Chapter 10
Cloud-Based Service Delivery Architecture with Service-Populating and Mobility-Aware Mechanisms

Fragkiskos Sardis
Middlesex University, UK

Glenford Mapp
Middlesex University, UK

Jonathan Loo
Middlesex University, UK

ABSTRACT

Advances in Mobile and Cloud technologies have redefined the way we perceive and use computers. Mobile devices now rely on Cloud technology for storage and applications. Furthermore, recent advances in network technology ensure that mobile devices in the future will have high-bandwidth connectivity at all times. This drives the incentive of doing all the processing and storage in the Cloud and using mobile devices to access the services. In this chapter, the authors argue that always-on connectivity along with increased demand of Cloud services will contest the Internet backbone and create problems in the management of Cloud resources. Client mobility is also a factor that should be taken into account when providing Cloud services to mobile devices. The authors therefore propose a new service delivery architecture that takes into account client mobility as well as the distance between clients and services in order to manage Cloud and network resources more efficiently and provide a better Quality of Experience for the user.

DOI: 10.4018/978-1-4666-4781-7.ch010

1. INTRODUCTION

In recent years, Cloud computing has taken the front line in terms of service provisioning. The ability to provide affordable services on-demand along with its elastic nature in terms of resource management and billing has made it a very popular technology for individual users and enterprises alike. Cloud technology, like other distributed architectures, is highly scalable and gives its owners the ability to add resources easily in order to increase performance and enhance the offered services. These characteristics made Cloud computing a lucrative business that has brought many service providers into the marketplace, all competing with their own unique set of services.

Meanwhile, High-Definition (HD) content has now become more easily accessible than ever. What we define as HD content is typically high resolution images and videos as well as multi-channel audio recordings. This type of multimedia content requires high storage capacity and when viewed online, generates a high volume of traffic. As a result, until a few years ago, it was almost impossible to stream such content directly from the Internet. Bandwidth limitations made it prohibitive to watch HD movies online, while at the same time, content providers would have increased storage costs in order to host such content. However, Cloud technology and the evolution of networks, now allows us to store and access this content on the Internet and even watch live events via HD streaming.

However, HD content streaming is not the only thing that Clouds have made possible. Among the latest service offerings, Cloud gaming is becoming increasingly popular in recent years with many providers already on the market. This type of Service-On-Demand performs all the intensive game computations on the Cloud and streams the rendered images to the user, effectively transforming the player's node into a thin gaming client. Since the processing is done on the Cloud, one advantage of this technology is that it makes cross-platform gaming possible without having to programme a game for multiple platforms. Naturally, in order to make a game playable on the Cloud, we have to provide a low latency connection. As a consequence, this technology can potentially generate a large amount of traffic that is also very sensitive to latency and jitter.

As we will see in the following sections, the increasing popularity of real-time entertainment services is putting a very high traffic load on the Internet and the evolution of network technologies is not as fast-paced as it used to be, leaving us with constantly increasing demand for high-bandwidth and low-latency but without the infrastructure to support it. This however is not a new problem and the use of caches brought the solution in an era when most of the content on the Internet was static. However, this is not applicable to real-time services and a new solution is needed for this type of content. Furthermore, as mobile devices are becoming more popular and increasingly make more extensive use of Cloud resources, we are faced with the second aspect of the problem which is the provision of consistent connection characteristics to mobile users. This is particularly challenging since it is often impossible to predict user mobility patterns and compensate for the changes in their connections. As a result, a mobile node, switching between networks may experience largely varying conditions to their connections and consequently, the user's Quality of Experience for real-time entertainment may deteriorate. Therefore, if in the future we are to provide QoS-sensitive Cloud services to mobile devices, we must first guarantee that the connection between the mobile node and the Cloud is reliable and meets certain criteria.

To provide a potential solution to the above problems, in this chapter, the authors investigate a new service-delivery framework that makes use of Cloud Interoperability mechanisms in order to move Cloud services closer to their clients. This approach can minimise the distance that data has to travel between a Cloud and its clients and