Using Metadata to Control Content Viewing on Mobile Devices

Takeshi Sasaki†, Masahito Kawamori, Jin Hatano, and Katsuhiko Kawazoe

Abstract

This article describes technology for controlling the viewing of content on cellular phones and other mobile devices. Compressed metadata transmission and metadata handling techniques are introduced as core technologies for achieving diverse content viewing modes in mobile environments.

1. Problems specific to mobile environments

Server-based broadcasting is gaining momentum and promises to make the idea of viewing TV programs stored on large-capacity hard disks commonplace. Mobile environments are no exception, and the infrastructure for server-based broadcasting to mobile terminals is under active discussion. The recent surge in popularity of personal video recorders (PVRs), which combine a DVD player/recorder and a large-capacity hard disk, is part of a larger trend toward offering users greater convenience. Storing programs on large-capacity hard disks frees viewers from time constraints, letting them enjoy the program they want to see, any time they want. Technologies supporting such services through the convergence of broadcasting and communications will become increasingly important in the future [1].

Among the technologies supporting content viewing are those for controlling content delivery and metadata control technology. Attempts to deploy these technologies in a mobile environment sometimes run up against problems specific to such an environment, such as narrow transmission bandwidth and limited terminal resources. Bandwidth limitations are on the way to being resolved through a series of steps, starting with services in the multi-Mbit/s class such as HSDPA (high-speed downlink packet access) and CDMA2000 1x EV-DO, then the so-called 3.5G services representing the halfway stage between third- and fourth-generation services, and then 4G services themselves, which will offer a bandwidth of 100 Mbit/s, which is comparable to today's FTTH (fiber to the home) service. However, there are still many problems to be overcome before 4G services are implemented and doing so may take several years. Moreover, since the mainstream business model today is designed around flat-rate packet communications services, carriers are looking for new profit structures not tied to traffic. Another factor is that today's mobile terminals are increasingly feature rich and used for a variety of network-based services, making it undesirable to consume large amounts of network resources for content viewing alone. A key issue, therefore, is how to utilize the available bandwidth efficiently.

At the same time, while advances in terminal functionality have made available more storage space and memory for use by applications, there is also a trend toward lighter, more compact devices with slimmer cases, which tends to limit the number of applications that can be installed. As with bandwidth, it is important to keep terminal resource requirements to the minimum necessary.

At NTT Cyber Solutions Laboratories, we have developed a new content viewing control technology for mobile environments utilizing metadata. This technology does more than simply control content viewing on mobile devices. By incorporating the means to exchange metadata among individual users

NTT Cyber Solutions Laboratories
† Present address: NTT East Corporation
Shinjuku-ku, Tokyo, 163-8019 Japan
E-mail: takeshi.sasaki@east.ntt.co.jp

and by facilitating the formation of communities around Internet sites, it enables diverse, new forms of content viewing in mobile environments.

2. Using segment metadata

Metadata is data about data; that is, it is information that describes the contents of data. The metadata for video content, for example, might include the title of the video, its data format, information about the service provider, and the fees. There are also many other kinds of metadata, such as that used for digital rights management and electronic program guides.

Our technology mainly uses segment metadata in XML (extensible markup language) format, which has been standardized by the TV-Anytime Forum [2] for digital broadcasting, as shown in **Fig. 1**. Segment metadata applies to small units of content data divided into individual scenes. Each unit is called a segment, and the metadata for each segment is called segment metadata.

Information about a content segment is recorded in a SegmentInformationTable element. A SegmentList element contains SegmentInformation for each segment along with a segment ID for uniquely identifying a particular segment. The metadata for individual segments includes such information as the title, a summary, a content reference identifier (CRID) identifying the applicable content, and a thumbnail image

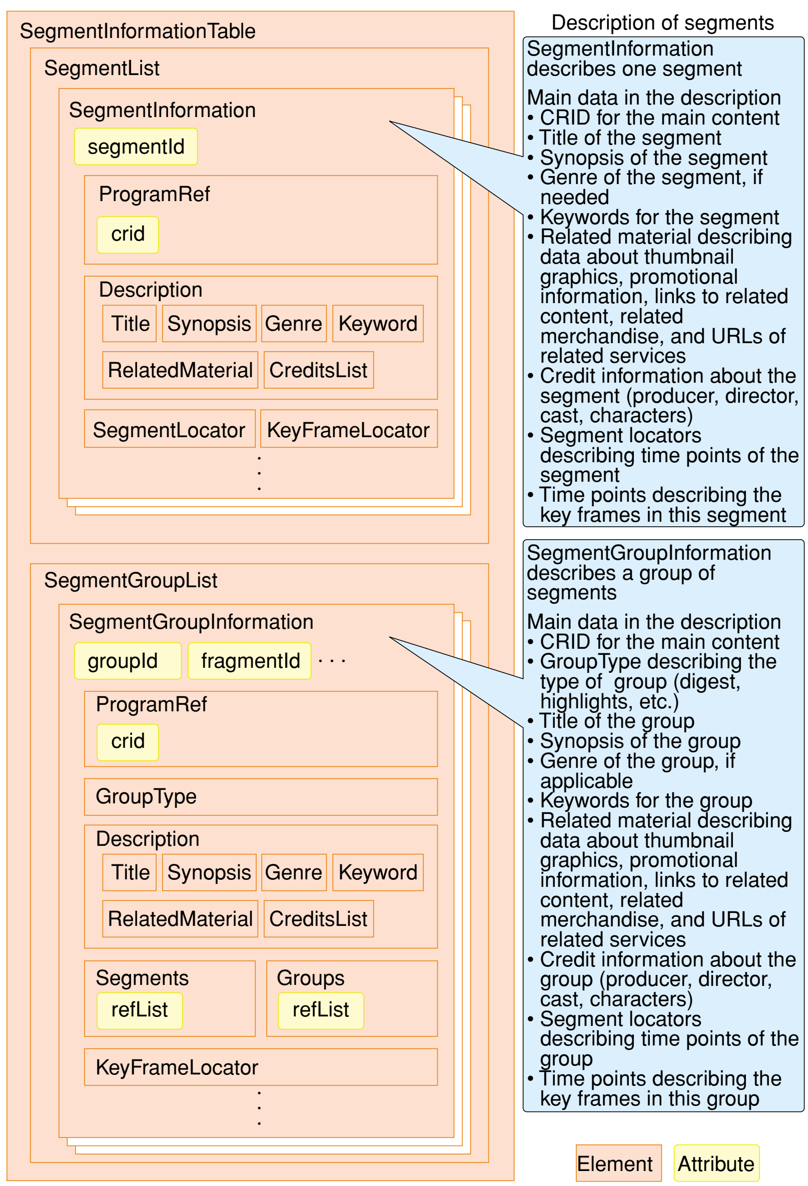

Fig.1. Structure of TV Anytime segment metadata.

providing a simple representation of the segment. These are used to create a chapter list, and enable the user to select particular segments for viewing.

A SegmentGroupList can be used to form a segment group made up of the segments indicated in SegmentList, which enables such features as digest viewing or highlight viewing. The segments belonging to a group can be designated by listing their segment IDs in the SegmentList. It is also possible to implement digest or highlight viewing across multiple programs by means of CRID designation.

3. Dealing with metadata bloat

As the services to be implemented using metadata become more advanced and diverse, more and more information is needed. Consequently, the amount of metadata describing that information expands in proportion to the sophistication of the services, which can result in large files, which in turn require more transmission bandwidth. Furthermore, high throughput and a large memory capacity will be required in the terminal that processes large-scale metadata. This is a problem that cannot be ignored in a mobile environment having limited bandwidth and terminal resources.

What is needed is an efficient way to transmit and process even large-scale metadata with limited resources. Here, we discuss a compressed metadata transmission technology as one approach to achieving this.

4. Compressed metadata transmission technology

XML-format metadata is coded in conformance with a strictly defined schema (e.g., XML Schema) enabling seamless distribution control across various platforms. Since this coding was originally defined to facilitate processing in various uses, many of its aspects have a redundant structure. These redundant parts can be compacted using a compression algorithm.

This technology is based on the BiM format (binary format for MPEG-7 data) [3] specified in MPEG-7, which standardized the description method for effectively searching for content, such as video and audio, as shown in **Fig. 2**. By converting text-format XML to binary format, we can compress metadata to around 10% of its original size. **Table 1** gives measurement results for the compression ratio based on a number of different samples of XML structures. These results show that XML structures other than real text data can generally be compressed by more than 99%. The compression ratio of real data varies with the content; however, in XML documents, where XML structures make up a large part of the total size, compressing just these structures makes a significant difference to the overall data size.

Since the same compression algorithm is used by both the sending codec and the receiving codec, it is only necessary to extract the data that needs to be conveyed and to send the compressed binary XML as is. Then, when the binary XML is accessed at the receiving end, there is no need to decode it to text format (XML format). Instead, the binary XML can be accessed directly to reference and update the data, in a manner similar to how an application at the receiving end manipulates the document object model (DOM) object that is the application programming

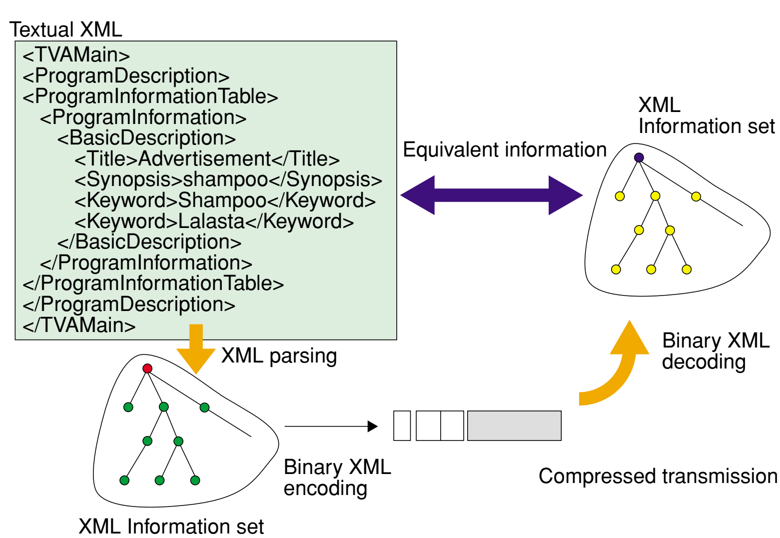

Fig. 2. Binarization of XML document.

interface for XML access.

This technology not only enables efficient transmission by means of XML compression, but also allows more information to be stored in the terminal, so it is useful for achieving effective use of terminal resources.

5. Metadata handling technology

Next, we describe a metadata handling technology for enabling a new kind of content viewing control. It can process a large amount of compressed metadata at high speed in a mobile terminal and match it with preference metadata, as explained below.

A typical mobile user is expected to view content not in long stretches but selectively, watching only items of interest over short time spans. To enable users to view only desired content of a short duration in a limited time, we must provide some means of matching content to individual preference attributes. The metadata handling technology shown in **Fig. 3** automatically generates this individual preference attributes information as metadata (called "preference metadata" here), enabling playback control tailored to the individual in combination with official metadata provided by broadcasters or content providers. The main things this technology can be used to accomplish are as follows.

(1) Assign an index to scenes of interest and play content based on it.

(2) Introduce friends and acquaintances to recommendable scenes while indicating the degree of interest and appending comments.

(3) Publish information about recommendable content on Internet sites, etc.

Sometimes the official metadata may define basic services such as digest or highlight viewing of the corresponding content, but these are viewing modes decided by the content provider and do not necessar-

Table 1. Some examples of metadata structure compression.

Textual XML example description	Textual XML data size (bytes)	Binary XML data size (bytes)	Compression rate (%)
Multimedia contents 1 (group)	768 345	542	99.9
Multimedia contents 2 (group)	3 847	24	99.4
Multimedia contents 3 (group)	148 847	255	99.8
Still region image 1	7 258	15	99.8
Still region image 2	3 208	15	99.5
Still region image 3	6 826	15	99.8
Video segment 1 (frames)	173 627	574	99.7
Video segment 2 (frames)	107 338	353	99.7
Video segment 3 (frames)	11 801	38	99.7
Video segment 4 (frames)	259 252	860	99.7
Video segment 5 (frames)	695 727	2 315	99.7
Still image + video frames 1	8 131	43	99.5
Still image + video frames 2	12 962	78	99.4
Still image + video frames 3	61 992	329	99.5

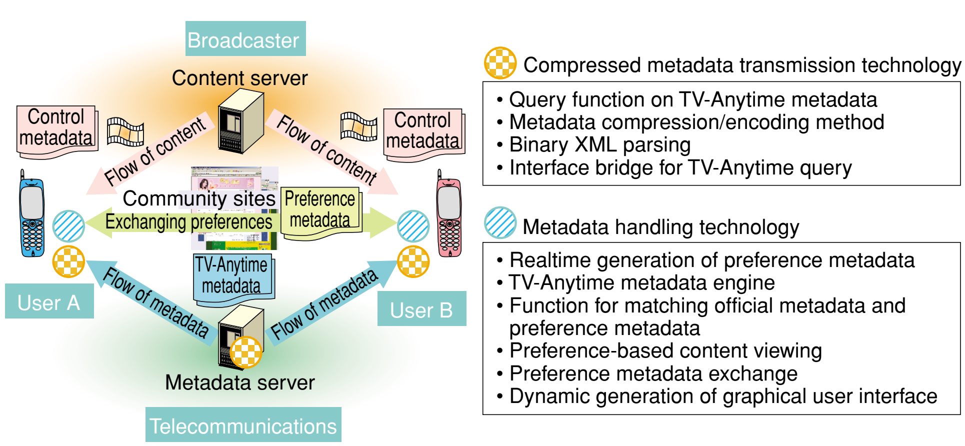

Fig. 3. Use scenario of metadata compression/handling technology.

ily present a collection of scenes that the viewer wants to see. The technology introduced here gives the viewer the means to view content as he/she desires while protecting the provider from the risk of the content being used in unauthorized ways. This is done by controlling the official metadata and preference metadata based on certain definitions, so that the user can enjoy the convenience of making private use of the content within the allowed scope, without infringing on the rights of the content provider.

In addition, preference metadata can be exchanged among individual users or published on Web sites. The recipient of such metadata can then view the same content in different ways by applying the preference metadata received from friends. One example is viewing by popularity ranking, based on multiple sets of preference metadata taking into account the degree of interest indicated by friends or their comments about the scenes being viewed. This will lead to interactive services that offer new ways of viewing content.

6. Extensive service scenarios

Web log services, commonly known as blogs, have advanced remarkably in a short time. Along with services such as social networking, they provide rapidly growing opportunities for individuals to create and post information. These phenomena attest to the strong desire of people to form communities, not only ones in real society but also virtual communities, centered on their interests. Such virtual communities tend to form naturally around sites such as blogs, where detailed information is published in specific fields and users interested in the same field gather. By exchanging preference metadata in forums of this nature, people will be able to express “points of interest” regarding various images related to a given theme, and this will give rise to new kinds of content uses.

To enable the various systems for achieving diverse services like those described above, to be incorporated into mobile terminals, a variety of technologies such as metadata compression and storage are being devised for compactness.

7. Developing new content services

NTT Cyber Solutions Laboratories is carrying out research and development on core technologies applicable to various communications-based services being undertaken by NTT, such as video delivery services and the OnQ Project of NTT DoCoMo [4].

References

[1] K. Kawazoe and S. Hayama, “Cooperative Broadcasting and Communication Technology for a New Way of Viewing TV,” NTT Technical Review, Vol. 2, No. 8, pp. 50-51, 2004.

[2] http://www.tv-anytime.org/

[3] http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm#3.5_BiM_(Binary_Format_for_MPEG-7)

[4] http://www.nttdocomo.co.jp/new/contents/05/whatnew0311.html (in Japanese).

Takeshi Sasaki

NTT East Corporation.

He received the B.E. degree in communication engineering from Tokai University, Kanagawa in 2002. He joined NTT Laboratories in 2003. He worked in research areas such as technology for transmitting MPEG-2 images encrypted by contents encoding technology and technology for distributing advertisements that performs filtering and recommendation according to viewer attributes. In addition he researched image contents control technology using metadata in mobile terminals, such as cellular phones and PDAs. He moved to NTT East Corporation in 2005.

Jin Hatano

NTT East Corporation.

He received the B.S. and M.S. degrees in computer science from Keio University, Kanagawa in 1994 and 1996, respectively. In 2001, he joined NTT Cyber Solutions Laboratories, Kanagawa and engaged in research on metadata and cooperative broadcasting and communication technology. He moved to NTT East Corporation in 2005.

Masahito Kawamori

Senior Research Engineer, NTT Cyber Solutions Laboratories.

He joined NTT Laboratories in 1989. He has worked in research areas such as artificial intelligence, language processing, and interactive agents using speech recognition. His current research area is metadata and rights language, media delivery for broadcasting and broadband communications, and the convergence of mobile and fixed communications services. Since 2000, he has been the leader of the Metadata Task Group of the Association of Radio Industries and Businesses (ARIB) Working Group for Broadcasting Systems based on a Home Server. He is also a convener of the TV-Anytime Forum, chairing the Metadata Working Group.

Katsuhiko Kawazoe

Senior Research Engineer, Supervisor, NTT Cyber Solutions Laboratories.

He received the B.E. and M.E. degrees in engineering from Waseda University, Tokyo in 1985 and 1987, respectively. Since joining NTT in 1987, he has mainly been engaged in R&D of radio communication systems, satellite communication systems, and the personal handy-phone system (PHS). His specialty is forward error correction systems. He is currently a co-chairman of ARIB Working Group for Broadcasting Systems based on a Home Server. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and received the Young Engineer Award from IEICE in 1995.