



Dynamic Optimization of Resources

Technical Note

Cisco TelePresence Server 3.1

D15085.01

June 2014



Contents

Introduction	4
Architecture	4
New service levels	6
Locally managed mode service levels	6
Remotely managed mode service levels	6
Benefits	7
Simultaneous service levels	8
Locally managed mode simultaneous service levels	8
Remotely managed mode simultaneous service levels.....	9
Benefits	10
Optimizing resources	11
Locally managed mode resource usage	11
Remotely managed mode optimized resources	11
Benefits	13
Summary	14

Tables

Table 1: TelePresence Server 8710 3.1 conferencing capacity in local managed mode.....	6
Table 2: TelePresence Server 8710 3.1 conferencing capacity in remotely managed mode	6

Figures

Figure 1: Locally managed mode architecture	5
Figure 2: Remotely managed mode architecture	5
Figure 3: Locally managed mode simultaneous service levels.....	9
Figure 4: Remotely managed mode simultaneous service levels.....	10
Figure 5: Locally managed mode resource usage in Full HD mode	11
Figure 6: Remotely managed mode resource optimized resources	13

Introduction

In releases prior to 3.0 TelePresence Server supported fixed port counts with set maximum resolutions. Any call to the TelePresence Server, even if it were made by an endpoint that only supported a lower resolution and required fewer resources, would use up a screen license or equivalent portion of one. For instance with a TelePresence Server in Full HD mode an endpoint supporting a maximum of 720p and a second endpoint supporting a maximum of 1080p would both use the same number of licenses when they connected, one screen license each.

Note: While a single screen system uses a single screen license in Full HD mode, a multi-screen system uses multiple screen licenses. For example a TX 9000 series endpoint would use three screen licenses in Full HD mode.

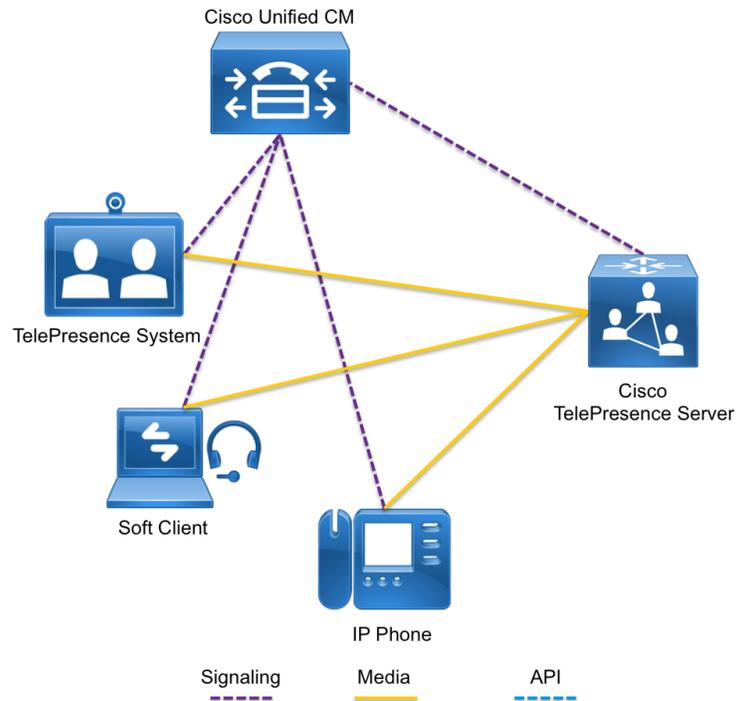
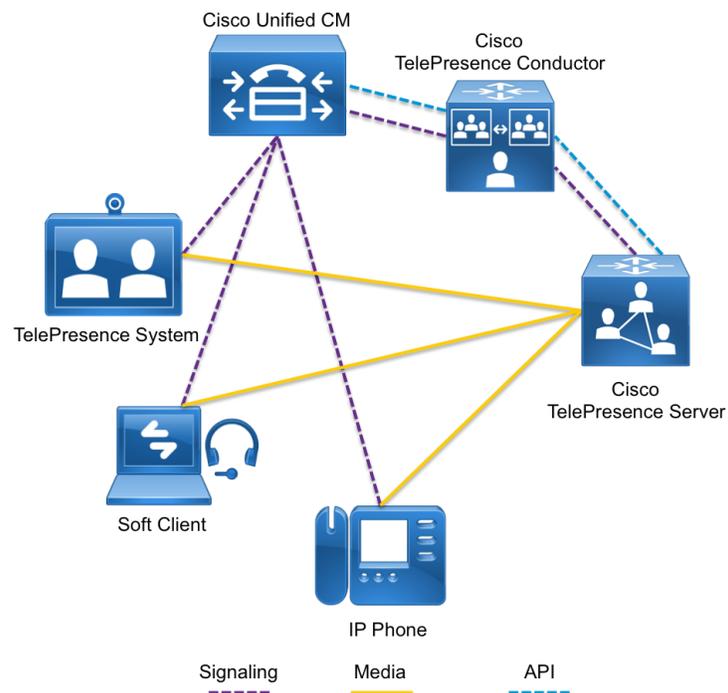
In release 3.0 and later the TelePresence Server's operation mode can be configured to be remotely managed via the new API, allowing its resources to be optimized dynamically by a remote management system. As a result there are several core benefits realized:

- New service levels with a range of main video qualities, content qualities and combinations thereof are available
- Multiple conferences of different maximum resolutions can run on the same unit, allowing an administrator greater control over the number and size of conferences per server
- Endpoints with maximum capabilities lower than the conference cap can be optimized to use less resource

This document will provide an overview of the technical features that provide these benefits.

Architecture

This document considers two architectures, a TelePresence Server running 3.1 in locally managed mode and a TelePresence Server 3.1 in remotely managed mode with TelePresence Conductor. Locally managed mode provides a fixed screen count while remotely managed mode can provide dynamic optimization of resources. For the purposes of this document only non-scheduled call scenarios are considered.

Figure 1: Locally managed mode architecture**Figure 2: Remotely managed mode architecture**

Note: While the architectures above show Cisco Unified CM, the concepts remain the same if using Cisco TelePresence Video Communications Server only or a combination of Cisco Unified CM and Cisco VCS.

New service levels

In locally managed mode for TelePresence Server versions prior to 3.0 there are only two main modes with one content quality for each mode. However many endpoints support lower main video resolutions while others support higher content qualities.

Locally managed mode service levels

In locally managed mode TelePresence server has two service levels, these are Full HD (1080p) mode with 720p15 content and HD (720p) mode with 720p5 content. Participants are free to connect at resolutions below these maximums at each service level but they will always use one whole resource (or multiples of this resource if they are using a multiscreen system), equivalent to the maximum possible quality.

Table 1: TelePresence Server 8710 3.1 conferencing capacity in local managed mode

Main video	Audio	Content	Maximum calls
720p30	Stereo	720p5	24
1080p30	Stereo	720p15	12

Remotely managed mode service levels

In remotely managed mode a TelePresence Server running software version 3.1 or above has four main video quality levels, Full HD (1080p), HD (720p), SD (480p) and 360p. It also has four content quality levels, 1080p30, 720p30, 720p15, 720p5 and off. These main video and content quality levels can be mixed as desired, so for instance 1080p main video with 720p30 content is a valid service level as is 1080p main video with 720p15 content.

Table 2: TelePresence Server 8710 3.1 conferencing capacity in remotely managed mode

Main video	Audio	Content	Maximum calls
None	Mono	None	104
360p30	Mono	In main video	97
480p30	Mono	In main video	48
480p30	Mono	720p5	36
720p30	Stereo	720p5	24
720p30	Stereo	720p30	12
720p60	Stereo	720p15	12
1080p30	Stereo	720p15	12

Main video	Audio	Content	Maximum calls
1080p30	Stereo	720p30	8
1080p30	Stereo	1080p30	6

Note: The above table considers single screen systems only, while the TelePresence Server supports multiscreen systems too. When multiscreen systems are used more resource is consumed. When not using TIP each screen of an endpoint counts as a “call” while when using TIP a single call is used, this is important to consider as far as the maximum supported call count of TelePresence Server is concerned.

Benefits

For the first time main video quality modes below 720p are available to participants allowing administrators to provide lower cost and higher scale services. In addition new content quality modes are available that allow higher quality than has been available previously along with the flexibility to match any main video quality with any content quality.

Simultaneous service levels

It is common for a deployment to require multiple service levels for users. The core reasons for this requirement tend to be cost and scale. When high quality 1080p ports are provided a higher percentage of the conferencing resources are consumed per call and so the cost is higher than a service providing a lower maximum quality.

While it would be desirable to offer the highest quality video to all users there is often a compromise made as to what quality levels a customer will run. For instance, it may be that certain users require Full HD (1080p) video while the majority of users will use an HD (720p) service and occasional users are allocated a maximum of SD (w448p/480p) quality.

Locally managed mode simultaneous service levels

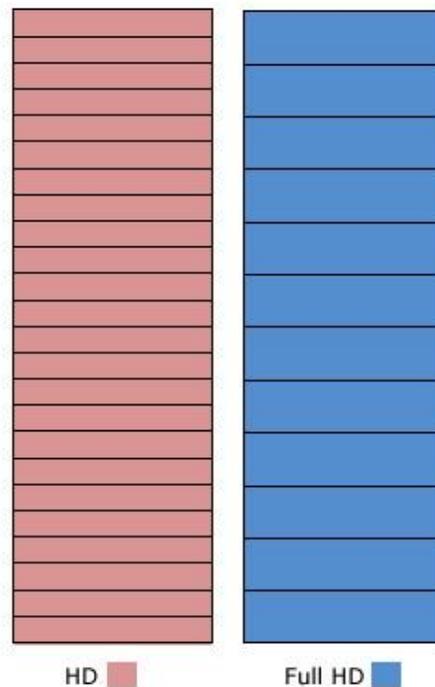
In locally managed the TelePresence Server has just two service levels available, only one of which may be used at a time per unit or per unit cluster. These modes are Full HD (1080p) and HD (720p). In order to provide a video service that supports both 1080p and 720p one of two choices can be made.

The first option is to run all TelePresence Servers in Full HD mode. In Full HD mode there are no restrictions to join the unit at a resolution lower than 1080p but each connection will use a Full HD resource regardless of the resolution used. This option may not be very attractive as there is no capability to limit users to a lower resolution than 1080p and therefore no ability to save conference resource or reduce cost.

If a deployment has multiple TelePresence Servers or TelePresence Server clusters the second option is to run TelePresence Servers in different modes. For instance deploy one cluster in Full HD mode and another cluster in HD mode. This provides two levels of service and brings the cost reduction that the service required. There are trade-offs to this approach as well:

- Requires multiple units or multiple clusters of units and so is unsuitable for smaller deployments.
- Requires careful planning to ensure that the right number of devices are purchased to satisfy both services
- As cascading is not currently supported it is not possible to mix service levels
- It can lead to excessive resource wastage. Since multiple units are used there could be overutilization on one unit running in HD but underutilization on another unit running in Full HD – excess can never be used for a different service level.

The figure below shows two fully licensed TelePresence Servers, one running in 24 port 720p HD mode and the other running in 12 port 1080p Full HD mode.

Figure 3: Locally managed mode simultaneous service levels

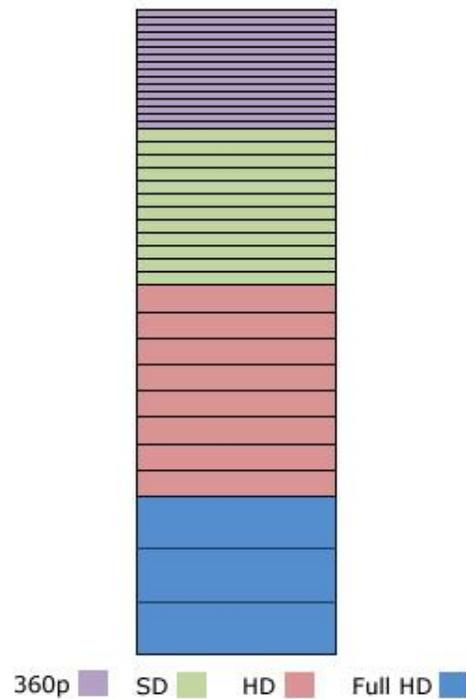
Remotely managed mode simultaneous service levels

As of software release 3.1, in remotely managed mode the TelePresence Server has four main video quality levels available, all of which can be used simultaneously on the same unit or unit cluster. These modes are Full HD (1080p), HD (720p), SD (480p) and 360p. In order to provide a video service that supports both 1080p and 720p a simplified approach is available.

A single TelePresence Server can now be used to provide both service levels. TelePresence Conductor is used to define a maximum main video resolution of 1080p for one template and 720p for a second template. Any number of aliases can now be linked to one of these templates.

When an alias linked to the 1080p template is called, TelePresence Conductor starts a conference on the TelePresence Server that allows participants to connect at up to 1080p. When an alias linked to the 720p template is called, TelePresence Conductor starts a conference on the TelePresence Server that allows participants to connect at up to 720p. Participants using the 720p aliases will use half the resource of the participants using the 1080p aliases. This can be extended by offering SD and 360p aliases too.

The figure below provides an example depiction of running multiple services on a single TelePresence Server simultaneously. In this example four conferences are running, each with a different maximum quality – 360p, SD, HD and Full HD.

Figure 4: Remotely managed mode simultaneous service levels

Benefits

The ability to run multiple conference service levels on the same TelePresence Server brings greater efficiency and scale to customers:

- Provide different quality levels to satisfy conference owners
- As few as one physical device needed to provide all quality levels
- More efficient resource usage
- The ability to trade off lower quality conference limits for higher participant counts
- Using multiple service levels on one device instantly allows more concurrent connections than possible in locally managed mode

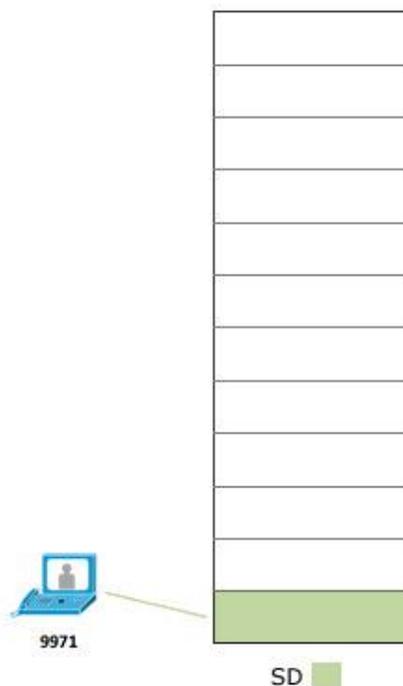
Optimizing resources

Setting a maximum quality level for a conference ensures predictable and efficient usage of resources. In some cases endpoints will join a conference that has no ability to reach the maximum quality allowed in the conference and with these cases it is possible to gain further efficiencies by using optimized resources.

Locally managed mode resource usage

When the TelePresence Server is locally managed in Full HD mode, each connection uses one whole screen license worth of resources regardless of the capability of the endpoint. For example when in Full HD mode even if an endpoint such as a 9971 phone joins that can send and receive a maximum of 480p video and no content, one 1080p video resource along with its 720p15 content resource will be used.

Figure 5: Locally managed mode resource usage in Full HD mode



Remotely managed mode optimized resources

Conference templates created by TelePresence Conductor have a setting called **Optimize Resources**. When enabled this allows resources that were initially allocated on the TelePresence Server and that the endpoints do not require, to be freed up if five seconds pass and no new participants have joined the conference.

In this case, endpoint resolution support is classed as the maximum resolution an endpoint advertises in its capability negotiation. It should be noted that it is common for endpoints to advertise their maximum supported resolution regardless of the maximum resolution they can achieve in the current call based on the bandwidth and settings used, i.e. An endpoint that is 1080p capable that joins a 1080p conference will always advertise 1080p capability even at low call rates such as 128kbps and so will continue to use a full 1080p resource in TelePresence Server after optimization.

Resources are optimized and freed up on the TelePresence Server in the following situations:

- If the maximum capability an endpoint advertises is lower than allowed by the conference template (with the consideration noted above).
- If an endpoint uses fewer screens than allowed by TelePresence Conductor conference template with the following considerations.
 - In deployments using the Cisco VCS's external policy service, Conductor does not know how many screens a participant will use until the call has been received by the bridge. (In B2BUA deployments TelePresence Conductor is able to detect the number of screens required from the SIP signalling and allocates resources for the correct number of screens up front, so no late optimization is required in the B2BUA case.)
 - For ad hoc calls TelePresence Conductor does not know how many screens a participant will use until the TelePresence Server has received the call. (Cisco Unified CM reserves the conference resources with TelePresence Conductor before the call is seen at the B2BUA, and before the call reaches the TelePresence Server).
 - TelePresence Conductor considers screen allocation for reserved Chairpersons separately to resources that can be used for Guests. If you configure the template to support 3 screens and a Chairperson dials in which is optimized down to a single screen resource, Conductor will only allow the reclaimed screens to be used with further Chairpersons, they cannot be used for Guests in the conference.

Resources are not optimized for auto-dialed participants or for pre-configured endpoints because when configuring these entities the desired quality is defined in their configuration. The configuration overrides the capabilities defined in the template which is used for non-pre-configured endpoint incoming calls.

Note: Conductor only checks pre-configured endpoints if **Allow multiscreen** is set to Yes in the template.

In order to allow a call to be placed on the TelePresence Server, enough resource must be available to cater for the conference quality setting e.g. if a conference is set to 1080p quality TelePresence Conductor will only allow the call to continue if there is enough resource on the TelePresence Server to provide the full conference quality as it assumes each call being made will use 100% of the assigned resource. If the endpoint is only capable of a lower resolution (note this is based on the endpoint's maximum supported resolution not the current call

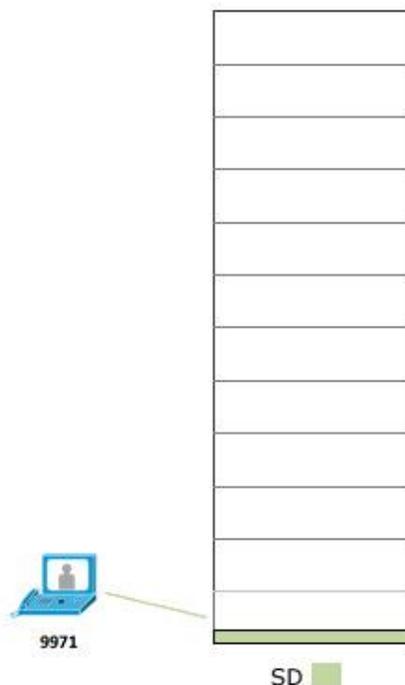
resolution) resource optimization will occur 5 seconds after the last join event. Where many connections happen very quickly this may cause later calls to be rejected until optimization has occurred and TelePresence Conductor is aware of the free resource.

For the best results Cisco recommends that the conference quality is limited where possible. In the case where conferences are left with excessively high quality settings, calls may be rejected unnecessarily and resources may be reserved unnecessarily.

Note: When limiting audio quality it is important to understand the impact of doing so. For instance limiting audio to mono will prevent TC based endpoints receiving directional audio from the TelePresence Server. In addition multichannel audio is a requirement in order to negotiate TIP for multiscreen endpoints and stereo is required for single screen TIP endpoints.

In a similar use case to locally managed mode if an endpoint such as a 9971 phone joins that can send and receive a maximum of 480p video and no content, one 1080p video resource along with its 720p15 content resource will be initially allocated but 5 seconds after the last person joins the connection will be optimized down to use only an SD main video resource with no content.

Figure 6: Remotely managed mode resource optimized resources



Benefits

Optimizing resources allows even greater efficiencies over limiting conferences to a maximum resolution by allowing resource that the endpoint does not advertise it can use to be reclaimed and used by other participants.

Summary

Dynamic optimization of resources is just one element of Cisco's Pervasive Conferencing architecture with Cisco TelePresence Server and Cisco TelePresence Conductor. Dynamic optimization of resources offers a number of features and benefits including:

- New main video service levels
- New higher content video qualities
- New combinations of main video and content service levels
- Per conference selection of maximum video and content resolutions
- As few as one physical device needed to provide all quality levels
- Resource optimization where endpoints with maximum capabilities lower than the conference cap will use less resource

These allow greater quality, scale and efficiency than previously available; reducing costs while enhancing user experience.

THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS IN THIS MANUAL ARE SUBJECT TO CHANGE WITHOUT NOTICE. ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS MANUAL ARE BELIEVED TO BE ACCURATE BUT ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. USERS MUST TAKE FULL RESPONSIBILITY FOR THEIR APPLICATION OF ANY PRODUCTS.

THE SOFTWARE LICENSE AND LIMITED WARRANTY FOR THE ACCOMPANYING PRODUCT ARE SET FORTH IN THE INFORMATION PACKET THAT SHIPPED WITH THE PRODUCT AND ARE INCORPORATED HEREIN BY THIS REFERENCE. IF YOU ARE UNABLE TO LOCATE THE SOFTWARE LICENSE OR LIMITED WARRANTY, CONTACT YOUR CISCO REPRESENTATIVE FOR A COPY.

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED "AS IS" WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Cisco and the Cisco Logo are trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1005R)

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

© 2014 Cisco Systems, Inc. All rights reserved.