We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,300
Open access books available

116,000
International authors and editors

125M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Mobile Services Meet Distributed Cloud: Benefits, Applications, and Challenges

Tien-Dung Nguyen, Yunkon Kim, Xuan-Qui Pham, Tri D.T. Nguyen and Eui-Nam Huh

Abstract

As the explosive growth of smart devices and enormous new applications, the variety of corresponding cloud services has been growing quickly. The conventional centralized cloud was faced with an overhead on backhaul links and high latency. Accordingly, a decentralized cloud paradigm including edge computing, mobile edge computing, cloudlet, and so on, was introduced to distribute cloud services to the edge network which located in proximity to mobile devices few years ago. However, this paradigm was not paid attention at that time since cloud technology and mobile network communication were immature to motivate mobile services. Recently, with the overwhelming growth of mobile communication technology and cloud technology, distributed cloud is emerging as a paradigm well equipped with technologies to support a broad range of mobile services. The 5G mobile communication technology provides high-speed data and low latency. Cloud services can be automatically deployed in the edge networks quickly and easily. Distributed cloud can prove itself to bring many benefits for mobile service such as reducing network latency, as well as computational and network overhead at the central cloud. Besides, we present some applications to emphasize the necessity of distributed cloud for mobile service and discuss further technical challenges in distributed cloud.

Keywords: distributed cloud, mobile service, edge cloud, mobile edge cloud

1. Introduction

As cloud computing becomes an essential technology on Information and Communication Technology (ICT) and easy to use for the public, mobile industry has been gradually expanding.
Mobile devices are increasing rapidly with diverse kinds, such as smart glass, smart watch, smart bracelet, or even smart plaster. Although mobile devices are much more powerful, they are still hard to run computing intensive applications due to limitation of computing capacity and battery consumption. For example, machine learning service needs high-performance computing capability, and disaster analysis needs real-time capability. Mobile cloud computing (MCC) [1] has been proposed to offload heavy computing from mobile devices to centralized core cloud as well as to provide huge data storage capabilities to the cloud service customers (CSCs).

According to the rapid growth of mobile devices, the mobile services are also developed overwhelmingly. Some cloud services require low latency, and others require high traffic volume. These requirements apparently degrade performance of the current centralized cloud. If a large number of users simultaneously request cloud services at the same geographical location, the user demands would become a heavy burden on the backhaul links and make long latency for other cloud services [2]. Especially, in wide area network (WAN) environment, cloud service latency is quite difficult to meet quality of service (QoS) due to many uncontro
- able causes [3, 4]. Moreover, one of the most important performance metric is measured by user experience that requires low latency to satisfy customers [5–7]. In the recent years, the latency-related issues have been studied for the purpose of bringing best performance for users [2, 8, 9]. Looking at the above limitations of centralized cloud, a next generation cloud is necessary to leverage mobile cloud market.

From the above necessity, the decentralized cloud has been widely researched and developed in the recent years. Many proposed concepts or techniques have been introduced to innovate a next generation cloud. We now alternately take into account these recent approaches to have an overview of the decentralized cloud.

First, several decentralized cloud concepts, such as fog/edge computing [10, 20], mobile edge computing (MEC) [11], cloudlet [12], micro data center [13], and cloud radio access network (C-RAN) [7], are described in Table 1.

The above concepts bring different views of decentralized cloud. However, each concept only represents an aspect of distributed cloud. While FC/EC focuses on Wi-Fi communication and supports mainly for Internet of thing (IoT), MEC, on the other hand, focuses on radio access network. Cloudlet does not mention about communication and concentrates on cloud computing resource. Micro data center focuses on data-related features and C-RAN tries to separate an edge network into two parts: cloud computing and radio access network. Each concept cannot draw a general view of distributed cloud. Therefore, in this chapter, we will present a fully distributed cloud that can support all above concepts.

Second, the fifth generation (5G)-based cellular network is a significant technology for distributed cloud, especially for multimedia cloud service and in vehicular ad hoc network (VANET). 5G technology is promising to provide high reliability and low latency. For example, the current communication link in VANET uses IEEE 802.11p (bandwidth limitation to 10 Mb/s per channel), which impacts augmented reality (AR) applications on video transmission (required data rate 40 Mb/s) [15]. The 5G cellular network with wide
spectrum (28 Ghz), multiple-in/out and ultra-dense network technologies can support 7.5 Gb/s data rate and achieve stable connection at 1.2 Gb/s in a mobile environment at the speed of 100 km/h [16].

Third, Software Defined Network (SDN) [17] represents an emerging network paradigm that has potential ability to join Wi-Fi/cellular networks and cloud. Originally, SDN was designed for and deployed in wired network environment with high-speed switch. SDN has data plane to transmit data and control plane to indicate where data should go. By integrating SDN, routers not only execute the basic network functions such as send/receive IP-based data packets but provide cloud services by itself or interact with other routers for computing resources as well. Therefore, the next generation network nodes will be considered to combine network functions and cloud functions.

Lastly, hierarchical network topology plays an important role to efficiently support service delivery and mobility. Proxy Mobile IPv6 (PMIPv6) [18] represents a host-based mobility management protocol that provides continuous services for mobile users. PMIPv6 includes: mobile Taccess gateway (MAG), local mobility anchor (LMA), and authentication-authorization-accounting (AAA). The MAG and LMA cooperate with each other to manage the movements of a mobile node. The AAA server is used for security and payment purpose.

<table>
<thead>
<tr>
<th>Issued by</th>
<th>Concept</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CISCO 2011</td>
<td>Fog/edge computing (FC/EC)</td>
<td>• ends the cloud to the edge of networks, in particular wireless networks for the Internet of Things (IoT) • pushes the computing applications, data, and services away from centralized nodes to the network edge, enabling analytics and knowledge generation to occur close to the data sources.</td>
</tr>
<tr>
<td>ETSI 2014</td>
<td>Mobile edge computing (MEC)</td>
<td>• pushes the cloud computing capabilities close to base station, • applied for Radio Access Networks in 4G, 5G, • is renamed to multi-access edge computing in 2016 [14]</td>
</tr>
<tr>
<td>Carnegie Mellon Univ. 2013</td>
<td>Cloudlet</td>
<td>• is known as limited-resource cloud node located close to users.</td>
</tr>
<tr>
<td>Microsoft 2015</td>
<td>Micro-data center (mDC)</td>
<td>• represents the middle tier of a three-tier hierarchy: mobile device cloudlet cloud • is an extension of hyper-scale cloud data centers (e.g. Microsoft Azure)</td>
</tr>
<tr>
<td>China Mobile Research Institute 2010</td>
<td>Cloud Radio Access Network (C-RAN)</td>
<td>• is combination of legacy radio access network and cloud computing to improve performance of base station</td>
</tr>
</tbody>
</table>

Table 1. Survey of decentralized cloud concepts.
According to the above related works, we will present a new type of distributed cloud based on the above decentralized cloud concepts and techniques. We believe that the distributed cloud has capability to provide high quality and real-time cloud services (augmented reality/virtual reality (AR/VR) or autonomous car, etc.) by reducing network latency and congestion.

In the following section, we will describe an overview of distributed cloud and its benefits. We then present some mobile applications that were emerged from distributed cloud. The open challenging research issues will be discussed in the later section.

2. Distributed cloud overview

2.1. Concept of distributed cloud

The purpose of distributed cloud is to bring a cloud service to a proximity of users geographically in order to reduce latency, network congestion, and guarantee QoS for mobile users. As mentioned above, a hierarchical network topology is considered to set up edge nodes or edge cloud (network node close to user) and to support mobility. The distributed cloud consists of core, regional, and edge cloud as shown in Figure 1. A core cloud is conventional cloud. A regional cloud and an edge cloud are improved from network node. An edge cloud is located in the proximity of users. A regional cloud locates between a core cloud and an edge cloud. Generally, a core cloud has the biggest scale, scale of a regional cloud is smaller than that of a core cloud, and the edge cloud has the lowest scale.

Three main layers of distributed cloud are described in detail as follows:

- Core cloud: It exists on the top layer of the distributed cloud. It has capabilities of whole works of conventional cloud including management and provision of cloud services and data. Examples of core cloud are Google Cloud, Amazon AWS, Microsoft Azure, and so on.

- Regional cloud: It exists on the middle layer of the distributed cloud. It has capabilities of proxy to support high mobility of mobile devices as well as self-deploy cloud service images and data caching. Examples of regional cloud are local mobility anchor (LMA) on Proxy Mobile IPv6 (PMIPv6), middle node on content delivery networks (CDN), etc.

- Edge cloud: It exists on the bottom layer of the distributed cloud and locates in the proximity of users. It has capabilities of being aware of mobile device, service, location, preferences, and so on, and then deploying and providing real-time cloud services. Examples of edge cloud are mobile edge computing (MEC), fog/edge computing (FC), cloudlet, and so on.

On the other hand, core, regional, and edge cloud can be managed on different administration domains because starting points of them are totally different. For example, a cloud service provider (CSP) (e.g., Google, AWS, Azure, and etc.) manages core cloud but a CSP does
not have network infrastructure, so if regional and edge cloud are managed by a network provider (e.g., SK Telecom, Korea Telecom, LG Telecom, etc. in Korea), a CSP and a network provider may cooperate with each other. Besides, in case of a company has both core cloud

Figure 1. Concept of distributed cloud.

Figure 2. Administrator domains on distributed cloud.
and network infrastructure, the company can also cooperate with other CSPs for leasing their network infrastructure. Because the edge and regional cloud are deployed from the network node, they can provide two functions: cloud and network, managed by network providers and cloud service providers. Therefore, at least two administration domains exist on distributed cloud environment as in Figure 2.

2.2. Distributed cloud and conventional cloud

The concept of distributed cloud was started from an idea to combine computing resource on network node. In conventional cloud shown in Figure 3, between an end device and core cloud, there are many network nodes. Since network nodes cannot process cloud service and must forward to core cloud, it clearly causes congestion across the network, increases latency, and degrades cloud service performance. Therefore, conventional cloud environment is gradually changed to the distributed cloud environment as shown in Figure 4. The network nodes are enhanced with cloud functions, so-called edge cloud and regional cloud. Edge cloud is located on access/edge network where it is close to users. Regional cloud is located on distribution network. Core cloud is located on core network. All kinds of clouds can provide cloud services. As compared with the conventional cloud environment, the distributed cloud environment is able to provide a cloud service near cloud service customers. Therefore, network latency of distributed cloud is clearly lower than conventional cloud. As a result, the distributed cloud is proper to provide responsive cloud services.

![Figure 3. Conventional cloud.](image-url)
2.3. Benefits of distributed cloud

The distributed cloud is emerged in order to support high real-time or high-mobility constraint of cloud services including augmented reality/virtual reality (AR/VR), autonomous car, and so on. The distributed cloud has some main benefits to provide cloud services as followings:

- **Location transparency**: A cloud service is provided from any cloud node, that is, an edge cloud, a regional cloud, a core cloud, or combination of them. Location of the cloud services can be hidden from a cloud service customer.

- **High real-time**: A cloud service on conventional cloud environment is provided through far from the network distance. It leads to nondeterministic network latency. However, a cloud service on the distributed cloud environment can be provided from edge cloud located in the proximity of the cloud service customers, so that it can guarantee the QoS of high real-time applications.

- **High mobility**: In order to support mobility in conventional cloud, the path of a cloud service is changed, but cloud service is still provided from a main cloud (i.e., core cloud). However, on the distributed cloud environment, cloud services and data can be floated by following target mobile devices.

- **Load detachability**: When a cloud service customer requests a cloud service to a core cloud on the distributed cloud environment, the edge clouds and the regional clouds between the core cloud and the cloud service customer can be aware of the cloud service and accomplish the cloud service instead of the core cloud. It leads to reducing network traffic in the vicinity of the core cloud.

![Distributed cloud](http://dx.doi.org/10.5772/intechopen.75818)
2.4. Edge cloud deployment

In this section, we will take into account a network node in a distributed cloud, so-called regional cloud or edge cloud. Basically, an edge cloud consists of two main parts: network cloud service, and service aware module. Network function has responsibilities to receive/send signal through antenna and route packages to destination. A cloud service is packaged under an instance (e.g., docker instance) and automatically installed in edge cloud. A cloud service has its own local IP address. Service aware module helps edge cloud to filter cloud service packages (if cloud service exists) and forward to the cloud service in edge cloud. The other packages will route to destination normally. Figure 5 describes three different types of edge cloud. Network of edge cloud can be a wireless access point at home or office or a base station to provide cellular network. To deploy cloud service, we present three ways:

1. A cloud service is packaged as an image and embedded inside network node (e.g. router).
2. A PC is used to run cloud service and connect to router through Ethernet port.
3. A virtual machine technology (e.g., docker, etc.) is exploited to execute cloud services. Virtual machines communicate with each other by internal switch and connect to base station (antenna) by an external switch. This way they can make edge cloud scalable depending on the number of incoming requests, and often use them for cellular network.

3. Distributed cloud-based applications

The trending edge cloud technology will leverage many new mobile services that are equipped with cloud media data. In this section, we introduce some significant applications which are assisted by distributed cloud.
3.1. VR-assist rental service

Finding an apartment/house to rent can be a time-consuming process. Virtual reality tours can help tenants to view a range of properties without ever needing to set foot in the actual house.

Deploying VR-assist rental service on conventional cloud, all databases of in/out-side house viewing are stored at core cloud, and it will cause a network congestion if many tenants access this service at the same time. Distributed cloud is a key solution to greatly mitigate network congestion and enhance performance of this service by distributing geography-based data to edge cloud. Each edge cloud will handle a part of the database from all houses located in same area (as shown in Figure 6). Tenants in every area tend to look up apartments/houses in their area instead of other areas. In case of looking for apartments/houses in another area, the edge clouds will communicate with each other to find matched data for tenants. We believe that this kind of application could be popular in forthcoming years as distributed cloud is mature.

3.2. Video streaming

Another application is video streaming in VANET. The application allows watching movie or enjoying multimedia in car/bus/limousine or even autonomous car while moving. The current popular solution is to use plugin DVD player to play in-car video resources. This solution is quite simple, but it requires given video sources and needs to update frequently. The second solution is to use phone/tablet to watch online favorite video through LTE connection. This solution can cause network congestion as many vehicles try to access this service at the same time. Transferring video data costs high bandwidth network that causes congestion easily and increases service delay. By applying distributed cloud, each base station will become edge cloud and can stream video by itself (as shown in Figure 7). With the benefit of mobility and caching, distributed cloud supports arranging edge clouds along user path in order to reduce latency, handover, and enhance quality of service.

Figure 6. VR-assist rental service.
3.3. Emergency care assistant

An Emergency Care Assistant (ECA) is a system that assists doctors to monitor and save patients during the journey to hospital (Figure 8). The challenge of this ECA is how to deliver symptoms and signs of disease to cloud so that the core service will analyze this information from medical knowledge and patients data to diagnose and give some treatments in real time. Besides, all symptoms and signs of patients are also sent to hospital so that doctors can monitor patients on the way to hospital and give the urgent treatment right away. Thus transferring information is very important and requires low latency. Therefore, with distributed cloud, edge cloud, and regional cloud will be arranged along the path to hospital to receive data from patients, analyze and respond right away.
4. Open research challenges

Although its promises and advantages as discussed above, the distributed cloud faces a new set of fundamental challenges that need to be addressed.

4.1. Context aware

The concept of context-awareness appeared initially in the field of ubiquitous computing, where devices and sensors are distributed pervasively within a given environment. In [19], context is defined as any information of environment that affects to user and the system, and even condition of user and the system (such as battery remaining, or number of user connecting to the system, etc.) can be considered as a context. Using the context information, the performance of services is much enhanced and the system could understand users’ situation to satisfy quality of experience (QoE).

Contextual information in distributed cloud can be categorized as environmental context, application context, user context, device context, and network context. Environmental context includes location of user and time user request. Application context is related to latency of application and type of application (focus on computing or database). User context is mobility, activities, and even interactions with other users. Device context contains information of user devices including CPU, memory, and battery remaining. Network context is the bandwidth of the network and connection signal (Wi-Fi or cellular signal).

To efficiently deploy a distributed cloud, the system has to be aware of all the above contexts to distribute cloud services to edge nodes efficiently so that we can save energy and resources as well as guarantee quality of service.

4.2. Caching

Caching in many research areas has been proved beneficial [21]. In distributed cloud, caching is also a key challenge because cloud services and data are distributed to various places in the proximity of users instead of storing at core cloud far away from users. In distributed cloud, not only data are cached, but cloud services are cached as well. If the number of available cloud services is large, users will receive response from edge cloud in low latency. However, it is a trade-off between performance and economic benefit. The cloud resources of edge cloud are typically limited and costly. Accordingly, we need to achieve an efficient strategy to gain maximum benefit. Another issue related to caching is synchronization. How we can keep updating version of cloud service at all edge clouds and synchronizing data between edge/regional/core clouds of the same cloud service? This synchronization issues needs to be considered carefully.

4.3. Heterogeneous compatibility

Similar to mobile cloud computing, in distributed cloud, we have to deal with various types of heterogeneity, in which the following two main heterogeneity needs to be focused on:
• Heterogeneity in infrastructures: Core, regional, and edge clouds provide different services with custom-built policies, infrastructures, platforms, and APIs that make the cloud landscape heterogeneous. For example, Cloud service is running Linux-based core cloud, but if edge cloud uses ARM-based, the cloud service cannot execute properly and it needs to be recompiled in order to support compatibility.

• Heterogeneity in network: integrate a wide variety of network technologies for radio access with wireless and wired transport solutions interconnecting a huge number of vastly different end devices and users. For example, if the conventional cloud service is running on core cloud that is equipped with high bandwidth network, running cloud service on edge cloud that is equipped Wi-Fi or cellular signal needs to be considered.

In such a non-uniform environment like distributed cloud, facilitating interoperability, compatibility, and portability among heterogeneous clouds is considered.

4.4. Self-deployed service

The next generation of cloud systems will have the ability to automatically provision capacity and distribute workload dynamically in a cloud-to-region-to-edge continuum [22, 23]. Depending on number of incoming requests or history data, edge/regional cloud will be able to preemptively take reconfiguration and remediation actions in a fully automatic fashion to make distributed cloud more efficient.

On the other hand, distributed cloud needs to build a standard API so that core cloud can deploy their cloud services on edge/regional cloud by only a function call. It means that cloud providers just only build their cloud services into images and put them on cloud. Whenever edge cloud wants to execute a cloud service, it just downloads that image and runs quickly. Simplifying such self-deployed service will attract more and more cloud providers to join distributed cloud.

4.5. Service routing

In conventional cloud, mobile application has to ask where the cloud service server is at the initial time and then connect to that service. Contrarily in distributed cloud, mobile application just sends request to cloud service normally, and edge cloud will detect the nearest destination that is running the requested cloud service. To deploy this kind of service, we need to construct a good service routing mechanism so that edge/regional cloud can easily find the target cloud service in short time. A combination of information centric networking (ICN) [24] and software defined network (SDN) [17] can be considered to shape an efficient service routing mechanism for distributed cloud.

4.6. Cooperative orchestration

Integrating a cloud service into a mobile network environment brings a number of challenges related to service orchestration, mainly due to the fluctuation of resources and the evolving radio and network conditions caused by user mobility. Since user mobility across edge/regional clouds cooperating the same cloud service, or even different cloud service on different edge/regional clouds, emerges as an open challenge, how to share current status and variables between cloud services, how to let other clouds know timestamp of video streaming...
when user moves to other areas, are things that need to be considered for assuring efficient network resource utilization, QoE, and reliability.

4.7. Hierarchical distribution

A hierarchical network topology enables mobility feature as well as provides the necessary computation power for latency-tolerant scenarios at the higher tiers and minimum delay for delay-sensitive applications at the lower tiers. To efficiently manage the whole system, it is necessary to build a good protocol between core-regional-edge cloud so that we can save more energy and avoid wasting cloud resources.

On the other hand, an open research challenge is using peer-to-peer network model for distributed cloud instead of hierarchical topology. By this way, managing cloud nodes will get more difficult as well as in routing for cloud service. However, peer-to-peer topology has its own advantage such as not depending on core cloud and reducing network congestion on backhaul link. Therefore, it has promising potential in distributed cloud.

5. Conclusion

The cloud computing is on the way to the next generation cloud, so-called distributed cloud. Distributed cloud is an emerging technology that brings forward the technical benefits of edge cloud computing with networking and support multi-tenancy allowing third party cloud providers to provision applications and services on-demand through standardized APIs. In this chapter, we have shown that the impact of distributed cloud in reducing latency, network congestion, and user mobility. Besides, some mobile applications are described to emphasize the necessity of distributed cloud. We have also introduced some significant open research challenges in distributed cloud in order to improve the next generation cloud. Eventually, considering its potential, it is obvious that distributed cloud will significantly leverage the mobile application and mobile communication.

Acknowledgements

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00294, Service mobility support distributed cloud technology).

Author details

Tien-Dung Nguyen, Yunkon Kim, Xuan-Qui Pham, Tri D.T. Nguyen and Eui-Nam Huh*

*Address all correspondence to: johnhuh@khu.ac.kr

Department of Computer Science and Engineering, Kyung Hee University, South Korea
References


tion. IEEE Communications Surveys & Tutorials. 2017;19:1657-1681. DOI: 10.1109/COMST.2017.2705720


[22] Östberg PO et al. Reliable capacity provisioning for distributed cloud/edge/fog computing applications. 2017 European Conference on Networks and Communications (EuCNC); Oulu. 2017. pp. 1-6. DOI: 10.1109/EuCNC.2017.7980667

