Broadcasting and Communication Convergent Network Based on MPEG-21: Design and Implementation of Multimedia Service Framework

Yongju Cho, Jae-Gon Kim, and Jin Woo Hong

In this paper, we present a practical implementation of the MPEG-21 multimedia framework for broadcasting and communication convergent services. MPEG-21 standard technology was exploited to build a convergent service framework. Using this framework, a service model and several scenarios have been successfully designed and implemented. In addition, interoperability, which is the main objective of a multimedia framework, especially in a convergent environment consisting of heterogeneous networks and various types of devices, has been addressed in detail. The experimental results show that the implemented test bed provides a next-generation multimedia service; that is, universal multimedia access (UMA), meeting the requirements of a broadcasting and communication convergent environment.

Keywords: MPEG-21, digital item, broadcasting and communication convergence, digital item adaptation, universal multimedia access.

Manuscript received Feb. 22, 2006; revised June 27, 2006.

Yongju Cho (phone: + 86 42 860 6138, email: yongjucho@etri.re.kr), Jae-Gon Kim (email: jgkim@etri.re.kr), and Jin Woo Hong (email: jwhong@etri.re.kr) are with Radio & Broadcasting Research Division, ETRI, Daejeon, Korea.

I. Introduction

Recently, digital media has had a tremendous impact on the development of the Internet, telecommunications, and broadcasting areas. The rapid popularization of the Internet, wireless communication systems, and digital broadcasting networks have led us to an epochal framework for content services with an end-to-end delivery chain of content generation, distribution, and consumption. Multimedia devices and digital TVs have accelerated the easy purchase and consumption of a vast number of media content types. These developments, however, have forced us to consider the delivery of content over heterogeneous networks to diverse types of devices. As a result, it has become clear that there is a need for a standard multimedia framework to support interoperability across a wide range of different networks and terminals in content service.

For these reasons, our research incorporates an advanced multimedia framework that satisfies the needs of convergent heterogeneous networks in an interoperable manner for universal multimedia access (UMA). The presented broadcasting and communication convergent service framework demonstrates three key aspects: 1) practical realization of a convergent-content service framework based on the MPEG-21 standard, 2) provision of UMA, and 3) provision of one-source multi-use in the framework. Our solution will eventually lead to a user ubiquitous service that enables content access from anywhere at anytime with ubiquitous networks and terminals. It may also be used in the near future as a reference model for convergent service solutions and may be applied in many different fields.

This paper consists of the following: Section II depicts an overview of the MPEG-21 standard. Section III introduces the core technologies applied to the proposed multimedia framework. Section IV describes the design and implementation of a test bed for broadcasting and communication convergence with the functionality of each functional block described in detail. Finally, with the exploration and simulation results, the strength of the proposed framework is indicated in section V.

II. Overview of MPEG-21 Standards

Commencing in late 1999, the MPEG standardization group analyzed the framework needs of users and the status of the multimedia world. Based on that analysis, the MPEG-21 (ISO/IEC 21000) began formal standardization activity at the beginning of 2000.

The work of MPEG-21 started from the initiation of the objectives of standardization. In October 1999, the objective of MPEG-21 was proclaimed in the first official document of MPEG-21 as being simply "to define multimedia contents in order to support commerce of digital contents." However, in December 1999, the goal of MPEG-21 was redefined as "to reveal common understanding about a multimedia framework in distributing digital contents with the perspectives of consumers." Therefore, the importance of the final contents users' perspectives had been emphasized. The MPEG-21 multimedia framework is related not only to the simplification of e-commerce of digital contents but also to the distribution of multimedia contents. The objective of the standardization was continuously discussed, and finally refined "to enable transparent and augmented use of multimedia resources across a wide range of networks and devices." [1], [2]. Thus, the stated objective of MPEG-21 standardization has developed and expanded from the original purpose of defining digital contents to support commerce; the current stated objective focuses simply on facilitating the effective use, generation, and distribution of digital contents.

The basic unit of transaction in the MPEG-21 multimedia framework is the digital item (DI), which packages resources along with identifiers, metadata, licenses and methods that enable interaction. Another key concept is that of a user, which stands for any entity that interacts in the MPEG-21 multimedia framework or makes use of DIs. At the most basic level, MPEG-21 can be seen as providing a framework in which one user interacts with another user and the object of that interaction is a DI.

Seven architectural key elements of MPEG-21 include digital item declaration (DID), digital item identification and description (DIID), content handling and usage, intellectual property management and protection (IPMP), terminal and network, content representation, and event reporting (ER) [1]. MPEG-21 is subsectioned into 18 parts. Parts for relevant new technologies and more investigative areas are still in progress for standardization [3].

It is important to remember that MPEG-21 differs from other MPEG standards in terms of defining a normative open framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain based on the above seven key elements, whereas others narrow their standard scope down to terminal or CODEC specification [3].

The following seven key elements are the core technologies of MPEG-21 standardization.

1. Digital Item Declaration

DID provides a uniform and flexible abstraction and interoperable schema for declaring a DI defined as a fundamental unit of distribution and transaction in the multimedia framework. The DI is a structured digital object with standard representation, identification, and metadata within the MPEG-21 framework. In other words, it is a combination of media resources, metadata, and structure. The resources are the individual assets or distributed content, and metadata describes distributed content about or pertaining to the DI as a whole or also to the individual media resources in the DI. The structure relates to the relationships between the parts of the DI–both media resources and metadata [4]. An example of a DI might be my album including media resources such as photos, audios, videos, and descriptions for them.

2. Digital Item Identification and Description

A DIID is given to each digital product for the purpose of searching and protecting copyright. In the existing standard, universal resource name (URN) is simply adapted for DII [1]; URN includes all the existing identification schemes such as digital object identifier (DOI), contents IDentification forum (cIDf), and ISBN.

While it is obviously necessary to have a standardized mechanism to describe the characteristics of media resources included in a DI MPEG-21, DIID does not define a new metadata scheme for the description. Instead it calls out existing standards such as MPEG-7, Dublin Core (DC), and resource description framework (RDF) by employing a namespace.

3. Content Handling and Usage

In content handling and usage, the interface and the delivery specifications are standardized. They are needed in creating, searching, having access to, archiving, routing, and sending content – media resource and/or metadata, throughout the life cycle of the DI between its generation and destruction.

4. Content Representation

The goal of the content representation item is to provide, adopt, or integrate content representation technologies able to efficiently represent MPEG-21 content, in a scalable and error resilient way. The content representation of the media resources should be synchronizable and multiplexed and allow interaction. This key element should allow the multimedia framework to optimally use existing and ongoing developments of media coders in MPEG [3].

The background and scope of intellectual property management and protection (IPMP), terminals and networks, and ER are well described in [5]-[9].

III. Core Technologies

Our objective is to implement a test bed to develop a multimedia framework based on the MPEG-21 standard for supporting efficient and reliable multimedia content generation, delivery, and usage in a broadcasting and communication convergent environment [10]. In the proposed test bed, some key technologies such as interoperable content authoring, content adaptation, and delivery are applied to provide UMA service that enables a user to access a large variety of content over heterogeneous networks in an interoperable manner [1].

First, interoperable content authoring aims to generate a basic unit, a DI, for distribution and transaction in the proposed multimedia framework [3]. A DI meets the DID specification to support interoperability and is used as content for the UMA service framework. Secondly, content adaptation aims to manipulate content to suit terminal or network characteristics or end users' requests by applying the MPEG-21 DIA concepts and specifications. Thirdly, content delivery supports delivery of multimedia content to any device no matter which network the device is connected to: this leads to one-source multi-use in the framework.

Finally, UMA service is the main goal of the research and combines all the technologies described previously. It supports transparent and ever-increasing use of multimedia resources by users across a wide range of networks and devices as is the objective of MPEG-21 standard.

1. Interoperable Content Authoring

According to the definition in MPEG-21, DID provides an unambiguous and flexible abstraction. Interoperable schema for defining a DI, a fundamental unit of distribution and transaction in the multimedia framework; that is, a DI can be regarded as an interoperable content that can be accessible by any device unless it does not follow the DID standard [3].

As mentioned above, the object is to define an interoperable content in the multimedia framework where various types of contents and networks coexist but lack the unity in use of contents. For this reason, a DI has been chosen to achieve the

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002/01-DIDL-NS"
        xmlns:mpeg7=" urn:mpeg:mpeg7:schema:2001"
        xmlns:RDF="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:dc="http://purl.org/dc/elements/1.1/">
 <Item>
  <Descriptor>
   <Statement type="text/xml">
    <mpeg7:Mpeg7>
     <mpeg7:DescriptionUnit
          xsi:type="mpeg7:CreationInformationType">
      <mpeg7:Creation>
         <mpeg7:Title>When the Thistle Blooms</mpeg7:Title>
       <mpeg7:Creator>
         <mpeg7:Role
           href="urn:mpeg:mpeg7:cs:MPEG7RoleCS:PERFORMER"/>
           <mpeg7:PersonGroup>
             <mpeg7:Name>Always Red</mpeg7:Name>
           </mpeg7:PersonGroup>
       </mpeg7:Creator>
       <mpeg7:Creator>
         <mpeg7:Role
           href="urn:mpeg:mpeg7:cs:MPEG7RoleCS:PUBLISHER"/>
           <mpeg7:Organization>
             <mpeg7:Name>PDQ Records</mpeg7:Name>
           </mpeg7:Organization>
       </mpeg7:Creator>
      </mpeg7:Creation>
     </mpeg7:DescriptionUnit>
    </mpeg7:Mpeg7>
   </Statement>
  </Descriptor>
  <Descriptor>
   <Statement type="text/xml">
    <RDF:Description>
      <dc:title>When the Thistle Blooms</dc:title>
      <dc:creator>Always Red</dc:creator>
      <dc:publisher>PDQ Records</dc:publisher>
    </RDF:Description>
   </Statement>
  </Descriptor>
  <Component>
      <Resource ref="rtsp://telemedia1:/v11.aac"/>
  </Component>
 </Item>
 </DIDL>
```

Fig. 1. An example of DI containing MPEG-7 and RDF-based Dublin Core descriptors.

requirements with its interoperability and extensibility to include all of the existing content formats.

A DI is composed of metadata, media resources, and structure as described in section II. The metadata is DID specification based and capable of containing many other forms of standard or non-standard metadata. MPEG-7 descriptions, Rights Expression Language/Rights Data Directory (REL/RDD), IPMP, and ER are good examples of elements that can be contained in a DI to represent media resources in terms of description, right, reporting, and protection.

Not only can users be content producers but they also can be end users in the MPEG-21 framework. In other words, a user is able to easily generate DIs by editing metadata based on the DID schema as shown in Fig. 1 and to efficiently access any DI with any device.

The proposed broadcasting and communication convergent framework utilizes the interoperable content, a DI, as a fundamental content unit of distribution and transaction for UMA service as described in section III.4.

2. Content Adaptation

A. MPEG-2 to MPEG-4 Video Transcoding

Due to the variety of video sources, terminals, and networks, several types of conversions are required for video delivery. In order to deliver digital content to diverse terminals over various networks, it is necessary to adaptively transcode the pre-encoded video into another form to meet specific terminal and network requirements [11].

This subsection considers MPEG-2 to MPEG-4 conversion and three types of adaptations: temporal, spatial, and rate adaptation. First, a temporal adaptation module changes temporal frame rate by sub-sampling decoded frames. Motion vectors are re-calculated to compensate skipped frames. We use the forward dominant vector selection (FDVS) algorithm for motion vector composition [12]. For example, at 30 fps to 15 fps conversion, we skip one frame per every two frames and compose motion vectors using previous and current motion vectors using FDVS.

Next, the spatial adaptation module reduces the spatial dimension of the input frame by sub-sampling a frame along with its horizontal and vertical axis. Motion vectors and coding modes are re-mapped for a new macroblock. If at least one macroblock is encoded as an intra mode, a new macroblock has an intra mode. Otherwise, an inter mode is selected and a motion vector is set by the median of motion vectors.

Rate adaptation is then performed by increasing a quantization parameter value so as to reduce the bitrate of the encoded video. The selection of quantization parameters is controlled by the MPEG-4 rate control algorithm.

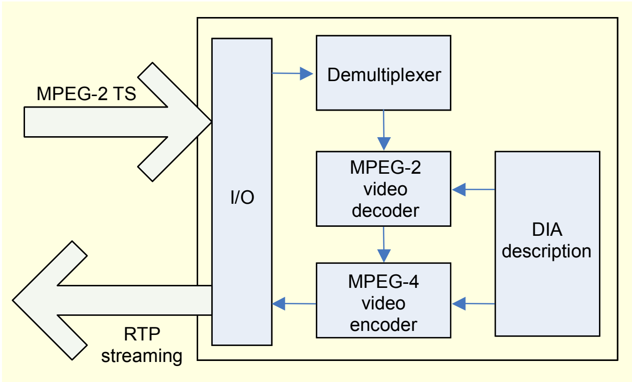


Fig. 2. Functional block diagram of video transcoder.

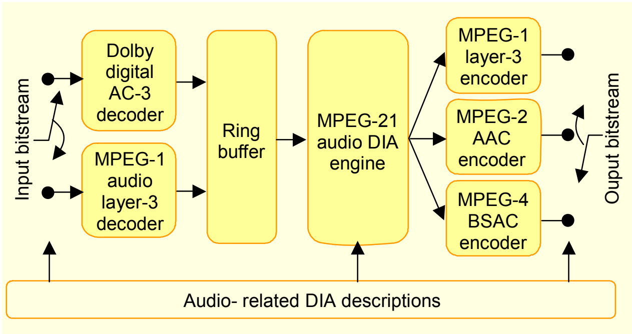


Fig. 3. Functional block diagram of audio transcoder.

B. Audio Transcoding

In the audio transcoder, an AC-3 or MPEG-1 layer III bitstream is transcoded into a form of the MPEG-1 layer III, MPEG-2 AAC, or MPEG-4 bit-sliced arithmetic coding (BSAC) bitstream. Also, the audio transcoder includes a resource adaptation tool called an MPEG-21 audio DIA engine, by which the audio is adapted according to the audio-related DIA descriptions specified in the MPEG-21 DIA.

The audio-related DIA descriptions include the audio presentation preferences, auditory impairment, and audio environment descriptions. Also, several descriptions in the terminal capabilities and network characteristics are associated with audio. In the audio transcoder, a volume controller, a set of lowpass and highpass filters, and a 10-band graphics equalizer are implemented to deal with the volume control, audio frequency range, and frequency equalizer attributes, respectively, in the audio presentation preferences description. Also, 7-band and 10-band equalizers are implemented in order to process the auditory impairment and audio environment descriptions, respectively, since the two descriptions are described by sets of gain values at several frequency locations.

The functional block diagram of the audio transcoder is depicted in Fig. 3. First, the input bitstream is decoded into PCM audio data, where an appropriate decoder is chosen according to the input bitstream type. Note that the frame size of the input audio may be different from that of the output audio. In order to cope with different frame sizes, a ring buffer is included for storing the decoded audio. Subsequently, the decoded PCM audio is adapted in the MPEG-21 audio DIA engine depending on the DIA descriptions. Finally, the adapted PCM audio is encoded by a proper encoder with a given bit rate. Here, three different encoders including the MPEG-1 layer III, MPEG-2 AAC, and MPEG-4 BSAC encoders are implemented. The selection of an encoder and its bit rate is made depending on the terminal capability and network characteristics descriptions.

C. Video Color Adaptation

Color vision deficiency (CVD) is described by deficiency type and deficiency degree in MPEG-21 DIA. The deficiency types are classified as red-deficiency, green-deficiency, and blue-deficiency. The degree of color vision deficiency can be measured in the textual degree type and numerical degree type. In the numerical degree type, between 0.1 and 0.9 represents mild color vision deficiency, and 1.0 is severe color vision deficiency.

The resource adaptation for severe color vision deficiency concentrates on conveying visual information to color deficient people by assisting them in discriminating the confused edge of each object in the color image. Thus the process to detect the necessary adaptation region should precede the adaptation itself. To adapt the resource for color vision deficiency, adaptations are performed to compensate for the lack of one cone for the severe color vision deficiency. Figure 4 shows the procedure of adaptation for severe color vision deficiency.

1) Transformation of the RGB pixel value in the image into LMS[1)] color space, because LMS specification, which decorrelates cone responses, is a very powerful colorimetric tool.
2) Detection of the region from which color deficient people cannot perceive visual information. Region detection is performed by identifying the change of the pixel value when we project the pixel value in the normal LMS color space onto the defective LMS one.
3) Changing either hue, intensity, or saturation pixel value in the detected region. For example, a changed color in the detected region is the same as the one in the background. We can change saturation, intensity, or pixel value in the image.
4) Conversion of the adapted pixel value in an image into an RGB color space for display [13].

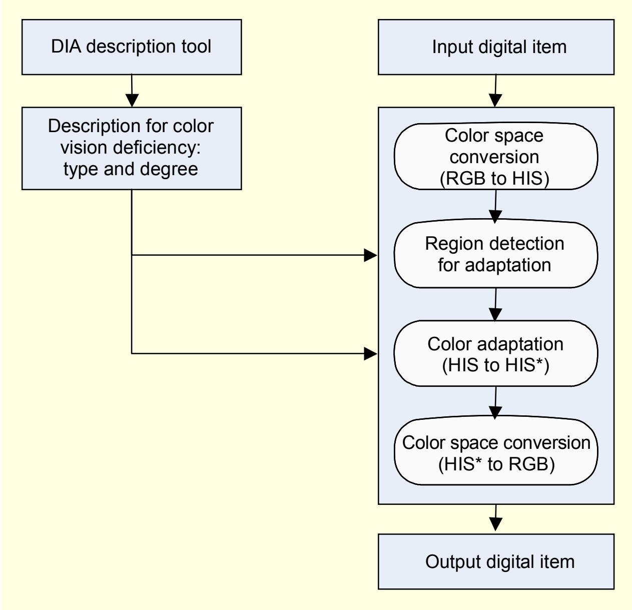


Fig. 4. Functional block diagram of CVD adaptation.

D. 2D to 3D Stereoscopic Adaptation

Stereoscopic Video Adaptation supports video adaptation from 2D video to stereoscopic video. A user can describe her or his own display presentation preferences by specifying the following descriptors: *ParallaxType*, *DepthRange*, and *MaxDelayedFrame* [14].

ParallaxType represents the type of parallax being composed of either positive parallax or negative parallax [15]. This description can be used by the resource adaptation of stereoscopic video in order to deliver the perception of 3D depth. In the negative parallax, 3D depth is perceived between the monitor screen and human eyes. On the contrary, 3D depth is perceived behind the monitor screen in the positive parallax. DepthRange indicates the range of 3D depth perceived by the user and is defined as the distance between the monitor screen and the object in 3D. It applies identically to the positive and negative parallaxes. The amount of DepthRange is varied within [Min., Max.], which can be normalized to [0, 1]. For positive and negative parallaxes, shifting a right image to the right increases the range of depth. On the contrary, shifting it to the left decreases the range of depth.

One of the schemes for stereoscopic video adaptation makes use of a previous (delayed) image. Suppose the image sequence is $(\cdots, I_{K-3}, I_{K-2}, I_{K-1}, I_K, \cdots)$ and I_K is the current frame. One of the previous frames, I_{K-i} $(i \geq 1)$ is chosen. Then, a stereoscopic image consists of I_K and I_{K-i}. If the current and previous images are appropriately presented to both human eyes according to

1) LMS is a color space represented by the response of three types of cones of the human eye, named after their sensitivity at long, medium, and short wavelengths.

camera and object motions, the user then feels the stereoscopic perception. MaxDelayedFrame determines the amount of *i* value. Thus, the larger it is, the more depth the user feels.

There are two ways of selecting an appropriate eye to receive the previous image. The first one is to determine the current image as the left image and the previous image as the right image. The second sets the current image as the right image and the previous image as the left image. The factor that determines a previous frame plays an important role in the stereoscopic perception. MaxDelayedFrame determines one of the previous images according to its value. The greater it is, the better the depth perception. ParallaxType is controlled by switching the left and right images. For varying values of DepthRange, a left image is horizontally shifted to the left.

The resource adaptation concentrates on generating stereoscopic video from 2D video. Among many kinds of image motion, a horizontal-motion image sequence can be used to verify the resource adaptation process. Figure 5 is the procedure of adaptation for stereoscopic adaptation.

1) From a given image sequence, a motion vector for each image block is estimated. The motion vector may be obtained from any motion estimation method or directly from a compressed bitstream of MPEG-encoded data.
2) The image motion type is analyzed so that horizontal motion, non-horizontal motion, and static image are determined for each image.
3) Based upon the image motion determined, a stereoscopic image composed of left and right images is generated.
4) According to descriptions such as ParallaxType, DepthRange, and MaxDelayedFrame, stereoscopic images are adapted.

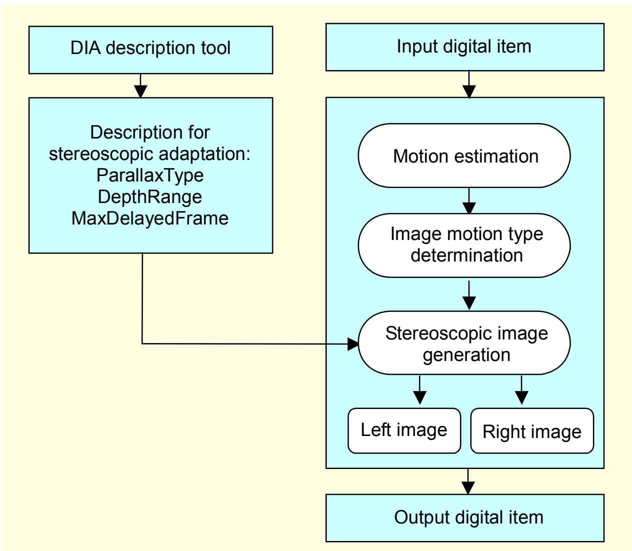


Fig. 5. Functional block diagram of stereoscopic adaptation.

3. Delivery

A. Digital TV Broadcasting

Digital television (DTV) is a technology for transmitting and receiving digital broadcast television signals via the broadcast spectrum. A digital standard is superior to analog because of its greater accuracy, versatility, efficiency, and interoperability with other electronic media. Digital signals also have the advantage of generating less noise and being more resistant to signal interference. Within the range of the signal, this results in a perfect signal.

One of the primary rationales for DTV broadcasting is high-definition television, or HDTV. This higher resolution and different aspect ratio makes HDTV images substantially more vivid and engaging than the images produced by the existing television format, and that effect is enhanced by five discrete channels of CD-quality audio. But DTV is not just about HDTV. As a digital (and not analog) signal, DTV allows broadcasters to offer a variety of innovations. Instead of sending an HDTV signal of 19.3 Mbps in an Advanced Television Systems Committee (ATSC) channel, for example, a broadcast station can send as many as five digital standard definition television (SDTV) signals, each of which might consist of 4 to 5 Mbps. This new capacity, known as multicasting or multiplexing is expected to allow broadcasters to compete with other multi-channel media such as cable and direct broadcast satellite systems.

Another DTV capability is the ability to provide new kinds of video and data services such as interactive services with data transmissions via a broadcasting channel. In the proposed framework, SDTV content with DI is multiplexed and transmitted at 19.3 Mbps via a terrestrial broadcasting channel.

B. Internet Protocol Streaming

Streaming technology allows digital media, whether audio or video, to be transferred in compressed formats across a network and played in real time, rather than after an entire file has downloaded. Along with the standardization of media formats and advancements in network infrastructures, streaming has fostered the emergence of numerous media-rich applications, which serve a wide range of purposes. For example, ESPN (a sport Web site) might provide instant replays of a live football game, including replays not broadcast on TV. Another site might use webcasts to allow viewers around the world to watch Broadway shows live.

Media streaming has several advantages over terrestrial broadcasts, including user interaction, fewer geographical constraints, and customized content delivery. As the availability of high-speed Internet access grows, the market presence of media streaming is expanding. Furthermore, in the foreseeable future, third-generation (3G) wireless networks capable of handling speeds between 384 kbps and 2 Mbps will provide handheld devices with the bandwidth needed for truly dynamic services and applications such as videoconferencing and multiplayer gaming.

Implementation of streaming in the proposed test bed is based on the specifications of the Internet Streaming Media Alliance (ISMA) derived from the international standard, MPEG-4 on IP. For streaming, the session should be initialized through a network connection using the Real-Time Streaming Protocol (RTSP) designed to work with a time-based media resource or any other application where application-controlled, time-based delivery is essential, with Session Description Protocol (SDP) containing the detailed information of server, Real-time Transport Protocol (RTP), and content information; that is, Initial Object Descriptor (IOD). A streaming server, a DIA server in the proposed test bed, starts streaming real-time audiovisual data using RTP while each encoded audiovisual frame is packetized in RTP format on the fly.

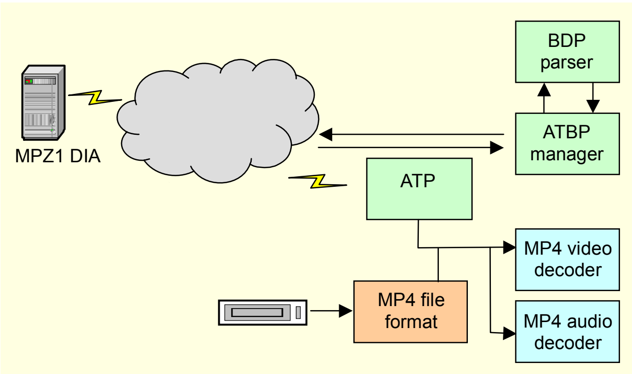


Fig. 6. Network interface modules for IP-based streaming.

4. UMA Service

Universal multimedia access (UMA) refers to the ability to access by any user the desired multimedia content over any type of network with any device from anywhere and anytime. UMA is a key framework for multimedia content delivery service using metadata. The key issues in UMA include quality of service (QoS) in multimedia delivery and service based on adaptations of multimedia content to bridge the gap between content creation and consuming; standardized metadata descriptions that facilitate the adaptation such as MPEG-7, MPEG-21 DIA, and UMA system design considering its target application [16]. In other words, UMA is about how users can access the same media resources with different terminal equipment and preferences over heterogeneous networks [17]. Therefore, the media resources have to be adaptable and flexible according to the users' needs and capabilities. In order to realize this, a multimedia framework has been developed and proposed in this article which takes the concepts of MPEG-21 and has very similar goals in terms of access and distribution of media using heterogeneous terminals and networks [10].

IV. Test Bed Implementation

This section gives a detailed explanation of a test bed that exploits advanced technologies for the proposed convergent multimedia framework based on MPEG-21. This section describes a use case scenario, systems in the framework, and a walkthrough in terms of technologies as well as UMA service.

1. Use Case Scenario

A service scenario for the proposed multimedia framework is described in detail in Fig. 1 representing the UMA services required in heterogeneous networks [18].

The Education Broadcasting System (EBS) produces an educational program, "Top of the English," for high school students. The main program is broadcast along with its digital item (a part of Electronic Program Guide, EPG - a set of digital items), and additional media resources such as "Textbook" (doc), "Daily English" (VRML, graphics), "Listening Practice" (Audio only, AAC), and "Speaking Practice" (Video only, MPEG-2) are also provided via a return channel at the user's request.

Kim notices the education program from EPG when he turns a PC on. He chooses to watch the program with a textbook after downloading it from an EBS server. Then, he also downloads other additional data in order to practice listening and speaking with them. When he watches an animation video for speaking practice, he notifies a DIA server that he is color-blind so that he can get color-modified content. He also downloads an AAC player from a tool server to listen to a listening practice audio exercise, since it is not available in his PC.

Lee also watches the program with his mobile terminal (PDA phone) on his way home on the bus along with the downloaded textbook. Since his mobile terminal cannot play broadcast contents, adapted media resource in decodable format streams to his terminal in real time from a DIA server.

Fig. 7. Use case scenario for the proposed multimedia framework.

2. Test Bed Configuration

The proposed test bed consists of a broadcasting server, a DIA server, a broadcasting terminal, a communication terminal, a tool server, and a license server as shown in Fig. 8. It provides UMA service in broadcasting and communication convergence network by satisfying users' demands. For instance, accessing

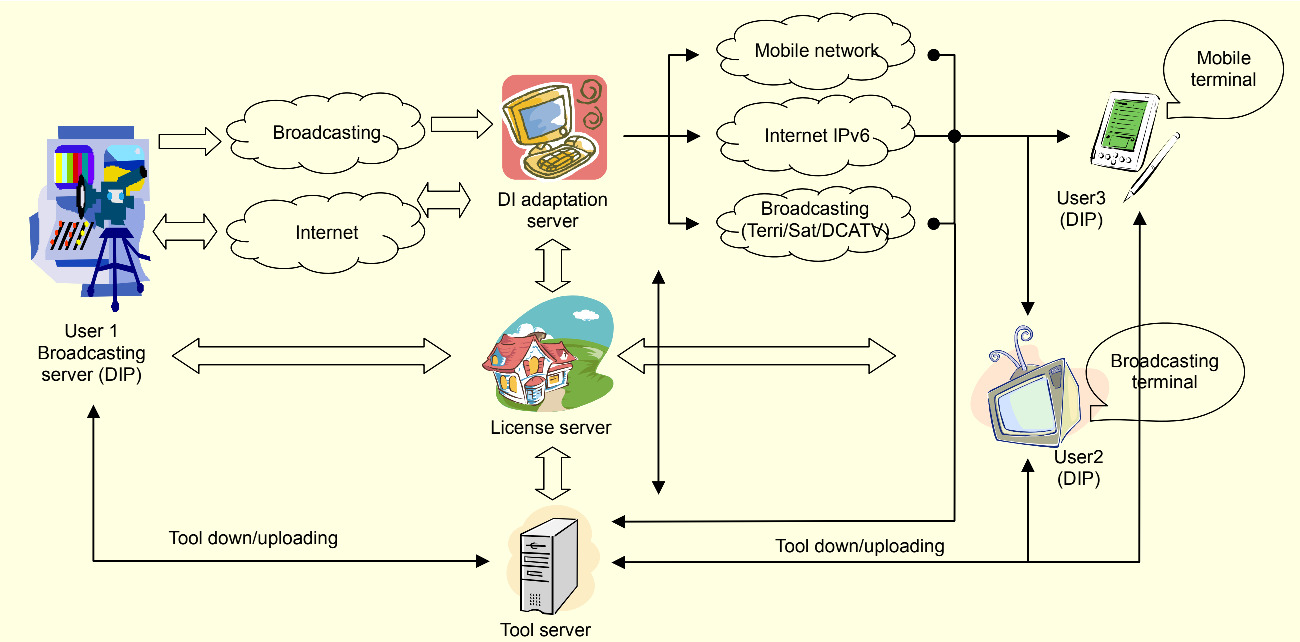


Fig. 8. UMA service in broadcasting & communication convergent environment based on MPEG-21.

broadcasting content in a mobile terminal connected in a telecommunication network or enabling the disable on CVD to watch any content with their own distinguishable color by adaptation of content would be good examples of UMA service. Also, the most important aspect, interoperability of content among various devices is guaranteed when the multimedia framework is based on MPEG-21.

A. Broadcasting Server

In the proposed framework, a broadcasting server acts as content or a service provider along with a DIA server. It generates interoperable content, a DI, and transmits it via a terrestrial broadcasting channel. The DI as broadcast content is then adapted so that a communication terminal can access it in accordance with the DIA server in the framework. This achieves one of the main objectives of the proposed framework; that is, one-source multi-use.

A broadcasting server comprises modules of a DI authoring engine (Fig. 10), a Remultiplexer, an MPEG-21 file generator, and an I/O. A DI authoring engine generates DIs that contain DID specification based metadata and media resources located either internally or externally, such as TV programs with ancillary data, as for example, a textbook and speaking practice (mpg file) as indicated in the use case scenario. An MPEG-21 file generator encodes XML-based DIs to a binary format and storage devices. DIs are then multiplexed with broadcasting content (MPEG-2 TS) by a Remultiplexer and broadcast, whereas ancillary data requested by users is streamed via the Internet.

As a content provider, a broadcasting server generates DIs,

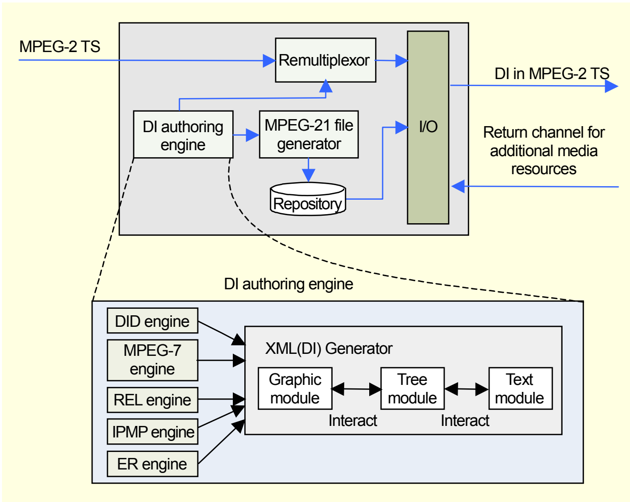
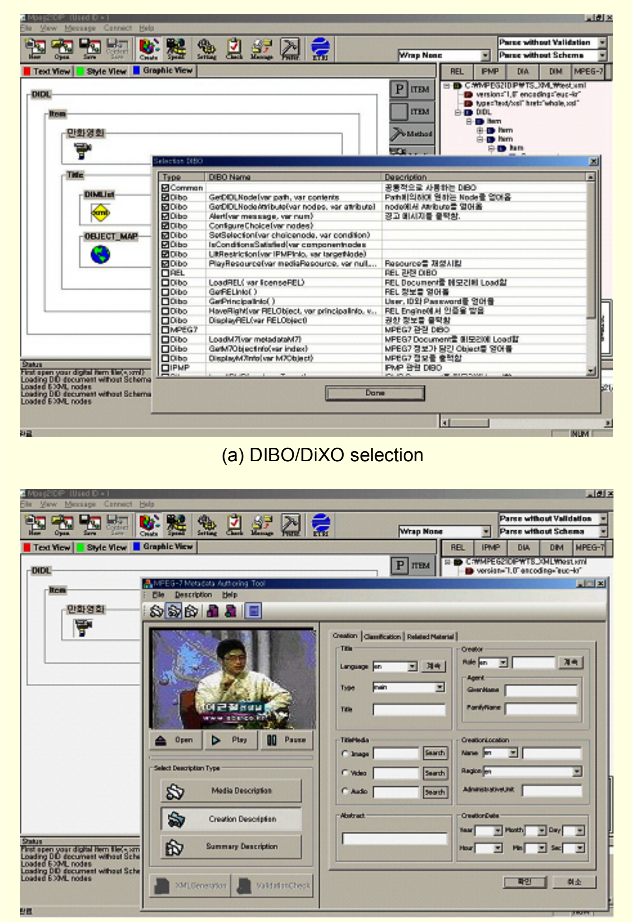


Fig. 9. Functional block diagram for a broadcasting server.

basic content units of distribution and transaction in the multimedia framework as described in section III.1. The DI is then multiplexed with an incoming MPEG-2 TS so as to be streamed to a broadcasting channel or possibly to be encoded into a storage format, which is MPEG-21. A broadcasting server also supports data broadcasting service[2] by handling users' requests from the return channel.

A broadcasting server plays an important role in providing UMA service by generating and delivering interoperable

2) It is also known as 'Interactive TV', 'two-way transmission TV', or value-added information service.



(a) DIBO/DiXO selection

(b) MPEG-7 description generation for a video content

Fig. 10. Procedure of a digital item authoring: (a) DIBO selection for a media resource within a digital item, (b) MPEG-7 authoring engine to generate MPEG-7 metadata such as summary description for a media resource.

multimedia content, or DIs, which can be manipulated and adapted by a DIA server for one-source multi-use in the proposed convergent networks.

B. Digital Item Adaptation Server

As a core component of the implemented broadcasting and communication convergent framework, the DIA server plays the important roles of connecting broadcasting, communication, and the Internet, and of adapting content depending on users' requests, terminal capability, or network characteristics. The DIA server follows the MPEG-21 DIA specification which was defined to adapt DIs to users' consumption environments.

The DIA server is composed of following modules – an I/O to receive and demultiplex broadcasting content (MPEG-2 TS) and to stream the adapted broadcasting content in the format of RTP, a metadata engine to adapt metadata, a media resource adaptor engine to process the adaptation of media resources, and a streamer to packetize the adapted media resource and to

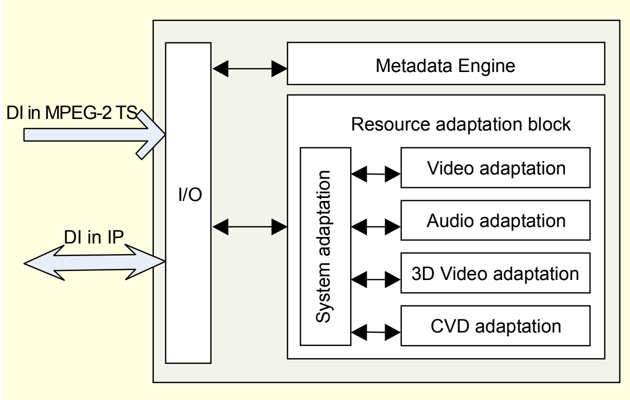


Fig. 11. Functional block diagram for DIA server.

stream it to the communication network.

The DIA server receives an MPEG-2 TS containing a DI via a terrestrial broadcasting channel. Then, it adapts metadata (a DID instance) and media resource comprising a DI after the demultiplexing of the MPEG-2 TS. As described in section 3, broadcasting content is adapted in real-time to a format that a mobile terminal is able to process; that is, MPEG-4 video and mp3 audio with lower bitrates. In the case of a media resource adaptation request in the form of a DI by a user, the DIA server also performs the adaptation process depending upon his/her preference such as CVD or 3D adaptation described in detail in Section 3. The adapted DI is then packetized into RTP and streamed at a variable bitrate for the network availability to a mobile terminal after a session is made. A mobile terminal can access the DI via Internet, WLAN, or CDMA.

Therefore, the DIA server enables a mobile terminal to access broadcasting content and provides a UMA environment to the user. Also, it leads us to achieve the one-source multi-use of content comparing to the present multimedia framework

Fig. 12. An example of adaptation processes in DIA server.

service where various types, sizes, and formats for the same content are generated and delivered for different networks or terminals. Figure 12 shows the list of adaptation processes in a DIA server.

C. End User Terminal

An end user terminal is the device by which a user can access UMA service within the multimedia framework. A broadcasting terminal and a mobile device enable a user to access various kinds of media resources with valuable information within a DI. Moreover, it provides personalized content services by requesting to and interfacing with the DIA server that adapts content format as the user prefers.

A terminal consists of the following modules: an I/O to receive and demultiplex MPEG-2 TS or streaming RTP packets (only in a mobile device); a metadata engine to parse, process, and generate DIs; and a media handler to access digital resources (Fig. 13).

A broadcasting terminal receives MPEG-2 TS multiplexing with a DI via a terrestrial broadcasting channel and then parses the DI after the demultiplexing of the MPEG-2 TS.

The digital item method (DIM)[3] list displays as shown in Fig. 13 after a DI is parsed and the user selects one that represents a particular media resource. The metadata engine checks the rights of the user on the chosen media resource, and displays

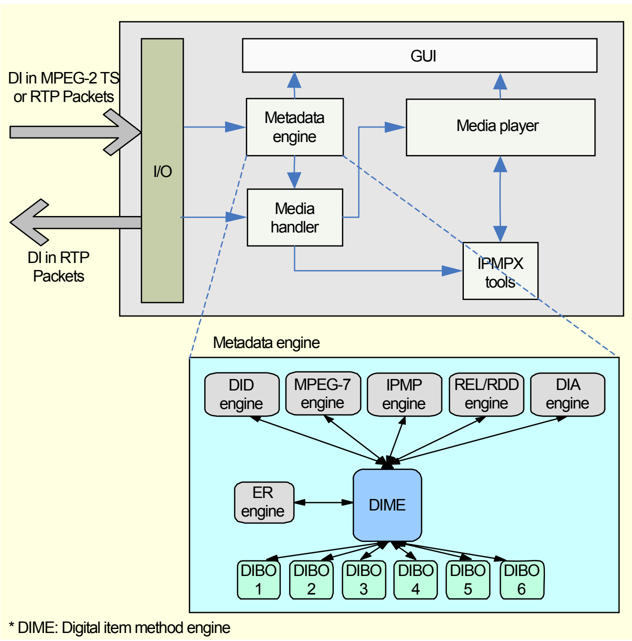


Fig. 13. The functional block diagram of an end user terminal, a broadcasting terminal, or mobile terminal with limited functionality.

3) DIM represents a middleware API to access a media resource defined within a DI. In the test DI, there exist as many DIMs as there are media resources.

it with its REL, IPMP, and MPEG-7 information as shown in Fig. 14 if provided in the DI. It repeats the same procedure when the user selects another media resource, a DIM in the proposed terminal.

It often happens that an end user wants to have better quality, or a terminal cannot afford to process a certain type of media resource due to its processing capability or the lack of tools such as the necessary media codec or player. In this case, the end user terminal then resolves the limitation by coupling with a DIA server that manipulates the requested media resource to overcome the terminal capability or by the user's request, uses the tool server to search for and download the missing tool to the terminal.

A mobile terminal as an end user terminal has many functionalities similar to a broadcasting terminal in terms of the handling of a DI and providing UMA service in the multimedia framework; however, it differs in that it receives and transmits DIs via a communication network, such as a CDMA or a WLAN. Moreover, compared to a broadcasting terminal such as a set-top, it has more constraints, such as terminal capability

Fig. 14. A broadcasting terminal.

Fig. 15. A mobile terminal (PDA phone).

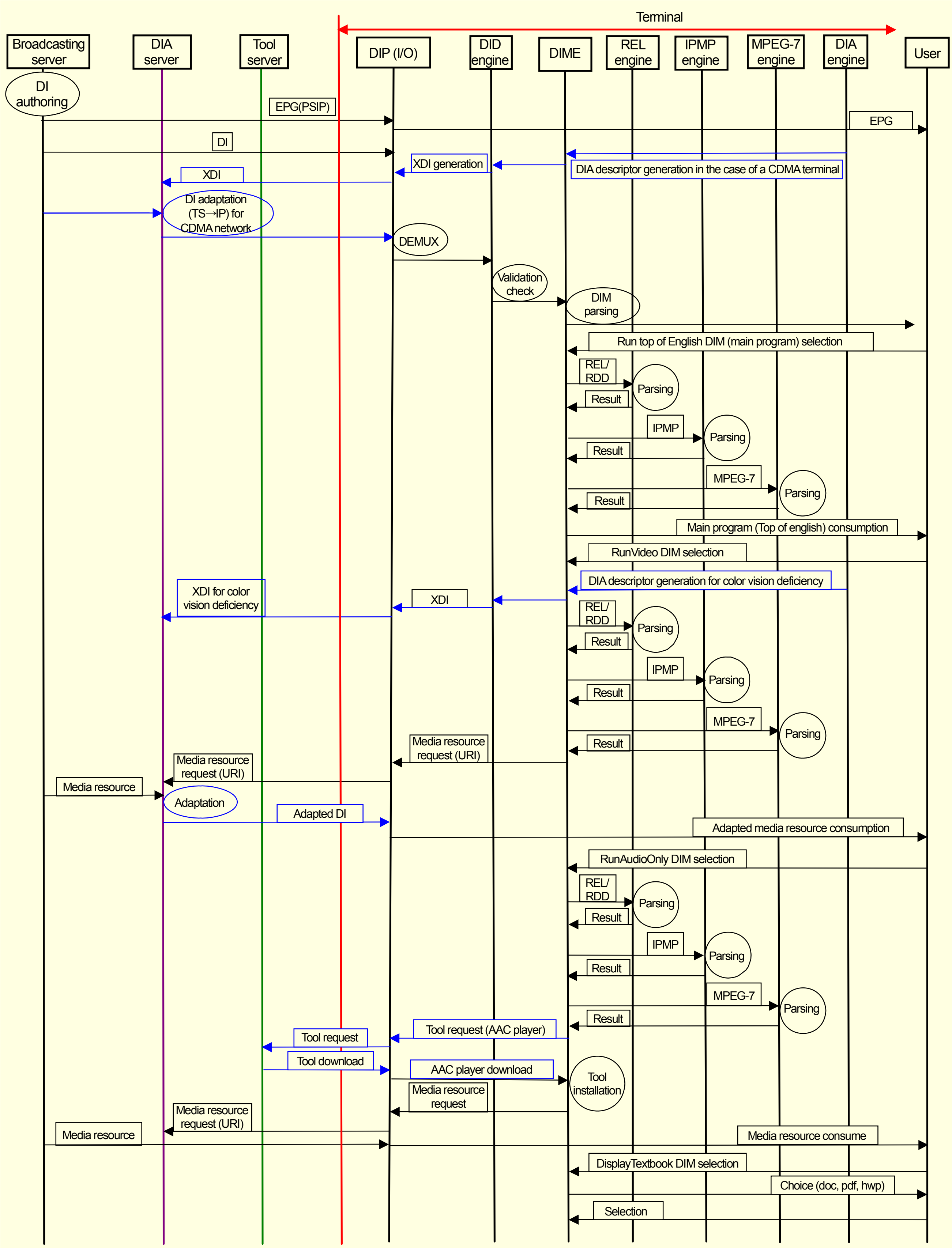


Fig. 16. Functional walkthrough of the proposed test bed.

and network bandwidth. This results in the downsizing of metadata engines and the related media resource handler for IPMP to reduce its processing power. We, however, expect that it will be simply resolved as the capability of mobile terminals increases.

In short, an end user terminal also plays an important role in the proposed multimedia framework. Moreover, it has unique features; that is, it has very simple terminal structure compared to any commercial devices and extensibility to utilize existing players since it mainly handles metadata and connects a specified media player so that a media resource is processed by the player.

D. Tool Server

All devices within the multimedia framework are capable of handling various types of content, although each cannot function perfectly to handle all different formats of media resources. Therefore, a device such as a tool server to help search for and download tools is definitely required in the multimedia framework, especially in the event that any device lacks tools such as a media codec or decryption modules for content adaptation or consumption.

E. License Server

A license server verifies the validation of content ID (DII) and copyright information (REL) when each device generates, adapts, and consumes tools related to digital contents, that is, DIs.

Throughout this practical implementation of the MPEG-21 multimedia framework for broadcasting and communication convergent network, the goal and value of MPEG-21 is exploited, that is, it defines the technology needed to support users in exchanging, accessing, consuming, trading, and otherwise manipulating DIs in an efficient, transparent, and interoperable way [1] for which MPEG-21 standard is needed.

Due to various technologies combined in this experiment, we limited our focus to providing users with UMA service in a convergent network with ease. Improving, for instance, higher throughput in a delivery chain or reducing complexity in CVD adaptation is not of interest in this experiment, although simple A/V rate adaptation has been introduced for the convergent network purpose. However, that will be intensively researched later along with scalable A/V to enhance the MPEG-21 multimedia framework even more.

3. Walkthrough

Figure 16 shows the procedure of information exchanges among devices in the proposed test bed. The procedure represents UMA service from the technical point of view as our service scenario depicts.

V. Conclusion

The fast development of digital communication has brought a great deal of convenience to users and has recently resulted in the growth of digital content related industries. This growth has enabled digital content to be easily produced, copied, and distributed anywhere in a short period of time. However, it has also highlighted the need to resolve how to converge various heterogeneous networks, and how to facilitate users to access multimedia contents by supporting their diverse devices and preferences. The development of a convergence framework to effectively resolve these problems is of vital importance.

As a result, many standardization groups such as MPEG and industries are currently putting their effort into the research and development of convergence framework to provide UMA service. We have proposed a test bed for a broadcasting and communication convergent service framework based on MPEG-21 as a possible solution for these important issues.

The proposed system provides UMA service in the convergence network, systematically and efficiently building on the MPEG-21 standard which leads to interoperable solutions in the deployment of advanced multimedia packaging and distribution applications [1]. The concepts of DI and DIA play important roles in UMA service for the proposed convergent framework. As core technologies, they provide a user with personalized service and result in one-source multi-use in the proposed broadcasting and telecommunication convergent network by allowing transparent and interoperable way of adapting multimedia contents. Consequently, the proposed implementation enables a user to experience the concept of UMA based on MPEG-21, that is, the ability for any type of terminals to access and consume a rich set of multimedia content by enabling various types of QoS management [19].

In the future, we will conduct further research on improving the performance of content adaptation and higher throughput for more advanced multimedia frameworks.

References

[1] ISO/IEC 21000-1, *Information Technology-Multimedia Framework (MPEG-21) – Part 1: Vision, Technologies and Strategy*, Nov. 2004.

[2] W. Kim and J. Kim, “MPEG-21 Multimedia Framework Summary and Its Trend,” *Broadcasting and Technology*, vol. 81, 2001 (in Korean).

[3] J. Bormans and K. Hill, *MPEG-21 Overview v.5,* ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, Oct. 2002.

[4] I. Burnett, R.V. de Walle, K. Hill, J. Bormans, and F. Pereira, "MPEG-21: Goals and Achievements," *IEEE Multimedia*, vol. 10, Oct. 2003, pp. 60-70.

[5] I. Burnett, S.J. Davis, and G.M. Drury, "MPEG-21 Digital Item Declaration and Identification–Principles and Compression," *IEEE Trans. Multimedia*, vol. 7, no. 3, June 2005, pp. 400-407.

[6] ISO/IEC 210001-15, *Information Technology-Multimedia Framework (MPEG-21)–Part1 5: Event Reporting*, Feb. 2006.

[7] P. Ruskin, "MPEG-21-Frequently Asked Questions ver 1.0," July 2001.

[8] H. Kim, "MPEG-21 Multimedia Framework," *TTA Journal*, no. 82, Aug. 2002, pp. 67-71.

[9] Y. Cho, J. Seok, J. Hong, and C. Ahn, "Broadcasting System Compliant with MPEG-2/4 IPMPX," *ETRI Journal*, vol. 26, no. 2, Apr. 2004, pp. 83-91.

[10] P. van Beek, J.R. Smith, T. Ebrahimi, T. Suzuki, and J. Askelof, "Metadata-Driven Multimedia Access," *IEEE Signal Proc. Mag.*, vol. 20, no. 2, Mar. 2003, pp. 40-52.

[11] J.-G. Kim, Y. Wang, S.-F. Chang, and H.-M. Kim, "An Optimal Framework of Video Adaptation and its Application to Rate Adaptation Transcoding," *ETRI Journal*, vol. 27, no. 4, Aug. 2005, pp. 341-354.

[12] J. Youn, M.T. Sun, and C.W. Lin, "Motion Vector Refinement for High Performance Transcoding," *IEEE Trans. Multimedia*, vol. 1, Mar. 1999, pp. 30-40.

[13] S.J. Yang, Y.M. Ro, J.H. Nam, J.W. Hong, S.Y. Choi, and J.H. Lee, "Improving Visual Accessibility for Color Vision Deficiency Based on MPEG-21," *ETRI Journal*, vol.25, no.3, June 2004, pp.195-202.

[14] M.B. Kim, J. Nam, W. Baek, J. Son, and J. Hong, "The Adaptation of 3D Stereoscopic Video in MPEG-21 DIA," *Signal Processing: Image Commun.*, B.V. Elsevier ed., vol. 18, no. 8, Sep. 2003.

[15] "Development of Service Framework Technology Based on MPEG-21 for Broadcasting and Communication Network Environment," Final project report, Jan. 2005.

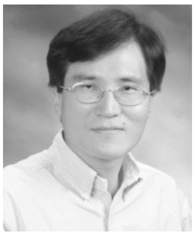
[16] E. Kasutani, "New Frontiers in Universal Multimedia Access," *ITS-22-04*, Sep. 2004.

[17] J. Bormans, J. Gelissen, and A. Perkis, "MPEG-21: The 21st Century Multimedia Framework," *IEEE Signal Proc. Mag.*, vol. 20, no. 2, Mar. 2003, pp.53-62.

[18] Y. Cho, D. Suh, J. Hong, and J. Kim, "MPEG-21 Based Broadcasting Communication Converged Framework over the Next Generation Internet," *Telecommunication Review*, vol. 13, no. 4, Aug. 2003, pp. 527-539 (in Korean).

[19] A. Vetro and C. Timmerer, "Digital Item Adaptation: Overview of Standardization and Research Activities," *IEEE Trans. Multimedia*, vol. 7, no. 3, June 2005, pp. 418-426.

Yongju Cho received the BS and MS degrees in electrical and computer engineering from Iowa State University in the USA in 1997 and 1999. He joined the Electronics and Telecommunications Research Institute (ETRI) in Korea as a Member of Engineering Staff in 2001 while going on for a doctorate. Since then, he has been engaged in the development of a data broadcasting system based upon MPEG-4 systems, IPMP, and in the MPEG-21 area. His research interests include digital signal processing in the field of data broadcasting and multimedia systems.



Jae-Gon Kim received the BS degree in electronics engineering from Kyungpook National University, Korea, in 1990, the MS and PhD degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Korea, in 1992 and 2005, respectively. Since 1992, he has been a Senior Member of Research Staff, in the Broadcasting Media Research Group of Electronics and Telecommunications Research Institute (ETRI), Korea. He is currently the Team Leader of the Convergence Media Research Team. From 2001 to 2002, he was a Staff Associate at the Department of Electrical Engineering, Columbia University, New York. His research interests include scalable video coding, video adaptation, networked video, multimedia applications, and MPEG-7/21.

Jin Woo Hong is a Director and Principal Member of the Technical Staff in the Digital Broadcasting Media Research Division of ETRI, Korea. He received his BS and MS from Kwangwoon University, Korea, in Feb. 1982 and 1984, and the PhD from the same university in Aug. 1993. Since he joined ETRI in 1984 he has been involved in developing digital audio systems, IPMP, and broadcasting technology. His recent work has been focused on developing a multimedia framework for convergence of broadcasting and telecommunication based on MPEG-21 technology. His main interests are in the areas of multimedia signal processing, IPMP, and digital audio broadcasting.