MOBILE EDGE COMPUTING BASED IMMERSIVE VIRTUAL REALITY STREAMING SCHEME

Juyong Lee
Department of Electronic and Information System Engineering
Sangmyung University, Cheonan, Korea
e-mail: juyonglee0208@gmail.com

Daeyoub Kim
Department of Information Security
Suwon University, Hwaseong-si Gyeonggi-do 18323, Korea
e-mail: daeyoub69@suwon.ac.kr

Jihoon Lee*
Department of Smart Information and Telecommunication Engineering
Sangmyung University, Cheonan, Korea
e-mail: vincent@smu.ac.kr

Abstract. Recently, new services using virtual reality (VR)/augmented reality (AR) have appeared and then exploded in entertainment fields like video games and multimedia contents. In order to efficiently provide these services to users, an infrastructure for mobile cloud computing with powerful computing capabilities is widely utilized. However, existing mobile cloud system utilizes a cloud server located at a relatively long distance, so that there are problems that a user is not effectively provided with personalized immersive multimedia service. So, this paper proposes the home VR streaming system that can provide fast content access time and high immersiveness by using mobile edge computing (MEC).

* Corresponding author
Keywords: Mobile edge computing, multi-access edge computing, mobile edge computing video service, mobile cloud computing, virtual reality streaming, video streaming

1 INTRODUCTION

Due to the widespread adoption of mobile and Internet of Things (IoT) devices, there are many emerging entertainment services where users can create and share content anytime and anywhere. Also, cloud computing such as Google Cloud Platform (GCP), Amazon Elastic Compute Cloud (EC2) and Microsoft Azure has emerged as a new computing paradigm with the explosive spread of mobile networks. Recently, thanks to the computing paradigm of cloud computing and the explosion of mobile devices, there are growing interests in new applications and services such as real-time online games, AR, VR and ultra-high definition (UHD) streaming that require very low latency and high access speeds. In the meantime, due to the need for high computing power from such immersive and high quality multimedia services, cloud computing platform has been considered as a solution to meet the service requirements.

Even though existing cloud computing platforms have performed well computation intensive task with its powerful computing capabilities and scalability of cloud service, it has trouble supporting immersive services such as AR, VR streaming and multimedia streaming. Furthermore, the enormous volume of data exchanged be-
between user devices and remote cloud servers cause the data tsunami, which leads to saturate and bring down the backhaul networks. In order to solve such problems, a new network architecture to place a small cloud with various computing functions close to the user devices is under active research named as Fog Computing and MEC [1, 2, 3, 4].

The concept of Fog Computing was proposed by Cisco, which is a service infrastructure that accesses the computation and storage resources [5, 6, 7]. It provides computing, storage, and a variety of application services, like the cloud, but it does not have a centralized structure on its central server. That is, Fog Computing seamlessly extends cloud computing to the edge for the secure control and management of domain specific hardware, software, storage, and network functions within the domain and enables secure rich data processing applications across the domain. Meanwhile, the concept of MEC was proposed by the European Telecommunications Standard Institute (ETSI) as a new platform that provides IT and cloud computing capabilities within a radio access network (RAN) in close proximity to mobile consumers [8, 9, 10]. So, MEC architecture can support latency-sensitive services with the backhaul capacity of limited mobile networks [11, 12, 13]. However, there is a lack of researches on the effect to mobile consumers for large capacity multimedia transmission such as VR and UHD in MEC environment.

<table>
<thead>
<tr>
<th>Cloud Computing</th>
<th>Fog Computing</th>
<th>Mobile Edge Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Central processing based model</td>
<td>– Extending cloud to the edge of the network</td>
<td></td>
</tr>
<tr>
<td>– Accessed through Internet</td>
<td>– Decentralized computing</td>
<td></td>
</tr>
<tr>
<td>– Easy to scale</td>
<td>– Realtime data analysis</td>
<td></td>
</tr>
<tr>
<td>– Low cost storage</td>
<td>– Edge can work without cloud or fog</td>
<td></td>
</tr>
<tr>
<td></td>
<td>– Decentralized computing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>– Low latency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>– Realtime service</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Cloud computing vs Fog computing vs Mobile edge computing

So, this paper proposes a scheme that enables users to provide low-latency VR contents streaming experience by utilizing MEC environment. The rest of the paper is organized as followed. Section 2 describes the existing cloud computing problems. Section 3 presents the proposed virtual reality streaming scheme based on MEC environment and then, we present the evaluation results in Section 4. Finally in Section 5 we conclude this paper.

2 PROBLEM STATEMENTS

In this section, we will look into the problems that arise created by the services requiring high computing power from the view point of users and cloud. First, we
will look at the propagation distance of cloud computing caused by the centralized structure and also the effects of when multiple users try to access the cloud server. In addition, we will describe the resource management problem in the cloud environment. Finally, we show the energy efficiency problem created by the services requiring high computing resources.

2.1 Propagation Distance and Multi User Access Problem

For immersive multimedia content, the propagation distance, which affects the transmission latency, is very critical. However, in existing mobile cloud computing environment, mobile consumer devices need to deliver or download their desired data from remotely located servers in the data center or core network. The centralized mobile cloud structure creates many challenges. First, it brings about long latency because various types of data are transmitted across multiple networks including wireless access networks, backhaul networks, and the Internet, which require its own traffic control, routing, and various network management tasks. Second, as the number of mobile consumer devices connected to the cloud server increases, the network resources such as the bandwidth are excessively used, making it impossible to deliver contents to the content requester quickly. Finally, the mobile cloud computing needs to share its computing resources with a much larger number of mobile consumer devices, which aggravates the computation latency due to increased processing delay and high sharing of system resource. Therefore, it leads to the situation that it cannot be assured of a low-latency content transmission as the content requestor is located at a distance from content source.
2.2 Cloud Computing Resource and Task Management Problem

Figure 3. Cloud computing resource and task management problem

Immersive multimedia services require not only many computationally intensive tasks (i.e., encoding, decoding, and transcoding, etc.) but a large storage capacity. Moreover, computing-related services that require robust memory and powerful computing performance (i.e., machine learning, image object detection, video transcoding, live streaming, etc.) lead to high energy consumption. However, sufficient energy supply is required to meet the above-mentioned requirements in order to use these services, but there is a great limitation in mobile and IoT devices using resource constrained batteries. Because of their compact size, mobile and IoT devices have limited energy capacity; therefore, computing power and high energy consumption are key factors. To solve these problems, cloud computing systems use a method of delegating tasks to be performed on mobile devices to a cloud having powerful computing functions. That is, existing cloud computing services are operated in a way that users delegate their own processing to a data center that is rich in computing resources specific to a task.

However, due to the characteristics of this cloud computing architecture, two problems arise; computing resource management and task migration problem. There is a possibility for unnecessary energy consumption, resulting from wasted computing resources, since the provision of services suited to the characteristics of the cloud server and the efficient server arrangement structure are not considered [14, 15, 16]. So, excessive energy consumption and inefficient processing overhead happen when a large number of user’s service requests are received because existing schemes adopt a fixed resource allocation policy for user requests. Therefore, the computing re-
sources are not properly allocated and long service latency is caused. This has become a critical factor for multimedia services.

2.3 Mobile Energy Efficiency Problem

Immersive multimedia services such as VR and AR require high computing power as well as a lot of computation-intensive tasks (i.e., encoding, decoding, transcoding, and head tracking, etc.), which results in high energy consumption. For mobile and IoT devices, high energy consumption is a critical factor. That is, running high computationally demanding applications at mobile user devices is constrained by limited battery capacity and energy consumption of the mobile user devices. In order to solve such a problem, a method of collecting and transmitting data without repeated transmission was proposed [17, 18, 19, 20]. However, these methods are not appropriate for both IoT devices and real-time services in which up-to-date information needs to be managed and monitored. In addition, although this solution can reduce the energy required for network transmission, efficient energy usage cannot be achieved while intensive computing services are used due to limited constrained energy when performing computationally intensive computing operations.

3 VIRTUAL REALITY STREAMING SCHEME BASED ON MOBILE EDGE COMPUTING ENVIRONMENT

To solve the structural limitations of existing cloud computing systems and at the same time, enhance the utilization efficiency of computing resources (i.e., memory, storage, network bandwidth, and so on), the proposed scheme has the hybrid scheme that utilizes both cloud computing and MEC simultaneously. As shown in Figure 4, the proposed scheme assumes the MEC environment to provide low latency VR content streaming to mobile consumer devices. The localized mobile edge computing server (MECS) is a key component to VR content encoding and transmission in the proposed scheme. That is, the MECS located at the network edge takes over performing computation-intensive tasks to be originally handled by the mobile consumer and IoT devices, thereby reducing the computing operation and service provision latency as well as allowing the service provision near the mobile consumer device. The proposed scheme starts by periodically transmitting the network status of wireless access point (WAP) to the relevant MECS in order to transmit the VR content quickly. That is, MECS recognizes the network status and then helps to select the MECS that best suits the specific task. The proposed scheme consists of two methods: live streaming of VR content and streaming of stored VR content.

3.1 Live Streaming of VR Content

The proposed scheme starts when the content source, that wants to stream VR contents in real time, sends a content streaming request message to its access MECS.
The access MECS receiving the message from the content source performs the task of selecting the MECS suitable for VR streaming by checking the computing resources and network status of the neighboring MECs. The proposed approach delegates computation-intensive task to the neighboring MEC when excessive computation-intensive task is concentrated on MECS. Since network and computing capacity status of each MECS are monitored in advance by exchanging information beforehand among MECSs, each MECS can make a decision on encoding process. In addition, when the state of each MECS’s network and computing resource is changed due to the user’s VR streaming request, the MECS transmits its resource variation information to the Broker.

Upon receiving the request, the access MECS determines whether VR content encoding and streaming is possible using its own computing resources. If the access MECS can encode and stream VR content, it provides encoding and streaming service to users using its own computing resources and transmits to the Broker its streaming information to generate a content list. In contrast, if the access MECS
cannot provide the service, it selects the serviceable MECSs using the status information of the neighboring MECs and the Broker, and transmits the VR streaming delegation message to the selected MECs. The neighboring MECS receiving the VR content streaming delegation request message from the access MECS performs VR content encoding and streaming service using its own computing resources. Figure 6 shows the operation procedures of the proposed MECS based on VR live streaming service. That is, since the proposed scheme utilizes the MEC architecture based on the computing power, it can efficiently cope with a large number of VR streaming requests through the resources of the MECS. If the neighboring MECS suitable for the VR streaming operation is selected, the access MECS delegates the VR content encoding and streaming operation to the selected MECS. Therefore, as the proposed method delegates the task processing to a specific MECS that has enough computing capacity and network resources, it enables providing low service latency and improved energy efficiency to mobile consumer device. The rest of the process is the same as using existing cloud-based normal streaming services. That is, since the MECS provides the encoding and streaming service, the operation flow is simplified and the processing burden of the mobile device is reduced.

Figure 6 shows how the proposed scheme works for VR encoding and streaming services. First, mobile user sends VR content streaming request message to the
access MECS. After receiving the VR content streaming request message, the MECS decides whether it can process the VR streaming service. If the MECS can handle the service request, it is in charge of the service request. Otherwise, it forwards the VR streaming request to neighboring MECS. After that, the mobile user wanting to deliver live VR streaming content simply streams the VR content to the MECS without encoding it on his/her own mobile device. Before watching the VR content, each content consumer accesses the VR content list after a registration/login process and then selects VR content to view. After that, the MECS delivers encoded VR content to the content consumer device. As the MECS is responsible for encoding and streaming VR content, the overall operation is simplified and at the same time, the processing burden of the mobile consumer device is reduced. Moreover, as the MECS is located in a relatively short distance from VR streaming source, it leads to low service latency critical to immersive VR content delivery.

3.2 Streaming of Stored VR Content Files

To store VR content in the cloud data center and to share them with other consumers, the proposed scheme is composed of two kinds of use cases as shown in Figure 7. The proposed scheme for streaming stored VR content files uses existing cloud computing architecture and MEC environment. Part 1 covers with the encoding of VR content in a distributed MECS environment.

In part 1, the mobile user trying to record specific VR content first transmits the VR content to its access MECS. The access MECS receiving the VR content
performs a task to check whether its own computing resources are sufficient to perform the VR content fragmentation operation. If the computing resources of the access MECS are sufficient, it performs the VR content encoding task itself. If the size of the VR content file is large, the VR content is fragmented and the encoding task for the VR content is delegated to the neighboring MECS. After VR content fragmentation, the neighboring MECS, to which the encoding task is delegated, encodes the fragmented VR content and then forwards the encoded VR content to the central cloud server. Since the proposed scheme can work encoding VR content together with neighboring MECS, computing resources are efficiently utilized and at the same time, the encoding is completed fast. Also, the central cloud server simply acts as a distribution point for the cached VR content. All fragmented VR content is stored at the data center in the central cloud and then is consumed at the subsequent requests. Moreover, the proposed scheme assumes the popularity-driven content caching policy for low latency service provision. That is, popular VR content is cached at local MECS. After that, when the stored VR content is requested, it is delivered from local MECS, not from the central cloud, which results
in low service latency and high immersiveness. In other words, the proposed scheme can provide multiple content deliveries from both the remotely-located cloud data center and the local MECS near the content consumer so that low service latency and content quality are guaranteed when compared with the existing cloud computing architecture.

On the other hand, part 2 indicates when the requested VR content is received from the repository of the central cloud due to the lack of computing resources at local MECS. In other words, if access MECS and neighboring MECS cannot meet the service requests from multiple content consumers, the requested content is forwarded from the cloud data center. The main difference between the proposed scheme and existing cloud scheme is to provide low service latency by localized transmission through MECS’s caching functionality. Whenever content is delivered toward content consumers, it is cached at MES, which leads to lower network resource utilization and lower service latency than conventional cloud computing platforms.

Figure 8 shows an example of saving and streaming VR content. User sends VR content to his/her access MECS to store VR content. The access MECS receiving the VR content performs fragmentation of the VR content. After completing the VR content fragmentation, the access MECS sends the fragmented VR content to the neighboring MECS. When receiving the fragmented VR content, the neighboring MECS performs the encoding operation and transmits the encoded file to the cloud data center and the neighboring MECS. By caching the encoded VR content at the local MECS, the proposed scheme can distribute the VR content faster than the existing cloud computing architecture.

4 PERFORMANCE EVALUATION

To evaluate the effectiveness of the proposed scheme, the computing power consumption, average service latency and the content download time are measured against various resolutions of VR content and content size, respectively.

4.1 VR Content Download Time

In order to compare the content download time of the proposed scheme with existing cloud architecture, variable sizes of VR content are considered. The MECS function is implemented by using Docker virtualization image [21]. Figure 9 shows the VR content download time against various sizes of VR content. When VR content size is small, the service delivery time is not significantly different between central cloud data center with far distance and the proposed scheme because the transmission distance is not a critical factor due to low volume of the content data. However, as shown in Figure 9 as the number of mobile users at the edge networks and the number of content requests for VR content increases, congestion of the core network and edge network occurs, and therefore an additional operation such as content migration operation must be performed, it may take a long download time.
Figure 8. Operation procedure for storing VR content and delivering stored VR content in the proposed scheme

Figure 9. VR content download time
Also, since the proposed scheme uses the local MEC caching function and the cloud computing architecture, it is possible to provide VR contents faster than the existing cloud and MEC environment. Meanwhile, as the VR content size increases, the existing cloud computing architecture experiences longer transmission delay due to long distance between content consumer and the central cloud server. In addition, the proposed scheme utilizes cached content in a localized MECS, which leads to shorter transmission delay. In other words, the proposed scheme provides the local accesses to the VR content and delivers the cached data from the access MECS locally located at the content consumer.

4.2 Average Response Latency

Figure 10 shows the average response latency when a user requests VR streaming services to a server based on the number of users. The response latency is assumed to exclude VR encoding time done after receiving VR streaming service request. As the number of users who request VR streaming services to the server increases, existing cloud computing architecture has significantly higher latency than the proposed scheme. It comes down to the fact that the data centers in existing cloud architecture are remotely located from the users. Meanwhile, the proposed scheme can provide low response latency because much processing load are performed at MECS close to users. Moreover, as the proposed scheme tries to delegate VR streaming tasks to neighboring MECSs, it can keep a low response latency no matter what the number of user request increases.

![Figure 10. Average response latency](image-url)
4.3 Energy Consumption

Finally, let us focus on the system efficiency of the proposed scheme in terms of the energy consumption. The energy consumption is measured in Watts unit according to the encoding types of 240p, 480p, and 1080p VR content, respectively. Figure 11 shows the computing power consumed to encode each VR content. When VR content resolution is low, the computation intensive task of the VR content encoding is not so large. So, a relatively small amount of computation is required, which consumes less computing power. However, as VR content resolution is high, the computation intensive task increases by VR content encoding, thus leading to much more power consumption. Meanwhile, the proposed scheme performs VR encoding by distributing Docker images based on virtualization using MEC environment. In other words, VR encoding is performed by allocating the appropriate resources required for VR encoding, thus enabling efficient utilization of computing resources. Therefore, the proposed scheme can reduce the workload such as setting up computing resources and task delegation. From that, it can reduce the energy consumption rate when compared to the existing cloud architecture.

Figure 11. Computing power consumption against various VR content size

5 CONCLUSION

This paper shows the following main points. First, the existing central cloud computing architecture cannot fully support real-time applications such as VR streaming/gaming because the central cloud servers are mostly far apart from content consumers. Moreover, high volume VR streaming content requires a lot of computation
load for content encoding. It is critical to mobile consumer devices with battery lifetime constraint. Second, to reduce energy consumption and service latency of consumer devices, the proposed scheme uses MEC that can provide powerful computing power in a close location to the user. It provides low latency service by virtue of its proximate position to the mobile user, and also reduces the computing load of content consumers by delegating to MECS the high intensive computation tasks like VR content encoding. Moreover, it extends lifetime of resource-constrained mobile/IoT devices. Therefore, the proposed scheme can support immersive services requiring fast access time and ultra-high computing performance.

6 ABBREVIATIONS

The following abbreviations are used in this manuscript.

VR: Virtual Reality
AR: Augmented Reality
GCP: Google Cloud Platform
EC2: Amazon Elastic Compute Cloud
IoT: Internet of Things
UHD: Ultra-High Definition
ETSI: The European Telecommunications Standard Institute
RAN: Radio Access Network
MEC: Mobile Edge Computing
MECS: Mobile Edge Computing Server
WAP: Wireless Access Point

Acknowledgement

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grants No. 2016R1D1A1B03931628 and No. 2017R1D1A1B03034215).

REFERENCES


J. Lee, D. Kim, J. Lee


Juyong Lee received his B.Sc. and M.Sc. degrees in information and telecommunication engineering from Sangmyung University, Korea in 2014 and 2016, respectively. He is currently a Ph.D. candidate in electronic information system engineering at Sangmyung University, Korea. His research interests include multi-access edge computing, content centric networking, mobile cloud computing, future internet, software defined networking, and network security.

Daeyoub Kim received his B.Sc., M.Sc., and Ph.D. degrees in mathematics from Korea University, Seoul, in 1994, 1997, and 2000, respectively. From 1997 to 2002, he worked for Telemann and Secui.Com as a researcher. From 2002 to 2011, he was a senior research member in Samsung Electronics. In 2012, he joined the University of Suwon, Korea, where he is currently Associate Professor in the Department of Information Security. His research interests include DRM/CAS, home networks, and Internet security.

Jihoon Lee received his B.Sc., M.Sc., and Ph.D. degrees in electronics engineering from Korea University, Seoul, Korea in 1996, 1998, and 2001, respectively. From 2002 to 2011 he worked at Samsung Electronics as a senior research member. He is currently Associate Professor at the Department of Smart Information and Telecommunication Engineering, Sangmyung University. His research interests include information centric networking, context-aware networking, mobile cloud, and network security.