GENESYS™

GVP Deployment Guide

How the MRCP Proxy Works

4/21/2025

How the MRCP Proxy Works

This section provides information about the following topics, to explain how the MRCP Proxy (MRCPP) Server performs its role in a GVP deployment:

- Operational Overview
- Resource Management
- High Availability
- Data Collection and Logging
- Support for Multiple Speech Servers

Important

MRCP Proxy means only MRCP v1 Proxy. For information on MRCP v2 Proxy, see How the MRCP v2 Proxy works.

Operational Overview

MRCP Proxy accepts client requests (from Media Control Platform) and sends requests to ASR and TTS speech servers by using MRCPv1.

MRCP Proxy also supports a subset of RTSP over persistent TCP connections for client or server interactions. MRCPv1 does not require full support of RTSP, therefore, SETUP, TEARDOWN, DESCRIBE, and ANNOUNCE requests only are supported.

Resource Management

MRCP Proxy obtains a list of MRCPv1 resources from Management Framework to maintain an up-to-date picture of the resource pool. ASR and TTS speech server are added as connections in the MRCP Proxy Application object and become the resource access point. The MRCP Proxy uses the information that is configured in the provision section of the speech resource Application object to determine how the requests for resources will be routed.

Resource Load Balancing

MRCP Proxy supports round-robin load balancing for eligible speech resources in the deployment. Eligibility is determined in the following manner:

- MRCP Proxy sends ping messages (RTSP DESCRIBE) regularly to the speech resources.

- If the ping fails, the speech resource is isolated and set to the unavailable state. Requests are not routed to that resource until the resource becomes available. The time between ping requests is controlled by the provision.vrm.proxy.ping_interval parameter in the ASR or TTS resource configuration option.

- Only the speech resources with the engine name (for example, provision.vrm.client.resource.name) that are requested by the Media Control Platform are considered for matching.

Resource and Application Updates

The MRCP Proxy receives and processes periodic updates from Management Framework for its configured Application objects and resources in the following way:

- If it receives an update from the Management Framework for its own Application object, all changes to the vrmproxy.timeout.* parameters are accepted and the MRCP Proxy uses the new timeout values.

- If it receives an update that a new connection is added to its Application object, and the connected object is identified as an ASR or TTS resource, the resource is added to its resource pool.

- If it receives an update that an existing connection is deleted from the its Application object, and the connected object is identified as an ASR or TTS resource, the resource is deleted from the resource pool.

- If it receives an update that the provision section of an ASR or TTS resource object has changed, it takes effect immediately.

High Availability

MRCP Proxy can be deployed in HA warm active standby mode. Two servers are required for this configuration one configured as the primary server and the other as the backup server. Management Framework s Solution Control Server determines which server is active and which is on standby at any given time.

The standby MRCP Proxy puts itself into suspended mode and does not submit data to the Reporting Server nor does it respond to incoming RTSP requests; only the active server performs these functions.

When failover occurs, the existing TCP connections are terminated and all of the existing ASR/TTS sessions are lost. In addition, when the standby machine becomes active, the peak ASR and TTS usage counter is reset to zero.

Data Collection and Logging

MRCP Proxy uses the Operational Reporting (OR) interface to send ASR and TTS usage data to the Reporting Server in real time. It submits the following usage information:

- ASR and TTS peak data for a specific speech resource

- ASR and TTS arrival data for a specific speech resource

- ASR and TTS peak data for a specific tenant

- ASR and TTS arrival data for a specific tenant

- ASR and TTS peak data for a specific IVR Profile

- ASR and TTS arrival data for a specific IVR Profile

- ASR and TTS peak data for the entire deployment

MRCP Proxy reports ASR and TTS usage data for tenants, IVR Profiles, or the entire deployment. The MRCP Proxy receives the tenant and profile information from the Media Control Platforms for each speech resource request. This information is sent to the Media Control Platform originally from the Resource Manager, based on the tenant/profile mapping.

Support for Multiple Speech Servers

MRCP Proxy supports Primary and Backup speech server lists, based on the configuration.

- MRCP Proxy will route the request to one of the primary speech servers in round-robin fashion.

- If none of the speech servers listed in the primary group is available, then MRCP Proxy will route the request to one of the backup speech servers.

MRCP Proxy supports connecting Speech Servers in the connection tab. See the procedure Adding a Speech Server as Primary or Backup.