MEDIACTRL BOF

Media Server Control

Eric Burger
e.burger@ieee.org

Remote Logistics

- mp3 Feed: SPEAK INTO THE MICROPHONE!!!
  - http://videolab.uoregon.edu/events/ietf/ietf673.m3u
- Jabber
  - Server: jabber.ietf.org
  - Room: mediactrl
  - Logs: http://www.ietf.org/meetings/ietf-logs/mediactrl/
- Meeting Materials (Presentations, Agenda, etc.)
  - https://datatracker.ietf.org/public/meeting_materials.cgi?meeting_num=67
- Supplemental Web Site and Mail List Info / Archive
  - http://flyingfox.cantata.com/i-d/mediactrl

Scribes and Transcribes

?? For Jabber
?? For Minutes

Note Well

Any submission to the IETF intended by the Contributor for publication as all or part of an IETF Internet-Draft or RFC and any statement made within the context of an IETF activity is considered an "IETF Contribution". Such statements include oral statements in IETF sessions, as well as written and electronic communications made at any time or place, which are addressed to:

- the IETF plenary session,

- ## any IETF working group or portion thereof,

- the IESG or any member thereof on behalf of the IESG,

- the IAB or any member thereof on behalf of the IAB,

- any IETF mailing list, including the IETF list itself, any working group or design team list, or any other list functioning under IETF auspices,

- the RFC Editor or the Internet-Drafts function

All IETF Contributions are subject to the rules of RFC 3978 (updated by RFC 4748) and RFC 3979.

Statements made outside of an IETF session, mailing list or other function, that are clearly not intended to be input to an IETF activity, group or function, are not IETF Contributions in the context of this notice.

Please consult RFC 3978 (and RFC 4748) for details.

Agenda

- Agenda Bashing
- Brief History of IETF Work in the Area
- What We Are About: Requirements Document
- How We Will Get There: Framework Document
- Charter Discussion

Brief History

- Independent Drafts for SIP-Control of Media Servers
  - Netann (RFC 4240) introduced in 2001
  - MSCML introduced in 2002
  - MSML introduced in 2003
  - MSCP introduced in 2005

- Ad Hoc Work Group
  - Formed Just Before XCON in 2003
  - Mail List in 2004
  - Met face-to-face at every IETF, starting with IETF 62
  - 11 active drafts

- Supplemental Web Site and Mail List Info:
  `http://flyingfox.cantata.com/i-d/mediactrl`

Problem Space

Media Control Protocol Requirements
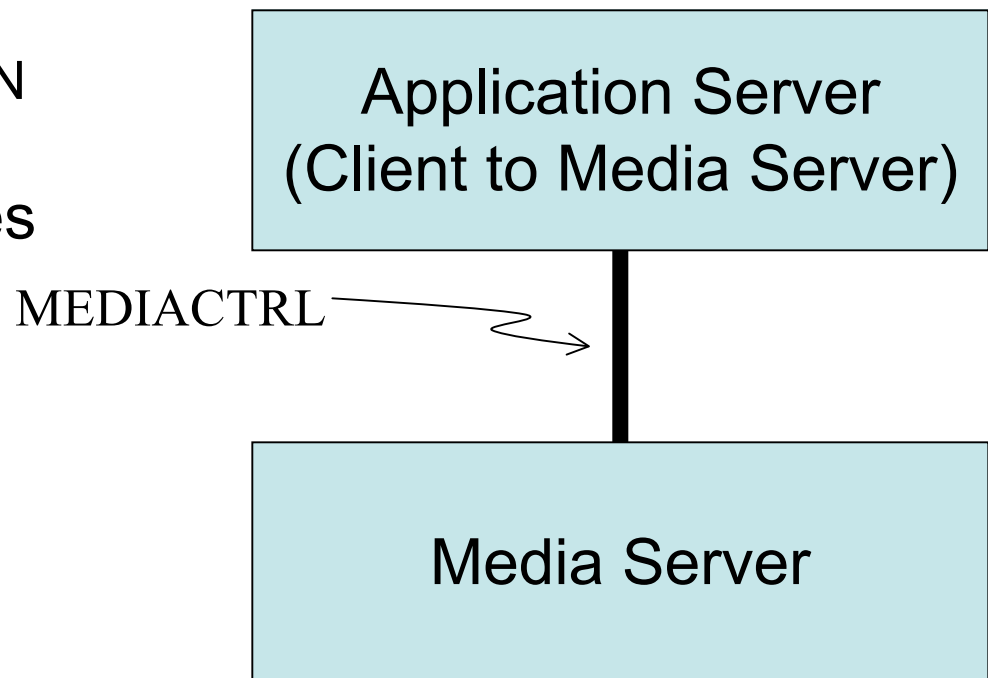
draft-dolly-xcon-mediacntrlframe-02

Dolly, Munson, and Rafferty

Market / Technical Needs

- High Level Protocol for Creating of Rich, Real-Time Interactive Multimedia Applications
  - Rules Out H.248
  - Stimulus Markup Model Leads to SIP For Session Initiation(draft-ietf-sipping-app-interaction-framework-05)

- Conference/Mixing Control
  - VoiceXML Excellent for User Interfaces
  - VoiceXML, By Design, Does Not Help for Conference Control / Topology (Level Below XCON)

- Need Protocol and Markup for Conference Control and Light IVR for Audio / Video Media Servers

Components

- ## AS, MS
  - Example of AS is XCON Conference Focus
- ## Implemented Interfaces to Date
  - SIP + MSCML, MSML, MSCP
  - H.248
- ## Others Considered
  - MRCP
- ## All SIP Implementations Agree On:
  - RFC 4240 (netann)
  - VoiceXML

Application Server (Client to Media Server)

MEDIACTRL

Media Server

Requirements

- Described Fully in draft-dolly-xcon-mediacntrlframe-02
- Things Like
  - Cardinality of AS:MS
    - Shared Resources, Multiples Resources
    - Independence of MS from Application
  - High Availability Support
  - Server in the Internet, Not Just IP Interfaces on Intelligent Peripheral
    - Servers Across Administrative Domains
    - Internet Scale
    - Privacy and Security Must Be Addressed

Solution Framework

A Control Framework for the Session
Initiation Protocol (SIP)

draft-boulton-sip-control-Framework-04

Boulton, Melanchuk, McGlashan, Shiratzky

Overview

- **Framework makes use of inherent SIP Protocol mechanisms**
  - Session Agnostic
  - Security Mechanisms
  - Service Location
    - RFC 3263 – Locating Sip Servers
    - RFC 3840 – User Agent Capabilities
    - RFC 3841 – Caller Preferences
  - Connection Establishment/Maintenance
    - RFC 4145 – COMEDIA

Extant Use of SIP

- Current SIP-based proposals use SIP INFO for transport of Media Server control PDU's

- Not intended use of SIP INFO

- Commands traverse signalling path

- UDP Default – Reliable protocol preferred

- MTU conversion on large payloads not ideal

Framework Approach

- Use SIP for intended purpose – Session Initiation and management

- MEDIACTRL Model Based on MSRP/MRCPv2

  – Use COMEDIA to establish control channel

  – MEDIACTRL PDU's traverse control channel

  – Use existing SIP extensions for service location, etc.

Charter Discussion

What is a Media Server?

- Real-time multi-media applications often need the services of media processing elements. It is true that modern endpoints are capable of media processing. However, the physics of some media processing applications dictate that it is much more efficient for the media processing to occur at a centralized location. By media processing, we mean media mixing, recording and playing media, and interacting with a user in the audio or video domains. The commercial market calls these media processing network elements "media servers."

Why Media Servers?

- Some services achieve significant efficiencies when a central node performs media processing. Because of these efficiencies, media servers are widely used for conference mixing, multimedia messaging, content rendering, and speech, voice, key press, and other audio and video input and output user interface modalities. Given the wide acceptance of the media server, we need a standard way to control them.

What do Media Servers Do?

- A media server contains media processing components that are able to manipulate RTP streams. Typical processing includes mixing multiple streams, transcoding a stream (e.g., from G.711 to MS-GSM), storing or retrieving a stream (e.g., from RTP to HTTP), detecting tones (e.g., DTMF), converting text to speech, and performing speech recognition. Note that an MRCPv2 server may offer the low-level processing for the last two services, where the media server is a client to the MRCPv2 server. Also note it is common to call the package of detecting user input, recording media, and playing media "Interactive Voice Response," or IVR.

With Respect to MRCPv2

- A media server contains media processing components that are able to manipulate RTP streams. Typical processing includes mixing multiple streams, transcoding a stream (e.g., from G.711 to MS-GSM), storing or retrieving a stream (e.g., from RTP to HTTP), detecting tones (e.g., DTMF), converting text to speech, and performing speech recognition. Note that an MRCPv2 server may offer the low-level processing for the last two services, where the media server is a client to the MRCPv2 server. Also note it is common to call the package of detecting user input, recording media, and playing media "Interactive Voice Response," or IVR.

- Media services offered by the media server are addressed using SIP mechanisms, such as described in RFC 4240. Media servers commonly have a built-in VoiceXML interpreter. VoiceXML describes the elements of the user interaction, and is a proven model for separating application logic (which run on the clients of the media server) from the user interface (which the media server renders). Note this is a fundamentally different interaction model from MRCPv2, where media processing engines offer raw, low-level speech services.

What We Will and Won't Do

- Since the media server is a centralized component, the work group will not investigate distributed media processing algorithms or control protocols.

- The work group will examine protocol extensions between media servers and their clients. However, modifying existing standard protocols, such as VoiceXML or SIP towards clients or MRCPv2 towards servers, is not in the work group's charter.

Will and Won't Do (Continued)

- The model of interest to this group is where the endpoint solely plays audio or video, transmits audio or video towards the server, and possibly transmits key press information towards the server. Alternate architectures, where the endpoint executes user interface commands, is outside the scope of the work group. For example, WIDEX/BEEP, with its distributed user interface description, is not in scope.

- The only model of user interface processing the work group will consider is where the media server performs all of the media processing. A caveat here is the media server, in interpreting a VoiceXML page, may make MRCPv2 requests to a MRCPv2 server for speech services. However, to the media server client and the media end point, the single point of signaling and media interaction is the media server.

An Internet Protocol

- Any protocol developed by this group will meet the requirements for Internet deployment. This includes addressing Internet security, privacy, and scale. The protocol will not assume a private administrative domain, also known as a "walled garden."

- There is broad market acceptance of the stimulus/markup application design model for the application server - media server protocol interface. Thus this work group will focus on the use of SIP and XML for the protocol suite.

Deliverable #1

- A requirements document. This document will identify and enumerate requirements for a suite of media server control protocols. Given that one of the common media server clients is a conference application server, we will consider the application server - media server requirements developed by the XCON work group. Likewise, we will consider media server control requirements from other standards groups, such as 3GPP SA2 and CT1.

- Candidate: draft-dolly-xcon-mediacntrlframe-02

Deliverable #2

- A framework document. This document will describe the different network elements, their interrelationship, and the broad set of message flows between them.

- Candidate: draft-boulton-sip-control-framework-04

Deliverable #3

- A protocol suite describing the embodiment of the framework document. There may be separate protocol PDU's for audio conference control, video conference control, interactive audio (voice) response, and interactive video (multimedia) response. The separation and negotiation of different PDU's is a working group topic. However, there will be one and only one (class) of PDU's defined by the work group.

- Candidates
    - draft-vandyke-mscml-09
    - draft-saleem-msml-02
    - draft-mcglashan-mscp-02
    - draft-boulton-ivr-control-package-02
    - draft-boulton-ivr-vxml-control-package-01

Deliverable #4

- Means for locating, and possibly establishing sessions to, media servers with appropriate resources at the request of clients. By appropriate, we mean the characteristics of a given media server required or desired for handling a given request and the overall efficient use of a pool of media servers. The expectation is such a means would build upon existing SIP, SNMP, and other protocol facilities. Such a means may or may not be an integral part of the item 3 deliverables above.

Related Activities In the IETF

- Given the above-mentioned conferencing example, the work of this group is of interest to the XCON work group, as this protocol will describe the "Protocol used between the conference controller and the mixer(s)." Thus we expect to work closely with XCON.

- If the work group requires extensions to SIP, the work group will forward those extensions to the SIP work group for consideration and refinement.

Related Activities Outside IETF

- The protocol suite also is an embodiment of the ISC/Mr interface from the 3GPP IMS architecture. Thus we expect to liaise with, and gather requirements from, 3GPP.

- This work may be of interest to the ITU-T, as it is a more modern embodiment of the conference control models put forth by H.248.

Milestones

- Mar 2007 - Requirements Document
- Jun 2007 - Framework Document
- Nov 2007 - Conference Control Protocol
- Mar 2008 - IVR Control Protocol
- Jun 2008 - Broker Protocol or BCP

Discussion

What is Out There

- H.248
  - **Media Gateway Control**, not Media Server Control
  - Programming Model Very Different and Difficult
  - Overkill for Solution Space
  - Not Commercially Accepted
- MRCPv2
  - Optimized for ASR/TTS/SI/SV Resources
  - Not Optimized for Conference Control, non-Speech Resources; IVR Possible but Complex
- SIP+MSCML/MSML/MSCP
  - Commercially Accepted
  - Similar Concepts
  - 5, 3, 1 years of road time (working code)
- Everyone Agrees on RFC 4240 (netann)