

# MMoH through CUBE Operation, Configuration, and Troubleshoot Guide



Document ID: 116392

Contributed by Bakthavatsal Muralidharan, Cisco TAC Engineer.  
Sep 12, 2013

## Contents

### Introduction

#### Prerequisites

- Requirements
- Components Used

#### Background

- MoH Overview
- Disable the Media Stream
- Connect to MoH
- When Media Resources are Used in a Call
- Resume the Call
- SDP Attribute

#### MoH on CUBE

- Codec Considerations

#### MMoH

- SIP Direction Attribute Manipulation
- Address Manipulation
- Stream from a Flash
- Stream from a Live-Feed
- Configure MMoH
- When MTP is Used in a Call

#### Performance Considerations

#### Restrictions

#### Troubleshoot

- Show and Debug Commands
- Scenario 1
- Scenario 2
- Scenario 3
- Scenario 4

#### Related Information

## Introduction

This document describes operation, configuration, and troubleshooting information for Multicast Music-on-Hold (MMoH) through Cisco Unified Border Element (CUBE).

Although the focus of this document is *multicast* Music-on-Hold (MoH), a substantial part is devoted to describing how MoH works in general. This additional information helps build a base-knowledge for the beginner in order to better recognize and appreciate the issues that are specific to MMoH.

**Note:** While the principles are same, Cisco Unified Border Element-Service Provider Edition (CUBE-SP) does not fall within the scope of this document, nor does CUBE usage in environments that do not involve

Cisco Unified Communications Manager (CUCM).

## Prerequisites

### Requirements

There are no specific requirements for this document.

### Components Used

This document is not restricted to specific software and hardware versions.

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, make sure that you understand the potential impact of any command.

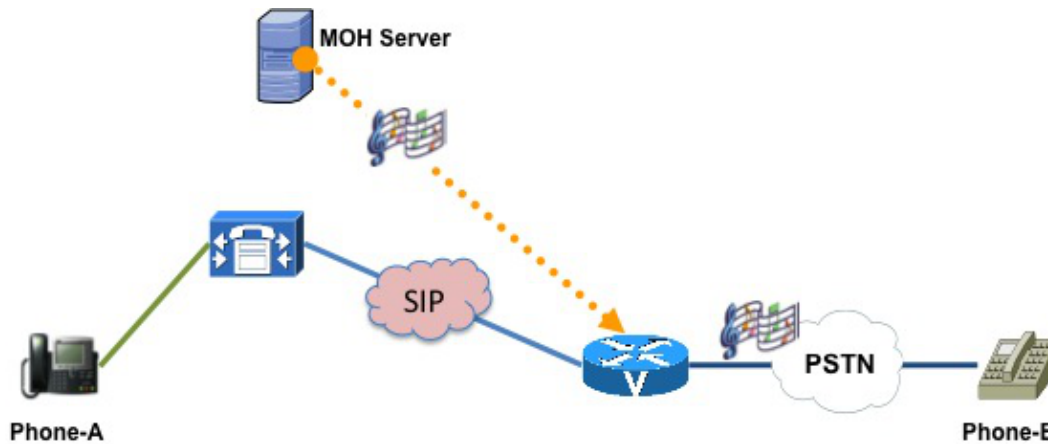
## Background

*Note:* With the exception of a couple of scenarios that are illustrated for H.323, Session Initiation Protocol (SIP) signaling is used throughout most of this document.

### MoH Overview

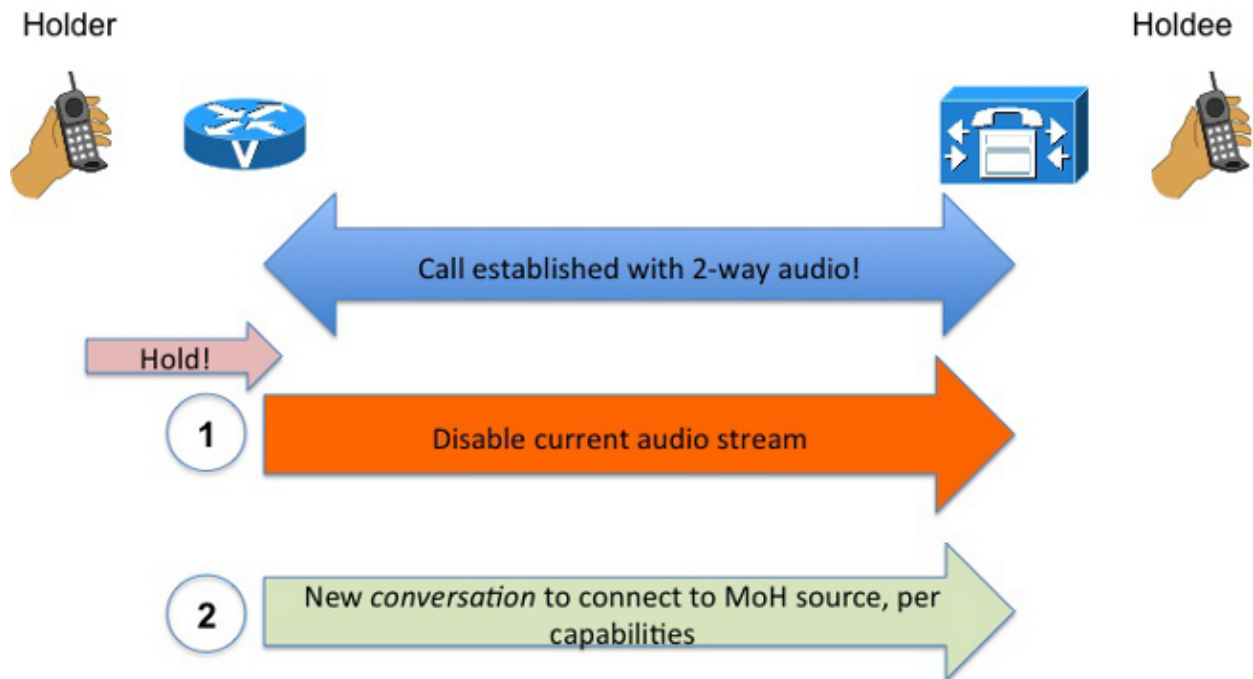
MoH is played whenever a caller is placed on–hold. Call–hold is initiated by either the user or by the network when a supplementary service process is implemented, such as call forward or transfer. The former is referred to as *user–initiated hold*, *user–hold*, or *user hold*. The latter is referred to as *network–initiated hold*, *network–hold*, or *network hold*.

Here is a review of how MoH works with Time Division Multiplexing (TDM) Gateways. This image illustrates the components and connections involved in a call–hold scenario:



1. Phone-A(Holder) puts Phone-B(Holdee) on Hold
2. CUCM tells Gateway to listen to MOH audio source
3. Gateway receives the MOH and passes it on to Phone-B

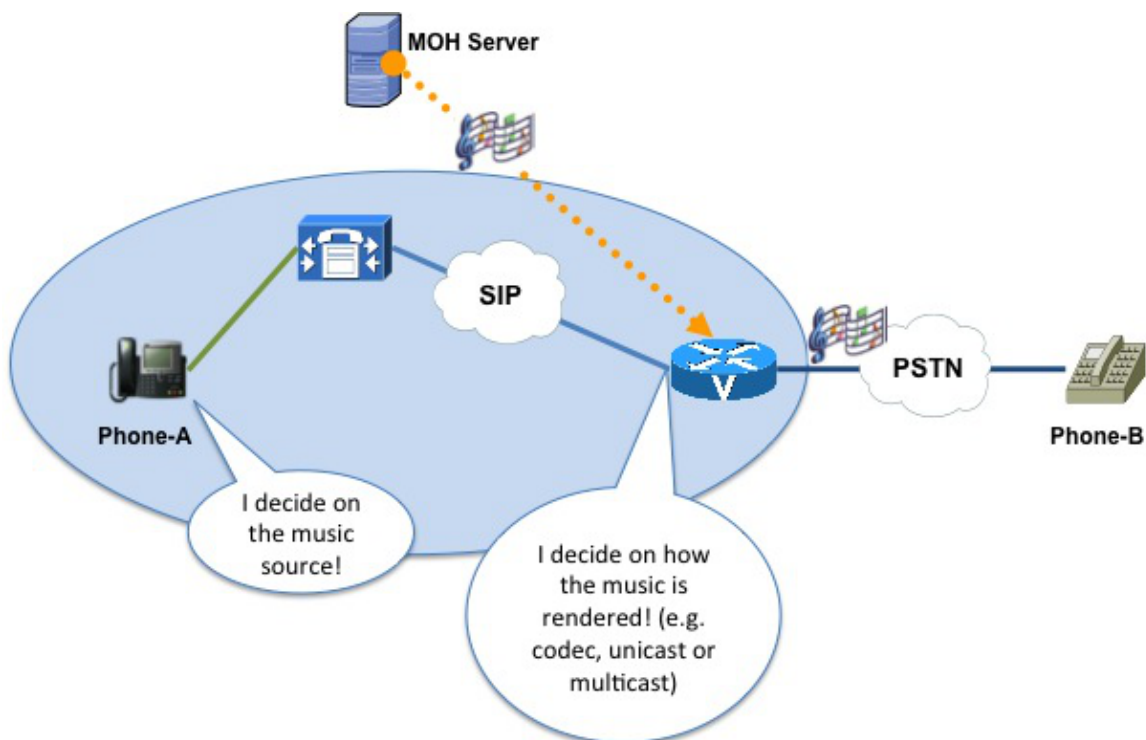
In order to place a call on–hold, a two–step process is needed. This image illustrates the two steps involved:



**Tip:** Remember this two-step process when you attempt to sort through MoH configuration and troubleshoot issues.

### MoH Sources

The user who places a call on-hold is referred to as the **holder**, and the user who is placed on-hold (and hears MoH) is referred to as the **holdee**. Each side decides certain aspects of the music that is played.



The music source is determined by the **holder**. The determination follows this hierarchy:

1. The music source configured on the Domain Name (DN)

2. The music source configured on the device
3. The music source on the device profile (user–hold music source only)
4. The music source at the global level (service parameter, or example)

There are two sets of music sources, called user–hold and network–hold. Whenever reference is made to music source, it could mean either a user–hold or network–hold music source.

### ***MoH Endpoints***

For MoH purposes, the endpoint on the CUCM side is the MoH server. This is important to understand because codec determination (based on inter–region codec configuration) is based on:

- The MoH server region
- The trunk/gateway region

The general recommendation is to assign the MoH server a dedicated region, so that inter–region codec between that region and all other regions is *g.711* (or other codec that you want to stream out for MoH).

From the CUCM perspective, the endpoints involved in the call are not the two phones, but rather:

- The IP phone registered to CUCM
- The gateway/CUBE

Thus, CUCM treats the trunk that points to the gateway/CUBE in question as the endpoint, and looks into the resources associated with it in order to determine how to render the music stream.

### ***MoH VoIP Protocol***

MoH, by definition, is a one–way audio conversation. How it is signaled depends on the VoIP protocol used. For example, on SIP, this is conveyed via the *direction* attribute. On H.323, the CUCM specifies *00000000* as the network address and *0* as the port (tsapIdentifier) of the MoH server in the H.245 Open Logical Channel Ack (OLCAck) message.

*Note:* For MMoH, CUCM sends the multicast address (239.1.1.1, for example) as the network address.

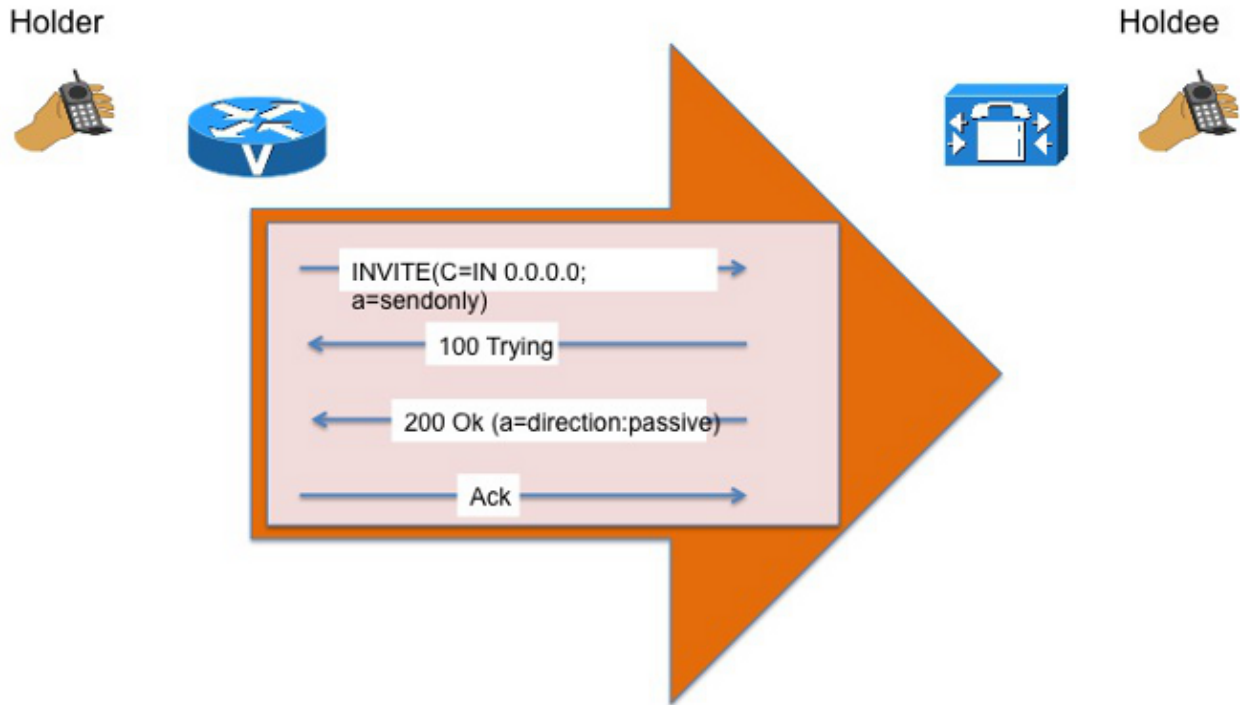
In call flows that involve CUBE, the CUCM has no knowledge about the call–leg between CUBE and Internet Telephony Service Provider (ITSP). The CUCM is only concerned with the call–leg between the IP phone and the SIP trunk (leading to CUBE).

The process of signaling for MoH is similar to signaling for a new conversation, with reduced scope. In SIP, for example, the conversation takes place within the context of the dialogue that already exists .[1]

## **Disable the Media Stream**

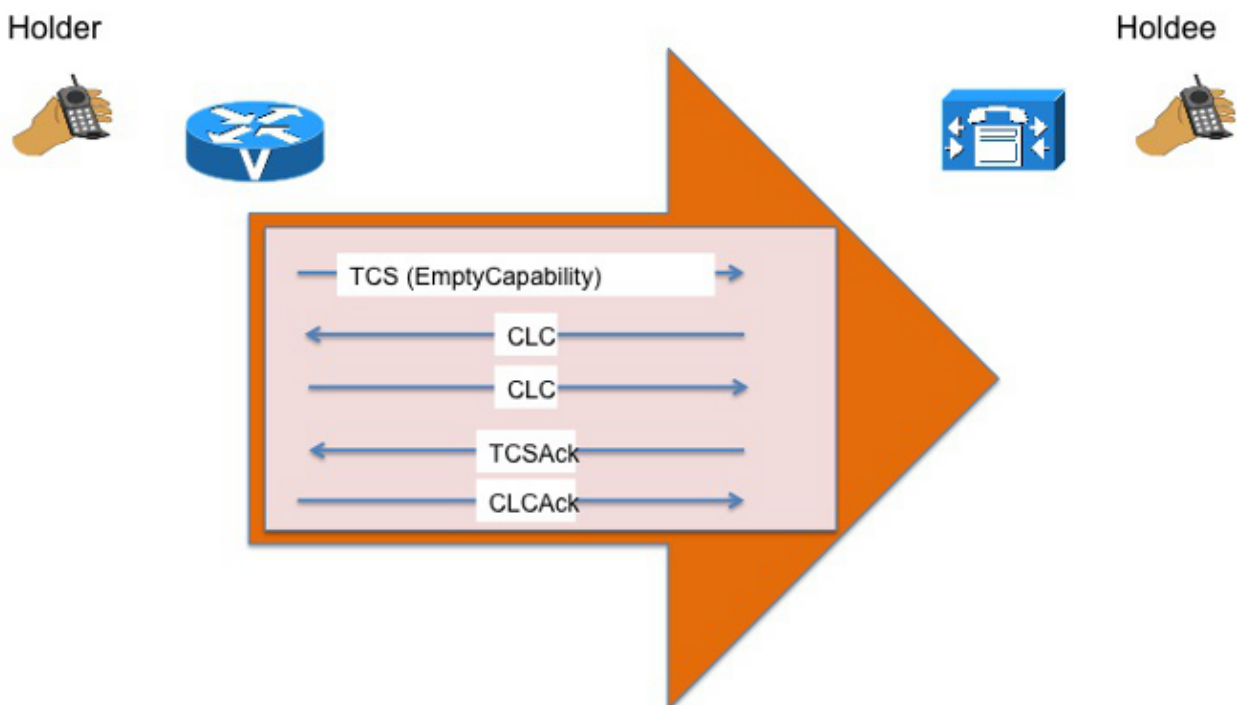
The first step in the previously mentioned two–step process is to disable the media stream.

This image illustrates how the media stream is disabled in SIP:



SIP implementations vary as to whether one or both attributes (*?a=?* and *?C=IN ?*) are used in order to indicate that the media stream is disabled.

This image illustrates how the media stream is disabled in the H.323:



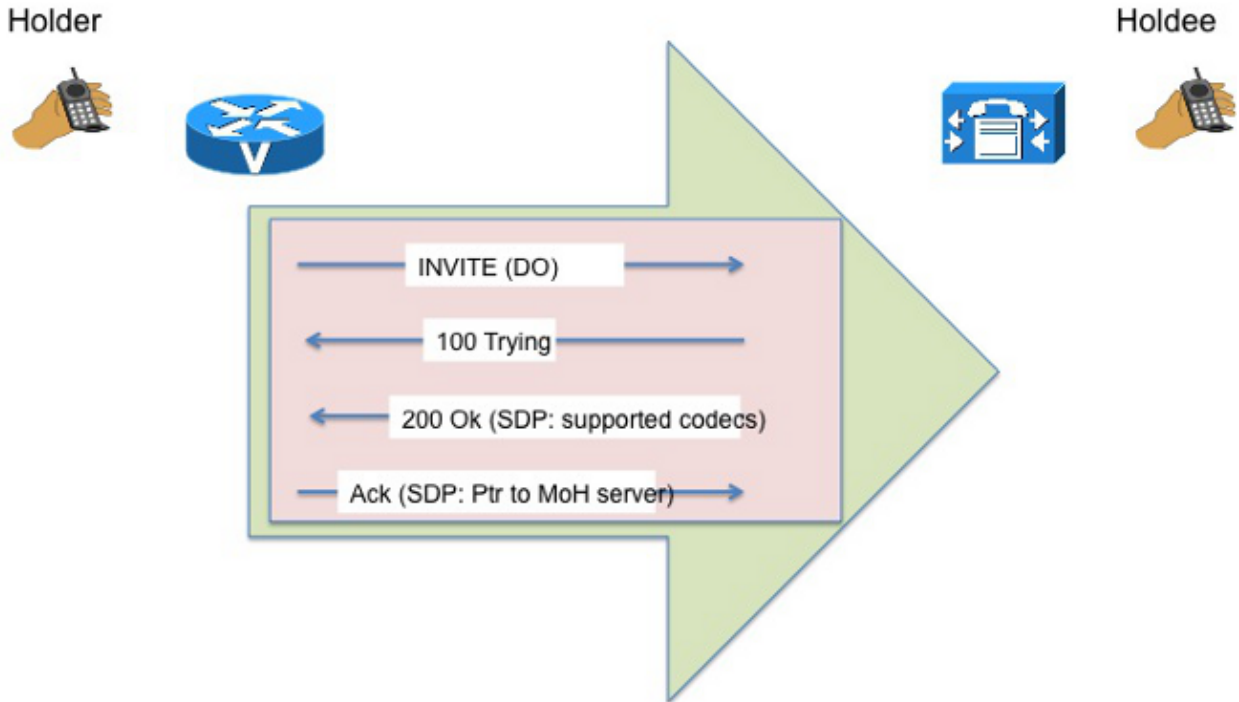
## Connect to MoH

The second step in the previously mentioned two-step process is to connect to MoH. Once the audio stream is disabled, CUCM signals the one-way MoH conversation that causes the *holdee* to listen to the MoH source.

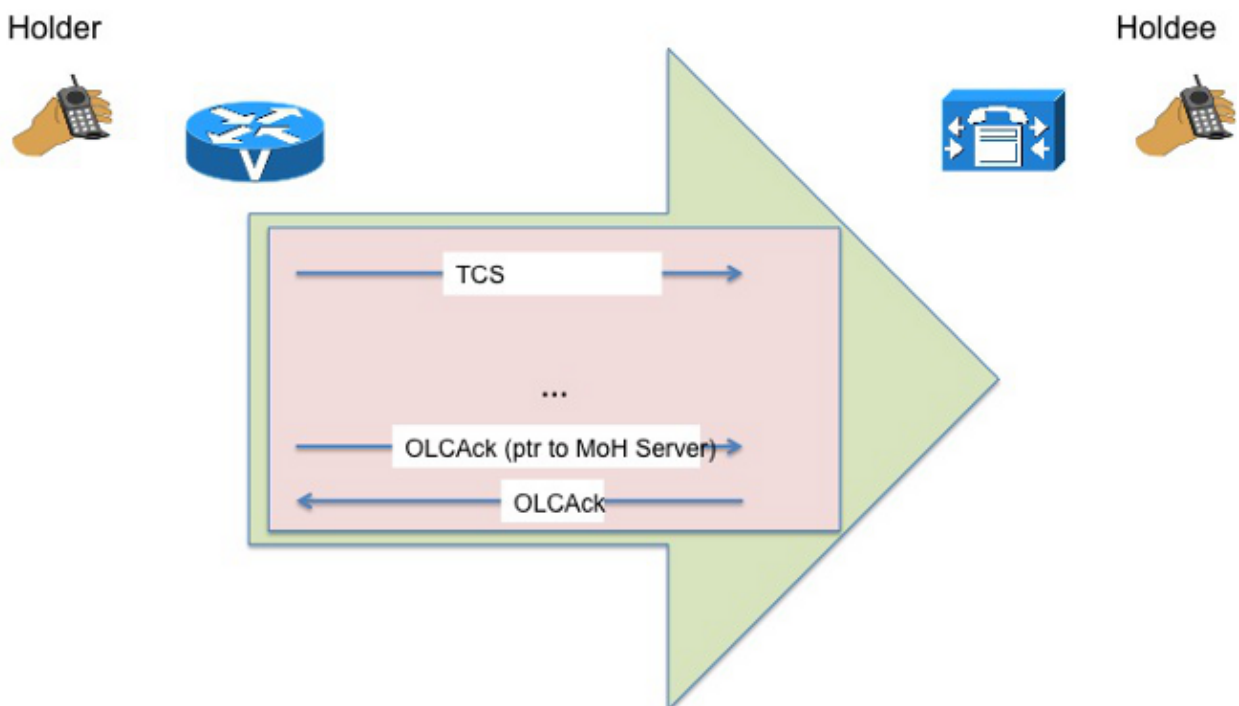
As part of this process, CUCM takes into account the media capabilities of the *holdee* and the Media Resource Group List (MRGL) associated with the trunk before it determines the parameters for streaming. Accordingly, signaling for this is always *Delayed Offer* (DO)[2] (in SIP).

The actual number of INVITE transactions vary. For example, CUCM connects the *holdee* to MoH with just one DO INVITE transaction. Alternatively, the DO INVITE is used in order to gather the media capabilities of the *holdee*, and a subsequent EO INVITE is used in order to actually connect the *holdee* to MoH.

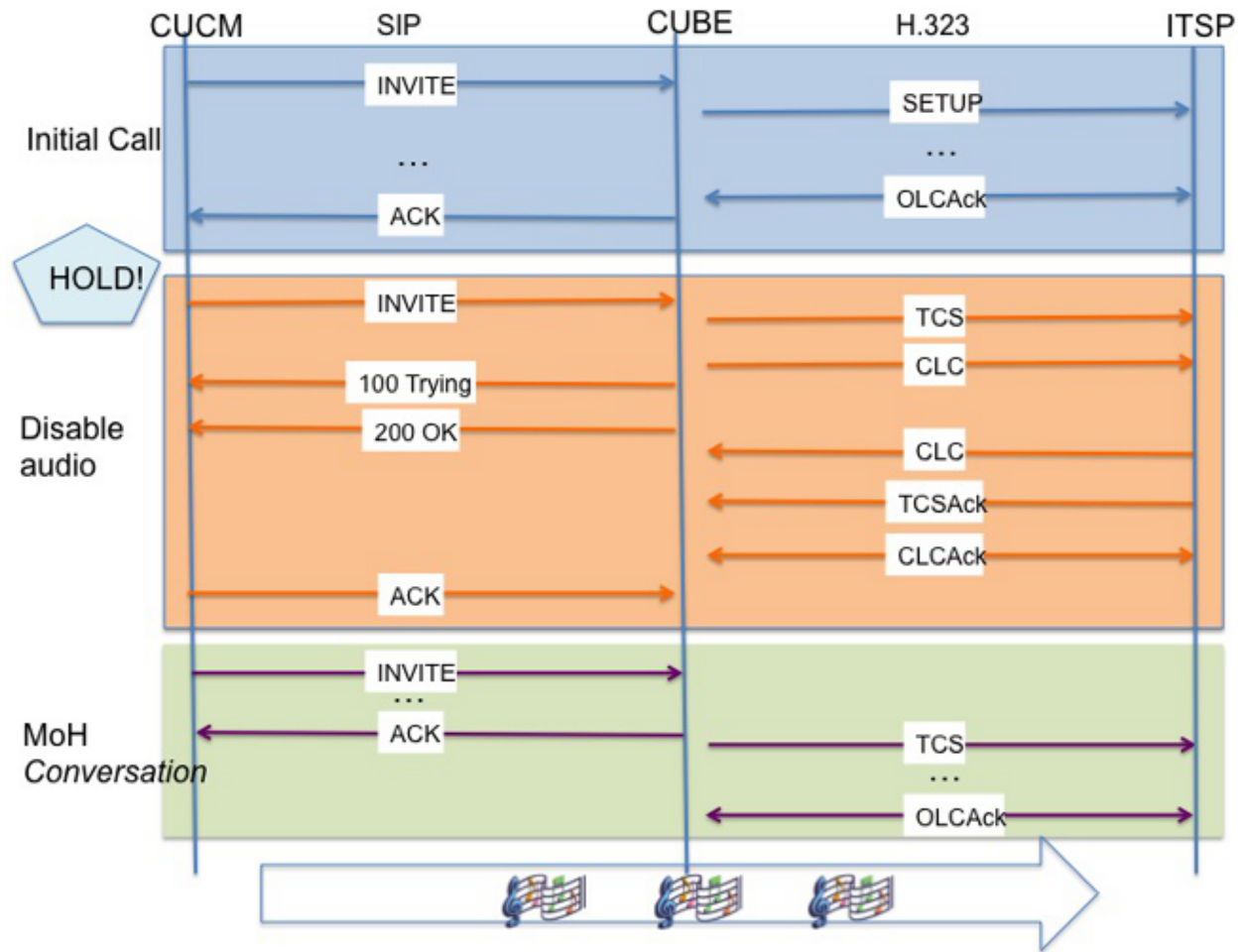
This image illustrates the transaction for SIP:



This image illustrates the transaction for H.323:



This image illustrates the signaling message sequence in an interworking environment (when one side of CUBE is SIP and the other side H.323, for example):



## When Media Resources are Used in a Call

Media Resources (Media Termination Point (MTP) / Transcoders) shield the CUBE-to-IT Service Provider (ITSP) call-leg for the most part. When a media resource is used in a call through CUBE, signaling for MoH mostly involves Skinny Client Control Protocol (SCCP) messages between CUCM and the Media Resource. Notice that it is the media resource that is put on hold, not the CUBE trunk. After the MTP/Transcoder is signaled to listen to MoH (assuming SIP), CUCM sends an SIP UPDATE message to CUBE. This updates the *branch* parameter, which identifies the new transaction (the MOH conversation).

## Resume the Call

The resume process is similar to the hold process, except that the order is reversed:

1. The current audio stream is disabled.
2. Another DO re-INVITE is sent in order to reconnect the *holdee* to the phone that placed the call on-hold.

## SDP Attribute

The *X-cisco-media:umoh* attribute in the Session Description Protocol (SDP) was introduced in order to simplify MoH signaling over Inter-Cluster Trunks (ICTs)[3]. With interoperation between endpoints that use different protocols, CUCM often performs awkward and intermediate signaling that is non-intuitive. In order

to avoid the guesswork, and make the signaling context–explicit, a proprietary SDP attribute, named *X-cisco-media*, is used.

With CUCM Versions 8.5 and later, MoH can [4] be signaled with this attribute set to either *Unicast Music on Hold (UMoH)* or *MMoH*, which removes the reliance on a fake port–value to indicate a MoH scenario to the held–party.

*Note:* This does not affect MoH signaling with CUBE.

## MoH on CUBE

With CUBE, the basic process remains the same; however, it is important to consider that [5] CUBE does not transcode MoH until Cisco IOS® Version 15.3T. This means that you must be careful with the factors that influence the codec selection in the CUCM–to–CUBE leg so that a transcoder is not needed.

*Note:* The transcoder referenced here is inserted by CUBE (as opposed to CUCM). As far as CUCM is concerned, the CUBE is the *destination*, and it does not involve any transcoder in the MOH Server–to–CUBE path.

## Codec Considerations

In general, several factors affect the codec used in the CUCM–to–CUBE leg, but these considerations apply for MoH:

- MoH cannot be transcoded.[5]
- MoH sounds good only with G.711.

*Note:* This topic is outside the scope of this document because many good documents already exist on codec considerations, and it would be redundant to cover it here.

## MMoH

*Note:* Most of the information described in this document thus far is relevant whether the MoH is streamed with unicast or multicast IP packets.

MMoH conserves system resources and bandwidth. Multicast allows multiple users to use the same audio source stream in order to provide music on–hold. MMoH is desirable in any corporate network where bandwidth savings are important.

Here are some concerns and issues when CUBE passes MMoH over the internet to ITSP:

- **Reach of multicast traffic** – Cisco uses the range 239.0.0.0 to 239.255.255.255 for multicast music. This range is known as *administratively scoped addresses*. This block is considered private, which means that it is used by enterprise networks, and should never be forwarded outside of the enterprise. The boundary routers are usually configured accordingly.
- **Multicast over VPN** – By default, IP security does not support MMoH.

This is how CUBE supports MMoH:

1. CUBE receives the MMoH packets from the MoH server.
2. It converts the packets into unicast IP packets.
3. CUBE forwards the packets to ITSP.



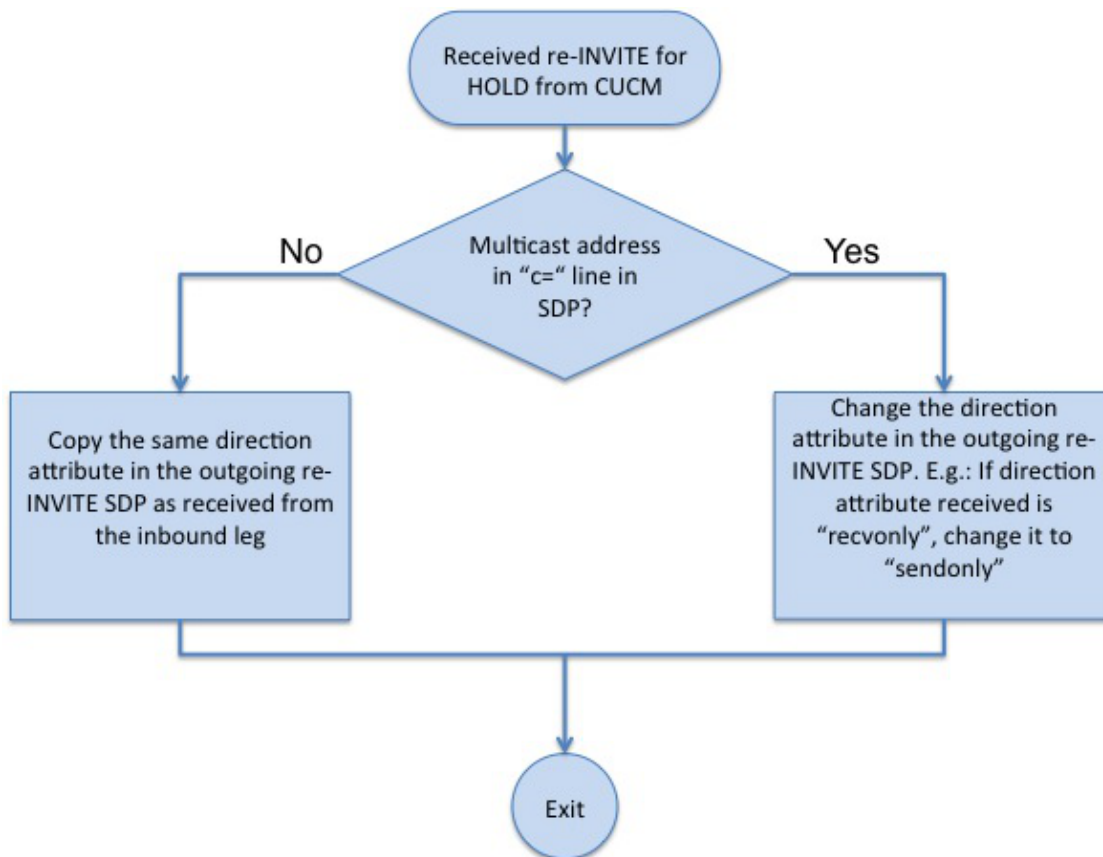
## SIP Direction Attribute Manipulation

As described in *RFC 3264*:

"If a session description contains a multicast media stream which is listed as receive (send) only, it means that the participants, including the offerer and answerer, can only receive (send) on that stream. ***This differs from the unicast view, where the directionality refers to the flow of media between offerer and answerer.*** Beyond that clarification, the semantics of an offered multicast stream are exactly as described in RFC 2327 [1]"

Accordingly, when CUCM sends a re-INVITE with a multicast IP address, it sets the direction attribute to ***recvonly***; however, since CUBE converts the multicast packets into unicast packets, it must set the direction attribute to ***sendonly*** on the leg with ITSP.

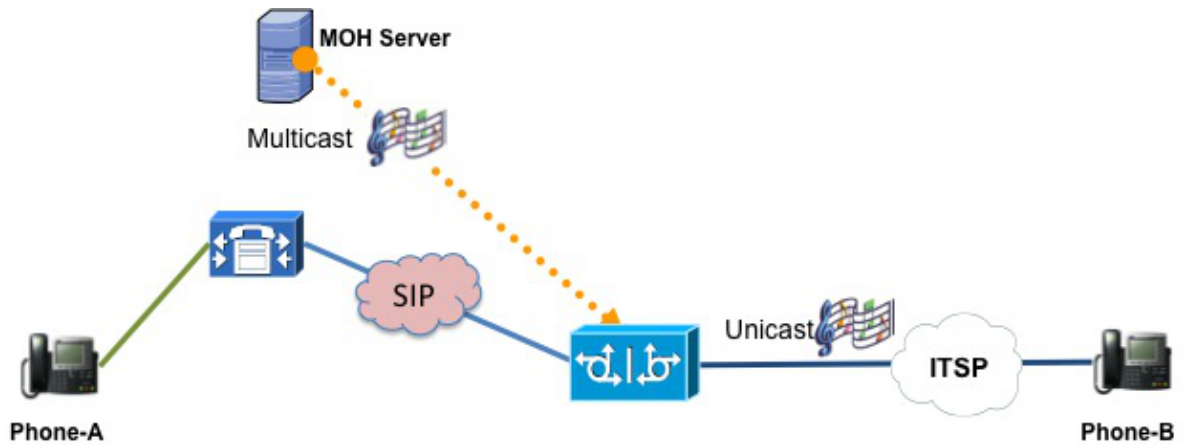
This image illustrates the logic:



## Address Manipulation

In the DO[6] re-INVITE sent in order to connect the ITSP caller to the MoH source, CUBE sends its own IP address in the SIP SDP *C=IN* field. This is a unicast address.

This image provides the end-to-end view:



1. Phone-A(Holder) puts Phone-B(Holdee) on Hold
2. CUCM tells CUBE to listen MOH at multicast IP address
3. CUBE receives the MMOH, converts to UMOH and passes on to Phone-B

*Note:* CUBE must run Cisco IOS Version 15.2(2)T or later in order to support MMoH.

## Stream from a Flash

With TDM gateways, additional WAN bandwidth savings are realized by streaming the multicast music right from the gateway. Thus, if the MoH server is at the headquarters, and the gateway is at a remote branch across a WAN connection, multicast traffic that carries MoH does not have to traverse the WAN (from headquarters to the branch) and use valuable WAN bandwidth.

CUBE is a trunk-side device that is not capable of streaming MMoH that is sourced from the local flash or via any analog TDM interface. It is still possible to realize WAN bandwidth. This is accomplished with use of another voice-enabled router at the remote branch as the source of the MMoH stream. This router streams MMoH from the flash. The CUBE can then be signaled in order to receive those packets, convert them, and pass them on as unicast packets.

## Stream from a Live-Feed

In order to stream from a live-feed, another router must be configured because CUBE is not a line-side device, as discussed in the previous section.

## Configure MMoH

This section describes how to configure MMoH on CUBE, CUCM, and L3-capable switches.

### *Configure MMoH on CUBE*

Use these commands in order to configure MMoH on CUBE:

```
ccm-manager music-on-hold
ip multicast-routing
```

### *Configure MMoH on CUCM*

Follow these steps in order to configure MMoH on CUCM:

1. Enable multicast capability on the MoH source, MoH server, and Media Resource Group (MRG).
2. Assign a MRGL to the trunk with the MRG configured in step 1.
3. Configure the codec in IP voice-streaming application service parameters.

**Note:** Refer to the Music on Hold section of the Cisco Unified Communications System 9.0 SRND – Media Resources article for detailed configuration steps.

### ***Configure MMoH on L3-Capable Switches***

Use these commands in order to configure MMoH on L3-capable switches:

```
ip routing
ip multicast-routing
```

## **When MTP is Used in a Call**

MTPs do not support multicast music. The *holdee* receives only dead-air.[7]

**Note:** Transcoders are MTPs also.

## **Performance Considerations**

All MMOH packets are *process switched* in Cisco IOS. This is fine for a small deployments, but has a significant impact on the performance of CUBE for large installations.

## **Restrictions**

Here is a list of restrictions with the use of MMoH:

- CUBE must be at Cisco IOS Version 15.2(2)T or later.
- MMoH is not supported on AS54xx.
- MMoH is not supported on ISR-G1s (28xx, 38xx series)
- Be aware of the codecs that are supported.

## **Troubleshoot**

Use this section in order to troubleshoot MMoH.

## **Show and Debug Commands**

Here is a list of show and debug commands, and their meanings:

- **Show *ccm-manager music*** – Helps confirm that CUBE knows where to listen for multicast music packets, and also whether it receives them. .

```
R1#show ccm-manager music
Current active multicast sessions : 1
```

Multicast	RTP port	Packets	Call	Codec	Incoming
Address	number	in/out	id		Interface

=====  
239.176.201.1 16384 956/956 237 g711ulaw Se0/1/0

- **Show ip igmp members** – Used in order to check if CUBE successfully joined the multicast group when signaled to listen to multicast music.
- These three commands are used in order to check the negotiated codec, IP address, and port numbers of the endpoints:

```
Show call active voice compact  
Show voip rtp conn  
Show sip calls
```

Here is an example output from the first command:

```
R1#show call active voice compact
```

```
<callID> A/O FAX T<sec> Codec type Peer Address IP R<ip>:<udp>  
  
Total call-legs: 2  
  
236 ANS T53 g711ulaw VOIP P1003 239.176.201.1:16384  
237 ORG T53 g711ulaw VOIP P919789362814 200.200.200.2:17808
```

- **Show call active voice brief**– Issue this command when the call is on–hold in order to check if the rx/tx counts increment.

```
0 : 236 29262010ms.1 (*22:34:23.659 UTC Fri May 10 2013)  
+4190 pid:1000 Answer 1003 connected  
dur 00:01:38 tx:919/147040 rx:918/146880 dscp:0 media:0 audio tos:0xB8 video tos:0x0  
IP 239.176.201.1:16384 SRTP: off rtt:0ms pl:0/0ms lost:0/0/0 delay:0/0/0ms  
g711ulaw TextRelay: off Transcoded: No  
media inactive detected:n media contrl rcvd:n/a timestamp:n/a  
long duration call detected:n long duration call duration:n/a timestamp:n/a
```

```
0 : 237 29262010ms.2 (*22:34:23.659 UTC Fri May 10 2013)  
+4190 pid:2000 Originate 919789362814 connected  
dur 00:01:38 tx:8910/1425600 rx:919/147040 dscp:0 media:0 audio tos:0xB8 video tos:0x0  
IP 200.200.200.2:17808 SRTP: off rtt:0ms pl:0/0ms lost:0/0/0 delay:0/0/0ms  
g711ulaw TextRelay: off Transcoded: No  
media inactive detected:n media contrl rcvd:n/a timestamp:n/a  
long duration call detected:n long duration call duration:n/a timestamp:n/a
```

- **Show perf query class "Cisco MOH Device"**– This CUCM CLI command is used in order to quickly check if an MoH resource is allocated, and what kind (unicast or multicast). This command is not very useful when you have multiple calls no–hold, as the counts change dynamically when calls are put on–hold and resumed.

```
admin:show perf query class "Cisco MOH Device"
```

```
==>query class :
```

```
- Perf class (Cisco MOH Device) has instances and values:
```

```
MOH_2 -> MOHHighestActiveResources = 0  
MOH_2 -> MOHMulticastResourceActive = 0  
MOH_2 -> MOHMulticastResourceAvailable = 250000  
MOH_2 -> MOHOutOfResources = 1  
MOH_2 -> MOHTotalMulticastResources = 250000
```

```

MOH_2          -> MOHTotalUnicastResources      = 250
MOH_2          -> MOHUnicastResourceActive           = 0
MOH_2          -> MOHUnicastResourceAvailable      = 250

```

- **Debug ccm-manager music-on-hold** – This command is used in order to trace how the call-legs are changed (when you disable the current audio and connect MoH, for example), as well as verify whether CUBE joins the Internet Group Management Protocol (IGMP) group as instructed by CUCM.
- **Debug ip packet** – This command is used as an alternative to Wireshark for checks. However, this command can quickly overwhelm the CPU. Use it only when absolutely necessary; turn off console logging, and do not run it for more than a second.

## Scenario 1

**Symptom** – A call from the Public Switched Telephone Network (PSTN) establishes fine with two-way audio. However, when the IP phone places the PSTN caller on-hold and then resumes the call, one-way audio results: the IP phone hears the audio from PSTN, but the PSTN user cannot hear the IP phone.

First, make sure that **Require SDP Inactive Exchange for Mid-Call Media Change** is NOT disabled on the SIP trunk in question[5]. This is what enables CUCM to send a re-INVITE with **a=inactive** in SDP, in order to break the media path that exists.

When the call is put on-hold, CUCM does not send a re-INVITE with an inactive SDP in order to break the media path if the **Send send-receive SDP in mid-call INVITE** check box is enabled for the SIP trunk[8]. That configuration is checked only for devices that cannot provide a full (send-recv) offer after the media mode is set to inactive.

Here are images that illustrate the available check boxes:

**SIP Profile Information**

Name\*

Description

Default MTP Telephony Event Payload Type\*

Early Offer for G.Clear Calls\*

SDP Session-level Bandwidth Modifier for Early Offer and Re-invites\*

User-Agent and Server header information\*

Accept Audio Codec Preferences in Received Offer\*

Dial String Interpretation\*

Redirect by Application

Disable Early Media on 180

Outgoing T.38 INVITE include audio mline

Enable ANAT

Require SDP Inactive Exchange for Mid-Call Media Change

Use Fully Qualified Domain Name in SIP Requests

Assured Services SIP conformance



*Note:* Refer to Cisco bug ID CSCtx84013 for additional information.

## Scenario 2

*Symptom* – There is only a tone when callers are placed on–hold instead of MMoH.

Generally, this suggests that CUCM did not allocate MMoH.

- Use the *show perf query class ?Cisco MOH Device?* CUCM CLI command in order to verify if the *MOHOutOfResources* count increments.
- Ensure that multicast is enabled on the MMoH source, server, and group.

## Scenario 3

*Symptom* – Only dead–air is heard when a caller is placed on–hold.

Ensure that:

- Multicast–routing is enabled on the CUBE and other routers in the audio path.
- IP routing and multicast–routing are enabled on the L3 switches in the audio path.
- The *ttl* (hop count) is configured on the MoH server on CUCM, and is large enough to cover the hops.
- If a transcoder is required, it is allocated successfully.
- The list of codecs configured on the IP Voice Streaming Application supports the codec used for MoH.

## Scenario 4

*Symptom* – A call fails in the flow–around mode for *Call hold & Resume*.

In order to support flow–around, you must send a re–INVITE or an update from IPIPGW; however, this is currently not supported. Hence, flow–around with DO–EO calls is not supported. If there is such a call–flow requirement from marketing, a consideration for support will be made. The Cisco bug, *SIP SIP SS DO–EO: Call fails in Flow around mode for Call hold & Resume*, is marked as an enhancement for consideration in the future.

## Related Information

- *Multicast Music–on–Hold Support on Cisco UBE*
- *CallManager Music on Hold Frequently Asked Questions*
- *Technical Support & Documentation – Cisco Systems*

---

[1] This can be confusing– How can you have a different conversation within a dialogue? Well, in SIP, dialogue refers to the 3–tuple <*To tag, From tag, and Call-ID*>. This 3–tuple remains the same during the hold phase.

[2] DO – Delayed Offer.

[3] Inter–cluster trunk

[4] Starting from CUCM 8.5.

[5] Transcoding works for MMoH in Cisco IOS Versions 15.3T and later.

[6] DO – Delayed Offer

[7] Cisco Unified Communications Manager Features and Services Guide, Release 8.6(1)

[8] These are settings on the SIP profile used in order to configure the SIP trunk.

---

Updated: Sep 12, 2013

Document ID: 116392

---