

Video Streaming Gateway Supporting Open V-Live

Harumi Aoyama, Atsuto Miyata and Ryohei Ohgushi

We developed a video streaming gateway that supports video distribution on the Internet from various open servers belonging to content providers. With this gateway, we are able to provide mechanisms that will lead to an expanded range of V-Live content and increased video phone traffic.

1. Introduction

For conventional V-Live services [1], DoCoMo provided a video streaming server and Content Providers (CP) would sign a contract with DoCoMo concerning the usage of equipment, registration of content, and other provisions, whereupon the CP would be allowed to provide video streaming services to FOMA terminals by using DoCoMo's server. However, with Open V-Live, it is now possible for CPs to use a variety of unique open streaming servers and Web servers that they operate themselves on the Internet to easily provide services without having to sign contracts with DoCoMo. Consequently, it is hoped that available content will be significantly expanded. Against this background, we developed a new video streaming gateway that supports Open V-Live, in addition to existing V-Live services. This article describes the unique efforts and technologies involved in this development, including efforts made to achieve more simultaneous connections and higher performance, support for a new interface for open servers, as well as technologies related to the issuance of called party subaddresses (content IDs) and technologies related to ensuring video quality.

2. Service Concepts and Important Considerations in Implementation

Figure 1 shows an overview of the service processing in Open V-Live. A user accesses a Website of a CP and goes

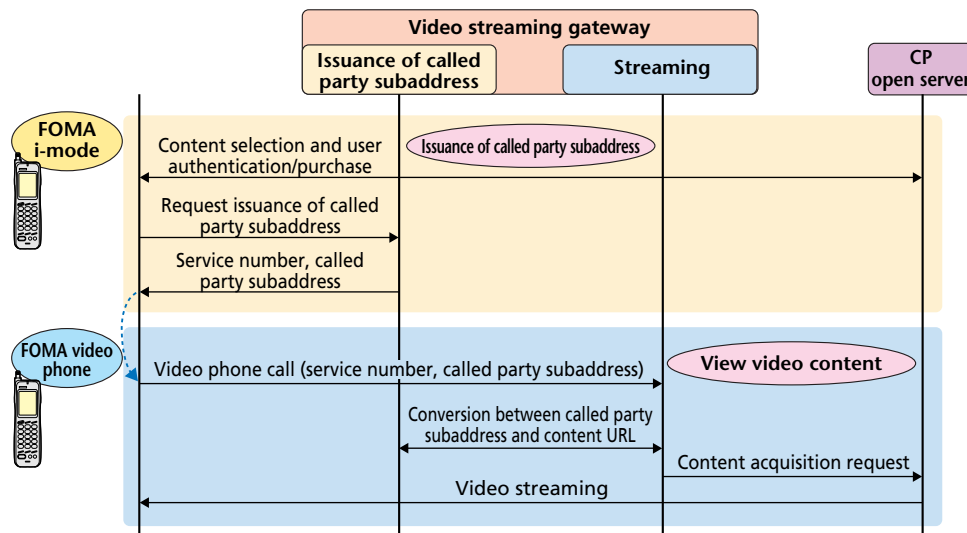


Figure 1 Overview of Open V-Live service processing

through the procedure for user authentication and/or purchase with the CP, ensuring that the user is allowed to view the video content provided on that Website. The user then requests a called party subaddress to be issued by the server for the desired content. The user then makes a video phone call to the video streaming gateway by using the service number and called party subaddress issued to view the video content. Whereas predetermined called party subaddresses are used in conventional V-Live, Open V-Live adopts dynamically issued called party subaddresses that are generated automatically each time a user requests the viewing of content and valid only for a fixed period of time.

Open V-Live eliminates the need for signing contracts with and being examined by DoCoMo as required in the past, and allows CPs to provide unique content via their own equipment and set up methods of payment independently of DoCoMo. This development raises expectations for the possible lifting of entry barriers for CPs, which will lead to significant improvements in both the quality and quantity of content that can be provided via V-Live and, consequently, the demand for V-Live services can be expected to increase dramatically.

We thus conducted development by focusing on expanding the capacity of simultaneous users connecting to video-stream-

ing gateway and improving equipment performance to address the expected rise in demand for V-Live services, implementing an open interface that can be used by a diverse range of CP servers, issuing and managing highly reliable and secure called party subaddresses, and guaranteeing high streaming video quality that is not affected by the varying capacity and performance levels of the CP's equipment.

3. System Architecture and Efforts toward more Simultaneous Connections and Higher Performance

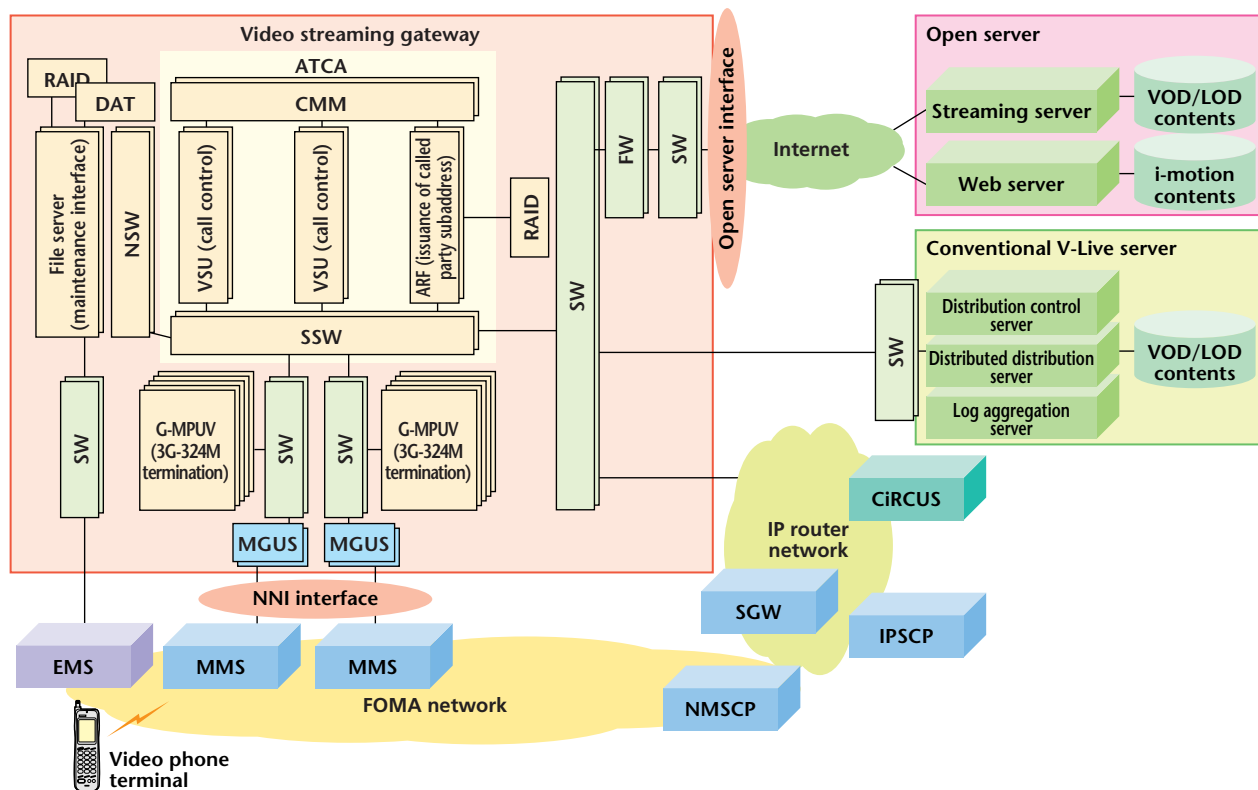
Figure 2 shows the system and network architecture of the video streaming gateway supporting Open V-Live. The equipment design of the conventional video streaming gateway can accommodate trends in user demand in details. For the interface with the core network, such design adopts a conventional User Network Interface (UNI) for small capacity that can be quickly adapted. Since Open V-Live requires greater processing capacity as mentioned earlier, however, Network Network Interface (NNI) for large capacity is adopted to connect the core network for this service.

In the video streaming gateway that supports Open V-Live, various applications for handling the synchronization between audio and video (extracting audio and video from Real-time Transport Protocol (RTP)^{*1}[2] packets and the time-series arrangement of audio and video using timestamps), which in the past were run on a general-purpose server, are moved to a dedicated Gateway-Multimedia Processing Unit for Video (G-MPUV) devices and integrated with the multiplexing processing of audio and video data using H.223^{*2}[3] to improve performance. Moreover, a Digital Signal Processor (DSP)^{*3} now han-

*1 RTP: A communication protocol for distributing audio/video streaming data in real time.

*2 H.223: A multiplexing protocol for low bit rate multimedia communication. It multiplexes video data compressed using MPEG-4 (see *14), audio data compressed by AMR (see *15), and control messages for multimedia communication.

*3 DSP: A processor specialized in the processing of audio, video, and other digital signals.



CMM (Chassis Management Module): A module that manages the entire chassis including motherboard slots.
 DAT (Digital Audio Tape unit): A device for recording data on magnetic tape, developed for the purpose of storing digitized audio.
 NSW (Node inside layer 2 SWitch): A layer 2 switch used to connect between a file server, SSW, CMM, etc.
 RAID (Redundant Arrays of Inexpensive Disks): A device that manages multiple hard disks at the same time.
 SSW (Shelf SWitch): A switch for connecting the content of each module within a shelf.
 VOD/LOD (Video On Demand/Live On Demand) contents: VOD refers to saved video content; LOD refers to live video content.

Figure 2 System and network architecture of video streaming gateway

dles multiplexing processing, which used to be implemented in software, to accelerate processing speed and reduce the equipment cost per channel.

In the general-purpose server, section that used to handle call control and session control with streaming servers, we adopted hardware conforming to Advanced Telecom Computing Architecture (ATCA), an industry standard established by the PCI Industrial Computer Manufacturers Group (PICMG)^{*4} for next-generation communication devices targeted for common communication carriers, and a Carrier-Grade Linux Operating System (OS)^{*5}. This arrangement allows the sharing of hardware and middleware among serving/gateway General packet radio service Support Node (xGSN)^{*6}, IP Service Control Point (IPSCP)^{*7}, and other devices that adopt ATCA to reduce

equipment and software development costs, guarantee the quality of applications including middleware, and reduce the maintenance costs by unifying maintenance procedures. The target devices include Video Streaming Unit (VSU) that perform such call control processing as call connection/disconnection by using ISDN User Part (ISUP)^{*8} signals, and the Address Registration Function (ARF) that issues called party subaddresses and maintains information related to the Uniform Resource Locators (URL) of corresponding content. Additional devices include a file server (operation device) that connects to an Element Management System (EMS) and a Media processing Gateway Unit for Synchronous transfer mode (MGUS), which is used for communicating user data to the core network. Dual redundant configurations are also utilized for each device

*4 PICMG: An industry group that formulates bus and board standards that can be used in industrial circuit boards and systems based on the Peripheral Component Interconnect (PCI) standard for inexpensive input/output buses used for personal computers.

*5 Linux OS: An open-source Unix-type OS that can be freely redistributed under GNU Public License (GPL).

*6 xGSN: A packet communication processing device in the FOMA network, equipped with both the Serving General packet radio service Support Node (SGSN) function and the Gateway General packet radio service Support Node (GGSN) function defined by the 3GPP.

*7 IPSCP: A device in the FOMA network/PDC network that integrates the functions of both NMSCP that accumulates subscriber information and Mobile Multimedia service Infrastructure (M²In) that provides mobile multimedia services, and manages advanced services within future core networks.

to enhance reliability.

Other related systems include CiRCUS (treasure Casket of i-mode service, high Reliability platform for CUSomer)^{*9}, which handles user requests for the issuance of called party sub-addresses, Mobile Multimedia switching System (MMS), which are the switchboards of the FOMA network, Signaling GateWay (SGW) that handle Signaling System No.7 (SS7) via IP, New Mobile Service Control Point (NMSCP) or IPSCP nodes that convert V-Live service numbers to routing numbers to the video streaming gateway, as well as the V-Live servers that provide conventional V-Live content, and other open servers.

4. Support for New Open Server Interface

The gateway supports most major servers as open streaming servers, including the Helix Mobile Server^{TM*10} that is gradually becoming the mainstream open platform for the creation and distribution of source digital media content. Some of these servers have specifications conforming to official standards, but certain individual vendors utilize implementations not compliant with standards, where the documentation of detailed specifications is typically not disclosed. For this reason, we adopted a method whereby each server can be processed independently, so that the other servers are affected as little as possible in case of the repeated trial-and-error verification and review of a process already deployed on one server. The processing by the conventional V-Live server is reused to generate the multiplexed streaming data from audio and video packets received from a streaming server adopting 3G-324M^{*11}[4], except for the processing of certain detailed parameters used to control sessions with the server [5].

Furthermore, the gateway supports most major open Web servers, such as the free and open source Apache^{TM*12}, which is currently the most widely used Web server in the world. Mobile MP4^{*13}, which supports the same type of pseudo-streaming as used by i-motion, is the file format used for targeted video content. The server acquires such files via HTTP, extracts Moving Picture Experts Group phase 4 (MPEG-4)^{*14} video and Adaptive Multi Rate (AMR)^{*15} audio data from the files, synchronizes the

data in time series based on timestamps calculated from the audio and video track control information, and then generates streaming data in which audio and video are multiplexed.

For the acquisition of Mobile MP4 files, the gateway supports the Range header of HTTP1.1, which allows a partial downloading, to keep track of file acquisition, thus making the processing more efficient in case of acquisition failure by acquiring data only from the point of failure.

Figure 3 shows an overview of processing until the generation of multiplexed audio and video streaming data from audio and video packets received from a streaming server, as well as from Mobile MP4 files received from a Web server. Audio and video packets received from a streaming server are sorted according to sequence numbers. The format of audio payload data is converted into audio frames of the size used in video phone terminals, and video payload data is gathered in the Video Object Plane (VOP), which is the unit of video frame composition, and then divided into the size used for mobile terminals. Audio and video in Mobile MP4 files received from a Web server are divided into audio frames and video data of the size used for video phone terminals according to individual items of track control information, to which timestamps calculated from the control information are attached. The divided audio frames and video data from the streaming server or Web server are sorted in time series according to the timestamps and transmitted to the video phone terminal as a audio/video stream using H.223.

5. Technologies for Managing the Issuance of Called Party Subaddresses

In order to place importance on reducing equipment cost and allow for small-scale expansion, we set an upper limit on the number of called party subaddresses that can be issued by one ARF server. Since only a small amount of information is to be managed, we decided to use both memory and disk files for storage, instead of a database. In this way, it is expected that security attacks targeting databases can also be avoided.

In certain cases, multiple ARF servers may be involved in issuing called party subaddresses. To handle such cases, we

*8 ISUP: Part of the SS7 common line signal method, which is a communication protocol used to control subscription telephone networks through connection processing in ISDN networks.

*9 CiRCUS: An i-mode gateway system.

*10 Helix Mobile ServerTM: Trademark of RealNetworks, Inc.

*11 3G-324M: An audio-visual communication standard adopted in IMT-2000 video-phones. It uses H.223 as the multiplexing method of audio, video and control signals, H.245 as the communication control method, H.263 or MPEG-4 (see *14) as

the video encoding scheme (MPEG-4 only in V-Live), and AMR (see *15) as the audio encoding scheme.

*12 ApacheTM: A Web server developed by and a Trademark of Apache Software Foundation.

*13 Mobile MP4: A video file format for mobile terminals. It uses MPEG-4 (see *14) as the video encoding scheme and AMR (see *15) as the audio encoding scheme.

*14 MPEG-4: A coding scheme for moving-picture data used for delivering video over relatively slow communication circuits as in mobile terminals.

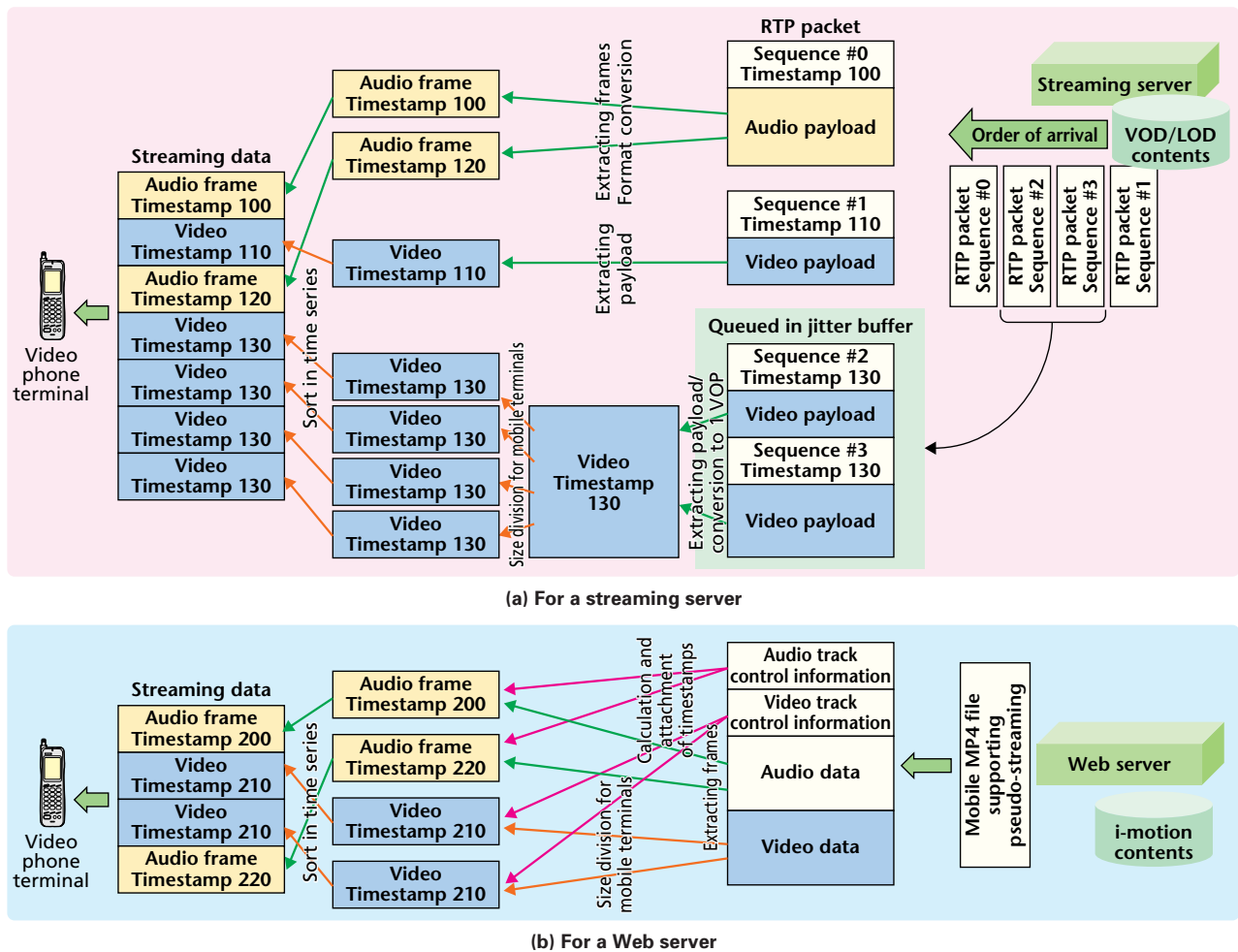


Figure 3 Overview of audio/video streaming data generation processing

decided to add a server identifier to each called party subaddress. A called party subaddress is an array of random numbers consisting of a random number of digits within a specified range, so that information other than the server identifier cannot be easily figured out. Moreover, each issued called party subaddress is assigned an expiration date to improve the security against illegal access by those other than the user who request issuance. A sequential number is appended to specific digits of an issued number to speed up search operations. Moreover, a Transmission Control Protocol (TCP) session will not be established with the mobile terminal requesting issuance if a number cannot be issued due to congestion on the server or access regulations, and if the mobile terminal does not support V-Live, an

error is reported about called party subaddresses not being issued in order to eliminate wasteful packet fees.

6. Technologies to Guarantee Video Quality

Figure 4 shows the processing sequence from a video phone call to initiate the generation of audio/video streaming data until disconnection of the communication.

The user of a video phone terminal uses the issued service number and called party subaddress to call the video streaming gateway. An NMSCP or IPSCP converts the service number into a routing number and connects to the video streaming gateway. The video streaming gateway uses the content URL converted from the issued called party subaddress and sends a

*15 AMR: An audio encoding scheme used for Third-Generation mobile communication formulated by 3GPP. It allows changing the transfer rate in a flexible manner according to the line type and conditions.

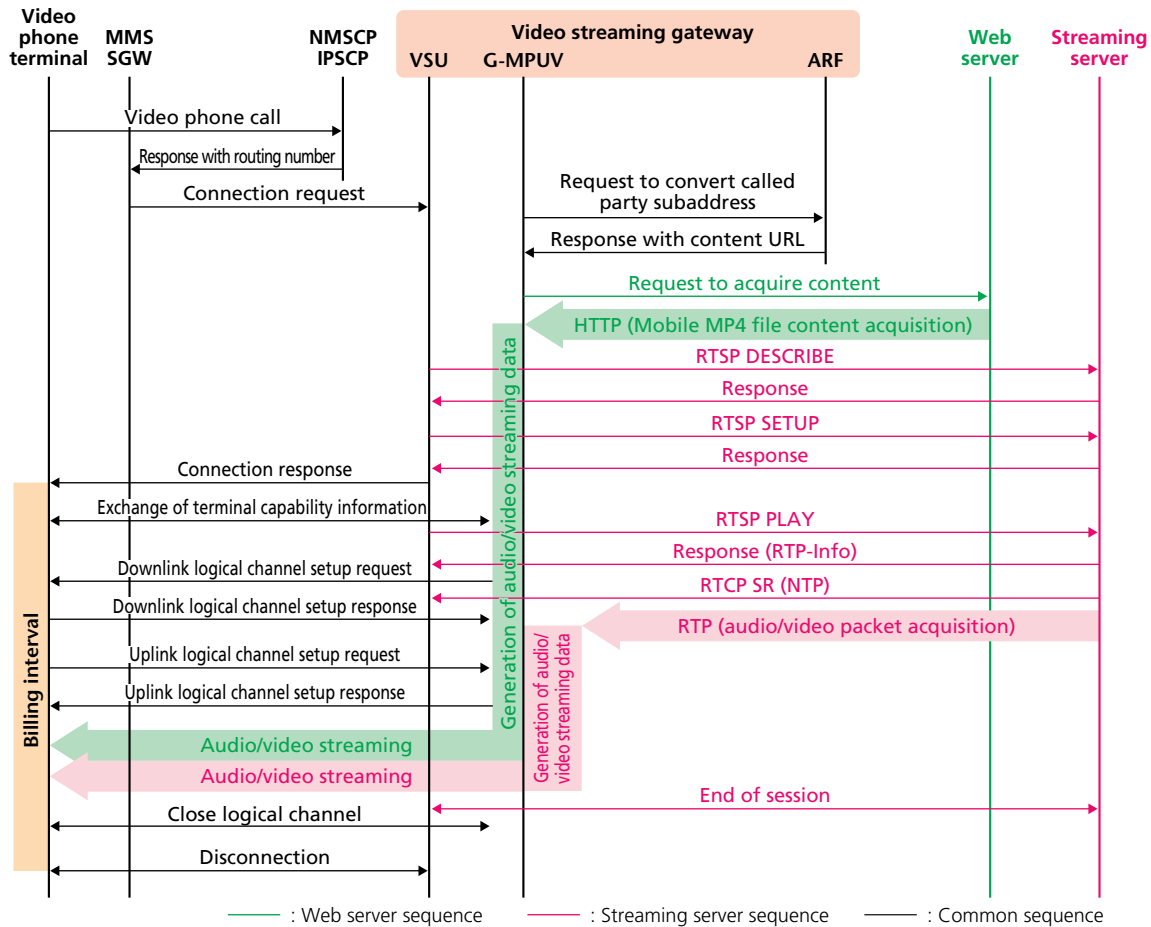


Figure 4 Processing sequence from video telephone call to initiating generation of audio/video streaming data until disconnection of communication

request to the corresponding Web server or streaming server to acquire the desired content. In the meantime, the gateway exchanges information about the mobile terminal capability for performing audio/video streaming and sets up a logical channel with the video phone terminal. When Mobile MP4 files are received from the Web server or audio and video packets are received from the streaming server, the files or packets are converted into audio/video streaming data and sent to the video phone terminal. When transmission is completed, the session with the server terminates, the logical channel with the video phone terminal is closed, and communication disconnected.

Conventionally, the normal condition of the server side was judged based on whether the Real Time Streaming Protocol (RTSP)^{*16} [6] (DESCRIBE) response to a content acquisition

request was normal, whereupon a communication path would be established and billing started. Conversely, communication is performed via the Internet for open streaming servers, and billing is not started until all processing has been completed up to the session establishment request RTSP (SETUP) response, thus further improving reliability. Web servers are judged to be normal when no problems are encountered in the content information of the HTTP header and Mobile MP4 header information. If this cannot be confirmed, the reason for disconnection is reported to the video phone terminal and the user can easily check the problematic event by reading a message displayed on the video phone terminal screen.

The amount of buffered audio/video data received from the server side is tuned to match the minimum amount required to

*16 RTSP: A communication protocol that controls the playback/stopping of content on streaming servers.

avoid any underflow^{*17} of streaming data in the network to the video phone terminal, thus ensuring the real-time delivery of live video. To avoid the impact of communication delays on the tuned buffer size due to conditions on the video phone terminal side, the audio and video information acquisition start request, RTSP (PLAY), is not issued until after the terminal capability is exchanged with the video phone terminal. Moreover, a logical channel setup response is not sent to the video phone terminal until immediately prior to the start of sending streaming data in which audio and video are multiplexed. This avoids intervals where video is not displayed after the completion of negotiation with the video phone terminal.

To address possible changes in the arrival order of audio and video packet data, the jitter^{*18} control that queues the order shortens the queuing time if the remaining amount of data in the initial storage buffer waiting for transmission is small, thus avoiding video feed disturbances due to an underflow of streaming data. Conversely, if the amount of data waiting to be transmitted is large, the control only waits for a preset time to avoid changes in the packet order as much as possible. In this way, a mechanism that can guarantee both real-time delivery and good video quality is achieved.

Conventionally, the starting point of synchronization between audio and video was at the reception of Network Time Protocol (NTP)^{*19} timestamps contained in a RTP Control Protocol (RTCP^{*20} Sender Report (SR)), which reports conditions on the transmission side. However, this process tended to delay the start of synchronization because RTCP (SR) did not arrive immediately. The video streaming gateway that supports Open V-Live shortens the delay before the start of synchronization by synchronizing using RTP timestamps embedded in the RTP-Information (RTP-Info) reported in response to the streaming start request RTSP (PLAY) and sequence numbers.

7. Conclusion

This article described the unique efforts made for and the technologies applied to the video streaming gateway that supports Open V-Live. In the future, we intend to study and evalu-

ate technologies that will allow us to increase and improve serviceability and convenience, including a wider range of billing options that can be supported on the telecommunication carrier side, such as billing for content according to viewing time, and further reducing the connection time by decreasing the negotiation time with the video phone side.

REFERENCES

- [1] K. Okubo et. al: "Video-Streaming Gateway for FOMA V Live service," NTT DoCoMo Technical Journal, Vol. 11, No. 3, pp.42-47, Oct. 2003 (In Japanese).
- [2] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson: "RTP: A Transport Protocol for Real-Time Applications," RFC1889, Jan. 1996.
- [3] ITU-T Recommendation H.223: "Multiplexing Protocol for Low Bit Rate Mobile Multimedia Communication," Jul. 2001.
- [4] 3GPP, 3G TS 26. 110: "General Description," Jun. 2002.
- [5] H. Aoyama et. al: "Latest Technology for Video-Streaming Gateway of M-stage V Live —Assuring Video Quality and Usability—," NTT DoCoMo Technical Journal, Vol. 6, No. 4, pp.13-18, Mar. 2005.
- [6] H. Schulzrinne, A. Rao and R. Lanphier: "Real Time Streaming Protocol (RTSP)," RFC2326, Apr. 1998.

*17 Underflow: Conditions where the streaming data throughput falls below the level where video quality is guaranteed (64 kbit/s). When this occurs, the amount of data received becomes insufficient with noticeable disturbance in the displayed video.

*18 Jitter: Fluctuations and deviations along the time axis in delay time in signals and similar. In this article, jitter refers to changes in the arrival order of audio and video packet data.

*19 NTP: A communication protocol for correcting the internal clock of computers via networks.

*20 RTCP: A communication protocol for exchanging data reception conditions from a streaming server and controlling the transmission rate. It is used in combination with RTP, which is a communication protocol for distributing audio and video streaming data in real time.