                  Requirements for a media server control protocol
                       draft-even-media-server-req-01.txt

Status of this Memo

Copyright Notice

Abstract

   This document addresses the communication between an application
   server and media server. The current work in SIPPING and XCON
   working groups show these logical functions but do not address the
   physical decomposition and the protocol between the entities.

   The document presents the architecture and the requirements from the
   protocol. The document lists current work that is relevant to the
   problem.

Table of Contents

[1](#). Introduction

The IETF SIPPING conferencing framework[CARCH] presents an
architecture that is built of several functional entities. The
framework document does not specify the protocols between the
functional entities since it is considered out of scope.

There is an interest to work on a protocol that will enable one
physical entity that includes the conference/media policy server,
notification server and the focus to interact with one or more
physical entities that serves as mixer or media server.

The document will present the requirements for such a protocol. It
will address all phases and aspects of media handling in a
conferencing service including announcements and IVR functionality.

2. Terminology

   The Media Server work uses, when appropriate, and expands on the
   terminology introduced in the SIPPING conferencing framework[CARCH].
   The following additional terms are defined for use within the Media
   Server work.

   Application Server (AS) - The application server includes the
   conference policy server, the focus and the conference notification
   server as defined in
   draft-ietf-sipping-conferencing-framework.[CARCH]

   Media Server (MS) - The media server includes the mixer as defined in
   draft-ietf-sipping-conferencing-framework[CARCH] The media server
   source media streams for announcements, it process media streams for
   functions like DTMF detection and transcoding. The media server may
   also record media streams for supporting IVR functions like
   announcing participants.

3. Architecture

   The proposed architecture is composed of an application server (AS)
   and a media server (MS).

   This section does not define any specific model for the interaction
   between the AS and MS. It does assume that every interaction from
   the participant to the MS will be controlled by the AS. The MS only
   handles data and may tunnel controls received in the media streams to
   the AS (e.g. DTMF).

   The assumption is that the external protocols to these entities will
   be based on the IETF work. The Conference aware participants will
   use XCON protocols to the AS. The signaling protocol between the
   Participants and the AS will be SIP. The media between the
   Participants and the MS will be RTP based. The solution may work for
   other signaling protocols like H.323.

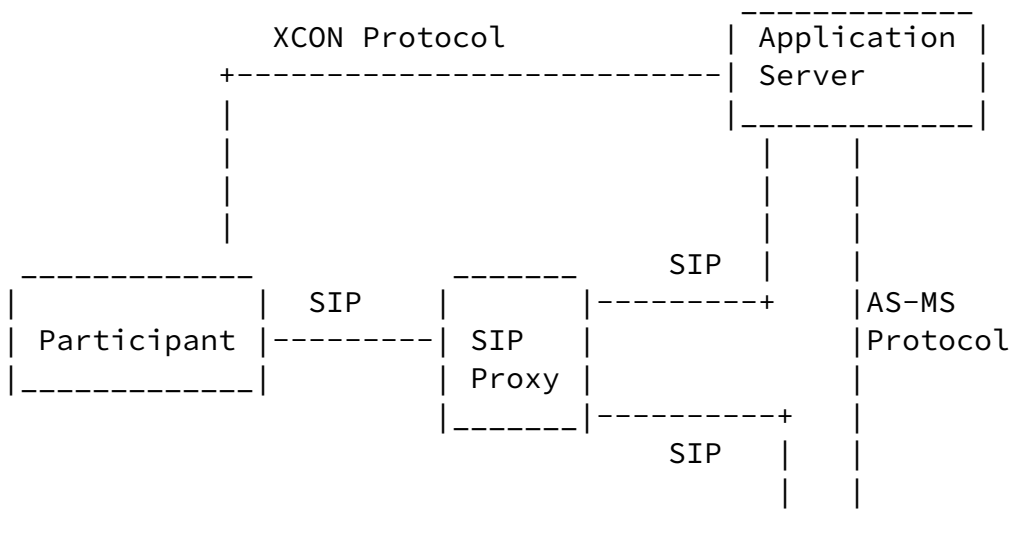   The MS functionality includes:

   - Control of the RTP streams.

- Mixing of incoming media streams.

- Media stream source (for multimedia announcements).

- Media stream processing (e.g. transcoding, DTMF detection).

- Media stream sink (Support announcing participants names)

The AS functionality includes:

- Creation and management of conferences by conference aware
participants.

- Creation of conference service logic using other mechanism which
may be standard or non-standard. Example is to create an IVR based
conference service.

- Manage the conference flow in the MS from start to finish.

The following diagram describes the architecture. The purpose of the
work is to address the AS-MS protocol.

```
                                              _____
                       XCON Protocol         | Application |
                   +--------------------------| Server      |
                   |                          |_____|
                   |                              |   |
                   |                              |   |
                   |                              |   |
        _____                   _____  SIP |   |
       |            |  SIP    |       |       |------+   |AS-MS
       | Participant |---------| SIP   |          |Protocol
       |_____|         | Proxy |          |   |
                              |_____|---------+   |
                                        SIP    |   |
                                               |   |
                                               --------
```

```
                                          |        |
                                          | Media  |
                                          | Server |
                                          |_____|
```

[4](#). Requirements

   This section addresses only the requirements. The requirements will
   be divided to general protocol requirements and to specific service
   logic requirements.

   General protocol

1. The Media server control messages shall be sent over a reliable
    connection.

2. The protocol shall enable one AS to work with multiple MS.

3. The protocol should enable many AS to work with the same MS

4. The AS should be able to find the MS and connect to it.

5. The MS shall be able to inform the AS about it status.

6. The protocol should be extendable.

7. The MS shall be able to tell the AS its capacities.

8. The MS shall be able to tell the AS its functionality (Mixing,
    IVR, Announcements)

9. The AS shall be able to request the MS to create, delete, and
    manipulate a mixing, IVR or announcement session

10. The MS shall supply the media addresses (RTP transport address)
     to be used to the AS.

11. The MS should send a summary report when the session is
     terminated by the AS.

12. The AS should be able to request call/session and conference
     state from the MS.

13. The MS should support DTMF detection (in band tones and [RFC2833](RFC2833))

14. The protocol shall include redundancy procedures.

15. The protocol shall include security mechanisms.

16. The AS should be able to reserve resources on the MS. The
     resources models should be simple. (this requirement needs more
     discussion)

17. The MS may support resource reservation and shall report the

support in the initial connection to the AS.

18.   The MS shall inform the AS about any changes in it capacities.
The changes may be due to reservation, internal usage or due to some
malfunction.

19.   The AS shall be able to tell the MS which stream parameters to
use on incoming and out going streams. Stream parameters may be for
example codec parameters (video codec features) or bit rates. This
requirement will help the MS to allocate the right resources.

20.   The AS shall be able to define operations that the MS will
perform on streams like mute and gain control.

21.   The MS shall supply the AS with sufficient information for the
event package.

Announcements

Announcements may include voice, audio, slides or video clips.

1. The AS shall be able to instruct the MS to play a specific
announcement.

2. The MS shall be able to retrieve announcements from an external
connection.

3. The AS shall be able to tell the MS if the message can be delayed
if the MS cannot play it immediately.

4. The AS shall be able to instruct the MS to play announcements to
a single user or to a conference mix.

Media mixing

1. The AS shall be able to define a conference mix.

2. The AS may be able to define a separate mix for each participant.

3. The AS shall be able to define the relationship between two
mixes, for example a pair of audio and video for lip-synch or for
voice activated video switch

4. The AS may be able to define a custom video layout built of
rectangular sub windows.

5. For video the AS shall be able to map a stream to a specific sub-

window or to define to the MS how to decide which stream will go to
each sub window. The number of sub-windows will start from one.

6. The MS shall be able to inform the AS who is the active speaker.

7. The AS may be able to cascade mixers (side bar with whisper
mode)

8. The MS shall be able to inform the AS which layouts it supports.

IVR

1. The AS shall be able to load an IVR script to the MS and receive
the result

2. The AS shall be able to mange the IVR session by sending
announcements and receiving the response (DTMF)

3. The AS should be able to instruct the MS to record a short
participant stream and play it back to the conference. This is not a
recording requirement.

5. Current protocols

   Currently there are a few protocols that try to address this
   architecture. The IETF drafts and ITU standards include:

   1. draft-vandyke-mscml-05 [MSCML]

   2. draft-melanchuk-sipping-msml-03[MSML]

   3. draft-melanchuk-sipping-moml-03[MOML]

   4. draft-burger-sipping-netann-10[NETANN]

   5. draft-levin-xcon-cccp-00[CCCP]

   6. ITU H.248.19 - Decomposed multipoint control unit, audio, video
   and data conferencing packages[ITU.H248.19]

   Note: The list is to the best of my knowledge and the order is
   meaningless

   A short overview of the drafts based on my poor understanding is
   given here please feel free to correct my mistakes.

   MSML[MSML]

   Convedia MSML (Media Sessions Mark-up Language)[MSML]. The latest
   version added support for video. MSML addresses the relationships of
   media streams MSML defines an XML schema that enables the AS to
   create sessions on the MS. The draft outlines how to use SIP as a
   transport for the XML schema by using SIP invite to create a control
   session between the AS and the MS. Subsequent control messages
   between the MS and AS will be done using INFO or INVITE. The control
   connection is only used for the transporting the XML schema.

   MSML supports several models for client interaction. When clients
   use 3PCC to establish media sessions on behalf of end users, clients
   will have a SIP dialog for each media session. MSML may be sent on
   these dialogs. However the targets of MSML actions are not inferred

from the session associated with the SIP dialog. The targets of MSML
actions are always explicitly specified using identifiers previously
defined.

The signaling from the SIP users is going to the AS. The AS is using
3pcc (third party call control) procedures to direct the SIP
signaling messages to the MS. The SDP is used to open the media
channel between the user and the MS while the call control is handled
by the AS. This is how the users join a media session on the MS that

was established using the MSML schema. Convedia has a second
protocol called Media Objects Mark-up Language (MOML)[MOML] that can
be used to specify individual user dialog or media control commands.

MSCML[MSCML]

Brooktrout Technology has suggested MSCML, Media Server Control
Mark-up Language. This current version supports only audio.

The general functionality is similar to MSML. MSCML is using SIP
Invite and Info to send the communication between the AS and MS.
Like MSML it opens a control connection for the conference. The
difference is that MSCML messages sent in the control connection are
for the entire conference while if they are sent over the users
dialogs they apply to that user.

Netann[NETANN]

This protocol provides a simple way to initiate an announcement, IVR
or mixing session on the media server using the URI parameters. The
AS can create the session bur has less control on what is happening
during the session itself.

Centralized Conference Control Protocol (CCCP)[CCCP]

This document may be of interest since it suggests a transaction
protocol that can serve the AS to MS communication. The data schema
is based on the conference event package.

MEGACO / H.248[ITU.H248.19]

The H.248[ITU.H248.19] protocol opens a channel between a controller

and a device (these can be AS and MS in our implementations). In
this channel it sends command to the MS to create context and to open
connections (terminations) for the media channel. The specific
functionality is defined by packages that can be extended. The MS
connects to it AS when it starts up and notify the AS which packages
it supports and its capacities. The H.248 packages include also
support for announcement and IVR.

[6](#).   IANA consideration

   None.

7. Security Considerations

   The security section will be added later

8. Informative References

   [CARCH]   Rosenberg, J., "A Framework for Conferencing with the
               Session Initiation Protocol",
               draft-ietf-sipping-conferencing-framework-03 (work in
               progress), October 2004.

   [CCCP]   Levin, O. and G. Kimchi, "Centralized Conference Control
               Protocol (CCCP)", draft-levin-xcon-cccp-00 (work in
               progress), October 2004.

   [ITU.H248.19]
               International Telecommunications Union, "Gateway control
               protocol: Decomposed multipoint control unit, audio, video

                    and data conferencing packages", ITU-T Recommendation
                    H.248.19, March 2004.

   [MOML]   Sharratt, G. and T. Melanchuk, Ed., "Media Objects Markup
                Language (MOML)", draft-melanchuk-sipping-moml-03 (work in
                progress), August 2004.

   [MSCML]   Van Dyke, J. and Eric. Burger, Ed., "Media Server Control
                Markup Language (MSCML) and Protocol",
                draft-vandyke-mscml-05 (work in progress), October 2004.

   [MSML]   Sharratt, G. and T. Melanchuk, Ed., "Media Server Markup
                Language (MSML)", draft-melanchuk-sipping-msml-03 (work in
                progress), August 2004.

   [NETANN]   Van Dyke, J. and Eric. Burger, Ed., "Basic Network Media
                Services with SIP", draft-burger-sipping-netann-10 (work
                in progress), October 2004.

Author's Address

   Roni Even
   Polycom
   94 Derech Em Hamoshavot
   Petach Tikva 49130
   Israel

   Email: roni.even@polycom.co.il

Intellectual Property Statement

   The IETF takes no position regarding the validity or scope of any
   Intellectual Property Rights or other rights that might be claimed to

pertain to the implementation or use of the technology described in
this document or the extent to which any license under such rights
might or might not be available; nor does it represent that it has
made any independent effort to identify any such rights. Information
on the procedures with respect to rights in RFC documents can be
found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any
assurances of licenses to be made available, or the result of an
attempt made to obtain a general license or permission for the use of
such proprietary rights by implementers or users of this
specification can be obtained from the IETF on-line IPR repository at
http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any
copyrights, patents or patent applications, or other proprietary
rights that may cover technology that may be required to implement
this standard. Please address the information to the IETF at
ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an
"AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS
OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET
ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED,
INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE
INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED
WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement

Copyright (C) The Internet Society (2006). This document is subject
to the rights, licenses and restrictions contained in BCP 78, and
except as set forth therein, the authors retain all their rights.

Acknowledgment

Funding for the RFC Editor function is currently provided by the
Internet Society.