Quality Assessment and Measurement for Internet Video Streaming

ZHANG Xinggong, XIE Lan, and GUO Zongming

(Peking University, Beijing 100871, China)

Abstract: Benefiting from the improvements of Internet infrastructure and video coding technology, online video services are becoming a new favorite form of video entertainment. However, most of the existing video quality assessment methods are designed for broadcasting/cable televisions and it is still an open issue how to assess and measure the quality of online video services. In this paper, we survey the state-of-the-art video streaming technologies, and present a framework of quality assessment and measurement for Internet video streaming. This paper introduces several metrics for user's quality of experience (QoE). These QoE metrics are classified into two categories: objective metrics and subjective metrics. It is different for service participators to measure objective and subjective metrics. The QoE measurement methodologies consist of client-side, server-side, and in-network measurement.

Keywords: Internet video streaming; QoE; QoE assessment and measurement; HTTP adaptive streaming

DOI: 10.12142/ZTECOM.201901003

http://kns.cnki.net/kcms/detail/34.1294.tn.20190314.0916.002.html, published online March 14, 2019

Manuscript received: 2018-09-04

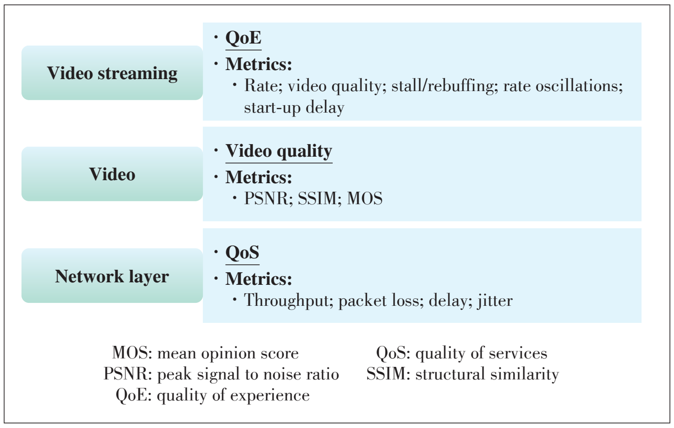
1 Introduction

Past few years have witnessed the booms of Internet video services. The Internet unicorns, such as Youku, Tencent, Toutiao from China and YouTube, Amazon, Hulu from US, are becoming main players in the video entertainment market. Mobile phones, over-the-top (OTT) devices, and online streaming are replacing broadcast/cable TVs as new favorable video entertainment for the generation born after 1980. According to the data from China Internet Network Information Center (CNNIC) [1], the total number of online video users in China is about 751 million, more than the population of Australia. At the same time, the users of broadcasting/cable are going steadily downhill. In 2018, the total number of cable TV users was about 295 million, dropping 19% from that of 2017. Online video services have totally changed the status quo of video transmission. Video streaming for delivering and playing multimedia at the same time emerges as one of the main technologies for Internet video transmission.

Although online video services have been widely deployed, they have not been standardized on the assessment and measurement of the quality of services. Unlike broadcast/cable televisions with dedicated infrastructure, online video streaming systems have to compete for network resources over the Internet. They provide services without quality guarantee. The existing methods of quality assessment are mainly designed for legacy broadcast/cable TVs, which is no more applicable to online video services. It is needed to propose a new framework for video streaming quality assessment.

As shown in **Fig. 1**, the quality metrics are different for network layer, video layer and streaming layer. Quality of service (QoS) is defined by ITU [2] to measure the performance of network, not the actual experience of user. The common QoS met-

This work was supported by National Key R&D Program of China No. 2018YFB0803702, Beijing Culture Development Funding under Grant No.2016-288 and Toutiao Funding No. ZN20171224003.



▲**Figure 1. Quality metrics for network, video, and streaming.**

rics are throughput, packet losses, delay, jitter, etc. The video quality is assessed by comparing original videos with outcome content, pixel by pixel. The metrics of video quality are mainly designed for video coding or legacy broadcast/cable TVs, such as the peak signal to noise ratio (PSNR), structural similarity (SSIM), and subjective metric of the mean opinion score (MOS).

However, video streaming is an end-to-end video delivery and playback. Its quality depends on video coding and on network conditions as well. Mostly, its quality is assessed by quality-of-experience (QoE), a user-centric metric that measures the performance subjectively perceived by the user.

The QoE of video streaming is influenced by the following factors:

- Video level: video quality (PSNR), frame rate, and resolution
- Network level: start-up delay, bitrate, stall/rebuffering, and rate oscillations
- Application-level: video buffering, browser/player, and screen size.

Due to the various factors that affect the QoE, it is needed to standardize the quality assessment for video streaming. This paper presents a framework of quality assessment and measurement, and introduces it from three perspectives: video streaming technologies, QoE metrics, and measurement methodology. This survey paper tries to present an overall framework of quality assessment and measurement, and provide tools to quantify QoE of Internet video streaming.

The paper is organized as follows. The framework of QoE is illustrated in Section 2. Then, three of the most used video streaming technologies are introduced in Section 3. The subjective and objective QoE metrics are given in Sections 4 and 5. The measurement methods are introduced in Section 6. At last, conclusions are given in Section 7.

2 Quality Assessment Framework for Video Streaming

The quality of video streaming is impacted by several factors as video coding, network, and video streaming technologies. It is needed to capture user's QoE to assess the quality of video streaming. A framework of quality assessment and measurement for video streaming is illustrated in **Fig. 2**. It mainly consists of three parts: video steaming technologies, quality metrics, and measurement methods.

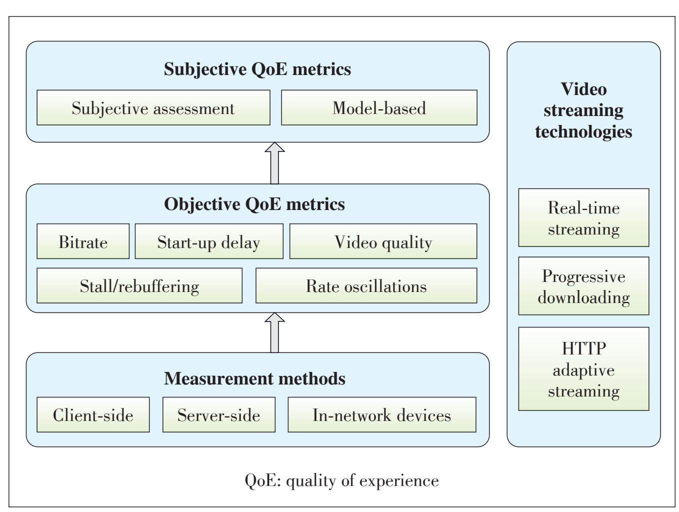
There are various video streaming technologies, which may result in different quality impairments. The streaming technology is one of the most important factors affecting QoE. The widely used streaming technologies includes real-time streaming (RTS), HTTP progressive downloading (HPD), and HTTP adaptive streaming (HAS). All of them are able to enable users to start the playback once the part of the video is downloaded. However, due to their different transmission technologies, their quality impairments are not same. For example, RTS is mainly used in low-latency interactive applications, such as live streaming and video chatting. It is not only sensitive to video quality, but also to round-trip delay. The quality impairments have different impacts on various streaming technologies.

The QoE metrics are used to assess the quality of video streaming. It can be classified into two categories: objective metrics and subjective metrics.

Objective metrics are the QoE metrics which can be quantified with a measurement tool, such as bitrate and delay. These metrics are objective and easy to be measured. However, they have only indirect impacts on users' experience with the service.

Subjective metrics are the direct QoE feedbacks from users. Users rate the video service on a standard measuring way. However, subjective metrics are susceptible to bias because the users' QoE could be varied from one subject to another.

The techniques to measure QoE are also important for video streaming. Video streaming is an end-to-end service. There are multiple parties participating in it, such as content providers, content-distribution-network (CDN) providers, network operators, and users. They view the end-to-end streaming from differ-



▲**Figure 2. Framework of quality assessment and measurement for video streaming.**

ent perspectives, thus the measurement methodologies and tools are also different.

3 Video Streaming Technologies

Video streaming is a delivery technology, which enables users to playback the video while it is being downloaded. For online video services, there are mainly three video streaming technologies: RTS, HPD, and HAS.

3.1 Real-Time Video Streaming

RTS is mainly used for low-latency video applications such as video chat, video conferences, and live video. RTS achieves low latency by a stateful protocol through User Datagram Protocol (UDP) or Transmission Control Protocol(TCP). The streaming server maintains the status of each connection and feedbacks the status to clients.

The implementation of RTS depends on public standardization and proprietary protocols. Real-Time Streaming Protocol (RTSP) was developed by RealNetworks, Netscape and Columbia University. It was standardized as the IETF RFC 2326 standard in 1998. It works with Real-Time Transport Protocol (RTP) and Real-Time Control Protocol (RTCP) together to transmit video data. Real-Time Messaging Protocol (RTMP) was initially a proprietary protocol developed by Macromedia (Adobe). It is a stateful protocol which streams audio/video between a Flash Player and Flash Server. RTMP runs on the TCP protocol and supports the parallel transport of video, audio, data, user commands, and control information.

3.2 HTTP Progressive Downloading

In HPD, a video file is downloaded as a regular file using HTTP from a web server. A client can playback the video while the downloading is going on. HPD is a stateless transmission. The server need not maintain session status. The use of HTTP greatly simplifies the traversal of firewalls and proxy server. Current Internet infrastructure and CDN are fully reusable for HPD. Thus, its deployment cost is relatively low.

However, HPD video playback may be interrupted under poor bandwidth or high packet loss situations. This leads to playback rebuffering or stall. Even more, HPD downloads video files at the fastest speed and stores them on the local hard disk, therefore, once the user exits early, the data that has been downloaded but not watched are wasted.

Many websites using Flash Player, such as Youku, use HPD for the streaming. However, in recent years, more and more websites have given it up for it is not adaptable to bandwidth variation.

3.3 HTTP Adaptive Streaming

HAS technique is proposed to support adaptive streaming over HTTP. An HAS server does not maintain any state information during the streaming. The rate adaptation is done at the client side. This provides scalability with better QoE experience to users. The diagram of HAS is illustrated in **Fig. 3**.

In HAS, media files are divided into "segments", which can be encoded into multiple bitrate versions and assigned to a unique URL. Different versions may have different bitrates, resolutions, formats, languages, and other characteristics. An HAS client requests the proper bitrate version to adapt to bandwidth variation.

Many online video services have already supported HAS, such as Netflix, Hulu, and Amazon. Some products, such as Adobe's HTTP Dynamic Streaming (HDS), Microsoft's HTTP Smooth Streaming (HSS), and Apple's HTTP Live Streaming (HLS), also provide HAS functions.

4 Objective Quality Assessment

The quality of video streaming can be quantified by some tools and objective metrics. The most used objective QoE metrics are listed as follows.

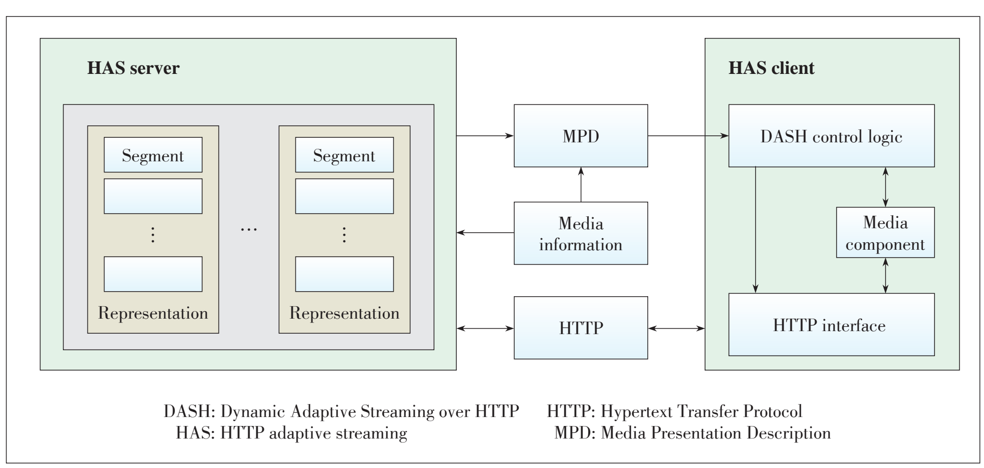
4.1 Video Quality

Video quality in streaming refers to the distortion caused by encoding and transmission compared with the original video. It is often measured with the metrics of PSNR, SSIM, and video quality metric (VQM).

Bitrate is one of the simplest ways to assess the video quality without reference. It is another metric of video quality.

4.2 Start-up Delay

Start-up delay is the time of users' clicking a video and



▲**Figure 3. Diagram of HTTP adaptive streaming.**

waiting before the video starts playing. The start-up delay includes the time of HTML page loading, script loading, video clip buffering, etc.

The start-up delay is an important factor which affects QoE. Some online video services (such as YouTube) tend to initially download data faster and fill the play buffer as soon as possible. In [3], a large-scale user study shows that the start-up delay has a significant impact on a user's online time, and if the start-up delay exceeds two seconds, the user may be stopping watching video.

4.3 Playback Rebuffering or Stall

Stall occurs temporarily when the playback buffer is exhausted. The stall duration is the time that a player waits for the buffer to be filled. In addition, the frequency of stall is also an indicator of video streaming performance.

The rebuffering events during playback will result in a poor user experience. In [3], the authors found that the users with four or more video interruptions were more likely to watch short videos. Also, when the stall duration was more than three seconds, the dissatisfaction increased [4]–[6].

4.4 Bitrate Fluctuation

Frequent bitrate switching will drop users' QoE [7]. Bitrate switching events occur during dynamic adaptive streaming. When network bandwidth deteriorates, a player will reduce the video bitrate and ensure continuous playback. Vice versa, the player increases the bitrate when network becomes better. The bitrate switching can improve bandwidth utilization, but with a bad impact on users' QoE.

5 Subjective QoE Assessment

The other way to assess the quality of video streaming is using subjective QoE metrics. Subjective QoE metrics are used to measure the satisfaction of users in video streaming sessions. The subjective assessment methods are divided into two categories: QoE feedback and model-based QoE.

5.1 QoE Feedback

The QoE score is decided by the feedback scores collected from the human subjects based on their experience of video playback. However, the feedback score can be biased across human subjects, since they are different in physical and psychological confounding factors. To obtain an unbiased and general QoE score, the introduction of statistical analysis techniques is necessary.

One of the most popular subjective QoE metrics is Mean Opinion Score (MOS). For getting MOS, limited sets of human subjects are exposed to watch a video under a controlled testbed and are asked to rate the experience of streaming session. The MOS is a five-point discrete value (Excellent, Good, Fair, Poor, and Bad). And the QoE score is calculated by averaging the MOS given by the users.

5.2 Model-Based Subjective QoE

Collecting feedbacks from the users is time consuming and has limitation on real practice. Therefore, it is feasible to establish a QoE model to estimate the subjective QoE scores from objective metrics. This is an automatic, quantitative and repeatable manner.

There are two model-based methods: (1) learning-based, which uses learning techniques to map the objective metrics to MOS; (2) heuristic methods, which estimate the subjective QoE scores by some manual functions.

5.2.1 Learning-Based QoE Models

The learning-based QoE model uses machine learning and regression analysis to estimate users' MOS. Meanwhile, some objective metrics such as rebuffering and video quality are recorded as well. The subjective ratings and objective metrics are used to train predictive models to estimate the subjective QoE.

In [8], the authors use Random Neural Network (RNN) to map objective metrics to MOS and train a predictive model. In [9], the authors model the correlations between MOS and objective metrics, including video quality level Q_k, rebuffering times F_{freq}, and average rebuffering duration F_{avg}. They use regression analysis to obtain the weights of each term. Furthermore, Maxim et al. [10] define the influence of the average quality level μ, quality variation σ and rebuffering event ϕ on the estimated MOS:

$$\mu = \frac{\sum_{k=1}^{K} \frac{Q_k}{N}}{K},$$

$$\sigma = \sqrt{\frac{\sum_{k=1}^{K}\left(\frac{Q_k}{N} - \mu\right)^2}{K-1}},$$

$$\phi = \frac{7 \times \max\left(\frac{\ln\left(F_{freq}\right)}{6} + 1, 0\right) + \left(\frac{\min\left(F_{avg}, 15\right)}{15}\right)}{8},$$

$$\mathrm{eMOS} = \max\left(5.67 \times \mu - 6.72 \times \sigma - 4.95 \times \phi + 0.17, 0\right), \tag{1}$$

where N is the number of video bitrate levels and K is the number of video segments.

5.2.2 Heuristic-Based Predictive Models

Heuristic-based predictive models manually establish relationship between QoE and objective metrics. Yin et al. [11] consider video quality, quality variation, rebuffering, and start-up delay as the objective factors. They define a QoE function between the objective factors and MOS :

$$QoE=\sum_{k=1}^{K}q(R_k)-\lambda\sum_{k=1}^{K-1}\left|q(R_{k+1})-q(R_k)\right|-\mu\sum_{k=1}^{K}\left(\frac{d_k(R_k)}{C_k}-B_k\right)_+-\mu_s T_s, \tag{2}$$

where R_k is the bitrate of k-th segment, $q(*)$ is the relationship between video bitrate and video quality, $d_k(R_k)/C_k$ is the download time of k-th segment, B_k is the buffer occupancy, and T_s is the startup delay. Therefore, the estimated MOS is a linear increasing function of the average video quality, and it is a linear decreasing function of the video variation, the rebuffering times, and the startup delay. Besides, *l*, *m* and *ms* are the weights on the objective factors.

6 QoE Measurement Methodologies

Using QoE to represent user satisfaction has been widely recognized by the industry, but there is no unified standard for measuring and obtaining the QoE for online video streaming services. According to the methodology and location of data-collection in the network, we classify QoE measurement methodologies for online video services into the following three categories: client-side, in-network, and server-side measurement.

6.1 Client-Side Measurement

There are passive measurement and proactive measurement in the client-side, where some tools are used to measure users' QoE directly.

Passive measurement tools [12], [13] collect the objective QoE metrics when users are watching videos. In this case, the measurement is completely depended on the users and the tools have no control on the video content or duration. Such QoE monitoring tools have been developed for YouTube [13] and Windows Media Player [14] users. By collecting information such as buffer status, TCP rates, and packet loss, they predict QoE metrics like start-up delays and stall times.

Proactive measurements typically use crawlers or bots that crawl through the websites and collect the QoE metrics for a large number of videos. The advantage of using such tools is that they can avoid user participation, thus eliminating any subjective bias. In [15], the authors used a tool called Pytomo to crawl video data on YouTube websites, collecting the network latency, startup-delay, number of stall, and the CDN information.

6.2 In-Network Measurement

Measurements of QoE within network [16] do not require modifying client or server software. It just overhears IP packets passing through links, and estimates the QoE of video streaming in the application layer. It is easy for network operators to deploy these measurement tools.

According to the type of data, in-network measurement can be divided into two categories: TCP layer measurement and HTTP layer measurement.

TCP layer measurement collects on-line or off-line packet information from the TCP layer or lower layer, such as throughput and Round-trip Time (RTT). By tracking the packet-level information of each session, the objective QoE can be estimated, including stall duration, start-up delay, etc.

HTTP layer measurement tracks the performance of HTTP sessions on the application layer. By analyzing the HTTP requests and responses of video data packets, the objective QoE can be obtained.

6.3 Server-Side Measurement

Server-side measurements [17] collect each HTTP data packet on server side and rebuild the HTTP session, by which they can obtain information such as rebuffering frequency, start-up delay, stall time, and bitrate switching frequency.

7 Conclusions

With the boom of online video services, it has attracted more and more interests from industry and academia how to measure and assess the service quality. This paper presents a whole image of the state-of-the-art quality assessment methods of online video streaming. It introduces three streaming technologies and the corresponding quality assessment methods: subjective quality assessment, objective quality assessment, and quality measurement. There still exists a large gap between the industry requirements and the existing academic works. More studies on the QoE modeling and measurement should be carried out in future.

References

[1] China Internet Network Information Center. The 40th CNNIC Statistical Survey Report on Internet Development in China [R]. Aug. 2017

[2] ITU-T. Vocabulary for Performance and Quality of Service: Recommendation P.10/G.100-Amendment 3 [S]. 2012

[3] KRISHNAN S S, SITARAMAN R K. Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs [J]. ACM Transactions on Networking, 2013, 21(6): 2001–2014. DOI: 10.1109/tnet.2013.2281542

[4] HOSSFELD T, EGGER S, SCHATZ R, et al. Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea [C]//Fourth International Workshop on Quality of Multimedia Experience, Yarra Valley, Australia, 2012: 1–6. DOI: 10.1109/QoMEX.2012.6263849

[5] HOßFELD T, SEUFERT M, HIRTH M, et al. Quantification of YouTube QoE Via Crowdsourcing [C]//IEEE International Symposium on Multimedia, Dana Point, USA, 2011: 494–499. DOI: 10.1109/ISM.2011.87

[6] QI Y N, DAI M Y. The Effect of Frame Freezing and Frame Skipping on Video Quality [C]//International Conference on Intelligent Information Hiding and Multimedia, Pasadena, CA, USA, 2006: 423–426. DOI: 10.1109/IIH-MSP.2006.265032

[7] LIU Y T, YUN S, MAO Y N, et al. A Study on Quality of Experience for Adaptive Streaming Service [C]//IEEE International Conference on Communications Workshops (ICC), Budapest, Hungary, 2013: 682- 686. DOI: 10.1109/ICCW.2013.6649320
[8] CHERIF W, KSENTINI A, NÉGRU D, et al. A-SQA: Efficient Real-Time Video Streaming QoE Tool in a Future Media Internet Context [C]//IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 2011: 1- 6. DOI: 10.1109/ICME.2011.6011993
[9] MOK R K P, CHAN E W W, CHANG R K C. Measuring the Quality of Experience of HTTP Video Streaming [C]//12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops, Dublin, Ireland, 2011: 485–492. DOI: 10.1109/INM.2011.5990550
[10] CLAEYS M, LATRE S, FAMAEY J, et al. Design and Evaluation of a Self-Learning HTTP Adaptive Video Streaming Client [J]. IEEE Communications Letters, 2014, 18(4): 716–719. DOI: 10.1109/lcomm.2014.020414.132649
[11] YIN X Q, JINDAL A, SEKAR V, et al. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP [J]. ACM SIGCOMM Computer Communication Review, 2015, 45(5): 325–338. DOI: 10.1145/2829988.2787486
[12] YANG F Z, WAN S. Bitstream-Based Quality Assessment for Networked Video: A Review [J]. IEEE Communications Magazine, 2012, 50(11): 203- 209. DOI: 10.1109/mcom.2012.6353702
[13] STAEHLE B, HIRTH M, PRIES R, et al. YoM: A YouTube Application Comfort Monitoring Tool [R]. Berlin, Germany: Inst. Informatik, Tech. Rep. 467, 2010
[14] DALAL A C and PERRY E. A New Architecture for Measuring and Assessing Streaming Media Quality [C]//3rd Workshop on Passive and Active Measurement Workshop (PAM), La Jolla, USA, 2003: 223–231
[15] JULURI P, PLISSONNEAU L, ZENG Y, et al. Viewing YouTube from a Metropolitan Area: What do Users Accessing from Residential ISPs Experience? [C]// IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), Ghent, Belgium, 2013: 589–595
[16] MAZHAR M H, SHAFIQ Z. Real-Time Video Quality of Experience Monitoring for HTTPS and QUIC [C]//IEEE INFOCOM 2018—IEEE Conference on Computer Communications, Honolulu, USA, 2018: 1331–1339. DOI: 10.1109/INFOCOM.2018.8486321
[17] MANGLA T, HALEPOVIC E, AMMAR M, et al. MIMIC: Using Passive Network Measurements to Estimate HTTP-Based Adaptive Video QoE Metrics [C]// Network Traffic Measurement and Analysis Conference (TMA), Dublin, Ireland, 2017: 1–6. DOI: 10.23919/TMA.2017.8002920

Biographies

ZHANG Xinggong (zhangxg@pku.edu.cn) received the Ph.D. degree from Department of Computer Science, Peking University, China in 2011. He has been an associate professor with the Institute of Computer Science and Technology of Peking University since 2012. His research interests lie in multimedia networks, video communications, information - centric network, and dynamic adaptive HTTP streaming over HTTP (DASH). He was a senior researcher at Founder R&D Center, Beijing, China from 1998 to 2007, and a visiting scholar at the Polytechnic Institute of New York University, USA from 2010 to 2011.

XIE Lan received her master's degree from Department of Computer Science, Peking University, China, and currently is a researcher at Hulu, China. Her research interests lies in multimedia streaming, adaptive video streaming, and panoramic video streaming.

GUO Zongming received the bachelor, master and Ph.D. degrees from Department of Computer Science, Peking University, China in 1987, 1990 and 1994, respectively. He has been a professor with the Institute of Computer Science and Technology of Peking University since 2002. His research interests lie in multimedia streaming, image/video compression, image and video retrieval, watermarking, IPTV, and mobile multimedia. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), a senior member of The Society of Motion Picture and Television Engineers (SMPTE) for China, and a senior member of China Computer Federation (CCF).