Video Pre-Caching Joint Hand-Off and Content Delivery in Multi-Access Edge Computing Based EONs

Yutong CHAI[1], Shan YIN, Lihao LIU, Liyou JIANG, Shanguo HUANG
State Key Lab of Information Photonics and Optical Communications, China

Abstract. Multi-access Edge Computing (MEC) performs as a feasible solution when it comes to content delivery, for it can bring contents much closer to users. However, the hand-off (HO) and latency that occur in user movement reduce the users' quality of service. In this work, we consider the problem of high mobility handoff and content delivery of video streaming in the MEC based EONs. We propose a video pre-caching algorithm considering handoff and content delivery. The algorithm firstly selects the content delivery method and chunks the video accordingly using a preset threshold. Secondly, the algorithm chooses the shortest transmission path and calculates the latency time using Dijkstra method. Simulation results show that our algorithm significantly reduces the latency time and balances the server load compared to the other two baselines.

Keywords. MEC, handoff, content delivery, metro optical network

1. Introduction

The user case of video streaming required by vehicular users in the smart city falls in the category of ultra-reliable and low latency communication (URLLC) of 5G service type, making existing networks face new challenges. As a prospective computing paradigm, Multi-access Edge Computing (MEC) greatly decreases the waiting time between request and response through the way of pushing computational resources and storage resources to the network edges (e.g., Cellular Base Stations (BSs)). Furthermore, the 5G Automotive Association (5GAA) defined the concept of Cooperative Intelligent Transportation Systems (C-ITS) in 2017, stating that MEC would be the enabling technology for V2X (Vehicle to Everything) communications [1]. To make MEC construction more efficient, a heterogeneous network called Macro-Small cell network has become an emerging network deployment method.

However, handoffs (HOs) of vehicular users between BSs and even MECs happen frequently due to users' high mobility, decreasing the constant QoE of content [2]. Concurrently, how to minimize the latency time to make the content delivery process more efficient is another important issue. A few related solutions are aiming at solving the HOs problems in literature, such as edge caching based on Markov prediction [3], mobility prediction based RAN cache Hos [4]. Also, several studies are exploring the

---

[1] Corresponding author: Yutong Chai; Email: chaiyutong@bupt.edu.cn

content delivery strategy. In literature [5], a distributed algorithm based on matching theory to minimize latency in the MEC-caching service network was provided. In literature [6], an MEC-based architecture was proposed to improve the performance of adaptive content delivery. This research led us to hypothesize that considering the handoff problem in advance can reduce the latency time of content delivery, which has not been studied in the previous research.

In this paper, we especially use video streaming as the required content of vehicular users in the smart city with MECs, which is a computation-intensive and caching-intensive service. Since the cost of storage resources is lower than the cost of bandwidth resources, distributing content in Multi-access Edge Computing is used as an effective method, allowing users to access the chunk of large-size video streaming from multiple MEC servers in advance. In addition, the main features of Elastic Optical Networks (EONs) like large bandwidth, low latency and flexible scheduling capabilities, make it a powerful technique to transmit information between BSs, thereby reducing the transmission cost [7]. Considering the above factors, we combine macro base stations (MBSs) and small-cell base stations (SBSs) which are all equipped with an MEC server to bring storage resources closer to users. Based on the structure, the video streaming can be pre-cached in the MEC server according to the traveling route of vehicles. Therefore, instead of migrating the whole content from one server to another, this paper proposes an algorithm which pre-caches the divided chunks in the BS MEC server covering the area where the vehicle is located. In this way, the content transfer time and HO time supporting service continuity in the dynamic and heterogeneous MEC network can be reduced effectively. The details of the design and simulation results are presented in this paper.
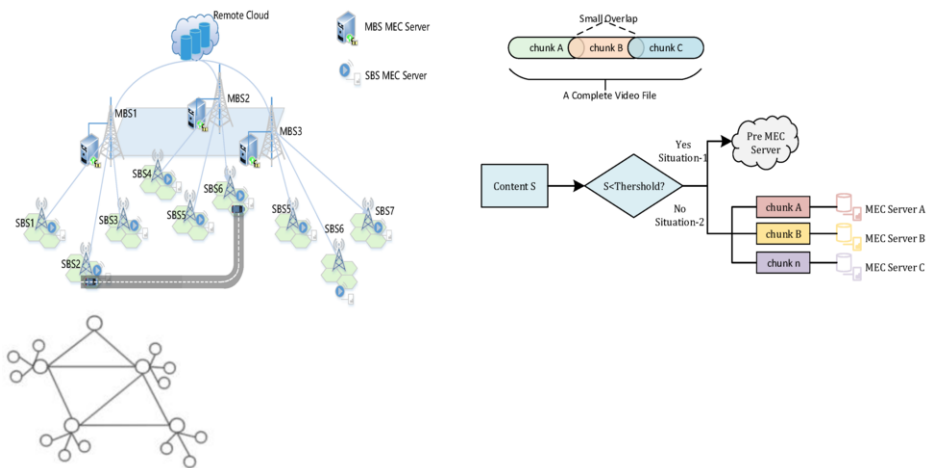

Figure 1. Illustration of (a) MEC based EON, (b) Content fetch decision and (c) Network Topology

2. The MEC Based Macro Small Cell Network Architecture

We consider a communication structure including multiple MBSs and SBSs which are all deployed with an MEC server as shown in Figure1. (a), so that they can monitor content requests sent by all users within the coverage area and make corresponding content delivery decisions. Also, there is a remote cloud. The service range of the MBS

covers all SBSs, which are deployed without overlapping with each other. From previous research [8], we know that users can only get the content from SBSs with the nearest geographic location and the video streaming is either delivered fully from the original storage place or stored separately in different MEC servers. To illustrate the problem, we can divide it into 2 situations which are shown in Figure 1. (b). Situation-1 indicates that the video is fetched from the MEC server of the original MBS and directly sent to the user. Situation-2 indicates that the video is divided into multiple chunks and pre-cached in the MEC server corresponding to the position of the vehicular user. During each user HO, the content must be migrated from the source MBS node to the new MBS node. Accordingly, the chunked video streaming should be pre-cached in the new MEC(NMEC) server before the vehicular user arrives at the new SBS cell. It should be noted that each chunked video fragment should have a small overlap to guarantee the integrity of the video streaming. There have been lots of research previously on predicting the traveling route of vehicles. For example, a convolution neural network can be applied to predict the road traffic situation in literature [9]. Therefore, we can assume that the route information of the car is predicted already so that we can mainly deal with the content delivery problem which determines the source of the content and the appropriate request scheduling process.

3. Pre Caching Algorithm Joint Handoff and Content Delivery

We have a set of MBS node $M=\{1,2,...m\}$ with the maximum storage capacity S_m, and a set of SBS node $N=\{1,2,...n\}$ with the maximum storage capacity S_n. The network topology $G(m,n,l)$ consists of m MBS nodes , n SBS nodes, and optical link l, as shown in Figure.1(c). The j_{th} video request is denoted as $R_j(P(t), C_j, T_{max, j})$, $P(t)$ is the vehicular users' location changing with time, while C_j is the required video's size (in bits) and $T_{max, j}$ is the maximum latency time for the request. It should be noted that the transmission latency from user to SBS node can be ignored as a relatively small constant cause the communication is wireless. The modulation level l_j is assigned to R_j according to the distance of delivery. f_j is the frequency slot to which the request needs to be allocated. B is the capacity of a frequency slot. Then the latency time LT_j for request R_j is given in Eq. (1), which is the time of transferring required content from the server storing it to the vehicular user through different servers. To determine whether a video streaming should be chunked, a threshold S_c is defined in advance according to the video size in practical.

$$LT_j = C_j / \left(l_j \times B \times f_j \right) \tag{1}$$

We propose a pre-caching algorithm joint handoff and content delivery in Table 1. Before the algorithm starts, we randomly initialize the load of each server and access the vehicular users' request to the SBS nodes at *t1*. In Line 4-8, we can get the position of each user and the MBS they access to before and after their movement. Thus, we can decide which MEC server the requiring content would be delivered to after the user moves. In Line9-11, the algorithm decides whether the content is delivered from the remote cloud or nearby MEC server. Once the content delivery method is scheduled, we need to calculate the latency time, which is shown in Procedure1 using Dijkstra

method. The distance between nodes A and B is defined as $d(A, B)$. Through our proposed algorithm, the amount of data to be transferred during the HO stage can be reduced since we deliver the chunked video segment instead of the whole one. It also should be noticed that the chunked video should have a small overlap with the ones before and after it to assure the integrity of the video. In this way, the resulting interruption time can be significantly reduced.

Table 1. Algorithm 1

Algorithm1: Pre-caching algorithm joint handoff and content delivery	**Procedure1**：Calculate the latency time

1: **Input:** Content Request R_j and network topology G
 2: **Output:** The source of the requested content, path,
 3:and latency time of the request
 4: **for** each request of the vehicular users in R_j **do**
 5: search its position in t1 $P(t1)$ and its connected
 6:MBS node $M1$
 7: search its position in t2 $P(t2)$ and its connected
 8:MBS node $M2$
 9: **if** $C_j<S_c$ **then:** deliver video from the remote cloud;
 10: calculate the latency time in **Procedure1**
 11: **else:**
 12: divide the video into chunks according to $P(t)$
 13: Pre-cache the chunk in MBS node M
 14: calculate the latency time in **Procedure1; end**
 15: **for**
	Select the closet node to $P(t1)$ as start node N
	 for all node S in graph G **do:**
	  for all neighbor nodes of S **do:**
	  if d (*neighbour node*, N)< d (N,S)+d (*S, neighbour node*) **then:**
	   Update d (N, *neighbour*)= d (N,S)+d (*S, neighbour node*)
	  find next neighbor node; **end for**
	 end for;
	 Determine the modulation level l
	 Calculate the latency time by Eq.1

4. Simulation Results and Discussions

The network topology graph shown in Figure.1 (c) with 17 nodes, 4 MBS MEC servers and 12 SBS MEC servers are used for simulations. The width of the FS is set to 12.5GHz, the size of a video file is 10MB. For comparison, we deployed two baseline algorithms. One is migrating the required video from the MEC server which cached the content in time *t1* and delivering it to the user, the other is delivering the required video straightly from the remote cloud. Since vehicular users are randomly accessed into our network, we test 10 samples to produce an overall performance of our pre-caching algorithm. The simulation result in Figure2. (a) shows that the performance of our algorithm is significantly better than the other two strategies by achieving a reduction in latency cost. In the meantime, the HO interruption time can be reduced since the required video chunks have been already cached in the MEC server corresponding to users' driving route which has been predicted before. What's more, we can observe that delivering contents from the remote cloud is most time costing for most cases. The reason is that the remote cloud is the furthest away from users, which causes more transmission time than the other two methods. Figure2. (b) indicates the server load of 4 MBS nodes, from the result we can see the load is more balanced by applying our algorithm than using the other two strategies, such that the servers' storage space utilization is higher. Consider a situation when one MEC server is almost full, it is

more feasible to accommodate a chunked video streaming than a complete one. In the meantime, the risk of content cannot be delivered successfully due to the crash of servers is reduced by the proposed algorithm.
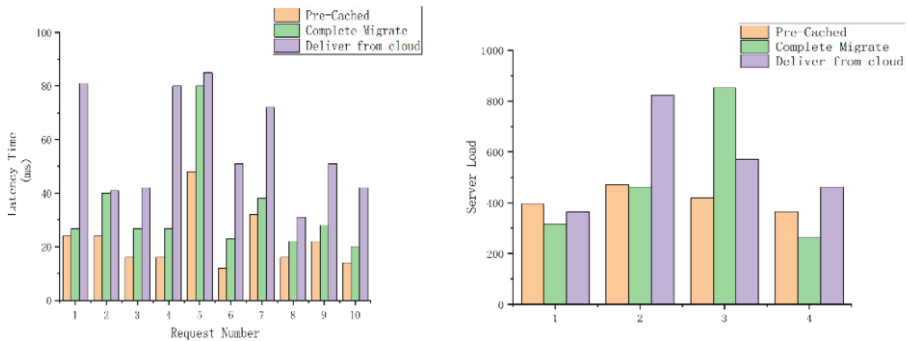

Figure 2. Calculation results of (a) Latency Time and (b) Server Load

5. Conclusion

In this paper, we propose a pre-caching algorithm joint handoff and content delivery, including pre-caching the video streaming in the corresponding server node, deciding the source of the video streaming, choosing the shortest transmission path and calculate latency time. Simulation results show the algorithm could reduce HO time and latency time at the scenario of dealing with high mobility vehicular users' requests. In the future, some more works following this paper can be considered, such as how to automatically decide the threshold S_c for each content to be delivered, and how to decide the size of the small overlap to make the resource utilization more efficient.

Acknowledge

References

[1] Francesco Spinelli,et al, "Toward Enabled Industrial Verticals in 5G: A Survey on MEC-Based Approaches to Provisioning and Flexibility", IEEE,2020

[2] Nandish P., et al, "Mobility Awareness in Cellular Networks to Support Service Continuity in Vehicular Users ", ICOIACT,2020

[3] K. Poularakis and L. Tassiulas, Exploiting user mobility for wireless content delivery, International Symposium on Information Theory, Istanbul, Turkey, 2013.

[4] H. Li and D. et, al, "Hu, Mobility prediction based seamless RAN-cache handover in HetNet", IEEE WCNC,2016.

[5] T. Liu, J. Li, B. Kim, C.-W. Lin, S. Shiraishi, J. Xie, and Z. Han,"Distributed file allocation using matching game in mobile fog-cachingservice network," inProc. IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Honolulu, HI, Apr. 2018, pp.499–504.

[6] Y. Li, P. A. Frangoudis, Y. Hadjadj-Aoul, and P. Bertin, "A mobile edge computing-based architecture for improved adaptive http video delivery," inProc. IEEE CSCN'16, Oct. 2016, pp. 1–6

[7] Z. Liu, et al., "Joint jobs scheduling and lightpath provisioning in fog computing micro data center networks," JOCN (2018):10(7): B152.

[8] Zizheng Guo,et al," Distributed Caching and Lightpath Provisioning in Multi access Edge Computing based Elastic Optical Networks" OSA,2020

[9] Jinglin Li, et,al,"An End-to-End Load Balancer Based on Deep Learning for Vehicular Network Traffific Control"，IEEE INTERNET OF THINGS JOURNAL,2019