Interworking Between SIP and MPEG-4 DMIF For Heterogeneous IP Video Conferencing

Toufik Ahmed[1], Ahmed Mehaoua[1] and Raouf Boutaba[2]

[1] University of Versailles, CNRS-PR*i*SM Lab.
45 av. des Etats-Unis, 78000, Versailles, France
[2] University of Waterloo, Dept. of Computer Science
200 University Avenue West, Waterloo,
Ont. N2L 3G1, Canada

Abstract— **This article discusses technical issues related to delivery and control of IP multimedia services, such as video-conferencing, involving heterogeneous end terminals. In particular, it describes the design and implementation of an experimental system for interworking between IETF SIP (Session Initiation Protocol) and ISO MPEG-4 DMIF (Delivery Multimedia Integration Framework) session and call control signaling protocols. This IP videoconferencing interworking system is composed of two core units for supporting delivery of audio-video streams from a DMIF domain to a SIP domain (i.e. DMIF2SIP unit) and from a SIP domain to a DMIF domain (i.e. SIP2DMIF unit). These units perform various translation functions for transparent establishment and control of multimedia sessions across IP networking environment, including, session protocol conversion, service gateway conversion and address translation.**

Keywords: IP, SIP, MPEG-4, Session and Call control.

I. Introduction

Internet video conferencing and IP telephony has grown rapidly in the last few years. This rapid expansion and potential underlies the importance of an enabling and unifying standard. Actually, it appears likely that both IETF SIP (Session Initiation Protocol) [1] with SDP (Session Description Protocol) [2] and the ITU-T recommendation H.323 [3][4] will be used for setting up Internet multimedia conferences and telephone calls. While the two protocols are comparable in their features, SIP provides a higher flexibility to add new features; a relatively easier implementation; and a better integration with other IP related protocols. Recently, The 3rd Generation Partnership Project (3GPP) has selected SIP as the session establishment protocol for the 3GPP IP Multimedia Core Network Subsystem. 3GPP has defined signalling flows for the IP multimedia call control based on SIP and SDP [5].

In the other hand, the recent ISO MPEG-4 standards [6][7][8] target a broad range of low-bit rates multimedia applications: from classical streaming video and TV broadcasting to highly interactive applications with dynamic audio-visual scene customisation (e.g. e-learning, videoconferencing). In order to reach this objective, advanced coding and formatting tools have been specified in the different parts of the standard ISO 14496 (i.e. Audio, Visual, and Systems), which can be configured according to profiles and levels to meet various application requirements. A core component of the MPEG-4 multimedia framework is the "Delivery Multimedia Integration Framework"[9]. DMIF offers content location independent procedures for establishing and controlling MPEG-4 audiovisual sessions and access individual media channels over RTP/UDP/IP.

Therefore, in order to achieve universal IP connectivity and seamless IP multimedia services, standardized interworking procedures between MPEG-4 and SIP terminals is required [10].

The remainder of the article is as follows. Section 2 analyses IP videoconferencing heterogeneity issues. Section 3 introduces the targeted multimedia networking and service environment. In particular, we focus on scenario that involve IP MPEG-4 and SIP terminals. Section 4 is devoted to the design and implementation of the proposed multimedia signaling interworking system. Both SIP2DMIF and DMIF2SIP session signaling translators are described as well as call sequence mapping and address translation. Finally, we conclude in section 5.

II. IP Videoconferencing Heterogeneity

Several points of heterogeneity should be addressed for permitting IP multimedia Interworking Service:

(1) Video and audio formats: digital audio formats will be characterized by many factors such as sampling rates or compression algorithms. Video formats may differ in spatial and temporal resolutions, color depth or compression schemes. This format adaptation issue is addressed by multimedia gateways.

(2) Synchronizing and system encoding: each elementary (video, audio, data) stream should be merged to form a single

0-7803-7400-2/02/$17.00 © 2002 IEEE

bit-stream for transmission, and they should be synchronized. In Internet, the Real-time Transport Protocol [11] provides temporal and encapsulation functions.

(3) Application control: users should be able to control the received stream to simulate interactive VCR-like function (e.g. play, pause, fast rewinding, etc.). IETF Real-Time Streaming Protocol (RTSP)[12], DAVIC-S3 [13] and ISO MPEG DSM-CC [14] are examples of signalling protocols developed for that purpose and require interoperability.

(4) Connection control/QoS control: in next generation Internet, QoS could be negotiated by way of different signalling (e.g. RSVP, MPLS LDP, etc.), while some domains will not provide any QoS guarantee and still perform in best effort. Thus, there should be a function that translates QoS requests from a domain to another.

(5) Call/session control: different IP network domains may adopt different method for reserving resources and maintaining session information. There should be a way of managing two independent sessions to form a composite multimedia session (e.g. a SIP compliant phone call and an MPEG-4 DMIF compliant video call). This can be performed by many ways like SLA (Service Level Agreement) negotiation between two domains using COPS Protocol for instance [15].

This paper addresses the later point of service heterogeneity (i.e. call and session control). It describes the design and implementation of an experimental system for IP videoconferencing interworking between ISO MPEG-4 DMIF and IETF SIP signaling protocols. This interworking system is composed of two core units for supporting delivery of audio-video streams from a DMIF domain to a SIP domain (i.e. DMIF2SIP unit), and from a SIP domain to a DMIF domain (i.e. SIP2DMIF unit). These two components perform various translation functions for transparent establishment and control of multimedia conferencing sessions across IP networking environment. Including, session protocol conversion, service gateway conversion, and address translation.

III. TARGET IP VIDEOCONFERENCING SERVICE

We consider an IP videoconferencing service involving numerous heterogeneous terminals like illustrated in Fig. 1. IP video-conferencing can serve as a typical multimedia service for testing SIP, H.323 and MPEG-4 DMIF call control signalling interworking.

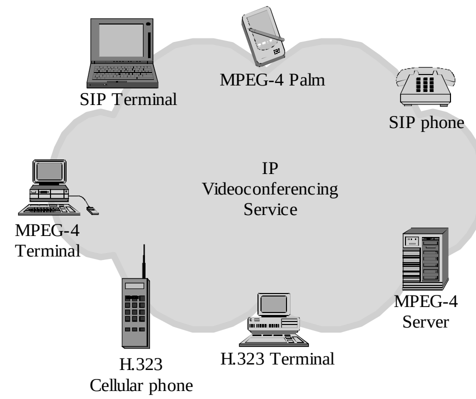

Fig. 1. Heterogeneous IP Videoconferencing

The specific motivation underlying this multimedia service is to help define a framework to enable interoperability between heterogeneous terminals which use different standard and to allow each technology to talk transparently to others without knowing in advance what type of session control protocol is used by peer terminals.

A videoconferencing session may involve multiple video or multiple audio streams, addressing multiple codecs with multi-rate requirements.

To permit all connected user to joint the conference, one solution consists to use a common protocol for session establishment, which is not obvious. Another solution consists to use new components known as Media Gateways that essentially perform feature translation between heterogeneous signaling protocols such as SIP, H.323 and DMIF.

The term media gateway is not new. It was proposed first for interconnecting telephone circuits and data packets network carried over the Internet or over other IP networks.

IV. PROPOSAL OF A DMIF-SIP INTERWORKING MODEL

We propose a functional architecture of logical entity that performs interworking between MPEG-4 DMIF and SIP. This entity is called **DMIF-SIP IWF** (DMIF-SIP InterWorking Function). Fig. 2 illustrates our purpose. The DMIF-SIP IWF is a server composed of two sides: SIP side and DMIF side performing two-ways signaling translation between SIP and DMIF domains.

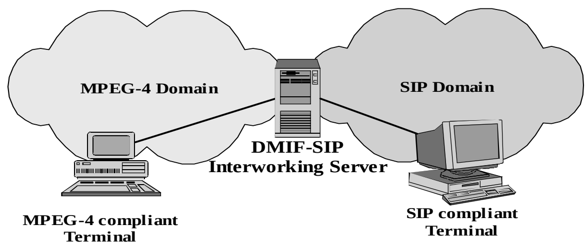

Fig. 2. Interworking between SIP and DMIF

A. *SIP side*

The SIP side of DMIF-SIP IWF is part of the IWF that terminates and originates SIP signaling from and to SIP network respectively. We call this function **SIP2DMIF** (SIP to DMIF) signaling. The SIP2DMIF signaling allows a SIP UAC to call MPEG-4 DMIF Terminal. SIP UAC talks to DMIF-SIP IWF with SIP specification. When SIP2DMIF IWF receives an "INVITE" message from SIP UAC, it sends DMIF Signaling Message (DS_Message) to DNI of MPEG-4 Server. When the session and the service are done, the SIP2DMIF IWF sends "200 OK" message back to SIP UAC. An Acknowledgment Message sends by SIP UAC confirms the connection of SIP UAC to the MPEG-4 Server. Fig. 3 illustrates the messages exchanged between SIP UAC, DMIF-SIP IWF and DNI of MPEG-4 Terminal.

The Steps of the call between User A (SIP Terminal) and User B (MPEG-4 DMIF Terminal) is described in what bellow:

Step 1: SIP User A sends an "INVITE" request to User B. this later is an invitation to User B to participate to videoconferencing. The "INVITE" request contains:

a). The identifier (mail address, phone number) of User B is inserted in the "Request-URI" field in the form of SIP URL

b). SIP User A identifier is recognized as the call session initiator in the "From" field.

c). A unique numeric identifier is assigned to the call and is inserted in the "Call-ID" field.

d). The transaction number within a single call leg is identified in the "Cseq" filed.

e). The Media capability User A is ready to receive is specified via SDP.

Step 2: SIP2DMIF IWF receives "INVITE" request from User A. it translate the identifier of User B in form of DMIF URL which was obtained when the MPEG-4 DMIF Terminal was turned-on. We suppose that the User B addressee match the DMIF URL. The mapping between SIP addresses and DMIF URL is described later. SIP2DMIF IWF passes "DS_SessionSetupRequest" message to the remote DMIF peer in order to activate a network session. This message contains SIP User A capability of handling media. The MPEG-4 Capability Descriptor is defined in [14].

Step 3: DMIF2SIP IWF sends a "100 TRYING" message to SIP User A.

Step 4: DMIF-SIP IWF receives a "DS_SessionSetupConfirm" message. Both peers (SIP UAC and DMIF Terminal) have knowledge of each other. The received message contains also a common set of User B capability descriptor in preferred order of choice.

Step 5: SIP2DMIF IWF sends "DS_ServiceAttachRequest" which contains DMIF URL

Step 6: DNI Layer Informs the MPEG-4 Terminal that User A would to establish a video conferencing with User B.

Step 7: DMIF-SIP IWF receives a "DS_ServiceAttachConfirm" message indicating that User B is capable of handling videoconferencing.

Step 8: SIP2DMIF IWF sends a "200 OK" message back to SIP User A. The response message notifies SIP User A that the connection has been established. This message contains also the intersection of the two terminals capabilities. If there is no support media between the two terminals a "400 Bad Request" response with "304 Warning" header field is transmitted.

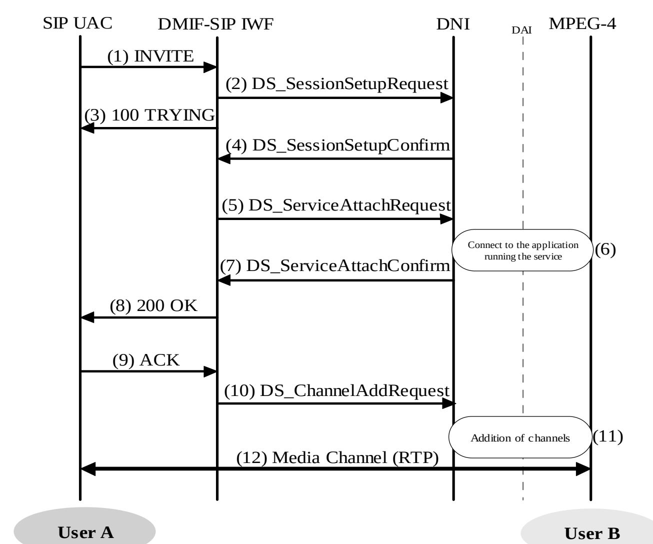

Fig. 3: SIP2DMIF Interworking

Step 9: SIP User A sends an "ACK" message to DMIF-SIP IWF

Step 10: By receiving the "ACK" message Media channel must be done. For this fact, a "DS_ChannelAddRequest" is sent to DNI of MPEG-4 Terminal.

Step 11: The MPEG-4 Terminal notifies the creation of the requested channels.

Step 12: When RTP channel is opened between SIP User A and DMIF User B, media can flows and videoconferencing can begin.

B. DMIF side

The DMIF side of the DMIF-SIP IWF is the part of the IWF that terminates and originates DMIF signalling from and to DMIF network respectively. We call this function **DMIF2SIP** (DMIF to SIP) signaling. The DMIF2SIP signaling allows a MPEG-4 DMIF Terminal to call SIP UAC. Processing steps for establishing connection between an MPEG-4 DMIF terminal and a SIP Terminal are illustrated in Fig. 4 and explained in the following:

Step 1: The MPEG-4 Terminal passes a "DA_ServiceAttach()" primitive indicating the User B address (email address, phone number, etc.). DNI layer assigns a local session to this request.

Step 2: DNI Layer sends a "DS_SessionSetupRequest" to DMIF2SIP IWF to establish a network session with SIP terminal

Step 3: Upon receiving the Session establishment request, the DMIF2SIP IWF sends an "INVITE" message to SIP Terminal to participate in videoconferencing. The "INVITE" request contains:

a). The address of User B is inserted in the "Request-URI" field as a SIP URL

b). User A address is recognized as the call session initiator in the "From" field.

c). DMIF2SIP checks whether its own address is contained in the "Via" field (to prevent loops), otherwise, it copies it own address in "Via" filed.

d). DMIF2SIP IWF creates a unique numeric identifier that is assigned to the call and is inserted in the "Call-ID" field.

e). The transaction number within a single call leg is identified in the "CSeq" filed.

f). The supported Media Capability of User A (DMIF terminal) is transformed to form a SDP message.

Step 4: After sending "TRYING" and "RINGING" messages, User B (SIP terminal) sends "200 OK" Message which notifies DMIF2SIP IWF that the connection has been established. If SIP User B supports the media capability advertised in the "INVITE" message send by DMIF2SIP IWF, it advertises the intersection of its own and User A's capability in the "200 OK" response. If SIP User B does not support the media capability advertised by DMIF2SIP IWF, it sends back a "400 Bad" Request response with a "304 Warning" header field.

Step 5: Upon receiving "200 OK" response, the DMIF2SIP IWF sends a "DS_SessionSetupConfirm" message back to MPEG-4 DMIF Terminal and specifies the common set of capability advertised by "200 OK" response.

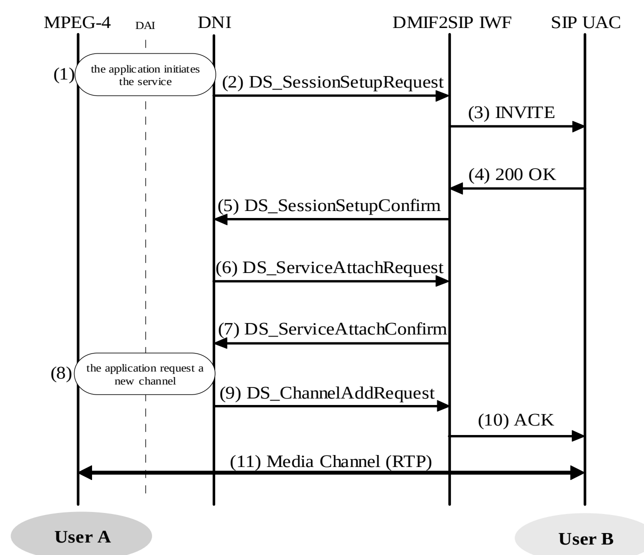

Fig. 4: DMIF2SIP Interworking

Step 6: The MPEG-4 DMIF terminal is now capable to assign a local significance of the network session and must request a service within the network session. Sending "DS_ServiceAttachRequest" message performs this task.

Step 7: DMIF2SIP IWF receives message of type "DS_ServiceAttachRequest" that identifies services at the DMIF2SIP side.

Step 8: The MPEG-4 DMIF Terminal request the establishment of a new media channel.

Step 9: The "DS_ChannelAddRequest" message sends by MPEG-4 DMIF Terminal requests the establishment of new media channel.

Step 10: DMIF2SIP IWF sends "ACK" message to SIP User B and it confirms that the two peers are capable of sending and receiving multimedia stream.

Step 11: Media channels are now initialised, media can flows between Terminals.

C. Finding common subset of capabilities between IP videoconferencing terminals

The capability set of a terminal or a user agent refers to the set of algorithms for audio, video and data that it can support. It also conveys information about constraints in the selection of algorithms it may have. For example, due to limited bandwidth, a terminal may indicate that it can use either G.711 without video or G.723.1 with H.261 video [16].

Terminal Capability matching is required to check the ability of two peer and-systems to setup connections between them, and select the supported protocol stack. In case of DMIF and SIP, this stage is performing at session setup.

SIP uses SDP as a protocol to describe SIP terminal capability.

When SIP UAC initiates the session with DMIF Terminal, it sends an “INVITE” request to DMIF. The capability descriptor is carried in this “INVITE” request throw SDP message. When DMIF Terminal initiates the session, it sends a “*DS_SessionSetupRequest*” to SIP terminal at this stage the capability descriptor is carried in the “*comptabilityDescriptor”* field part of “DS_*SessionSetupRequest*” message.

The algorithm to find the common subset of capability descriptor maximal intersection of any two-capability sets C1 and C2 is described in [17]and is given here:

1. Set the result C to the empty set.

2. For each pair of capability descriptors (d1, d2), where d1 is from C1 and d2 is from C2, derive the permutations of alternative sets, s1 and s2. For each such permutation, where s1 is from d1 and s2 is from d2, intersect s1 and s2 (written as s=s1 ^ s2) and add s to C.

3. Remove duplicate entries from C.

D. Conversion Between SIP Address And DMIF URL

SIP address can be converted to DMIF URL facially. SIP-URL is copied verbatim to DMIF URL without any additional or supplement information.

V. CONCLUSION

Diversity and heterogeneity of multimedia terminals and services characterize today IP networking environment. Consequently, interworking becomes a critical issue for resolving the differences in these elements and enabling seamless provision of audiovisual applications across networks. Interworking is a way of making different communicating systems cooperate to perform a particular service. Thus, this article emphasizes technical issues on call control and session establishment interworking between ISO DMIF-compliant terminals (e.g. MPEG-4 video servers) and IETF SIP-compliant terminals (e.g. mobile IP phones). We designed and implemented an interworking multimedia gateway that performs various translation functions between two major multimedia signaling protocols, namely DMIF and SIP, that performs session protocol conversion, service gateway conversion, and address translation. Multimedia Internet designers may then transparently combine interactive MPEG-4 multimedia content with IP telephony for advance videoconferencing services such as e-learning, e-commerce or collaborative conferencing. Internet users might well have the impression to interact with integrated IP multimedia services regardless to data content and feature location. This is what we call a “seamless interactive IP videoconferencing service interworking”.

REFERENCES

[1] M. Handley H. Schulzrinne, E. Schooler J. Rosenberg “RFC 2543: Session Initiation Protocol (SIP)”, March 1999.

[2] M. Handley, V. Jacobson, “RFC 2327 SDP: Session Description Protocol”, IETF, April 1998.

[3] IUT-T Recommendation H.323- “Visual telephone systems and equipment for local area networks which provide a non guaranteed quality of service”, November 1996.

[4] IUT-T Recommendation H.323 Draft v4– “Packet-based multimedia communications systems”, November 2000.

[5] 3GPP Technical Specification Group Core Network, “Signalling flows for the IP multimedia call control based on SIP and SDP (Release 5)”, 3GPP, December 2001.

[6] ISO/IEC 14496-1 “Coding of audio-visual objects --Part 1: Systems -final committee draft”, May 1998.

[7] ISO/IEC 14496-2 “Coding of audio-visual objects – Part 2: Visual -final committee draft” May 1998.

[8] ISO/IEC 14496-3 “Coding of audio-visual objects -- Part 3: Audio -final committee draft” May 1998.

[9] ISO/IEC 14496-6 “Coding of audio-visual objects -- Part 6: Delivery Multimedia Integration Framework (DMIF) final committee draft.”, May 1998.

[10] T. Ahmed A. Mehaoua and R. Boutaba “Interworking Between SIP and MPEG-4 DMIF”, IETF, draft-ahmed-dmif-sip-00.txt, work in progress, Sept. 2001.

[11] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson “RFC1889 RTP: A Transport Protocol for Real-Time Applications”, IETF, January 1996.

[12] H. Schulzrinne, A. Rao, R. Lanphier “RFC2326: Real Time Streaming Protocol (RTSP)”, IETF, April 1998.

[13] H. Yasuda and H.J.F. Ryan, “DAVIC and Interactive Multimedia Services”, IEEE Comm. mag; vol. 36, n°9., Sept. 1998, pp. 137-143.

[14] ISO/IEC 13818-6:1998 “Generic coding of moving pictures and associated audio information -- Part 6: Extensions for DSM-CC”, October 1998.

[15] D. Durham, Ed, J. Boyle, R. Cohen, S. Herzog, R. Rajan,w, A. Sastry “RFC 2748: The COPS (Common Open Policy Service) Protocol”, IETF, January 2000.

[16] H.Agrawal,R.R Roy, V.Palawat, A.Johnston, C.Agboh, D.Wang, H.Schulzrinne, K.Singh, J.Maeng, “SIP-H.323 Interworking”, Internet Draft, IETF, Work in progress, July 2001, available at http://search.ietf.org/internet-drafts/draft-agrawal-sip-h323-interworking-01.txt.

[17] Singh, Schulzrinne “Interworking between SIP/SDP and H.323” Internet draft, IETF, Work in progress.