Media Resource Control Protocol v2

Sarvi Shanmugham,
Editor: MRCP v1/v2
Technical Leader, Cisco Systems
Roadmap

• Overview of the IETF Speechsc WG Effort
• MRCP – Short Summary
• MRCP – Architecture Diagram
• MRCP - Usage
• MRCP v1 & v2 – Current Status
Overview of the IETF Speechsc WG Effort

• IETF Working group - formed in 2002
• Aimed to develop a protocol that allows distributed speech processing (speech recognition, speaker recognition, verification and text-to-speech)
• Work with VoiceXML and SALT
• Leverage existing protocols as much as possible
• Leverage existing W3C standards for markup
MRCP – Short Summary

• Control Plane only
  Media transmission and media pipe setup not addressed. Uses another protocol such as RTP/RTCP instead.
  Client/Server style of interaction
  Messages, format, headers and resource state-machines based on MRCPv1
  Uses a separate TCP/TLS pipe for MRCP message communication.

• “Embedded” Protocol Model
  Rendezvous and session setup done with SIP
  Uses SIP and SDP to setup the media pipe.
  Uses SIP and SDP to setup a separate MRCP control channel for each resource in a session.
  Uses SIP and SDP to negotiate the establishment Establish separate TCP or TLS pipe to communicate MRCPv2 messages.
MRCP – Short Summary (contd.)

• Basic Speech Services defined
  
  Speech Recognition
  Text-to-Speech
  Speaker Identification
  Speaker Verification
  Recording
MRCP – Short Summary (contd.)

• Makes use of W3C standards for markup

• SSML
  Speech Synthesis Markup Language
  Input to TTS Engines

• SRGS
  Speech Recognition Grammar Specification
  Input to ASR Engines

• NLSML
  Natural Language Semantic Markup Language
  Output from ASR Engines
MRCP – Short Summary (contd.)

- MRCPv2 defines some additional XML markup not yet addressed by the W3C.
- Recognition Results – XML markup based on an early draft of NLSML
- Additional support in the XML result markup for Speaker Identification
  Speaker Verification
MRCP – Architecture Diagram

SpeechSc Client

Application Layer
- Media Resource API
- SIP Stack
- MRCPv2
- TCP/IP Stack

Media Source/Sink

SpeechSc Server

- SIP Stack
- MRCPv2
- TCP/IP Stack
- Media Resource Management
- TTS Engine
- ASR Engine
- SV Engine
- SI Engine

SIP

MRCPv2

RTP
Use Case: VXML-based ASR

- Users call into the service in order to obtain stock quotes.
- Media Server fetches VoiceXML to drive user interaction.
- Media Server INVITEs Speechsc server for ASR
- VoiceXML interpreter on the Media Server directs the user's media stream to the ASR server and uses MRCPv2 to control the ASR server.
- Results come back and the application proceeds.
Use Case: Speaker Verification

- A user speaks into a SIP phone to "log in" to that phone to make and receive phone calls using his identity and preferences.
- IP phone uses SIP and MRCPv2 to set up an RTP stream between the phone and the SPEECHSC SI/SV server and request verification.
- SV server verifies the user's identity and returns the result via MRCPv2.
- The IP Phone may either use the identity directly to identify the user in outgoing calls, to fetch the user's preferences from a configuration server, request authorization from a AAA server, etc.
Current WG Status

• Requirements Document passed IESG Review - soon to be published as an RFC
  
  draft-ietf-speechsc-reqts-05.txt

• MRCPv2 Protocol Document in second revision - expect last call in late fall
  
  draft-ietf-speechsc-mrcpv2-04.txt

• MRCPv1 Protocol Document is pending IESG review for publication as an Informational RFC.
  