MULTIMEDIA SERVICES IN CLOUD COMPUTING ENVIRONMENT

Andreja Samčović

Faculty of Transport and Traffic Engineering,
University of Belgrade
Belgrade, Serbia

Abstract:

For multimedia applications and services over the internet, there are strong demands for cloud computing, due to the massive storage and computation required for serving millions of wired internet or mobile network users. In this new multimedia cloud computing paradigm, users can store and process their multimedia data in the cloud in a distributed manner, eliminating full installation of the media application software on local computers or devices. An overview of cloud computing will be provided in this work, focusing on its impact on multimedia services. The case study of *Netflix* video provider is presented, as well.

Keywords:

cloud computing, multimedia services, video streaming, gaming.

1. INTRODUCTION

It is necessary to have an installed application for every computer operation (word processing, listening to music, making presentations, photo processing...). More complex applications are required for complex computer tasks, and therefore a stronger machine is necessary, which further requires more investment in the computer itself. If there is a need to access all data from different locations, an older solution was to personally carry an external hard drive or laptop, which is not practical et al.

The concept that allows these problems to be avoided is computing in the cloud, or cloud computing. The advantage of this technology is that it provides a number of new business opportunities. For example, simple data sharing and collaborative work on documents increase flexibility and faster execution of tasks. In addition, the extremely important characteristic of cloud computing technology is its lower cost. This technology actually responds to the growing need for increasingly complex IT (Information Technologies) services, causing a reduction in budgets reserved for IT.

Cloud has made great improvements, especially for small and medium-sized enterprises, as they can use the new opportunities offered by clouds and have previously not invested in expensive infrastructure.

Correspondence:
Andreja Samčović

e-mail:
andrej@sf.bg.ac.rs

DOI: 10.15308/Sinteza-2019-176-183

We can say that cloud is a revolution in the IT world. This technology can be freely called the technology of saving, for many reasons. However, organizations are still not completely safe with this technology, and the main reason is that information security is beyond their environments [1]. This is also evidenced by the research of the Statistical Office of the Republic of Serbia in 2017, which showed that only 19.5% of the internet population uses cloud storage and data services [2]. In many organizations, business policy bans the storage of data on non-owned systems. Such organizations generally have a private cloud computing format, which certainly has great advantages.

Speaking about cloud computing, we need to define what is cloud computing. It is well known in the IT world that cloud computing has many definitions. One of them was defined by the US National Institute of Standards and Technology (NIST) in October 2011. This definition is often taken as a reference point for any discussion about cloud: “Cloud computing is a model that allows a common presence, a flexible network access at the user’s request, a shared set of computer resources (e.g. network resources, servers, hard disk space, applications and services) that can be quickly enabled with minimal effort to manage or interact with a service provider” [3].

On the other hand, research marketing agencies, such as Gartner and Forrester offer the following definitions: “The field of computing where scalable and elastic information capabilities are delivered in the form of internet services to a number of external consumers” [4].

Simply, cloud computing involves the use of renting of computer resources from specialized companies that offer these services. The essence is that companies that use cloud services pay just as much computer resources as they are using.

Different definitions exist because of the comprehensive concept from which it is difficult to derive an individual element which is the most important factor for the very idea of the cloud. The term cloud comes from the internet experts, in the early period of the global network. The space and all properties among networked computers were labeled as a cloud. Hence, we can simplify that the name cloud computing comes from the applications that exist on the “cloud” of the web server [5].

Cloud computing has emerged over the past few years as a new environment for hosting and delivering multimedia services over the internet. Cloud computing has became attractive to service providers, because there is a rise in multimedia service demand. However, despite that cloud computing offers important advantages, there are many questions still to be answered.

This paper is organized as follows: some important multimedia services based in a cloud are described. The characteristics of multimedia cloud are presented in the next section, followed by the explanation of media sharing. Advanced multimedia services, such as interactive multimedia gaming and video streaming, are presented further. Concluding remarks are discussed in the final part of the paper.

2. CLOUD COMPUTING FOR MULTIMEDIA SERVICES

The emergence of cloud computing has largely changed existing service models for modern information and communication applications. By utilizing elastic resources in powerful data centers, cloud provides end-users with practical access to computer infrastructure, platforms and software. It is usually provided by remote cloud providers (e.g. *Amazon, Google* and *Microsoft*) on a pay-as-you-go method or on a long-term lease. This new generation of computer paradigms, which offer reliable, elastic and cost-effective resource allocation, can significantly reduce the costs of building and maintaining own computer, storage and network infrastructure. Cloud provides a large number of options for new and existing applications.

Existing applications, from content sharing applications and file synchronization to media transfer applications, have made progress in terms of system efficiency and usability through the cloud computing platform. These improvements come mainly from the exploitation of massive cloud resources, with elastic reservation, pricing and by unloading local devices from demanding computing operations. On the other hand, start-up companies can easily implement their new ideas into real products with minimal investments in the initial phase and can later expand their business without much effort.

Representative of such platforms is *Dropbox*, a typical cloud storage space and a file synchronization provider, which is realized on *Amazon*’s S3 cloud storage servers from the very beginning. *Dropbox* uses the *Amazon* EC2 cloud platform to provide key features such as synchronization and collaboration among different users. Relatively new multimedia cloud services such as cloud-based VoD (Video on Demand) and cloud-based games that appear on the market have also emerged and can change the business model in the coming years [6].

A prominent example of such services is *Netflix*, the main internet provider of video streaming content that switched its infrastructure to the *Amazon* cloud platform in 2010 and has become one of the most important cloud users. Totally, *Netflix* has over one petabyte of multimedia data stored in *Amazon*'s cloud. *Netflix* pays for bandwidth and storage resources, and therefore long-term costs become lower than those with excessive security on owned servers.

Another example of multimedia cloud platforms is *Cloudcoder*, an adaptive video transcoding service that transfers most of the processing into the cloud. Built on the *Microsoft's Azure* cloud platform, *Cloudcoder* can support a large number of transcoder services at the same time, which can also be automatically configured to resolve the burst of the request.

Since 2000, *Amazon* has played a leading role in the development of cloud computing by modernizing its data centers. *Amazon* launched in 2006 the new products to provide cloud computing to external customers, such as AVS (*Amazon Web Services*), which has become one of the most common used cloud computing platforms.

Multimedia cloud

There are strict requirements for cloud computing due to the high storage and required calculations to provide multimedia applications and services to internet users. They can store and process multimedia data in the cloud in a distributed manner, whereas installation of application software can be done on local computers or devices. Multimedia cloud computing shares many common features with cloud computing in general. However, multimedia services are very heterogeneous. There is a wide range of media types and related services, such as Voice over IP (VoIP), video-conferencing, image sharing and editing, image retrieval, or image-based rendering [7]. Clouds should support different types of media and their services for large user databases at the same time. In addition to service heterogeneity, different types of devices, such as smart TVs, computers and smart phones, have different multimedia processing capabilities. The cloud should have a flexible ability to fit into these types of devices, in terms of CPU (*Central Processing Unit*) and GPU (*Graphics Processing Unit*), displays, memory and storage.

A multimedia cloud also needs to provide a certain level of QoS (*Quality of Service*) to meet different requirements. There are two ways to provide QoS for multimedia: one is adding QoS to the current cloud computing infrastructure, and another is adding QoS between cloud infrastructure and multimedia applications. The first approach focuses on design and improvement within the cloud infrastructure. The other way focuses on improving QoS in middle layers, such as QoS on the transport layer, and QoS mapping between cloud infrastructure and media applications.

In short, the strict requirements for access, processing and transmission of multimedia data would create bottlenecks, or cloud congestions for general purposes. Today's cloud design has mostly focused on allocating computer and warehouse resources through various usable mechanisms, while QoS requirements in terms of bandwidth, latency, and jitter have yet to be resolved. In order to realize a multimedia cloud, there is a need for inertia between cloud and multimedia services; that is, multimedia-aware cloud computing (a cloud that is "conscious" multimedia) is needed with enhanced QoS support for multimedia applications and ease of content storage, processing, adaptation and rendering with optimized use of resources.

3. MEDIA SHARING BY USING CLOUD

The first association on cloud is its use for media exchange. Media sharing services like *YouTube* are developing extremely fast. In general, it is difficult, perhaps impossible, to predict in advance the impact and development of new multimedia services. Providing resources is a huge challenge, as each service comes with new ideas. Moreover, advanced techniques and smart marketing strategies can progress in a similar proportion to *YouTube*. However, there is a great dilemma at an early stage of service delivery. On the one hand, providing enough large resources at the beginning is expensive and risky, and if the service is not as popular as it was expected to be, the resources would be lost. On the other hand, the launch of the new service usually encounters problems of scalability. New features and increased user base will lead to high pressure on insufficient infrastructure, which reduces the QoS.

Cloud, which offers a reliable, resilient and cost-effective resource with a pay-as-you-go payment service, is an elegant solution that allows designers to launch service at a lower level. In addition to launching new services, migration of content into the cloud is also useful for existing media services, in the presence of scalability challenges.

Content sharing is an integral part of the cloud service. The demand for easy and scalable interchange is the main reason why multimedia content now occupies a large part of the storage space in the cloud. Centralized data exchange systems can make highly efficient one-to-many exchanges, and synchronous content can be easily shared by large populations, either immediately or with a little delay. Cloud computng can also offer better QoS because links to data centers are generally good and avoid firewall and Network Address Translation (NAT) problems commonly encountered by peer-to-peer networks.

Figure 1 shows a generic framework that facilitates the migration of existing streaming services to a cloud-assisted solution. There is a split into two layers, namely, *Cloud Layer*, and *User Layer*. *Cloud Layer* consists of content sources that transmit live and dynamically booked cloud servers. Upon receipt of user requests, *Cloud Layer* will redirect users to the properly selected cloud server. Such a redirect is, however, transparent to the user, i.e. the entire *Cloud Layer* is considered to be one of the original servers from the perspective of users. Since user requirements change over time and are also dependent on location, *Cloud Layer* will adjust accordingly the amount and locations of rented servers. It will intuitively lend more server resources to increase demand during peak times and terminate leases after reduction. Application of *User Layer* can be flexible. It can be made by individual users that rely solely on *Cloud Layer*, or users connected to a peer-to-peer or CDN (*Content Distribution Network*) infrastructure, but they require additional support from the cloud during load. In other words, *Cloud Layer* can migrate from different existing live streaming systems into the cloud [4].

There are, however, a number of critical theoretical and practical issues to be addressed. Although cloud services are improving, there are still delays in providing resources, for example, to start or shut down a virtual machine may take a few minutes in the current implementation of Amazon EC2. Although such delays are gradually reduced with improved cloud design, they can be difficult to eliminate. Therefore, the system must make good predictions when to hire new servers to meet changing requirements, and when to terminate the server lease in order to reduce costs. This can be done by the cloud forecasting algorithm.

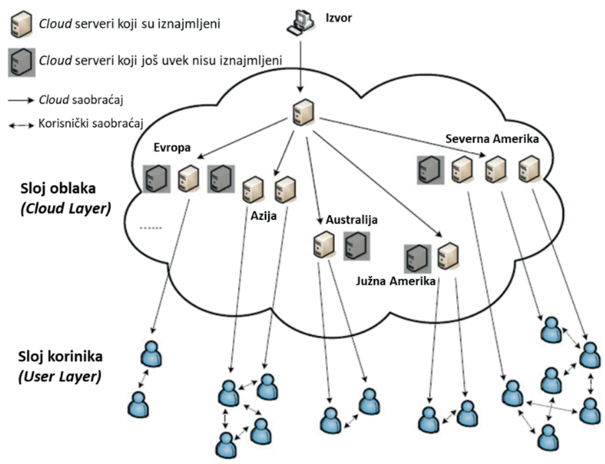

Figure 1. Generic framework for migration of existing live streaming services into a cloud

Netflix application

One of the most successful migration applications for the media sharing into the cloud is *Netflix*, which now occupies one-third of the US internet download traffic during the peak load period, the so-called peak traffic hours. Founded in 1997, *Netflix* has begun to develop from its original business model - sending DVDs (*Digital Versatile Disks*) by post to video-on-demand via internet in early 2007. The original *Netflix* digital video distribution is based on several large *Oracle* servers with *Java* front end. Later in 2008, there were errors in the storage of data that was dropped by the service. Rapid development has created many challenges, which can hardly be foreseen when building private server clusters, without mentioning the staff and skills needed to run a large data transfer infrastructure with a high rate of development. Since 2009, Netflix has started to use *Amazon*'s AWS (*Amazon Web Services*) for part of its services and moved the entire technology infrastructure to AWS in 2012.

To support a combination of massive traffic and unpredictable demand, *Netflix* has developed a global video distribution system that uses the AWS cloud. Figure 2 shows the cloud-based *Netflix* architecture that includes the following key modules:

- *Content conversion* - *Netflix* purchases master copies of digital movies from movie studios and, using powerful EC2 cloud machines, converts them into over 50 different versions with different video resolutions and sound quality, targeting a variety of client video players on computers, smart phones, even on DVD players or game consoles connected to television;

- *Content storage* - main copies and many converted copies are stored in S3. In total, *Netflix* has over 1 petabyte of data stored on *Amazon*;
- *Distribution of content*. To serve users around the world data is sent to content delivery networks CDN (including *Akamai, Limelight and Level* 3) that distribute content to local internet service providers.

With different formats and data rates, dynamic adaptive streaming over HTTP DASH (*Dynamic Adaptive Streaming over* HTTP) is available for streaming to end users. All these services are distributed in the three AWS availability zones. *Netflix* only maintains a minimum hardware infrastructure for user registration and credit card payment. Using Amazon's cloud and CDNs, *Netflix* can respond much better and faster to the users' needs. Such costs, without clouds, can be extremely high for dedicated content providers.

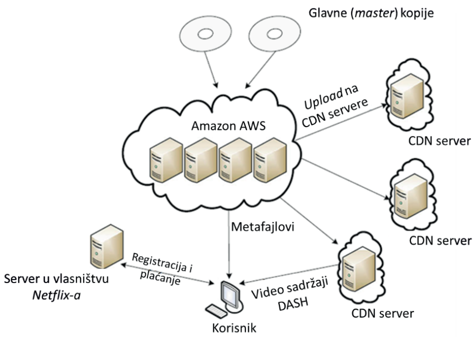

Figure 2. Netflix architecture based on a cloud

Computer requirements for multimedia services

In addition to storing, computing is another rich resource that cloud offers. Many tasks that require intensive computing can now migrate to the cloud, and users do not need to maintain ultra-expensive high-performance servers or clusters of servers, but only to pay the cost on-demand. Computer deployment effectively extends the usability of local devices beyond their physical limitations, which is particularly useful for mobile devices. Today's smart phones and tablets increasingly penetrate into everyday life of people as efficient and practical tools for communication and entertainment. The touch screen and all types of sensors provide even richer user experiences than desktop computers.

Despite the rapid development of key components such as CPU, GPU, memory, wireless access technology and efforts to integrate both handheld and desktop computers, there is a general conclusion that mobile terminals will not completely replace laptop and desktop computers in the near future. The migration of popular software to mobile platforms is still limited due to limited computing, as well as the uniqueness of operating systems and hardware architectures. To make things worse, the battery, as the only power source for most mobile terminals, has achieved relatively little improvement over the past decade. This is what constitutes a major obstacle in providing reliable and sophisticated mobile applications to meet user needs.

By combining the strength of the cloud and the benefits of mobile terminals, there is great progress. For example, *Apple's Siri* service - after launching a voice message over the iPhone, a local recognizer will perform speech recognition and decide whether to turn back to the end cloud to respond in the right way. Other examples include MAUI and *CloneCloud*. The first allows for energy-efficient offloading of mobile codes into the cloud, based on the history of energy consumption. *CloneCloud* uses input variables and offline cost model for dynamic sharing of applications on "bad" devices and cloud.

QoS requirements must also be taken into consideration for multimedia applications. It is necessary to respect critical QoS requirements (e.g., delay, image/video quality, accuracy of computing), so offloading does not affect the user experience. A simple solution for calculating offloading is to move the entire computing mechanism to a distant cloud, then upload all the data needed to calculate and download the calculated results. Although this is common for many implementations, complex applications for organizations are usually composed of multiple service components, which can hardly migrate to the cloud as one entity.

This is further complicated by wireless communication with mobile terminals. Compared to its wired counterparts, mobile terminals are generally more limited in resources; in particular the capacity of wireless communication and the capacity of the battery are inherent bottlenecks. Energy consumption depends on a large extent of the workload characteristics, the data transmission scheme and the used technologies. As such, offloading of the entire application's computer module on a remote cloud is not necessary efficient, if the data volume is large. This may not be a serious problem for users who have a high speed wired connection, but can greatly reduce the benefits for mobile users.

It is necessary to consider video encoding/compression as an important task in a wide range of mobile multimedia applications. The local user uses his mobile terminal to record video in real time, expecting to encode video and then transfer to other users in real time. Directly uploading "raw" video signals without efficient coding inevitably leads to high bandwidth costs and high power consumption for transmission. On the other hand, for video encoding, complex computing is often required, which leads to high energy consumption. For example, for a video coding of 5 seconds (30 frames per second with a resolution of 176×144 and a pixel depth of 8 bits) using a H.264/AVC encoder requires a total of 1×1010 CPU cycles total or 2×109 CPU cycles per second in average, which means that a 2 GHz CPU is required for real-time encoding [4]. Considering that the latest smart phones and tablets are equipped with high definition cameras, the workload of the processor can be 5-10 times higher than in the previous example.

However, offloading the entire cloud video compression task is not practical, because it is identical to the direct transmission of raw video data. Wireless transmission can be either too expensive or simply impossible due to limited bandwidth. For instance, it is well known that motion estimation in video signals is the most computationally demanding operation. Certainly, this operation should be in the offloading focus.

4. INTERACTIVE CLOUD GAMING

Cloud technologies are expanded to enable offloading not only for traditional computing, but for complex tasks such as 3D high definition display, which brings cloud gaming from the level of idea to the level of reality. Cloud gaming in the simplest form represents an interactive game on a remote cloud and involves streaming the scenes over the internet to the user who plays the game. The cloud gaming player interacts with the application through the thin client, which is responsible for displaying the video from the cloud server, as well as for collecting player controls and sending cloud interactions [8]. Figure 3 shows the high-level architecture of this cloud system with thin clients and cloud-based display.

Cloud gaming can bring great benefits by expanding the user base to a large number of less powerful devices that support only thin clients, especially smart phones and tablets. As such, mobile users can enjoy quality video games without performing computer-intensive display of images at the local level.

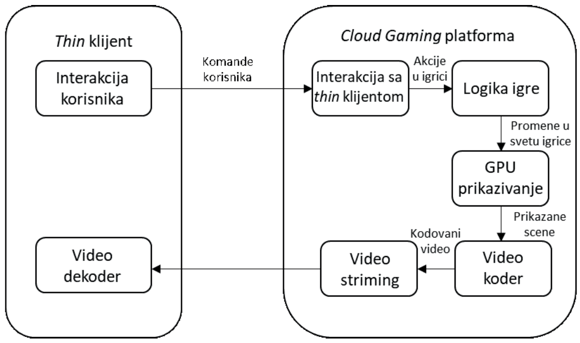

Figure 3. Interaction of a thin client with cloud gaming platform

Moreover, mobile devices have different hardware/software architecture from personal computers, lower memory, limited bandwidth, and limited battery capacities [9]. The traditional console model of games is not always feasible for such devices, which on the other hand becomes the goal of cloud gaming. Cloud gaming also masks the discord between different operating systems through standard web development tools such as HTML5, *Flash*, and *JavaScript*. It further reduces customer support costs because computer hardware is now under the complete control of cloud hosting providers.

As shown in Figure 3, player controls in cloud gaming must be sent over the internet from a thin client to a cloud gaming platform. When the commands arrive at a distant cloud, they are converted into corresponding actions in the game. The logic game turns these actions into certain changes in the game world. Changes in the world of games are processed by GPUs cloud system in the displayed scene. The scene is then compressed by the video encoder and it is sent to the video streaming module, which delivers the video stream back to the thin client. Finally, the thin client decodes the video and displays the video frame of the player.

In order to ensure interactivity, all these operations must be carried out during the millisecond period. This time period is defined as a delay in interaction, and it must be as short as possible that a rich user experience can be passed to cloud players. However, there are always compromises: the player has less tolerance for the interaction delay, the system has less time to perform critical operations such as displaying scenes and video compression. On the other hand, as this time threshold is lower, it is more likely that the delay in the network will adversely affect the player's experience in interaction.

5. VIDEO STREAMING AND CODING

Cloud gaming's requirements for video streaming are similar to live streaming requirements. Cloud gaming and live video streaming must quickly encode/compress incoming video and distribute to end users. In both cases, only a small set of new video frames is interesting and there is no need or ability to access future frames before they are produced, which suggests that encoding must be done in relation to very few frames.

However, conventional live video streaming and cloud gaming have important differences. First, compared to live video streaming, cloud gaming practically does not have the capacity to buffer video frames on the user's side. This is because, when a player issues a command to a local thin client, the command must reach the cloud through the internet, must be processed with the game, displayed by the processing unit, compressed by the video encoder and streamed back to the player. Since all has to be done in less than 100-200 ms, it is obvious that there is not much room for the buffer. Live video streaming, on the other hand, can afford a buffer of several hundred milliseconds, even a few seconds with a very small loss of QoS to the end user.

Sensitive coding needs of cloud gaming in real time make the choice of video encoders to be of importance to any cloud gaming provider. Currently, the main providers of cloud gaming, such as *Gaikai* and *Onlive*, use versions of H.264/AVC (*Advanced Video Coding*) encoders [4]. *Gaikai* uses an encoding-based approach, while *Onlive* uses specialized hardware to compress video stream in cloud gaming. In any case, the choice of the H.264/AVC encoder is motivated by the fact that the encoder has not only a very high compression rate, but also can be configured to work well with strict real-time requirements.

Onlive and *Gaikai* are two industrial pioneers of cloud gaming. *Sony*'s new *Playstation* 4 console will also use *Gaikai*'s cloud platform. *Gaikai* is implemented using two public clouds, namely *Amazon* EC2 and *Limelight*. Figure 4 provides a practical view of *Gaikai*'s workflow. When a user selects a game at *Gaikai* (step 1 in Figure 4), the virtual machine EC2 will first deliver the *Gaikai* game to the user (step 2). It then forwards the proxy IPs ready to run the game by the user (step 3). The user will then select a proxy to run the game (step 4). The proxy starts running the game and the game images will be transmitted to the user via the UDP (*User Datagram Protocol*) (step 5 and step 6). For multiplayer online games, these proxies will also forward user operations to gaming servers and send feedback to users (step 7).

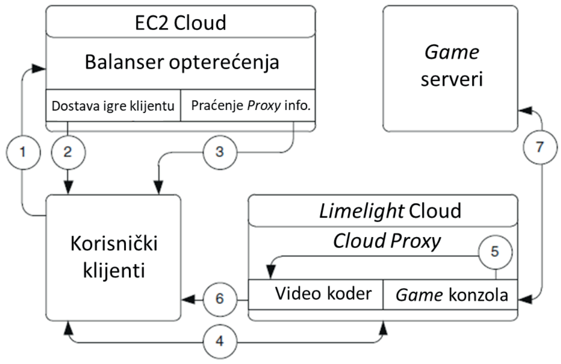

Figure 4. *Gaikaki* cloud gaming platform

Onlive's workflow is similar, but it is applied in a private cloud [10]. The use of public clouds reduces implementation costs and increases scalability; yet a private cloud can offer better performance for playing games.

Table 1 illustrates the effect of compression on a single frame of the video sequence for the popular game, *Batman Arkham Asylum*, streamed at 1280x720 pixels (720p) by *Onlive*. The video quality from the local game console is compared. One can see that the compression effect is noticeable, especially when the amount of available bandwidth is reduced. Although the video quality is good with bit rate 10 Mb/s, there is still room for improving cloud gaming performances.

Table 1. Comparison of video quality for the game Batman Arkham Asylum [4]

Video	PSNR (dB)	SSIM
Master image	n/a	n/a
Local image (without cloud)	33,85	0,97
Onlive (10 Mb/s)	26,58	0,94
Onlive (6 Mb/s)	26,53	0,92
Onlive (3 Mb/s)	26,03	0,89

Video frames are analyzed using two classical metrics, i.e. Peak Signal-to-Noise Ratio (PSNR) ratio and structural similarity (SSIM). The SSIM method calculates similarities in image structures between two video frames. As can be seen, local recording was achieved by high PSNR and SSIM; but that's not perfect, which indicates a certain difference between the recorded video and the master file. Much of this difference is probably due to the slightly different illumination and color settings used by the internal video player in the *Batman* gaming engine. When the local recording is compared with *Onlive* at any connection speed, it can be seen that there is a large drop in terms of PSNR and SSIM. Since PSNR and SSIM are not on a linear scale, the decrease actually indicates a significant degradation of video quality.

6. CONCLUSION

A huge amount of multimedia data can be found on the internet, taking into account rapid development and spread of social networks, smart phones with high resolution cameras, video games, podcasts and other multimedia services. Multimedia services play now a big role in different applications, such as e-health, smart grid, smart cities, image and video forensics, digital cinema and similar. Multimedia cloud systems represent a logical solution for big multimedia data for storage, processing, distribution and service.

One of the rapidly increasing multimedia services is cloud gaming, with many exciting features. In addition to software and service providers, hardware vendors also showed strong interest in cloud gaming and began working on dedicated hardware solutions to address such issues as fast and simultaneous displaying and encoding game scenes, which surely illuminates the future of cloud gaming.

ACKNOWLEDGMENT

This paper is partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Projects No. 32025, 32048).

REFERENCES

[1] T. Mather, S. Kumaraswamy, S. Latif, „Cloud Security and Privacy“, USA, O'Reilly, 2009.

[2] http://www.stat.gov.rs/publikacije/

[3] https://csrc.nist.gov/publications/detail/sp/800-145/final

[4] L. Ze-Nian, D. Mark, L. Jiangchuan, „Fundamentals of Multimedia“, Second edition, USA, Springer, 2014.

[5] A. Samčović, "Cloud computing okruženje u mrežama naredne generacije", Simpozijum o novim tehnologijama u poštanskom i telekomunikacionom saobraćaju – Postel, str. 317-326, Beograd, 2012.

[6] J. Dantas, et. Al., "Hierarchical model and sensitivity analysis for a cloud-based VoD streaming service", 46th IEEE/IFIP Annual international conference on dependable systems and networks workshop, Touluse, France, 28 June – 1 July 2016.

[7] A. Samčović, "Multimedijalne komunikacije", Univerzitet u Beogradu – Saobraćajni fakultet, Beograd, 2015.

[8] R. Shea, J. Liu, E.C.H. Ngai, Y. Cui: "Cloud gaming: architecture and performance", IEEE Network, Vol. 27, No. 4, pp. 16-21, 2013.

[9] K. Kumar, Y.H. Lu, "Cloud computing for mobile users: can offloading computation save energy?", IEEE Computer, Vol. 43, No. 4, pp. 51-56, 2010.

[10] J. Nick, D. Cohen, B. Kaliski, „Key enabling Technologies for Virtual Private Clouds“, USA, Springer, 2010.