Special Article on Mobile Multimedia Signal Processing Technologies

Multimedia Delivery Technology

This article provides an explanation of multimedia delivery technologies for music and video, with reference to multimedia transport technology, delivery control technology and MPEG-7, which describes the metadata of multimedia contents. Multimedia delivery is expected to become the main application in next-generation mobile networks based on the International Mobile Telecommunications-2000 (IMT-2000) standard, in which all of those technologies would have great importance.

Takeshi Yoshimura, Toshiro Kawahara, Shun-ichi Sekiguchi and Minoru Etoh

1. Introduction

Since i-mode enabled mobile phones to access the Internet, users have been able to access various types of content on the Internet with the use of mobile phones. Although applications are primarily limited to e-mail exchange and Web browsing for the time being, applications to access music, video and other types of multimedia content are expected to become dominant when high-speed wireless access is enabled under IMT-2000.

This article introduces and explains the technologies that enable multimedia delivery as such. Chapter 2 describes the transport technologies that enable the transmission of multimedia content over the Internet. Chapter 3 introduces various technologies controlling multimedia sessions, and Chapter 4 provides a detailed commentary on Moving Pictures Experts Group-7 (MPEG-7) — the international standard of metadata for multimedia content description — as well as its applications to multimedia delivery. Chapter 5 provides a summary of this article.

2. Multimedia Transport Technology

2.1 Real-time Transport

E-mail and Web browsing are tolerant to transmission delay to a certain extent, but they require highly reliable packet transmission that involves no packet loss. The Transmission Control Protocol (TCP) is adopted by such applications, providing highly reliable packet transmission based on retransmission.

In contrast, the transmission of multimedia content is tolerant to packet loss to a certain extent, but requires packet transmission with low transmission delay. To these applications, User Datagram Protocol (UDP) and Real-time Transport Protocol (RTP) [1] are normally applied, in place of TCP. UDP is a simple protocol that merely delivers packets, without guaranteeing any reliability as in the case of TCP. RTP is a protocol prescribed in Request for Comment (RFC) 1889 by the Internet Engineering Task Force (IETF) with the aim to transmit speech, video and other real-time media. The key functions of RTP are as follows.

(1) Determination of Payload Type

RTP determines the coding algorithm used for encoding the RTP payload.

(2) Assignment of Sequence Numbers

RTP provides information for the receiver to reorder data in the order that was in at the source, as packets are not necessarily transmitted in order over the Internet.

(3) Time Stamping

RTP provides information for the receiver to play back the multimedia data at the appropriate timing in the event of any fluctuations in transmission delay (jitter).

RTP is a protocol framework that does not work by itself — it is fully prescribed by a profile document that determines the payload type and the payload format mapping, and a document that describes the payload format for each application. Currently, the Audio Video Transport (AVT) Working Group (WG) at IETF is working on the standardization of the Adaptive Multi Rate (AMR) and MPEG-4 video, which are the audio and video coding technologies in IMT-2000.

RTP Control Protocol (RTCP) is prescribed in conjunction with RTP, in order to transmit RTP session information. Participants in an RTP session send RTCP packets that

include the participant-specific information on a regular basis. In addition, the number of lost packets, jitters and other information concerning reception quality can be passed on by the RTCP receiver report. The information can be used to learn the communication conditions and control the Quality of Service (QoS), including transmission rate control (Figure 1). AVT WG at IETF is also working on the enhancement of RTCP to enable the retransmission of RTP packets based on RTCP information and the alteration of audio and video coding parameters.

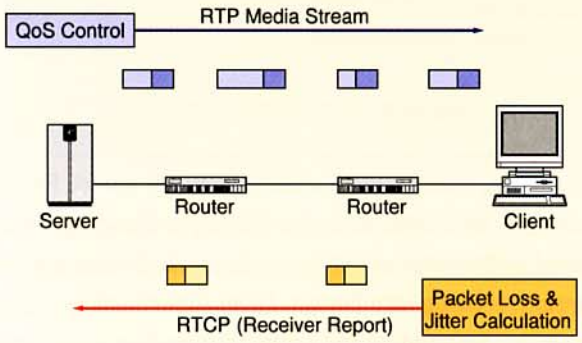

Figure 1 Multimedia Packet Transmission by RTP and RTCP

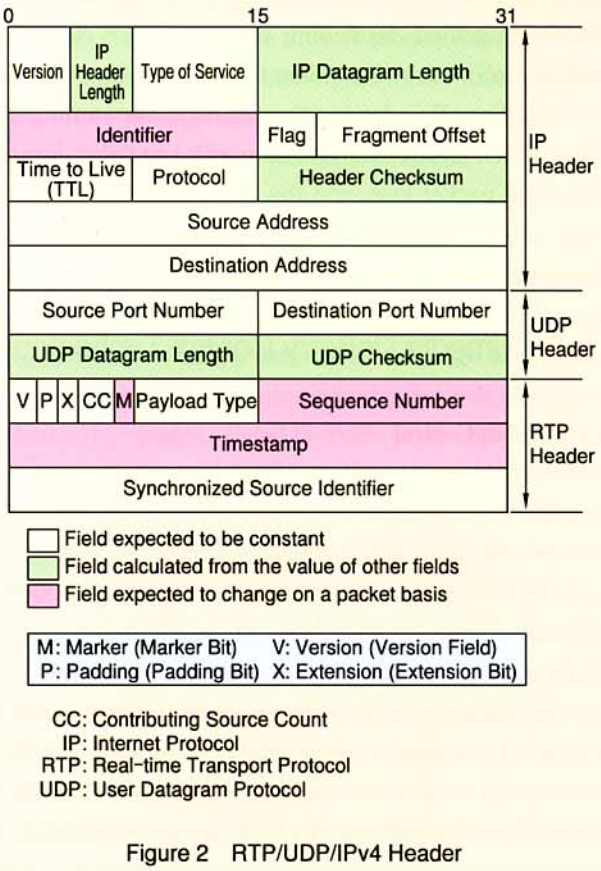

Figure 2 RTP/UDP/IPv4 Header

2.2. Robust Header Compression

The header length of an RTP packet adds up to as much as 40 bytes, including UDP and the Internet Protocol (IP) headers. In the case of voice packets, the RTP/UDP/IP header overhead is very large since the payload length per packet is only about 20-30 bytes. The header overhead is particularly problematic in cases where low-speed modems or radio links are used for communication. For this reason, much attention is paid to RTP/UDP/IP header compression, in view of the efficient use of network resources.

At present, RTP/UDP/IP header compression is prescribed in IETF RFC2508 (CRTP) [2], which assumes modems or other low-speed connections. CRTP takes advantage of the fact that most of the RTP/UDP/IP header information is constant, or that its differential is constant, and enables efficient header compression by calculating the differential from the previous header information. Using CRTP, an RTP/UDP/IP header worth 40 bytes can be compressed to 2-4 bytes.

The problem with CRTP is that it does not assume packet loss or any other transmission error, meaning that it is difficult to apply it to a mobile communications environment. As CRTP compresses the header based on the previous header information, a packet loss may cause the loss of synchronization between the sender and receiver, and make it impossible to recover the subsequent compressed packets. Since a single packet loss is likely to lead to packet discard in bursts at the receiving end, frequent packet losses, as in mobile communications environments, will substantially deteriorate transmission quality.

To resolve this issue, IETF Robust Header Compression (ROHC) WG is working on RTP header compression technologies that are robust against packet losses. The header compression technology being studied by ROHC WG enhances packet loss properties based on the following techniques.

① If there is an irregular alteration made to a specific field value, the subordinate bits of the field value, instead of the differential with the field concerned, will be transmitted several times. This enables the recovery of compressed data from packets other than the previous one.

② The checksum calculated based on the RTP/UDP/IP header before compression will be added to the compressed header, so that the receiver can check whether

the data can be recovered properly.

③ If perfect robustness against packet loss is required, the receiver will send back an Acknowledgment (ACK) to the sender; the sender will then construct the compressed header so that it can be recovered from the packet referred to in the ACK.

In addition, DoCoMo has proposed a technique to ROHC WC that reduces the number of lost packets even in the event of the loss of synchronization between the sender and receiver, by recovering the header in the reverse order after synchronization recovery. The ROHC scheme including DoCoMo's proposal will be standardized in the near future.

2.3 QoS Control

At present, the Internet is a best-effort network in that no QoS control is done in regard to throughput, delay or packet loss. However, multimedia content transmission requires some kind of QoS control, as throughput and delay requirements are strict. To meet these requirements, many QoS-aware network architectures are being proposed, out of which Integrated Service (IntServ) and Differentiated Service (DiffServ) are attracting a great deal of attention.

IntServ reserves the buffer, bandwidth and other network resources for each session, with the use of signaling protocol called Resource Reservation Protocol (RSVP) [3]. It is a network model that satisfies the QoS requirements based on the reservation of resources (Figure 3). IntServ can provide Guaranteed Service, which guarantees End-to-End delay. However, some shortcomings have been pointed out, such as RSVP's large signaling traffic and the need to control the session status at each router.

In contrast, DiffServ [4] does not control QoS at each session. Instead, it controls QoS for a predefined number of service classes, constituting a scalable network model. Figure 4 illustrates QoS control under DiffServ. The edge router, located at the edge of DiffServ's network domain (DiffServ Domain), refers to the Service Level Agreement (SLA) with the sender host and/or the header information of each packet, and determines the service class of the packet concerned. Then, it marks the service class to the Type of Service (TOS) field in the IPv4 header or the Traffic Class field in the IPv6 header (collectively referred to as the "DS field"). In the DS domain, sessions in the same service class are aggregated, and a different forwarding process called Per Hop Behavior (PHB) is applied depending on the service class. This enables QoS control depending the service class. Currently, IETF DiffServ WG defines the Expedite Forwarding (EF) Class, which ensures transmission with low delay, low jitter and small packet loss, and the Assured Forwarding (AF) class, which consists of 4 priority classes and 3 drop classes for each priority class.

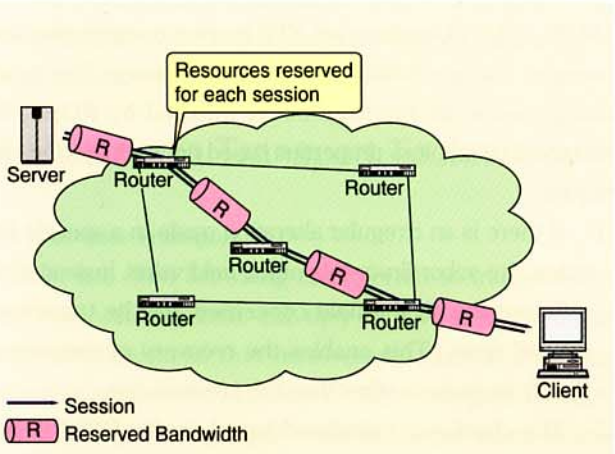

Figure 3 QoS Control by IntServ

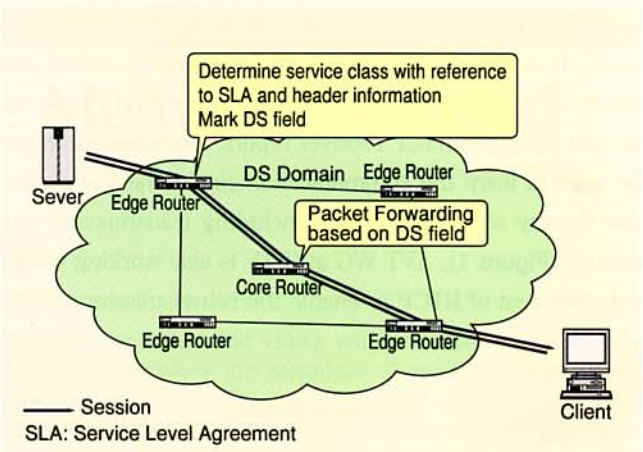

Figure 4 QoS Control by DiffServ

3. Multimedia Delivery Control Technology

3.1 Session Control

When listening to music and watching videos-or enjoying any other multimedia content-it would be preferable if various user-requests could be processed real-time, such as midway playback, pause, fast forward and slow playback. IETF RFC2326 defines the Real Time Streaming Protocol (RTSP) [5] for controlling multimedia sessions as such.

RTSP is a protocol for a remote client to control the transmission of multimedia contents from the server. The RTSP client sends request messages describing a method to the server. Then, the server sends back a response message and executes a process in compliance with the method. Table 1

Table 1 List of RTSP Methods

Method	Description
SETUP	Reserve resources, Start session
PLAY	Start playback
PAUSE	Stop temporarily
TEARDOWN	Release resources, End session
RECORD	Start recording
DESCRIBE	Acquire description
ANNOUNCE	Update description (e.g., Add media during session)
OPTIONS	Enquire available options
REDIRECTION	Order connection with other servers (e.g., Load distribution, Content transfer)
GET_PARAMETER	Acquire parameters
SET_PARAMETER	Setup parameters

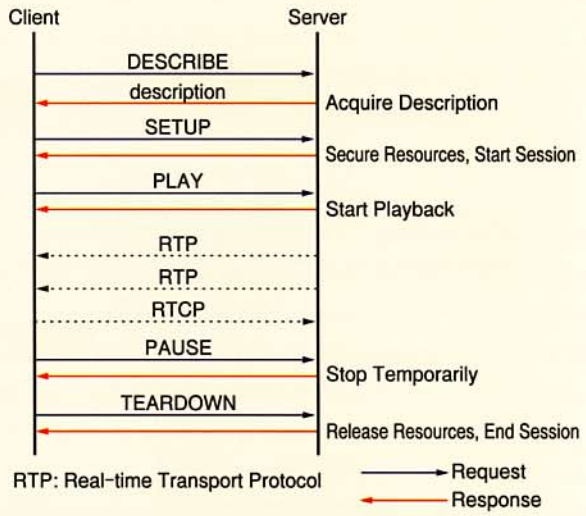

Figure 5 Example of RTSP Sequence

shows the methods prescribed in RFC2326. In addition to these methods, more detailed requests, such as the playback time and the playback speed, can be included in the message header.

Figure 5 illustrates an example of an RTSP sequence. If the client sends a request message specifying the DESCRIBE method, the server sends back the description of a session relating to the media concerned, with the use of a Session Description Protocol (SDP) [6]. If the client sends a SETUP message, the server will reserve the resources and start the session at the same time. A subsequent PLAY message will start the transmission of the media data. With the TEARDOWN message in the end, the server's resources will be released and the session will come to an end.

3.2 Media Description

Multimedia delivery penetrated the market with the help of Hyper Text Markup Language (HTML) over the Internet. The importance of media description language to integrate audiovisual media for presentation purposes will increase in the future. Languages related to media presentation description over the Internet, which are standardized by the World Wide Web Consortium (W3C), are evolving into highly flexible and scalable languages, based on the Extensible Markup Language (XML) [7]. The replacement of HTML by XHTML, which is based on XML, enables media presentation functions that are not inherent in HTML by supplementing them with other languages. In terms of multimedia, the important language is Synchronized Multimedia Integration Language (SMIL) [8], which can prescribe the synchronization between content media and describe the content attributes that are available depending on the content viewing/listening environment. While the SMIL1.0 recommendation is already being applied in practice, the next version (SMIL2.0) will have XHTML integration and mobile subset features.

On the other hand, the International Organization for Standardization (ISO)/MPEG has standardized the Binary Format for Scene (BIFS) to describe the presentation of media objects in MPEG-4 [9]. This is based on the Virtual Reality Modeling Language (VRML), and expresses the temporal and spatial positioning of multimedia objects in the scene space in binary format. It is worth noting that it enables multimedia presentations to have user interaction. Currently, MPEG is working on a text description format that represents BIFS in XML and integrates SMIL components. MPEG-7, which is discussed in Chapter 4, is also based on XML notation, and might be integrated with the XML-based media description languages described above.

In addition, SDP [6] is drawing much attention for its attribute to negotiate media transmission and reception capabilities. SDP can be used for the exchange of media playback capabilities in the course of transmission and reception in RTSP, as mentioned in Section 3.1. Some efforts are being made to integrate SDP into the next version of SMIL, as part of the media description language.

3.3 File Format

Multimedia delivery involves storing multimedia content in the server in the form of files, and transmitting them to users upon request. This means that the format of the files stored in the server must be prescribed in the same manner as the transmission protocol. Normally, a protocol for the transmission of multimedia contents over a packet network

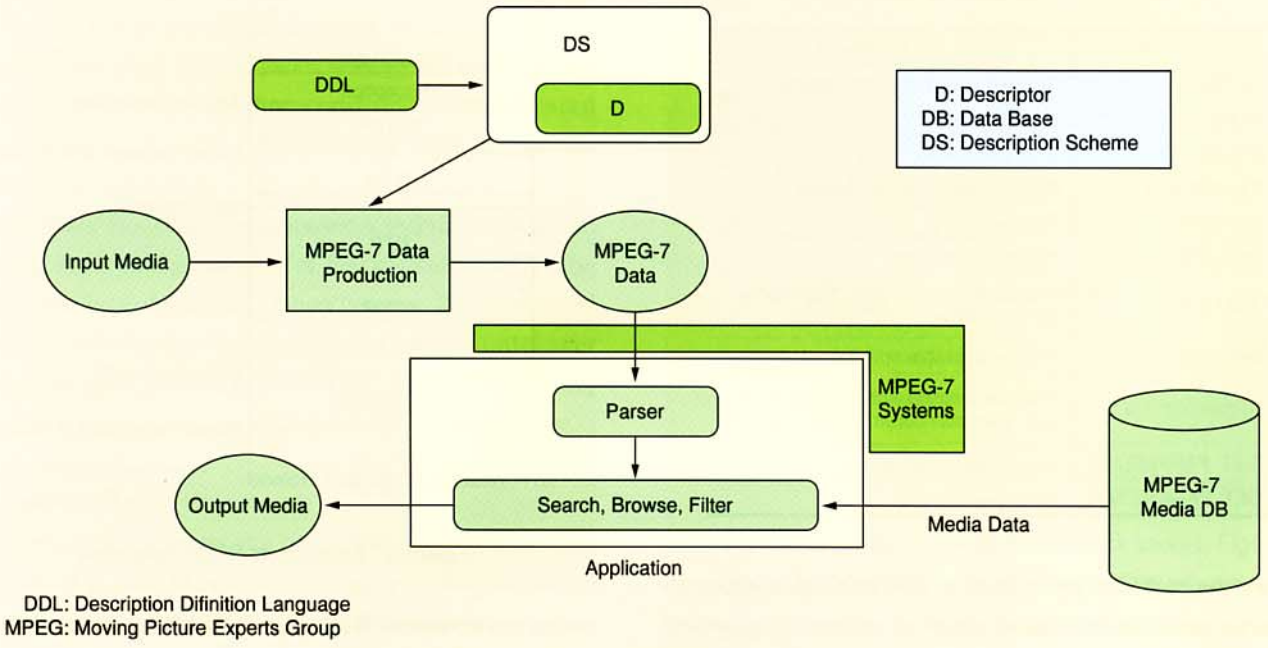

Figure 6 Scope of MPEG-7 Standard

is dedicated to transmission purposes only, as stated in Chapter 2. Video and audio data are packetized individually with additional information for synchronized playback, and then transmitted. This means that the format of the files stored in the server must:

① Keep the synchronization information of the multimedia contents,

② Be able to make packets easily, in the format prescribed by the transmission protocol, and

③ Multiplex media content (i.e., keep more than one media content in one file).

A file format standard that meets these requirements, prescribed by ISO/MPEG, is the MPEG-4 File Format in MPEG-4 System Version 2, which is designed with special consideration given to requirement(2). It stores media data in a domain called "mdat" in free format, and stores the time interval between media data, the media data size, the sample numbers, the offset from the beginning of the file and other data in a header domain called "moov."

Other than this, de facto standards with the same functions include Advanced Streaming Format (ASF) by Microsoft Corp., and QuickTime by Apple Computer Inc.

4. MPEG-7

4.1 Overview

ISO/MPEG is in the process of standardizing MPEG-7 [10], following standardizing MPEG-4. Unlike the conventional MPEG series, which served as media compression standards, MPEG-7 is a standard for a metadata format used to efficiently access multimedia content (e.g., searching, filtering, browsing). One of the milestones will be the Final Committee Draft (FCD) to be issued in March 2001, in which the technical specifications will be finalized. The standardization process is due to be completed in September 2001. By associating the metadata standardized by MPEG-7 with contents, users will be able to access contents dispersed across various platforms by following a standard set of procedures. This will help boost the utility value of the content.

Figure 6 illustrates the components of standardization. MPEG-7 basically treats metadata as instances in XML document form, and incorporates Description Definition Language (DDL) as the language that defines each metadata format (or syntax). DDL is based on the XML schema that W3C is working on for standardization.

In regard to metadata types, Descriptor (D) and Description Scheme (DS) will be included in the standard. D is mainly for describing the features of the media that can be identified at the signal level, whereas DS describes the temporal and spatial structure of the media based on a combination of various Ds or DSs. For example, D can be applied to a color histogram of an image, and DS to a structure description of a video program (definition of scenes and shots). In general, D is designed for the retrieval or the categorization of similar contents. DS is expected to be used for efficiently accessing the required portion of long-hour contents (ran-

dom access) and customizing the presentation of the content depending on the user's taste and viewing/listening environment (e.g., terminal performance, network QoS).

Normative system components required to operate D and DS in practice will be included in the Systems part of the standard. It may include binary expression for efficient transmission of D/DS, and provisions for transmitting D/DS in access units.

4.2 Key Technologies

MPEG-7 may have extremely diverse description tools, each of which is standardized as D or DS. The detailed definitions of D and DS may be referred to in the reference [10]. The key technologies that make MPEG-7 practical are the acquisition technology for D/DS descriptions, and the application technology for D/DS in practice. As both of these technologies are beyond the scope of the standard, the choice will depend on the user. For D/DS acquisition, automatic extraction of the media feature and structure is essential, in order to add metadata to a massive volume of content. For example, the extraction of the video structure requires such processes as the automatic extraction of the key frames in the video [11]. Meanwhile, an application system needs to be developed to efficiently utilize the acquired D/DS. For the image search system, a D/DS matching process algorithm that meets the system requirements is necessary. It should be remembered that MPEG-7 plays a substantial role in realizing a one-source, multipurpose environment as well. Important technologies include automatic adaptation and delivery of content to different infrastructures, using the added MPEG-7 description as hint information.

4.3 Multimedia Delivery Applications

In a mobile environment, MPEG-7 could be applied to services that efficiently deliver multimedia content despite the restricted mobile terminal's media display and playback capabilities and the limited network resources. Figure 7 illustrates an application example. In this system, MPEG-7 metadata is used to produce a summary of the multimedia content, so that it can be transmitted to users based on push technology, giving an idea what the content is about to users. Considering that location information applications are expected to become powerful services for mobile terminals in the future, this system could be used to send appropriate content to the terminal according to its location information, through push technology. A user can watch and listen to portions of the content customized to satisfy his/her tastes, with reference to the user preference set for the terminal in advance, and the MPEG-7 description of rich content that provides the summarized version of the content. If multimedia content can be provided in various stages in this manner, both users and content providers will be able to provide and gather the information they want whenever necessary, which should promote multimedia delivery over mobile networks.

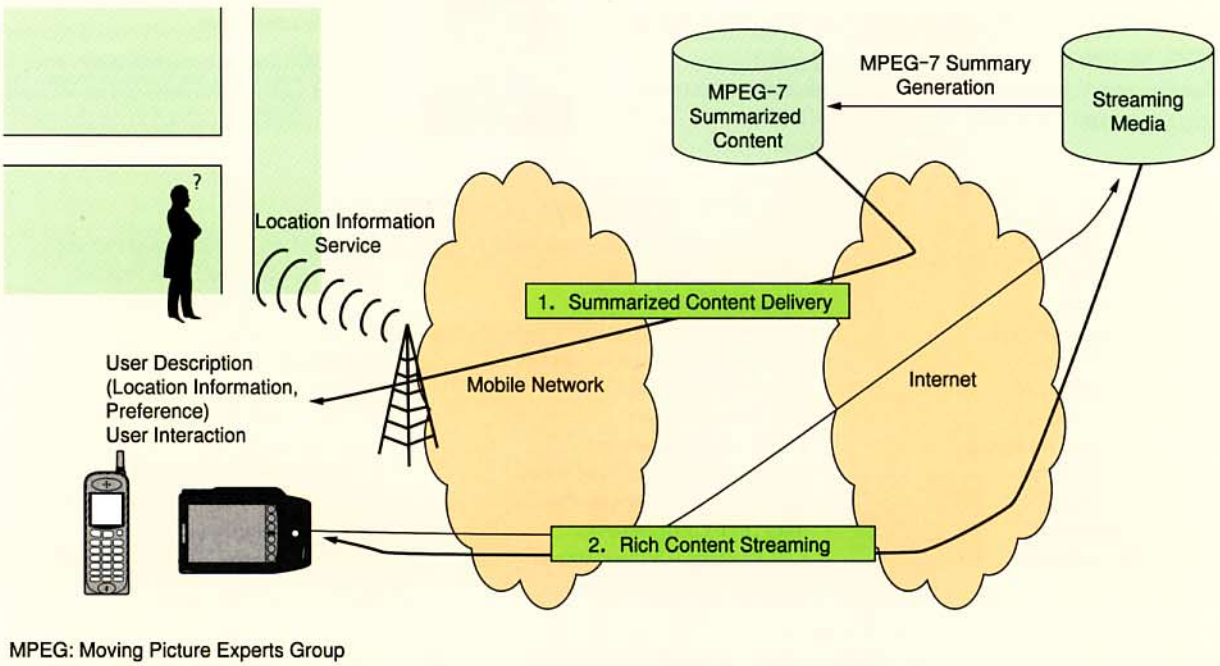

Figure 7 Example of MPEG-7 Mobile Multimedia Delivery Application

5. Conclusion

This article served as an introduction to transport and control technologies for multimedia delivery and MPEG-7, and provided a brief explanation of standardization trends. It described the multimedia transport technologies with reference to RTP (which aims at real-time multimedia transmission), robust header compression (which reduces the overhead of the header), and QoS control technologies such as DiffServ. It explained multimedia delivery control technologies with reference to RTSP (which enables remote session control), SMIL (which provides multimedia presentation functions) and the format for storing files in a server. In the end, it provided an overview of the standardization trends in regard to MPEG-7, which prescribes the metadata format, in addition to its applications to multimedia delivery. As all of these technologies are important to make multimedia delivery possible in a mobile environment, much attention will have to be paid to future trends.

References

[1] H.Schulzrinne, S.Casner, R.Frederick, and V.Jacobson: RTP: A Transport Protocol for Real-Time Applications, RFC1889, Jan.1996.

[2] S.Casner and V.Jacobson: Compressing IP/UDP/RTP headers for low-speed serial links, RFC2508, Feb.1999.

[3] R.Braden, L.Zhang, S.Berson, S.Herzog, and S.Jamin: Resource ReSerVation Protocol (RSVP), RFC2205, Sep.1997.

[4] S.Blake, D.Black, M.Carlson, E.Davies, Z.Wang, and W.Weiss: An Architecture for Differentiated Services, RFC2475, Dec.1998.

[5] H.Schulzrinne, A.Rao, and R.Lanphier: Real Time Streaming Protocol, RFC2326, Apr.1998.

[6] M.Handley and V.Jacobson: SDP: Session Description Protocol, RFC2327, Apr.1998.

[7] Extensible Markup Language (XML) 1.0, W3C Recommendation, 6 Oct.2000 (http://www.w3.org/TR/2000/REC-xml-20001006).

[8] Synchronized Multimedia Integration Language (SMIL) 1.0 Specification, W3C Recommendation 15 June 1998 (http://www.w3.org/TR/REC-smil/).

[9] Information technology — Coding of audio-visual objects — Part 1: Systems, ISO/IEC 14491-1.

[10] Introduction to MPEG-7, ISO/IEC JTC1/SC29/WG11/N3545, July, 2000.

[11] R.Brunelli et.al, "A Survey on the Automatic Indexing of Video Data", Journal of Visual Communication and Image Represenation 10, 78-112, 1999.

Glossary

ACK: Acknowledgment
AF: Assured Forwarding
AMR: Adaptive Multi Rate
ASF: Advanced Streaming Format
AVT: Audio Video Transport
BIFS: Binary Format of Scence
CC: Contributing Source Count
DDL: Description Difinition Language
DiffServ: Differentiated Service
EF: Expedite Forwarding
FCD: Final Committee Draft
HTML: Hyper Text Markup Language
IETF: Internet Engineering Task Force
IntServ: Integrated Service
IP: Internet Protocol
ISO: International Organization for Standardization
MPEG: Moving Picture Experts Group
PHB: Pre Hop Behavior
QoS: Quality of Service
RFC: Request for Comment
ROHC: Robust Header Compression
RSVP: Resource Reservation Protocol
RTCP: RTP Control Protocol
RTP: Real-time Transport Protocol
RTSP: Real-Time Streaming Protocol
SDP: Session Description Protocol
SLA: Service Level Agreement
SMIL: Synchronized Multimedia Integration Language
TCP: Transmision Control Protocol
TOS: Type of Service
UDP: User Datagram Protocol
VRML: Virtual Reality Modeling Language
W3C: World Wide Web Consortium
WG: Working Group
XML: Extensible Markup Language