New Technology Reports

Latest Technology for Video-Streaming Gateway of M-stage V Live –Assuring Video Quality and Usability–

Harumi Aoyama, Atsuto Miyata, Ryohei Ohgushi and Hiroshi Ishimaru

In the case of a system for delivering "live streaming" video and audio from the internet under a videophone mechanism in real-time, the major issue is to assure video quality and usability at the same time. To solve this issue, NTT DoCoMo has developed technology that utilizes the "video-streaming gateway" provided by "M-stage V Live" service.

1. Introduction

NTT DoCoMo has been providing a service for delivering real-time live streaming of video and audio, M-stage V Live, since May 2003 [1]. Using a transmission protocol for streaming over the internet, the Real-time Transport Protocol (RTP) [2], and utilizing the audio-visual communication standard 3G-324M for 3G mobile visual-type terminals such as Freedom Of Mobile multimedia Access (FOMA) terminals, the M-stage V Live service delivers video and audio under a videophone (or "TV phone") mechanism. As for such a service, from the viewpoint of "liveliness", it is important to assure the real-time characteristics. This means that there must be little to choose between the video and audio play back time on a FOMA terminal and the actual time they were transmitted from their source. Moreover, although it conflicts with this "real-timeliness", it is also important to use buffering to absorb RTP payload size variations and fluctuations in the delay between video and audio delivered from the internet. And by shortening the time before video images are displayed on the FOMA terminal and the disconnect time after completion of playback of archived contents by means of eliminating additional charged time, it is also an important challenge to improve usability by shortening the time taken to connect the FOMA terminal to the M-stage V Live service, in terms of compliance with standards for communication with TV phones (e.g., protocol H.245) and with real-time streaming standards like Real Time Streaming Protocol (RTSP) [3].

Solutions among new technologies adopted by video-streaming gateway of M-stage V Live devised to assure video quality are buffering method for absorbing RTP payload size variation and fluctuations in the delay between video and audio delivered from the internet, and a method to configure 3G-324M streaming data from RTP payload by accounting for error tolerance. Furthermore, as solutions to assure usability, two more technologies are described: a technology for shortening the time delay before video images are displayed on a FOMA terminal, and a technology for shortening the disconnecting time after completing playback of archived contents.

2. System Configuration

The system configuration of the video-streaming gateway of M-stage V Live is shown in **Figure 1**. The video-streaming gateway is positioned as a gateway for combining the internet world with the mobile world; that is, it is located between the DoCoMo core network and the Mobile OPEration Radio Assistant (mopera) [4] server cluster. Note that the video-streaming gateway consists of Interface Converter Equipment (ICE) and the gateway main unit.

ICE is located between the DoCoMo core network and the gateway main unit. It is considered as a specialized device, since it provides an interface with DoCoMo's core network by terminating a Primary Rate Interface (PRI) which provides the interface with the DoCoMo-core-network, and also provides the control protocol for multimedia communication with H.245 [5] which is a multimedia transmission control protocol used in standard 3G-324M, and also provides multiplexing/demultiplexing of video and audio, and H.245 signals under multiplexing protocol H.223 for multimedia communication [6].

Positioned between the ICE and a cluster of servers on the mopera platform, the gateway main unit performs several operations: call control, certification of connection to mopera servers, control of streaming-contents selection, and termination of RTP protocols by serving as a streaming client. It also performs synchronization of video and audio by aligning frames in a time series as well as operational control of the whole system. For the gateway main unit, a general-purpose server was chosen for

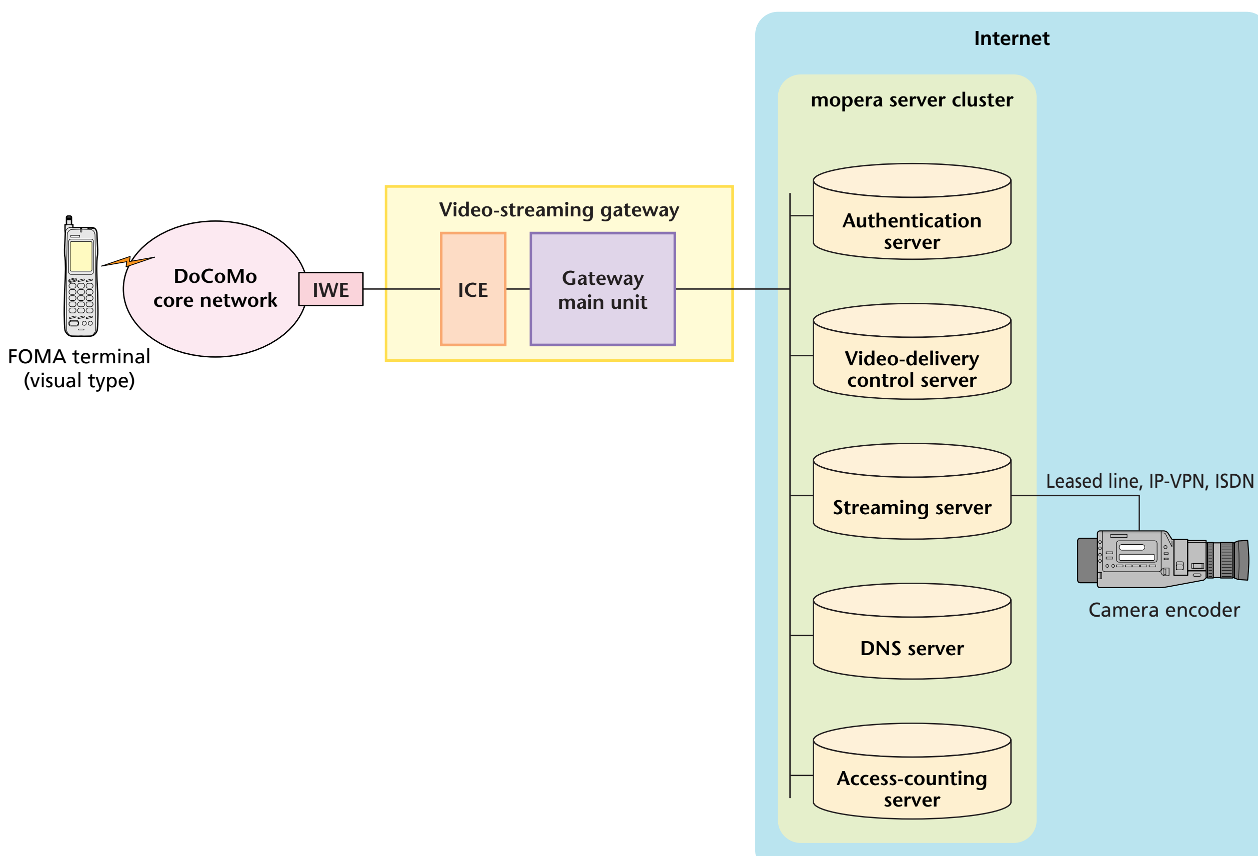


Figure 1 Video-streaming gateway

two reasons: it is generally suitable when the amount of data processed for synchronization of video and audio is relatively large; and it provides better cost performance than the hardware platform. Connected systems include FOMA visual-type terminals, the DoCoMo core network, Inter Work Equipment (IWE), mopera servers (including authentication server, video-delivery control server, streaming server, Domain Name System (DNS) server, and access-counting server), and a live-camera encoder.

3. Technology to Assure Video and Audio Quality

As a technology for assuring high quality for video and audio, the video-streaming gateway of M-stage V Live utilizes technology that absorbs the variations of RTP payload size and the arrival delay between video and audio transmitted in real-time by RTP protocol, and another which configures 3G-324M streaming data from RTP payload by accounting for error characteristics during streaming playback. Technology overviews are shown in **Figure 2**, and explained in the following sections.

3.1 Absorption of Fluctuation in Arrival Delay of Video/Audio and RTP Payload Size Variation

In conventional video-streaming systems, two standard methods are used to absorb fluctuations in arrival delay of video and audio data and RTP payload size variations. That is, on the client side, video and audio are played back after buffering; and on the transmission side, data transmission rate is adjusted in response to the reception status on the receiver side. However, in case where real-time characteristics are emphasized such as the live contents streaming, and the data rate cannot be adjusted because of performance restrictions on the client side, it is difficult to apply the above-mentioned rate-control method as it is. That is to say, in the case of live contents streaming on a TV phone, with narrow-bandwidth transmission channel at 64

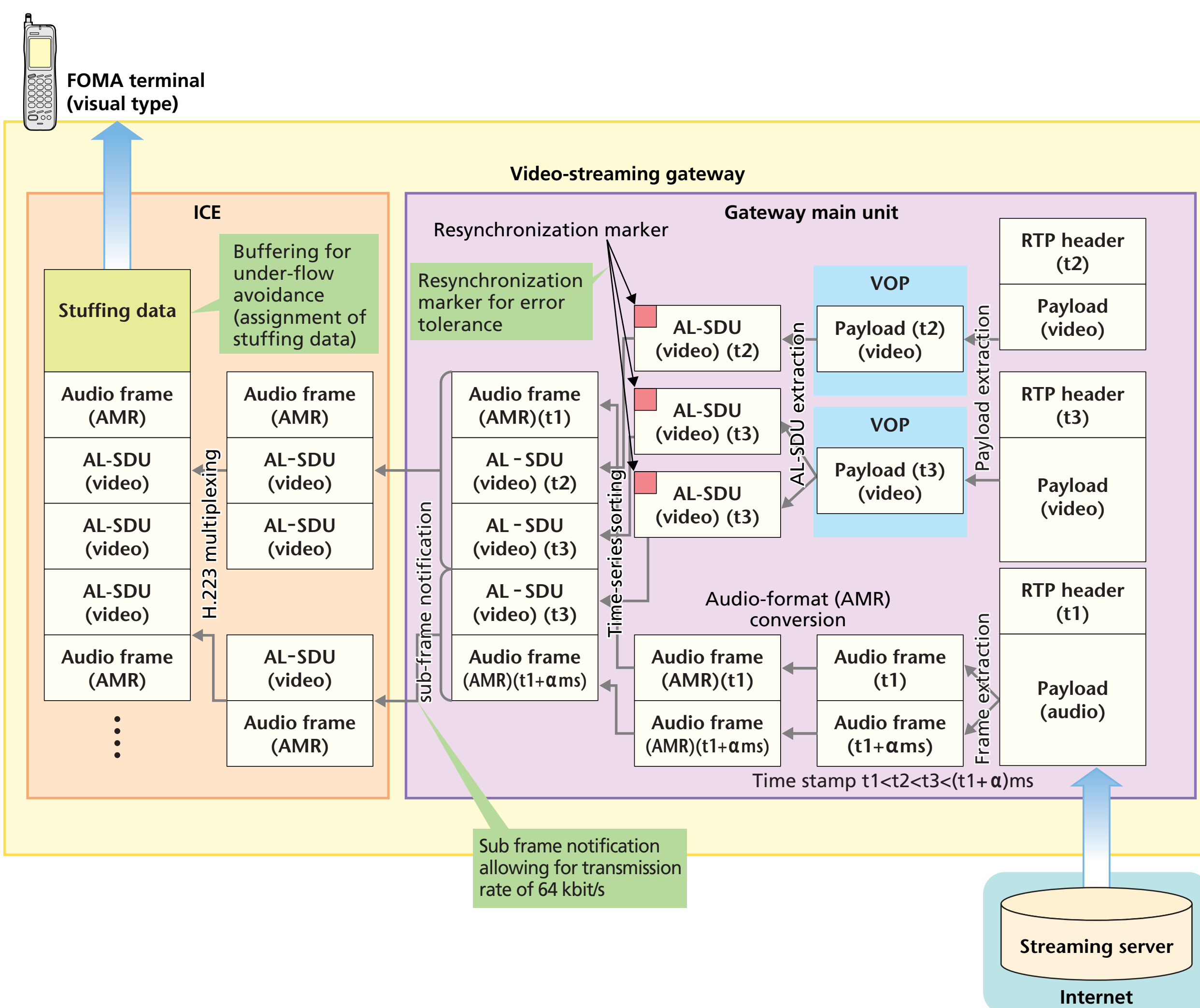


Figure 2 Technology to assure video images quality

kbit/s, which is a calculated average value for length of time, fluctuation of RTP packet arrival and RTP payload size variation mean that a constant 64 kbit/s cannot be assured. The transmission rate for FOMA terminals is guaranteed at 64 kbit/s, and to accommodate this rate, streaming-data rate control that does not depend on fluctuations in arrival delay of video and audio, or on variations in RTP payload size from the streaming server, is needed. Regarding the video-streaming gateway of M-stage V Live, the gateway main unit receives and processes RTP packets with fluctuation, and the ICE sends video and audio data multiplexed at a constant bit rate of 64 kbit/s to a FOMA terminal. Under this configuration, buffering is performed in the ICE and gateway main unit, but flow control is performed only in the gateway main unit. In this way, fluctuations in arrival delay of video and audio RTP packets and RTP payload size variations can be suppressed as much as possible. And, from the internet, the gateway main unit can receive, RTP packet of video and audio affixed with time-stamp information (t1, t2, and t3 in Fig. 2).

As regards audio, the RTP audio payload is extracted as multiple frames according to frame type, and each extracted frame is converted to an appropriate multi-rate Adaptive Multi-Rate (AMR) frame corresponding to the format for FOMA terminals. The converted frame is attached with a time stamp "α" on a time-series at intervals of α ms.

As regards video, the RTP video payload is converted into Video Object Plane (VOP) segments. Each VOP is split up into Adaptation-Layer Service Data Units (AL-SDUs), which are data units used when transmitting streaming data to a FOMA terminal for a maximum size designated for the FOMA terminal under multiplexing protocol H.223. Regarding the split of VOP and audio frames, the time-stamp information present in the RTP header is matched, and each frame corresponding to absolute time is aligned with the time series and sent to the ICE. At that time, so as not to produce overflow or underflow of data on the ICE side, equalization of delivery-transmission and buffering rates—namely, match data rates to 64 kbit/s—is performed for a fixed time period. Video distortion on the FOMA terminal due to underflow of video or audio data is avoided by means of further buffering in the ICE for a fixed time period. This certain fixed time can be set in a "tuning" manner to the shortest possible value in order to have a minimal effect on the real-time transmission characteristic.

3.2 Configuration Technology for 3G-324M Streaming Data Allowing for Error Tolerance

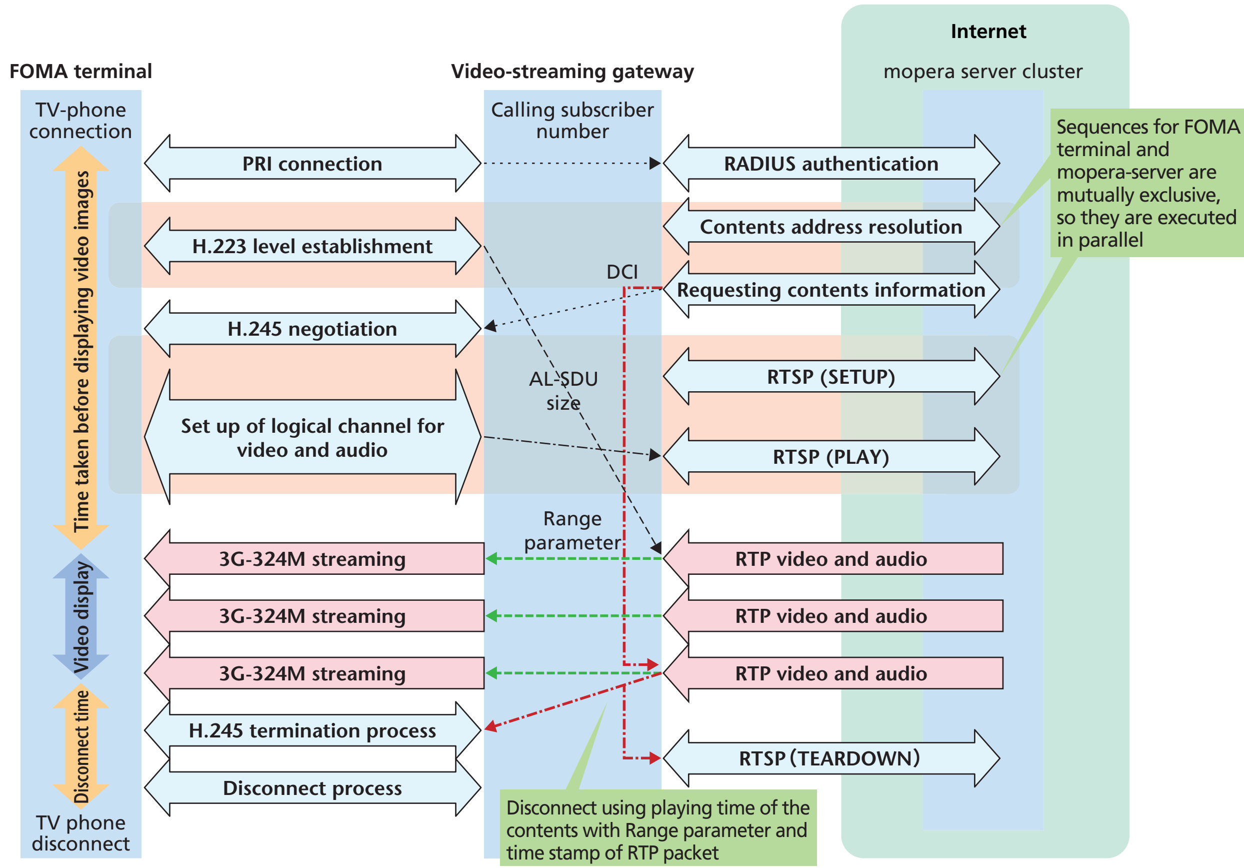
The gateway main unit receives video in the form of RTP packets from the internet, the video payload sent over RTP is converted to a VOP, which is then divided into multiple AL-SDUs (i.e., data units used when transmitting to FOMA terminals). In the case that errors are mixed in with the encoded data of the video, the encoded data cannot be synchronized, and decoding is not possible. As a result, encoded data including the error is skipped and, to denote the start position of the next encoded data to be decoded, a re-synchronization marker is used. As regards the division into AL-SDUs, each re-synchronization marker in a VOP is arranged at the head of each AL-SDU. As a consequence of this re-synchronization, in the case that video images are distorted because bit errors are generated under a wireless transmission environment, on the FOMA terminals, the video images can be refreshed from the instant they get distorted, thus minimizing video degradation.

4. Technology for Assuring Usability

As technology to assure usability in the case of the video-streaming gateway of M-stage V Live, two methods—one for shortening the time elapsed before video images are displayed on the FOMA terminal and another for shortening the disconnect time after completion of playback of archived contents—are explained in the following sections. Technology overviews are shown in **Figure 3**.

4.1 Shortening Delay Before Video Images are Displayed on FOMA Terminal

Video and audio received from the internet are delivered in compliance with real-time streaming standards such as RTP and RTSP. And they are in compliance with TV phone transmission standards such as Recommendation H.245, so they can be streamed to visual-type FOMA terminals. Before streaming starts, according to a connection request from the FOMA terminal, Remote-Authentication Dial-In User Service (RADIUS) authentication is executed by using the calling subscriber number, and contents Uniform Resource Locator (URL) solutions corresponding to a sub-address number designated by the caller, and so on are negotiated between the video-streaming gateway and the FOMA terminal, and between the video-streaming gateway and multiple servers. From the viewpoint of establishing logical channels on the FOMA terminal and avoiding wasted



RADIUS authentication by means of calling subscriber number obtained at PRI connection, and H.245 negotiation using DCI available as response of contents-information request.

Splitting of VOP by using AL-SDU size obtained under a H.223 level establishment

On completion of logical channel set-up, time needed for acquisition of first video and audio RTP packets is measured, and RTSP (PLAY) is sent

In response to recived video and audio as RTP packets, AL-SDUs are extracted, and after the audio format is converted, they are delivered as time-series-sorted 3G-324M streaming data.

When actual elapsed time corresponding to a time stamp exceeds the playing time in terms of the range parameter, RTSP (TEARDOWN) and H.245 termination/call disconnect processing are executed

Figure 3 Technology for assuring usability

communication processing with each server, processing on the FOMA terminal and processing on each server can be serialized. However, time is required before streaming starts, so there is a problem regarding usability. Consequently, except sequences which must be sequentially processed, the processing is done in parallel so that the time elapsed before steaming starts is shortened.

As the sequence performing parallel processing, there is transfer of contents information from the internet, such as Decoder Configuration Information (DCI) needed for decoding video frames on the FOMA terminal, and there is transfer of maximum required size of AL-SDU when streaming video and audio from the internet to the FOMA terminal. As a result of acquiring video and audio from the internet too quickly, real-time characteristic of video and audio will be lost. Therefore, as for parallelization, timing for completing the set-up of logical channels on the FOMA terminal and timing of acquisition of the first RTP packets of video and audio must be synchronized. At the same time, a RTSP (PLAY) method, which indicates the starting of content streaming, is issued. As a result, video and audio data from the internet are attained in a timely manner.

4.2 Shortening Disconnect Time of Archive Contents Playback Completion

From the internet, video and audio data are acquired in compliance with RTP and RTSP standards for real-time streaming. In the case of Video On Demand (VOD) contents, TEARDOWN processing of RTSP, which indicates ending of a streaming session from the internet, is not always performed. In that case, disconnection has to be done by the user or determine whether RTP packets of video and audio are not being sent from the internet. However, in this way, there will be a waiting time to determine whether it is the end of the RTP packets or a delay in the network, so it requires time to end the VOD contents.

In contrast, in the case of the video-streaming gateway, playing time acquired from the "range" parameter, which indicates the time length of the contents, and elapsed time acquired from actual time corresponding to the time stamp of the sequentially received RTP packets are compared. When actual elapsed time corresponding to the time stamp exceeds the playing time by the range parameter, TEARDOWN processing of RTSP and H.245 protocol termination and disconnection processing are executed. As a result of this unique process, completion of contents streaming is detected, and the call will be disconnected in a short time, minimizing the charged communication time.

5. Conclusion

The video-streaming gateway of M-stage V Live is based on two main technologies: one for assuring video quality by taking up variations in RTP payload size and fluctuations in time delay between video and audio RTP packets; and another for assuring usability in connection by means of shortening the time elapsed before video images are displayed. From now onwards, DoCoMo is planning to continue investigating and evaluating technologies that further expand the service by developing a service with high-quality audio codecs and by devising contents browsing control for making contents provision simpler.

References

[1] K. Okubo, et al. "Video-streaming gateway for the M-stage Live service", NTT DoCoMo Technical Journal, Vol. 11, No. 3, pp. 42–47, Oct. 2003.

[2] Schulzrinne H., et al: "RTP: A Transport Protocol for Real-Time Applications," RFC1889, Jan. 1996.

[3] Schulzrinne H. Rao A., and Lanphier R: "Real Time Streaming Protocol (RTSP)," RFC2326, Apr. 1998.

[4] K. Kubo, et al. "mopera STEP2 and Future Developments", NTT DoCoMo Technical Journal, Vol. 7, No. 4, pp. 32–39, Jan. 2000.

[5] ITU-T Recommendation H.245: "Control Protocol for Multimedia Communication," Jul. 2001.

[6] ITU-T Recommendation H.223: "Multiplexing Protocol for Low Bit Rate Mobile Multimedia Communications," Jul. 2001.

Abbreviations

AL-SDU: Adaptation Layer Service Data Unit
AMR: Adaptive Multi Rate
DCI: Decoder Configuration Information
DNS: Domain Name System
FOMA: Freedom Of Mobile muitimedia Access
ICE: Interface Converter Equipment
IWE: Inter Work Equipment
mopera: Mobile Operation Radio Assistant
PRI: Primary Rate Interface
RADIUS: Remote Authentication Dial In User Service
RTP: Real-time Transport Protocol
RTSP: Real Time Streaming Protocol
URL: Uniform Resource Locator
VOD: Video On Demand
VOP: Video Object Plane