Load Balancing Logic on Cisco Meeting Server

Contents

Introduction

This document describes the load balancing logic of the Cisco Meeting Server (CMS) (formerly Acano product) which is covered on the Load Balancing white paper. This document visualizes this process in a flowchart and goes in more detail on the selection algorithm.

Prerequisites

Requirements

Cisco recommends that you have knowledge of these topics:

- Cisco Meeting Server Call Bridge component (and clustering of it)
- Cisco Meeting Server API configuration

Components Used

The information in this document is based on Cisco Meeting Server, Version 2.4.x.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.

What is the Load Balancing Algorithm of the CMS?

Load balancing has been introduced in Version 2.1 of CMS in order to make efficient use of conference resources. It tries to minimize the number of distribution calls between the Call Bridges

that host the same space. This mechanism is based on the Replaces header in Session Initiation Protocol (SIP) and is supported in Cisco Unified Communications Manager (CUCM) as the call control. It is also supported with Expressway Version X8.11 (or later), in combination with a CMS Version 2.4 or later. CMA calls (both thick client and WebRTC type) can be load balanced as well from CMS Version 2.3 onwards.

> **Note**: Load balancing of Lync/Skype calls is not supported in any CMS version at this moment in time and thus this flowchart does not apply.

> **Note**: The Load Balancing logic only applies for calls to CMS spaces and thus not for gateway calls (P2P calls) or dual-home calls at this moment in time.

The load balancing process is highlighted in the [white paper](#) in the section **How load balancing uses the settings** under **Configuring Call Bridges for load balancing incoming calls.** It is shown in text format and is visualized here in the flowchart (download).

The flowchart makes use of some abbreviations and terminology:

- CB = Call Bridge
- ExistingConferenceLoadLimit = existingConferenceLoadLimitBasisPoints * loadLimit
  (by default the existingConferenceLoadLimitBasisPoints equals 8000, which corresponds to 80%)
- NewConferenceLoadLimit = newConferenceLoadLimitBasisPoints * loadLimit
  (by default the newConferenceLoadLimitBasisPoints equals 5000, which corresponds to 50%)

If MediaProcessingLoad is referenced, it is seen in regards to that particular Call Bridge where the call has landed. This load value can be verified with an API GET on /system/load in real-time and gives a representation of the actual load processed by this Call Bridge at that moment in time.