A Business Model for Video Transmission Services using Dynamic Adaptation Streaming over HTTP

Demóstenes Zegarra Rodríguez, Renata Lopes Rosa, Graça Bressan
Laboratory of Computer Architecture and Networks - Computer and Digital System Engineering Department
University of São Paulo
São Paulo, Brazil
{demostenes,gbressan}@larc.usp.br , rrosa@usp.br

***Abstract*—A business model for video transmission services is proposed that uses the Dynamic Adaptation Streaming over HTTP (DASH) mechanism using different versions of the same video. These video versions are encoded using different parameters, such as spatial resolution or number of frames per second. The transmission of each video version depends on the capabilities of the network at the end user point. The business model proposed considers different costs for each version of video fragments transmitted. The costs of video resolution upgrade are predefined by the service provider and used during the video streaming. Also, in this model, the user can either accept or reject an upgrade. Subjective tests were performed using different spatial resolutions of the same video, in which the interest level of users to perform a video upgrade is evaluated. Furthermore, the network architecture of the proposed solution is presented and the technical feasibility to deploy the proposed solution in commercial networks is shown. As a consequence, the proposed solution can improve the earnings of the video service provider.**

Keywords-Video Streaming; DASH; Businees Model; Video Quality; QoE; MOS.

I. INTRODUCTION

Nowadays, there are several types of video services applications over IP networks. In this context, video streaming services over the Internet have gained popularity, such as, YouTube [1], Dailymotion [2] among others. This is one of the reasons why the traffic estimated for video on Internet, excluding shared files and games, will reach 61% of the total Internet traffic in 2015 [3].

It is worth noting that, depending on the requirements of video applications and the information to be transmitted, a specific transport protocol can be used. The User Datagram protocol (UDP) is not a connection oriented protocol and therefore the data is not guaranteed to reach its destination; it is generally used for real-time services such as VoIP, some video services or applications of simple request/response e.g., Domain Name System (DNS). The Transmission Control Protocol (TCP) is a connection-oriented mechanism, thereby ensuring that data arrives correctly and in the right order.

Most of the video streaming applications run over HyperText Transfer Protocol (HTTP) which uses TCP as the transport protocol. This is because communication services based on UDP are in some cases intercepted and blocked by firewalls or Network Address Translation (NATs), for this reason, UDP is rarely offered for this type of service.

One problem that arises in the Internet network is congestion. In order to minimize these problems, several congestion control mechanisms for TCP are implemented. Hence, there are many studies aiming to improve its performance, as presented in [4]-[6].

These algorithms can reduce the impairments but none of them solves the problem when TCP detects packet losses. The algorithm reduces the congestion window size and, consequently, the TCP transfer rate decreases. If this new rate is smaller than playback rate, the device' player takes all the buffer information and after that it will enter it into a rebuffering process. During this rebuffering period, no information is displayed and this causes degradation of users' Quality of Experience (QoE).

Different players, such as the open source Adobe Dynamic Streaming [7], Microsoft Smooth Streaming [8], Apple HTTP Live Streaming [9], as well as the players developed by Netflix, and others, use the DASH technique. In adaptive streaming, the video server maintains different versions of the same video, encoded in different bitrates considering spatial and temporal configurations. Also, the video to be transmitted is partitioned into fragments. Then, the client can request different video fragments at different encoding bitrates, considering the network conditions or the users' QoE.

Subjective tests are relevant to quantify the human perception on the quality of voice and video services. As a result, we obtain an index value known as Mean Opinion Score (MOS), which is the mean of the scores granted by at least 15 subjects. Furthermore, providers improve their services based on subjective test results [10].

The most important contribution of this paper is to introduce a new business model for video streaming service using the DASH approach. A test scenario was implemented, in which a video streaming service using DASH was complemented with a billing system. Different resolutions of the same video were stored on the video server for testing purposes. Subjective tests were conducted in a laboratory environment. The goal of these test was to evaluate the users' interest in accepting the suggested video upgrade. Results show that the majority of real users are motivated to improve their QoE watching better spatial resolutions. Additionally, each video version (each resolution) has an identifier (id). For commercial purposes, these identifiers are associated with a predefined cost, and the

 ISBN: 978-1-61208-279-0

upgrade cost is the difference between the cost of two video versions.

In this context, the remainder of this paper is structured as follows. Section II presents the Methods to assess Video Quality considering subjective and objective methodologies. Section III shows the DASH messages and main characteristics. Section IV introduces the Proposed Business Model for video streaming services. Section V presents the results, indicating the users' interest in performing the upgrade and the main functionalities of the proposed solution. Finally, Section VI presents the conclusions.

II. Video Quality Assessment

Video evaluation methods can be classified considering different criteria. The evaluation method can be classified by either using the score given by a subject or the score obtained by an algorithm:

- Objective: these methods use an algorithm or mathematical model to estimate the video impairments. The output of the algorithm could be expressed in different scales. The output of these methods are known as objective metrics.
- Subjective: based on the subject's perception, who grants a video quality score.

The subjective test methodologies and objective metrics are described next.

A. Subjective Test Methodologies

Subjective Methods have been used since the beginning of video quality assessment, and are still valid. They are described in Recommendation ITU-R BT.500-8-11 [11]. This recommendation presents the number of tests and methodologies to correctly conduct the subjective test. The methods are:

- Single Stimulus Continuous Quality Scale (SSCQE);
- Double Stimulus Impairment Scale (DSIS);
- Double Stimulus Continuous Quality Scale (DSCQS);

The ITU-T Recommendation P.910 [12] describes methods of subjective evaluation of video in multimedia applications. The methods presented in this recommendation are:

- Absolute category rating (ACR);
- Absolute category rating with hidden reference (ACRHR);
- Degradation category rating (DCR);
- Pair comparison method (PC).

The description of both, conditions and procedures to produce a video reference is presented in ITU-T Recommendation P.930 [13]. A model to determine the quality of a video transmission using ITU-T P.910 is presented in [14], and its results show how the network variations affect the user's QoE.

B. Objective Metrics

There are different criteria to classify the video quality assessment methods, and they are treated in this section. The ITU defines different categories for the objective methods depending on the type of information considered as the input of the evaluation algorithm, and these categories are the following [15]:

- Media Layer models use either the voice signal or video as an input to estimate the signal quality that determines the end-user QoE. For this method is not necessary to know the network parameters.
- Parametric packet-layer models do not consider the total information content in the packets transmitted. They consider only the headers to estimate the video signal quality
- Planning packet-layer models. In order to perform the estimation quality using this method, prior knowledge of network parameters is required. This method can help the network administrator to optimize the network resources. A proper network planning helps to ensure a good quality of services.
- Bitstream-layer models consider the bit transmitted and, therefore, also use the information used in the parametric packetlayer models.
- Hybrid models are methods that combine two or more of the models described above.

As shown above, there are different objective metrics for video quality assessment, but most of them are not suitable for video streaming over TCP, such as: Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), Structural Similarity (SSIM) [16], and algorithms based on Region of Interest (RoI) [17] or attentions maps [18, 19]. This reason is because they do not take into account intrinsic characteristics of pauses.

In recent years, few studies have presented metrics based on the application layer parameters, such as the pause intensity metric introduced in [20] or Video Streaming Quality Metric (VsQM) presented in [21]. In this work, we use the VsQM metric in our test scenario.

The VsQM metric proposed in [21] was determined by the parameters: number of pauses, pauses length and weight of the temporal segment during which the pauses occur. These parameters represent the network layer conditions.

The metric $VsQM$ is defined by equation [21]

$$VsQM = \sum_{i=1}^{k} \frac{Wi * Ni * Li}{T} \quad (1)$$

Where:

- N_i is the number of pauses;
- L_i is the average length of pauses, in seconds, that happened in the same temporal segment;
- W_i is a weigh factor which represents the degree of degradation that each segment adds to the total video degradation;
- T_i is the time period in seconds of each segment, and
- k is the number of temporal segments of a video.

III. Video Streaming Service and DASH

Currently, video streaming services do not use any mechanism based on IP networks conditions to improve the QoE.

End user devices are implemented with adaptive video players that use rate adaptation algorithms to minimize the network problems. However, in some cases, this is not enough. In this arena, adaptive streaming plays an important role to improve the user QoE [22]. DASH optimizes and adapts the video characteristics during the video transmission taking into account the network conditions at end user point. Thus, DASH is able to switch from one video fragment to another; the video fragment length is usually between two to ten seconds [23]. Figure 1 shows the messages sent between the client and video server during a DASH process.

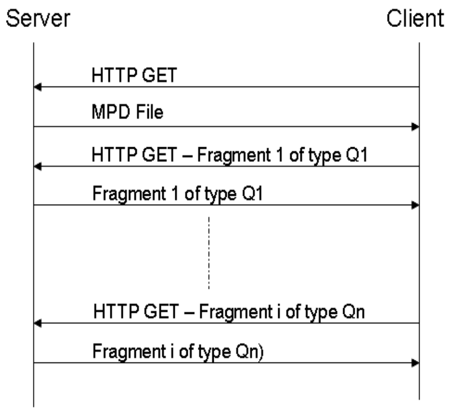

Fig. 1. Messages sent in DASH

The DASH technique presents new challenges and business opportunities for video service providers, telecommunication companies and device developers. Additionally, the research on both video and voice quality metrics are very important, because the adaptive algorithms will choose the video fragment based on these metrics.

The procedure of video streaming over IP network using DASH is described as follows. First, different characteristics of the same video are generated and stored on the video server. These video characteristics can be, for example, the resolution in both spatial and temporal domain, video quality and audio quality. As a second step, each video is logically divided into several fragments; each fragment is identify by a Metadata in the file header. Thus, the client accesses a specific fragment considering its characteristics.

Later, the client sends a HTTP request to the video server ready to attend client requirements. The video server starts the communication sending a DASH-capable video player to the web browser of the client. If an adaptation algorithm is implemented in the player, this algorithm returns a quality metric, which is used to send a new request (standard HTTP GET). Thus, the player requires a video fragment that maps the quality level obtained by the metric. All these steps are performed continuously during the video streaming without any user's action.

IV. The Business Model Proposal

In order to propose a business model for video streaming service, this paper takes into account the DASH technique. The proposal methodology assigns an index value to each version of the same video available in the server. Every video fragment can be identified, and a specific cost can be associated to each fragment.

Figure 2 introduces the solution architecture of the business model proposed for streaming video service using DASH. As depicted in this figure, there are four different versions (A,B,C and D) of the same video. Also, each video fragment is named as *FXi*. Where, *X* represents de video version, and *i* is the fragment number.

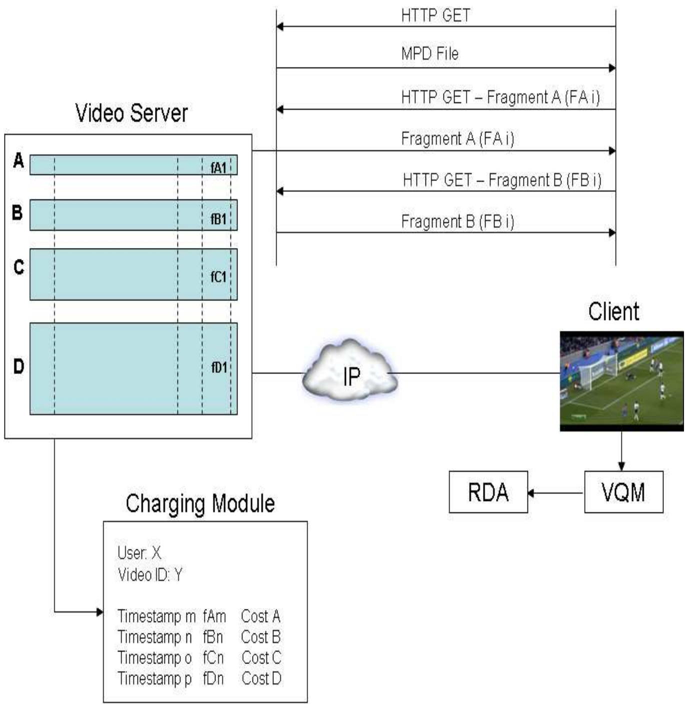

Fig. 2. Architecture of the Business Model for Video Streaming using DASH

From Figure 2, a charging module connected directly with the video server can be observed. In this module, the information sent to the client, regarding all the video fragments is stored and processed. Later, this information needs to be processed in the billing system of the service provider, and finally, charged to the user's account.

The VQM module represents the video quality assessment metric, which is implemented at the application layer, and it is based on the VsQM [21]. In order to estimate the user's QoE, this metric takes in consideration the different buffer status (initial buffering, playing and rebuffering status) in the user's device player. The rebuffering statuses are directly related with the number of pauses and their duration. Then, if the number and duration of pauses increase during a specific period of time, the user's QoE decreases. Further detail about VsQM metric is provided in Section 2.

The module named RDA represents the video Resolution Determination Algorithm that, based on the VQM metric, sends a request for the most appropriated video quality fragment. To perform this action, RDA needs the information contained in the Multimedia Protocol Description (MDP) file.

Some additional considerations about the proposed solution are described below:

- When the RDA module determines that a better video resolution can be received, it sends a message request (HTTP GET - Fragment Ai). The applications running in the video server send the video fragment required (Fragment Ai), and also sends a message with the information regarding the upgrade costs.
- The new video resolution is displayed on the user's device for a time period. This period is programmable, for example, 1 or 3 minutes.
- The user is free to accept or reject the proposed upgrade. Furthermore, there is an option to require not to receive this type of suggestion again.
- If the user accepts the proposed upgrade, the billing module saves the time and the video type transmitted.
- Another input could be used for the RDA module; this new input is related to the networks parameters, such as bandwidth, throughput among others, and is currently under investigation.
- There are other commercial elements that need to be included in the business model, such as content providers and network providers.

V. Test Scenario and Results

In order to determine how the user reacts against a change in video resolution, subjective assessment tests were performed, regarding the user's intention to perform an upgrade. It should be clear that the objective of these tests is not to assess the monetary cost of the upgrade, but the degree of user's interest in performing the video resolution upgrade.

Considering the architecture of Figure 2, tests were performed using only three categories of video resolution, which were previously stored in the video server. These categories are resolution A (A), resolution B (B) and resolution C (C), where A is the lowest resolution, B is the intermediate resolution and C is the highest resolution.

We used the following video quality upgrade scenarios: from A to B, B to C, and A to C. Each upgrade scenario was evaluated using a score into a four-point scale, in which score one represents that user has no interest. Score four represents the case in which the user is very interested in performing the upgrade.

The scale used in the subjective test is presented in Table I.

Carrying out subjective tests involved 23 volunteers who are economically independent, for example, they do not need a third person's authorization to make a payment. Additionally, none of them have any sight problem.

It is considered that the new resolution video (for example, resolution B, in the case of an upgrade of type A - B) is shown for a period of time, i.e., one minute, and in parallel, the upgrade cost is sent to the user screen. If the user accepts the upgrade cost, the charging module of Figure 2 starts recording the new resolution.

TABLE I
Scale used in Subjective Tests

Grading Value	Interest on video resolution upgrade
1	I am sure I have no interest
2	Not right now, maybe later
3	I have interest
4	I am very interested

The global subjective test results of video resolution upgrade intention are presented in Table II.

TABLE II
GLOBAL RESULTS OF SUBJECTIVE TESTS ON VIDEO RESOLUTION UPGRADE

Upgrade Scenario	Score (Mean)
A - B	2.91
B - C	2.13
A - C	3.30

Table III presents the subjective test results in more detail, indicating the number of votes for each option.

TABLE III
RESULTS OF SUBJECTIVE TESTS ON VIDEO RESOLUTION UPGRADE

Scale	Upgrade A-B (number of votes)	Upgrade B-C (number of votes)	Upgrade A-C (number of votes)
1	3	7	2
2	5	9	5
3	6	4	8
4	9	3	10

It can be concluded from the results that users have more interest in performing the upgrade; when they are watching the smallest resolution and the possibility to watch the highest resolution appears. Thus, for example, upgrade scenario "A to C" reached the highest score. Also, Table II allows observing that in all upgrade cases, the mean score is higher than 2; this fact indicates that many users intend to accept the upgrade.

VI. Conclusions and Future Work

The DASH technique permits deploy different solutions, such as the business model proposed for video streaming services. The network architecture of the solution proposed only adds the charging module to the DASH architecture. For this reason, the deployment in commercial networks over Internet or cellular networks is feasible. The implementation of the solution does not represent great costs, and the technical deployment is not complicated.

In the solution proposed, the user is free to accept or reject the video resolution upgrade offered. Thus, the user has the control and takes the decision.

Results of subjective tests show that real users are interested in performing an upgrade, and they can accept to pay a bit more for the spatial resolution upgrades. Thus, video streaming service providers using the business model proposed can increase their income due to the different video rates.

This work shows that video quality metrics play an important role in video streaming services, because their results can be used in different solutions in order to improve the user's QoE.

As future work, conducting subjective tests using more spatial resolution of the same video as test material are proposed, and also different video contents. Also, videos with different temporal resolutions will be included in the test material. Additionally, to have statistically valid results, the number of volunteers to perform the test will be increased.

ACKNOWLEDGMENTS

The authors thank the Laboratory of Computer Architecture and Networks (LARC) at University of São Paulo for the motivation to research in the area of Computer and Telecommunication Systems. This work was supported by FAPESP (Foundation for Researching Support of São Paulo -Brazil). FAPESP project number: 2011/12724-8

REFERENCES

[1] *Youtube, available at http://www.youtube.com*. Retrieved Apr. 15, 2013.
[2] *Dailymotion, available at http://www.dailymotion.com/*. Retrieved Apr. 15, 2013.
[3] *Cisco System, Visual Networking Index*. White Paper. Jun. 2011.
[4] O. Hiroki, H. Hisamatsu and H. Noborio, *Design and Evaluation of Hybrid Congestion Control Mechanism for Video Streaming*. 11th IEEE International Conference on Computer and Information Technology, 2011.
[5] H. Hisamatsu, G. Hasegawa, and M. Murata, *Non bandwidth-intrusive video streaming over TCP*. Proceedings of 2011 Eighth International Conference on Information Technology, pp. 78-83, Apr. 2011.
[6] L. Cai, X. Shen, J. Pan, and J.Mark, *Performance analysis of TCP friendly AIMD algorithms for multimedia applications*. IEEE/ACM Transactions on Networking, pp. 339-355, Apr. 2005.
[7] *Adobe, HTTP Dynamic Streaming on the Adobe Flash Platform, http://www.adobe.com/products/httpdynamicstreaming*. Retrieved Apr. 18, 2013.
[8] *Microsoft, Smooth Streaming technical overview*.
[9] *Apple, HTTP Live Streaming., http://developer.apple.com/resources/http-streaming*. Retrieved Apr. 18, 2013.
[10] H-J. Park and D-H. Har, *Subjective Image Quality Assessment based on Objective Image Quality Measurement Factors*, IEEE Trans. Consumer Electron., Vol. 57, no. 3, pp. 1176-1184, Aug. 2011.
[11] ITU-R Recommendation BT.500-11, *Methodology for the Subjective Assessment of the Quality of Television Pictures*.
[12] ITU-T Recommendation-P.910, *Subjective Video Quality Assessment Methods for Multimedia Applications,* Geneva, Sep. 1999.
[13] ITU-T Recommendation-P.930, *Principles of a reference impairment system for video,* Geneva, Sep. 1996.
[14] ITU-T Recommendation-P.931, *Multimedia communications delay, synchronization and frame rate measurement,* Geneva, Nov. 1998.
[15] A. Takahashi, D. Hands, and V. Barriac, *Standardization activities in the ITU for a QoE assessment of IPTV*. IEEE Commun. Mag., Vol. 46, no. 2, pp. 78, Feb. 2008.
[16] A. Wang, and C. Bovik, and H. Sheikh, *Image Quality Assessment: From Error Visibility to Structural Similarity*. IEEE Transactions on Image Processing, Vol. 13, no. 4, pp. 600-612, Apr. 2004.
[17] H. Kwon, H. Han, S. Lee, W. Choi, and B. Kang, *New Video Enhancement Preprocessor Using the Region-Of-Interest for the Videoconferencing*. IEEE Trans. Consumer Electron., Vol. 56, no. 4, pp. 2644-2651, Nov. 2010.
[18] J. You, A. Perkis, M. Gabbouj, and M. M. Hannuksela, *Perceptual quality assessment based on visual attention analysis*, in Proc. ACM Int. Conf. Multimedia, Beijing, China, pp. 561-564, May 2009.
[19] A. K. Noorthy and A. C. Bovik, *Visual importance pooling for image quality assessment*, IEEE J. Select. Topics Signal Processing, Vol. 3, no. 2, pp. 193-201, Apr. 2009.
[20] T. Porter and X. H. Peng, *An Objective Approach to Measuring Video Playback Quality in Loss Networks using TCP*, IEEE Communications Letters, Vol. 15, no. 1, pp. 76 - 78, Jan. 2011.
[21] D. Rodríguez, J. Abrahão, D. Begazo, R. Rosa, and G. Bressan, *Quality Metric to Assess Video Streaming Service over TCP considering Temporal Location of Pauses*, IEEE Transaction on Consumer Electronic, Vol. 58, Issue: 3, pp. 985 - 992, Aug. 2012.
[22] O. Oyman and S. Singh, *Quality of experience for HTTP adaptive streaming service*, IEEE Commun. Mag., Vol. 50, Issue:4, pp. 20 - 27, Apr. 2012.
[23] A. Begen, T. Akgul, and M. Baugher, *Watching video over the Web: Part 1: Streaming Protocols*, IEEE Internet Computer, Vol. 15, Issue:2, pp. 54-63, Mar. 2011.