machines, the energy consumption of each virtual machine is express by Cvm, then the total energy consumption of the server is Esi. The formula is as follows:

$$Esi = Eb + n * Cvm \tag{1}$$

Definition 2 (Cost model). The cost here includes migration costs and communication costs. The cost of migration Cm is related to the distance of migration dsi,sj and unit migration cost Wm; and communication energy consumption Cc is related to the unit cost of communication Wc and the distance of communication dsi,sj. The specific formula is as follows:

$$Cm(si,sj) = Wm * dsi,sj \tag{2}$$

$$Cc(si,sj) = Wc * dsi,sj \tag{3}$$

Definition 3 (Latency Model). The delay here refers to the entire process time when the user issues a request to receive the server data, including the communication delay Tp and the computing latency Tc.

$$T = TP + TC \tag{4}$$

The communication delay is the sum of the send delay Ts, propagation delay Tt, and router queuing delay Trq. In addition, data is the size of the packet sent and BW is the rate of transmission of data packets. The transmission time is related to the transmission distance and the speed. In the formula (8), d is the distance of transmission and v is the electromagnetic wave transmission rate.

$$TP = Ts + Tt + Trq \tag{5}$$

$$Ts = Data / BW \tag{6}$$

$$Tt = d / v \tag{7}$$

C. MADA Algorithm

1) Build decision matrix

There are five evaluation indicators need to be considered: bandwidth, computing ability, latency, costs and energy consumption.

The decision matrix is constructed based on the candidate set and the evaluation indicators as follows:

$$X = \begin{array}{c} \\ X1 \\ X2 \\ X3 \end{array} \begin{array}{c} \begin{array}{ccccc} b1 & b2 & b3 & b4 & b5 \end{array} \\ \begin{bmatrix} x11 & x21 & x31 & x41 & x51 \\ x12 & x22 & x32 & x42 & x52 \\ x13 & x23 & x33 & x43 & x53 \end{bmatrix} \end{array} \tag{8}$$

Among them, X1, X2, X3 represent three kinds of migration plans. b1, b2, b3, b4, b5 represent bandwidth, computing ability, latency, costs and energy consumption, respectively.

2) Standardize decision matrix

In order to eliminate the differences between the evaluation indicators, we often need to standardize them.

Decision indicators are divided into three types: beneficial indicators, cost indicators and neutral indicators.

In the decision matrix, for beneficial indicators:

$$x_j^* = \max_{1\le i\le m} x_{ij} \neq 0 \tag{9}$$

$$y_{ij} = \frac{x_{ij}}{x_j^*} \quad (1 \le i \le m) \tag{10}$$

For negative indicators:

$$x_j^* = \min_{1\le i\le m} x_{ij} \tag{11}$$

$$y_{ij} = \frac{x_j^*}{x_{ij}} \quad (1 \le i \le m) \tag{12}$$

3) Determine the weight of the index

In information theory, entropy is not only a measure of uncertainty, but also used to indicate how much information is included in the data.

The normalized matrix Y is normalized according to the following formula:

$$p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}} \quad \begin{pmatrix} i = 1, 2\cdots, m \\ j = 1, 2\cdots, n \end{pmatrix} \tag{13}$$

Calculate the entropy of the j index:

$$e_j = -k\sum_{i=1}^{m} p_{ij} \ln p_{ij} \quad (1 \le j \le n) \tag{14}$$

Calculate the difference coefficient of the j index:

$$g_j = 1 - e_j \quad (1 \le j \le n) \tag{15}$$

Calculate the weight of the j indicator:

$$w_j = \frac{g_j}{\sum_{j=1}^{n} g_j} \quad (1 \le j \le n) \tag{16}$$

4) TOPSIS

TOPSIS sorts the schemes according to the distance between the evaluation object and the idealized goal. The TOPSIS process is as follows. Calculate the ideal solution and the negative ideal solution:

$$z_{ij} = y_{ij} * w_j \tag{17}$$

$$z_j^* = \begin{cases} \max z_{ij} & \textit{beneficial indicators} \\ \min z_{ij} & \textit{negative indicators} \end{cases} \tag{18}$$

$$z_j^0 = \begin{cases} \min z_{ij} & \textit{beneficial indicators} \\ \max z_{ij} & \textit{negative indicators} \end{cases} \tag{19}$$

a. Calculate the distance from the candidate network to the ideal solution and the negative solution:

$$d_i^* = \sqrt{\sum_{j=1}^{n} (z_{ij} - zj^*)^2} \tag{20}$$

$$d_i^0 = \sqrt{\sum_{j=1}^{n} (z_{ij} - z_j^0)^2} \tag{21}$$

b. Calculate the proximity index of each candidate server:

$$C_i^* = d_i^0 \big/ (d_i^0 + d_i^*) \tag{22}$$

c. Sort by the Ci*, select the server with the maximum proximity as the target.

D. MADA Algorithm

Fig. 2 shows VM migration and interoperability between different edge servers. First, the mobile terminal is connected to one of the data centers. Next, the user's move triggers the migration decision. MADA is used to determine migration. If the service needs to be migrated, the virtual machine is performed. Finally, recovery service is made in the new data center.

In the process of moving, users exceed the pre-set threshold value and start the migration strategy based on multi-attribute decision making. The multi-attribute decision matrix is constructed and sorted through the above methods. The optimal MEC server is selected for migration.

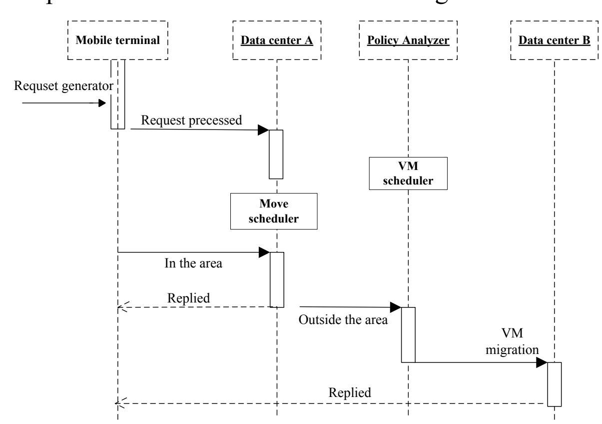


Figure 2. Sequence diagram of VM-migration in mobile edge cloud environments.

IV. Simulation and Evaluation

A. Simulation Environment

The simulation experiment is based on MATLAB. As shown in Fig. 3, the server distribution obeys the Poisson's point distribution, and users are connected to 8 data centers with different performances. In addition, the black circle in the Fig. 3 represents the best service area for each data center. Migration strategy is initiated whenever the connected user exceeds the service scope. The user moves counterclockwise along the fixed track, where the solid dot is the starting point of the mobile terminal.

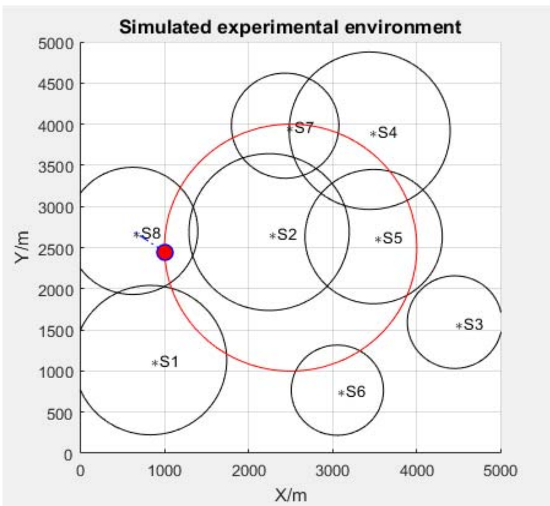


Figure 3. Simulation scenario.

B. Main Parameters

The simulation parameters has a direct effect on the experimental results. In this paper, we set the parameters as realistic as possible. We set the queue delay of router as 1ms in the simulation. The best service area of each data center is subject to Gaussian distribution. The experimental parameters are shown in Table I.

TABLE I. Simulation Parameters

Parameters	Meaning
1M	Task upload volume
800MIPS	Task calculation
[2,10]	Energy consumption of server
2M	Task migration
5M	Task download
1ms	Router queuing delay
[60000,100000]	Computing ability
[4,6]*10(-4)	the unit cost of the task

C. Simulation Results and Analysis

The experiments were carried out under three different load conditions: heavy load, normal, light load. Respectively, compared the performance of non-migration strategy, MDP (Markov decision process) and MADA-based migration strategy. We compare the task completion time, calculation cost, energy consumption, and bandwidth usage under different policies. Tasks are divided into computational intensive and transport intensive.

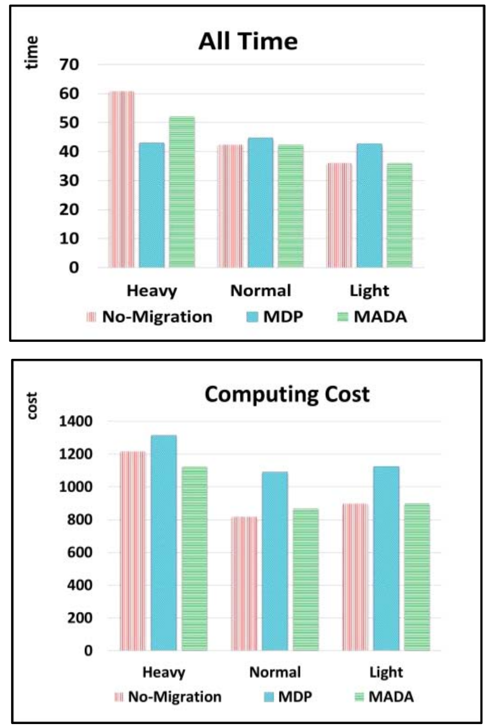
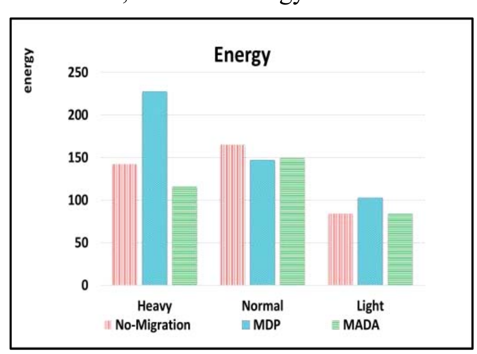


Figure 4. Time and cost of transport intensive tasks under different strategies.

As shown in Fig. 4 and Fig. 5, for transport-intensive tasks, migration is beneficial, no matter what kind of migration strategy. In addition, MADA-based migration strategy is more advantageous than MDP in general. This is because the MADA strategy makes decisions based on a variety of factors, while MDP only makes decisions based on distance. Therefore, MADA strategy is more effective.



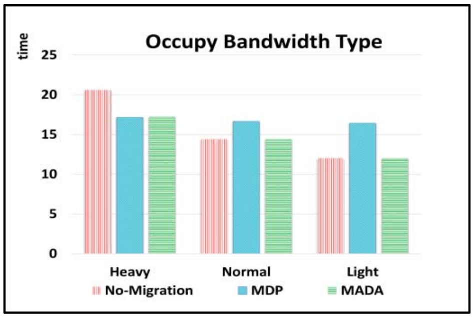


Figure 5. Energy consumption and Occupied bandwidth time of transport intensive tasks under different strategies .

As shown in Fig. 6 and Fig. 7, for computationally intensive tasks, computing power is decisive. As a result, the task is better in the normal load range of the server compared to the case where the server is overloaded. From the simulation results, MADA strategy slightly better than MDP.

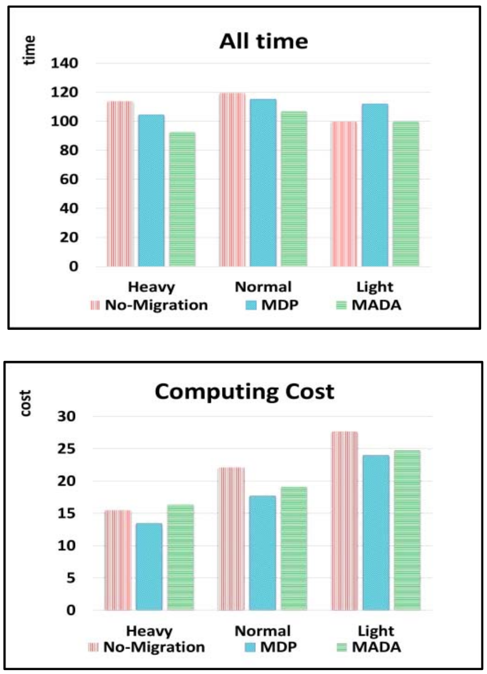
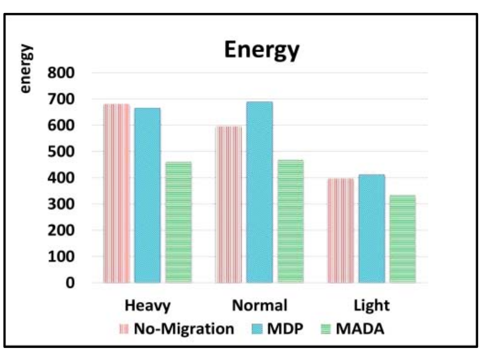


Figure 6. Time and cost of computing intensive tasks under different strategies.



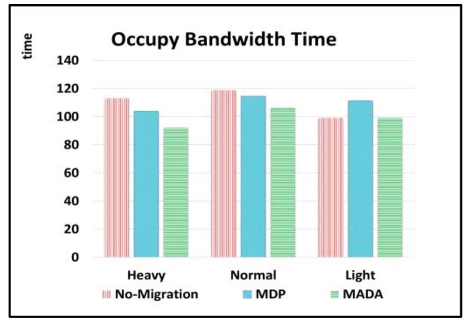


Figure 7. Energy consumption and Occupied bandwidth time of computing intensive tasks under different strategies.

V. Conclusion

In this paper, we propose a mobile edge computing migration strategy based on multiple attribute decision making. The strategy can solve the problem whether, when and where to migrate efficiently. Simulation results show that the mobile edge computing migration strategy can choose the appropriate server, significantly reduce the user's response time and improve the quality of user’s experience. More simulation will be done in the future, and we will focus on optimizing the migration strategy that can be applied to real networks which are more complex.

Acknowledgment

This paper is supported by State Key Laboratory of Network and Switching Technology.

References

[1] Index C V N. Forecast and Methodology, “2015-2020 White Paper[R]. Technical Report,” Cisco, 2016.

[2] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile Edge Computing A key technology towards 5G," ETSI White Paper, 2015.

[3] A. Ahmed and E. Ahmed, "A survey on mobile edge computing," ACM SIGCOMM Computer Communication Review, vol. 8, pp. 1-18, 2016.

[4] Y. Cao, P. Hou, D. Brown, J. Wang, and S. Chen, "Distributed Analytics and Edge Intelligence:Pervasive Health Monitoring at the Era of Fog Computing," The Workshop on Mobile Big Data ACM, pp. 43-48, 2015.

[5] T. Taleb and A. Ksentini., "Follow Me Cloud: Interworking Federated Clouds and Distributed Mobile Networks," Communications Magazine IEEE, pp. 12-19, 2013.

[6] O. Salman, I. Elhajj, A. Kayssi, and A. Chehab, "Edge computing enabling the Internet of Things," Internet of Things IEEE, pp. 603-60.

[7] M. Mishra, A. Das, P. Kulkarni, and A. Sahoo, "Dynamic resource management using virtual machine migrations," Communications Magazine IEEE, pp. 34-40, 2012.

[8] H. Dubey, J. Yang, N. Constant, A. M. Amiri, Q. Yang, and K. Makodiya, "Fog Data: Enhancing Telehealth Big Data Through Fog Computing," the ASE BigData & SocialInformatics 2015, 2015.

[9] A. Ksentini, T. Taleb and M. Chen, "A Markov Decision Process-based service migration procedure for follow me cloud," Communications (ICC), pp. 1350-1354, 2014.

[10] S. Wang, R. Urgaonkar, T. He, M. Zafer, and K. Chan, "Mobility-Induced Service Migration in Mobile Micro-clouds," Military Communications Conference (MILCOM), pp. 835-840, 2014.