Avaya Aura® Solution Deployment

Release 6.1
03-603978
Issue 1
June 2011

Notice

While reasonable efforts have been made to ensure that the information in this document is complete and accurate at the time of printing, Avaya assumes no liability for any errors. Avaya reserves the right to make changes and corrections to the information in this document without the obligation to notify any person or organization of such changes.

Documentation disclaimer

“Documentation” means information published by Avaya in varying mediums which may include product information, operating instructions and performance specifications that Avaya generally makes available to users of its products. Documentation does not include marketing materials. Avaya shall not be responsible for any modifications, additions, or deletions to the original published version of documentation unless such modifications, additions, or deletions were performed by Avaya. End User agrees to indemnify and hold harmless Avaya, Avaya's agents, servants and employees against all claims, lawsuits, demands and judgments arising out of, or in connection with, subsequent modifications, additions or deletions to this documentation, to the extent made by End User.

Link disclaimer

Avaya is not responsible for the contents or reliability of any linked Web sites referenced within this site or documentation provided by Avaya. Avaya is not responsible for the accuracy of any information, statement or content provided on these sites and does not necessarily endorse the products, services, or information described or offered within them. Avaya does not guarantee that these links will work all the time and has no control over the availability of the linked pages.

Warranty

Avaya provides a limited warranty on its Hardware and Software (“Product(s)”). Refer to your sales agreement to establish the terms of the limited warranty. In addition, Avaya’s standard warranty language, as well as information regarding support for this Product while under warranty is available to Avaya customers and other parties through the Avaya Support Web site: http://support.avaya.com. Please note that if you acquired the Product(s) from an authorized Avaya reseller outside of the United States and Canada, the warranty is provided to you by said Avaya reseller and not by Avaya.

Licenses

THE SOFTWARE LICENSE TERMS AVAILABLE ON THE AVAYA WEBSITE, HTTP://SUPPORT.AVAYA.COM/LICENSEINFO/ ARE APPLICABLE TO ANYONE WHO DOWNLOADS, USES AND/OR INSTALLS AVAYA SOFTWARE, PURCHASED FROM AVAYA INC., ANY AVAYA AFFILIATE, OR AN AUTHORIZED AVAYA RESELLER (AS APPLICABLE) UNDER A COMMERCIAL AGREEMENT WITH AVAYA OR AN AUTHORIZED AVAYA RESELLER. UNLESS OTHERWISE AGREED TO BY AVAYA IN WRITING, AVAYA DOES NOT EXTEND THIS LICENSE IF THE SOFTWARE WAS OBTAINED FROM ANYONE OTHER THAN AVAYA, AN AVAYA AFFILIATE OR AN AVAYA AUTHORIZED RESELLER; AVAYA RESERVES THE RIGHT TO TAKE LEGAL ACTION AGAINST YOU AND ANYONE ELSE USING OR SELLING THE SOFTWARE WITHOUT A LICENSE. BY INSTALLING, DOWNLOADING OR USING THE SOFTWARE, OR AUTHORIZING OTHERS TO DO SO, YOU, ON BEHALF OF YOURSELF AND THE ENTITY FOR WHOM YOU ARE INSTALLING, DOWNLOADING OR USING THE SOFTWARE (HEREINAFTER REFERRED TO INTERCHANGEABLY AS “YOU” AND “END USER”), AGREE TO THESE TERMS AND CONDITIONS AND CREATE A BINDING CONTRACT BETWEEN YOU AND AVAYA INC. OR THE APPLICABLE AVAYA AFFILIATE (“AVAYA”).

Avaya grants End User a license within the scope of the license types described below. The applicable number of licenses and units of capacity for which the license is granted will be one (1), unless a different number of licenses or units of capacity is specified in the Documentation or other materials available to End User. “Designated Processor” means a single stand-alone computing device. “Server” means a Designated Processor that hosts a software application to be accessed by multiple users. “Software” means the computer programs in object code, originally licensed by Avaya and ultimately utilized by End User, whether as stand-alone Products or pre-installed on Hardware. “Hardware” means the standard hardware originally sold by Avaya and ultimately utilized by End User.

Concurrent User License

Concurrent User License (CU). End User may install and use the Software on multiple Designated Processors or one or more Servers, so long as only the licensed number of Units are accessing and using the Software at any given time. A “Unit” means the unit on which Avaya, at its sole discretion, bases the pricing of its licenses and can be, without limitation, an agent, port or user, an e-mail or voice mail account in the name of a person or corporate function (e.g., webmaster or helpdesk), or a directory entry in the administrative database utilized by the Software that permits one user to interface with the Software. Units may be linked to a specific, identified Server.

Copyright

Except where expressly stated otherwise, no use should be made of materials on this site, the Documentation, Software, or Hardware provided by Avaya. All content on this site, the documentation and the Product provided by Avaya including the selection, arrangement and design of the content is owned either by Avaya or its licensors and is protected by copyright and other intellectual property laws including the sui generis rights relating to the protection of databases. You may not modify, copy, reproduce, republish, upload, post, transmit or distribute in any way any content, in whole or in part, including any code and software unless expressly authorized by Avaya. Unauthorized reproduction, transmission, dissemination, storage, and or use without the express written consent of Avaya can be a criminal, as well as a civil offense under the applicable law.

Third-party components

Certain software programs or portions thereof included in the Product may contain software distributed under third party agreements (“Third Party Components”), which may contain terms that expand or limit rights to use certain portions of the Product (“Third Party Terms”). Information regarding distributed Linux OS source code (for those Products that have distributed the Linux OS source code), and identifying the copyright holders of the Third Party Components and the Third Party Terms that apply to them is available on the Avaya Support Web site: http://support.avaya.com/Copyright.

Preventing Toll Fraud

“Toll fraud” is the unauthorized use of your telecommunications system by an unauthorized party (for example, a person who is not a corporate employee, agent, subcontractor, or is not working on your company's behalf). Be aware that there can be a risk of Toll Fraud associated with your system and that, if Toll Fraud occurs, it can result in substantial additional charges for your telecommunications services.

Avaya Toll Fraud Intervention

If you suspect that you are being victimized by Toll Fraud and you need technical assistance or support, call Technical Service Center Toll Fraud Intervention Hotline at +1-800-643-2353 for the United States and Canada. For additional support telephone numbers, see the Avaya Support Web site: http://support.avaya.com. Suspected security vulnerabilities with Avaya products should be reported to Avaya by sending mail to: securityalerts@avaya.com.

Trademarks

Avaya® , Avaya Aura®, Avaya™, and Avaya Aura ™ are registered trademarks of Avaya Inc. in the United States of America and/or other jurisdictions.

Downloading Documentation

For the most current versions of Documentation, see the Avaya Support Web site: http://support.avaya.com.

Hardware Support

For full hardware support, please see *Avaya Support Notices for Hardware Documentation*, document number 03–600759 on the Avaya Support Web site, http://support.avaya.com.

Contact Avaya Support

Avaya provides a telephone number for you to use to report problems or to ask questions about your Product. The support telephone number is 1-800-242-2121 in the United States. For additional support telephone numbers, see the Avaya Web site: http://support.avaya.com.

Contents

Chapter 1: Introduction

Avaya Aura® overview

Avaya Aura® is the flagship communications solution for next generation, people-centric collaboration. It is based on Session Initiation Protocol (SIP) architecture that unifies media, modes, networks, devices, applications and real-time, actionable presence across a common infrastructure. This creates an environment where users have on-demand access to advanced collaboration services and applications that deliver enhanced customer access, improved employee efficiency, and lower total cost of ownership. It is Avaya's core communications platform and is available in two, pre-packaged software editions, each one based on simple, per user licensing that supports mid- to large enterprises.

The Avaya Aura® solution includes the following key capabilities:

- Total scale of up to 10 Session Managers and 100,000 SIP users supported in an enterprise
- Centralized Avaya Aura® core-supported converged SIP voice and video call admission control
- SIP features that include desk-level SIP E911 location reporting
- CS 1000 SIP networking and feature transparency
- Session Manager SIP routing adaptations
- A single, central management application (System Manager) for many Avaya Aura® applications and Avaya Communication Server 1000 with single authentication and single navigation

Avaya Aura® redefines and simplifies existing voice and video communications architectures, including multivendor network by introducing a platform that unifies all forms of communication (voice, messaging, e-mail, voice mail, and more), without sacrificing any of the resiliency, security and performance for which Avaya communications systems have always been known.

Avaya Aura® core components

Avaya Aura® includes five core applications:

- Avaya Aura® Session Manager, a SIP-based Session Manager capability
- Avaya Aura® Communication Manager, the highly-reliable and extensible communication software
- Avaya Aura® Application Enablement Services, an integration capability to extend communications to business applications
- Avaya Aura® Presence Services, a presence aggregation solution
- Avaya Aura® System Manager, a common management framework that optimizes centralized management functions for provisioning, operations and fault and performance monitoring.

This combination of applications makes it possible to unify media, modes, networks, devices, applications and real-time, actionable presence across a common infrastructure, creating the web-style, on-demand access to services and applications that users increasingly expect from their enterprise communications solution.

Avaya Aura® Session Manager overview

Companies typically have a diverse set of communications products within their corporate intranet that cannot communicate with each other. A standard signaling protocol is required to make these products work together. Avaya adopted the Session Initiation Protocol (SIP) as the signaling protocol for communication.

Avaya Aura® Session Manager is a SIP routing and integration tool and the core component within the Avaya Aura® Enterprise Edition solution. It integrates all the SIP devices across the entire enterprise network within a company. Session Manager offers a new perspective on enterprise communication where individual locations are no longer managed as separate units within the enterprise. Each location, branch, or application is part of the overall enterprise, managed as an enterprise, and seen as an enterprise. Session Manager offers:

- a simplified network-wide feature deployment
- centralized routing, SIP trunking, and user profiles
- cost-effective scalability (from small to very large deployments)
- high availability with geographic redundancy
- a secure environment that conforms to specific SIP standards and practices

What does Avaya Aura® Session Manager do?

Avaya Aura® Session Manager offers a core communication service that builds on existing equipment and adds a SIP-based architecture.

Session Manager connects Avaya Aura® Communication Manager as a feature (SIP-only) server or evolution (SIP and non-SIP) server, Avaya enterprise PBX and small key PBX systems within branch offices, third-party PBXs, gateways, service providers, SIP-enabled adjuncts, and SIP and non-SIP telephones. Specifically, Session Manager

- Normalizes disparate networks
- Routes SIP sessions across the network
- Integrates with third-party equipment and endpoints
- Offers centralized management, including user profiles, through Avaya Aura® System Manager
- Supports SIP survivable branches
- Communicates with a session border controller and provides protection at the edge of the enterprise network
- Enables 3rd party E911 emergency call service that supports up to 100,000 users.
- Supports direct SIP connectivity with Avaya Aura® Presence Services and makes Avaya one-X® Communicator, Avaya Soft Flare Core, and 96XX phones as presence-enabled devices.
- Serves as the master control point for Avaya and Polycom video domains.
- With Session Manager and Avaya SIP endpoints, Session Manager provides the ability to search contacts in the enterprise-wide user database for calling, instant messaging, and presence.

Each Session Manager installation combines several or all of the following configurations:

- Centralized routing and dial plan management
- Policy-based routing
    - Time of day routing
    - Alternate routing
    - Load balancing
    - Call admission control
- Tail end hop off (TEHO)
- Centralized SIP trunking

- Centralized applications
  - Registrar, Event State Compositor, Proxy and application sequencing functionality for SIP phones.
  - Geographic redundancy for SIP phones.
- Sequenced applications

It also handles all call redirection, internal network call accounting feeds, toll bypass, interoffice routing, and international least-cost routing.

Communication Manager overview

Avaya Aura® Communication Manager organizes and routes voice, data, image and video transmissions. It can connect to private and public telephone networks, Ethernet LANs, and the Internet.

Communication Manager is a key component of Avaya Aura®. It delivers rich voice and video capabilities and provides a resilient, distributed network for gateways and analog, digital and IP-based communication devices. In addition, Communication Manager delivers robust PBX features, high reliability and scalability, and multi-protocol support. It includes advanced mobility features, built-in conference calling and contact center applications and E911 capabilities.

Communication Manager seeks to solve business challenges by powering voice communications and integrating with value-added applications. It is an open, scalable, highly reliable and secure telephony application. Communication Manager provides user and system management functionality, intelligent call routing, application integration and extensibility, and enterprise communications networking.

Communication Manager enables the virtual enterprise with:

- Robust voice and video call processing capabilities
- Advanced workforce productivity and mobility features
- Built-in conferencing and contact center applications
- Centralized voice mail and attendant operations across multiple locations
- Connectivity to a wide range of analog, digital, and IP-based communication devices
- Support for SIP, H.323, and many industry standard communications protocols over a variety of different networks
- More than 700 powerful features
- High availability, reliability, and survivability.

Evolution server

Communication Manager configured as an evolution server is equivalent to the traditional Communication Manager. It provides Communication Manager features to both SIP and non-SIP endpoints. It uses the full call model.

The connection from the evolution server to the Session Manager server is a non-IMS signaling group. Communication Manager is administered as an evolution server by disabling IMS on the signaling group to Session Manager. Session Manager handles call routing for SIP endpoints and allows them to communicate with all other endpoints that are connected to the evolution server.

With Communication Manager configured as an evolution server:

- H.323, digital, and analog endpoints register with Communication Manager
- SIP endpoints register with Session Manager
- All endpoints receive service from Communication Manager

Branch gateways provide connection preserving failover and failback to Survivable Core and Survivable Remote processors. Communication Manager as an evolution server can support IP-connected port networks, but they are not connection preserving.

Feature server

Communication Manager configured as a feature server provides features to SIP endpoints. It only supports SIP endpoints that are registered to an Avaya Aura® Session Manager. Communication Manager configured as a feature server uses the IP Multimedia Subsystem (IMS) half call model that allows full application sequencing. It is connected to Session Manager via an IMS-enabled SIP signaling group and an associated SIP trunk group.

The Communication Manager feature server has the following constraints:

- The dial plan for IMS users must route all PSTN calls back to Session Manager over the IMS trunk group. Routing of such calls directly to ISDN trunks is not supported.
- Traditional phones such as DCP, H.323, ISDN, and analog are not supported.
- Port networks are not supported.

G430 and G450 gateways provide connection preserving failover and failback to Survivable Core and Survivable Remote processors.

Communication Manager templates overview

Communication Manager as a template is a virtualized version that runs on System Platform. This image has all the features that Communication Manager supports whether it is on a

duplicated server or a branch server. The templates support Communication Manager duplication on S8800, HP ProLiant DL360 G7, or Dell™ PowerEdge™ R610 Server. The templates support Communication Manager which configures as Main, Survivable Core (formerly known as Enterprise Survivable Server - ESS), or Survivable Remote (formerly known as Local Survivable Processor - LSP). In addition, the templates allow customers to use their network infrastructure without dedicated control networks.

Note:

The Communication Manager installation and administration Web pages refer to Survivable Core as Enterprise Survivable Server (ESS) and Survivable Remote as Local Survivable Processor (LSP), respectively.

The advantages of using a solution as a template on System Platform are as follows:

- Simplified and faster installation of the solution
- Simplified licensing of applications and solutions
- Web Console with a common Avaya look and feel
- Remote access and alarming for Avaya Services and Avaya Partners
- Coordinated backup and restore
- Coordinated software upgrades

The Communication Manager templates come in two categories: Avaya Aura® for Communication Manager Main/Survivable Core and Avaya Aura® for Communication Manager Survivable Remote. The templates in each category are listed below:

- Avaya Aura® for Communication Manager Main/Survivable Core template category contains the following templates:
  - Simplex CM Main/Survivable Core
  - Duplex CM Main/Survivable Core
  - Embedded CM Main
- Avaya Aura® for Communication Manager Survivable Remote template category contains the following templates:
  - Simplex Survivable Remote
  - Embedded Survivable Remote

Avaya Aura® for Communication Manager Main/Survivable Core

The Communication Manager Main/Survivable Core templates include the following applications:

- Communication Manager
- Communication Manager Messaging

Note:

Communication Manager Messaging is available only if Communication Manager is configured as the main server. Communication Manager Messaging and Utility Services are not available on Duplex Main/Survivable Core.

- Utility Services

Both Simplex Main/Survivable Core and Duplex Main/Survivable Core templates can be installed on an S8800, HP ProLiant DL360 G7, or Dell™ PowerEdge™ R610 Server.

Note:

The S8800 Server is no longer being sold. The S8800 Server can only be installed as an upgrade.

The Simplex Main/Survivable Core can be installed on an S8510 Server with a total 8 Gb memory as an upgrade only. The Embedded Main template is installed on an S8300D Server in either a G250, G350, G430, G450, or G700 Branch Gateway.

Avaya Aura® for Communication Manager Survivable Remote

The Communication Manager Survivable Remote templates include the following applications:

- Communication Manager
- Branch Session Manager
- Utility Services

The Simplex Survivable Remote is installed on an S8800, HP ProLiant DL360 G7, or Dell™ PowerEdge™ R610 Server. Simplex Survivable Remote can be installed on an S8510 Server with 8 Gb memory as an upgrade only. Embedded Survivable Remote is installed on S8300D Server in either a G250, G350, G430, G450, or G700 Branch Gateway. Both templates are used in the following two scenarios:

- Communication Manager Evolution Server
- Communication Manager Feature Server

Note:

For information on template capacities, refer to the *Avaya Aura® Communication Manager System Capacities Table*.

Port network and gateway connectivity

Communication Manager supports the following connectivity features:

- Circuit switched
- Internet Protocol

  H.248 gateway control. Communication Manager uses standards based H.248 to perform call control to Avaya media gateways, such as the G430. H.248 defines a

framework of call control signaling between the intelligent Avaya 8XXX Servers and multiple “unintelligent” media gateways.

- Separation of Bearer and Signaling. The Separation of Bearer and Signaling (SBS) feature provides a low cost virtual private network with high voice quality for customers who cannot afford private leased lines. SBS utilizes QSIG and replaces DCS + VPN for those customers who need Dial Plan Expansion (DPE) functionality. SBS also utilizes QSIG for communication between Communication Manager systems.

Application Enablement Services overview

Avaya Aura® Application Enablement Services (AES) is a server-based software solution that provides an enhanced set of telephony application programming interfaces (APIs), protocols, Web services, and direct IP access to media. It also supports standards such as Computer Supported Telecommunications Applications (CSTA), Java Telephony API (JTAPI) and Telephony Server API (TSAPI) that utilize Communication Manager features. This makes the full-functionality customization capabilities of Avaya communication solutions accessible to corporate application developers, third party independent software vendors (ISVs), authorized business partners, and systems integrators. All of these services are integrated into a single, secure, scalable, software application with management, redundancy and fail-over capabilities to support mission-critical business needs.

Presence Services overview

Presence is information that conveys a person's ability and willingness to communicate across a set of services, such as telephony and instant messaging. A person's presence is characterized by states such as *busy* and *away*. These states impact the individual's ability and availability to communicate with other users. *Presentity* is the overall presence information for a person whose presence is being reported. *Watcher* is a user who is interested in a presentity. A watcher who is interested in receiving presence updates for a given persentity must subscribe to that presentity for presence updates.

Avaya Aura® Presence Services is a single point of presence collection. It supports presence information gathering from a diverse range of sources. This information is aggregated on a per user basis and then made available to presence aware applications.

Applications interested in a user's presence must first subscribe to receive presence information. Presence aware applications may use the Local Presence Service (LPS) to subscribe to Presence Services. In doing so, they will receive presence change notifications containing aggregated presence for a user and the communication resources that the user has available to them. LPS runs co-resident on the application server. This information can be used to provide visual indications about a users' presence to an end user client Graphical User Interface (GUI). Presence Services uses LPS to efficiently transfer presence information

between the Presence Services server and the application servers. Presence Services uses presentities and watchers to do this.

Presence Services facilitates the secure exchange of telephony availability and instant messaging (IM) information between applications.

System Manager overview

System Manager is a central management system that delivers a set of shared management services and a common console for System Manager and its components. System Manager includes the following shared management services:

Service	Description
Elements	Provides features offered by individual components of System Manager. Except some links that provide access to generic features provided by System Manager, most of the links provides access to features provided by different components of System Manager.
Events	Provides features for administering alarms and logs generated by System Manager and other components of System Manager. You can view and change the status of alarms. For logs, you can view logs, harvest logs for System Manager and its components, and manage loggers and appender.
Groups & Roles	Provides features for administering groups and roles. You can create and manage groups, roles, and permissions.
Licenses	Provides features for administering licenses for individual components of Avaya Aura Unified Communication System.
Routing	Provides features for managing routing applications. You can create and manage routing applications that includes Domains, Adaptations, SIP Entities, Entity Links, Time Ranges, Policies, Dial Patterns, and Regular Expressions to configure your network configuration.
Security	Provides features for configuring certificates.
System Manager Data	Provides features for: • Backing up and restoring System Manager configuration data. • Monitoring and scheduling jobs. • Replicating data from remote nodes. • Configuring data retention settings and profile for various services provided by System Manager.
Users	Provides features to administer users, shared address, public contact list and system presence access control list information. You can create and manage user profiles. You can associate the user profiles

Service	Description
	with groups, roles, communication profiles, create a contact list, add address, and private contacts for the user.

About this book

This book describes the Avaya Aura® solution, IP and SIP telephony product deployment, and network requirements for integrating IP and SIP telephony products with an IP network. The book can be used as a tool to provide a better understanding of the benefits of Avaya IP and SIP solutions and of the many aspects of deploying IP and SIP Telephony on a customer's data network.

This book does not contain procedural information for installing, configuring, administering or maintaining IP or SIP telephony products. Procedural information is contained in other product documentation available at http://www.avaya.com/support .

Audience

The primary audiences for this book are:

- Avaya employees and Avaya Partners working in sales and sales-support organizations.
- Customers considering the purchase of an Avaya Aura® solution products.
- Avaya customers who have purchased Avaya Aura® solution products and are seeking suggestions for their implementation.

Secondary audiences include the Technical Service Center (TSC), training, and development.

What the book covers

The book covers the following main areas:

- Supported architectures
- Avaya Aura® components (applications)
- Hardware components
- Network support
- Call processing
- Sample deployments
- Licensing
- Secure Access Link (SAL)
- Security

- Survivability, reliability, and recovery
- Supported tools
- Traffic engineering

Additional resources

To complete the deployment of the Avaya Aura® solution requires additional information than what's contained in this book. Each component of the Avaya Aura® solution has its own documentation set that provides details on implementing and administering the application. All of the documents are available on the Avaya Support site: http//support.avaya.com under the respective product names.

The following sections provide a brief list of books required for a typical implementation. Check Avaya Support for the complete list of documents required for a complete implementation.

Hardware implementation

- *Installing the Dell™ PowerEdge™ R610 Server*
- *Installing the HP ProLiant DL360 G7 Server*
- *Installing the Avaya G650 Media Gateway*
- *Mediant 5000 Installation and Operation* (Avaya G860 Media Gateway)
- *Mediant 3000 Setup Guide*
- *Quick Start for Hardware Installation: Avaya Branch Gateway G430*
- *Quick Start for Hardware Installation: Avaya Branch Gateway G450*
- *Implementing Avaya B5800 Branch Gateway*

Application implementation

- *Installing and Configuring Avaya Aura® Session Manager*
- *Installing and Configuring Avaya Aura® Communication Manager*
- *Implementing Avaya Aura® Application Enablement Services on Avaya Aura® System Platform*
- *Implementing Avaya Aura® Presence Services 6.1*
- *Installing and Upgrading Avaya Aura® System Manager Release 6.1*
- *Installing and Configuring Avaya Aura® Session Border Controller*
- *Implementing Avaya Aura® Messaging*
- *Communication Server 1000E Installation and Commissioning - Avaya Communication Server 1000*
- *Avaya Video Conferencing Manager Deployment Guide*

Endpoint implementation

- *Implementing Avaya one-X® Communicator*
- *Implementing Avaya one-X® Portall*
- *Implementing and Administering the Avaya A175 Desktop Video Device with the Avaya Flare™ Experience*

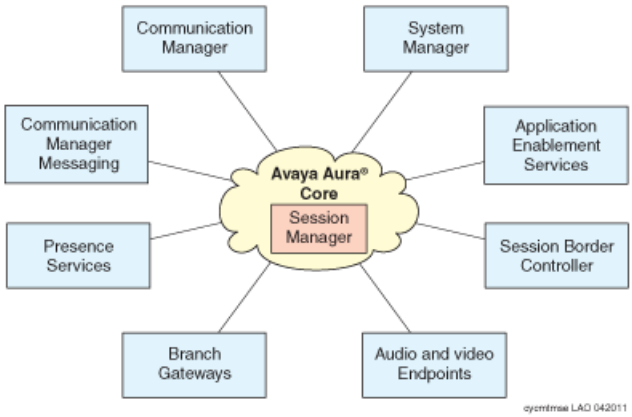
Chapter 2: Avaya Aura solution architectures

Architecture overview

The solution offers several architectures with Session Manager at the core. These architectures support mid- to large enterprises and include a video option.

Standard Edition architecture

The Standard Edition is for mid-to-large enterprises that require comprehensive voice, video, messaging, SIP, and Presence communications capabilities with standard survivability at remote locations. With Standard Edition, customers may easily add licensing for enterprise-wide SIP session management and Unified Communications applications for targeted users, including Microsoft and IBM integration, and mobile worker and teleworker support. The following graphic depicts a typical architecture.



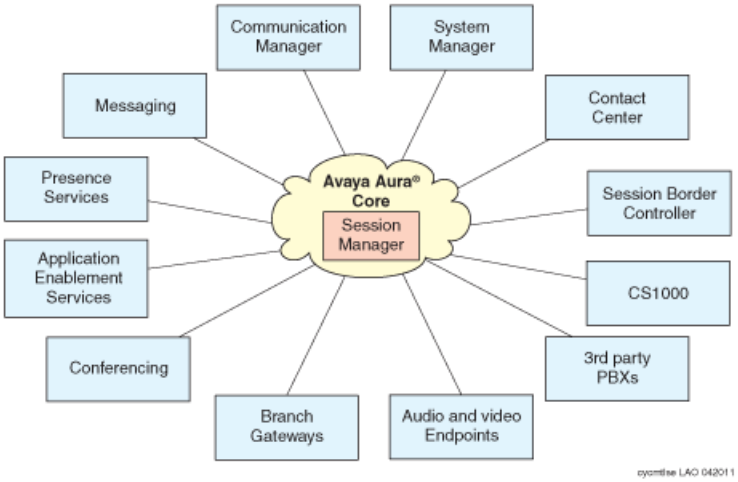
Enterprise Edition architecture

The Enterprise Edition is for highly distributed enterprises requiring the same comprehensive communications capabilities as Standard Edition with increased high availability options, including 100% feature survivability at remote locations. Enterprise Edition also includes, with no additional licensing, enterprise-wide SIP session management and Unified Communications applications for all users, including Microsoft and IBM integration, and mobile worker and teleworker support.

Additional Avaya applications for messaging, conferencing, collaboration, video communications, customer service, and contact centers can all be incrementally added. An extensive array of certified third-party products is also available.

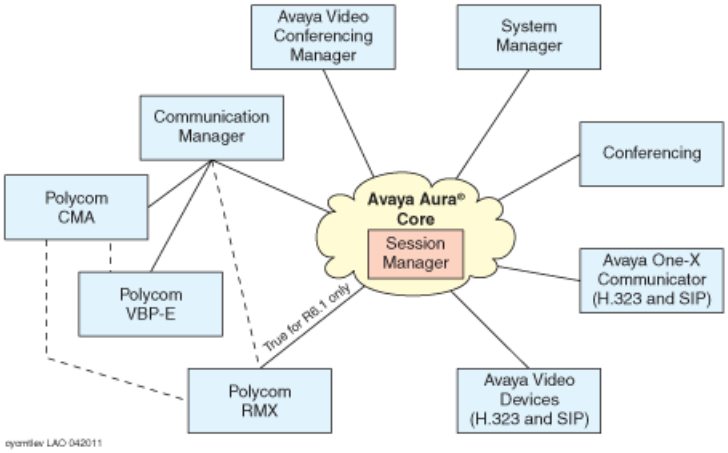
Example enterprise architecture

The following graphic depicts a typical architecture.



Example enterprise with video architecture

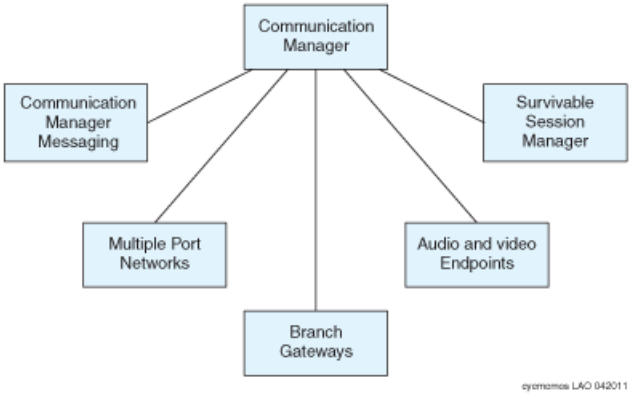
The following graphic depicts a typical architecture for video conferencing.



Communication Manager only architecture

The Communication Manager portfolio covers small, medium, and large enterprises with advanced communications needs between 2 and 48,000 ports per system. This is for customers that do not need the capabilities offered in a complete Avaya Aura® solution.

The following graphic depicts a typical architecture.



Chapter 3: Avaya Aura components

Software components overview

The Avaya Aura® solution consists of several Avaya software applications beyond the core components. The following products are part of the Avaya Aura® solution:

- Avaya Aura® Session Border Controller
- Avaya Aura® Messaging
- Avaya Aura® Conferencing Standard Edition
- Avaya Communication Server 1000
- Avaya Video Conferencing Manager

Many of these applications are installed on Avaya's virtualization software: Avaya Aura® System Platform.

One of the advantages of the Avaya Aura® solution is its ability to work with third-party applications. Supported third party applications include:

- Polycom CMA
- Polycom VBP-E
- Polycom RMX

Avaya Aura® supports several software mobility endpoints. These endpoints include:

- Extension to Cellular (EC500)
- Avaya one-X® Communicator (H.323 and SIP versions)
- Avaya one-X® Portal
- Avaya A175 Desktop Video Device with the Avaya Flare™ Experience

System Platform overview

Avaya Aura® System Platform technology delivers simplified deployment of Unified Communications and Contact Center applications. This framework leverages virtualization technology, predefined templates, common installation, licensing, and support infrastructure.

The advantages of System Platform include:

- Ability to install predefined templates of one or more Avaya software applications on a single server in a virtualized environment
- Simplified and faster installation of software applications and solutions
- Simplified licensing of applications and solutions
- Web Console with a common Avaya look and feel
- Remote access and alarming for Avaya Services and Avaya Partners
- High Availability Failover option for failover using active and standby servers
- Coordinated backup and restore
- Coordinated software upgrades

System Platform enables real-time communications solutions to perform effectively in a virtualized environment. System Platform effectively manages the allocation and sharing of server hardware resources, including the CPU, memory, disk storage, and network interfaces. To continue delivering the high reliability of real-time communications that Avaya customers expect, System Platform is being delivered solely through an *appliance* model, which includes an Avaya Server, System Platform, and the Avaya software applications.

Easy installation

Using solution templates on System Platform significantly reduces the installation time. During the installation, the installer program installs the predefined solution template, which takes less time then installing the applications individually. The installation process is simple and requires that the installer possesses basic software installation skill. System Platform allows remote installation of product-specific templates.

Solution templates

A solution template is a set of one or more applications to be installed on System Platform. Installers must download these templates using the Product Licensing and Delivery System (PLDS) (http://plds.avaya.com). PLDS allows Avaya customers, Avaya Partners, and associates to manage software licensing and to download software for various Avaya products.

System Platform provides an installation wizard for the template. The installation wizard makes it possible for you to configure template-specific parameters, including network and server details, or to upload a preconfigured Electronic Preinstallation Worksheet (EPW) created in a stand-alone version of the installation wizard.

Note:

You must install System Platform before installing the solution template software on a single server.

Remote serviceability

System Platform can be serviced remotely, potentially eliminating the need for a service technician to visit the customer site.

System Platform uses Secure Access Link (SAL), which is an Avaya serviceability solution for support and remote management. SAL provides remote access and alarm reception capabilities for Avaya and participating Avaya Partners.

SAL uses your existing Internet connectivity to facilitate remote support. All communication is outbound from your environment using encapsulated Hypertext Transfer Protocol Secure (HTTPS).

Avaya Partners without a SAL Concentrator must provide their own IP-based connectivity (for example, B2B VPN connection) to deliver remote services.

Important:

Avaya Partners and customers must ensure that SAL is always configured and registered with Avaya during installation. Avaya support will be delayed or not possible if SAL is improperly implemented or not operational. System Platform and SAL do not support modem connections.

High Availability Failover

All System Platform applications can have high availability through redundant hardware and software setup. The secondary node is constantly updated with any configuration changes. In the event of a failure of the primary server the secondary server restores all functionality to the system in a matter of minutes. See System Platform High Availability Failover Overview section in *Installing and Configuring Avaya Aura™ System Platform*.

Note:

This feature is available only for the solution templates supporting this feature.

Virtual machines

System Platform includes a base, which is the host, operating system, and a virtual machine, Console Domain (cdom), that is used to manage the platform. A particular product or solution then provides a template (prebuilt and installed set of virtual appliances) that is installed on the platform. After template installation, the system looks similar to the following diagram.

System Domain (Dom-0):

In addition to exporting virtualized instances of CPU, memory, network and block devices, the system exposes a control interface to manage how these resources are shared between the running domains. Access to the control interface is restricted to one specially-privileged virtual machine, known as domain 0 or System Domain.

Console Domain (cdom):

Console Domain is a virtual machine, which is a part of System Platform and has many platform elements, including the System Platform Web Console.

The Console Domain is capable of deploying and running the following plug-ins:

- Virtual Appliance (VA) plug-ins, which interact with virtual appliances for operations such as backing up and restoring data, providing sanity heartbeats, and getting version information
- Preinstallation plug-ins, which accept parameters that are configured by the user at the time of installation
- Post-installation plug-ins
- Backup plug-in
- Restore plug-in
- Patch plug-ins, which install or uninstall patches in the system

Networking in System Platform

System Platform uses software bridging to support networking for virtual machines. Software bridging works like a network switch inside the system. During installation, System Platform creates two software bridges: avpublic and avprivate.

The avpublic bridge is connected to a physical interface and is intended to be the default connection to your LAN. Most virtual machines have a virtual interface on the avpublic bridge to connect to your network. When connected to your network, these virtual machines can be reached by ping.

The avprivate bridge is not connected to any physical interface and is intended for communication among the virtual machines in a single server. The IP addresses used on avprivate cannot be reached from your network.

Some templates require additional connections to your network. In some cases, this results in System Platform creating another software bridge. This bridge contains the name specified by the template, and this name is displayed during template installation or in the Network Configuration page.

If a virtual machine has high or real time traffic requirements, it can be assigned a dedicated network interface card (NIC) in the template file. This means the virtual machine is assigned another physical NIC on the system (for example, Avaya Aura® Media Services uses eth3) and does not use avpublic. See the respective template documentation for more information.

If a virtual machine in the installed template requires a dedicated NIC, it must have a separate cable connection to your network. Both the avpublic interface and the dedicated NIC must be connected to the network for those machines to communicate in the same way that they would if they were separate physical machines. For example, in the Solution for Midsize Enterprise, the Console Domain is on the avpublic bridge and Application Enablement Services has a dedicated NIC (eth 3). So in this case you must connect eth0 and eth3 to the network before attempting to ping the Application Enablement Services virtual machine from the Console Domain.

Avaya Aura® Session Border Controller overview

Session border controllers (SBCs) are network elements that reside between Session Manager and public SIP service providers. SBCs connect to Session Managers via SIP entity links.

The Avaya Aura® Session Border Controller controls and delivers secure interactive communications, such as voice, video and multimedia sessions, across an enterprise's IP network borders. The Avaya Aura® Session Border Controller supports applications for Unified Communications and Contact Centers, including remote worker and agent solutions. Session Border Controller is powered by Acme Packet, the market leader in session border controller solutions.

Avaya CS 1000E overview

The Avaya Communication Server (CS) 1000E is a robust and highly scalable Internet Protocol (IP) Private Branch eXchange (PBX) that supports traditional Meridian features as well as new IP telephony features, including Session Initiation Protocol (SIP).

With the CS 1000E, customers can evolve from a traditional Time Division Multiplexing (TDM) network to a converged IP network. Deployment is seamless because the CS 1000E integrates with existing PBX systems from Avaya and third parties. This enables customers to expand the size and functionality of their networks while preserving their investment in legacy equipment, such as Meridian 1, Option 11C, and Communication Server 1000 systems.

Being IP-based, the CS 1000E supports distributed architecture. This enables customers to locate systems and components where they fit best. For example, using the Branch Office feature, customers can establish Branch Office Media Gateways (MG 1000B) in remote sites to extend complete feature sets across multiple locations and time zones. Customers can also configure the CS 1000E to support Campus Redundancy and Geographic Redundancy to increase system availability.

Like other Enterprise Solutions from Avaya, the CS 1000E delivers business-grade availability, security, reliability, and scalability. And as always, CS 1000E customers receive industry leading support services from Avaya to ensure successful implementation.

Related topics:

Key attributes on page 32

Key attributes

The Avaya Communication Server 1000E has the following key attributes:

- Adaptable to meet current and future needs. It delivers investment protection and an evolution path to the next-generation multimedia communications
- Superior IP Telephony experience. It has a more open platform to take advantage of innovative applications and feature-rich next generation clients
- Improved reliability and security. It offers business continuity improvement from a reliable and secure environment
- Simplified convergence solution. It offers a product portfolio that is simplified for easier deployment, configuration, and management

Unified communications for business users

Avaya Aura® Conferencing overview

Avaya Aura® Conferencing 6.0 is a fully integrated audio and data conferencing solution for your organization. Conferencing consists of a number of components which provide booking engines, account management utilities, data sharing functionality, billing outputs, directory server integration capabilities, and audio management for all calls.

Typically, the Standard Edition of Conferencing suits smaller deployments. In the Standard Edition of Conferencing, the media server and the application server reside on a single server.

Avaya also supports another conferencing server, called Avaya Aura® Meeting Exchange 5.2 Enterprise Edition. Typically, the Enterprise Edition of Meeting Exchange 5.2 suits larger, more complex deployments. The Enterprise Edition of Meeting Exchange 5.2 supports complicated installations, such as those with multiple application servers, a global distribution of servers, and redundancy requirements. The Enterprise Edition of Meeting Exchange 5.2 also supports additional functionality, such as self registration for conferences, reseller and wholesaler users, and Avaya Web Conferencing recording and playback.

Related topics:

Avaya Aura Conferencing components on page 33

Avaya Aura® Conferencing components

The Avaya Aura® Conferencing Standard Edition contains the following components.

Audio bridge

Avaya Aura® Conferencing Standard Edition allows you to easily deploy on-premises audio-only or audio and Web conferencing across your entire enterprise, providing all of the capabilities you need, in one cost-effective solution. The Dell R610 server is the audio Conferencing server, software-based IP conferencing platform that supports Avaya Conferencing features on standard Intel-based servers. The Dell R610 is interoperable with Avaya and third party IP-PBXs and IP phones.

Client Registration Server

This is a database and server. It utilizes an MS-SQL database to support a wide range of functions including: system administration, booking, scheduling, reporting and billing applications. The end user application CRS Front End is used to interface to the CRS system.

Web Portal

This is a Web based user interface to the CRS and audio bridge. It enables secure and streamlined conference scheduling, management, and administration through an intuitive browser-based interface. Additionally, Web Portal carries the Audio Console application, which can be integrated with Web Portal or be used standalone with the Conferencing bridge. The Audio Console provides moderators with real-time control during conferences, using icons representing each participant and intuitive buttons for tasks such as mute, inbound call intercept, dial out, sub conferencing, and recording/playback controls. In the Standard Edition of Conferencing, the Web Portal cannot reside in a DMZ (as all components of the system reside on a single server) so it’s recommended for internal use only.

Avaya Web Conferencing

Avaya Web Conferencing enables Web conferencing features that include white boarding, PowerPoint presentations, video steaming of moderators, and application sharing functionality using a Web browser.

Audio Codes Mediant 2000

This is a TDM to SIP gateway. It is used to integrate Conferencing into a legacy TDM telephony environment. It supports T1, T1 ISDN, E1 ISDN.

Virtualization

With the exception of the Audiocodes Mediant 2000, the applications above run as Virtual Machines (VM) on a single server. Collectively the VMs are known as the Conferencing template.

- You must install System Platform on the server before the template can be applied. System Platform consists of two components, system domain & console domain.
- As well as CRS, Bridge, Webportal and AWC, there is a virtual machine called Conferencing Manager which provides a single interface to configure the Conferencing applications. Consequently, seven sets of networking data (IP address, hostname, gateway, DNS, NTP) are required for the Conferencing installation. All IPs must be on the

same subnet. The Conferencing Manager virtual machine has been scaled for use in deployments where there is no System Manager installed. Avaya has created Conferencing Manager to manage one and only one Avaya Aura® Conferencing 6.0 Standard Edition server. Customers who have multiple servers require Avaya Aura® System Manager 6.1.x server to manage multiple Avaya Aura® products.

Avaya Aura® Messaging overview

Avaya Aura® Messaging, also referred to as Messaging, is Avaya's next generation messaging product.

Messaging is flexible, scalable, resilient, and easy to deploy on standard Linux-based servers.Messaging is an enterprise-class messaging system targeted for flexible deployment options in single site and multisite environments.

Messaging enables quick and effective communication and collaboration across an enterprise, to enhance employee productivity. Using a variety of features and capabilities that the solution offers, employees can receive and respond to calls and contacts from customers, partners, and coworkers faster and more efficiently.

Messaging can improve your business by enabling employees to work faster and make better decisions while lowering acquisition and operating costs, with unique and powerful messaging capabilities that deliver tangible benefits:

- Allowing important calls to get to the right person, at the right time
- Alerting employees to critical new messages
- Providing fast and easy access to all messages
- Lowering the cost of acquisition, implementation, and ownership of the Messaging systems through standards-based interfaces that allow easy integration with the existing networks, administrative systems, and security processes
- Providing multiple configuration choices for scalability to enable system consolidation — significantly lowering total cost of ownership (TCO) while offering new business continuity options

Message Networking

Message Networking allows networking customers to simplify their network topology and administration by supporting store-and-forward message protocols. With Message Networking, you can exchange messages between supported multimedia messaging systems. Features include:

- Support for multiple network configurations, including hub and spoke, bridge, and hybrid. The bridge and hybrid configurations take advantage of Message Networking's bridging feature
- Support for MultiSite-enabled Modular Messaging remote machines.
- Transport and protocol conversion that automatically transcodes message formats between all supported networking protocols
- Directory views that allow a subset of names and subscriber remote pages to be downloaded from the Message Networking system to a specific location
- Variable-length numeric addressing from Modular Messaging MultiSite and Avaya Aura® Messaging systems
- Dial Plan Mapping, which allows you to map existing mailbox addresses to unique network addresses
- Enterprise Lists that are created using a unique virtual mailbox on the Message Networking system to which subscribers can forward multimedia messages. This mailbox has a voice name and ASCII list name that can be administered. Messages can be addressed by list number or list ASCII name. On receipt of a list message, the system checks the appropriate permissions for use of the list. Once the system verifies the permissions, the Message Network sends the message to all recipients defined in the list.

Avaya Video Conferencing Solution overview

The Avaya Video Conferencing Solution (AVCS) enhances and extends the use of Avaya Aura® Communication Manager, Avaya Aura® Session Manager, and Avaya Aura® System Manager as a SIP-based platform for video communications. Communication Manager is an H.323-to-SIP and a SIP-to-SIP video signaling gateway that supports single and mixed protocol video deployments.

The key video components of AVCS are:

- *Gatekeepers* that provide call admission control and bandwidth management for all SIP and H.323 video calls and video telephony features. Communication Manager and Session Manager serve that role within AVCS. The Polycom CMA is supported only in a neighbored gatekeeper configuration that uses CMA to provide the H.323 call control for all H.323-to-SIP video calls.
- *Video Border Proxy* that combines the function of a gatekeeper with a gateway to proxy SIP and H.323 video calls. The video border proxy currently supported is the Polycom VBP-E series for H.323 video calls. The Polycom VBP-E series is supported only when the VBP-E is trunked to the Polycom CMA and the CMA is neighbored to Communication Manager configured as an evolution server with a branch gateway, such as the Avaya G430 Branch Gateway.
- *Gateways* that provide SIP-to-H.323 video interworking. Communication Manager configured as an Evolution server with a branch gateway, such as the Avaya G430 Branch

Gateway, serves that role. The Polycom RMX is supported when configured as an H.323-to-H.320 gateway. It provides interoperability with video endpoints on public ISDN communication services.

- *Conferencing bridges*, aka Multimedia Conferencing Units (MCU), that support content sharing between video endpoints. The Avaya 1040 and 1050 video endpoints provide that functionality. They have embedded 4- and 6/8-port MCUs, respectively. The Polycom RMX-series HD MCU is also supported.
- *Conferencing scheduling and endpoint management* applications that manage and monitor video endpoints. Avaya Video Conferencing Manager manages and monitors the Avaya 1000-series endpoints on Session Manager. The Polycom CMA manages and monitors Polycom video endpoints.
- *Endpoints* that are the desktop and conference room devices that deliver audio and video. Within AVCS, several video endpoints register directly with Session Manager: the Avaya 1000-series, Avaya one-X® Communicator, Avaya A175 Desktop Video Device with the Avaya Flare™ Experience, and Polycom HDX series. The Polycom HDX H.323 and BSX H.323-series register to Communication Manager configured as an evolution server with a branch gateway. Additional supported H.323 endpoints include the Tandberg MXP.

Other Applications

Communication applications

Communication Manager supports a large number and variety of communication capabilities and applications, including:

Call Center

The Avaya Call Center provides a total solution for a customer's sales and service needs. Building on the performance and flexibility of the Communication Manager, customers can

select from a powerful assortment of features, capabilities, and applications that are specially designed to enhance Call Center operations.

The objective of this offer, which involves new and existing versions of Avaya servers and Communication Manager, as well as a host of attached Call Center peripherals, is to improve Avaya's Call Center offers by supporting increased capacities. These capabilities include 6-digit and 7-digit extensions, LAN backup of Call Management System for the High Availability offer, and customer requested enhancements to be made available in a single global release.

Avaya Call Center applications are designed to efficiently connect each caller with the representative who is best suited to serve that caller. Communication Manager begins the process by capturing information about the caller even before the call is routed. That information is integrated with existing databases, and the combined data is used to match caller to agent.

Communication Manager integrates with a variety of Call Center applications like the Avaya Call Management System for real-time reporting and performance statistics, and with Avaya Business Advocate for expert predictive routing according to incoming calls, not just historical data.

Unified Communication Center

Unified Communication Center lets mobile, remote and office workers easily access important communications tools and information via any telephone using simple and intuitive speech commands.

Avaya Call Management System overview

The Avaya Call Management System is a software product for businesses and organizations that have Communication Manager and receive a large volume of telephone calls that are processed through the Automatic Call Distribution (ACD) feature. CMS collects call-traffic data, formats management reports, and provides an administrative interface to the ACD feature on Communication Manager.

An administrator accesses the CMS database, generates reports, administers ACD parameters, and monitors call activities to determine the most efficient service for the calling customers.

Dual links provide an additional TCP/IP link to a separate CMS for full, duplicated data collection functionality and high availability CMS configuration. The same data are sent to both servers, and the administration can be done from either server. The ACD data is delivered over different network routes to prevent any data loss from such conditions as ACD link failures, CMS hardware or software failures, maintenance, or upgrades.

Computer Telephony Integration (CTI)

Computer Telephony Integration (CTI) enables Communication Manager to be controlled by external applications, and allows integration of customer databases of information with call control features. CTI is a LAN-based solution that consists of server software that runs in a client/server configuration.

CTI opens up Application Programmer Interfaces like ASAI, Telephony Services Application Programming Interface (TSAPI), and Java Telephony Application Programming Interface (JTAPI), which can be used to control the server from an external application.

Application Programming Interfaces (APIs)

Communication Manager supports the following APIs to interface with other applications:

- Adjunct Switch Application Interface (ASAI) allows adjunct applications to access a collection of Communication Manager features and services. Integration with adjuncts occurs through APIs. ASAI is part of Avaya Computer Telephony.
- DEFINITY Application Programming Interface (DAPI) for accessing control and data paths within Communication Manager.
- Java Telephony Application Programming Interface (JTAPI) is an open API supported by Avaya Computer Telephony that enables integration to Communication Manager ASAI.
- Telephony Application Programming Interface (TAPI).
- Telephony Services Application Programming Interface (TSAPI) is an open API supported by Avaya Computer Telephony that allows integration to Communication Manager ASAI.

Best Services Routing (BSR) polling

Best Service Routing (BSR) polling over QSIG Call Independent Signaling Connections (CISCs) and Temporary Signaling Connections (TSCs) provides the ability to do BSR polling between multiple sites over H.323 IP trunks without requiring an ISDN PRI B-channel. QSIG CISC/TSCs are used by BSR polling software to reduce the need for the IP Media Processor circuit pack, thereby making BSR a cost-effective, multi-site solution for an enterprise-wide Contact Center.

Soft clients

Avaya one-X® Communicator overview

Avaya one-X® Communicator helps users manage their communication tasks by providing enterprise users with simple, intuitive access to all of their everyday communications tools. Enterprises can offer Avaya one-X® Communicator to their users in one of the following ways:

- A standalone client that provides basic and advanced telephony features and Instant Messaging and Presence support when integrated with Avaya Aura® Presence Services.
- A Unified Communications client that is integrated with Avaya one-X® Client Enablement Services for 24*7 call logs, with Conferencing Enterprise to provide live audio conference services, and with Avaya Aura® Messaging or Avaya Modular Messaging with the Message Storage Server (MSS) to provide voice message services. This client is integrated with Avaya Aura® Presence Services server and Microsoft Office Communication Server (OCS) to provide Instant Messaging and Presence support across Avaya one-X® Communicator and Microsoft Office Communicator.
- A communications client that connects to the Avaya Communication Server 1000 (CS 1000) and provides basic telephony and video capability through SIP protocol. This client supports instant messaging and presence capability with the use of Extensible Messaging and Presence Protocol (XMPP) with Presence server.

Avaya one-X® Communicator benefits

Avaya one-X® Communicator allows users to increase their productivity with tools that:

- Enhance collaboration with assurance of security
- Improve responsiveness
- Allow them to work anywhere, yet never miss important calls
- Allow them to exchange instant messages
- Allow them to know when a contact is available, on a call, busy, or away
- Make high definition video calls
- Lower costs for IT and end-user support

Avaya one-X® Communicator feature sets

Depending on the communication requirements of your organization, you can select a feature set for Avaya one-X® Communicator. The Avaya one-X® Communicator offer is available with the following feature sets:

- Basic
- Basic with video

- Unified Communications
- Unified Communications with video

Avaya one-X® Portal overview

Avaya one-X® Portal gives you access to Avaya telephony, messaging, conferencing, and presence services. With Avaya one-X® Portal, you do not need multiple applications to gain access to the features provided by Avaya Aura®Communication Manager, Avaya Aura®Presence Services, Avaya Modular Messaging, and Avaya Aura® Conferencing.

Avaya one-X® Portal is a browser-based application. You can gain access to Avaya one-X® Portal from any computer that meets your corporate security requirements. You do not have to install software on the computer to place and receive calls, listen to your voice mail, select the telephone where you want to receive and handle calls, or participate in conferences. If your computer has the needed prerequisites, you can also use the advanced features, such as accessing your business telephone through your computer with VoIP, accessing your corporate contacts, and recording responses to voice mail messages.

Avaya one-X® Portal gives you the flexibility to use the telephone that works best for you, whether you are in the office, at home, in a temporary office, or in a remote location when you travel on business. Therefore, you can use your Avaya telephony, messaging, mobility and conferencing features from any browser on the public Internet, including your home computer, business centers, hotels, and Internet cafes.

Note:

You must connect to Avaya one-X® Portal through your corporate security gateways, such as a virtual private network (VPN) or firewall.

Avaya one-X Agent overview

Avaya one-X® Agent is an integrated telephony softphone solution for agents in contact centers. Avaya one-X Agent provides seamless connectivity to at-home agents, remote agents, out-sourced agents, contact center agents, and agents interacting with clients with speech and hearing impairments.

Avaya one-X Agent 2.5 offers a number of enhancements in addition to the features available in Avaya one-X Agent 2.0.

Avaya one-X Agent 2.5 is compatible with Call Center Elite 6.0, Avaya one-X Agent Central Management 2.5, and Avaya Aura® Communication Manager 2.x and later. Avaya one-X Agent 2.5 also offers interoperability with other IM and presence clients, namely, Avaya one-X® Communicator 6.1 and Microsoft Office Communicator 2007. Avaya one-X Agent 2.5 supports Avaya Aura® Presence Services 6.1 with Avaya Aura® System Manager 6.1 SP1.1.

The availability of features depends on the Avaya one-X Agent user type you deploy. Avaya one-X Agent 2.5 retains all the enhancements with the same user interface so that the existing users of Avaya one-X Agent can adapt easily to the new features.

Avaya one-X® Attendant overview

Avaya one-X® Attendant is the PC-based operator console for Communication Manager. It allows the integrated linking of the telephony with customer data and internal employee's information as free/busy, present or absent, meeting and vacation. Quick, competent and individual service comfort for Attendants in one-X style. Avaya one-X® Attendant extends the classic Attendant system to the multimedia system of information.

Mobility

IP/SIP telephones and softphones

IP and SIP telephones allow access to the features of Communication Manager without having to be tied to one location. One of the major benefits of IP and SIP telephones that you can move the telephones around in the network by unplugging them and plugging them in somewhere else. One of the main benefits of IP softphones is that you can load them on a laptop computer and connect them to Communication Manager from almost anywhere. Users can place calls and handle multiple calls on their PCs.

IP telephones support the following features — Time-To-Service (TTS) capable, gratuitous ARP reply, and accept incoming TCP connection from a server going active. The following table provides the corresponding capabilities of the IP telephones:

IP telephones	TTS aware	Incoming TCP	Gratuitous ARP
96x0 series 9610, 9620/C/L, 9630/G, 9640/G, 9650/C, 9670G 96x1–series 9608, 9611, 9621, 9641	Yes	Yes	Yes
46xx Broadcom series 4601+, 4602SW, 4602SW+, 4610SW, 4620SW, 4621SW, 4622SW, 4625SW, 4630	Yes	Yes	No
46xx Agere series	No	No	No

IP telephones	TTS aware	Incoming TCP	Gratuitous ARP
4601, 4602, 4606, 4612, 4620, 4624			
16xx series 1603, 1603SW-I, 1603SW, 1608, 1616	No	No	No
IP wireless (Polycom) 3641, 3645	No	No	No
IP conference (Polycom) 1692, 4690	No	No	No

SIP telephones include the following audio endpoints:

- 9600–series, including 9620/C/L, 9630/G, 9640/G, 9650/C with Deskphone SIP firmware version 2.6; 9608, 9611, 9621, 9641 with Deskphone SIP firmware version 6.0; 9601 with Deskphone SIP firmware version 6.1(v6.1 supports only the 9601 telephone.
- Avaya A175 Desktop Video Device with the Avaya Flare™ Experience
- Avaya one-X® Communicator in three audio modes:
  - Telecommuter—audio is at the other telephone number provided during registration
  - Road Warrier—audio is on the PC
  - Shared control— audio is on the hard telephone registered with the same extension. Shared control adds an additional registration and subscription to the capacity load.
- Blaze SIP endpoint
- Avaya 1000–series video devices.

Extension to Cellular overview

The Avaya Extension to Cellular feature provides users with the capability to have one administered telephone that supports Communication Manager features for both an office telephone and up to four outside telephones. An office telephone is a telephone that is directly under the control of Communication Manager, such as a desk telephone in an office. The outside telephone is a cellular or wireless telephone and is referred to in this text as a *cell phone.* Extension to Cellular works with any type of wireless or cellular service.

With Extension to Cellular, users can receive and place office calls anywhere, any time. People calling into an office telephone can reach users even if the users are not in the office. In addition, users can access Communication Manager features through the cell phone. Users can enable and disable Extension to Cellular so that the cell phone does not always receive office telephone calls. Users can also switch between the cell phone and office telephone during an ongoing Extension to Cellular telephone call.

When Extension to Cellular is administered and active, a call to the office telephone extension alerts both the office telephone and the cell phone simultaneously. In addition, Extension to Cellular maintains consistency in contact information. The cell phone takes on the identity of the office telephone when calls are made from the cell phone to another number on the same switch. In this circumstance, the cell phone sends the caller ID information of the office telephone. Therefore, calls from the cell phone appear to be from the office telephone number.

A user operates a cell phone as if it were a standard, caller ID-enabled telephone extension connected directly to the Avaya server running Communication Manager. The cell phone acts as an extension because it is mapped to the main office telephone. All other types of cell phone calls, such as direct calls to and from the published cell phone number, are not affected by Extension to Cellular. The cell phone performs exactly as it did before enabling Extension to Cellular. If your Cellular Service Provider (CSP) provides this service, Extension to Cellular is always enabled. You can also enable or disable Extension to Cellular by using a Feature Name Extension (FNE), as described in Setting up Feature Name Extensions set.

Note:

EC500 and CSP work only with ISDN-PRI, ISDN-BRI, H.323, Multi Frequency Compelled (MFC), and SIP trunks.

Cellular service providers who resell the Extension to Cellular service use the CSP or SPFMC (Service Provider Fixed-Mobile Convergence for dual mode phones) application type. CSP/SPFMC support ISDN, H.323, and SIP trunks. CSP/SPFMC is essentially the same as the Extension to Cellular application. Unlike Extension to Cellular, CSP/SPFMC is always enabled. Under CSP/SPFMC, users cannot disable Extension to Cellular.

The Extension to Cellular feature also supports the Fixed Mobile Applications (FMC), Public Fixed Mobility (PBFMC) and Private Fixed Mobility (PVFMC). The FMC applications are used for wireless endpoints that support a one-X Mobile Client application and have two modes called SMode (Single Mode) and DMode (Dual Mode). The FMC applications (PBFMC, PVFMC, and SPFMC) are the only OPTIM application that supports the CTI Mobility Integration feature.

When both PBFMC and the PVFMC applications are administered for a station, calls to that station extend calls out to both the public and private destinations specified in the station-mapping administration. If the private FMC application receives a message indicating that the far-end is alerting, the public FMC application cancels the call. Reception of an alerting indication means that the wireless endpoint must be present in the private wireless network and therefore cannot be in the cellular network.

See also Application RTUs for Fixed Mobile Convergence.

Chapter 4: Hardware

Hardware components

The Avaya Aura® solution includes supported hardware. This hardware includes servers, gateways, desk telephones, and video devices.

Servers

The Avaya software applications are installed on the following supported servers:

- Avaya S8300D Server, an embedded server that resides in the G430 and G450 Branch Gateways.
- Avaya S8800 Server, a standalone server that comes in a 1U or 2U configuration.
- HP ProLiant DL360 G7, a standalone server that comes in a 1U configuration.
- Dell™ PowerEdge™ R610, a standalone server that comes in a 1U configuration.

Gateways

The Avaya Aura® solution uses the following supported gateways.

- Avaya G650 Media Gateway, a traditional gateway that houses TN circuit packs and used in port networks
- Branch gateways
  - Avaya G430 Branch Gateway, a gateway that provides H.248 connectivity and houses media modules.
  - Avaya G450 Branch Gateway, a gateway that provides H.248 connectivity and houses media modules.
- Avaya B5800, a SIP Gateway
- Avaya M3000 Media Gateway, a high-density trunk gateway that provides SIP connectivity

Circuit packs and media modules

Communication Manager often times uses port networks made up of Avaya G650 Media Gateways that houses TN circuit packs. The following circuit packs support IP connectivity:

- TN2312BP IP Server Interface (IPSI), with Communication Manager on a server provides transport of control (signaling) messages.
- TN799DP Control LAN (C-LAN), provides TCP/IP connectivity over Ethernet or PPP to adjuncts

- TN2302AP IP Media Processor (MedPro), the H.323 audio platform
- TN2501AP voice announcements over LAN (VAL), an integrated announcement circuit pack that uses announcement files in .wav format
- TN2602AP IP Media Resource 320, provides high-capacity voice over Internet protocol (VoIP) audio access

Communication Manager also uses branch gateways in lieu of or in addition to port networks. The G430 and G450 Branch Gateways house media modules. The following media modules support IP connectivity:

- MM340 E1/T1 data WAN Media Module, provides one WAN access port for the connection of an E1 or T1 data WAN
- MM342 USP data WAN Media Module, provides one USP WAN access port

Note:

MM340 and MM 342 are no longer sold but if an existing customer already has them they can be used in the G430 and G450 Branch Gateways.

ing

For more information on circuit packs and media modules, see *Avaya Aura® Communication Manager Hardware Description and Implementation* (555-245-207).

Telephones, endpoints, and video devices

The Avaya Aura® solution supports the following Avaya and third party IP (H.323/H.320) and SIP telephones and video devices:

- Avaya IP telephones and devices
  - Avaya IP deskphone series
  - Avaya one-X deskphone series
  - Avaya 1600/9600–series specialty handsets
  - Avaya 4600–series IP telephones
  - Avaya IP conference telephones
  - Avaya 1000–series video devices
  - Avaya A175 Desktop Video Device with the Avaya Flare™ Experience
- Third-party telephones and video devices
  - Polycom VSX/HDX endpoints
  - Tandberg MXP endpoint

For more information on telephones and video devices, see *Avaya Aura® Communication Manager Hardware Description and Implementation* (555-245-207) and documentation on the individual telephones and video devices.

Supported servers

Avaya S8800 Server overview

The Avaya S8800 Server supports several Avaya software applications. The server is available in a 1U model or 2U model and with various hardware components. The server model and specific hardware components in your server depend on the requirements of the software application that will run on the server.

Avaya one-X Speech uses the Avaya S8800 2U Server.

Message Networking uses the Avaya S8800 2U Server.

Collaboration Server uses the Avaya S8800 1U Server.

Solution for Midsize Enterprise uses the Avaya S8800 1U Server.

Related topics:

Avaya S8800 1U Server specifications on page 47
Avaya S8800 2U Server specifications on page 48
S8800 Server environmental requirements on page 50

Avaya S8800 1U Server specifications

Type	Description
Dimensions	Height: 43 mm (1.69 inches, 1U) Depth: 711 mm (28 inches) Width: 440 mm (17.3 inches)
Weight	Maximum weight: 15.4 kg (34 lb.) when fully configured.
Heat output	Approximate heat output: • Minimum configuration: 662 Btu per hour (194 watts) • Maximum configuration: 1400 Btu per hour (400 watts) Heat output varies depending on the number and type of optional features that are installed and the power-management optional features that are in use.
Acoustic noise emissions	Declared sound power, operating: 6.1 bel The sound levels were measured in controlled acoustical environments according to the procedures specified by the

Type	Description
	American National Standards Institute (ANSI) S12.10 and ISO 7779 and are reported in accordance with ISO 9296. Actual sound-pressure levels in a given location might exceed the average values stated because of room reflections and other nearby noise sources. The declared sound-power levels indicate an upper limit, below which a large number of computers will operate.
Electrical input requirements	• Sine-wave input (47–63 Hz) required • Input voltage low range: - Minimum: 100 V AC - Maximum: 127 V AC • Input voltage high range: - Minimum: 200 V AC - Maximum: 240 V AC • Input kilovolt-amperes (kVA), approximately: - Minimum: 0.194 kVA - Maximum: 0.700 kVA
Front connectors	• Two USB • Video
Back connectors	• Two Ethernet (RJ 45). Optionally, two or four additional Ethernet. • Serial • Two USB • Video • Systems management Ethernet (IMM)

Avaya S8800 2U Server specifications

Type	Description
Dimensions	Height: 85.2 mm (3.346 inches, 2U) Depth: 729 mm (28.701 inches) Width: 482.0 mm (18.976 inches)
Weight	Maximum weight: 29.03 kg (64 lb.) when fully configured

Type	Description
Heat output	Approximate heat output: • Minimum configuration: 307 Btu per hour (194 watts) • Maximum configuration: 1700 Btu per hour (500 watts) Heat output varies depending on the number and type of optional features that are installed and the power-management optional features that are in use.
Acoustic noise emissions	Declared sound power, operating: 6.5 bel The sound levels were measured in controlled acoustical environments according to the procedures specified by the American National Standards Institute (ANSI) S12.10 and ISO 7779 and are reported in accordance with ISO 9296. Actual sound-pressure levels in a given location might exceed the average values stated because of room reflections and other nearby noise sources. The declared sound-power levels indicate an upper limit, below which a large number of computers will operate.
Electrical input requirements	• Sine-wave input (50–60 Hz) required • Input voltage range automatically selected • Input voltage low range: - Minimum: 100 V AC - Maximum: 127 V AC • Input voltage high range: - Minimum: 200 V AC - Maximum: 240 V AC • Amperes: - 7.8 A at 100–127 V AC - 3.8 A at 200–240 V AC • Input kilovolt-amperes (kVA), approximately: - Minimum: 0.12 kVA - Maximum: 0.78 kVA
Front connectors	• Two USB • Video
Back connectors	• Two Ethernet (RJ 45). Optionally, two or four additional Ethernet. • Serial • Two USB

Type	Description
	• Video • Systems management Ethernet (IMM)

S8800 Server environmental requirements

Server status	Air temperature	Maximum Altitude	Relative humidity
Server on	10º C to 35º C (50º F to 95º F) at altitude of up to 914.4 m (3,000 feet)	2,133 m (7,000 feet)	8% to 80%
	10º C to 32º C (50º F to 90º F) at altitude of 914.4 m to 2,133 m (3,000 to 7,000 feet)		
Server off	10°C to 43°C (50.0°F to 109.4°F)	2,133 m (7,000 feet)	8% to 80%

HP DL360 G7 Server overview

The Avaya Common Servers category includes the HP ProLiant DL360 G7 1U server that supports several Avaya software solutions, some requiring additional hardware and memory requirements beyond the standard configuration. This book covers the standard configuration only—consult specific Avaya product documentation for application-specific or solution-specific server configurations.

Related topics:

HP DL360 G7 Server physical specifications on page 50
HP DL360 G7 Server environmental specifications on page 51

HP DL360 G7 Server physical specifications

Type	Description
Dimensions	Height: 4.32 cm (1.70 in)
	Width: 42.62 cm (16.78 in)

Type	Description
	Depth: 69.53 cm (27.38 in)
Weight (maximum; two processors, two power supplies, eight hard disk drives)	15.97 kg (35.20 lb)
Weight (minimum; one processor, one power supply, no hard drives)	14.51 kg (32.00 lb)
Weight (no drives installed)	14.06 kg (31.00 lb)

HP DL360 G7 Server environmental specifications

Specification	Value
Temperature range	**Note:** All temperature ratings shown are for sea level. An altitude derating of 1°C per 300 m (1.8° per 1,000 ft.) to 3048 m (10,000 ft.) is applicable. No direct sunlight allowed.
Operating	10°C to 35°C (50°F to 95°F)
Shipping	-40°C to 70°C (-40°F to 158°F)
Maximum wet bulb temperature	28°C (82.4°F)
Relative humidity (noncondensing)	**Note:** Storage maximum humidity of 95% is based on a maximum temperature of 45° C (113°F). Altitude maximum for storage corresponds to a pressure minimum of 70 kPa.
Operating	10% to 90%
Non-operating	5% to 95%

Dell R610 Server overview

The Avaya Common Servers category includes the Dell™ PowerEdge™ R610 1U server that supports several Avaya software solutions, some requiring additional hardware and memory requirements beyond the standard configuration. This book covers the standard configuration

only—consult specific Avaya product documentation for application-specific or solution-specific server configurations.

Related topics:

Dell R610 Server physical specifications on page 52
Dell R610 Server environmental specifications on page 52

Dell R610 Server physical specifications

Type	Description
Dimensions	Height: 4.26 cm (1.68 in)
	Width: • 48.24 cm (18.99 in) with rack latches • 42.4 cm (16.99 in) without rack latches
	Depth: • 77.2 cm (30.39 in) with power supplies and bezel • 73.73 cm (29.02 in) without power supplies and bezel
Weight (maximum configuration)	17.69 kg (39 lb)
Weight (empty)	13.25 kg (29.2 lb)

Dell R610 Server environmental specifications

Specification	Value
Temperature	
Operating	10° to 35°C (50° to 95°F) with a maximum temperature gradation of 10°C per hour **Note:** For altitudes above 2,950 feet, the maximum operating temperature is de-rated 1°F per 550 ft.
Storage	-40° to 65°C (-40° to 149°F) with a maximum temperature gradation or 20°C per hour

Specification	Value
Relative Humidity	
Operating	20% to 80% (non-condensing) with a maximum humidity gradation of 10% per hour
Storage	5% to 95% (non-condensing) with a maximum humidity gradation of 10% per hour
Altitude	
Operating	-16 to 3,048 m (-50 to 10,000 ft.) **Note:** For altitudes above 2,950 ft, the maximum operating temperature is de-rated 1°F per 550 ft.
Storage	-16 to 10,600 m (-50 to 35,000 ft.)

Avaya S8300 Server

An S8300 Server is an Intel Celeron-based processor that runs on the Linux operating system. It resides in one of the following gateways: G250, G350, G430, G450, G650, or G700.

Related topics:

S8300D Server configuration on page 53

S8300D Server configuration

The S8300D Server is supported by Communication Manager Release 5.2 and later.

An S8300D Server is an Intel Core 2 Duo U5700 processor that runs on the Linux operating system. The S8300D Server resides in Slot V1 of a gateway and includes:

- 80-GB hard disk
- 4-GB DRAM (with one 1 GB DIMM)
- 8-GB Internal Solid State Drive (SSD)
- Three USB ports and a 10/100 Base-T port
  - One USB port supports a readable DVD/CD-ROM drive, which is used for system installations and upgrades.
  - Another USB port can be used for a USB modem.
  - A third USB port can be used for a Compact Flash drive.

- One services port
- One internal Compact Flash drive which is used as the primary reboot device
- Modem support for alarming

Supported gateways

Avaya G650 Media Gateway overview

The Avaya G650 Media Gateway is a 14-slot, rack-mounted carrier configured for TN circuit packs. The media gateway has redundant, hot-swappable power supplies and offers AC or DC power. The backplane supports 14 circuit packs and 2 power supplies and provides monitoring of system fans, power supplies, and temperature. Up to five G650 Media Gateways can be mounted in an EIA-310 standard 19-inch (48 cm) rack.

The G650 Media Gateways are used as port networks and work with standalone servers, such as the Avaya S8800 server, the HP ProLiant DL360 G7 server, and the Dell™ PowerEdge™ R610 server.

Avaya G860 Media Gateway

Avaya G860 Media Gateway overview

The Avaya G860 Media Gateway is a DS3-capable, high-channel-density, standards-compliant, VoIP media gateway system. It provides a robust, scalable, and modular solution designed for a large campus or call center with high availability and reliability. To support high availability, the Avaya G860 Media Gateway features automatic protection switching and full redundancy for all common equipment.

The media gateway is a high capacity, cost-effective IP telephony trunking system that supports up to four T3s (redundant 3+1 media gateway board configuration). The media gateway supports SIP for interoperability with a wide range of communications applications.

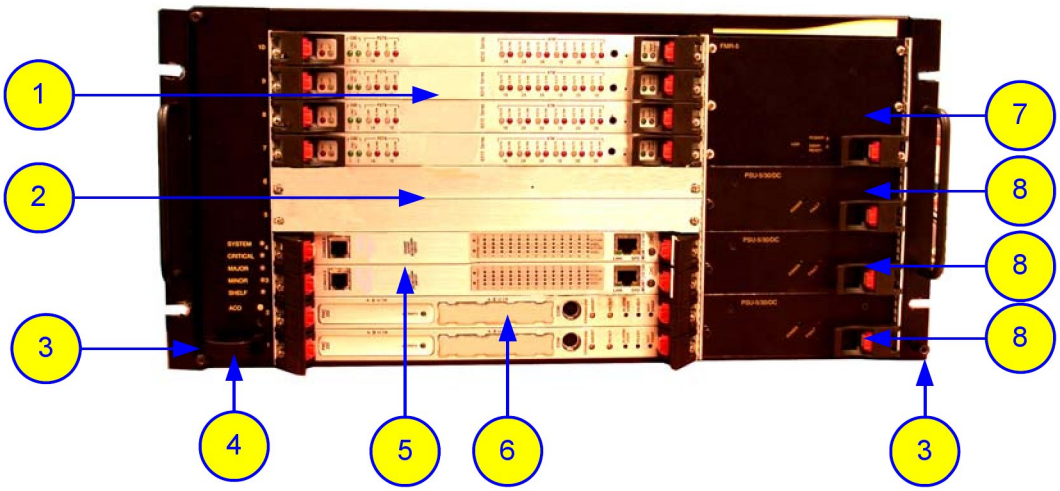
G860 features include:

- Support of up to 6000 voice channels in a Communication Manager configuration
- Supports multiple DS3s
- Redundant power supply units, system controller, Ethernet switch

- Optional N+1 media gateway boards
- Scalable density options
- Open, scalable SIP-based architecture

Avaya G860 Media Gateway — front view

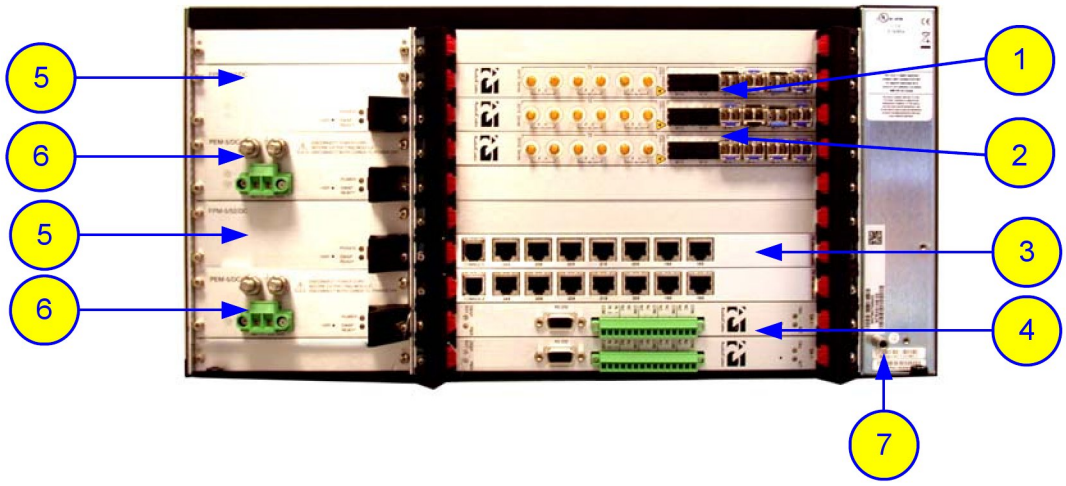
The figure shows the front view of the G860 Media Gateway.



1	Trunk Processing Module (TPM) boards
2	Blank and baffled panels
3	ESD connectors on the attachment brackets
4	Fan Tray Module (FTM) with alarm LEDs
5	ES 6600 Ethernet switch boards
6	System controller (SC) boards
7	FMR Auxiliary fan module
8	Power supplies

Avaya G860 Media Gateway — back view

The figure shows the back view of the G860 Media Gateway.



1	TPM I/O Rear Transition Module (TPM/RTM/Redundant)
2	TPM/RTM
3	ES/6600/RTM
4	SA/RTMs
5	PEM units
6	FPM advanced fan module
7	ESD connections

G860 Components

The G860 media gateway hardware consists of the chassis, TP6310 (media gateway board), Ethernet Switch, System Controller, and the corresponding Rear Transition Modules (RTM).

System Controller Board

The G860 contains two System Controller (SC) boards, which control and monitor the G860 media gateway operation. The SC boards are installed into their dedicated slots. Each controller contains an on-board hard disk, which stores the SC software and the configuration and performance database.

The SC board incorporates a 650 Mhz UltraSparc processor with 512 MB memory and uses the robust Solaris operating system environment enhanced for advanced high-availability features. The SC board is designed in compliance with the PICMG CompactPCI standards for high-availability systems. It supports hot-swap operation, system management, and environmental monitoring.

Two 10/100 Base-TX redundant Ethernet ports connect the SC boards with two Ethernet Switch boards. Each SC board is accompanied by a Synchronization and Alarm (SA) Rear Transition Module (RTM) board.

G860 Trunk Media Processing Module (TP-6310)

The G860 Trunk Processing Module (TP-6310) is a high-density, hot-swappable, compact PCI resource board with a capacity of 672 DS0 channels, supporting all necessary functions for voice, data, and fax streaming over IP networks. TP-6310 provides STM-1/OC-3 (future), PSTN, ATM, and T3 interfaces via its Rear transition Module (RTM).

Slots 7 to 10 of the G860 chassis are used for up to 4 trunk processing modules (including the redundant TP-6310) according to customer requirements. The PSTN interface and the ATM interfaces are provided with 1+1 protection.

For redundant N+1 protection, the 6310/RTM/HA/Redundant Standby board is provided. It contains no port connection and occupies slot 10.

Note:
The Trunk Processing Module is hot-swappable for redundant systems.

Mediant 3000 Media Gateway

The Mediant 3000 is a feature rich high density SIP trunk gateway. The Mediant 3000 offers channel scalability of up to 1932 DSO's (to carry voice calls) and 84 D-channels in a compact 19" - 2U chassis. Additionally, Mediant 3000 delivers the same carrier-grade availability that service providers are accustomed to on their legacy equipment. Mediant 3000 also provides trunking and access protocol, such as PRI. Mediant 3000 meets the needs of wireline, cable, cellular, and mixed service providers.

Enterprises migrate to VoIP due to cost considerations and for a richer, integrated telephony service. An enterprise can choose to connect to a PSTN Service Provider or to an Internet Telephony Service Provider (ITSP) or both. Large enterprises deploy business critical contact centers where the high availability of Mediant 3000 is a key factor. Mediant 3000 supports high-density PSTN interfaces, such as T3 and OC3 (future release).The proven interoperability of Mediant 3000 with different PBXs and PSTN switches facilitates smooth deployment.

Branch gateways

Avaya G430 Branch Gateway

The Avaya G430 Branch Gateway is a multipurpose Branch Gateway targeting small and medium branches of 1 to 150 users. The G430 Branch Gateway supports two expansion modules to support varying branch office sizes. It works in conjunction with IP telephony Communication Manager software running on Avaya S8xxx Servers to help deliver intelligent communications to enterprises of all sizes.

The G430 Branch Gateway combines telephone exchange and data networking by providing PSTN toll bypass and routing data and VoIP traffic over WAN. The G430 Branch Gateway

features a VoIP engine, an optional WAN router, and Ethernet LAN connectivity. The G430 Branch Gateway provides full support for Avaya IP and digital telephones, as well as analog devices such as modems, fax machines, and telephones.

Avaya G450 Branch Gateway

The Avaya G450 Branch Gateway is a multipurpose Branch Gateway that can be deployed in medium to large sized branch locations or in wiring-closets servicing buildings and floors in a campus environment. It works in conjunction with Communication Manager IP telephony software running on Avaya S8xxx Servers to help deliver intelligent communications to enterprises of all sizes.

The G450 Branch Gateway combines telephone exchange and data networking by providing PSTN toll bypass and routing data and VoIP traffic over WAN. The G450 Branch Gateway features a VoIP engine, an optional WAN router, and Ethernet LAN connectivity. The G450 Branch Gateway provides full support for Avaya IP and digital telephones, as well as analog devices such as modems, fax machines, and telephones.

B5800 Branch Gateway Overview

The Avaya B5800 Branch Gateway is a single-platform solution with multiple deployment options that enable seamless, user-centric access to Avaya Aura® Messaging, Avaya Aura® Conferencing, Avaya Aura® Presence services and much more. It's complimentary to any existing networking solution, adding communications and collaboration functionality in a “thin” device designed for branch use. Supporting either distributed, centralized, or concurrent network deployments, the B5800 Branch Gateway is adaptable to meet the needs of specific features and applications of individual employees in each branch location. The result is a smooth migration between architectures. In addition to centralized SIP endpoints, the B5800 Branch Gateway can concurrently support other IP and TDM endpoints for a community of centralized and distributed users on the same platform. Ideal for customers wanting applications deployed in customer data centers and/or in the branch itself, the B5800 Branch Gateway enables the branch to cost effectively deliver the range of communication tools without complex infrastructure and administration.

Chapter 5: Network Support

Voice quality network requirements

In addition to the influence of the telephony terminals at either end of a connection, there are several network parameters that can affect voice quality. This chapter lists some of the more important ones. The concept of voice quality has different aspects that need to be properly understood and considered. IP Telephony quality can be engineered and administered to several different levels to accommodate differing business needs and budget. Avaya therefore provides network requirements options to allow the customer to choose which *voice quality* level best suits their specific business needs.

Before implementing IP Telephony, Avaya recommends a network assessment to measure latency, jitter, and packet loss to ensure that all values are within bounds.

Network delay

In IP networks, packet delay (latency) is the length of time for a packet to traverse the network. Each element of the network, including switches, routers, WAN circuits, firewalls, and jitter buffers, adds to packet delay.

Delay can have a noticeable effect on voice quality but can be controlled in a private environment (LAN/WAN). For example, delay can be reduced by managing the network infrastructure or by agreeing on a Service Level Agreement (SLA) with a network provider. An enterprise has less control over the delay when using the public Internet for VoIP.

Previously, the ITU-T recommended 150 ms one-way delay (including endpoints) as a limit for conversations. However, this value was largely misinterpreted as the limit to calculate a network delay budget for connections. Depending on the desired voice quality, network designers might choose to exceed this number for their network.

Some of the issues that must be considered when designing a network for VoIP are:

- One-way delays in excess of 250 ms can cause the well-known problem of *talk-over*. This occurs when both parties talk at the same time because the delay prevents them from realizing that the other person has already started talking.
- In some applications, delays less than 150 ms can impact the perceived quality, particularly in the presence of echo.
- Long WAN transports must be considered as a major contributor to the network delay budget, averaging approximately 10-20 ms per 1000 miles. Some transport mechanisms,

such as Frame Relay, can add additional delay. Thus, staying within 150 ms, end to end, may not be possible for all types of connections.

- Finally, one-way delay over 400 ms on signaling links between port networks and the S8xxx series Server can cause port network instability.

Again, there is a trade-off between voice quality and the technical and monetary constraints which businesses confront daily. For this reason, Avaya suggests the following guidelines for one-way LAN/WAN delay between endpoints, not including IP phones:

- 80 ms delay or less yields the best quality.
- 80 ms to 180 ms delay can give Business Communication quality. This delay range is much better than cell-phone quality if echo is properly controlled and, in fact, is very well suited for the majority of businesses.
- Delays exceeding 180 ms might still be quite acceptable depending on customer expectations, analog trunks used, codec type, and the presence of echo control in endpoints or network equipment.

Codec delay

In addition to delay incurred in the network, codecs in the endpoints also add some delay. The delay of the G.711 codec is minimal. However, the G.729 codec, for example, adds:

- approximately 10 ms of algorithmic delay in each direction
- another 5 ms look-ahead
- plus signal processing delays.

The compression algorithm in G.723.1 uses multiple blocks (called frames) of 30 ms voice samples per packet. This results in increased latency over codecs configured to use 20 ms or less samples per packet.

The G.722 codec adds a 0.82 ms delay.

Jitter

Jitter is the statistical average variance of the arrival time between packets received from the IP network. To compensate for jitter, a de-jitter buffer is implemented in VoIP endpoints. The purpose of the jitter buffer is to hold incoming packets for a specified period of time such that voice samples can be played out at a regular rate to the user. In doing so, the jitter buffer also adds packet delay.

Excessive jitter might cause additional delay if the jitter still fits the size of the jitter buffer. Excessive Jitter might also result in packet discard creating audible voice-quality problems when the variation is greater than the jitter buffer size. Dynamic jitter buffers give the best quality. Static jitter buffers should generally be sized at twice the largest statistical variance between packet arrivals. However, care needs to be taken in the design of the resizing algorithm of dynamic buffers in order to avoid adverse effects. Dynamic jitter buffering can

exacerbate problems in an uncontrolled network. The network topology can also affect jitter. The existence of multiple paths between endpoints with load balancing enabled in routers can contribute significant amounts of jitter.

The following Avaya products have dynamic jitter buffers to minimize delay by automatically adjusting the jitter buffer size:

- Avaya G350, G430 and G450 Branch Gateways and the G700 and G650 Media Gateway with the TN2302AP IP Media Processor or TN2602 IP Media Resource 320 circuit pack
- Avaya IP SoftPhone software
- Avaya 4600 Series IP Telephones

Packet loss

Packet loss occurs when packets are sent but not received, or are received too late to be processed by the endpoint jitter buffer. Too much delay or packet mis-order can be perceived as lost packets. It may appear that the network is losing packets when in fact they have been discarded intentionally because of late arrival at the endpoint. IP networks are characterized by unintentional packet loss in the network as well as by discarded packets in the jitter buffers of the receiving endpoints.

Packet loss can be bursty or more evenly distributed. Bursty packet loss has a greater effect on voice quality than distributed loss. Therefore, a 1% bursty loss will have more adverse effect than a 1% distributed loss.

The effects of packet loss on VoIP service are multifold:

- Problems caused by occasional packet loss are difficult to detect because each codec has its own packet loss concealment method (PLC). Therefore, it is possible that voice quality would be better using a compression codec (G.729A), which includes its own PLC, compared to a full bandwidth G.711 codec without PLC.
- Packet loss is more noticeable for tones like fax or modem (other than DTMF) than for voice. The human ear is less able to detect packet loss during speech (variable-pitch), than during a tone (consistent pitch).
- Packet loss is more noticeable for contiguous packet loss than for random packet loss over time. For example, the effect of losing ten contiguous packets is worse than losing ten packets evenly spaced over an hour time span.
- Packet loss is generally more noticeable with larger voice payloads per packet than with smaller packets, because more voice samples are lost in a larger payload.
- In the presence of packet loss, the time for a codec to return to normal operation depends on the type of codec.

- Even small amounts of packet loss, 0.12%, can greatly affect a TTY/TDD device's (for hearing-impaired people) ability to work properly.
- Packet loss for signaling traffic increases network traffic substantially when the loss exceeds 3%, possibly impacting voice quality.

Network packet loss

Like packet delay, Avaya offers customers a tiered approach of recommendations to deal with network packet loss to balance new network costs and the constraints of business directives.

The maximum loss of IP packets (or frames) between endpoints should be:

- 1% or less for best quality.
- 3% or less for Business Communications quality. Again, this quality is much better than cell-phone quality.
- More than 3% may be acceptable for voice but may negatively impact signaling, which might degrade voice quality due to increased traffic. More information on signaling bandwidth requirements can be found in white papers on the Avaya Support website (http://support.avaya.com.

Many 3rd party vendor tools such as Prognosis can measure packet loss for ongoing calls.

Packet loss concealment (PLC)

Some packet loss can be dealt with by attempting to conceal the loss by generating voice samples to take the place of the missing samples. ITU standards G.711 Annex I and the G.729 standard define methods by which packet loss concealment can be provided. Excessive packet loss cannot be disguised so, ultimately, PLC gives way to comfort noise generation (CNG) if too many packets are lost in succession.

Ramping down to silence is a typical way that PLC is performed. Loss of six consecutive packets is considered to be the maximum number of packets over which PLC can be sensibly applied.

Echo

The two main types of echo are acoustic echo and electrical echo caused by hybrid impedance mismatch. Usually, in a 2-party call, only the speaker hears an echo but the receiver does not. However, in a conference call many parties might hear an echo.

Acoustic echo occurs when a talker’s voice traverses through the airpath in the acoustic environment of the receiver and feeds back to the microphone of the receiver’s terminal. The

severity of the echo effect depends on the acoustic properties of the remote room. For example, room size and wall reflection characteristics.

Electrical echo is also a reflection effect but is due to an impedance mismatch between four-wire and two-wire systems or in the interface between a headset and its adapter.

The user's perception of echo increases with delay. In practice, most echo received within 30 ms is ignored by the human auditory system. However, if the level of the received echo signal is extremely high, even a couple of milliseconds delay will cause adverse perception of echo. Echo received after 30 ms will generally be perceived as an annoyance. Because of the end-to-end latency in some IP Telephony implementations exceeds the latency in some circuit-switched systems, the perception of echo can be greater in the IP Telephony system.

One strategy for dealing with echo is the deployment of echo cancellers at strategic places in phones or network equipment. Echo cancellers, which have varying amounts of memory, store incoming voice streams in digital form in a buffer and compare the received voice with the previously transmitted voice patterns stored in memory. If the patterns match, the canceller considers it an echoed signal and attempts to remove it.

Because echo cancellers are not perfect, there is a residual level of echo left even in optimal operating conditions. Echo cancellers operate properly only if the one-way delay between the echo canceller and the echo source (for example, the acoustic airpath at the telephone set or electrical hybrid) is not larger than the echo canceller can handle, otherwise, the echo canceller will not find a pattern to cancel.

Avaya's G350, G430 and G450 Branch Gateways and G700 Media Gateway, the Avaya MM760 VoIP Media Module, the Avaya TN2302AP IP Media Processor (in the G650 Media Gateways), the Avaya TN2602AP IP Media Resource 320, the Avaya IP SoftPhone, and all IP Telephones incorporate echo cancellation designed for IP Telephony to improve voice quality.

Signal levels

In order to provide more natural-sounding conversations, voice communication systems attempt to emulate a typical communication scenario where the two parties are speaking directly and are separated by one meter. To achieve these conditions, an acoustic loss of 10 dB is added between speaker and listener. Any significant differences from this loss level will be audible as a signal level that is either too soft or too loud and thus may result in some degree of listener discomfort.

In IP Telephony networks, the end-to-end loss of 10 dB is implemented as 8 dB in the speaker's telephone, 0 dB in the IP network, and another 2 dB loss in the listener's telephone. To account for personal preferences or the presence of background distortions, listeners may adjust relative to the 10 dB loss value by adjusting the volume control of their telephone. The IP Telephony loss values are globally identical and specified in ITU-T Recommendations.

In traditional circuit-switched networks the telephone transmit, receive, and inter-port line/trunk losses are country-dependent. The end-to-end country-specified losses often also differs

somewhat from the 10 dB loss value for historical reasons. The country-dependency of loss values makes it more challenging to guarantee a proper listener signal level when the PSTN is involved or when country borders are traversed.

IP Telephony gateways should provide proper signal level adjustments in the direction from the IP network to the circuit-switched network and in the reverse direction, and also between circuit-switched ports.

To allow for multi-country deployment of Avaya telephones and gateways, these devices facilitate programmable loss values. In order to ensure that the signal levels are controlled properly within the scope of a voice network consisting of Avaya systems, the appropriate country-dependent loss plan should be administered.

In addition to administering loss for two-party calls, Communication Manager allows country-dependent conference call loss administration. Loss is applied depending on the number of parties involved in the conference.

Echo and Signal Levels

As mentioned before, in circuit-switched telephony, echo may be caused by acoustic reflection in the remote party's environment, or by electrical reflection from 2-to-4 wire analog hybrid impedance mismatches. Impedance mismatch can occur in analog telephones and analog line/trunk cards, electrical cross-talk in circuitry, or in telephony wiring (particularly in low-cost headsets). For this reason, in circuit-switched analog and digital phones, a relatively large transmit loss is implemented in order to minimize the perceived echo due to electrical reflection and cross-talk effects. In principle, the transmit loss of telephones could be made very large followed by signal amplification in the receiving telephone. In practice however, the transmit loss should be limited to prevent the electrical voice signal from dropping below electrical background noise. This has resulted in the adoption of transmit loss and receive loss values around 8 dB and 2 dB, respectively, although country-specific values may actually deviate quite significantly from these values.

The loss plan administration provided by Avaya Communication Manager software is primarily intended to control signal losses in telephones and gateways, and not intended to control echo. However, in case of severe echo, the administered loss can be changed to a different plan. In general, an increase of the loss between two endpoints by a certain amount will decrease the echo level by twice this amount. It is not advised to use loss plan administration in this way without consultation with Avaya Services personnel. It is better to reduce the echo by using Avaya products with echo cancellers to minimize echo.

Tone Levels

The level of call progress and DTMF tones played out through telephones must adhere to specified levels in order to be satisfying for the average user. Again, respective standards are country specific and can be set in Communication Manager by administration. The volume of received call progress tones can be adjusted by the telephone volume control.

Audio codecs

Codecs (Coder-Decoders) convert between analog voice signals and digital signals. Avaya supports several different codecs offering varying bandwidth usage and voice quality, including the following codecs:

- G.711 produces uncompressed audio at 64 kbps
- G.729 produces compressed audio at 8 kbps
- G.723.1 produces compressed audio at 5.3 or 6.3 kbps
- G.722 produces compressed audio at 64, 56, or 48 kbps
- G.726 produces compressed audio at 32 kbps

Table 1: Comparison of speech coding standards (without IP/UDP/RTP overhead) on page 65 provides a comparison of voice quality considerations associated with some of the codecs supported by Avaya products.

Toll-quality voice must achieve a MOS (mean opinion score) of 4 or above. The MOS scoring is a long-standing, subjective method of measuring voice quality.

Table 1: Comparison of speech coding standards (without IP/UDP/RTP overhead)

Standard	Coding Type	Bit Rate (kbps)	MOS-LQO (Mean Opinion Score - Listening Quality Objective)
G.711	PCM	64	4.37
G.729	CS-ACELP	8	3.94
G.723.1	ACELPMP-MLQ	6.3 5.3	3.78 3.68
G.726	ADPCM	32	4.30

[1] As predicted. Measured according to ITU-IT Recommendation P.862 (PESQ). See draft Recommendation P.862.2, application guide for PESQ.
[2] Given MOS-LQO values for American English.

Because it does not use compression, G.711 offers the highest level of voice quality assuming proper operation of the IP network. Unfortunately, there is a trade-off with higher bandwidth usage. In situations where bandwidth is scarce, such as across WAN links, G.729 offers a good compromise with lower bandwidth usage, but still good fidelity audio.

In general, compression codecs use twice as many signal processing resources than the G.711 codec. On the TN2302AP IP Media Processor circuit pack (and on the G700 VoIP engine)

there are 64 DSP resources. Thus, the number of calls supported by one Media Processor circuit pack or G700 Branch Gateway is:

- 64 G.711 calls
- 32 compressed calls (for example, G.729)
- Some number in-between for a call mix.

The formula for calculating the number of calls one Media Processor board supports is

(Number of uncompressed calls) + 2 × (Number of compressed calls) ≤ 64

The TN2602AP circuit pack supports:

- 320 channels of G.711 (u/a-law)
- 320 channels of G.729A/G.729AB
- 320 channels of G.726 (32 kbps only)
- 320 channels of T.38
- 320 channels of V.32 SPRT

The above channel counts are the same if Advanced Encryption Standard (AES) encryption and SHA-1 authentication are enabled.

The One-X Deskphones (96xx) support the G.722 codec with 64 kbps and with 20 ms packets

Generally, G.711 is used on LANs because bandwidth is abundant and inexpensive whereas G.729 is used across bandwidth-limited WAN links. G430 and G450 Branch Gateways can have varying amounts of DSP resources depending on the size and number or DAR daughter cards installed. These resources behave like the TN2602 IP Media Resource 320 circuit pack.

G.726 Codec and branch gateways

Media processing resources on branch gateways support the G.726 codec. Table 2: Number of simultaneous bidirectional connections supported on page 66 provides the corresponding capacities: G430 (20 to 80 connections supported); G450 (80 to 320 connections supported)

Table 2: Number of simultaneous bidirectional connections supported

Codec	G430	G450
G.726A Unencrypted	10	16
G.726A with Avaya Encryption Algorithm (AEA) encryption	10	16
G.726A with Advanced Encryption Standard (AES) encryption	10	12

Silence suppression/VAD

Voice Activity Detection (VAD) or silence suppression can also be used to save bandwidth. During a conversation, because only one party is speaking at any given time, more than 40% of the transmission is silence. Voice Activity Detection (VAD) in Avaya IP telephones monitor the locally produced voice signal for voice activity. When no activity is detected for the configured period of time, packets are no longer transmitted. This prevents the encoder output from being transported across the network when there is silence, resulting in bandwidth savings.

When silence suppression is enabled, the remote end is instructed to generate “comfort noise” when no voice is present to make the call sound more natural to users. The trade-off with silence suppression lies with the silence detection algorithm. If it is too aggressive, the beginnings and ends of words can be “clipped.” If not aggressive enough, no bandwidth is saved.

Silence suppression is built into G.729B. It can be enabled for other codecs from within Communication Manager. Because of voice quality concerns with respect to clipping, silence suppression is generally disabled (with the exception of G.729B).

The following Avaya products employ silence suppression to preserve bandwidth:

- Avaya Communication Manager software (for control)
- All Avaya IP Telephones
- Avaya IP SoftPhone
- Avaya Media Gateways

For procedures to administer QoS parameters, refer to *Administration for Network Connectivity for Avaya Communication Manager* (555-233-504).

Transcoding/tandeming

Transcoding or tandeming occurs when a voice signal passes through multiple codecs. This can be the case when call coverage is applied on a branch office system back to a centralized voice mail system. These calls might experience multiple transcodings including, for example, G.729 across the WAN and G.711 into the voice mailbox. Each transcoding action results in degradation of voice quality. Avaya products minimize transcoding using methods such as shuffling and hairpinning.

Voice Quality in a path with more than two transcodings is considered objectionable to most people.

IP Telephony network engineering overview

In the early days of local area networking, network designers used hubs to attach servers and workstations, and routers to segment the network into manageable pieces. Because of the high cost of router interfaces and the inherent limitations of shared-media hubs, network design was fairly simple. In recent years, with the rise of switches to segment networks, designers were able to hide certain faults in their networks and still get good performance. As a result, network design was often less than optimal. IP Telephony places new demands on the network. Suboptimal design cannot cope with these demands. Even with switches installed, a company must follow industry best practices to have a properly functioning voice network. Because most users do not tolerate poor voice quality, administrators should implement a well-designed network before they begin IP Telephony pilot programs or deployments.

Related topics:

Network engineering overview on page 68
Voice quality on page 70
Best practices on page 72
Common issues on page 72
LAN issues on page 74
WAN on page 77
Virtual private network on page 82
Network Address Translation on page 85
Converged network design on page 86

Network engineering overview

Industry best practices dictate that a network be designed with consideration of the following factors:

- Reliability and redundancy
- Scalability
- Manageability
- Bandwidth

Voice mandates consideration of the following additional factors when designing a network:

- Delay
- Jitter

- Loss
- Duplex

In general, these concerns dictate a hierarchical network that consists of at most three layers (Table 3: Layers in a hierarchical network on page 69):

- Core
- Distribution
- Access

Some networks can collapse the functions of several layers into one device.

Table 3: Layers in a hierarchical network

Layer	Description
Core	The core layer is the heart of the network. The purpose of the core layer is to forward packets as quickly as possible. The core layer must be designed with high availability in mind. Generally, these high-availability features include redundant devices, redundant power supplies, redundant processors, and redundant links. Today, core interconnections increasingly use 10 Gigabit Ethernet or higher.
Distribution	The distribution layer links the access layer with the core. The distribution layer is where policy like the QoS feature and access lists are applied. Generally, Gigabit Ethernet connects to the core, and either Gigabit Ethernet or 100base-TX/FX links connect the access layer. Redundancy is important at this layer, but not as important as in the core. This layer is combined with the core in smaller networks.
Access	The access layer connects servers and workstations. Switches at this layer are smaller, usually 24 to 48 ports. Desktop computers, workstations, access points, and servers are usually connected at 100 Mbps or 1 Gbps. Limited redundancy is used. Some QoS and security features can be implemented in the access layer. Mostly, Power over Ethernet (PoE) is included to power IP telephones and other access devices.

For IP Telephony to work well, WAN links must be properly sized with sufficient bandwidth for voice and data traffic. Each voice call uses 9.6 kbps to 120 kbps, depending on the desired codec, payload size, and header compression used. Additional bandwidth might be used if video or redundancy for Fax, Modem, and TTY is implemented. The addition of video can stress or overrun WAN links engineered for voice only. WAN links must be re-engineered when video is introduced to an existing network. The G.729 compression algorithm, which uses about 27 kbps of bandwidth, is one of the most used standards today. Traditional telephone metrics, such as average call volume, peak volume, and average call length, can be used to size interoffice bandwidth demands. See Introduction to traffic engineering on page 199 for more information.

Quality of Service (QoS) also becomes increasingly important with WAN circuits. In this case, QoS means the classification and the prioritization of real-time traffic such as voice, video, or FoIP. Real-time traffic must be given absolute priority through the WAN. If links are not properly

sized or queuing strategies are not properly implemented, the quality and the timeliness of voice and data traffic will be less than optimal.

The following WAN technologies are commonly used with IP Telephony:

- MPLS (Multiprotocol Label Switching)
- ATM (Asynchronous Transfer Mode)
- Frame Relay
- Point-to-point (PPP) circuits
- Internet VPNs

The first four technologies all have good throughput, low latency, and low jitter. MPLS and ATM have the added benefit of enhanced QoS. MPLS is a relatively new service offering and can have issues with momentary outages of 1 to 50 s duration.

Frame Relay WAN circuits can be difficult to use with IP Telephony. Congestion in Frame Relay networks can cause frame loss, which can significantly degrade the quality of IP Telephony conversations. With Frame Relay, proper sizing of the committed information rate (CIR) is critical. In a Frame Relay network, any traffic that exceeds the CIR is marked as discard eligible, and is discarded at the option of the carrier if it experiences congestion in its network. Because voice packets and other real-time packets must not be dropped during periods of congestion, CIR must be sized to maximum traffic usage. Also, Service Level Agreements (SLAs) must be established with the carrier to define maximum levels of delay and frame loss, and remediation if the agreed-to levels are not met.

Internet VPNs are economical but more prone to quality issues than the other four technologies because there is no control or SLA to modify the handling of voice packets over data packets.

Network Management is another important area to consider when implementing IP Telephony. Because of the requirements imposed by IP Telephony, it is critical to have an end-to-end view of the network, and ways to implement QoS policies globally. Products such as HP OpenView Network Node Manager, Prognosis, Concord NetHealth, and MRTG help administrators maintain acceptable service. Outsource companies are also available to assist other companies that do not have the resources to implement and maintain Network Management.

Voice quality

Voice quality is always a subjective topic. Defining *good* voice quality varies with business needs, cultural differences, customer expectations, and hardware and software used. The requirements set forth are based on the ITU-T and EIA/TIA guidelines and extensive testing. Avaya requirements meet or exceed most customer expectations. However, the final determination of acceptable voice quality lies with the customer’s definition of quality, as well as the design, implementation, and monitoring of the end-to-end data network.

Quality is not one discrete value where the low side is good and the high side is bad. A trade-off exists between real-world limits and acceptable voice quality. Lower delay, jitter, and packet loss values can produce the best voice quality, but also can come with a cost to upgrade the network infrastructure to get to the low values. Another real-world limit is the inherent WAN delay. An IP trunk that links the west coast of the United States to India could add a fixed delay of 150 ms into the overall delay budget.

Perfectly acceptable voice quality is attainable, but will not be toll quality. Therefore, Avaya presents a tiered choice of elements that make up the requirements.

The critical objective factors in assessing IP Telephony quality are delay, jitter, and packet loss. To ensure good and consistent levels of voice quality, Table 4: Factors that affect voice quality on page 71 lists Avaya's suggested network requirements. These requirements are true for both LAN only and for LAN and WAN connections. Note that all measurement values are between endpoints because this document assumes that IP Telephony is not yet implemented. All values therefore reflect the performance of the network without endpoint consideration.

Table 4: Factors that affect voice quality

Network factor	Measurement
Delay (one-way between endpoints)	• A delay of 80 ms or less can, but may not, yield the best quality. • A delay of 80 ms to 180 ms can yield business-communication quality. Business-communication quality is much better than cell-phone quality, and is well-suited for the majority of businesses. Also, business-communication quality is defined as less than toll quality, but much better than cell-phone quality. • Delays that exceed 180 ms might still be quite acceptable depending on customer expectations, analog trunks used, codec type, and so on.
Jitter (variability of the delay between endpoints)	20 ms, or less than half the sample size, for the best quality. **Note:** This value has some latitude, depending on the type of service that the jitter buffer has in relationship to other router buffers, the packet size used, and so on.
Packet loss (maximum packet/frame loss between endpoints)	• < 1% can yield the best quality, depending on many factors. • < 3% should give business-communications quality, which is much better than cell-phone quality. • > 3% might be acceptable for voice, but might interfere with signaling.

For more information see Voice quality network requirements on page 59.

Best practices

To consistently ensure the highest quality voice, Avaya highly recommends consideration of the following industry best practices when implementing IP Telephony. Note that these suggestions are options, and might not fit individual business needs in all cases.

- QoS/CoS

  QoS for real-time packets is obtained only after a Class of Service (CoS) mechanism tags voice packets as having priority over data packets. Networks with periods of congestion can still provide excellent voice quality when using a QoS/CoS policy. The recommendation for switched networks is to use IEEE 802.1p/Q. The recommendation for routed networks is to use DiffServ Code Points (DSCP). The recommendation for mixed networks is to use both. Port priority can also be used to enhance DiffServ and IEEE 802.1p/Q. Even networks with plentiful bandwidth should implement CoS/QoS to protect voice communications from periods of unusual congestion, such as a computer virus might cause. See Implementing Communication Manager on a data network for more information.
- Switched network

  A fully switched LAN network is a network that allows full duplex and full endpoint bandwidth for every endpoint that exists on that LAN. Although IP Telephony systems can work in a shared or hub-based LAN, Avaya recommends the consistently high results that a switched network lends to IP Telephony.
- Network assessment

  A Basic Network Readiness Assessment Offer from Avaya is vital to a successful implementation of IP Telephony products and solutions. Contact an Avaya representative or authorized dealer to review or certify your network.
- VLANs

  Placing voice packets on a separate VLAN or subnetwork from data packets is a generally accepted practice to reduce broadcast traffic and to reduce contention for the same bandwidth as voice. Note that Avaya IP Telephones provide excellent broadcast storm protection. Other benefits become available when using VLANs, but there can be a substantial cost with initial administration and maintenance.

Common issues

Some common negative practices that can severely impact network performance, especially when using IP Telephony, include:

- A flat, non-hierarchical network

  For example, cascading small workgroup switches together in a flat non-hierarchical network. This technique quickly results in bottlenecks, because all traffic must flow across

the uplinks at a maximum of 10 Gbps, versus traversing switch fabric at speeds of 256 Gbps or greater. The greater the number of small switches or layers, the greater the number of uplinks, and the lower the bandwidth for an individual connection. Under a network of this type, voice performance can quickly degrade to an unacceptable level.

- Multiple subnets on a VLAN

  A network of this type can have issues with broadcasts, multicasts, and routing protocol updates. This practice can have a significant negative impact on voice performance, and complicate troubleshooting.

- A hub-based network

  Avaya strongly recommends all hubs be replaced with switches if they will lie in the path of IP telephony. Hubs are half-duplex by definition and may degrade the performance of real-time communications over IP.

- Too many access lists

  Access lists slow down a router. While access lists are appropriate for voice networks, care must be taken not to apply them to unnecessary interfaces. Traffic should be modeled beforehand, and access lists applied only to the appropriate interface in the appropriate direction, not to all interfaces in all directions.

Avaya recommends caution when using the following:

- Network Address Translation (NAT)

  IP Telephony may not work across Network Address Translation (NAT), because if private IP addresses are exchanged in signaling messages, these addresses are not reachable from the public side of the NAT and cannot be used for the media sessions.

- Analog dial-up

  Be careful in using analog dial-up (56 kbps) to connect two locations. Upstream bandwidth may be limited to a maximum of 33.6 kbps, and in most cases is less. This results in insufficient bandwidth to provide quality voice. Some codecs and network parameters provide connections that are acceptable, but consider each connection individually.

- Virtual Private Network (VPN)

  Large delays are inherent in some VPN software products due to encryption, decryption, and additional encapsulation. Some hardware-based products that encrypt at near wire speed can be used. In addition, if the VPN is run over the Internet, sufficient quality for voice cannot be guaranteed unless delay, jitter, and packet loss are contained within the parameters that are listed above.

LAN issues

This section covers Local Area Network (LAN) issues, including speed and duplex, inline power, hubs versus switches, and so on.

General guidelines

Because of the time-sensitive nature of IP telephony applications, IP telephony should be implemented on an entirely switched network. Ethernet collisions, which are a major contributor to delay and jitter, are virtually eliminated on switched networks. Additionally, the C-LAN, Media Processor circuit, and IP telephones should be placed on a separate subnetwork or VLAN (that is, separated from other non-IP telephony hosts). This separation provides for a cleaner design where IP telephony hosts are not subjected to broadcasts from other hosts, and where troubleshooting is simplified. This separation also provides a routed boundary between the IP telephony segments and the rest of the enterprise network, where restrictions can be placed to prevent unwanted traffic from crossing the boundary. When personal computers are attached to IP telephones, the uplink to the Ethernet switch should be a 100 Mbps link or greater, so that there is more bandwidth to be shared between the telephone and the computer.

Sometimes enterprises are unable to follow these guidelines, and Avaya's solutions can be made to work in some less-than-ideal circumstances. If IP telephones will share a subnetwork with other hosts, the IP telephones should be placed on a subnetwork of manageable size (24-bit subnet mask or larger, with 254 hosts or less), with as low a rate of broadcasts as possible. If the broadcast level is high, 100 Mbps links are less likely to be overwhelmed by broadcast traffic then 10 Mbps links. Perhaps a worst-case example is the scenario where IP telephones are deployed on a large subnetwork that is running IPX or other broadcast-intensive protocol, with broadcasts approaching 500 per second. Although the performance of the IP telephones and the voice quality can be satisfactory in this environment, this type of deployment is strongly discouraged.

Ethernet switches

The following recommendations apply to Ethernet switches to optimize operation with Avaya endpoints. These recommendations are meant to provide the simplest configuration by removing unnecessary features.

- Enable spanning tree fast start feature or disable spanning tree at the port level. The Spanning Tree Protocol (STP) is a Layer 2 loop-avoidance protocol. When a device is first connected (or reconnected) to a port that is running spanning tree, the port takes 31 – 50 s to cycle through the Blocking, Listening, and Learning states. This delay is neither necessary nor desired on ports that are connected to IP endpoints. Instead, enable a fast start feature on these ports to put them into the Forwarding state almost immediately. If this feature is not available, disabling spanning tree on the port is an option that should be considered. Do not disable spanning tree on an entire switch or VLAN. Also, Rapid Spanning Tree Protocol (802.1w) is always preferred over STP (802.1D). When using

RSTP, the Ethernet switch ports connected to IP phones must be in the *Edge-Type* mode. This places the port in a fast-start mode. BPDU guard is also desirable if it is available on the Ethernet switch to protect against a loop created through the IP phone.

- Disable the vendor features that are not required. Some vendor features that are not required by Avaya endpoints include EtherChannel/LAG, cdp, and proprietary (not 802.3af) inline power. These features are nonstandard mechanisms that are relevant only to vendor specific devices, and can sometimes interfere with Avaya devices.
- Properly configure 802.1Q trunking on Cisco switches. When trunking is required on a Cisco CatOS switch that is connected to an Avaya IP telephone, enable it for 802.1Q encapsulation in the no-negotiate mode. This causes the port to become a plain 802.1Q trunk port with no Cisco autonegotiation features. When trunking is not required, explicitly disable it.

Speed and duplex

One major issue with Ethernet connectivity is proper configuration of the speed and duplex settings. It is important that the speed and duplex settings are configured properly and that they match.

A duplex mismatch condition results in a state where one side perceives a high number of collisions, while the other side does not. This results in packet loss. Although it degrades performance in all cases, this level of packet loss might go unnoticed in a data network because protocols such as TCP retransmit lost packets. In voice networks, however, this level of packet loss is unacceptable. Voice quality rapidly degrades in one direction. When voice quality problems are experienced, duplex mismatches are the first thing to investigate.

Best practice is to use autonegotiation on both sides of an IP connection. You can also lock down interfaces on both sides of a link. However, many a times, this practice requires a coordination between the Ethernet switch data team and the voice team. Gigabit links should ALWAYS use Auto-Negotiation. Details of all aspects of Auto-Negotiation and lockdown are found in this whitepaper: http://support.avaya.com/css/P8/documents/100121639

Whether you choose the autonegotiation mode or the lock down mode, make sure that both the ends of the link use the same mode which results in 100 Mbps and full duplex for 10/100 Mbps links and that Gigabit links results in 1 Gbps and full duplex in Auto-Neg mode.

Caution:

Do not use the auto negotiation mode on one side of the IP connection and the lock down mode on the other as it can result in a duplex mismatch and cause voice quality and/or signaling problems.

Virtual LANs

Virtual Local Area Networks (VLANs) are an often-misunderstood concept. This section begins by defining VLANs, and then addresses configurations that require the Avaya IP telephone to connect to an Ethernet switch port that is configured for multiple VLANs. The IP telephone is

on one VLAN, and a personal computer that is connected to the telephone is on a separate VLAN. Two sets of configurations are given: Cisco CatOS, and some Cisco IOS.

VLAN defined

With simple Ethernet switches, the entire switch is one Layer 2 broadcast domain that usually equates to one IP subnetwork (Layer 3 broadcast domain). Think of a single VLAN (on a VLAN-capable Ethernet switch) as being equivalent to a simple Ethernet switch. A VLAN is a logical Layer 2 broadcast domain that typically equates to one IP subnetwork. Therefore, multiple VLANs are same as logically separated subnetworks. This arrangement is analogous to multiple switches being physically separated subnetworks. A Layer 3 routing process is required to route between VLANs, just as one is required to route between subnetworks. This routing process can take place on a connected router or a router module within a Layer 2/Layer 3 Ethernet switch. If no routing process is associated with a VLAN, devices on that VLAN can only communicate with other devices on the same VLAN.

Port or native VLAN

Port VLAN and native VLAN are synonymous terms. The IEEE 802.1Q standard and most vendor switches use the term *port VLAN*, but Cisco switches use the term *native VLAN*.

Every port has a port VLAN or a native VLAN. Unless otherwise configured, it is VLAN 1 by default. It can be configured on a per-port basis or over a range of ports.

All untagged Ethernet frames (with no 802.1Q tag, for example, from a personal computer) are forwarded on the port VLAN or the native VLAN. This is true even if the Ethernet switch port is configured as an 802.1Q trunk, or otherwise configured for multiple VLANs.

Trunk configuration

A trunk port on an Ethernet switch is one that is capable of forwarding Ethernet frames on multiple VLANs through the mechanism of VLAN tagging. IEEE 802.1Q specifies the standard method for VLAN tagging. Cisco also uses a proprietary method called ISL. However, Cisco discourages the use of ISL. Avaya products do not interoperate with ISL.

A trunk link is a connection between two devices across trunk ports. This connection can be between a router and a switch, between two switches, or between a switch and an IP telephone. Some form of trunking or forwarding multiple VLANs must be enabled to permit the IP telephone and the attached personal computer to appear on separate VLANs. The commands in Table 5: Administration commands for VLAN trunking on page 76 enable trunking.

Table 5: Administration commands for VLAN trunking

Cisco IOS	Cisco CatOS
`switchport mode trunk` By default, all VLANs (1 to 4094) are enabled on the trunk port. Switches supporting ISL trunking have different commands for trunk setup. For more information, see the IOS manual.	`set trunk <mod/port> nonegotiate dot1q` By default, all VLANs (1 to1005) are enabled on the trunk port. VLANs can be selectively removed with the command `clear trunk <mod/port> <vid>`

Note that Cisco and other vendor switches can remove VLANs from a trunk port. This is a highly desirable feature because only two VLANs at most should appear on a trunk port that is connected to an IP telephone. That is, broadcasts from nonessential VLANs should not be permitted to bog down the link to the IP telephone. Cisco IOS switches can have an implicit trunk that contains only two VLANs, one for data and one for voice. You can configure an implicit trunk using the following commands:

- `switchport access vlan <vlan-id>`
- `switchport voice vlan <vlan-id>`

Trunking for one-X Communicator and other softphones

It is possible to set the Layer 2 priority on a softphone (or physical phone) using IEEE-802.1p bits in the IEEE-802.1Q VLAN tag. This is useful if the telephone and the attached personal computer are on the same VLAN (same IP subnetwork), but the telephone traffic requires higher priority (Figure 1: Trunking for softphones or physical phones on page 77). Enable 802.1Q tagging on the IP phone, set the priorities as desired, and set the VID to zero. As per the IEEE standard, a VID of zero assigns the Ethernet frame to the port VLAN or the native VLAN.

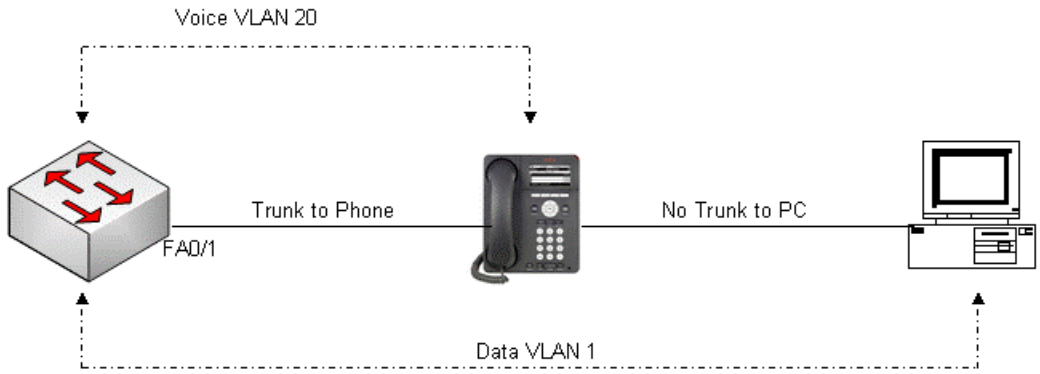


Figure 1: Trunking for softphones or physical phones

Cisco switches behave differently in this scenario, depending on the hardware platforms and OS versions.

Note:

Setting a Layer 2 priority is useful only if QoS is enabled on the Ethernet switch. Otherwise, the priority-tagged frames are treated no differently than clear frames.

WAN

Because of the high costs and lower bandwidths available, there are some fundamental differences in running IP telephony over a Wide Area Network (WAN) versus a LAN. As more problems occur in WAN environment, so it is important to consider network optimizations and proper network design.

Related topics:

WAN QoS on page 78
Frame Relay on page 80
Multiprotocol Label Switching on page 81

WAN QoS

In particular, Quality of Service (QoS) becomes more important in a WAN environment than in a LAN. In many cases, transitioning from the LAN to the WAN reduces bandwidth by approximately 99%. Because of this severe bandwidth crunch, strong queuing, buffering, and packet loss management techniques have been developed. These are covered in more detail in Quality of Service guidelines on page 96.

Recommendations for QoS

For many customers, including small and medium, simplicity is more effective than complex configurations when implementing QoS for voice, data, signaling and video. If traffic engineering is done properly and sufficient bandwidth is available, especially for WAN links, voice and voice signaling traffic can both be tagged as DSCP 46. This Class of Service (CoS) tagging will place both types of packets into the same High Priority queue with a minimum of effort. The key is to have enough bandwidth to prevent any packets from dropping.

For large enterprises and Multi-National companies, a stratified approach to CoS makes sense. This allows maximum control for many data and voice services. For this environment, Avaya recommends using DSCP 46 (Expedited Forwarding) for voice (bearer), but voice signaling and especially IPSI signaling could have their own DSCP values and dedicated bandwidth. This would prevent traffic, like voice bearer from contending with signaling. Although the configuration may be more complex to manage and administer, the granularity will give the best results and is recommended as a best practice.

At the routers, Avaya recommends using strict priority queuing for voice packets, and weighted-fair queuing for data packets. Voice packets should always get priority over non-network-control data packets. This type of queuing is called Class-Based Queuing (CBQ) on Avaya data networking products, or Low-Latency Queuing (LLQ) on Cisco routers.

Codec selection and compression

Because of the limited bandwidth that is available on the WAN, using a compression codec allows much more efficient use of resources without a significant decrease in voice quality. Avaya recommends that IP telephony implementations across a WAN use the G.729 codec with 20 ms packets. This configuration uses 24 kbps (excluding Layer 2 overhead), 30% of the bandwidth of the G.711 uncompressed codec (80 kbps).

To conserve even more bandwidth, RTP header compression (cRTP) can be used on point-to-point links. cRTP reduces the IP/UDP/RTP overhead from 40 bytes to 4 bytes. With 20 ms packets, this translates to a savings of 14.4 kbps, making the total bandwidth required for G.729 approximately 9.6 kbps. The trade-off for cRTP is higher CPU utilization on the router. The processing power of the router determines the amount of compressed RTP traffic that the router can handle. Avaya testing indicates that a typical small branch-office router can handle 768 kbps of compressed traffic. Larger routers can handle greater amounts. cRTP is available

on several Avaya secure routers (1000–series, 2330, 3120, and 4134) and on the Extreme, Juniper, Cisco, and other vendor routers.

Serialization delay

Serialization delay refers to the delay that is associated with sending bits across a physical medium. Serialization delay is important to IP telephony because this delay can add significant jitter to voice packets, and thus impair voice quality.

Network design

Routing protocols and convergence

When designing an IP telephony network across a WAN, some care should be taken when selecting a routing protocol or a dial-backup solution. Different routing protocols have different convergence times, which is the time that it takes to detect a failure and route around it. While a network is in the process of converging, all voice traffic is lost.

The selection of a routing protocol depends on several factors:

- If a network has a single path to other networks, static routes are sufficient.
- If multiple paths exist, is convergence time an issue? If so, EIGRP and OSPF are appropriate.
- Are open standards-based protocols required? If so, OSPF and RIP are appropriate, but not EIGRP or IGRP, which are Cisco proprietary.

In general, Avaya recommends the use of OSPF when routing protocols are required. OSPF allows for relatively fast convergence, and does not rely on proprietary networking equipment.

In many organizations, because of the expense of dedicated WAN circuits, dial-on-demand circuits are provisioned as backup if the primary link fails. The two principal technologies are ISDN (BRI) and analog modem. ISDN dial-up takes approximately 2 s to connect, and offers 64 kbps to 128 kbps of bandwidth. Analog modems take approximately 60 s to connect, and offer up to 56 kbps of bandwidth. If G.729 is used as the codec, either technology can support IP telephony traffic. If G.711 is used as the codec, only ISDN is appropriate. Also, because of the difference in connect times, ISDN is the preferred dial-on-demand technology for implementing IP telephony.

Multipath routing

Many routing protocols, such as OSPF, install multiple routes for a particular destination into a routing table. Many routers attempt to load-balance across the two paths. There are two methods for load balancing across multiple paths. The first method is per-packet load balancing, where each packet is serviced round-robin fashion across the two links. The second method is per-flow load balancing, where all packets in an identified flow (source and destination addresses and ports) take the same path. IP telephony does not operate well over per-packet load-balanced paths. This type of setup often leads to “choppy” quality voice. Avaya recommends that in situations with multiple active paths, per-flow load balancing is preferable to per-packet load balancing.

Balancing loads per-flow

In the presence of multiple links, data can be balanced across them by a round-robin fashion for either packets or a stream (flow) of data. Real-time media like voice and video should use flow balancing only.

Frame Relay

The nature of Frame Relay poses somewhat of a challenge for IP telephony, as described in this section.

Overview of frame relay

Frame Relay service is composed of three elements: the physical access circuit, the Frame, Relay port, and the virtual circuit. The physical access circuit is usually a T1 or fractional T1 and is provided by the local exchange carrier (LEC) between the customer premise and the nearest central office (CO). The Frame Relay port is the physical access into the Frame Relay network, a port on the Frame Relay switch itself.

The access circuit rate and the Frame Relay port rate must match to avoid discarded packets during the periods of congestion. The virtual circuit is a logical connection between Frame Relay ports that can be provided by the LEC for intra-lata Frame Relay, or by the inter-exchange carrier (IXC) for inter-lata Frame Relay. The most common virtual circuit is a permanent virtual circuit (PVC), which is associated with a committed information rate (CIR). The PVC is identified at each end by a separate data-link connection identifier (DLCI) in Figure 2: Data-link connection identifiers over an interexchange carrier Frame Relay network on page 80.

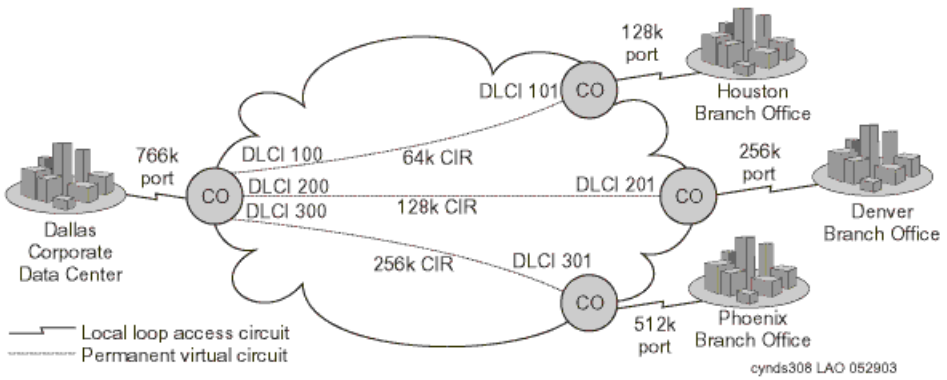


Figure 2: Data-link connection identifiers over an interexchange carrier Frame Relay network

This hypothetical implementation shows the Dallas corporate office connected to three branch offices in a common star topology (or hub and spoke). Each office connects to a LEC CO over a fractional T1 circuit, which terminates onto a Frame Relay port at the CO, and onto a Frame Relay capable router at the customer premise. The port rates and the access circuit rates match. PVCs are provisioned within the Frame Relay network between Dallas and each branch office. The CIR of each PVC is sized so that it is half the respective port rate, which is a common implementation. Each branch office is guaranteed its respective CIR, but it is also allowed to burst up to the port rate without any guarantees.

The port rate at Dallas is not quite double the aggregate CIR, but it does not need to be, because the expectation is that not all three branch offices will burst up to the maximum at the same time. In an implementation like this, the service is probably negotiated through a single vendor. But it is likely that Dallas and Houston are serviced by the same LEC, and that the Frame Relay is intra-lata, even if it was negotiated through an IXC, such as AT&T, or Sprint.

The service between Dallas and the other two branch offices, however, is most likely inter-lata.

A frame relay issue and alternatives

The obstacle in running IP telephony over Frame Relay involves the treatment of traffic within the CIR and outside of CIR, commonly termed the burst range.

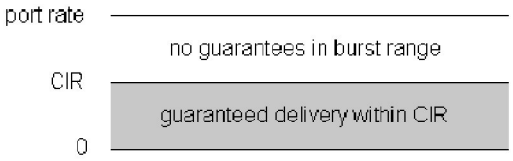


Figure 3: Committed information rate (burst range)

As Figure 3: Committed information rate (burst range) on page 81 shows, traffic up to the CIR is guaranteed, whereas traffic beyond the CIR usually is not. This is how Frame Relay is intended to work. CIR is a committed and reliable rate, whereas burst is a bonus when network conditions permit it without infringing upon the CIR of any user. For this reason, burst frames are marked as discard eligible (DE), and are queued or discarded when network congestion exists. Although experience has shown that customers can achieve significant burst throughput, it is unreliable and unpredictable, and not suitable for real-time applications like IP telephony.

Therefore, the objective is to prevent voice traffic from entering the burst range and being marked DE. One way to accomplish this is to prohibit bursting by shaping the traffic to the CIR and setting the excess burst size (B_e – determines the burst range) to zero. However, this also prevents data traffic from using the burst range.

Additional frame relay information

One interesting piece of knowledge is that most IXCs convert the long-haul delivery of Frame Relay into ATM. That is, the Frame Relay PVC is converted to an ATM PVC at the first Frame Relay switch after leaving the customer premise. It is not converted back to Frame Relay until the last Frame Relay switch before entering the customer premise. This is significant because ATM has built- in Class of Service (CoS). A customer can contract with a carrier to convert the Frame Relay PVC into a constant bit rate (CBR) ATM PVC. ATM CBR cells are delivered with lower latency and higher reliability.

Finally, under the best circumstances, Frame Relay is still inherently more susceptible to delay than ATM or TDM. Therefore, after applying the best possible queuing mechanism, one should still expect more delay over Frame Relay than is present over ATM or TDM.

Multiprotocol Label Switching

Multiprotocol Label Switching (MPLS) VPN service from service providers is commonly used by enterprises for WAN connectivity. The service is often available over different types of access links, and usually offers multiple classes of service. MPLS service is typically expected to provide good QoS and therefore to satisfy VoIP requirements, though this often depends on the Service Layer Agreement (SLA) and the actual quality delivered by the service provider.

With MPLS service, unlike private WAN, the enterprise controls QoS explicitly only on the access link —- that is, on the connection from each enterprise site to the MPLS network. Within the MPLS network QoS is controlled by the service provider. The enterprise affects the service given to its traffic by assigning the traffic to appropriate classes of service in the service provider's network. This is done with DiffServ Code Point (DSCP) marking in the packet's IP header. DSCP remarking by the enterprise edge routers may be required, mapping the DSCPs of enterprise traffic to the DSCP values designated by the MPLS service provider for the different classes of service in their service offering.

Virtual private network

Many definitions exist for Virtual Private Networks (VPNs). VPNs refer to encrypted tunnels that carry packetized data between remote sites. VPNs can use private lines, or use the Internet through one or more Internet Service Providers (ISPs). VPNs are implemented in both dedicated hardware and software, but can also be integrated as an application to existing hardware and software packages. A common example of an integrated package is a firewall product that can provide a barrier against unauthorized intrusion, as well as perform the security features that are needed for a VPN session.

The encryption process can take from less than 1 ms to 1 s or more, at each end. Obviously, VPNs can represent a significant source of delay, and therefore have a negative affect on voice performance. Also, because most VPN traffic runs over the Internet and there is little control over QoS parameters for traffic crossing the Internet, voice quality may suffer due to excessive packet loss, delay, and jitter. Users might be able to negotiate a service-level agreement with the VPN provider to guarantee an acceptable level of service. Before implementing IP telephony with a VPN, users should test their VPN network over time to ensure that it consistently meets the Avaya requirements.

Convergence advantages

For increasing numbers of enterprises, the VPN carries not only data, but voice communications. Though voice communication over IP networks (IP telephony) creates new quality of service (QoS) and other challenges for network managers, there are compelling reasons for moving forward with convergence over maintaining a traditional voice and data infrastructure:

- A converged infrastructure makes it easier to deploy eBusiness applications, such as customer care applications, that integrate voice, data, and video.
- Enterprises can reduce network costs by combining disparate network infrastructures, and eliminating duplicate facilities.
- A converged infrastructure can increase the efficiencies of the IT organization.
- Long distance charges can be reduced by sending voice over IP networks.

Voice over IP VPN is emerging as a viable way to achieve these advantages. The emergence of public and virtual private IP services promises to make it easier for customers, suppliers,

and businesses to use data networks to carry voice services. As with any powerful new technology, however, VPNs require skilled management to achieve top performance. The highest network performance becomes imperative when the VPN network must deliver high-quality voice communication. Not all IP networks can meet these quality requirements today. For instance, the public Internet is a transport option for voice communication only when reduced voice performance is acceptable, and global reach has the highest priority. When high voice quality is a requirement, ISPs and Network Service Providers (NSPs) can provide other VPN connections that meet required Service Level Agreements (SLAs).

Managing IP telephony VPN issues

This section provides information on communications security, firewall technologies, and Network Management as related to VPN issues.

Communication security

The public nature of the Internet, its reach, and its shared infrastructure provide cost savings when compared to leased lines and private network solutions. However, those factors also contribute to make Internet access a security risk. To reduce these risks, network administrators must use the appropriate security measures.

It is important to note that a managed service can be implemented either as a premises-based solution or a network-based VPN service. A premises-based solution includes customer premises equipment (CPE) that allows end-to-end security and Service Level Agreements (SLAs) that include the local loop. These end-to-end guarantees of quality are key differentiators. A network-based VPN, on the other hand, is provisioned mainly by equipment at the service provider's point-of-presence (PoP), so it does not provide equivalent guarantees over the last mile. For a secure VPN that delivers robust, end-to-end SLAs, an enterprise must demand a premises-based solution that is built on an integrated family of secure VPN platforms.

The "private" in virtual private networking is also a matter of separating and insulating the traffic of each customer so that other parties cannot compromise the confidentiality or the integrity of data. IPSec tunneling and data encryption achieves this insulation by essentially carving private end-to-end pipes or "tunnels" out of the public bandwidth of the Internet, and then encrypting the information within those tunnels to protect against someone else accessing the information. In addition to IPSec, there are two standards for establishing tunnels at Layer 2. These are the Point-to-Point Tunneling Protocol (PPTP) and Layer 2 Tunneling Protocol (L2TP), neither of which includes the encryption capabilities of IPSec. The value of IPSec beyond these solutions is that IPSec operates at IP Layer 3. It allows for native, end-to-end secure tunneling and, as an IP-layer service, it also promises to be more scalable than the connection-oriented Layer 2 mechanisms.

Also, note that IPSec can be used with either L2TP or PPTP, since IPSec encrypts the payload that contains the L2TP/PPTP data. Indeed, IPSec provides a highly robust architecture for secure wide-area VPN and remote dial-in services. It is fully complementary to any underlying Layer 2 network architecture, and with its addition of security services that can protect the VPN

of a company, IPSec marks the clear transition from early tunneling to full-fledged Internet VPN services.

An issue, however, is the fact that different implementations of IPSec confer varying degrees of security services. Products must be compliant with the latest IPSec drafts, must support high-performance encryption, and must scale to VPNs of industrial size.

Finally, a VPN platform should support a robust system for authentication of the identity of end users, based on industry standard approaches and protocols.

Firewall technologies

To reduce security risks, appropriate network access policies should be defined as part of business strategy. Firewalls can be used to enforce such policies. A firewall is a network interconnection element that polices traffic the flows between internal (protected) networks and external (public) networks such as the Internet. Firewalls can also be used to segment internal networks.

The application of firewall technologies only represents a portion of an overall security strategy. Firewall solutions do not guarantee 100% security by themselves. These technologies must be complemented with other security measures, such as user authentication and encryption, to achieve a complete solution.

The three technologies that are most commonly used in firewall products are packet filtering, proxy servers, and hybrid. These technologies operate at different levels of detail, and thus they provide varying degrees of network access protection. That means that these technologies are not mutually exclusive. A firewall product may implement several of these technologies simultaneously.

Network Management and outsourcing models

While enterprises acknowledge the critical role that the Internet and IP VPNs can play in their strategic eBusiness initiatives, they face a range of choices for implementing their VPNs. The options range from enterprise-based or do-it-yourself VPNs that are fully built, owned, and operated by the enterprise, to VPNs that are fully outsourced to a carrier or other partner. In the near term, it is generally believed that enterprise-operated and managed VPN services will hover around a 50/50 split, including hybrid approaches.

Increasingly, enterprises are assessing their VPN implementation options across a spectrum of enterprise-based, carrier-based/outsourced, or hybrid models. Each approach offers a unique business advantage.

- Enterprise based

  This option operates over a public network facility (most commonly the Internet) using equipment that is owned and operated by the enterprise. Its greatest benefit to the enterprise is the degree of flexibility and control it offers over VPN deployment, administration, and adaptability or change.

- Fully outsourced

This managed service could be implemented by a collection of partners, including an ISP and a security integration partner. Its advantages include quick deployment, easy global scalability, and freedom from overhead Network Management.

- Shared management

  With this hybrid approach, a partner can take responsibility for major elements of infrastructure deployment and management, but the enterprise retains control over key aspects of policy definition and security management.

Conclusion

Moving to a multipurpose packet-based VPN that transports both voice and data with high quality poses a number of significant management challenges. Managers must determine whether to operate the network using an enterprise-based model, an outsourced or carrier-based model, or a hybrid model. They must settle security issues that involve several layers of the network. And they must ensure that they and their vendors can achieve the required QoS levels across these complex networks. Yet the advantages of converged, multipurpose VPNs remain a strong attraction. The opportunity to eliminate separate, duplicate networks and costly dedicated facilities, avoid costly public network long distance charges, and reduce administrative overhead provides a powerful incentive. Most important, by helping integrate voice and data communication, multimedia Messaging, supplier and customer relationship management, corporate data stores, and other technologies and resources, converged networks promise to become a key enabler for eBusiness initiatives.

Network Address Translation

IP telephony may not work across Network Address Translation (NAT), because if private IP addresses (RFC-1918) are exchanged in signaling messages these addresses are not reachable from the public side of the NAT and cannot be used for the media sessions.

The problem is not encountered in all VoIP scenarios. It is avoided for VPN-based remote access, and NATs are usually not needed internally within the enterprise network. VoIP has to traverse NAT, usually at the border between the enterprise and a VoIP trunk to a service provider, as well as in hosted VoIP service.

If the network design includes a firewall within the enterprise network to protect certain servers or some part of the network, so that IP telephony traffic has to traverse the internal firewall, then it is preferable that the firewall not perform a NAT function. IP telephony will then work across the firewall once the appropriate ports are open on the firewall. For information on the list of needed ports for any Avaya product, contact the Avaya account team or Business Partner.

When connecting the enterprise to an IPT SP through a VoIP trunk — either SIP or H.323 — there will likely be a NAT at the enterprise border. The recommended solution for this scenario is to deploy a Session Border Controller (SBC) near the NAT (for example, in the enterprise DMZ). SBCs from multiple vendors have been tested for interoperability with Avaya’s IP

telephony solutions. For a list of Avaya Developer *Connection* members and Avaya compliance-tested solutions, refer to the solutions directory at: http://www.avaya.com/gcm/master-usa/en-us/corporate/alliances/devconnect/index.htm.

Alternatively, in certain cases an Application Layer Gateway (ALG) can be used. Another alternative is setting up a C-LAN and a Media Processor card in the DMZ, and using Communication Manager as a proxy server between the internal and external networks.

Solutions based on standards such as ICE and STUN are expected to be supported in some NAT traversal scenarios where applicable.

Figure 4: IP telephony without NAT on page 86 shows IP telephony without NAT scenario.

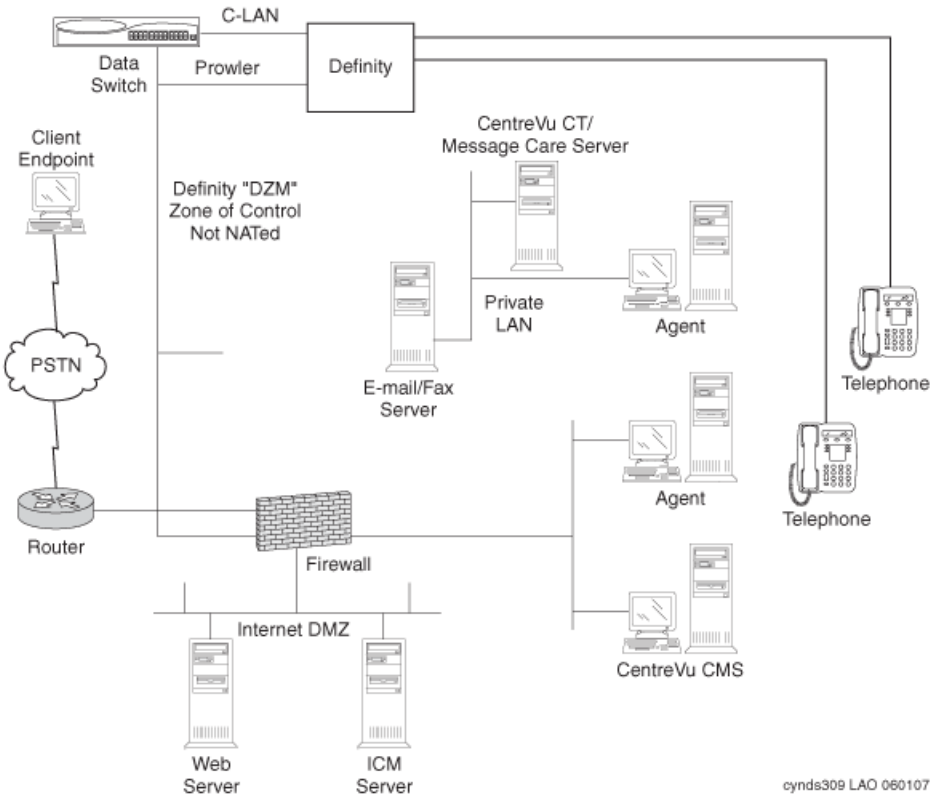


Figure 4: IP telephony without NAT

Converged network design

Converged networks require the application of good management and control practices to support and sustain the deployment of IP telephony. The first step in implementing an IP telephony system is making the commitment to provide a network capable of supporting a real time application such as voice. Many large enterprise networks averaged about 87 hours of downtime in 2009.

Design and management

Highly available networks do not just happen. They must be planned and maintained. The probability of success for both of those activities is improved by the application of three fundamental principles -- Simplicity, Manageability and Scalability.

According to an article published in Network World [1] 59% of network downtime was attributable to routing management with approximately 36% of that caused by configuration errors. In another instance a report published by Infonetics Research and reviewed in DMReview[2] attributed the single largest portion (32%) of downtime costs to Application problems with software failure (36%) and human error (22%) precipitating those outages over half of the time. In another article available through Infonetics Research, [3] the author cites human error as “the most troubling” cause of outages due to the time and cost of correcting the problems. Older studies[4] attribute as much as 80% of all mission-critical application service downtime to “people or process” failures.

It is clear from the available data that to deploy a business-critical IP-based service the network upon which it runs must be:

- Easy to configure.
- Easy to monitor and troubleshoot.
- Extensible with minimum reconfiguration. That is, designed with enough resources to grow with the business it supports.

Design for Simplicity

It is self evident that action without understanding is unpredictable. IT staff must interact with the network, so if the system is difficult to understand the probability of error increases. With this thought in mind it is advisable to reduce the number protocols and services on any network segment and reduce the number of decisions the network must make. Simple, documentable, reproducible and verifiable configurations are a must for IP telephony deployment. The IT staff responsible for the network needs to understand how it works, and new staff should be easy to train. A conscious choice to favor simplicity in design may be the single biggest factor in improving uptime due to its cascading effect on process, documentation and verification.

Design for Manageability

Studies of operator errors have identified several classes of errors typical of network service administrators. Most of these are the result of misconfiguration of new components and unintended actions such as restarts or disabling of hardware while diagnosing problems.

---

1 Pisello, Tom, and Quirk, Bill. 2004. “How to quantify downtime.” *Network World* Web. http://www.networkworld.com

2 DM Review 2/18/2004

3 Willson, Jeff. 2006. “Medium businesses lose $867,000 a year to network downtime” Press release, Infonetics Research. Web. http://www.infonetics.com

4 Scott, Donna. 2001. “NSM: Often the Weakest Link in Business Availability” *Gartner Research*. Web. http://www.gartner.com

Significantly, operators of all experience levels were found to introduce almost all classes of errors with roughly equal frequency.

Research conducted at Rutgers University [6] found that operator action-verification techniques allowed detection and prevention of over half of the errors typically introduced by operators. This data argues strongly for the implementation of reliable change control procedures, and change verification as requisites for highly available networks. To support these activities management tools must be in place to aid in detecting and reporting errors, both for validation of operator actions and diagnosing problems. Network documentation is typically inaccurate and outdated [7] (due in part to lack of change control) so management capabilities to verify configurations are essential.

Design for scalability

Other researchers have proposed mechanisms for reducing or eliminating the need for operator interaction by automating common tasks. The success of these approaches argues that reducing the scope of changes required to manage and expand network services will pay dividends in network uptime. Excess bandwidth, unused ports and available addresses are required to verify changes and to simplify network expansion. Expansion should begin well before these resources are exhausted.

Using designs that limit the impact of changes reduces the potential for errors. For example, if the administrator needs to change both an aggregation switch and a router configuration in order to accommodate new media gateway interfaces, the design has doubled the opportunity for configuration error. If a new subnet needs to be added to Access Control Lists throughout the network core the potential for outage is expanded further still.

The same principles used to reduce software complexity and improve software reliability are applicable to the network complexity problem. Modularity, design reuse, and testability are all attributes of highly reliable networks.

Topologies

The network topology most commonly recommended consists of a redundant core with building blocks of layered routers and switches as shown in the figure below. This is the defacto standard for network design supporting both modularity and reuse.

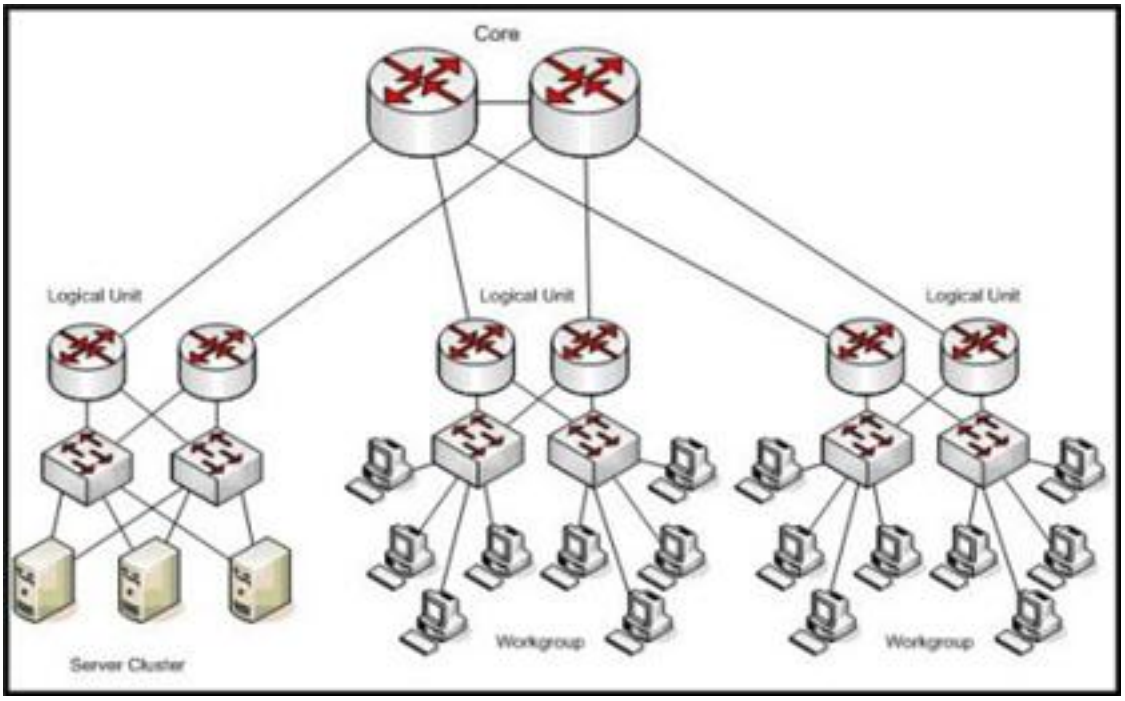


Figure 5: Typical Network Topology Design

Of course, real networks are far more complex with many more nodes and services. In addition, real deployments typically have legacy constraints, multiple sites, and heterogeneous equipment. It is beyond the scope of this document to detail solutions for the potential configurations of entire networks. To address those issues, Avaya provides a full range of service offers from assessment to outsourced management. Please visit the Avaya web site (http://support.avaya.com) for further information regarding these services.

Server Cluster

A review of the server cluster configuration as applied to a set of G650 IP-connected port networks will serve to illustrate the principles discussed and validate the modular topology.

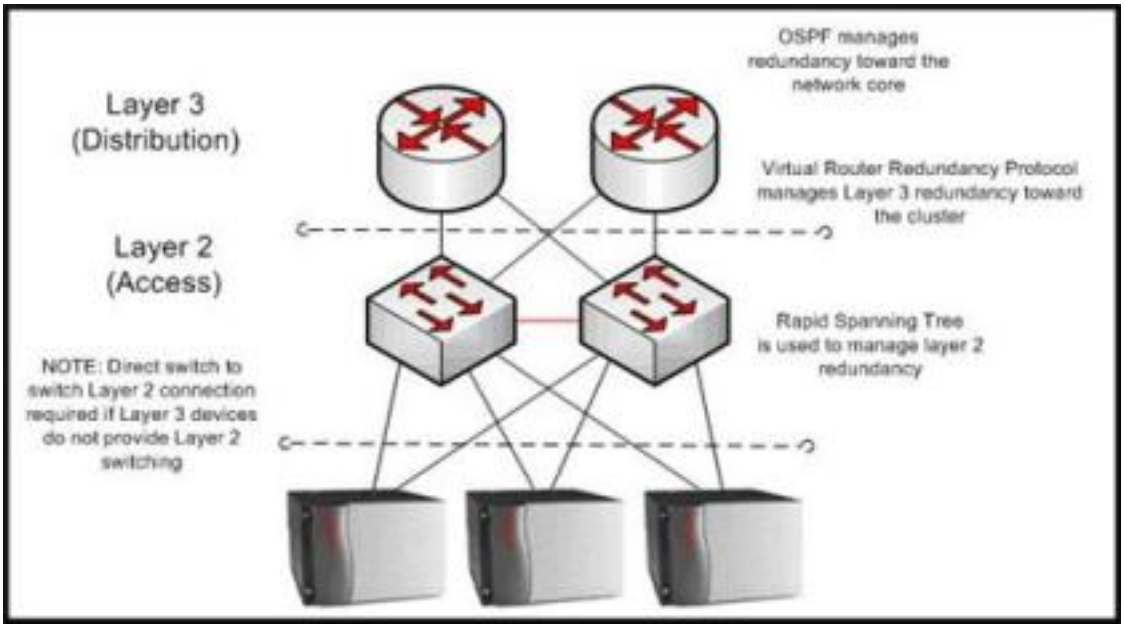


Figure 6: Layered Server Cluster Topology

Assume that each G650 is equipped with redundant TN2602AP Media Resource 320 circuit packs optionally configured for load balancing or IP bearer duplication. Each G650 is also assumed to contain duplicated TN2312BP IP Server Interfaces (IPSI). The number of TN799DP Control Lan (C-LAN) socket termination boards would be sized to accommodate the devices in the wider network and the call capacity of the cluster, but for this small configuration assume a C-LAN for each G650. Also assume the Layer 3 devices use hardware switching for Layer 3 forwarding and that they are also capable of Layer 2 switching between ports. It is important to remember that IP telephony LAN traffic consists primarily of smaller (approximately 218 octet) packets and it is the per packet overhead that impacts software based routing. Consider that telephony traffic is roughly an order of magnitude more packets per unit bandwidth than typical web page transfers.

Layers

The obvious question is why there is a separate Layer 2 access level if the devices at the Layer 3 distribution layer are capable of Layer 2 switching? In general, the access layer reduces the complexity of the network block by separating the functions of the devices, and provides scalability when more ports are required as the network grows. To ensure network modularity, the routers serving this cluster should be dedicated to the cluster and sized to the task. Simplification argues for the reduction of subnets and therefore routed interfaces in the cluster since the service is common. If remote IPSIs or multiple server clusters are implemented across the network, using a single subnet within the cluster simplifies the configuration of the entire network. The addition of static subnets in the direction of the cluster increases the configuration complexity with little benefit in terms of availability unless the subnets terminate on different routers, which in turn implies separate modular clusters. A separate management subnet would be expected but is unrelated to the service address configuration. An argument can be made that separate subnets simplify diagnostic activities, but this benefit is achievable

with address partitioning within the subnet. Port densities for smaller full featured routers may be inadequate to scale to the connectivity requirements of even this small cluster when the extra ports for management, troubleshooting, and testing are considered.

An alternative design uses the smaller high density integrated switching and routing platforms that are becoming popular as routing functions have moved into commodity ASICs.

When selecting this type of configuration, bandwidth and inter-switch traffic capacity must be considered. In a load balanced configuration under fault conditions, approximately ½ the call load may travel on the inter-switch link. The inter-switch link must be redundant to prevent a single failure from causing a bifurcated network and, if a Link Aggregation Group (LAG) is used to eliminate potential spanning tree loops, the individual link bandwidth must still be capable of supporting the required traffic.

Redundancy

Hardware redundancy is a proven and well defined tool for increasing the availability of a system. Avaya Critical Availability solutions have traditionally employed this technique to achieve 99.999% availability. One question to consider in the deployment of redundant hardware is symmetric (active-active) or asymmetric (active-standby) configurations. Well known reliability expert Evan Marcus recommends asymmetric configurations for pure availability. Avaya's control network and TDM bearer redundancy solutions follow that model. For IP-PNC designs, bearer duplication supports asymmetric redundancy for bearer flows but symmetric redundancy, or load balanced configurations, are the default. Because of the inherent complexity of TCP state replication, the C-LAN configurations are always symmetric.

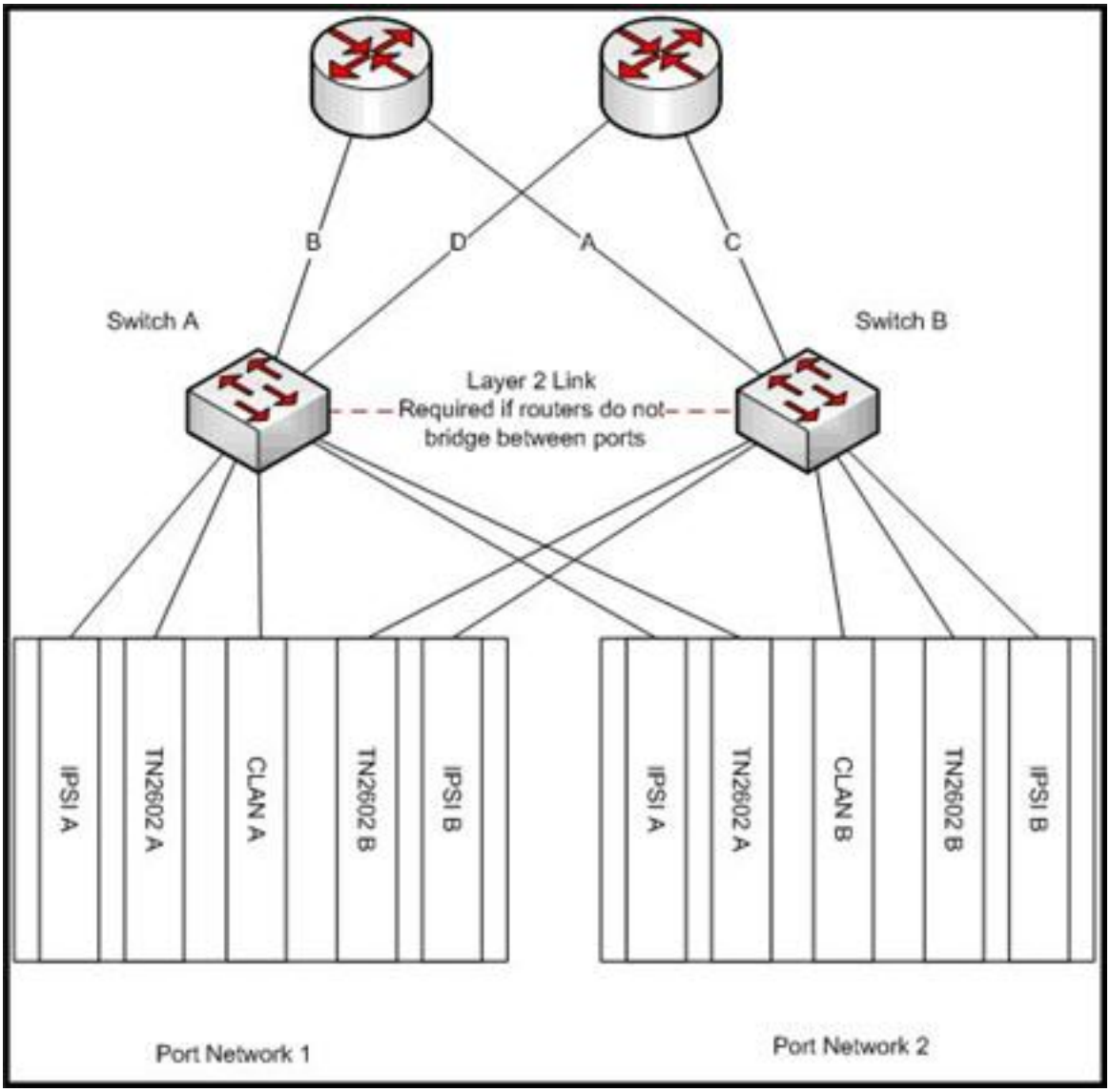


Figure 7: Redundant connections

It is good practice to have the redundant boards of each PN connected to redundant Layer 2 switches as shown to protect each PN from failure of the Layer 2 switch itself. If asymmetric redundancy is configured through IP bearer duplication, it is essential for proper fail over operation that the active and standby TN2602 circuit packs have equivalent Layer 2 connectivity. In the case of IP bearer duplication the secondary TN2602 circuit pack takes over by assuming both the L2 and L3 address of the connection terminations. This minimizes the disruption due to failover but requires the network be configured to accommodate the apparent move of an endpoint from one switch to the other, as it normally would for a spanning tree change.

Moving the L2 address to the standby device limits the disruption to the address forwarding tables of the L2 switches, which are designed to accommodate rapid connectivity moves.

Layer 2

Layer 2 configuration of the switches supporting the cluster should use IEEE 802.1w Rapid Spanning Tree Protocol (RSTP) to prevent loops and for selection between redundant links. Most modern switches implement this protocol. Selecting a device for Layer 2 access that does not support RSTP should be very carefully considered since those devices are likely to be obsolete and lacking in other highly desirable features in areas such as Quality of Service, security, and manageability. RSTP is also preferred over most alternative solutions that are typically not standards based and can cause problems with interoperability, scalability and configuration complexity. Whichever redundancy protocol is selected must be well understood by the IT staff responsible for maintaining the network.

It is good policy to enable RSTP on all ports of the Layer 2 switches, even those ports directly connected to hosts. Remember that misconfiguration and human error are more likely to occur than link failure and the added protection of loop avoidance is worth the minimal overhead. This is an additional argument in favor of RSTP as the redundancy protocol to use since other solutions may not be applicable uniformly to the subnet.

With modular configuration the spanning tree itself is kept simple and deterministic. Consider the sample spanning tree configuration in the figure below. The topology has been redrawn and the host connections have been removed to simplify the explanation. Assume the bridge priorities are assigned such that the VRRP primary router has the highest priority, the secondary router is next, Switch 1 is third, and Switch 2 is last. It is also important that the bandwidth of all links be equivalent and adequate to handle the aggregated traffic.

In the example below, links A and B are directly attached to the root bridge so they will be forwarding. Link C connects to a higher priority bridge than link D, so link D will be disabled and Switch 1 will be the designated root for the secondary router. In this configuration, traffic from the attached devices flows directly to the primary router on links A & B.

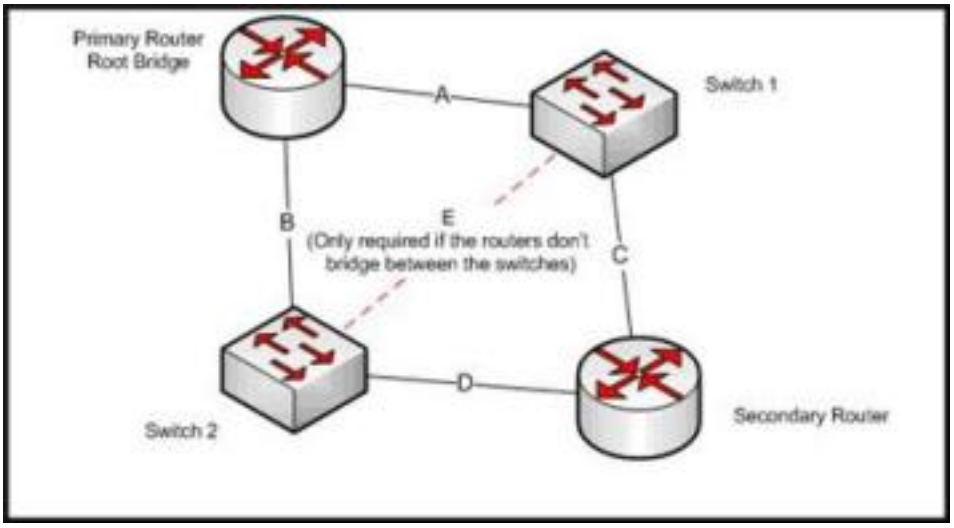


Figure 8: Sample spanning tree

If the primary router fails, the secondary router becomes both the active router and the root bridge, and traffic from the switches flows on the reconfigured spanning tree along links C & D. If bridge priorities are not managed, traffic from one switch may be directed through the secondary router and the other switch as normal operation.

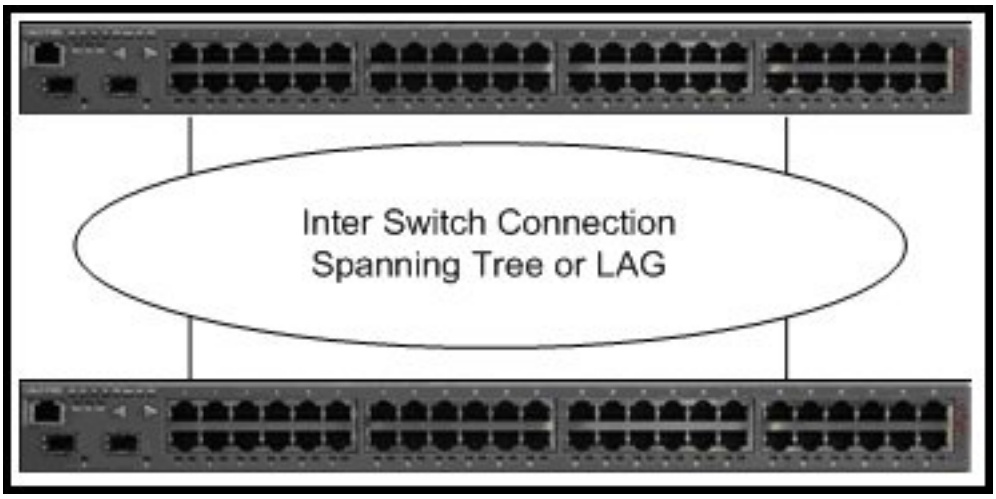


Figure 9: Alternate configuration - Layer 2

In the alternate integrated device configuration, bridge priority is less significant but other factors may add to complexity. In particular, link sizing becomes an issue if there are not enough Gigabit Ethernet aggregation ports. If a link aggregation group (LAG) is used, flow distributions must be understood to ensure correct behavior.

Layer 3

The symmetric or asymmetric question is linked to the configuration of redundancy for the routers serving this cluster. If the single subnet model is used, the router configuration in the

direction of the cluster will also follow the asymmetric model. Virtual Router Redundancy (VRRP) is configured with one router as the primary, and the other router as the secondary. If multiple subnets are configured, it is common practice to make one router the primary for some of the subnets and the other router as the primary for the rest. Note that VRRP should be configured with a failover latency greater than the latency required for the Layer 2 loop avoidance protocol to prevent LAN failures from perturbing the wider network. Typical defaults are between two and three seconds, which should be adequate in a simple well configured spanning tree.

The cluster subnet is exported to OSPF through the interfaces to the core so that the devices are reachable, but OSPF needs to know which router interface to use for the packets directed to the cluster. Proper operation requires that the link to the primary router is also the preferred OSPF path. If the primary router fails in such a way that the link to the core is not brought down, packets will not reach the cluster until the neighbor adjacency times out. Making these timeouts too small makes the protocol overly sensitive, and may still provide inadequate results.

Alternative configurations with static routes and VRRP operating in the direction of the core have also been proposed. The probability of a VRRP interchange that occurs asymmetrically is arguably lower than a router failure that leaves the physical link state unchanged. Some implementations address this by allowing the link state of different interfaces to be coupled. These techniques are also applicable to the OSPF solution, but are typically proprietary. Combining that observation with the benefit of decoupling route core disruption from local failure are arguments for this configuration.

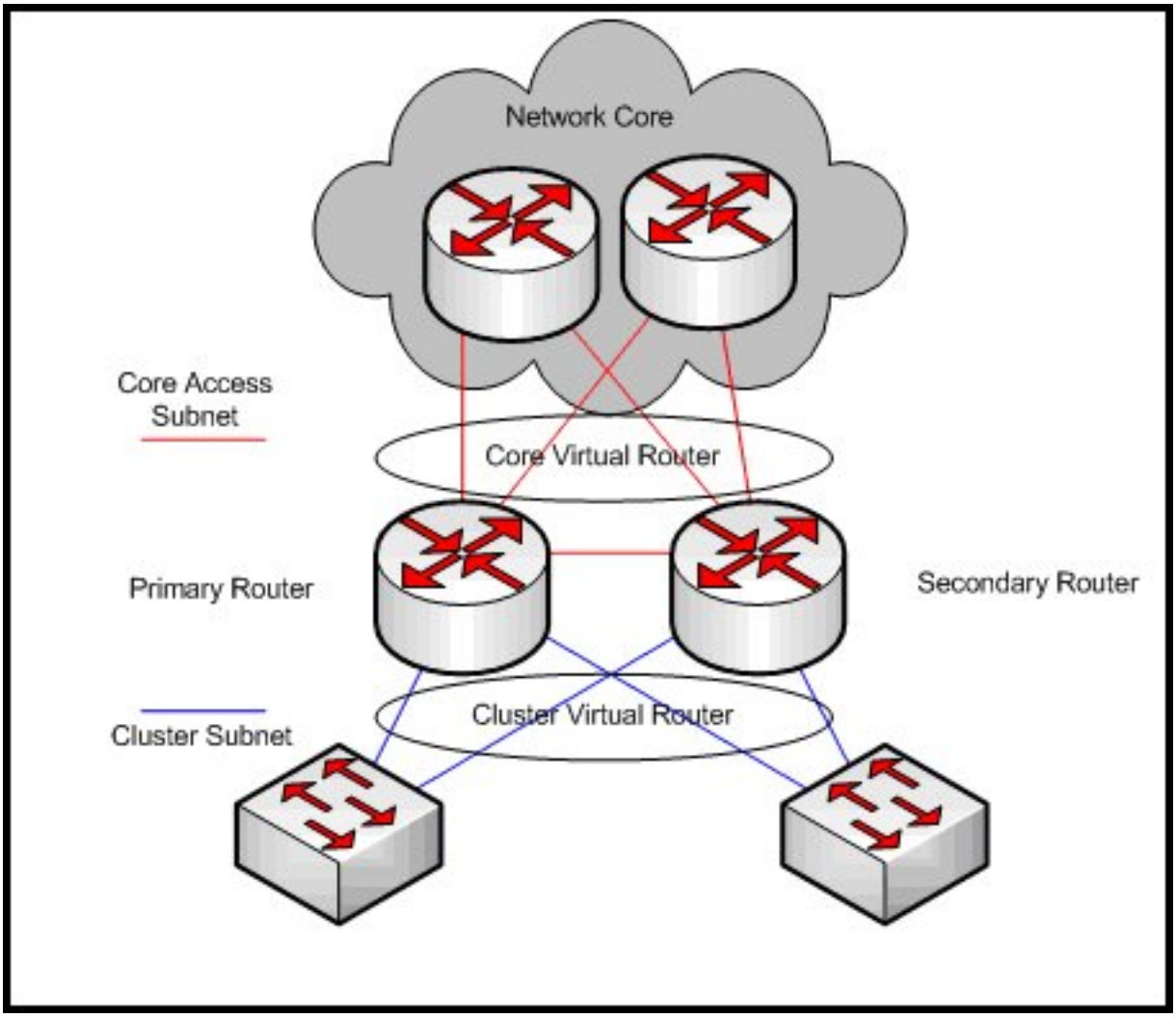


Figure 10: VRRP configured for Core Access

Quality of Service guidelines

This chapter contains guidelines for deploying Quality of Service (QoS) for an IP Telephony network. This chapter begins with an overview of Class of Service (CoS) versus QoS.

Class of Service refers to mechanisms that tags traffic in such a way that the traffic can be differentiated and segregated into various classes. *Quality of Service* refers to what the network does to the tagged traffic to give higher priority to specific classes. If an endpoint tags its traffic with Layer 2 802.1p priority 6 and Layer 3 Differentiated Services Code Point (DSCP) 46, for example, the Ethernet switch must be configured to give priority to value 6, and the router must be configured to give priority to DSCP 46. The fact that certain traffic is tagged with the intent

to give it higher priority does not necessarily mean it will receive higher priority. CoS tagging does no good without the supporting QoS mechanisms in the network devices.

Class of Service

IEEE 802.1p/Q at the Ethernet layer (Layer 2) and DSCP at the IP layer (Layer 3) are two standards-based CoS mechanisms that are used by Avaya products. These mechanisms are supported by the IP telephone, and the C-LAN and Media Processor circuit packs. Although TCP/UDP source and destination ports are not CoS mechanisms, they can be used to identify specific traffic, and can be used much like CoS tags. Other non-CoS methods to identify specific traffic are to key in on source and destination IP addresses and specific protocols, such as RTP. The Media Processor circuit pack and IP telephones use RTP to encapsulate audio.

Note that the 802.1Q tag changes the size and the format of the Ethernet frames. Because of this, older switches had to be explicitly configured to accept 802.1Q tagged frames. Otherwise, those switches might reject the tagged frames. This has not been a significant problem in many years. The two fields to be concerned with are the Priority and Vlan ID (VID) fields. The Priority field is the “p” in 802.1p/Q, and ranges in value from 0 to 7. (*802.1p/Q* is a common term that is used to indicate that the Priority field in the 802.1Q tag has significance. Prior to real-time applications, 802.1Q was used primarily for VLAN trunking, and the Priority field was not important.) The VID field is used as it always has been, to indicate the VLAN to which the Ethernet frame belongs.

The IP header with its 8-bit Type of Service (ToS) field, which was, and in some cases still is, originally used. This original scheme was not widely used, and the IETF developed a new Layer 3 CoS tagging method for IP called Differentiated Services (DiffServ, RFC 2474/2475). DiffServ uses the first 6 bits of the ToS field, and ranges in value from 0 to 63. Figure 11: Comparison of DSCP with original ToS on page 98 shows the original ToS scheme and DSCP in relation to the 8 bits of the ToS field.

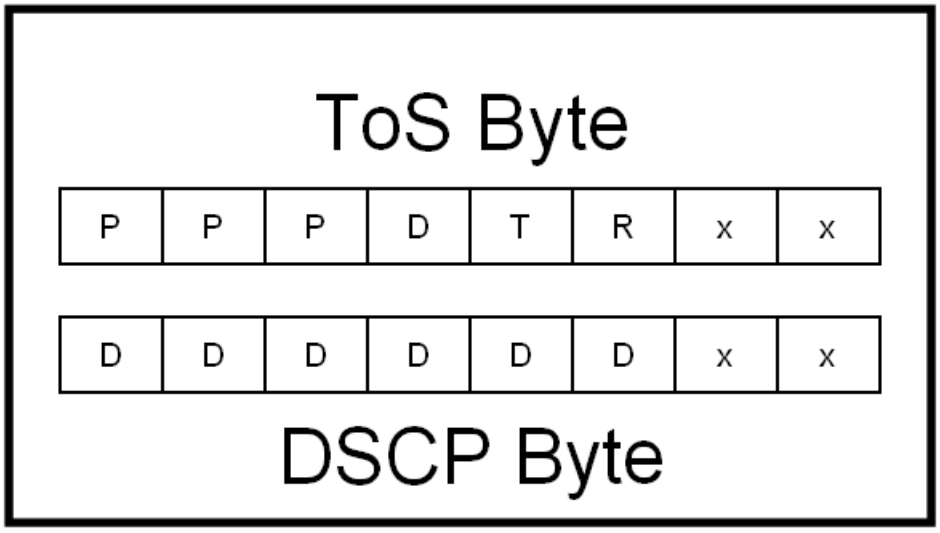


Figure 11: Comparison of DSCP with original ToS

Ideally, any DSCP value should map directly to a precedence and traffic parameter combination of the original scheme. This is not always the case, however, and it can cause problems on some older devices.

On any device, new or old, a nonzero value in the ToS field has no effect if the device is not configured to examine the ToS field. Problems arise on some legacy devices when the ToS field is examined, either by default or by enabling QoS. These legacy devices (network and endpoint) might contain code that implemented only the precedence portion of the original ToS scheme, with the remaining bits defaulted to zeros. This means that only DSCP values that are divisible by 8 (XXX000) can map to the original ToS scheme. For example, if an endpoint is tagging with DSCP 40, a legacy network device can be configured to look for precedence 5, because both values show up as 10100000 in the ToS field. However, a DSCP of 46 (101110) cannot be mapped to any precedence value alone. Another problem is if the existing code implemented precedence with only one traffic parameter permitted to be set high. In this case, a DSCP of 46 still does not work, because it requires 2 traffic parameter bits to be set high. When these mismatches occur, the older device (10 years older) might reject the DSCP tagged IP packet, or exhibit some other abnormal behavior. Most newer devices support both DSCP and the original ToS scheme.

Layer 2 quality of service

On Cisco and other vendor switches, IP telephony traffic can be assigned to higher priority queues. The number and the sizes of queues and how the queues function are device dependent, and beyond the scope of this document.

However, in general, a fixed number of queues exist, and the queues are usually not configurable. If the queues are configurable, it is typically not recommended. Older or lower end switches commonly have only two queues or none at all. Newer or higher-end switches commonly have four or eight queues, with eight being the maximum because there are only

eight Layer 2 priority levels. When configured to do so, the Ethernet switch can identify the high-priority traffic by the 802.1p/Q tag, and assign that traffic to a high-priority queue. On some switches, a specific port can be designated as a high-priority port, which causes all traffic that originates from that port to be assigned to a high-priority queue.

Layer 3 quality of service

It is usually more complicated to implement QoS on a router than on an Ethernet switch. Unlike Ethernet switches, routers do not just have a fixed number of queues. Instead, routers have various queuing mechanisms. For example, Cisco routers have standard first-in first-out queuing (FIFO), weighted fair queuing (WFQ), custom queuing (CQ), priority queuing (PQ), and low-latency queuing (LLQ). LLQ is a combination of priority queuing and class-based weighted fair queuing (CBWFQ), and it is Cisco's recommended queuing mechanism for real-time applications such as IP telephony. Each queuing mechanism behaves differently, is configured differently, and has its own set of queues.

First, the desired traffic must be identified using IEEE-802.1p/Q, DSCP, IP address, TCP/UDP port, or protocol. Then the traffic must be assigned to a queue in one of the queuing mechanisms. Then the queuing mechanism must be applied to an interface.

The interface itself might also require additional modifications, independent of the queuing mechanism, to make QoS work properly. For example, Cisco requires traffic shaping on Frame Relay and ATM links to help ensure that voice traffic is allotted the committed or guaranteed bandwidth. Cisco also recommends link fragmentation and interleaving (LFI) on WAN links below 768 kbps, to reduce serialization delay. Serialization delay is the delay that is incurred in encapsulating a packet and transmitting it out the serial interface. It increases with packet size, but decreases with WAN link size. The concern is that large low-priority packets induce additional delay and jitter, even with QoS enabled. This is overcome by fragmenting the large low-priority packets and interleaving them with the small high-priority packets, thus reducing the wait time for the high-priority packets. Table 6: Serialization delay matrix on page 99 lists serialization delay for a variety of packet sizes and line speeds. The formula for determining serialization delay is:

serialization delay = (packet size in bits / line speed)

Table 6: Serialization delay matrix

WAN line speed	Packet size					
	64 bytes	128 bytes	256 bytes	512 bytes	1024 bytes	1500 bytes
56 kbps	9 ms	18 ms	36 ms	72 ms	144 ms	214 ms
64 kbps	8 ms	16 ms	32 ms	64 ms	128 ms	187 ms
128 kbps	4 ms	8 ms	16 ms	32 ms	64 ms	93 ms
256 kbps	2 ms	4 ms	8 ms	16 ms	32 ms	46 ms
512 kbps	1 ms	2 ms	4 ms	8 ms	16 ms	23 ms

WAN line speed	Packet size					
	64 bytes	128 bytes	256 bytes	512 bytes	1024 bytes	1500 bytes
768 kbps	640 µs	1.28 ms	2.56 ms	5.12 ms	10.24 ms	15 ms

Because of all these configuration variables, properly implementing QoS on a router is no trivial task. However, QoS is needed most on the router because most WAN circuits terminate on routers.

QoS guidelines

There is no all-inclusive rule regarding the implementation of QoS because all networks and their traffic characteristics are unique. It is good practice to baseline the IP telephony response on a network without QoS, and then apply QoS as necessary. Avaya Professional Services (APS) can help with baselining services. Conversely, it is bad practice to enable multiple QoS features simultaneously, not knowing what effects, if any, each feature is introducing.

Generally, for newer network equipment, best practices involve enabling Layer 3 (DiffServ) QoS on WAN links traversed by voice. Tag voice and data with DiffServ Code Point 46 (Expedited Forwarding), and set up a strict priority queue for voice. If voice quality is still not acceptable, or if QoS is desired for contingencies such as unexpected traffic storms, QoS can then be implemented on the LAN segments as necessary.

There is one caution to keep in mind about QoS with regard to the processor load on network devices. Simple routing and switching technologies have been around for many years and have advanced significantly. Packet forwarding at Layer 2 and Layer 3 is commonly done in hardware (Cisco calls this fast switching, with switching being used as a generic term here), without heavy processor intervention. When selection criteria such as QoS and other policies are added to the routing and switching process, it inherently requires more processing resources from the network device. Many of the newer devices can handle this additional processing in hardware, and maintain speed without a significant processor burden. However, to implement QoS, some devices must move a hardware process to software (Cisco calls this process “process switching”). Process switching not only reduces the speed of packet forwarding, but it also adds a processor penalty that can be significant. This can result in an overall performance degradation from the network device, and even device failure. Each network device must be examined individually to determine if enabling QoS will reduce its overall effectiveness by moving a hardware function to software, or for any other reason. Since most QoS policies are implemented on WAN links, the following points increase the effectiveness of QoS remains:

- Hardware platforms such as the 2600, 3600, 7200, 7500 series, or later are required. Newer platforms such as the 1800, 2800 and 3800 series have no problem handling QoS because of powerful processors.
- Newer interface modules (WIC, VIP, and so on) are required.

Note:

If you are using Cisco devices with the interfaces mentioned above, consult Cisco to determine which hardware revision is required for any given module.

- Sufficient memory is required: device dependent.
- Recommended IOS 12.0 or later.

Several things should be examined whenever QoS is enabled on a network device. First, the network administrator should examine the processor load on the device, and compare it to levels before QoS was enabled. It is likely that the levels will have gone up, but the increase should not be significant. If it is, then it is likely that the QoS process is being done by software. Also, the processor load must remain at a manageable level (50% average, 80% peak). If the processor load is manageable, then the IP telephony response (for example, voice quality) should be checked to verify that it has improved under stressed conditions (for example, high congestion). If the IP telephony response has improved, the other applications should be checked to verify that their performances have not degraded to unacceptable levels.

IEEE 802.1Q standard

Surprisingly, many data network engineers are still not familiar with CoS/QoS. Data networks were never designed for real-time protocols and this section helps them to understand the protocols.

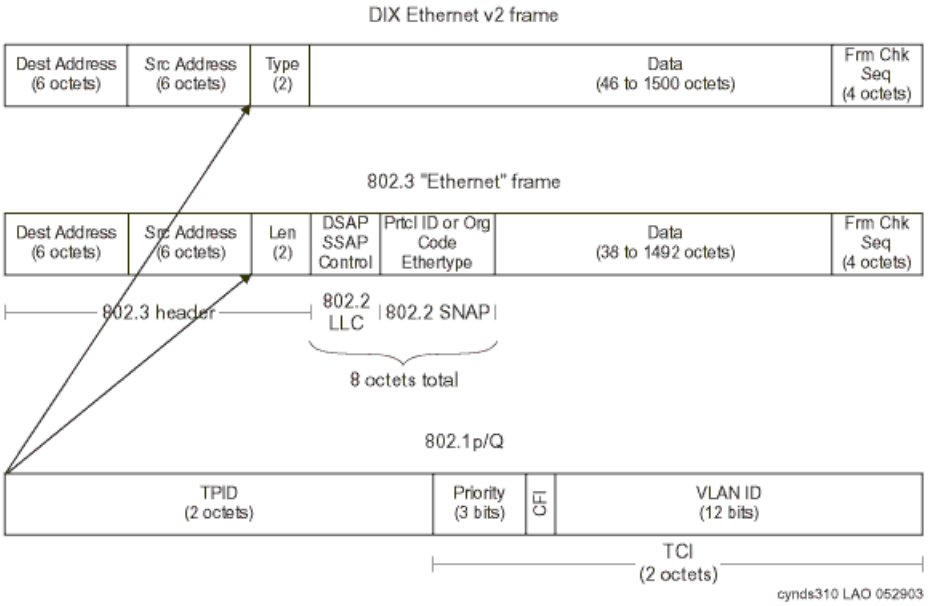


Figure 12: 802.1Q tag

The IEEE 802.1Q standard is a Layer 2 tagging method that adds 4 bytes to the Layer 2 Ethernet header. IEEE 802.1Q defines the open standard for VLAN tagging. Two bytes house 12 bits that are used to tag each frame with a VLAN identification number. The IEEE 802.1p standard uses 3 of the remaining bits in the 802.1Q header to assign one of 8 different classes

of service. Communication Manager users can add the 802.1Q bytes and set the priority bits as desired. *Avaya suggests that a priority of 6 be used for both voice and signaling* for simplicity. However, the default values are — 5-Video, 6-Voice, and 7-Control. IEEE 802.1p and IEEE 802.1Q are OSI layer 2 solutions, and work on frames.

Because 802.1Q is a Layer 2 (Ethernet) standard, it only applies to the Ethernet header. At every Layer 3 boundary (router hop), the Layer 2 header, including CoS parameters, is stripped and replaced with a new header for the next link. Thus, 802.1Q does not enable end-to-end QoS.

Recommendations for end-to-end QoS

When end-to-end QoS is desired, Avaya recommends using Differentiated services on page 102, a Layer 3 CoS method. Modern routers can map DiffServ Code Points (DSCP) to 802.1p priority values, so 802.1p tags can be recreated on each Ethernet link.

IEEE 802.1p states a standard according to which these bits are used for CoS. The precedence is listed in Table 7: IEEE 802.1 precedence and service mapping on page 102.

Table 7: IEEE 802.1 precedence and service mapping

User priority	Service mapping
000	Default, assumed to be best effort
001	Reserved, less than best effort
010	Reserved
011	Reserved
100	Delay sensitive, no bound
101	Delay sensitive, 100 ms bound
110	Delay sensitive, 10 ms bound
111	Network control

Differentiated services

The Differentiated Services (DiffServ) prioritization scheme redefines the existing ToS byte in the IP header (Figure 13: Differentiated Services (DiffServ) ToS byte on page 103) by combining the first 6 bits into 64 possible combinations. The ToS byte can be used by Communication Manager, IP telephones, and other network elements such as routers and switches in the LAN and WAN.

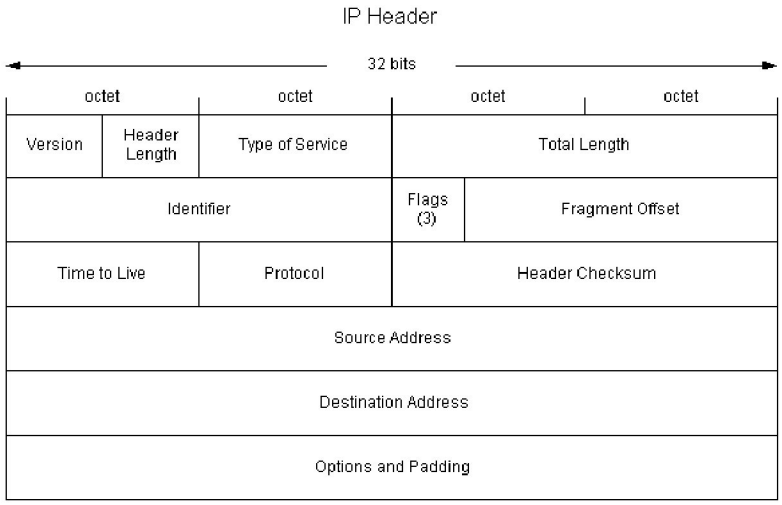


Figure 13: Differentiated Services (DiffServ) ToS byte

A DiffServ Code Point (DSCP) of 46 (101110), referred to as expedited forwarding (EF), is suggested for the proper treatment of voice packets. Signaling packets can also be marked with DSCP 46 if there is sufficient bandwidth to prevent dropped packets. To assure that voice and signaling packets are not in contention, mark signaling packets with a different DSCP value. With Communication Manager, you can set any DSCP value needed to work with a company's QoS scheme.

Some common DiffServ Code Points are defined in RFCs 2474 and 2475. Although DSCPs are specified in IETF RFCs, the treatment of packets that are tagged with DiffServ depends on implementation.

Note that older routers might require a DSCP setting of 40 (101000), which is backward compatible with the original ToS byte definition of critical. But again, Avaya products and software allows users to set any of the 64 possible DSCP values to work with your voice quality policy. Avaya recommends DSCP-46 for both bearer and control for simplicity. Default values are — Bearer-46, Control-34, Video-26. The ToS byte is an OSI model Layer 3 solution, and works on IP packets on the LAN and possibly the WAN, depending upon the service provider.

Table 8: Original ToS specification

Bit description	Value	Use
Bits 0-2 IP precedence	000 001 010 011 100 101 110 111	Routine Priority Immediate Flash Flash Override CRITIC/ECP Internetwork control Network control
Bit 3 delay	0	Normal

Bit description	Value	Use
	1	Low
Bit 4 Throughput	0 1	Normal High
Bit 5 reliability	0 1	Normal High
Bit 6 monetary cost	0 1	Normal Low
Bit 7 reserved		Always set to 0

Resource reservation protocol

Resource Reservation Protocol (RSVP) is a protocol that hosts can use to request specific QoS parameters through the network for a particular application data stream. A host can request guaranteed service through a network. If all routers have RSVP support enabled, and if there exists sufficient unreserved bandwidth, a reservation is established throughout the network. If insufficient bandwidth exists, the reservation fails and notifies the hosts. At that point, hosts can choose to send traffic without a reservation, or drop the connection.

RSVP can be enabled per network region on the network region form. If RSVP is enabled, endpoints including IP telephones and media processors attempt to establish a reservation for each call. If the reservation fails, Avaya endpoints still try to place a call, but lower the DiffServ priority of the call to the better-than-best-effort (BBE) DSCP that is defined on the network region form. By default, this value is 43.

If RSVP is enabled on a network region, it is very important that it also be enabled on associated routers. If not, all RSVP reservations fail, and all voice traffic in that region is marked with the BBE DSCP, which generally receives degraded service versus the EF (DSCP 46) DiffServ Code Point.

Queuing methods

This section discusses common queuing methods and their appropriateness for voice.

Weighted fair queuing

Weighted fair queuing (WFQ) is similar to first in, first out (FIFO) queuing, except that it grants a higher weight to small flows, and flows that are marked with higher DiffServ or IP TOS priorities. This queuing strategy does allow smaller (for example, telnet) and higher-priority (for example, IP telephony) protocols to squeeze in before high-flow (for example, ftp) packets, but

does not starve off any traffic. By itself, it is not appropriate for IP telephony traffic because high-flow traffic can still delay IP telephony traffic, and cause unacceptable latency and jitter.

Priority queuing

Strict priority queuing (PQ) divides traffic into different queues. These queues are usually high, medium, normal, and low, based on traffic type. This form of queuing services the queues in order of priority, from high to low. If there is a packet in the high-priority queue, it will always be serviced before the queue manager services the lower-priority queues. With priority queuing, however, it is possible to starve out lower-priority flows if sufficient traffic enters the high-priority queue. This mechanism works very well for IP telephony traffic (where IP telephony bearer and signaling are inserted in the high-priority queue), but might work less well for routine data traffic that is starved out if sufficient high-priority traffic arrives.

Round-robin

Round-robin (sometimes called *custom*) queuing sorts data into queues, and services each queue in order. An administrator manually configures which type of traffic enters each queue, the queue depth, and the amount of bandwidth to allocate to each queue.

Round-robin queuing is not particularly suited to IP telephony. It does not ensure strict enough priority to voice packets, so they may still wait behind other traffic flows in other queues. Latency and jitter can be at unacceptable levels.

Class-Based weighted fair queuing

Class-Based Weighted Fair Queuing (CB-WFQ) with Low-Latency Queuing (LLQ), which is sometimes called Class-Based Queuing (CBQ), combines the above-mentioned queuing mechanisms. Generally, there is one strict-priority queue, several round-robin queues, and weighted fair queuing for the remainder. This queuing mechanism works very well for converged networks. IP telephony bearer and signaling packets receive the priority they need, while there remains an equitable mechanism for distributing remaining bandwidth. In addition, limits can be set on the high-priority queue to prevent it from using more than a specified amount of bandwidth. Bandwidth that is reserved for the high-priority queue will be given to other queues if insufficient traffic enters the high-priority queue.

Random early detection and weighted random early detection

Although they are not queuing methods *per se*, Random Early Detection (RED) and Weighted Random Early Detection (WRED) are important queue management techniques. RED and WRED work by randomly discarding packets from a queue. RED takes advantage of the congestion control mechanism of TCP. By randomly dropping packets prior to periods of high congestion, RED causes the packet source to decrease its transmission rate. Assuming that

the packet source is using TCP, it will decrease its transmission rate until all the packets reach their destination, which indicates that the congestion is cleared. Some implementations of RED, called Weighted Random Early Detection (WRED), combines the capabilities of the RED algorithm with IP Precedence. This combination provides for preferential traffic handling for higher-priority packets. It can selectively discard lower-priority traffic when the interface begins to get congested, and provide differentiated performance characteristics for different classes of service.

RED and WRED are useful tools for managing data traffic, but should not be used for voice. Because IP telephony traffic runs over UDP, because IP telephony protocols do not retransmit lost packets, and because IP telephony transmits at a constant rate, the IP telephony queue should never be configured for WRED. WRED only adds unnecessary packet loss, and consequently reduces voice quality.

Traffic shaping and policing

Traffic shaping is a mechanism to reduce the rate at which data is transmitted over an interface. When people discuss traffic shaping, they are usually referring to the related technology of traffic policing. Policing works by either adjusting the priority of excess traffic to a lower queue, or discarding it. As with RED, discarding TCP traffic has the effect of throttling the stream by forcing window size to shrink, and decreasing its transmission rate. Because RTP is a fixed-bandwidth application, discarding RTP packets reduces voice quality without altering the transmission rate. Adjusting the priority of voice traffic removes the strict priority protection that reduces latency and jitter, and offers the highest voice quality. Thus, in most cases, it is beneficial to use the QoS mechanisms listed above, rather than traffic shaping and policing, to offer the highest quality for voice.

Frame Relay traffic shaping

Traffic shaping is important in technologies that implement virtual circuits (VCs), such as Frame Relay or ATM, where the Committed Information Rate (CIR) might be less than the physical speed of the interface, the port speed. In such scenarios, it is possible for traffic to burst above the CIR. Depending on the Service Level Agreement (SLA), a carrier might mark excess traffic as Discard Eligible (DE), and either delay or discard it if congestion is detected within the network of the carrier. This behavior is unacceptable for voice traffic, which must minimize delay and jitter to achieve optimal voice quality. To solve this issue, Frame Relay traffic shaping gives an administrator tools to limit the transmission rate on a Frame Relay virtual circuit to the CIR.

A popular misconception is that voice traffic can be confined to the CIR, while data traffic can be allowed to burst. Unfortunately, that is not how Frame Relay works. There is not a QoS mechanism for Frame Relay that is negotiated between service providers and customers. Service providers view all traffic equally, and mark any packet that exceeds the CIR as DE, even if the packet is high-priority voice. Thus, the only way to guarantee optimal performance for voice traffic is to restrict the traffic rate to the CIR.

Configuring Cisco router

1. Disable Frame Relay adaptive shaping.

   This technique reduces the CIR in response to backwards explicit congestion notification (BECN) messages from the service provider. Because traffic is being transmitted at the CIR in the first place, it does not need to be throttled.

2. Set *cir* and *mincir* to the negotiated CIR.

   If FRF.12 fragmentation is implemented, reduce the *cir* and *mincir* values to account for the fragment headers.

3. Set *be*, the excess burst rate, to 0.

4. Set *bc*, the committed burst rate, to cir/100.

   This accounts for at most a 10 ms serialization delay.

5. Apply this map class to an interface, subinterface, or VC.

Example

Thus, the complete configuration for Frame Relay traffic shaping looks like:

```
map-class frame-relay
    NoBurst no frame-relay adaptive shaping
    frame-relay cir 384000! (for a 384K CIR)
    frame-relay mincir 384000
    frame-relay be 0
    frame-relay bc 3840

    interface serial 0
    frame-relay class NoBurst
```

Fragmentation

One large cause of delay and jitter across WAN links is serialization delay, or the time that it takes to put a packet on a wire. For example, a 1500 byte FTP packet takes approximately 214 ms to be fed onto a 56 Kbps circuit. For optimal voice performance, the maximum serialization delay should be close to 10 ms. Thus, it can be problematic for a voice packet to wait for a large data packet over a slow circuit. The solution to this problem is to fragment the large data packet into smaller pieces for propagation. If a smaller voice packet comes in, it can be squeezed between the data packet fragments and be transmitted within a short period of time.

The sections that follow discuss some of the more common fragmentation techniques.

Maximum transmission unit

The maximum transmission unit (MTU) is the longest packet (in bytes) that can be transmitted by an interface without fragmentation. Reducing the MTU on an interface forces a router to fragment the large packet at the IP level. This allows smaller voice packets to squeeze through in a timelier manner.

The drawback to this method is that it increases overhead and processor occupancy. For every fragment, a new IP header must be generated, which adds 20 bytes of data. If the MTU is 1500 bytes, the overhead is approximately 1.3%. If the MTU is shortened to 200 bytes, however, the overhead increases to 10%. In addition, shortening the MTU to force fragmentation increases processor utilization on both the router and the end host that needs to reassemble the packet.

For these reasons, shortening the MTU is only recommended as a last resort. The techniques described later in this section are more efficient, and should be used before changing the values of the MTU. When changing the MTU, size it such that the serialization delay is less than or equal to 10 ms. Thus, for a 384 kbps circuit, the MTU should be sized as follows: 384 kbps *0.01 second (10 ms)/8 bits/byte = 480 bytes. As the circuit size diminishes, however, care should be taken to never reduce the MTU below 200 bytes. Below that size, telephony signaling and bearer (voice) packets can also be fragmented, which reduces the link efficiency and degrades voice performance.

Link fragmentation and interleaving

Link Fragmentation and Interleaving (LFI) is an enhancement to Multilink PPP (MLP) that fragments packets at the Layer 2 (PPP) level. Fragmenting at the IP layer, as with MTU reduction, forces the addition of a new 20 byte IP header and an 8 byte PPP header. However, fragmenting at the data link (PPP) layer only forces generation of an 8 byte PPP header, which greatly increases the efficiency of the link.

Avaya recommends use of LFI functionality instead of MTU manipulation when transmitting IP telephony packets over PPP links. As with MTU, Avaya recommends sizing packets so that the serialization delay is approximately 10 ms or less.

FRF.12

FRF.12 is a Frame Relay standard for fragmentation. It works for Frame Relay in the same way that LFI works for PPP, with similar increases in efficiency over MTU manipulation. When implementing a Frame Relay network, Avaya recommends using FRF.12 for fragmentation, and sizing the fragments so the serialization delay is no more than 10 ms.

Real-time transport protocol

RTP header compression is a mechanism that reduces the protocol overhead that is associated with IP telephony audio packets. It is a function of the network, and not a function of the IP telephony application. Along with the benefits of using RTP header compression, there are also cautions.

Application perspective

Table 9: Anatomy of 20 ms G.729 audio packet on page 109 shows the anatomy of a 20 ms G.729 audio packet, which is recommended for use across limited bandwidth WAN links. Notice that two-thirds of the packet is consumed by overhead (IP, UDP, and RTP), and only one-third is used by the actual audio.

Table 9: Anatomy of 20 ms G.729 audio packet

IP header	UDP header	RTP header	20 ms of G.729 audio
20 B	8 B	12 B	20 B

It is important to understand that all 20-ms G.729 audio packets, regardless of the vendor, are constructed like this. Not only is the structure of the packet the same, but the method of encoding and decoding the audio itself is also the same. This sameness is what allows an Avaya IP telephone to communicate directly with a Cisco IP telephone, or any other IP telephone, when using matching codecs. The packets from the application perspective are identical.

Network perspective

RTP header compression is a mechanism that routers use to reduce the 40 bytes of protocol overhead to approximately 2 to 4 bytes. Cisco routers uses this RTP header compression. The RTP header compression can drastically reduce the IP telephony bandwidth consumption on a WAN link when using 20 ms G.729 audio. When the combined 40 byte header is reduced to 4 bytes, the total IP packet size is reduced by 60% (from 60 bytes to 24 bytes). This equates to reducing the total IP telephony WAN bandwidth consumption by roughly half, and it applies to all 20 ms G.729 audio packets, regardless of the vendor.

Recommendations for RTP header compression

Enterprises that deploy routers that are capable of this feature might be able to benefit from it. However, Cisco recommends caution in using RTP header compression on its routers because it can significantly tax the processor if the compression is done in software. Depending on the processor load before compression, enabling RTP header compression can significantly slow down the router, or cause the router to stop completely. For best results, use a hardware/IOS/interface module combination that permits the compression to be done in hardware.

RTP header compression has to function with exactness or it will disrupt audio. If for any reason the compression at one end of the WAN link and decompression at the other end do not function properly, the result can be intermittent loss of audio or one-way audio. This has been very difficult to quantify, but there is some anecdotal evidence that cRTP sometimes leads to voice-quality issues. One production site in particular experienced intermittent one-way audio, the cause of which was garbled RTP audio samples inserted by the cRTP device. When, for experimentation purposes, RTP header compression was disabled, the audio problems went away.

Cisco configuration example

The following example shows the Cisco IOS Ethernet switch commands that demonstrate and optimize an environment needed for IP Telephony and video.

Connecting to the Ethernet switch

Ports 1 through 10 are assigned to the voice vlan and this config is suitable for stand-alone IP phones and video devices that connect to the Ethernet switch..

```
Switch> enable                                    change from user to privilege mode
Switch # configure terminal                      change to global config mode
Switch(config)# vlan 20 name v20                 create vlan 20 for voice traffic
Switch(config)# int range fa 0/1 - 10             context for ports 1 through 10
Switch(config)# des "IP phones with no
PCs attached"                                  Description of ports' use
Switch(config)# switchport access vlan 20 change native/port vlan from data (vlan-1)
to voice (vlan-20)
Switch(config)# no cdp enable                    disable CDP for ports 11-20 (remove
proprietary protocols)
Switch(config)# spanning-tree portfast          place ports in forwarding mode
immediately
Switch(config)# spanning-tree bpduguard
enable                                       enable bpdu guard in case of a layer-2 loop
```

Attaching a PC

Ports 12 through 20 are assigned to the voice vlan and this config is suitable for IP phones and video devices that have a PC attached to them..

```
Switch(config)# int range fa 0/11 - 20            context for ports 11 through 20
Switch(config)# des "IP phones with PCs
attached"                                        description of ports' use
Switch(config)# speed 100                              lock port speed to 100-Mbps
(optional setting to Auto-Neg)
Switch(config)# duplex Full                       lock port duplex to Full (optional
setting to Auto-Neg)
Switch(config)# switchport voice vlan 20         config implicit trunk for IP phones
or video endpoints
Switch(config)# no cdp enable                     disable CDP for ports 11 through 20
Switch(config)# spanning-tree portfast             place ports in forwarding mode
immediately
Switch(config)# spanning-tree bpduguard enable enable bpdu guard in case of a
layer-2 loop
```

Taking traffic to a router

These commands create a trunk (more than one vlan) to take voice and data traffic to a router.

```
Switch(config)# int fa 0/48                          context for port 48
Switch(config-if)# des "Uplink trunk to
router R1"                                          description of port usage
Switch(config-if)# switchport trunk encap
dot1q                                                define port 48 as a data trunk
using 802.1Q
Switch(config-if)# switchport mode trunk            enable trunking mode
Switch(config-if)# switchport nonegotiate           trunk port 48 will not negotiate a
trunk status with the
                                                    other end of the link
Switch(config-if)# switchport trunk allowed
vlan remove 2-19,21-4094                             remove unneeded vlans
```

Avaya ExpertNet™ VoIP Assessment Tool

Avaya ExpertNet™ VoIP Assessment Tool overview

Avaya Professional Services (APC) offer the Avaya ExpertNet ™ VoIP Assessment Tool (EVAT) which is used during network readiness assessments. These assessments are conducted by Avaya Professional Services engineers who use this tool to determine the readiness of a network to support telephony and video throughout the media path. Network readiness assessment is scheduled through Avaya account team or business partner.

Avaya strongly recommends a network readiness assessment to ensure the Avaya network best practices have been followed, to appropriately prioritize real time data and avoid impairments to voice and video traffic.

EVAT runs on an Avaya laptop server residing on a customer premise, along with strategically placed single board computers (SBCs, commonly referred to as babels) to collect Quality of Service (QoS) metrics across the media path.

The web interface built into the EVAT server allows for the setup, execution, and monitoring of simulated calls. The SBCs carry out the work of simulating calls, collecting QoS measurements, and aggregating the measurements back to the EVAT server for processing. All processed measurements are stored in the EVAT server database. An Avaya Professional Services engineer can then generate QoS metric graph reports from the stored information using the EVAT Web interface.

The graphs focus on the following factors that affect successful VoIP and video deployment:

- Bandwidth utilization
- Codec selection

- One-way delay
- Jitter
- Packet loss
- Packet prioritization
- Reliability

EVAT provides:

- Synthetic VoIP traffic generation and measurement of VoIP metrics.
- Analysis of integrated VoIP, video and data
- Graphical depiction of measured calls
- OSI layer-2/layer-3 topology obtained from Simple Network Management Protocol (SNMP)
- Layer-3 topology obtained from Traceroute probes
- SNMP data collection from network devices

Note:
The SNMP data collection and the network topology discoveries are optional features. If they are allowed, a more comprehensive analysis can be provided.

EVAT key differentiators

- Provides assessment of a live network 24x7, over a period of several days.
- Simulates IP calls, measures effectiveness of the QoS mechanisms, and optionally measures bandwidth utilization across the network in real time.
- Delivers both layer-2 and layer-3 topology discovery for a more complete network view.

EVAT features

Voice traffic generation and measurement

EVAT uses Real-time Transport Protocol (RTP) that simulates VoIP calls between two endpoints in a pattern appropriate to the agreed upon test plan. The calling pattern for minimum or maximum network load can be configured as appropriate for the VoIP/video call volume. The various parameters of the calls are:

- Codec selection
- DSCP value
- Call volume
- Port range
- Payload

The SBCs use a User Datagram Protocol (UDP) port in the range of 2048 to 3329 (configurable) to simulate synthetic RTP calls.

Video traffic generation and measurement

Avaya Professional Services engineer can configure the video call patterns between SBCs or Windows Agents for minimum or maximum network load. EVAT provides data which is analyzed by Avaya engineers, who conduct assessments and provide detailed network readiness reports based on parameters such as number of video calls to simulate and DSCP values for the synthetic video calls. The video calls use a configurable UDP port in the range of 2048 to 3329.

IPSI/TCP/SIP traffic generation assessment

In addition to voice and video traffic path analysis, EVAT also supports call signaling path analysis. For call signaling analysis, EVAT supports Transmission Control Protocol (TCP), IP Server Interface (IPSI), and Session Initiation Protocol (SIP) test patterns.

In a TCP test, a pair of SBCs can be selected to replicate the optimum message size and frequency with set characteristics for endpoints. The options for a TCP test include the DSCP values, message size, message frequency, bandwidth, and port number.

In an IPSI test, an SBC can be configured to simulate a Communication Manager server and another SBC to simulate an IPSI board. The SBC that simulates the Communication Manager server sends a message every second to the SBC representing the IPSI board. The SBC at the Communication Manager side then records an error if the time between responses from the IPSI SBC is longer than the specified time.

In a SIP test, a pair of SBCs can be selected to replicate the expected call traffic, SIP endpoints, and SIP trunks. The Avaya Professional Services engineer can then set the call volume selection to the expected value on the network. The call volume selection options include the DSCP value, call volume, and port number.

SNMP monitoring

EVAT can be configured to analyze SNMP data from routers and Ethernet switches that lie in the path of the synthetic calls. EVAT gathers the information while making the synthetic calls. This information can be divided into two categories:

- Device level management information bases (MIBs) that gather packet level counters for traffic and errors.
- Interface level MIBs that gather octet level traffic and errors.

Scheduled calls

The scheduled calls feature of EVAT enables the Avaya Professional Services engineer to start and end a call unattended. Call patterns can be scheduled to start and stop at specified dates and times.

Historical network visualization with QoS

A call ratio for a device is the ratio of the number of calls passed through the device that has exceeded the set threshold for one-way delay, jitter, packet loss, or mean opinion score (MOS) to the total number of calls that passed through that device.

The QoS parameters that can be configured for coloring include one-way delay, jitter, packet loss, and MOS.

Data Graph generation

After performing all of the operations, EVAT stores the findings on the EVAT database. Using the gathered information, the Avaya Professional Services engineer can then produce multiple, extensive data graphs from the EVAT web interface.

EVAT benefits

Real-time network assessment

EVAT has the capability of initiating synthetic IP calls, and measures QoS and utilization across the network in real time.

Powerful network analysis

EVAT supports call signaling and video simulation, thereby providing a powerful network analysis, used in concert with other gathered information to complete a network readiness assessment.

EVAT operation

Figure 14: Network schematic of Avaya ExpertNet VoIP assessment tool on page 115 show network schematic of Avaya ExpertNet™ VoIP Assessment Tool. The EVAT call placement software runs on SBCs.

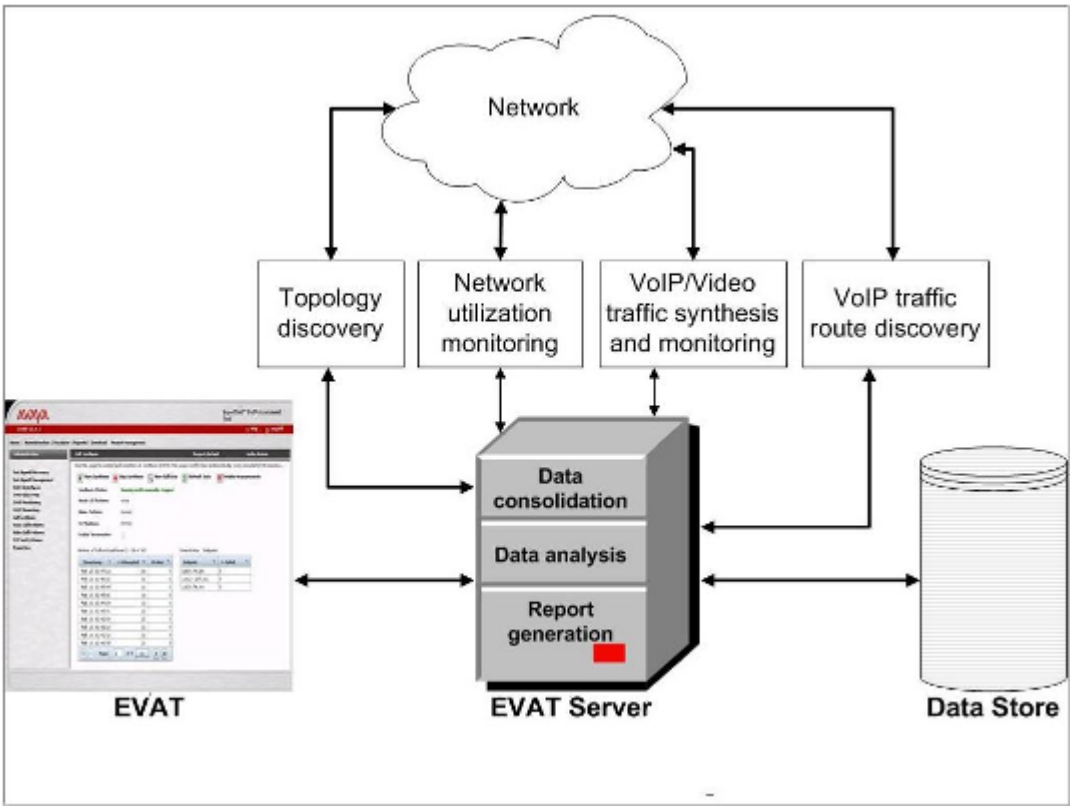


Figure 14: Network schematic of Avaya ExpertNet VoIP assessment tool

Reports

EVAT provides the following types of metrics graphs/charts:

- Time Series QoS
- Summary All
- Summary One To Many
- Summary SNMP Device Errors One Device
- Summary SNMP Interface Errors One Interface
- Summary SNMP Utilization
- Time Series SNMP Device
- Time Series SNMP Interface
- TCP Bandwidth Report
- TCP Delay Report
- IPSI Bandwidth Report

- IPSI Delay Report
- SIP Bandwidth Report
- SIP Delay Report
- Concurrent Calls
- Many To Many Concurrent Calls
- Group To Group Calls and Bandwidth
- Group To Group Total Bandwidth
- One Group To All Total Bandwidth
- Inventory Report
- Endpoint Health Report
- Per Hop DSCP Report

Chapter 6: Call processing

Avaya Aura® call processing

Avaya Aura® introduces important architectural differences in how Communication Manager call processing is distributed. The Avaya Common SIP Reference Architecture establishes roles that Communication Manager assumes in an Avaya Aura® configuration that includes a Session Manager SIP network. There are three roles:

- Communication Manager as a Feature Server with Trunk Gateway functionality
- Communication Manager as a Feature Server without Trunk Gateway functionality
- Communication Manager as an Evolution Server

In an Avaya Aura® Session Manager configuration, call processing is now decomposed into origination and termination processing. A single Communication Manager is no longer responsible for the complete processing of a call. This decomposition of call processing is referred to as the Half Call Model. Origination and termination processing for session establishment invokes pre-defined sequences of applications, of which Communication Manager is one alternative – for telephony features. Session Manager controls when and in which order Communication Manager, as one of the set of pre-defined applications, is invoked for origination and/or termination processing. The origination and termination sequences for a SIP user are not required to be the same. This process of routing session establishment requests through a series of applications is known as Application Sequencing.

Application sequencing

Application sequencing allows a provisioned series of applications to be invoked one at a time as each new call session is being established. This allows new applications to work with existing applications to provide enhanced functionality on each and every session. It is the Session Manager that binds different applications together to act on a same session. Typically, the sequenced applications are invoked by Session Manager upon receiving session establishing requests. Session Manager may receive such requests from clients, other Session Managers (for incoming requests for the same or an external domain), or from application servers. The name application sequencing is derived from the fact that Session Manager invokes these applications in sequence (not parallel). Application sequencing helps in rapidly

composing and delivering features by combining different (smaller) applications and features.

Related topics:

Half-call model on page 118
Full-call model on page 119

Half-call model

Session Manager allows applications to be invoked using a provisioned application list (users may have a unique application list or inherit a list, called an application-set that is assigned to a group of users). In any case, the provisioned applications are invoked in a particular sequence to provide the desired user experience. The order of applications—and even the type of applications—does not necessarily have to be the same for originating and terminating processing for the same user. For instance, when a particular user is initiating a session, the administrator might wish to invoke applications such as authorization, white-list/black-list verification, IVR and traditional Communication Manager features. Whereas, when a session terminates to the same user, the administrator may wish to invoke applications such as call assistance, call forwarding, voice messaging, and traditional Communication Manager features. This objective is achieved in Session Manager by designating different applications to be associated with the originating and terminating side of the call.

The concept of associating a session with originating or terminating processing is referred to as the *half-call model*. Note that the concept of the Session Manager's half-call model is different from the traditional telephony half-call model. In a traditional switch model, on receiving an incoming request to setup a call, the switch launches a state machine that represents the incoming or originating side. Similarly, when the switch is about to initiate an outbound call (either to a directly connected user or trunk), switch launches another state machine representing the outgoing or terminating side. These two independent state machines (called half-calls) are tied together by switch logic to complete the call setup.

Within Session Manager, it follows a different half-call model - unlike the traditional call model. The Session Manager-based half-call model is more elaborate than the traditional model, for example, in a traditional half-call model there is always an originating leg concatenated with a terminating leg. Whereas, with the Session Manager half-call model, while processing a request for the originating side or the terminating side, one can have multiple incoming and outgoing legs concatenated. On receiving an incoming request, Session Manager might have to invoke multiple applications on behalf of the calling user before it attempts the terminating logic. Similarly, before terminating the request to the target contact, Session Manager might have to invoke multiple applications on behalf of the terminating user.

Full-call model

In the full call model, the processing of a call request is done in one step. The origination and termination parts of the call are processed without a break. Classic Communication Manager adheres to the full-call model.

Because application sequencing expects the applications to follow the half-call model and because Communication Manager administered as an evolution server supports full-call model, no other sequenced application should be provisioned along with evolution server.

For the full-call model, the **IMS-enabled?** field on the SIP Signaling Group screen must be set to **n** (disabled).

IMS flag and SIP-ISC interface

This section clarifies the distinction and relationship between the IMS flag and SIP-ISC interface. First, the IMS Flag and SIP-ISC are not synonyms. The IMS flag is an attribute of SIP trunks that appears on the SIP Signaling Group screen in Communication Manager. The specific field is **IMS enabled?**, and the values are **yes** or **no**.

SIP-ISC is a set of SIP rules that Session Manager and the sequenced applications follow to communicate with each other. The use of SIP-ISC rules by Communication Manager is also an attribute of SIP trunks that appears on the SIP Signaling Group screen in Communication Manager. The specific field is **Peer Server:**. The value must be **SM** for the SIP-ISC rules to be in force. The **Peer Server** field value can be set administratively or via an automatic Peer Detection mechanism. There is another field on the SIP Signaling Group screen, **Peer Detection Enabled?**, that controls if the automatic detection mechanism is used or not.

The IMS flag can only be set to **Yes** when the Peer Server is a Session Manager.

When the IMS Flag is set to **Yes**, Communication Manager assumes the role of a Feature Server. That is, Communication Manager performs the half-call-model-based processing. While Communication Manager is acting as a Feature Server, it follows the SIP-ISC conventions, such as phase tags imsorig and imsterm, to communicate with Session Manager.

When the IMS flag is set to **No**, Communication Manager performs Evolution Server processing. If the incoming request contains SIP-ISC parameters, Communication Manager assumes the role of Evolution Server. For example, if an incoming request contains the phase tag imsorig, Communication Manager processes the request per the Evolution Server logic. If the Peer Server is not a Session Manager and the IMS flag is set to **No**, Communication Manager performs the regular SIP request processing.

Related topics:

Communication Manager Feature Server roles on page 120
Communication Manager as Evolution Server on page 121
How does Evolution Server differ from Feature Server and classic Communication Manager? on page 122

Communication Manager Feature Server roles

Communication Manager administered as a Feature Server is Communication Manager's realization of a role within the Avaya common SIP reference architecture that provides Communication Manager features to SIP phones using the Internet Multimedia Subsystem (IMS) half-call model that allows Application Sequencing. As part of Application Sequencing, Session Manager invokes Feature Servers using the SIP IMS Service Control (SIP-ISC) interface. In a single invocation, a sequenced application executes either the originating side or terminating side of the call but not both. Session Manager communicates the requested side (origination or termination) of the call to the sequenced applications using phase tags that Session Manager inserts into the request. When Communication Manager is finished with its origination processing, Communication Manager does not then terminate the call. The Feature Server, instead, passes the call back to Session Manager telling Session Manager that it is done with everything needed for origination, and Session Manager then hands it off to the next application in the sequence.

Session Manager can apply applications to a call before Communication Manager ever sees it—or afterwards for that matter. A Feature Server does not support non-SIP endpoints or traditional application interface like Application Enablement Services. Feature Server only supports IMS-SIP users, which are registered to Session Manager. The Feature Server itself is connected to via SIP signaling groups, which have **IMS enabled?** field set to **yes**. IMS-enabled tells the Feature Server to support the half-call model for the calls and features of the IMS users.

Communication Manager administered as a Trunk Gateway is Communication Manager's realization of a different role in the Avaya common SIP reference architecture for providing interworking for trunk calls between Session Manager's SIP network and non-SIP networks such as ISDN. These could be PSTN or private network calls. A Trunk Gateway is an optional subcomponent of a Feature Server. Special routing is required in the Feature Server and Session Manager to separate these two roles within Communication Manager. The configuration must ensure that the two roles are not internally mixed and that all calls are routed back to Session Manager. For example, an incoming public trunk call to an IMS user (that is, a SIP endpoint) must tandem out to Session Manager on the non IMS signaling group and then come back via the IMS signaling group to the Feature Server part of Communication Manager.

In the dual role of a Feature Server and Trunk Gateway, a Feature Server must have two connections to Session Manager to support these different roles. One is the ISC/IMS-enabled signaling group for the Feature Server part of Communication Manager and the other is the non-ISC signaling group to support the Trunk Gateway part. The IMS and ISC capabilities and

administration are explained below. Note that not all current Communication Manager features are defined in this new model for either phone features or trunk features.

Some rules for the Feature Server:

- Endpoints: supports only SIP endpoints.
- Trunks: supports only SIP trunks to Session Manager.
- Traffic: supports intra-Feature Server (users on same Feature Server) and inter-Communication Manager (users on different Communication Managers) calls.
- Dual-role with Communication Manager administered as a Trunk Gateway (optional)
- Both Feature Server and Trunk Gateway (if configured) are located on the same Communication Manager. These roles must separated; there is no administrative check.
- Supports only the following calls:
  - IMS to IMS (Feature Server)
  - non-IMS trunk to non-IMS trunk (Trunk Gateway)
- Session Manager number adaptation is not allowed for Feature Server but is for Trunk Gateway.

Communication Manager as Evolution Server

Communication Manager administered as an Evolution Server is a variant of Communication Manager that fulfills the needs of Avaya's solutions where SIP endpoint support is required along with the traditional endpoints (for example, H.323) and traditional applications (for example, Application Enablement Services). The Evolution Server offers a migration path for the SIP Enablement Services/Communication Manager customers to a Session Manager/ Communication Manager-based environment. The Evolution Server implements the functionality of classic Communication Manager. The Evolution Server's functionality allows classic Communication Manager to integrate with Session Manager using the traditional full-call model. This contrasts with Communication Manager administered as a Feature Server that integrates with Session Manager using the half-call model. At the interface level, Evolution Server and Feature Server both follow the SIP-ISC (IMS Service Control) conventions to communicate with Session Manager.

The Evolution Server uses the same set of conventions to interface with Session Manager while following the full-call-model-based call processing. The Evolution Server uses an approach to make Session Manager consider it is communicating with a half-call-model-based application.

To shield Evolution Server's full-call-model processing from Session Manager and not break the conventions that Session Manager uses to communicate with the half-call model based applications, the Evolution Server uses a unique approach. With this approach, the Evolution Server continues to process the incoming request using the traditional full call logic even if

Session Manager requested the originating-side processing. Before handing the request back over to Session Manager, the Evolution Server inserts a special tag in this request to avoid a duplicate invocation. When Session Manager invokes the Evolution Server again to process the terminating side of the call, the Evolution Server looks at the special tag it inserted during the first invocation. If the tag is present, the Evolution Server forwards the request to Session Manager (indicating to Session Manager a successful terminating-side processing). During the second invocation, the Evolution Server does not process the call and bounces the request at the interface level only.

How does Evolution Server differ from Feature Server and classic Communication Manager?

At a high-level, a Communication Manager administered as an Evolution Server is viewed as a hybrid of a Feature Server and a classic Communication Manager.

On an Evolution Server, SIP endpoints and all kinds of stations and trunks are supported and can communicate with each other. The establishment of calls between non-SIP endpoints are handled in the same way as is done in classic Communication Manager. Session Manager is not contacted for routing purposes. But the establishment of calls between SIP endpoints and non-SIP endpoints is handled differently. As the SIP endpoints register with Session Manager only – and not directly visible to Communication Manager administered as an Evolution Server, for calls involving the SIP endpoints, the Evolution Server hands over the requests to Session Manager.

In brief, the Evolution Server offers the following support for different types of endpoints:

- SIP endpoints connect with Session Manager. Session Manager provides SIP services like registrar, location service, and outbound proxy to Evolution Server SIP endpoints. Session Manager invokes the Evolution Server as a sequenced application for the Evolution Server SIP endpoints.
- Non-SIP endpoints connect with the Evolution Server directly.

Whereas the Feature Server is based on the half-call model, the Evolution Server is based on the traditional full-call model (a modified traditional full-call model). The Feature Server integration with Session Manager allows full application sequencing support (that is, multiple applications in a sequence), whereas the Evolution Server integration with Session Manager allows only a single application (that is, Communication Manager) in the application sequence. From Session Manager's perspective, it views both the Evolution Server and the Feature Server as half-call-model-based applications.

Comparing the Evolution Server with classic Communication Manager, both of them differ at the interface conventions with Session Manager. While classic Communication Manager integrates with Session Manager using the traditional SIP trunk interface , the Evolution Server allows the traditional SIP trunk as well as the SIP-ISC interface.

Voice and multimedia networking

Intelligent networking and call routing

With the advent of Session Manager, Communication Manager servers now use SIP trunks across a Session Manager network to communicate between switches without the need for dedicated leased lines. SIP trunks can use QSIG Services called QSIP to extend feature transparency, centralized voice mail, centralized attendant service, Call Center applications, and enhanced call routing across SIP trunks.

IP port network and Branch Gateway connectivity

IP port network connectivity allows servers and port networks and Branch Gateways to be connected over IP networks. Communication Manager uses a proprietary method to package signaling messages over IP. This method allows deployment of communications systems throughout a customer's data network.

H.248 gateway control

Communication Manager uses the standards-based H.248 gateway control protocol to perform call control of Avaya Branch Gateways. H.248 defines a framework of call control signaling between the intelligent servers and multiple Branch Gateways. H.248 controls both IP (H.323) and non-IP connections into a Branch Gateway. H.248 has been extended by Avaya to also tunnel proprietary CCMS messages to allow for enhanced call handling.

Communication Manager gatekeepers

A gatekeeper is an H.323 entity on the network that provides address translation and controls access to the network for H.323 endpoints. For Communication Manager platforms, these are the Linux-based servers. H.323 RAS (Registration, Admission, and Status) protocol messages are exchanged between the server and the IP endpoints for the endpoint registration.

All H.323 voice applications (IP softphones, IP Agents, and IP telephones) register with an Avaya gatekeeper before any calls are attempted. Communication Manager sets up call signaling (Q.931) and call control (H.245) channels from endpoints to the gatekeeper. This allows Communication Manager to provide many of its calling features to H.323 calls.

Related topics:

Registration and alternate gatekeeper list on page 124

Registration and alternate gatekeeper list

The RAS protocol is used by the IP endpoint to discover and register with the Communication Manager gatekeeper.

When registration with the original gatekeeper (C-LAN, Linux-based server) IP address is successful, Communication Manager sends back the IP addresses of all the gatekeepers or survivable remote servers in the IP telephone's network region. These addresses are used if the call signaling to the original gatekeeper IP address fails. Figure 15: Discovery and registration process to the gatekeeper on page 124 shows the registration process.

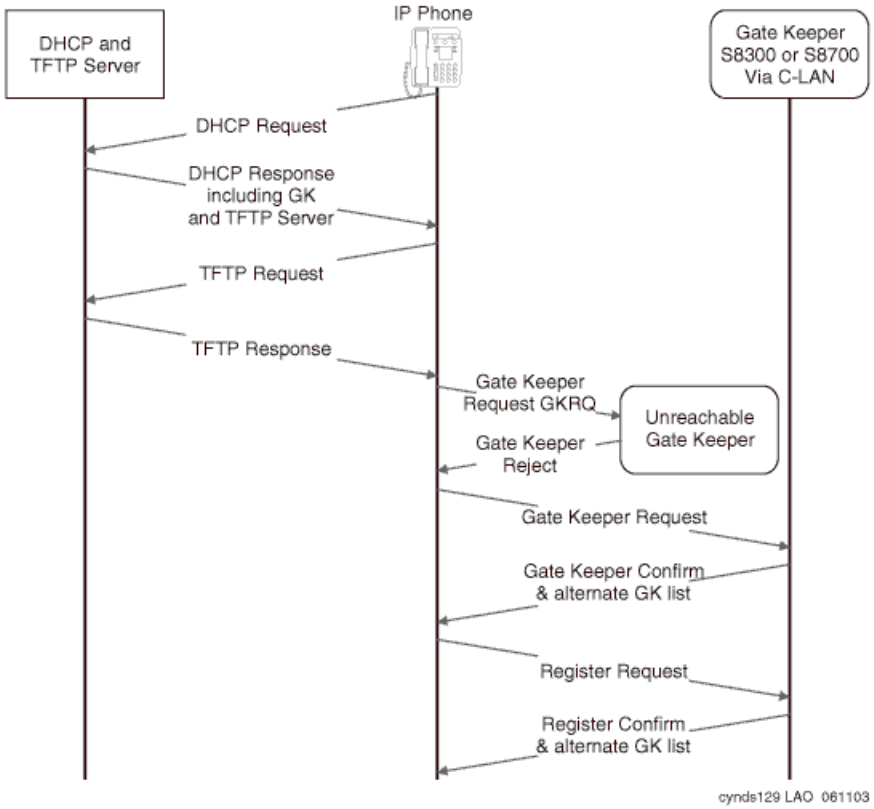


Figure 15: Discovery and registration process to the gatekeeper

Call signaling

Communication Manager implements the gatekeeper routed call model of H.323. The registration process allows the endpoint and the Communication Manager gatekeeper to exchange addresses to establish a TCP connection for a call-signaling channel (the H.323/

H.225 channel). Once the TCP connection is established for call signaling, the H.225.0/Q.931 signaling protocol is used over that connection to route the call and exchange addresses necessary to establish a second TCP connection. This second TCP connection is used for media control (the H.245 channel).

When Communication Manager chooses to route the media flow streams through the system, it selects and allocates available media processor resources and sets the corresponding circuit packs up to receive and send the media stream or streams from/to the endpoints using the negotiated capabilities for each terminal. Each terminal is told to send its media stream or streams to the appropriate media processor circuit pack. The switch connects the two media streams and thus completes the bearer path between the terminals.

Media stream handling

Media processor circuit packs (VoIP resources)

The basic functions of the TN2302AP IP Media Processor and TN2602AP IP Media Resource 320 circuit packs, and VoIP on the Branch Gateways, include:

- Taking media streams off the IP network, terminating RTP/UDP (adjusting for variable delay in arrival rate), and converting them into PCM audio for transmission on the TDM bus.
- Taking media streams from the TDM bus, encoding them with the proper codec, and transmitting them as RTP packets to an IP endpoint.
- Originating and terminating an RTCP control channel for each media stream.
- Encryption and decryption of media

The particulars of the media conversion that is to be performed on each media stream are controlled by Communication Manager. The Quality of Service (QoS) information obtained from the RTCP channel is passed from the circuit pack to Communication Manager.

DTMF tone handling

The media processor circuit pack listens for and detects DTMF tones coming from the TDM bus, strips them out of the audio stream, and sends a message to the server indicating that it has done so. The server in turn generates and sends the appropriate H.245 tone message to the endpoint that is receiving the audio stream. The receiving endpoint then plays the specified tone. Compressed codecs, such as G.729, generally do a poor job of passing DTMF tones. By sending tones out of band, fidelity is maintained. This method is useful when connecting to a voice mail or an integrated voice response (IVR) system, where DTMF digits are used to navigate through prompts.

When this capability is used on an H.323 tie trunk between Communication Manager servers, the server that receives the H.245 tone message plays the required tone onto all the ports receiving the audio stream.

Media stream for audio conferencing

When calls between IP endpoints are conferenced, the media streams must be routed through the media processor circuit pack.

Communication Manager allows the audio streams from different parties to come into different media processor circuit packs. Each media processor sends its received signal to the TDM bus in Pulse Code Modulation (PCM) format. All the other processors serving endpoints on the call can then receive and sum the audio signals coming from all parties and send the resultant composite audio stream to the IP parties that it supports.

Figure 16: Media Processor circuit pack support of a 3-party audio conference on page 126 provides an example of how the media processor circuit pack is configured for a 3-party H.323 audio conference using G.729. This conference is conventional in that it uses TDM bus timeslots to allow each party to listen to all of the other parties. Communication Manager balances the available media processing resources, effectively sharing load among multiple media processor circuit packs.

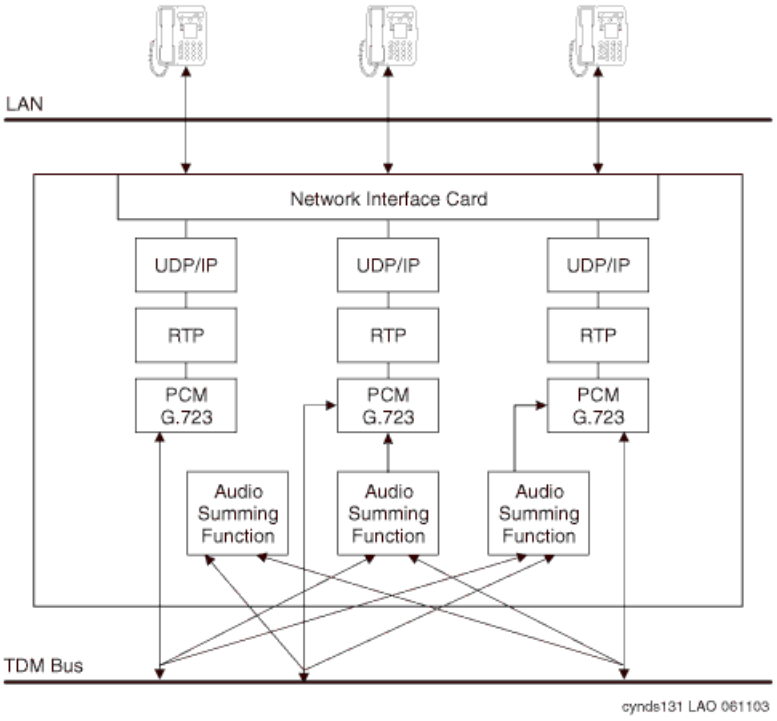


Figure 16: Media Processor circuit pack support of a 3-party audio conference

Separation of Bearer and Signaling (SBS)

In an Avaya IP telephony system, call signaling and bearer traffic may be routed over separate paths. This is useful for a remote branch office with only limited WAN bandwidth back to headquarters. Call signaling traffic can be routed across the WAN, while bearer traffic is sent over the PSTN.

Multilocation

Communication Manager allows a Linux-based server located in one country to control gateways located across national borders and provide appropriate country-specific tones and features. Specifically, these features include the following:

- A-law & Mu-law companding
- Call progress tone generation
- Loss plan
- Analog line board parameters
- Call detail recording
- R2-MFC (multifrequency-signaling) trunks

Multilocation functionality is subject to the following limitations:

- 25 Countries
- R2-MFC: 8 signaling sets
- No per-country alarms or traffic reports are available
- Survivable remote servers must be set to the same time zone as the main server
- This feature is intended for IP-remoted gateways.

Modem/Fax/TTY over IP

In the past, many organizations have experienced problems transporting modem, fax, and TTY tones over an IP Telephony network, regardless of vendor. Modems, faxes, and TTYs are very sensitive to latency and jitter and do not tolerate distortion induced through compression, expansion, and transcoding. To overcome these difficulties, Avaya has enhanced its modem, fax, and TTY-over-IP support.

There are two enhanced modes for supporting Modem-over-IP (MoIP): pass-through and relay. Pass-through is essentially a best-effort transmission and works by forcing the use of the G.711 (uncompressed) codec for the call. Retransmission is governed by the application. Pass-

through mode is suited to LAN environments where both ends of a call are synchronized using a common clock source. As the relay mode does not force the use of G.711, it requires less bandwidth than pass-through but required more DSP resource. So relay uses redundant packet transmission to protect against packet loss. Relay is more effective than pass-through across a WAN.

Avaya's TTYoIP support works by identifying TTY Baudot tones at the near-end media processor, removing them from the voice path and transporting them across the network in RFC 2833 messages. The far-end media processor receives the RFC 2833 messages and regenerates them for the far-end endpoint. By default, this feature is enabled on IP trunks and intergateway calls and is capable of toggling between text and voice modes.

Avaya's support for modem, fax, and TTY over IP is summarized as follows:

- TTY over IP continues to be supported
- Modem Pass-through is supported between Avaya gateways
- Modem Relay at 9.6K is supported between Avaya gateways
- Avaya supports sending multiple instances of the same packet

Redundant transmission mitigates the effects of packet loss, but requires additional bandwidth.

Avaya's modem, fax, and TTY over IP support is subject to the following limitations:

- QoS is required, even on LAN.
- Avoid MoIP where possible (especially over a WAN environment).
- Use circuit-based resources on the same gateway.
- Use different classes of service and restrictions.
- Use centralized modem pooling for larger communities.
- Only one TDM-to-IP-to-TDM conversion is allowed.
- Send duplicate streams, where practical.

Table 10: Fax, Modem, and TTYoIP options on page 128 summarizes Avaya's fax, modem, and TTY options.

Table 10: Fax, Modem, and TTYoIP options

Service	Options	Description
Fax	relay	Default, Avaya-proprietary mode, interoperates with previous releases
	Pass-thru	Proprietary mode; uses more bandwidth, fewer DSP resources
	off	System ignores fax tones, call remains in administered codec
Modem	off	Default, system ignores modem tones, call remains in administered codec

Service	Options	Description
	relay	Avaya-proprietary mode, most reliable modem-over-IP mode
	pass-thru	Similar to Fax pass-thru
TTY	USM	Default, 45.45 Baudot, interoperates with previous releases
	UK	50 Baudot
	pass-thru	Similar to Fax pass-thru
	off	System ignores TTY tones, call remains in administered codec

SIP

SIP stands for Session Initiation Protocol, an endpoint-oriented messaging standard defined by the Internet Engineering Task Force (IETF). SIP is a text-based protocol, similar to HTTP and SMTP, for initiating interactive communication sessions between users. Such sessions include voice, video, instant messaging, interactive games, and virtual reality.

Session Manager offers a core communication service that builds on existing equipment and adds a SIP-based architecture. Session Manager provides SIP routing, registrar, location server, and application sequencing capabilities.

Session Manager connects Communication Manager as a feature (SIP-only) server or evolution (SIP and non-SIP) server, Avaya enterprise PBX and small key PBX systems within branch offices, third-party PBXs, gateways, service providers, SIP-enabled adjuncts, and SIP and non-SIP telephones. Specifically, Session Manager provides the following features:

- Normalizes disparate networks
- Routes SIP sessions across the network
- Integrates with third-party equipment and endpoints
- Offers centralized management, including user profiles, through System Manager
- Supports SIP survivable branches
- Communicates with a session border controller and provides protection at the edge of the enterprise network
- Enables 3rd party E911 emergency call service that supports up to 100,000 users
- Supports direct SIP connectivity with Avaya Aura® Presence Services and makes Avaya one-X® Communicator, Avaya A175 Desktop Video Device with the Avaya Flare™ Experience, and 96XX phones as presence-enabled devices

- Serves as the master control point for Avaya and Polycom video domains
- With Communication Manager and Avaya SIP endpoints, Session Manager provides the ability to search contacts in the enterprise-wide user database for calling, instant messaging, and presence

Each Session Manager installation combines several or all of the following configurations:

- Centralized routing and dial plan management
- Policy-based routing
  - Time of day routing
  - Alternate routing
  - Load balancing
  - Call admission control
- Tail end hop off (TEHO)
- Centralized SIP Trunking
- Centralized applications
  - Registrar, Event State Compositor, Proxy and application sequencing functionality for SIP phones
  - Geographic redundancy for SIP phones
- Sequenced applications

It also handles all call redirection, internal network call accounting feeds, toll bypass, interoffice routing, and international least-cost routing.

IP-based trunks

In circuit-switched networks, trunks provide the means to interconnect PBXs with each other and to the PSTN. Connection to the public network allows PBX station users to call and be called by endpoints that are not part of the PBX private network. An analogous arrangement exists in packet-switched IP networks.

H.323 trunks connect H.323 systems or gateways over IP networks, similar to circuit-switched tie trunks. Similarly, SIP trunks connect SIP systems or gateways over IP networks.

A set of Communication Manager systems can each be attached to an IP network, and voice and fax calls can flow between them in the usual manner except that the call signaling and audio/fax streams are carried over the IP network. The signaling is carried through the C-LAN circuit packs or Processor Ethernet, and the audio and fax streams are carried between systems through the media processor circuit packs.

The benefits of using IP trunks include:

- Reducing long distance voice and fax expenses
- Facilitating global communications
- Providing a fully functional network with data and voice
- Converging and optimizing networks by using the available network resources

IP trunk calls can be compressed to save network bandwidth. Repeated compression and decompression (transcoding) results in a loss of data at each stage and degrades the final quality of the signal. The maximum recommended number of compression cycles on a call is three. Normal corporate voice calls or fax calls typically go through fewer than three compression cycles.

IP (H.323 and SIP) trunks can also connect to other vendors' compliant PBXs.

Related topics:

IP tie trunks on page 131
Trunk signaling on page 131

IP tie trunks

IP tie trunks are used to connect systems to one another. When an IP trunk is used to interconnect two systems, the trunk can also carry standard (QSIG) and proprietary (DCS+) signaling for intersystem feature transparency. The location of each other node (system) in the network is administered, and node selection is based on the dial plan and call routing features, such as AAR/ARS.

H.323 or SIP tie trunks are administered as a new type of trunk group. Instead of administering ports as members of the trunk group, only the number of channels must be specified. Each channel is analogous to a member trunk. In addition, an IP tie trunk can be made a member of a signaling group so that a virtual D-channel can be administered and used to carry feature transparency information.

For SIP trunk capacities, see SIP on page 129.

Trunk signaling

Several variations of IP signaling must be accommodated for the variety of trunks supported by Communication Manager. These are specified as options in the trunk group administration. When the IP trunk is used as a tie trunk to another vendor's PBX, gateway, or gatekeeper, Communication Manager sets up a separate TCP connection for each call.

Note:
The maximum number of members of a single IP trunk signaling group is 255.

Configuration of G860 with Communication Manager

The Avaya G860 Media Gateway provides SIP connectivity to Communication Manager and can work with the G650 Media Gateway or directly with the server's Processor Ethernet interface. This solution is ideal for large IP-based Contact Centers and campuses.

The G860 Media Gateway with Communication Manager provides non blocking SIP-VoIP capacity of up to 6000 channels. In this configuration the G860 Media Gateway eliminates the need for T1/E1 resources by providing multiple DS3, where each DS3 is equivalent to 28 DS1 interfaces.

When configured with a G650 Media Gateway, the following circuit packs are needed according to enterprise traffic.

- C-LAN — for the signaling links between a Communication Manager installed on a server and SIP adjuncts, such as the G860 Media Gateway
- IPSI — for signaling between a Communication Manager installed on a server and other Avaya branch gateways.
- MedPro — for calls that do not shuffle.

Related topics:

Example configuration for a call center on page 132

Example configuration for a call center

The Avaya G860 Media Gateway allows call center customers to consolidate facilities and reduce communications costs. The media gateway concentrates incoming PSTN traffic over several DS3 lines while supporting VoIP telephony in the call center itself.

The use of VoIP and conversion from DS1 to DS3 lines eliminate the large number of DS1 interfaces required to support the same amount of call traffic. The Avaya G860 Media Gateway supports up to 6000 channels of SIP VoIP telephony. It uses N+1 redundancy of media gateway, Ethernet switch, shelf controller, and power supply modules to achieve high availability in mission critical applications.

In the sample call center configuration shown in Figure 17: Sample call center configuration using the Avaya G860 Media Gateway on page 133, a simulated PSTN delivers customer calls using a DS3 interface to the Avaya G860 Media Gateway. The media gateway establishes calls with the server via SIP signaling and routes all Real-time Transport Protocol (RTP) traffic

to the appropriate media resources within an Avaya branch gateway or Avaya endpoints. Communication Manager delivers the calls to an agent phone.

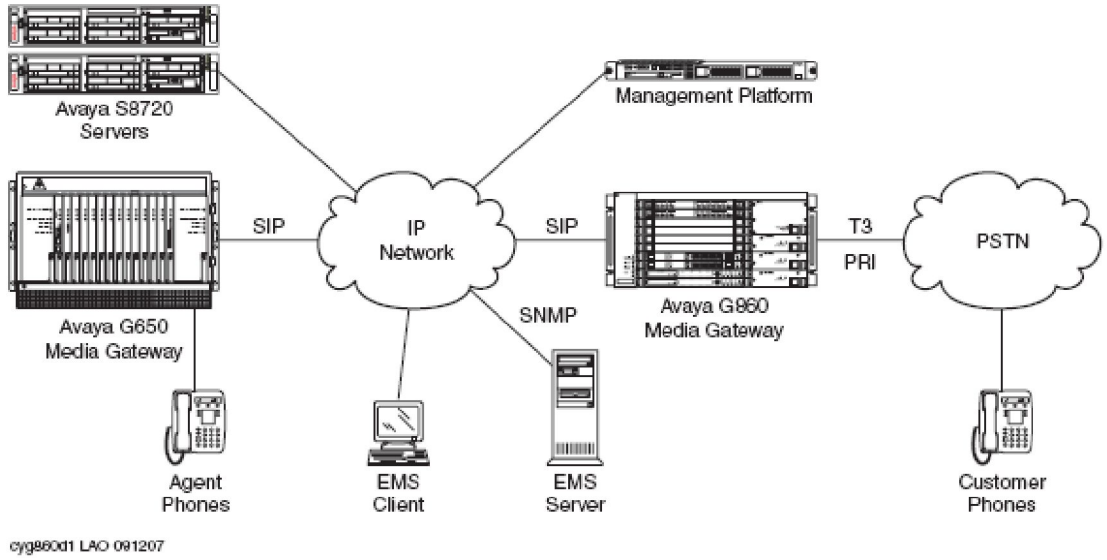


Figure 17: Sample call center configuration using the Avaya G860 Media Gateway

Agents can also make outbound calls using the same network. In the sample configuration, multiple TN799DP C-LAN circuit packs support alternate routing and permit load sharing of calls delivered by the Avaya G860 Media Gateway.

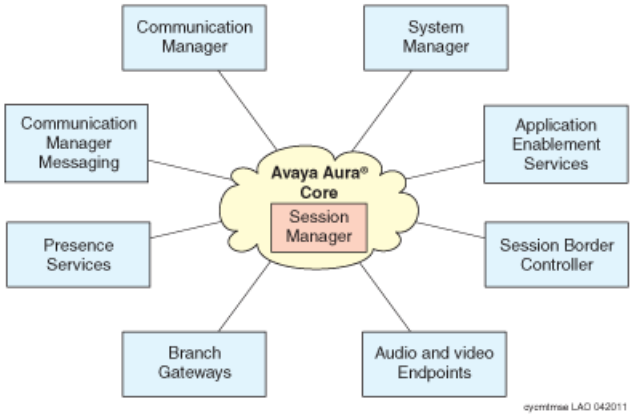
Chapter 7: Sample deployments

Each Avaya Aura® solution is customized for the customer. We are including some example deployments to help you determine what is needed to deploy the Avaya Aura® solution in the filed.

Deployments are basically two kinds: Greenfield deployments, where it is a brand new setup with hardware and applications or an enhancement to an existing Avaya setup.

Sample standard edition deployment

The Standard Edition architecture is for mid-to-large enterprises. The following graphic depicts a typical architecture for a mid-size enterprise deployment.



This particular architecture can be deployed in two ways: Each application is installed on separate servers, or all the applications are installed on one server. The latter deployment is called the Solution for Midsize Enterprise.

Applications

For this particular architecture, you need the following applications:

- Session Manager—The application is the Avaya Aura® core and includes the Linux operating system.
- Communication Manager/Communication Manager Messaging—The application includes the Linux operating system.
- Application Enablement Services—The application includes the Linux operating system.

- System Manager—The application is the sole management interface for Session Manager and Presence. It is also used to manage portions of Communication Manager, Communication Manager Messaging, Application Enablement Services, and Avaya Aura® Session Border Controller. It includes the Linux operating system.
- Avaya Aura® Session Border Controller—The application includes the Linux operating system.
- System Platform—The application includes the Linux operating system and is required for Communication Manager/Communication Manager Messaging, Application Enablement Services, System Manager, and Presence.
- Presence—The application includes the Linux operating system.
- (Optional) Solution for Midsize Enterprise—The application contains all of the above applications in one template.

Hardware

For this particular architecture, you need the following hardware:

- For these particular applications, you use one of the following Avaya-provided servers. If the application supports high availability (duplicated servers), then you need two of the same servers; you cannot mix servers. The servers are shipped with the correct memory and components for the application.
  - Dell™ PowerEdge™ R610
  - HP ProLiant DL360 G7
- For this particular architecture, you use the following Avaya-provided gateways. These gateways provide media services and the bearer network for telephones.
  - Avaya G650 Media Gateway if using port networks with Communication Manager. This media gateway includes TN circuit packs. Required circuit packs for IP connectivity are the TN2312BP IP Server Interface (IPSI), TN799DP Control LAN (C-LAN), TN2302AP IP Media Processor (MedPro) or TN2602AP Media Resource 320.
  - Avaya G430 Branch Gateway This branch gateway includes media modules.
  - Avaya G450 Branch Gateway This branch gateway includes media modules.

Audio and video endpoints

For this particular architecture, you need the following Avaya-provided audio and video endpoints.

- IP telephones
  - 11xx-series
  - 12xx-series
  - 16xx-series
  - 2007
  - 46xx-series (Broadcom and Agere)
  - 96xx-series

- 46xx-series (Broadcom and Agere)
- IP wireless (3641, 3645, 612 WLAN, 6140 WLAN)
- IP conference (1692, 2033)

• SIP telephones

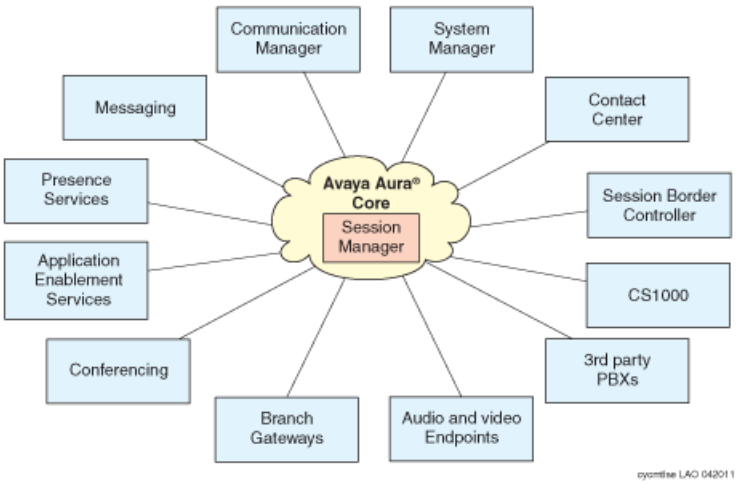
- 11xx-series
- 96xx-series

• Video devices

- Avaya 1000-series
- Avaya A175 Desktop Video Device

Sample Enterprise Edition deployment

The Enterprise Edition is for highly distributed enterprises. The following graphic depicts a typical architecture.



Applications

For this particular architecture, you need the following applications:

- Session Manager—The application is the Avaya Aura® core and includes the Linux operating system.
- Communication Manager/Communication Manager Messaging—The application includes the Linux operating system.
- Application Enablement Services—The application includes the Linux operating system.
- Avaya Aura® Messaging—The application includes the Linux operating system.

- Conferencing—The application includes
- Avaya Aura® Contact Center—The application includes
- CS 1000—The application includes
- System Manager—The application is the sole management interface for Session Manager and Presence. It is also used to manage portions of Communication Manager, Communication Manager Messaging, Application Enablement Services, CS 1000, and Avaya Aura® Session Border Controller. It includes the Linux operating system.
- Avaya Aura® Session Border Controller—The application includes the Linux operating system.
- System Platform—The application includes the Linux operating system and is required for Communication Manager/Communication Manager Messaging, Application Enablement Services, System Manager, and Presence.
- Presence—The application includes the Linux operating system.
- (Optional) Solution for Midsize Enterprise—The application contains all of the above applications in one template.

Third-party PBXs

The Enterprise Edition may include third-party PBXs. The following PBXs are known to work within the solution:

Cisco Unified CallManager

Hardware

For this particular architecture, you need the following hardware:

- For these particular applications, you use one of the following Avaya-provided servers. If the application supports high availability (duplicated servers), then you need two of the same servers; you cannot mix servers. The servers are shipped with the correct memory and components for the application.
  - Dell™ PowerEdge™ R610
  - HP ProLiant DL360 G7
- For this particular architecture, you use the following Avaya-provided gateways. These gateways provide media services and the bearer network for telephones.
  - Avaya G650 Media Gateway if using port networks with Communication Manager. This media gateway includes TN circuit packs. Required circuit packs for IP connectivity are the TN2312BP IP Server Interface (IPSI), TN799DP Control LAN (C-LAN), TN2302AP IP Media Processor (MedPro) or TN2602AP Media Resource 320.
  - Avaya G430 Branch Gateway This branch gateway includes media modules.
  - Avaya G450 Branch Gateway This branch gateway includes media modules.

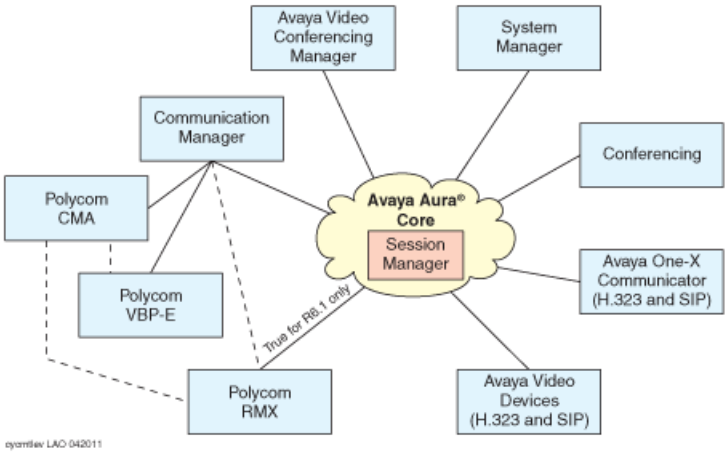
Audio and video endpoints

For this particular architecture, you need the following Avaya-provided audio and video endpoints.

- IP telephones
  - 11xx-series
  - 12xx-series
  - 16xx-series
  - 2007
  - 46xx-series (Broadcom and Agere)
  - 96xx-series
  - 46xx-series (Broadcom and Agere)
  - IP wireless (3641, 3645, 612 WLAN, 6140 WLAN)
  - IP conference (1692, 2033)
- SIP telephones
  - 11xx-series
  - 96xx-series
- Video devices
  - Avaya 1000-series
  - Avaya A175 Desktop Video Device

Sample Enterprise Edition video deployment

The Enterprise Edition may offer the video architecture. The following graphic depicts a typical video architecture.



Applications

For this particular architecture, you need the following applications:

- Session Manager—The application is the Avaya Aura® core and includes the Linux operating system.
- Communication Manager—The application includes the Linux operating system.
- Conferencing Manager
- System Manager—The application is the sole management interface for Session Manager. It is also used to manage portions of Communication Manager and Avaya Aura® Session Border Controller. It includes the Linux operating system.
- Avaya Aura® Session Border Controller—The application includes the Linux operating system.
- System Platform—The application includes the Linux operating system and is required for Communication Manager and System Manager.
- For this particular architecture, you may use the following third-party applications:
  - Polycom CMA
  - Polycom VBP-E
  - Polycom RMX
- Avaya one-X® Communicator (H.323 and SIP versions)

Hardware

For this particular architecture, you need the following hardware:

- For the Avaya applications, you use one of the following Avaya-provided servers. If the application supports high availability (duplicated servers), then you need two of the same servers; you cannot mix servers. The servers are shipped with the correct memory and components for the application.
  - Dell™ PowerEdge™ R610
  - HP ProLiant DL360 G7
- For the Polycom applications, you use Polycom-recommended servers.
- For this particular architecture, you use the following Avaya-provided gateways. These gateways provide media services and the bearer network for telephones.
  - Avaya G650 Media Gateway if using port networks with Communication Manager. This media gateway includes TN circuit packs. Required circuit packs for IP connectivity are the TN2312BP IP Server Interface (IPSI), TN799DP Control LAN (C-LAN), TN2302AP IP Media Processor (MedPro) or TN2602AP Media Resource 320.
  - Avaya G430 Branch Gateway This branch gateway includes media modules.
  - Avaya G450 Branch Gateway This branch gateway includes media modules.

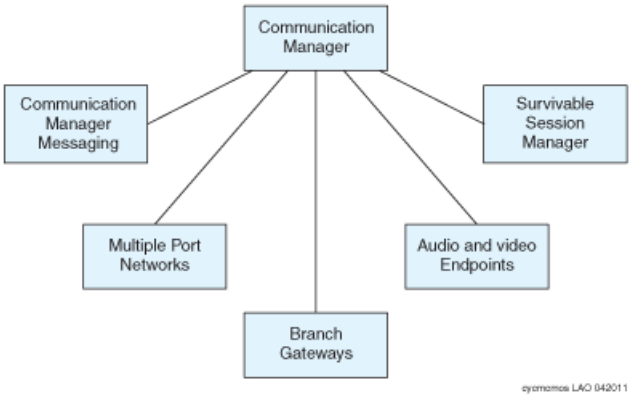
Audio and video endpoints

For this particular architecture, you need the following Avaya-provided audio and video endpoints.

- IP telephones
  - 11xx-series
  - 12xx-series
  - 16xx-series
  - 2007
  - 46xx-series (Broadcom and Agere)
  - 96xx-series
  - 46xx-series (Broadcom and Agere)
  - IP wireless (3641, 3645, 612 WLAN, 6140 WLAN)
  - IP conference (1692, 2033)
- SIP telephones
  - 11xx-series
  - 96xx-series
- Video devices
  - Avaya 1000-series
  - Avaya A175 Desktop Video Device

Sample Communication Manager only deployment

The Communication Manager architecture is for customers that do not need the capabilities offered in a complete Avaya Aura® solution. The following graphic depicts a typical architecture for a Communication Manager deployment.



Applications

For this particular architecture, you need the following applications:

- Communication Manager/Communication Manager Messaging—The application includes the Linux operating system.
- System Platform—The application includes the Linux operating system and is required for Communication Manager/Communication Manager Messaging.

Hardware

For this particular architecture, you need the following hardware:

- For these particular applications, you use one of the following Avaya-provided servers. If the application supports high availability (duplicated servers), then you need two of the same servers; you cannot mix servers. The servers are shipped with the correct memory and components for the application.
  - Dell™ PowerEdge™ R610
  - HP ProLiant DL360 G7
- For this particular architecture, you use the following Avaya-provided gateways. These gateways provide media services and the bearer network for telephones.
  - Avaya G650 Media Gateway for port networks. This media gateway includes TN circuit packs. Required circuit packs for IP connectivity are the TN2312BP IP Server Interface (IPSI), TN799DP Control LAN (C-LAN), TN2302AP IP Media Processor (MedPro) or TN2602AP Media Resource 320.
  - Avaya G430 Branch Gateway This branch gateway includes media modules.
  - Avaya G450 Branch Gateway This branch gateway includes media modules.

Audio and video endpoints

For this particular architecture, you need the following Avaya-provided audio and video endpoints.

- IP telephones
  - 11xx-series
  - 12xx-series
  - 16xx-series
  - 2007
  - 46xx-series (Broadcom and Agere)
  - 96xx-series
  - 46xx-series (Broadcom and Agere)
  - IP wireless (3641, 3645, 612 WLAN, 6140 WLAN)
  - IP conference (1692, 2033)
- SIP telephones
  - 11xx-series

- 96xx-series

• Video devices

- Avaya 1000-series
- Avaya A175 Desktop Video Device

Greenfield deployment

A Greenfield site is a business or an organization that does not have an existing communication system. Most Greenfield systems are deployed into new businesses and organizations, and these systems tend to be smaller in size. Occasionally, an established large organization may completely remove its existing system and install a new system. In these cases, the incumbent system is usually a leased service, such as a centrex service from a telephony service provider.

In general, most organizations want to protect their investment in their PBX communications system. Avaya provides ways for our circuit-switched PBX customers to evolve from circuit-switched systems to IP-enabled systems. This solution provides most of the advantages of IP telephony with minimal equipment upgrades to an enterprise's existing PBX.

Components needed for Greenfield deployment

In a Greenfield deployment, the primary connection medium is IP. To provide the greatest flexibility and the lowest costs for a converged solution, most endpoints should be SIP or IP telephones or soft clients (Soft clients would include IP Softphone, IP Agent, one-X Communicator and one-X Agent). A mixture of IP endpoints and circuit-switched endpoints places increased demand on media processor resources and thus increases the cost of the deployment. Intersite communications should also be IP based. This is done either through direct connections between SIP and IP telephones or through SIP and IP trunks. Circuit-switched or TDM-based communications should be kept to a minimum. The primary TDM connections should be for PSTN access, where necessary, and connections to any analog telephones, modems, or fax machines that exist. The figure below shows a typical Greenfield architecture..

Servers as IP-endpoint gatekeepers

The servers are responsible for running Communication Manager and controlling the branch gateways and endpoints. The servers control the dial plan translations and call routing, call setup and teardown, Call Detail Record (CDR) generation, and traffic management. The servers also offer IP-endpoints gatekeeper functionality, and provide the extensive telephony features that are included with Communication Manager. The Processor Ethernet (PE)

component of the server provides gatekeeper functionality for IP-endpoints. C-LANs can also provide this capability in port networks.

Avaya currently supports the following servers:

- Avaya S8300D Server (server resides in a branch gateway)
- Avaya S8800 Server
- HP ProLiant DL360 G7
- Dell™ PowerEdge™ R610

Communication Manager capabilities

Communication Manager IP capabilities and applications support voice over an IP network and ensure that remote workers have access to communication system features from their personal computers. Communication Manager also provides standards-based control between servers and branch gateways, which allows the communications infrastructure to be distributed to the edge of the network. The Communication Manager IP engine offers features that enable users to increase the quality of voice communications. Quality of Service (QoS) features enable users to optimize voice quality by assisting some routers in prioritizing audio traffic. Communication Manager media processors allow for hairpinning and shuffling. These features make voice communications more efficient by reducing both per-port costs and IP bandwidth usage. Avaya IP telephony solutions support trunks, IP communications devices, IP port networks, and IP control for branch gateways.

Avaya IP telephony solutions are implemented using various IP media processor circuit packs inside the port networks. These media processors provide H.323 trunk connections and H.323 voice processing for IP telephones and softphones. H.323 signaling is handled by a C-LAN circuit pack or native Processor Ethernet connectivity. The IP network can be extended across geographically disparate locations. With Communication Manager ISDN, Distributed Communication Services (DCS+), or QSIG services, Communication Manager can extend feature transparency, centralized voice mail, centralized attendant service, call center applications, and enhanced call routing across IP trunks.

For more information on Communication Manager architecture, see Call processing.

Branch gateways and port networks

Avaya branch gateways and port networks support voice and signaling traffic that is routed between circuit-switched networks and packet-switched networks. The branch gateways and port networks support all the applications and adjuncts that are supported by Communication Manager, accommodating call center and customer relationship management applications, messaging, remote workers, and remote offices. Avaya branch gateways and port networks work with standards-based IP networks and connect easily with the public switched telephone network (PSTN). The IP network infrastructure provides support for the communication between the servers and the gateways and port networks.

In a Greenfield installation, the recommended gateways are the Avaya G430 and G450 Branch Gateways. The Avaya G650 Media Gateway is used for port networks. The Avaya G650 Media Gateway houses traditional circuit switch boards and boards that support IP telephony. The Avaya G430 and G450, Branch Gateways house media modules that provide ports for non-IP endpoints, including analog and DCP telephones.

Chapter 8: Security

Security philosophy

This section describes the security-related considerations, features, and services for the Avaya Aura® solution and its various components. Avaya Aura® needs to be resilient to attacks that can cause service disruption, malfunction or theft of service. Avaya‘s products inherit a number of mechanisms from legacy communications systems to protect against toll fraud or the unauthorized use of communications resources. However, Unified Communications capabilities, which converge telephony services with data services on the enterprise data network, have the additional need for protections previously specific only to data networking. That is, telephony services need to be protected from security threats such as:

- Denial of Service (DoS) attacks
- Malware (viruses, worms and other malicious code)
- Theft of data
- Theft of service

To prevent security violations and attacks, Session Manager uses Avaya‘s multilayer hardening strategy:

- Secure by design
- Secure by default
- Secure communications

For more information on security design for the various Avaya Aura® components, see the following books:

- *Avaya Aura® Session Manager Security Design*
- *Avaya Aura® Communication Manager 6.1 Security Guide*
- *Avaya Aura® System Manager Security Design*
- *Avaya Aura® Messaging Security Design*
- *Avaya Aura® Presence Services 6.0 Security Design*

Secure by design

Secure by design encompasses a secure deployment strategy that separates Unified Communications (UC) applications and servers from the enterprise production network. Since all SIP sessions flow through the core (Session Manager), being the SIP routing element, it is able to protect the UC applications and servers from network and transport Denial of Service (DoS) attacks, SIP DoS attacks as well as protect against other malicious network attacks. For customers that deploy SIP trunks to SIP service providers, Avaya recommends the use of Avaya Aura® Session Border Controller to provide an additional layer of security between the SIP service provider and Session Manager.

The architecture is related to the trusted communication framework infrastructure security layer and allows for the specification of trust relationships and the design of dedicated security zones for:

- Administration
- Gateway control network
- Enterprise network
- Adjuncts
- SIP Elements

For Communication Manager, Avaya isolates assets such that each of the secure zones is not accessible from the enterprise or branch office zones. The zones are like dedicated networks for particular functions or services. They do not need to have access from or to any other zones because they only accommodate the data they are built for. This provides protection against attacks from within the enterprise and branch office zone.

Gateways with dedicated gatekeeper front-end interfaces (C-LAN) inspect the traffic and protect the server zone from flooding attacks, malformed IP packets, and attempts to gain unauthorized administrative access of the server through the branch gateways. This architecture and framework can also flexibly enhance the virtual enterprise and integrate branch offices into the main corporate network. The security zone from the branch office can terminate at the central branch gateway interfaces, again protecting the heart of Communication Manager.

Secure by default

Secure by default incorporates a hardened Linux operating system with inherent security features for Unified Communications applications and servers. This hardened operating system provides only those functions that are necessary for securing mission-critical call processing applications, and protect the customers from toll fraud and other malicious attacks.

In many cases, Avaya uses a modified kernel based on the Linux-community offering but changed it for secure, real-time telephony processing.

The Linux operating system limits the number of access ports, services, and executables and helps protect the system from typical modes of attack. At the same time, the reduction of Linux services limits the attack surface.

Secure communications

Secure communications uses numerous features and protocols to protect access to and the transmissions from Avaya communications systems. Avaya uses media encryption to ensure privacy for the voice stream. Alongside media encryption, integrated signaling security protects and authenticates messages to all connected SIP elements, IP telephones, and gateways, and minimizes an attacker's ability to tamper with confidential call information. These features protect sensitive information like caller and called party numbers, user authorization, barrier codes, sensitive credit card numbers, and other personal information that is keyed in during calls to banks or automated retailers.

Critical adjunct connections are also encrypted. IP endpoints additionally authenticate to the network infrastructure by supporting supplicant 802.1X protocols. Network infrastructure devices like gateways or data switches act as an authenticator and forward this authentication request to a customer authentication service.

Chapter 9: Licensing

PLDS Overview

The Avaya Product Licensing and Delivery System (PLDS) provides customers, Avaya Partners, distributors, and Avaya Associates with tools for managing license entitlements and electronic delivery of software and related license files. Using PLDS, you can perform operations such as license activations, license upgrades, license moves, and software downloads.

Installation software packages for Avaya products are available as ISO files on PLDS. Users can download the ISO images to a PC, and choose to either burn a DVD for installation or transfer the ISO file to the target server for installation.

You can check PLDS to determine if a later service pack or software release is available. If updates do exist, you should refer to the appropriate upgrade procedures, contact Avaya, or contact the Avaya Partner Service representative.

When you place an order for a PLDS-licensed software product, the license entitlements on the order are automatically created in PLDS. Once these license entitlements are created, you receive an e-mail notification from PLDS. This e-mail notification includes a license activation code (LAC). Using the LAC, you can quickly find and activate the newly purchased license entitlements in PLDS. You can then download the license file.

Important:

You must provide the WebLM host ID to activate the license file in PLDS. The WebLM host ID is the MAC address of the server and is obtained from the WebLM Web interface.

Examples of license management tasks that you can perform in PLDS include:

- Adding more license entitlements to an existing activation
- Upgrading a license file to a new major release
- Moving license entitlement activations between license hosts
- Regenerating a license file with an new host ID

Communication Manager license

Obtaining and installing the license file

Use the Avaya Product Licensing and Delivery System (PLDS) to generate and download license files for Communication Manager 6.0 and later Collaboration ServerSolution for

Midsize Enterprise licensing. PLDS is an online, Web-based tool for managing license entitlements and electronic delivery of software and related license files.

Earlier versions of Communication Manager, except Communication Manager 5.2.1 that is part of Avaya Aura® Midsize Business Template, continue to use the Remote Feature Activation (RFA) online tool for license files. Communication Manager 5.2.1 that is part of Avaya Aura® Midsize Business Template uses PLDS to manage licenses. After you obtain the license file, use WebLM to install it. WebLM is a Web-based application for managing licenses and is installed as part of System Platform in the Console Domain.

The license file is an Extensible Markup Language (XML) file. It contains information regarding the product, major release, and license features and capacities.

For Communication Manager 6.0 and later, license files are installed only on the Communication Manager main server. License files are not installed on survivable servers. Survivable servers receive licensing information from the main server.

If you are licensing a duplicated pair configuration, you must install the license file on both servers. The license file is not synchronized from the active server to the standby server.

A 30-day grace period applies to new installations or upgrades to Communication Manager 6.0Collaboration ServerSolution for Midsize Enterprise. You have 30 days from the day of installation to install a license file.

Customer Option features and the license file

Communication Manager has two categories of Customer Option features:

- Unlicensed Customer Option features that are available to all customers with the purchase of Communication Manager 6.0.
- Licensed Customer Option features that customers purchase and are controlled by the license file.

A specific feature in the Communication Manager license file may enable or map to multiple features on the Customer Options form. For example, ASAI Features (FEAT_CM_ASAI_PCKG) in the license file enables multiple ASAI-related features on the Customer Options form, including ASAI Link Core Capabilities and ASAI Link Plus Capabilities. Unlicensed features are available to all customers and are not included in the license file.

The Communication Manager System Management Interface (SMI) provides the ability to enable or disable individual Customer Option features that have an on or off (yes or no) setting. This capability is available only for the Customer Option features to which you are entitled. Features to which you are entitled include unlicensed features plus any licensed features that are entitled based on the license file. This capability does not apply to capacity features and other types of features that do not have an on or off setting.

How Communication Manager acquires licenses from WebLM

At startup, Communication Manager contacts WebLM and requests license release, features, and capacities. WebLM responds to Communication Manager, which then uses the acquired features and capacities to set license permissions in the Communication Manager software.

Communication Manager acquires all of the feature capacity from WebLM, regardless of actual usage. For example, if the Maximum Stations (VALUE_CM_STA) feature is set to 36,000 in the license file, Communication Manager acquires capacity for all 36,000 stations regardless

of the number of stations currently configured. Actual license usage can be viewed on the Customer Options form in the System Administration Terminal (SAT) interface.

Every 9 minutes, Communication Manager sends a request to WebLM to renew its license information. Because of this time interval, you may have to wait up to 9 minutes for a newly installed license file to take effect on Communication Manager.

Communication Manager license features

A specific feature in the Communication Manager license file may enable or map to multiple features on the Customer Options form. For example, ASAI Features (FEAT_CM_ASAI_PCKG) in the license file enables multiple ASAI-related features on the Customer Options form, including ASAI Link Core Capabilities and ASAI Link Plus Capabilities. Unlicensed features are available to all customers and are not included in the license file.

The following table summarizes the mapping of features in the Communication Manager license file to Customer Option features.

License feature	Communication Manager Customer Option features
Edition (VALUE_CM_EDITION)	`Standard` enables all unlicensed Customer Option features. `Enterprise` maps to the Multinational Locations Customer Option feature. Also enables all unlicensed Customer Option features.
Maximum Stations (VALUE_CM_STA)	Maps to multiple Customer Option features, notably Maximum Stations.
Maximum Analog Stations (VALUE_CM_ANALOG)	Specifies the number of analog stations to which the customer is entitled.
Maximum Survivable Processors (VALUE_CM_SP)	Maps directly to the Maximum Survivable Processors Customer Option feature.
Maximum ESS Stations (VALUE_CM_ESS_STA)	Specifies the number of Survivable Core station licenses to which the customer is entitled.
Maximum LSP Stations (VALUE_CM_LSP_STA)	Specifies the number of Survivable Remote station licenses to which the customer is entitled.
Maximum Mobility Enabled Stations (VALUE_CM_MOBILITY)	Maps to multiple Off-PBX Telephones Customer Option features.
Maximum Video Capable IP Softphones (VALUE_CM_VC_IPSP)	Maps to the Maximum Video Capable IP Softphones Customer Option feature.

License feature	Communication Manager Customer Option features
ASAI Features (FEAT_CM_ASAI_PCKG)	Maps to ASAI-related Customer Option features including ASAI Link Core Capabilities and ASAI Link Plus Capabilities.
Maximum Expanded Meet-Me Conference Ports (VALUE_CM_EMMC_PORTS)	Maps to the Maximum Number of Expanded Meet-Me Conference Ports Customer Option feature.
Access Security Gateway (FEAT_CM_ASG)	Maps to the Access Security Gateway Customer Option feature.
IP Endpoint Registration Features (for example, IP_Soft)	Map directly to Customer Option features of the same name, for example, IP_Soft.

Communication Manager Messaging license features

The following table shows the mapping of features in the Communication Manager Messaging license file to features.

License feature	Communication Manager Messaging features
CM Messaging Offer (VALUE_CMM_OFFER)	`EMBEDDED` allows up to 6000 mailboxes. `FEDERAL_MARKET` allows up to 15,000 mailboxes.
Maximum CM Messaging Mailboxes (VALUE_CMM_MAILBOX)	Maps directly to the CM Messaging Mailboxes feature.

Call Center license features

The following table summarizes the mapping of features in the Call Center license file to Customer Option features.

License feature	Call Center Customer Option features
Maximum Elite Agents (VALUE_CC_ELITE)	Maps to multiple Customer Option features, most notably Logged-In ACD Agents.
Maximum Advocate Agents (VALUE_CC_ADVOCATE)	Maps to multiple Customer Option features, most notably Logged-In Advocate Agents.

License feature	Call Center Customer Option features
Proprietary (FEAT_CC_PROPRIETARY)	Maps directly to the Proprietary Customer Option feature (renamed from Agent States in Communication Manager 6.0).
Call Center IP Endpoint Registration Features (for example, IP_Agent)	Maps directly to Customer Option features of the same name, for example, IP_Agent.

Chapter 10: Secure Access Link overview

Secure Access Link (SAL) is an Avaya serviceability solution for support and remote management of a variety of devices and products. SAL provides remote access and alarm reception capabilities. SAL uses the customer's existing Internet connectivity to facilitate remote support from Avaya. All communication is outbound from the environment of the customer and uses encapsulated Hypertext Transfer Protocol Secure (HTTPS).

SAL provides the following features:

- Enhanced availability and reliability of supported products through secure remote access.
- Support for service provision from Avaya, Avaya partners, system integrators, or customers.
- Administration of alarming through configuration changes.
- Elimination of the requirement for modems and dedicated telephone lines at the customer sites.

SAL provides the following security features:

- Communication initiated from customer networks (egress connectivity model).
- Detailed logging.
- Support for PKI-based (Public Key Infrastructure) user certificates for Avaya support personnel to remotely access managed devices.
- Customer-controlled authentication.
- Rich policy-based authorization management.
- Support for local access and management options.
- Reduced firewall and network security configuration.

SAL Gateway overview

SAL Gateway is a software package that:

- Facilitates remote access to support personnel and tools that need to access supported devices.
- Collects and sends alarm information to a Secure Access Concentrator Core Server, on behalf of the managed devices.
- Provides a user interface to configure its interfaces to managed devices, Concentrator Remote and Core Servers, and other settings.

The SAL Gateway is installed on a Linux host in the customer network and acts as an agent on behalf of several managed elements. It receives alarms from products and forwards them to the Secure Access Concentrator Core Server.

The SAL Gateway polls the Secure Access Concentrator Servers for connection requests and authorizes connection requests with the Secure Access Policy Server. The use of the Policy server is optional. The SAL Gateway also sends alarms to the Secure Access Concentrator Core Server as they are received and periodically polls to report availability status.

The SAL Gateway provides remote access to those devices that are configured for remote access within it. It controls connections to managed elements, new or updated models, and verifies certificates for authentication. The SAL Gateway also communicates with a Secure Access Concentrator Remote Server.

Summary of SAL Gateway features

The SAL Gateway user interface is used to administer the following SAL Gateway settings:

- Secure Access Concentrator Remote and Core Server host names
- Proxy Servers
- Managed device connectivity
- Policy server and LDAP authentication.
- Network Management Server details.
- The ability to view SAL Gateway logs.
- SAL Gateway status and diagnostic capabilities.

Other SAL components

Other SAL components include two concentrator servers:

- Secure Access Concentrator Core Server handles alarming
- Secure Access Concentrator Remote Server handles remote access and updates models and configuration.

Secure Access Policy Server

Customers can deploy an optional Secure Access Policy Server that centrally defines and manages access and control policies. Gateways enforce the policies. The SAL Gateway polls the policy server for updates on policies. The policy server provides active monitoring and termination of remote access sessions. The use of the Policy Server is optional; devices can still be serviced through the Avaya SAL solution, even if the customer does not install a Policy Server. However, a Policy Server provides flexibility and control to the customer. Through it, the customer establishes, and controls Avaya Secure Access Link permissions for the devices within the customer's network. When a Policy Server is installed and one or more SAL Gateways are configured to use the policies from the Policy Server, the customer can:

- Control who accesses their devices
- Control when the devices are accessed
- Control what remote session types (protocols) can be employed
- Monitor activity, with the ability to terminate any or all remote access sessions on an on-demand basis.

Besides controlling remote access, the Policy Server provides controls over other activities that can be initiated from the upstream servers for which the SAL Gateways communicate. The list of managed operations includes:

- Device-specific actions, for example, restarting a device or executing an application on a device
- Remote access connections to a device
- File uploads
- Script registration and execution
- Package execution

For more information on the Policy server, see *Avaya Secure Access Link, Secure Access Policy Server: Installation and Maintenance Guide*.

While policy decisions can be made in the SAL Gateway or the Secure Access Policy Server, it is the SAL Gateway that enforces all policies.

The policy server can support up to 500 managed devices, regardless of how many gateways are used. The combination can have many variations:

- One gateway with 500 managed devices.
- 100 gateways with the gateway and four additional managed devices each.
- 250 gateways, each with only the gateway and one managed device.
- 500 gateways, each with no managed devices.

How the SAL components work

The SAL Gateway relays alarms and heartbeats to the Secure Access Concentrator Core Server. A SAL Gateway can collect alarms through the receipt of Simple Network Management Protocol (SNMP) traps or the receipt of Initialization and Administration System (INADS) alarms. It provides the collected alarm information to the upstream Secure Access Concentrator Core Enterprise Server (see Figure 18: SAL Components on page 160).

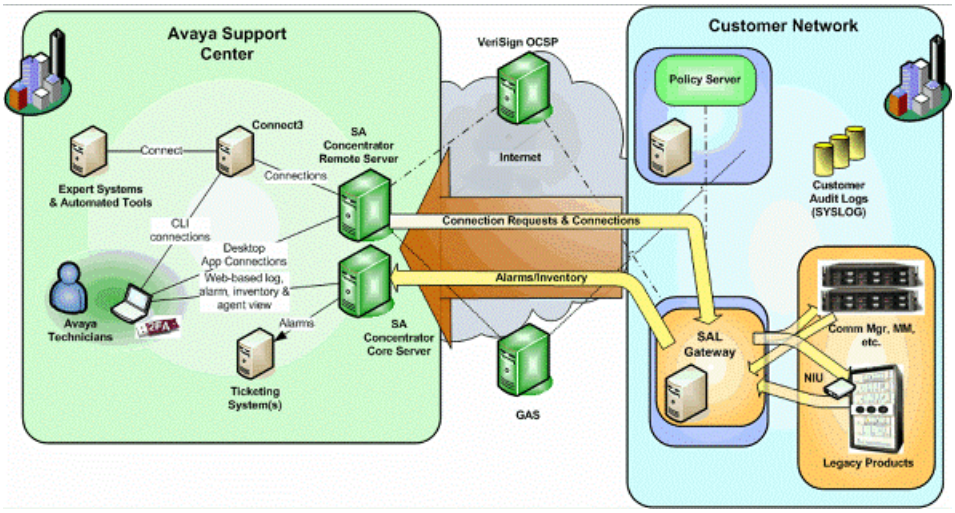


Figure 18: SAL Components

SAL provides remote access to managed devices through HTTPS requests originating inside a customer network. SAL Gateway customers have ultimate control over all SAL-facilitated access to their devices. All connectivity is originally established from the network of the customer, and customer-controlled SAL components enforce authorizations.

When a request for remote access reaches the Avaya Secure Access Concentrator Remote Enterprise Server, the request is sent to the gateway that authenticates the user and determines if the connection should be authorized.

The SAL Gateway frequently polls the Secure Access Concentrator Remote Server to determine if there are any remote access requests for it. If there is a request for remote access, the gateway consults local policy, either provided by a Policy server, or directly configured within, to check whether to allow the remote access request. The SAL Gateway does the authorization. If the policy permits access, it establishes end-to-end connection for remote access from the computer that initiated the request to the managed device.

NTP

The SAL Gateway uses Network Time Protocol (NTP) to synchronize its clock with the other SAL components over the network. NTP provides stability and reliability for remote access to devices. The SAL certificate-based authentication mechanisms rely on accurate clocks to check the expiration and signatures of the remote access requests. If the Gateway host does not use NTP, remote access to service the gateway or any managed device becomes unreliable.

SAL egress model

As egress filtering is considered an important best practice, SAL provides an egress model of remote access that includes customer policy management of remote access, file transfers, and egress data flow. This gives the customer complete control over whether access to their devices is permitted or not. All connectivity is fundamentally established from the network of the customer. As SAL provides egress communication from the SAL Gateway, customers need not expose the gateway with open ports on the Internet. SAL supports the following TCP protocols: SSH, HTTPS, telnet, sftp, ftp, and RDC.

Chapter 11: Survivability, redundancy, and recovery

Survivability, redundancy, and recovery

The Avaya Aura® solution offers several methods to ensure its reliability. Avaya has a long-standing commitment to high availability in hardware and software design and the architectural strength.

This section describes availability and its significance to a communications system. Hardware-design considerations, software-design and recovery considerations, and IP and SIP telephone and remote branch gateway recovery. The reliability methods include duplicated systems and backup systems available if there is a problem with the main system or a network outage.

This section covers the following methods:

- Availability
- Reliability
- Survivability
- Redundancy
- Recovery after an outage

Reliability

Customers need the full reliability of their traditional voice networks, including feature richness and robustness, and they want the option of using converged voice and data infrastructures. With the convergence of voice and data applications that run on common systems, a communications failure could bring an entire business to a halt. Enterprises are looking to vendors to help them design their converged infrastructure to meet their expected availability level.

Communication Manager reliability

Communication Manager supports a wide variety of servers, gateways and survivability features enabling maximum availability for any customer. The software is capable of mirroring processor functions, providing alternate gatekeepers, supporting multiple network interfaces and ensuring survivability at remote and central locations.

Communication Manager reliability features include:

- Alternate gatekeeper. The alternate gatekeeper can provide survivability between Communication Manager and IP communications devices such as IP telephones and IP softphones.
- Auto fallback to primary for branch gateways. This feature automatically returns a fragmented network, where a number of branch gateways are being serviced by one or more Communication Manager survivable remote sites, to the primary (main) server. This feature is targeted to branch gateways only.
- Connection preserving failover/failback for branch gateways. The Connection Preserving Migration (CPM) feature preserves existing bearer (voice) connections while an branch gateways migrates from one Communication Manager server to another. Migration might be caused by a network or server failure.
- Connection preserving upgrades for duplex servers. The connection preserving upgrades for duplex servers feature provides connection preservation on upgrades of duplex servers for:
  - connections involving IP telephones
  - connections involving TDM connections on port networks
  - connections on branch gateways
  - IP connections between port networks and branch gateways
- Communication Manager survivable core provides survivability by allowing backup servers to be placed in various locations in the customer network. The backup servers supply service to port networks in the case where the main server or server pair fails, or connectivity to the main server or server pair is lost.
  - Automatic return to primary (main) server. When the survivable core is in control due to a network fragmentation or catastrophic main server failure, the return to the main server is predicated by three options.
  - The Dial Plan Transparency for survivable remote and survivable core preserves users' dialing patterns if a branch gateway registers with survivable remote, or when a port network registers with survivable core.
- IP bearer duplication using the TN2602AP circuit pack. The TN2602AP IP Media Resource 320 circuit pack provides high-capacity voice over Internet protocol (VoIP) audio access to the switch for local stations and outside trunks.

  - Load balancing. Up to two TN2602AP circuit packs may be installed in a single port network for load balancing. The TN2602AP circuit pack is also compatible with and can share load balancing with the TN2302 and TN802B IP Media Processor circuit packs.
  - Bearer signal duplication. Two TN2602AP circuit packs may be installed in a single port network for bearer signal duplication. In this configuration, one TN2602AP is an active IP media processor and one is a standby IP media processor.

- IP endpoint Time-to-Service The IP endpoint time-to-service (TTS) feature improves a customer’s IP endpoint time to service, especially in cases where the system has a lot of IP endpoints trying to register or re-register. With this feature, the system considers that IP endpoints are in-service immediately after they register.
- A survivable router is an Internal Call Controller (ICC) with an integral branch gateway, in which the ICC is administered to behave as a spare processor rather than as the main processor. The standby Avaya S8300 Server runs in standby mode with the main server ready to take control in the event of a outage with no loss of communication.
- Handling of split registrations. Split registrations occur when resources on one network region are registered to different servers. For example, after an outage activates the survivable remote server, telephones in a network region register to the main server or survivable remote server, while the branch gateways in that network region are registered with a survivable remote server. The telephones registered with the main server are isolated from their trunk resources. Communication Manager detects a split registration and moves telephones to a server that has trunk resources.
- Power failure transfer provides service to and from the local telephone company central office (CO), including wide area telecommunications system, during a power failure. This allows you to make or answer important or emergency calls during a power failure. This feature is also called emergency transfer.
- Standard Local Survivability. Standard Local Survivability (SLS) provides a local Avaya G430 or G450 Branch Gateway and Juniper J4350 or J6350 gateways with a limited subset of Communication Manager functionality when there is no IP-routed WAN link available to the main server or when the main server is unavailable.

Availability

Availability is an associated service implementation that ensures a prearranged level of operational performance during a time period. For the Avaya Aura® solution it means users want their telephones and video devices to be ready to serve them at all times.

In this context, the term availability describes the use of duplicated servers. The term high availability describes specifically System Platform's use of duplicated servers.

Communication Manager availability

High availability communications require the system to work reliably with pre-existing transport infrastructures and to integrate with a wide variety of external connectivity options. As a result, the underlying architecture must be designed to support reliable performance at every level. Communication Manager uses a variety of techniques to achieve this high reliability and availability.

Communication Manager automatically and continually assesses performance, and detect and correct errors as they occur. The software incorporates component to subassembly self-tests, error detection and correction, system recovery, and alarm escalation paths. Its maintenance subsystem manages hardware operation, software processes, and data relationships.

Servers running the duplex template provide server redundancy, with call preserving failover, on the strength of a Linux operating system.

For more information about availability assessment and methodologies, see

- The white paper, *Avaya Communication Manager Software Based Platforms: High Availability Solutions, Avaya Media Servers and Gateways*, available on the Avaya Support Web site, http://support.avaya.com
- The white paper, *Building Survivable VoIP for the Enterprise*, available on the Tolly Group Web site http://tolly.com

Communication Manager availability consists of providing a duplicated server pair that can be collocated or separated. These servers contain the same system and data files and work in an active/standby mode. When the active server fails, the servers experience an interchange, and the standby server becomes the active server.

Collocated servers are generally in the same rack in the same room and connected by crossover cable or through the customer's LAN using software duplication. Separated servers allow the two servers to be geographically separated. Server separation offers an improved survivability option by allowing servers to reside in two different buildings across a campus or small Metropolitan Area Network.

Server interchange

Server interchange is the process within a duplex server pair of a standby server becoming an active server. An arbiter process analyzes the state-of-health of both the active and standby servers and initiates a server interchange if the state of health of the active server is less than the state of health of the standby server. During this process, the standby server sends a request for the alias address. The ARP module resolves the IP address and sends an ARP reply packet with its Ethernet MAC address. The active server is seen by all the devices in the same subnet.

Each server has a unique IP address for the Processor Ethernet interface. A separate shared alias IP address is assigned to this interface on the active server and is used for connections

to the Processor Ethernet interface on the active server. As part of the operations for a server interchange, the alias address is removed from the Processor Ethernet network interface on the server going standby, and it is added to the Processor Ethernet network on the server going active. After the interchange, a gratuitous ARP message is sent out from the Processor Ethernet interface on the server going active to update the MAC address in the ARP data cache stored in the IP endpoints on the local LAN that need to be connected to the PE interface.

The IP connection for the Processor-Ethernet-connected endpoints is not available during the server interchange. This is similar to a network outage. After the interchange, the Processor-Ethernet-connected endpoints use a short network IP address of the active server.

The IP connection for the C-LAN-connected endpoints is available during the server interchange. However, some messages may be lost during the interchange. Normal operation resumes after the interchange.

Fast server interchange

The fast server interchange process is available only for the devices connected to the Processor Ethernet on duplicated servers. The branch gateways and IP telephones must have the updated firmware. The active server preserves information about all the connections and connects to IP telephones and branch gateways before resuming normal operation. The IP telephones and branch gateways accept the incoming connection to replace the old connection.

In a scenario where some of the branch gateways and IP telephones are upgraded and others are not, the following statements are true:

- The upgraded branch gateways and telephones reconnect faster
- The other branch gateways and telephones take longer time to reconnect
- The other branch gateways and telephones may negatively impact the performance of the server following the server interchange

Connection preserving upgrades for duplex servers

This feature preserves stable bearer connections for TDM endpoints and IP stations during an upgrade of duplex servers. TDM and IP connection of branch gateways, with the duplex servers being the main call controller, are also preserved.

This feature is supported on all duplex servers and all port networks. It applies when upgrading to a newer release of Communication Manager.

This feature is not call preserving and only preserves connection on stable calls. Connection preservation does not apply to calls involving H.323 IP trunks; these are H.323 IP calls and SIP calls. Connection preservation does not apply to IP trunks and ISDN-BRI stations and trunks using branch gateway resources.

System Platform high availability

About High Availability

System Platform High Availability is an optional feature that provides different levels of services continuity. This feature is available with some, but not all, Avaya Aura® solution templates. For example, the Communication Manager template does not currently use the High Availability feature. To determine support for High Availability modes of operation, refer to your Avaya Aura® solution template feature information.

Note:

System Platform High Availability does not support IPv6 and cannot be configured with IPv6 addresses.

Physical configuration

The following table describes physical configuration requirements for the various modes of System Platform High Availability operation:

Note:

See alsoHigh Availability modes for more information.

Physical requirements	High Availability modes			
	FRHA	LMHA	MPHA	Geo-Redundancy
Servers	Two servers with exactly the same hardware configuration. The servers must have identical memory, number of processors, total disk space or free disk space. (See also Avaya Aura® solution template information about server requirements.)			
Network Interface Cards (NICs)	Both servers: • A Gigabit NIC to carry Ping and other control traffic over the IP network between the primary and secondary servers. • A spare Gigabit NIC to carry data replication traffic over the high-speed crossover cable between		Both servers: • A Gigabit NIC for Ping and other IP control traffic over the IP network between the primary and	Both servers: • A Gigabit NIC to carry Ping and other IP control traffic over the IP network between the primary and

Physical requirements	High Availability modes			
	FRHA	LMHA	MPHA	Geo-Redundancy
	the primary and secondary servers.		secondary servers. • A spare 10–Gigabit NIC for data replication traffic over the crossover cable between the primary and secondary servers.	secondary servers. • A spare Gigabit NIC to carry data replication traffic over the IP network between the primary and secondary servers.
Cables	Both servers: • A Gigabit-certified CAT5A Ethernet cable connected from server port Eth0 to a Gigabit Ethernet port on the local default gateway. This connection typically carries Ping and other control traffic over the IP network between the primary and secondary servers. • A Gigabit-certified CAT5A Ethernet crossover cable connected from primary server port Eth2 to secondary server port Eth2. This connection carries data replication traffic between the primary and secondary servers.		Both servers: • A Gigabit-certified CAT5A Ethernet cable connected from server port Eth0 to a Gigabit Ethernet port on the local default gateway. This connection typically carries Ping and other control traffic the IP network between the primary and secondary servers. • A 10–Gigabit-certified CAT6A Ethernet crossover cable connected	Both servers: • A Gigabit-certified CAT5A Ethernet cable connected from server port Eth0 to a Gigabit Ethernet port on the local default gateway. This connection typically carries Ping and other control traffic over the IP network between the primary and secondary servers. • Server port Eth2 to a Gigabit-certified Ethernet port on the local

Physical requirements	High Availability modes			
	FRHA	LMHA	MPHA	Geo-Redundancy
			from primary server port Eth2 to secondary server port Eth2. This connection carries data replication traffic between the primary and secondary servers.	default gateway. This connection carries data replication traffic over IPv4 routes between the primary and secondary servers.
Default gateways	One or more high-speed IPv4 routers with Gigabit-certified Ethernet NICs. (The System Platform High Availability feature does not support IPv6 routing at this time.)			
IP addressing	The primary and secondary servers are configured on the same IP subnetwork.			The primary and secondary servers are typically configured on geographically remote IP networks.
Ping targets	The primary and secondary servers are each configured with at least one Ping target. The ping target for the primary server is the secondary server. The ping target for the secondary server is the primary server. This is necessary for node arbitration under various failover or pre-failover conditions. (See High Availability events on page 171 and High Availability node arbitration for more information.)			

Software configuration

The software configuration for High Availability operation includes:

- The same version of System Platform installed on both servers.
- The System Platform High Availability option configured on both servers.
- The Avaya Aura® solution template installed only on the primary server. The primary server automatically replicates the solution template to the secondary server.

Note:

You must perform all System Platform, High Availability, and solution template configuration using the System Platform web console while logged on to the primary server.

High Availability events

High Availability events vary in terms of type and characterization, as follows:

- **Planned (manual) switchover** — An administrator can perform a planned switchover when the active and standby nodes become synchronized and therefore contain the same data. The administrator performs this action typically to complete maintenance on the currently active server, causing it to become the new standby node during the maintenance action. This action triggers a *graceful shutdown* of node resources, where the system sequentially and safely shuts down key processes on the active node just prior to the actual switchover. No loss of data should occur during a planned switchover.
- **Preemptive (automatic) failover** — An automatic and graceful failover of the two nodes, typically triggered by ongoing detection of an intermittent hardware failure or transient shortages of node resources (for example, insufficient disk or memory space). Like the planned switchover, a preemptive failover requires full disk data synchronization across the active and standby servers. The preemptive failover triggers a graceful shutdown of resources on the active node, and a transition of all node resources to the standby node while incurring no loss of data during the failover interval.
- **Unplanned (spontaneous) failover** — A non-graceful but instantaneous failover of nodes, commonly triggered by a loss of power, a sudden and severe hardware failure, or a sudden loss of connectivity. (The latter condition can cause *split-brain* operation. See Network link failure and recovery for more information.) A spontaneous failover will cause HA protection on all virtual machines to revert to Fast Reboot High Availability behavior. (For more information, see High Availability modes.)

Regardless of the High Availability event type, you can view the reason for failover on the High Availability page of the web console.

High Availability recovery sequence

Complete recovery from a High Availability switchover or failover event occurs in stages:

- Virtual machines restart first, depending mainly on data replicated from the active node, which in turn depends on each virtual machine's High Availability mode and current operational state.
- Applications (one per virtual machine) restart next, depending mainly on recovery of the underlying virtual machine, and on the internal complexities and efficiencies of the application itself. For example, a large and complex application may not recover at the

instant its host virtual machine recovers. This application may lag virtual machine recovery by a brief interval.

- The overall Avaya Aura® solution template recovers last, depending mainly on recovery of its underlying, interdependent, applications. For example, a solution template that incorporates only one or two efficient applications may recover more quickly than a template that includes two or more larger, more complex and interdependent applications. The latter example template may lag recovery of its applications by some additional increment of time.

The full recovery time of an Avaya Aura® solution after a switchover/failover event depends on the collective recovery times of the underlying virtual machines, the applications they support, and the overall solution template itself.

High Availability configurations

You can choose from three different configurations for High Availability protection against switchover or failover events:

- All virtual machines (except the dom0 and cdom virtual machines) configured with FRHA protection. (FRHA for all virtual machines is the default configuration when enabling High Availability operation.)
- One udom (application) virtual machine configured with MPHA protection, with all remaining virtual machines (except the dom0 and cdom virtual machines) automatically configured for LMHA protection.
- All udom (application) virtual machines configured with GRHA protection. You cannot configure Geographic Redundancy with other High Availability modes on the same server. Both the primary and secondary servers are configured with GRHA protection.

The solution template you deploy in your network typically determines the type of High Availability protection you can apply to virtual machines on the primary and secondary servers.

You cannot explicitly configure LMHA operation in the SP web console. The High Availability software automatically enables LMHA operation whenever you configure at least one udom (application) virtual machine with MPHA protection.

During spontaneous failover events, both MPHA and LMHA revert to FRHA behavior , with corresponding effects on virtual machine, application, and overall solution recovery times..

No Automatic Failback

High Availability modes do not automatically migrate resources back to the preferred node when system resources are running on the standby node when the preferred node becomes available again. If both servers are healthy, then running system resources on the preferred node offers no increased benefit.

Note:

If you want to migrate resources back to the preferred node after a switchover or failover event, use the **Manual Interchange** option on the High Availability page at an appropriate time. (See Manually switching High Availability server roles.)

Avaya Call Management System availability

Avaya Call Management System (CMS) provides a high availability configuration. Dual links to the CMS provide an additional TCP/IP link to a separate CMS for full, duplicated CMS data collection functionality. The same data are sent to both servers, and the administration can be done from either server. The Automatic Call Distribution (ACD) data is delivered over different network routes to prevent any data loss from such conditions as ACD link failures, CMS hardware or software failures, maintenance, or upgrades.

Survivability

Survivability is the ability of the components within the Avaya Aura® solution to function during and after a natural or man-made disturbance. Avaya qualifies survivability for a given range of conditions over which the solution will survive.

This section addresses Session Manager and Communication Manager survivability options.

Survivable core server

The survivable core server provides survivability to an Avaya configuration by allowing backup servers to be placed in various locations in the customer's network. This option is available for Communication Manager only.

Backup servers are given administered values that are advertised to each IPSI in the configuration. The IPSI places the survivable server on a priority list based on the administered values. If for any reason, the IPSI can no longer communicate with the main server, the IPSI requests service from the next highest priority survivable server on its list. The survivable server accepts the request and assumes control of the IPSI-controlled port network.

The IPSI request for survivable server service happens after an administered No-service timer expires. The value of the No-service timer determines the amount of time the IPSI will wait to request service from a survivable server, after losing communication with the main server or the controlling survivable server. The value for the No-service timer is administrable from 2 to 15 minutes.

During No-Service timer interval, stable calls remain up in the same state as they were before the outage occurred. The stable calls do not have access to any features, such as hold and

conference. After the No-Service timer expires, shuffled IP-to-IP calls stay up, but calls on DCP or analog phones terminate.

When service to the main server is restored, the IPSI(s) return to the control of the main server in the manner administered by the customer, which can be either manually or according to a scheduled time.

In a survivable core environment, there is one main server. The main server can be a simplex server or duplex servers. If the main server is a simplex server, all survivable core servers in the configuration must also be simplex servers.

Survivable core servers provide a survivability option for all IP port networks, as well as providing Processor Ethernet for registration of gateways and IP sets..

Through careful planning and consideration, servers are placed in various locations in the customer's network. Each survivable core server is administered on the main server. The IPSIs in the configuration contain a list (called a priority list) of survivable servers. The main server is always the highest ranking server on an IPSI's priority list.

For more information on survivable core servers, see *Avaya Aura® Communication Manager Survivability Options*, 03-300428.

Survivable core server system capacities

The survivable core server can be administered as local only or as enterprise-wide survivable server(s). When administered as local only, which indicates it will act as the survivable server for a community or a subset of port networks, up to 63 survivable server clusters can be configured as survivable core servers. This way the customer may configure some servers to serve only a few port networks to enable localization of failover where desired.

For enterprise-wide fail-over coverage, up to 7 survivable server clusters can be administered. The survivable core server that acts as a main server is called System Preferred server, and it must have the same capacity as the original main. For example, when a simplex server is the system-preferred server to duplex main server, it is configured to have the same capacities as the duplex servers. This can be done based on its license files.

Depending on the type of failure and how the survivable servers are configured, an individual survivable server may accept control of all port networks, several port networks, a single port network, or no port networks. When a LAN or WAN failure occurs in configurations where port networks are widely dispersed, multiple survivable servers may be required to collectively accept control with each survivable server controlling some portion of the set of port networks.

When a survivable core server accepts control, it communicates directly with each port network through the IPSI circuit pack.

Stable calls remain up in the same state as they were before the outage occurred. The stable calls do not have access to any features such as hold and conference. The state of the stable call cannot be changed.

Survivable core server and adjunct survivability

Some adjuncts register with the C-LAN circuit pack, which in the event of failure follows the port network IPSI to a survivable server. If the port network containing the C-LAN cannot get service from a survivable server, then the adjunct will not be survivable. Having a C-LAN circuit pack in IPSI-connected port networks gives the adjunct a higher probability of survival.

Communication Manager survivability

Communication Manager offers two survivability options: survivable core and survivable remote. Survivable core servers ensure business continuity in the event of connection failure or events leading to total failure of main server complex, such as natural disaster. Survivable remote servers enhance redundancy for branch gateways within networks. Survivable remote servers take over segments that have been disconnected from their primary call server and provide those segments with Communication Manager operation until the outage is resolved.

Branch gateways and IP endpoints

Branch gateways and H.323 endpoint registration on a survivable core server is allowed if you administer the **Enable PE for H.248 Gateways** and **Enable PE for H.323 Endpoints** fields on the Survivable Processor screen of the main server.

In the event of failure of a main server:

- The H.323 endpoints that are connected through the C-LAN circuit pack reregister to survivable core servers through the C-LAN circuit pack contained in the port network that has requested a survivable core server or it reregisters to a survivable remote server. See Figure 19: IP device with C-LAN on page 176.

Note:
Only one IP address is available to the IP endpoint regardless of the server (main or survivable) in control.

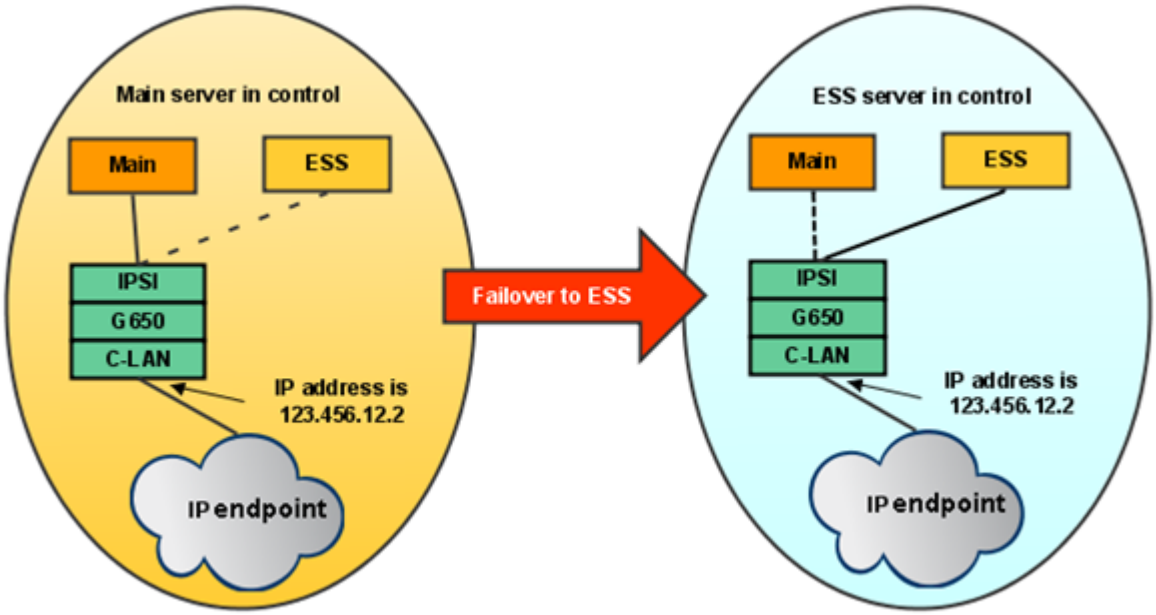


Figure 19: IP device with C-LAN

- The branch gateways and IP endpoints that are directly connected to the Processor Ethernet of the main server reregister to the Processor Ethernet on the survivable core or survivable remote server. See Figure 20: IP device with Processor Ethernet on page 177.

Note:

Two IP addresses are available to the IP endpoint: the IP address of the main server and the IP address of the survivable server. If the IP endpoint loses connectivity to its current primary gatekeeper, the IP device uses the alternate gatekeeper list for automatic recovery of service.

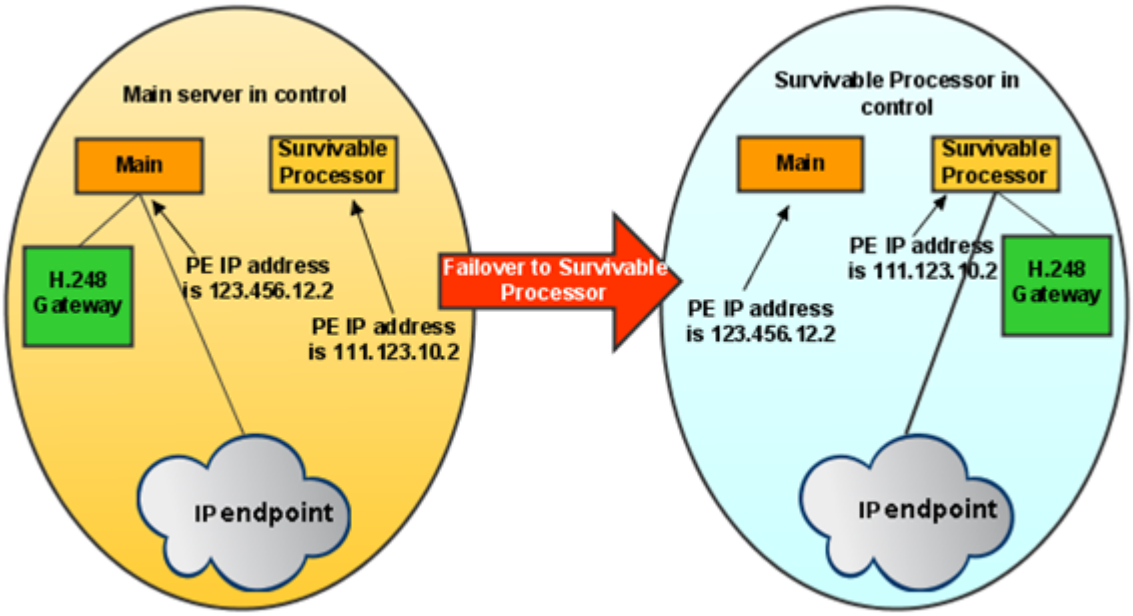


Figure 20: IP device with Processor Ethernet

IGAR and survivability

Inter-Gateway Alternate Routing (IGAR) enables systems with distributed gateways and distributed Call Centers an alternative means of providing bearer connection between port networks and branch gateways when the IP-WAN is incapable of carrying the bearer traffic. IGAR may request that bearer connections be provided by the PSTN under the following conditions:

- VoIP RTP resource exhaustion in a MG/PN is encountered.
- A codec set is not specified between a network region pair.
- Forced redirection between a pair of network regions is configured.
- The number of calls allocated or bandwidth allocated via Call Admission Control–Bandwidth Limits (CAC-BL) are reached.

IGAR takes advantage of existing public and private-network facilities provisioned in a network region.

Most trunks in use today are used for IGAR. Examples of the better trunk facilities for use by IGAR would be:

- Public or Private ISDN PRI/BRI
- R2MFC

IGAR is the next logical step in providing Quality of Service (QoS) to large distributed single-server configurations.

IGAR relies on Call Admission Control. When all VoIP RTP resources have been used, the next attempt to get a VoIP RTP resource results in denial of the VoIP connection. Communication Manager attempts to use existing applications and features to redirect the call accordingly. Each IP audio stream requires a VoIP RTP resource from either a TN2302AP IP Media Processor or a branch gateway. Exactly how many audio streams can be supported by these resources depends on the codec selection. Upon hitting the VoIP RTP resource limit, IGAR immediately attempts to use an alternative path for a bearer connection to the network region of the called party using PSTN facilities allocated for use by the IGAR feature.

Survivable remote server

The survivable remote server provides survivability to IP and SIP telephones and one or more branch gateways when communication to the core is lost. The survivable remote server provides survivability for both Communication Manager and Session Manager.

A typical survivable remote solution contains the following components:

- Survivable Session Manager that provides service to users in case there is a WAN failure between branch and core.
- Survivable remote server (Communication Manager) for the branch gateway. The survivable remote server starts to work when the branch gateway loses connectivity with main Communication Manager and register itself to survivable remote server.
- Branch gateway that provides the ability to connect the branch to the PSTN and media services as conferencing, tones, and announcements.
- End user devices (telephones and video devices) that register themselves to core Session Manager as a primary controller, but uses the survivable Session Manager as a third controller in case of WAN failure.

The survivable remote server template can be installed on a simplex standalone server, such as the HP ProLiant DL360 G7 or Dell™ PowerEdge™ R610 servers, or on an embedded server, such as the Avaya S8300D Server.

Communication Manager

For Communication Manager the survivable remote server takes control of branch gateways that has its address in the Media Gateway Controller (MGC) list. The IP telephones use an Alternate Gateway List (AGL) for branch gateway addresses. These addresses are automatically generated by Communication Manager and sent to the IP telephones upon registration. Because the survivable remote server does not manage the IP Server Interface (IPSI) circuit packs, it cannot control port networks.

In a survivable remote environment, each IP endpoint and branch gateway is manually configured with a list of call controllers during initialization. If for any reason, the communication between a branch gateway and its primary controller stops, the branch gateways and the IP endpoints register with a call controller on its list. If the survivable remote server is in the list of call controllers, the branch gateway and the IP endpoint registers with the survivable remote

server. The branch gateway registers with the survivable remote server first before the IP telephone registers with the survivable remote server.

For more information on the survivable remote servers as it relates to Communication Manager, see *Avaya Aura® Communication Manager Survivability Options*, 03-300428.

Session Manager

Session Manager for Survivable Remote is a set of software packages that acts as SIP routing and user relation elements when in survivable mode. It is built with the same specifications as the core Session Manager, providing survivability services for trunks, SIP stations and applications.

In the branch office are the branch SIP endpoints and a branch gateway. The endpoints are registered to both the core Session Managers and the survivable Session Manager. The endpoints have the concept of an active controller. The active controller is defined as the Session Manager to which the endpoints currently have subscriptions established. In sunny day operations, the core Session Manager is always the active controller. The survivable Session Manager receives no call traffic. The branch gateway is registered with the main Communication Manager. In rainy day operations, the survivable Session Manager is always the active controller. Currently, the only supported network outage is a complete branch WAN outage where all devices in the branch have lost contact with all devices in the core. Partial network outages are not guaranteed to exhibit desired redundancy behaviors.

For more information on the survivable remote server as it relates to Session Manager, see *Administering Avaya Aura® Session Manager*, 03-603324.

Telephone perspective

Session Manager supports simultaneous registration of telephones, a method that provides the greatest robustness using the SIP-outbound semantics. This means that SIP telephones simultaneously register with core Session Managers and the survivable Session Manager. The telephones accept incoming calls from any of these servers and automatically perform active controller selection according to existing algorithms. This means that although SIP telephones can receive calls from any of the registered controllers, telephones initiate calls through only the highest priority controller, the *active* controller. With an active controller outage, telephones mark the next controller in the list as the active controller for outbound services. Upon detecting the revival of the highest priority server, telephones move back to the revived controller as the active controller.

Alternate routing during rainy day

During a rainy day scenario, the survivable Session Manager provides the following three major tasks:

- Connects branch users to each other.
- Connects users to other non-survivable Session Manager users that reside on different branch.
- Connects users to other branches using PSTN trunks, such as emergency numbers and other branches.

Messaging during rain day

Calls to a user on a survivable Session Manager

If a call comes into the core Session Manager because of centralized trunking, and the core cannot reach the user within the branch, the call goes to the user's coverage path and to voice mail, if administered as part of the coverage path. If a call comes in directly to the survivable Session Manager, there is no local messaging support.

Access to voice mail for survived users

There are two ways to access the voice messaging system when in survivable remote mode:

- Direct call to voice messaging system.
- A coverage method using special characters through which station-to-station calls from a branch location to the main location can be redirected over a PSTN trunk. Here is a description of the special characters:
  - The character *L* at the beginning of a coverage remote entry ensures that only the survivable core or remote server uses the entry. The main server considers this entry as unavailable.
  - The character *%* signifies *wait for answer.*
  - The character , instructs Communication Manager to pause, which is useful after a *wait for answer* to ensure the far-end is prepared to receive subsequent digits.
  - The character *D* denotes the called party's extension and allows the same coverage path and coverage remote entry to be used by many users sharing common characteristics.

Survivability for branch gateways

Branch Gateway recovery via survivable remote server

If the link between the remote branch gateway and the branch gateway controller is broken or the controller is down, the survivable remote server activates and assume call processing for the branch gateway. The branch gateway controller can be any simplex or duplex server. The strategy by which the branch gateways change control from the primary to the survivable remote controller is driven by the gateway using the branch gateway controller list.

When main server is a standalone server

The connectivity path between the remote branch gateway and the Call Controller are in two configurations: via C-LAN or direct to the server Processor Ethernet interface. The connection path via C-LAN in the port network is as follows:

```
branch gateway ⇔ IP network ⇔ C-LAN ⇔ port network backplane ⇔ IPSI ⇔ IP network
⇔ duplex server
```

The connectivity path directly to the Processor Ethernet interface of the server is as follows:

```
branch gateway ⇔ IP network ⇔ PE interface of the server
```

Link connectivity between the main call controller and the branch gateway is monitored through the exchange of keep-alive messages between the two components. If the link between the active call controller and the branch gateway is severed, the branch gateway tries to reestablish the link using its alternate gatekeeper list. The alternate gatekeeper list is divided into primary and secondary addresses. All primary addresses receive priority weighting over the secondary addresses.

In the event of a WAN failure, any IP endpoint or branch gateway that cannot reach the primary controlling server registers with a survivable remote server controller in survivable mode. In the duplex server/branch gateway configuration, up to 50 survivable remote servers are available and ready for the fail-over process. The survivable remote server is always ready to acknowledge service requests from IP telephones and branch gateways that can no longer communicate with their main controller. Once the telephones and the branch gateway are registered, end users at the remote site have full feature functionality. This failover process usually takes less than 3 minutes. After failover, the remote system is stable and autonomous.

When main server is embedded server

In this configuration, the connectivity path between the branch gateway and the embedded S8300D server is:

```
Endpoint ⇔ IP Network ⇔ S8300D server
```

The link failure discovery and recovery process is the same as above, except there are no C-LAN addresses in the alternate gatekeeper list. In this configuration, up to 10 survivable remote servers can back up the branch gateways that are controlled by the S8300D Server.

Auto fallback to main server for branch gateways

Auto fallback to main server for branch gateways allows a branch gateway being served by a survivable remote server to automatically return to its primary gatekeeper. This feature is connection preserving; that is, stable bearer connections do not drop during this process.

The auto fallback process

Although the survivable remote server is the acting call controller, the branch gateway attempts to register with the main server every 30 s or whenever there are no active calls. This signaling also acts as keep-alive messages to the main server. The first registration request with the main server sets up encryption on the TCP link for H.248 messages. The branch gateway keeps the survivable remote registration until the branch gateway is accepted by the main server. Once registered with the main server, the branch gateway drops the survivable remote link. Once all branch gateways have migrated from the survivable remote server, that server unregisters all IP endpoints, which automatically reregister with the main server.

This automatic migration of branch gateways to the main server is administered to happen immediately (default), when there are no active calls, or at a scheduled time of a day.

Connection preserving failover/failback for branch gateways

This feature allows existing stable calls to be preserved when the branch gateway fails over to another server, or a survivable remote server, or returns to its main server. It is supported on all branch gateways. It applies to the failover and fallback of branch gateways to or from a survivable remote server and to or from a survivable core server.

During the failover/fallback process the bearer connection of stable calls are preserved. These include analog stations and trunks, DCP stations, digital trunks, IP stations using branch gateway resources, ISDN-PRI trunks, calls between gateways, IGAR, and previous connection-preserved calls.

Branch gateway standard local survivability

Standard local survivability (SLS) is survivable call processing engine that provides service to the branch gateway when the branch gateway cannot reach Communication Manager. This engine is resident in the branch gateway firmware and provides basic telephony functions at the branch without being registered to Communication Manager.

The SLS features are:

- Local station and outbound PSTN calling
- Inbound calls over the trunks to be delivered to available stations
- An H.323 gatekeeper for local IP phones to register
- Call Detail Recording in a syslog format

During transition to survivability mode, only local IP-IP calls are preserved.

The link recovery process follows these steps:

1. While SLS is enabled and processing, the branch gateway continues to seek an alternative branch gateway controller.
2. If Communication Manager accepts the registration, then the active IP-to-IP calls that shuffle are preserved.
3. The SLS application stops processing any new calls and goes to inactive mode.

Survivable Modular Messaging

Survivable Modular Messaging provides a disaster recovery solution for Modular Messaging. The survivable configuration consists of a primary Modular Messaging system at one site and a Survivable Modular Messaging at another site. The primary system actively handles the calls and all the Modular Messaging services, while the survivable system is kept on standby. The survivable system periodically duplicates the data and configuration from the primary system, which keeps the survivable system in a suitable standby state. On the event of a disaster of the primary system, the calls can be routed to the survivable system. The survivable system assumes the role of the active system and thereon handles the calls and all the Modular Messaging services.

Survivable Modular Messaging is supported only for the MSS and Exchange stores. Survivable Modular Messaging is designed to work with Communication Manager running on a survivable core or survivable remote server.

Note:

The survivable system for the MSS store is as good as the last restore of data and configuration backup. To continuously replicate the data and configuration in this configuration, use third-party mirroring applications like Message Mutare Mirror.

Redundancy

Avaya Aura® Messaging redundancy

Avaya Aura® Messaging provides a variety of redundancy options.

Resilience

The Messaging solution is very resilient as it can remain functional even with major failures in the environment or its own components.

N+1 Redundancy

The application servers do not maintain data that is unique to a single application server. Messaging provides N+1 redundancy: a single application server can be added to the Messaging application server cluster to provide redundancy for any of the existing application servers in the same cluster.

Offline handling

To deal with outages in the data network the application server can operate in offline mode. In offline mode, the application server continues to provide call answering for existing users, and callers cannot tell that there is no connectivity with the storage server. Messaging users can still use the TUI and access messages that were received in the past three days as well as any new call answering messages that were received since the outage. Individual user-to-user messages sent while offline can be retrieved as well.

Offline handling in the Messaging system is session-based. It only applies to those users who are affected by the outages in the data network. The Messaging system is capable of handling outages because the application server maintain a cache of relevant data, such as:

- directory information (both user directory and user personal distribution lists)
- user greetings and recorded names
- voice messages

The cache is a hybrid single or distributed cache. Some data, for example, directory data, is cached on each single application server, while other data, for example, voice messages, is stored in a cache that is distributed among the application servers in a cluster.

Disaster recovery

The application servers can be clustered over a WAN to provide disaster recovery capabilities. This is often done to match a clustered-over-WAN IP-PBX configuration. In a typical deployment, half of the cluster application servers are co-located with the corresponding half of the distributed IP-PBX in one geographical location. The other half of the cluster is co-located

with the other half of the IP-PBX in a different location. For more information, see Clustering on page 185.

Clustering

You can combine up to three application servers to form a cluster. You can add an additional N+1 server for redundancy (for a total of 4 servers in the cluster) as long as the active traffic does not exceed 300 ports. Each cluster connects to one storage server and supports the same telephony server.

Clustering application servers allows you to:

- Increase the system's capacity so it can support more users. Every application server you add to the cluster increases the number of available ports.
- Provide redundancy for any application server in the same cluster. Application servers within a cluster are configured identically and are, therefore, interchangeable.

Recovery

Network recovery

Conventional wisdom holds that network reliability is typically 3-9s (99.9%) on a LAN, and 2-9s (99%) on a WAN. The leading causes of network failure are a WAN link failure, administrator error, cable failure, issues that involve connecting new devices or services, and malicious activity, including DoS attacks, worms, and viruses. Somewhere lower down on the list are equipment failures. To achieve the highest levels of availability, it is important that a strong change control policy and network management strategy be implemented.

There are numerous techniques for improving the reliability of data networks, including spanning tree, self-healing routing protocols, network management, and change control.

Related topics:

Change control on page 185
Dial backup on page 188
Convergence times on page 188

Change control

Change control describes a process by which an organization can control nonemergency network changes and reduce the likelihood of administrator errors that cause network

disruption. It involves carefully planning for network changes (including back-out plans), reviewing proposed changes, assessing risk, scheduling changes, notifying affected user communities, and performing changes when they will be least disruptive. By implementing a strict change control process, organizations can reduce the likelihood of administrator errors, which are a major cause of network disruption, and increase the reliability of their networks.

Layer 2 mechanisms to increase reliability

Spanning tree

IEEE 802.1D spanning tree is an Ethernet loop avoidance protocol. It allows network managers to connect redundant network links within their networks. Before the advent of spanning tree, loops within a switched Ethernet network would forward traffic around the loop forever, which saturated the network and prevented new traffic from getting through. Spanning tree selects one switch as a root and creates a loop-free topology connecting to the root. If loops are discovered, one switch blocks that port until its alternate path to the root is disrupted. Then the blocked port is brought back into service. There are several drawbacks to spanning tree:

- By default, all switches have the same priority, which means that root bridge selection can be suboptimal in a network.
- Spanning tree is slow to converge. It typically takes at least 50 s from link failure for a backup link to become active. As Layer 2 complexity increases, so does convergence time.
- Although there are mechanisms for speeding up spanning tree, most are proprietary.
- Traditional spanning tree is not VLAN aware. Thus, it will block links even if VLAN provisioning would have prevented a loop.

To solve these issues, the IEEE has recently introduced 802.1s and 802.1w enhancements. 802.1w introduces rapid spanning tree protocol (RSTP). RSTP uses active handshaking to speed up convergence times. 802.1s introduces multiple spanning trees (MST), which is a way of grouping different VLANs into different spanning tree instances.

Link aggregation groups

Link aggregation groups (LAGs) is a mechanism for combining multiple real interswitch links (typically four; Avaya products are configurable from two to eight) into one point-to-point virtual interswitch link. The advantage of this mechanism over spanning tree is that an organization can have the redundant links in if a failure occurs in one of the LAG links, the two switches will quickly discover it and remove the failed link from the LAG. This reduces the convergence time to nearly instantaneous. Not all implementations interoperate, so care must be taken when the LAG connects switches from multiple vendors. Also, LAG links are a point-to-point technology. They cannot be used to connect a backup switch in case the primary fails. When available, this is a very good mechanism for improving the resiliency of LANs.

Layer 3 availability mechanisms

Routing protocols

Routing protocols allow routers to dynamically learn the topology of the network. Should the topology of the network change, routing protocols update their internal topology table, which allows them to route around failure.

There are two types of routing protocol, distance vector and link state. Distance vector protocols, including RIP and IGRP, exchange their entire routing table periodically. To each route, they add their metric (for RIP, this is hop count) and insert it in the routing table. If updates fail to arrive before the router's timer expires, it purges the route and looks for another path. These protocols are usually slow to converge. See Table 11: Sample convergence times (single link failure) on page 188.

Link-state protocols, such as OSPF, take a more holistic view of the network. They compute the entire topology of the network and insert the best path to a destination in the routing table. Link state protocols exchange their routing tables only once, when routers first establish a relationship. After that, they only send updates. They also send hello messages periodically to ensure that the other routers are still present. Link state protocols converge much more quickly than distance vector protocols, and thus are generally better suited to networks that require high availability.

Virtual router redundancy protocol

Virtual router redundancy protocol (VRRP) and the related Cisco proprietary hot standby router protocol (HSRP) provide a mechanism to deal with router failure without disrupting endpoints on the network. In essence, these protocols work by assigning a virtual IP address and MAC address for the routers. This address is given to endpoints as their default gateway. The two routers send periodic hello messages marked with a priority value between each other. The high-priority router assumes the virtual address, and traffic flows through it. If the primary router fails or its capabilities become degraded (such as if a WAN link fails), the secondary router takes over. This is a useful mechanism to protect endpoints from router failures, and works with IP Telephony endpoints.

Multipath routing

Modern routers and Layer 3 switches allow multiple routes for a particular destination to be installed in the routing table. Depending on the implementation, this can be as high as six routes. Some implementations require that all routes that are inserted in the routing table have the same metric, while others allow unequal metric routing. In cases where the metric for all installed routes are the same, the router will load balance traffic evenly across each path. When the metric for multiple routes vary, the traffic is load balanced in proportion to the metric (in other words, if one path is twice as good as another, two-thirds of the traffic travels down the good path, and one-third of the traffic selects the other one). Asymmetric routing is suboptimal for voice, so route-caching (described earlier) should be considered in this environment.

In addition to using all (up to 6) active paths and optimally using available bandwidth, multipath routing greatly improves convergence time. As soon as a router detects a path failure, it remove

it from the routing table, and sends all traffic over the remaining links. If this is a physical link failure, the detection time is nearly instantaneous. Therefore, Avaya recommends the use of multipath routing, where available, across multiple links to a particular location.

Dial backup

One cost-effective technique for installing backup WAN links is to use dial backup. This can be done using either ISDN-BRI or analog lines. ISDN lines typically take 2 s to connect, while 56-kbps analog modems take approximately 1 min. Although this strategy is effective for data traffic, it is less effective for voice. First, the bandwidth may have been greatly reduced. If this is the case, the number of voice channels that can be supported might have been reduced proportionally. Also, if QoS is not properly applied to the backup interface, high packet loss and jitter can adversely affect voice quality. Finally, the time that is required to establish the new link can be up to 1 minute, which disrupts active calls. However, providing that these considerations are taken into account, proper QoS is applied, and a compressed codec is chosen, dial backup can be an effective solution for two to four users.

Convergence times

Convergence is the time that it takes from the instant a failure occurs in the network until a new path through the network is discovered, and all routers or switches are aware of the new path. Convergence times vary, based on the complexity and size of a network. Table 11: Sample convergence times (single link failure) on page 188 lists some sample convergence times that are based on a single link failing in a relatively simple network. They reflect update and/or hello timers expiring. Dialup *convergence* times reflect the time that it takes to dial, connect, and authenticate a connection. These times do not take into account LAG, fast spanning tree, or multipath routing, which speed up convergence. This table shows the importance of carefully planning for fail-over in a network. For example, both OSPF and EIGRP (Layer 3) protocols converge faster than spanning tree (Layer 2). When designing a highly available data network, it is more advantageous to use Layer 3 protocols, especially link-state (OSPF) or hybrid (EIGRP) protocols, than Layer 2 (spanning tree).

Table 11: Sample convergence times (single link failure)

Protocol	Approximate convergence time (in seconds)
EIGRP (Cisco)	2
OSPF	6 to 46
RIP	210
Rapid spanning tree RSTP	10
Spanning tree (Layer 2)	50+

Protocol	Approximate convergence time (in seconds)
ISDN dialup (connect + authentication)	2
56-k dialup (connect + authentication)	60

IP endpoint recovery

Avaya's distributed IP-based systems experience increased availability by virtue of the alternate gatekeeper feature. When IP telephones register with Communication Manager, they are given a list of alternate gatekeepers to which they can re-register in the event of a failure. Thus, if a C-LAN fails or becomes unavailable, users registered to a particular C-LAN can re-register to another C-LAN that is unaffected by the failure.

The Avaya servers have a scalable architecture with different server components. These components provide processing and relay signaling information between Communication Manager and the Avaya IP endpoints. The system architecture is inherently distributed, providing the scalability to support a large number of endpoints and the flexibility to work in various network configurations.

This distributed nature of the architecture introduces additional complexity in dealing with endpoint recovery, since failure of any element in the end-to-end connectivity path between an IP endpoint and the switch software can result in service failure at the endpoint.

The recovery algorithm outlined here deals with detection and recovery from the failure of signaling channels for IP endpoints. Such failures are due to connectivity outages between the server and the endpoint, which could be due to failure in the IP network or any other component between the endpoint and the server.

The connectivity path between the endpoint and the server are:

```
Endpoint ⇔ IP network ⇔ C-LAN ⇔ PN backplane ⇔ IPSI ⇔ IP network ⇔ server
```

```
Endpoint ⇔ IP network ⇔ Server PE interface
```

In this configuration, IP endpoints register to the C-LAN circuit pack within the port network or directly register to the server Processor Ethernet interface.

A C-LAN circuit pack provides two basic reliability functions:

- A C-LAN hides server interchanges from the IP endpoints. The signaling channels of the endpoints remain intact during server interchanges and do not have to be re-established with the new active server.
- A C-LAN terminates TCP keep-alive messages from the endpoints and thus frees the server from handling frequent keep-alive messages.

Recovery algorithm

The recovery algorithm is designed to minimize service disruption to an IP endpoint in the case of a signaling channel failure. When connectivity to a gatekeeper is lost, the IP endpoint progresses through three phases:

- Recognition of the loss of the gatekeeper
- Search for (discovery of) a new gatekeeper
- Re-registration

When the IP endpoint first registers with the C-LAN circuit pack, the endpoint receives a list of alternate gatekeeper addresses from the DHCP server. The endpoint uses the list of addresses to recover from a signaling link failure to the C-LAN circuit pack/gatekeeper.

When the IP endpoint detects a failure with the signaling channel (H.225/Q.931), its recovery algorithm depends on the call state of the endpoint:

- If the user of the telephone is on a call and the endpoint loses its call signaling channel, the new IP robustness algorithm allows the telephone to reestablish the link with its gatekeeper without dropping the call. As a result, the call is preserved. Call features are not available during the time the telephone is trying to reestablish the connection.
- If the user of the telephone is not on a call, the telephone closes its signaling channels and searches for a gatekeeper using the algorithm defined below.

To reestablish the link, the telephone tries to register with a C-LAN circuit pack on its gatekeeper list. The C-LAN circuit pack load balancing algorithm looks for the C-LAN on the list with the least number of telephones registered to it. As a result, the recovery time is short, and there is no congestion due to too many telephones trying to register to a single C-LAN circuit pack.

In this configuration, the telephone registers to the server's Processor Ethernet Interface and the IP endpoint connects directly to the server Processor Ethernet (there is no C-LAN circuit pack). The connectivity path between the telephone and the server is:

```
Endpoint ⇔ IP network ⇔ Server
```

To discover connectivity failure, keep-alive messages are exchanged between the IP endpoint and the server. When the endpoint discovers that it no longer has communication with its primary gatekeeper, it looks at the next address on its list. If the next address is for a survivable remote server, then that server accepts the registration and begins call processing.

While the survivable server is not call preserving, the fail-over from primary gatekeeper to survivable server is an automatic process and does not require human intervention. The failback from a survivable server to a primary gatekeeper, however, is not currently automatic and requires a system reset on the survivable server. During the fallback to the primary gatekeeper, all calls are dropped with the exception of IP-to-IP calls.

IP endpoint time to service

The Time to Service (TTS) feature improves the time required to bring an IP endpoint into service by reducing the amount of required signaling for a telephone to reach the in-service state. Once a telephone is registered, TTS keeps the registration persistent for a relatively long Time to Live (hours) regardless of TCP connection failure, network outages, or even restarts of the endpoint. This significantly reduces the number of times that IP telephones need to re-register with Communication Manager due to outages. As a result, the TTS feature improves system availability after a network outage.

There are two functions in TTS that improves the availability of IP endpoints. One function is that the IP Endpoint Time-To-Service feature changes the way IP endpoints register with their gatekeeper, reducing the time to come into service. In the current Communication Manager architecture, there are two activities to bring the IP endpoints into service. The H.323 IP endpoint must register with Communication Manager and then it must establish a TCP socket connection between the server and the endpoint for call signaling. Since all the IP endpoints in a system strive to get into service as quickly as possible after an outage, the main server can be flooded with activity. In a system with a large number of IP endpoints, this flooding leads to delays not only for telephones trying to get into service but also for endpoints already in service trying to make calls.

The TTS separates the timing of the H.323 registration process from the timing of the TCP socket-connection setup process. This decoupling of the steps considerably improves the time for telephones to be in-service.

With TTS, after all the IP telephones within a system register to Communication Manager, the TCP socket is established when the processor occupancy level returns to normal. However, when the main processor occupancy level is high, the TCP socket is established on demand (when users make a first call or when a call needs to be delivered to a user) or via background maintenance. Once the TCP socket is established, the socket remains up for subsequent calls. In addition, with TTS, Communication Manager, rather than the IP endpoint, initiates the establishment of the TCP socket resulting in faster establishment of TCP sockets.

The second function of TTS significantly reduces the number of times that IP endpoints need to reregister with Communication Manager. This feature provides the capability to persist IP endpoint registrations across many network failures and other outages. Currently, whenever TCP sockets are dropped, the IP endpoints must reregister. With TTS, IP endpoints do not usually need to reregister for network outages that do not cause the system to failover to an survivable core or remote server. Since most issues with registration delays in the past have been after short network outages, this capability dramatically reduces the number of times that an IP endpoint needs to reregister with Communication Manager.

If reregistration is not required, only the re-establishment of the TCP socket is needed, which is also done in an on-demand fashion. Currently, in a call center environment, the agents must always log in again whenever the endpoint becomes unregistered. As a consequence of not requiring reregistrations after most outages, the agents' log-ins persist and they do not need to log in again.

Note that reregistration is still required for outages that cause the IP endpoints to failover to an survivable server (and then again when they recover back to the main server). In addition, a Communication Manager reset of level 2 (or higher) or a power cycle on the IP endpoints also requires IP endpoints to reregister because the information for the registration is erased under these conditions. For security reasons, IP endpoints also need to reregister with Communication Manager if they have not been able to communicate with Communication Manager over the RAS signaling channel for an extended period of time.

Changes in IP endpoints

Time to Service (TTS) features work only if corresponding changes are made to the Avaya H.323-based IP endpoints. The TTS algorithms are implemented in the IP endpoints. These TTS-enabled endpoints continue to support previous link recovery algorithms when communicating with a server that does not support TTS or does not have TTS enabled.

The TTS features works seamlessly with older IP endpoints. However, the benefits of the features are limited to the number of TTS-capable endpoints that supports TTS deployed with Communication Manager.

Note:

16xx-series endpoints do not support TTS.

Operation with NAT/firewall environment

With the Time to Service (TTS) algorithm, the TCP connection for the call signaling channel is initiated by the server, not by the endpoints. With server-based NAT or firewall environments, the firewalls must be configured appropriately to allow TCP connections from the server to the endpoints.

Network outage time line for port networks

The port network survivability during short network outages is relatively short at 15 s. This allows time for Communication Manager and the affected port network to recover from a network outage without closing the IPSI socket connection, which can cause data loss and port network warm restarts.

If the network outage is shorter than 15 s (interval A in the figure):

- All stable calls that go through the port network are preserved.
- Most transient calls will complete with a delay but some may fail

If the network outage is between 15 s and 60–s (interval B in the figure):

- Most stable calls that go through the port network are preserved.
- Most transient calls will fail

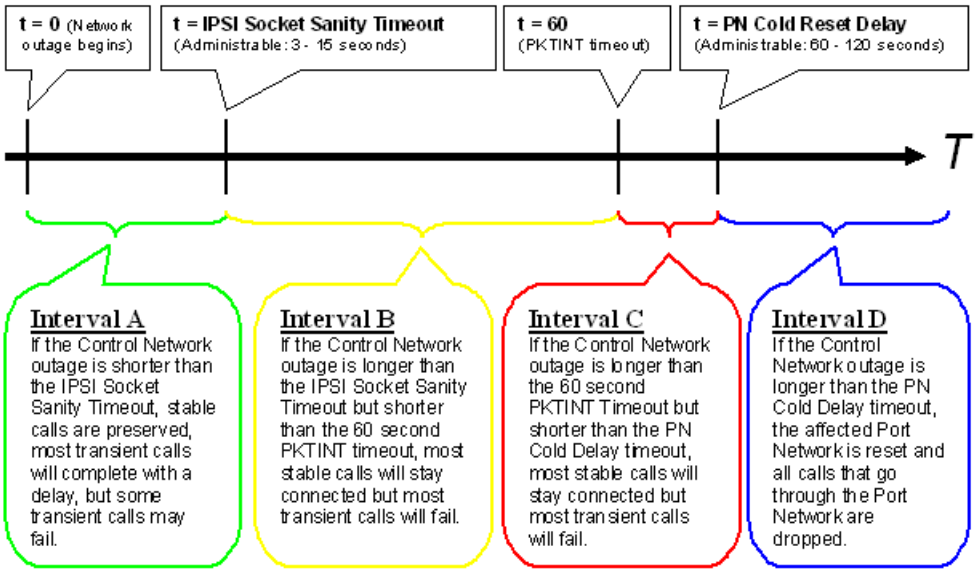
If the network outage is between 60–s but shorter than the port network cold reset delay timer setting (interval C in the figure):

- Most stable calls that go through the port network are preserved.
- Most transient calls will fail

If the network outage is longer than the port network cold reset delay timer setting (interval D in the figure):

- The port network is reset
- All calls are dropped

The following figure shows the survivability time line.



Chapter 12: Tools

Management tools

The Avaya Aura® solution offers many tools to manage the many components. These tools include

- Avaya Integrated Management Suite
- ART
- Third-party tools, such as HP ProView

Avaya Site Administration

Site Administration is a stand-alone application installed on a Microsoft Windows client computer. Use the comprehensive tools in Site Administration to simplify the day-to-day administration and maintenance tasks of Avaya media servers and Avaya messaging platforms. The Windows-based GUI of the application simplifies moving, adding, changing, and basic traffic analysis as easy as pointing and clicking.

Site Administration provides graphical station and system administration screens and easy-to-use wizards for basic administration tasks. With Site Administration, you can:

- View and administer system-wide settings.
- Add, change, swap, and delete stations.
- Change phone types and feature buttons. You can assign features to phone buttons by clicking a button on the picture.
- Create station templates to save time when adding stations.
- Print button labels for your phones.
- Add or delete subscriber accounts on your messaging systems.
- Look up extensions and query the voice system for stations matching specific criteria.
- Add change and delete Announcements.
- Perform backup and restore Announcements.
- Generate and schedule a Communication Manager command report that can be printed, exported, or sent as an e-mail.

- Export data from your voice system to a file that other software applications such as, spreadsheet or database, can use.
- Import data from a spreadsheet or database application into your voice system.
- Use the Graphically Enhanced DEFINITY Interface (GEDI) to administer voice systems. If there is an administration activity that you cannot do with one of Avaya Site Administration's wizards or shortcuts, you can generally do it with GEDI.
- Start a terminal emulation (cut-through) session to administer devices that are not supported by GEDI. When you use terminal emulation, you are making changes directly to the device.

Site Administration is provided in the Avaya Integrated Management Release 6.0 Site Administration offer.

Avaya Fault and Performance Manager

Avaya Fault and Performance Manager provides the following:

- Network map or system view of your converged network.
- Tabular tools to monitor the status and performance of the devices on your network.
- View the network to examine faults and performance data from the Avaya media servers on the network.

In addition, Avaya Fault and Performance Manager collects configuration, fault, and performance data from the Secure Services Gateway or directly from an IP-enabled voice system using OSSI, and then displays the data in text, tables, and graphic formats.

The Avaya Fault and Performance Manager:

- Gives alerts when voice system faults and performance problems occur.
- Aids in isolating and identifying fault and performance problems.
- Provides the tools to help fix fault and performance problems.
- Collects configuration, fault, and performance data from your systems (by a schedule set by you).
- Stores system exceptions data and performance measurements in a database.
- Generate reports on the data. Data can be presented as text, tables, and graphs.

In addition to monitoring your voice systems, Avaya Fault and Performance Manager also provides alarm management (not performance monitoring) for several adjuncts related to your voice systems including Avaya Aura® Messaging, Application Enablement Services (AES), Modular Messaging, Message Networking, and Secure Access Link. Avaya Fault and Performance Manager supports up to 400 systems.

Avaya Fault and Performance Manager provides the following features:

- Graphical User Interface (GUI) — Provides the following views of the managed nodes in your network: system groups, DCS trunk connectivity, IP trunk connectivity, and clusters.
- Alarm Reception and Forwarding — Detects alarms from your voice system (or other Avaya system) and relays them to the Avaya Technical Service Center. Also provides a ping service to inform if a system is unreachable.
- Configuration Display — Enables you to browse the configuration and administered properties of managed nodes in both a graphic view and a table view.
- Notes Page for each Voice System — Enables you to create your own knowledge database.
- Cut Through — Provides access to your voice systems to perform routine maintenance and administrative tasks.
- System Status Reporting — Creates reports about the status of your voice systems. Includes information such as connection state, connect times, attempts, requests, errors, and alarms. You can define a wide variety of tabular and graphical reports covering performance, configuration, and exception data from any combination of systems. You can restrict reports to components of interest; for example, a specific hardware location or a list of trunk groups.
- Report Manager — Enables you to define the parameters for individual reports for all or selected systems. The report options include performance, configuration, and exceptions. You can view the reports on screen in both the table and chart formats or direct the reports to a printer, HTML file, GIF file, or ASCII file.
- Data Collection — Enables you to specify the types of data to collect from each system, the schedule for collecting the data, and how long to store the data.
- Exception Logging — Enables you to specify conditions for exception logging. This includes performance thresholds and fault or error conditions.
- Exception Alerting — Enables you to specify an alert level for each exception. The level and location of the alert are displayed in the Configuration and Status window as long as the condition persists.

Avaya Integrated Management Release 6.0 Performance and Administration offer provides Avaya Fault and Performance Manager.

Network Management Console with VoIP System View

Avaya Network Management Console is the platform and the focal point for initiating all Network Management activities. Network Management Console provides a single system view of the entire deployment, which includes topology, faults, and status display. Network Management Console automatically detects devices on the network and displays them in an

intuitive, hierarchical navigation tree. The different views of the navigation tree helps the administrator to view the network by IP subnet, device type, or logically by voice systems.

The elements in the navigation tree are color coded to indicate the fault status. Network Management Console also serves as a focal point for viewing fault event notifications, which are collected and displayed in the Event Browser. The administrator can also configure actions to be executed upon receipt of particular types of notifications.

Network Management Console serves as a launch point for all other applications in the Network Management offer. Applications can be easily launched from the menu bar, either globally, or focused on a device selected in the navigation tree. Network Management Console also provides built-in capabilities to launch a telnet or web browser session on the selected device in the tree.

Avaya Network Management Console with VoIP System View is provided in the Avaya Integrated Management Release 6.0 Network Management offer. For information on Network Management Console, refer to the *Avaya Integrated Management Release 6.0 Network Management Console User Guide*.

Device Managers

Avaya Integrated Management Release provides a group of Network Management applications. The Network Management applications include device manager applications that are specific to individual media gateways.

Using these powerful, mouse-driven tools, network managers can set up and configure all device parameters, including:

- Standard port settings
- Port security
- Redundancy modes
- Device-specific functions

The device managers described below are provided in the Avaya Integrated Management Network Management offer.

Chapter 13: Traffic Engineering

Introduction to traffic engineering

In its most general sense, an Avaya Aura® enterprise solution consists of a network of various applications. Currently, the most prominent application is Communication Manager. Other applications include

- Experience Portal (Voice Portal)
- Expanded Meet-Me Conferencing
- Avaya Aura® Conferencing
- Avaya Aura® Messaging
- Avaya G860/M3K High-Density Trunk Gateway
- Voice Recording

The various applications supported by an enterprise can be interconnected in various ways. Circuit-switched trunking (TDM trunking) and H.323 trunking (IP trunking) are still widely used in the field today.

The interconnection of enterprise elements via SIP uses Avaya Aura® Session Manager, which controls the call routing between all SIP-enabled elements in an enterprise. Session Manager can also function as a registrar for SIP stations.

The primary purpose of this section is to provide methodologies by which the traffic-sensitive components in an Avaya communication enterprise can be properly engineered. Special emphasis is placed on elements associated with the most recent Session Manager and Communication Manager releases.

Design inputs

This section discusses the essential design elements to be specified by the customer. Those elements pertain to the configuration topology, the various endpoints involved, and the nature of the traffic flow between those endpoints.

Topology

An Avaya Aura® enterprise solution consists of a network of various applications, including Session Manager, Communication Manager, Experience Portal (Voice Portal), Messaging, Avaya G860/M3K High-Density Trunk Gateway, and Voice Recording. Communication Manager, which is currently the most prominent application, consists of a server and all of the components under that server's control. The various components can be placed into logical and/or physical groups.

A single Communication Manager system comprises one or more network regions. Each network region is a logical grouping of components such as endpoints, gateways, and certain circuit packs. The components of a Communication Manager system could also span various physical placements including gateways and geographical locations (sites).

Knowledge of the details of the configuration topology, from both logical and physical standpoints, is essential to properly conduct a traffic analysis. In particular, the topology often plays a role in determining the routes that are traversed by various call types.

Calls and endpoints

A call is normally thought of as a communication between two or more parties, across a set of communication facilities, and it is natural to think of those parties as the endpoints involved in the call. In fact, such parties are sometimes referred to as terminals, and they can include telephones, fax machines, voice recorders, IVRs (Interactive Voice Response units), and video devices. However, the term endpoint can also be used in certain circumstances to include facilities that do not represent the true points of termination of a call (most notably, trunks).

We use the term station to refer to a device being used by human beings in real-time to originate and receive calls (including voice calls, faxes, and text messaging). Such devices include circuit-switched telephones, H.323 hardphones and softphones, SIP hardphones and softphones, and fax machines. The people using them are referred to as users.

When referring to a domain such as an Avaya Aura® enterprise solution or a particular Communication Manager system, the term station is only used to refer to stations within that domain. For example, when performing traffic analysis on a particular Avaya Aura® enterprise, telephones in the PSTN are not considered to be stations in that enterprise. This represents a circumstance in which trunks are referred to as endpoints from the perspective of the enterprise of interest, even though they are not true points of call termination.

Normally, the set of enterprise configuration inputs includes a specification of the quantity, physical location, and logical association (for example, network region) of each type of station to be used in a particular enterprise. In some cases, the number of trunks is also specified, while in others, the number of trunks must be calculated as an output of the traffic-engineering process.

Traffic usages

Erlang and ccs definitions

Consider a stream of calls flowing across a group of trunks from one population of endpoints to another. The number of simultaneous calls traversing the trunks generally varies over time (that is, it increments by one every time a new call arrives on an available trunk, and it decrements by one every time an existing call terminates). The corresponding *carried load* (or *usage*), expressed in Erlangs, is defined as the *average number of simultaneous calls* that are traversing the trunks during a given time period (for example, during the busy hour). Note that in this example, the number of active calls always equals the number of busy trunks (since each active call requires exactly one trunk). Therefore, the *call usage* (that is, the average number of simultaneous active calls) equals the *trunk usage* (that is, the average number of simultaneous busy trunks) in this example.

If a call arrives while all trunks are busy, it is said to be blocked at the trunk group. In other words, not all calls that are offered to the trunks are actually carried by the trunks. Accordingly, the corresponding offered load, expressed in Erlangs, is defined as the average number of simultaneous calls that would have been traversing the trunks during a given time period (for example, during the busy hour), had there been enough trunks to prevent blocking. Note that in this example, the offered call load (that is, the average number of simultaneous active calls had there been enough trunks to carry all call attempts) equals the offered trunk load (that is, the average number of simultaneous busy trunks had there been enough trunks to carry all call attempts) in this example.

To summarize so far, the traffic load, expressed in Erlangs, represents the average number of simultaneous active calls or busy resources, during a given time period (for example, the busy hour).

Also note that the usage of a single station, when expressed in Erlangs, represents the fraction of time the station is in use. For example, a station that carries 0.1 Erlang of usage is busy 10% of the time (during the time interval of interest; for example, the busy hour).

Two fundamental characteristics of a stream of call traffic are the call rate (usually expressed in calls per hour) and the average call duration (usually expressed in seconds). The corresponding call usage can be defined as follows:

Usage (in Erlangs) = [(calls per hour)(seconds per call)]/3600

Note that in some traffic reports, the call rates are termed as call counts. If a particular report is associated with a period of time other than one hour, care must be taken not to mistakenly apply the call count as the calls per hour in the preceding formula. Be careful to convert call counts to calls per hour before applying the formula.

The term ccs stands for centum call seconds, which is a period of time 100 s in duration. To minimize confusion, although ccs is technically a unit of time and could be used as such, in this case it is only used to designate traffic loads.

Recall that a traffic load expressed in Erlangs is tacitly associated with a given time period (typically one hour). If that is the case, the relationship between a traffic load expressed in Erlangs and that same load expressed in ccs is:

Usage (in Erlangs) = Usage (in ccs)/36

However, consider a case in which a particular load, expressed in Erlangs, represents the average number of simultaneous active calls or busy resources, during a given time period other than one hour. In such a case, the denominator in the preceding expression should be set to equal the number of 100 s intervals in the time period of interest.

Finally, since a single station carrying one ccs of traffic is busy for 100 s during the busy hour, the maximum traffic that can be carried by a single station or trunk is 36 ccs.

Erlang B and C models

The Erlang B model is used to represent a situation in which calls that arrive when all resources (for example, trunk channels) are busy, are blocked and subsequently denied service. The model further assumes that the calls follow Poisson arrival and departure processes (which is typical for actual calls in real configurations), and that blocked calls never retry.

There are four parameters associated with the Erlang B model:

- Offered load (Erlangs)
- Carried load, which is sometimes referred to as usage (Erlangs)
- Number of resources
- GoS (grade of service, which is the probability of blocking at the resources)

Normally, when working with anticipated traffic loads (for example, for a new configuration), we work with the offered load, the number of resources, and the GoS. On the other hand, when working with measured traffic loads (for example, found on traffic reports run on existing configurations), we work with the carried load (usage), the number of resources, and the GoS. In either case, given any two of the three relevant values, the Erlang B model produces the third value.

The Erlang C model is used to represent a situation in which calls that arrive when all resources (for example, trunk channels) are busy, are blocked and subsequently queued. Like the Erlang B model, the Erlang C model is predicated on the assumption that the calls follow Poisson arrival and departure processes. Furthermore, the Erlang C model assumes an infinite amount of space in the queue.

There are three parameters associated with the Erlang C model:

- Offered load, which equals the carried load in this model (Erlangs)
- Number of resources
- GoS (grade of service, which is the probability of blocking at the resources)

Given the values of any two of those three parameters, the Erlang C model produces the third value.

Note that the GoS is often expressed as P01 or P001. P01 represents at most 1 call out of every 100 being blocked at the resource of interest (that is, 1% blocking), and P001 represents at most 1 call out of every 1000 being blocked at the resource of interest (that is, 0.1% blocking).

Consider a situation in which a call that is blocked is constantly retried until it receives service, meaning that as soon as a busy signal is heard, the caller hangs up and immediately redials. This is the most extreme form of retrial, and it is almost as if each blocked call is simply placed in queue and receives service as soon as a resource frees up for it. In other words, the Erlang C model is a reasonable approximation for constant retrials.

So, since Erlang B represents no retrials and Erlang C approximates constant retrials, the average of the two models is a reasonable approximation for moderate retrials. In this document, the pure Erlang B model is used when ignoring the effect of retrials, and the average of the Erlang B and C models (that is, a mixed Erlang B/C model) is used when the effect of retrials is deemed to be relevant.

Although the Erlang C model deals with queueing effects, it is not a particularly reasonable model for inbound Call Centers unless the number of trunks is significantly higher than (for example, several orders of magnitude greater than) the number of agents. The M/M/c/k Finite Queue model, which is beyond the scope of this discussion, should be used instead. A pure Erlang C model is never used in this discussion.

Endpoint usages

The three fundamental components of general business call traffic are intercom (that is, calls between two enterprise stations), outbound (that is, enterprise station to PSTN trunk), and inbound (that is, PSTN trunk to enterprise station). There are two possible approaches for determining default values for the corresponding per-station call usages; one approach typically applies if the number of PSTN trunks is unknown and needs to be sized, and the other can only be applied if the number of PSTN trunks is known (or assumed to be a specific value) a priori.

Endpoint usages in a 1/3-1/3-1/3 call mix

In a general business environment, the intercom, outbound, and inbound call usages are often assumed to be equal. In other words, each of those three components represents 1/3 of the traffic.

The average duration of a general business call is typically assumed to be 200 s (20 s for call set-up, and 180 s of talk time) as a default. Furthermore, the average station is assumed to induce the following call rates during the busy hour.

Light General Business Traffic:

- originate 0.25 intercom call per hour
- originate 0.25 outbound call per hour
- receive 0.25 inbound call per hour

Moderate General Business Traffic:

- originate 0.50 intercom call per hour
- originate 0.50 outbound call per hour
- receive 0.50 inbound call per hour

Heavy General Business Traffic:

- originate 0.75 intercom call per hour
- originate 0.75 outbound call per hour
- receive 0.75 inbound call per hour

The corresponding default per-station busy-hour usages can be calculated using the preceding call rates, a 200-s average hold time, and the formulas in the Erlang and ccs definitions section.

Light General Business Traffic:

- originate 0.5 ccs = 0.014 Erlang of intercom call usage
- originate 0.5 ccs = 0.014 Erlang of outbound call usage
- receive 0.5 ccs = 0.014 Erlang of inbound call usage

Moderate General Business Traffic:

- originate 1.0 ccs = 0.028 Erlang of intercom call usage
- originate 1.0 ccs = 0.028 Erlang of outbound call usage
- receive 1.0 ccs = 0.028 Erlang of inbound call usage

Heavy General Business Traffic:

- originate 1.5 ccs = 0.042 Erlang of intercom call usage
- originate 1.5 ccs = 0.042 Erlang of outbound call usage
- receive 1.5 ccs = 0.042 Erlang of inbound call usage

Endpoint usages driven by the number of trunks

If the number of PSTN trunks is known (or is assigned some assumed value as part of the given information), then an alternate approach to the one provided in Endpoint Usages in a 1/3-1/3-1/3 Call Mix can be used. Actually, the procedure for determining the per-station intercom usage is identical to the procedure used in the 1/3-1/3-1/3 model. The difference appears in the outbound and inbound usages; specifically, the outbound and inbound

components of the traffic are derived by assuming the trunks have been engineered to a P01 GOS. The results are as follows:

- The default per-station intercom usage either 0.5 ccs = 0.014 Erlang (light general business traffic), 1.0 ccs = 0.028 Erlang (moderate general business traffic), or 1.5 ccs = 0.042 Erlang (heavy general business traffic)
- The default per-station outbound usage is determined by calculating the carried load associated with the given number of outbound trunks, an assumed grade of service (P01 is standard for PSTN trunks), and the mixed Erlang B/C model
- The default per-station inbound usage is determined by calculating the carried load associated with the given number of inbound trunks, an assumed grade of service (P01 is standard for PSTN trunks), and the mixed Erlang B/C model

One drawback to using this method is that it assumes the trunks have been engineered to a P01 GOS. If the trunks are not being heavily used (for example, if a lot of extra trunks have been added solely for redundancy purposes), this model produces estimates for the outbound and inbound usages that are far greater than the actual usages.

Non-SIP Communication Manager

In this document, the term non-SIP Communication Manager refers to a Communication Manager supporting no SIP signaling groups. A non-SIP Communication Manager supports TDM stations (for example, DCP, analog, BRI) and H.323 stations. TDM stations can be administered to port networks and branch (H.248) gateways. H.323 stations can register to Communication Manager via C-LAN and IPSI circuit packs, or via Processor Ethernet. Since no SIP signaling groups are supported on non-SIP-enabled Communication Manager (by definition), such systems do not support SIP stations.

Note that the only way endpoints administered to a non-SIP-enabled Communication Manager can talk to endpoints elsewhere in the enterprise is via non-SIP trunks to a Communication Manager administered as either a feature server or evolution server.

Additional non-IMS elements

Communication Manager administered as an evolution server and embedded in a branch gateway connects to Session Manager via non-IMS SIP trunks. Other SIP elements connected to Session Manager via non-IMS trunks include:

- SIP stations
- SIP service providers (optionally via session border controllers)
- Avaya G860/Mediant 3000 media gateways
- SIP voice portals
- Avaya Aura® Conferencing Standard Edition

- Avaya Aura® Messaging
- Non-Avaya SIP gateways

Call types encountered in a Session Manager enterprise

For the purposes of identifying call flows and the corresponding Session Manager and Communication Manager SIP resources involved, the endpoints are consolidated into the following four categories:

- SIP stations

  SIP stations registered to Session Manager, using Communication Manager administered as either a feature server or an evolution server as feature source.

- Non-IMS SIP elements

  Non-SIP endpoints on a Communication Manager administered as an evolution server, endpoints on non-Avaya SIP gateways, SIP service providers, G860/M3K Trunk Gateways, SIP voice portals, and Messaging

- Non-SIP Communication Manager

  Endpoints on non-SIP-enabled Communication Manager

- Non-SIP PSTN trunks

  Endpoints in the PSTN that are connected to Session Manager via non-SIP trunking (the case of SIP trunking is covered in non-IMS SIP elements)

The call flows associated with the various combinations of the preceding endpoint types are described in more detail in the examples.

Session Manager call types: Example 1

Example 1 describes calls between two SIP stations within the same Session Manager instance with the same Communication Manager administered as either a feature server or evolution server.

Figure 21: Call between two SIP stations with same Session Manager instance and same Communication Manager on page 207 shows the signaling flow associated with a call between two SIP stations registered to the same Session Manager instance, and using the same Communication Manager as a feature source.

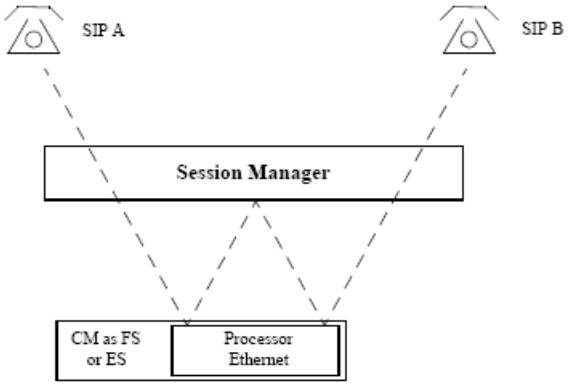


Figure 21: Call between two SIP stations with same Session Manager instance and same Communication Manager

The Session Manager resources associated with the call depicted in Figure 21: Call between two SIP stations with same Session Manager instance and same Communication Manager on page 207 include:

3 SIP sessions

- SIP A - Session Manager - Communication Manager
- Communication Manager - Session Manager - Communication Manager
- Communication Manager - Session Manager - SIP B

Communication Manager resources associated with the call depicted in Figure 21: Call between two SIP stations with same Session Manager instance and same Communication Manager on page 207 include:

- 2 SIP trunk channels if evolution server; 4 SIP trunk channels if feature server
- CPU for 2 SIP trunk call legs if evolution server; CPU for 4 SIP trunk call legs if feature server

Session Manager call types: Example 2

Example 2 describes calls between two SIP stations within the same Session Manager instance with different Communication Managers administered as either a feature server or evolution server.

Figure 22: Call between two SIP stations with same Session Manager instance and different Communication Managers on page 208 shows the signaling flow associated with a call between two SIP stations registered to the same Session Manager instance and using different feature servers, different evolution servers, or one of each.

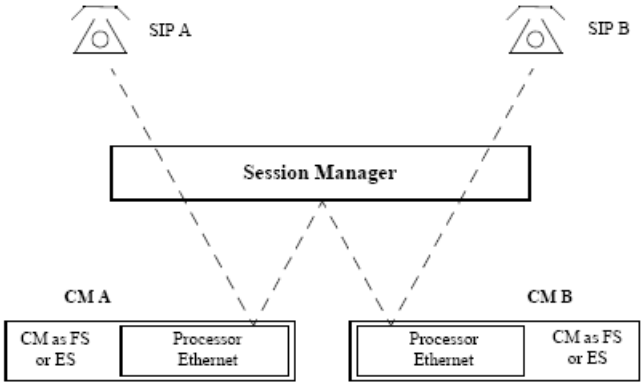


Figure 22: Call between two SIP stations with same Session Manager instance and different Communication Managers

Session Manager resources associated with the call depicted in Figure 22: Call between two SIP stations with same Session Manager instance and different Communication Managers on page 208 include:

3 SIP sessions

- SIP A - Session Manager - Communication Manager
- Communication Manager - Session Manager - Communication Manager
- Communication Manager - Session Manager - SIP B

Communication Manager resources associated with each Communication Manager for the call depicted in Figure 22: Call between two SIP stations with same Session Manager instance and different Communication Managers on page 208 include:

- 2 SIP trunk channels (for either feature server or evolution server)
- CPU for 2 SIP trunk call legs (for either feature server or evolution server)

Session Manager call types: Example 3

Example 3 describes calls between two SIP stations within different Session Manager instances with the same Communication Manager administered as either a feature server or evolution server.

Figure 23: Call between two SIP stations with different Session Manager instances and same Communication Manager on page 209 shows the signaling flow associated with a call between two SIP stations registered to different Session Manager instances and using the same Communication Manager as a feature server.

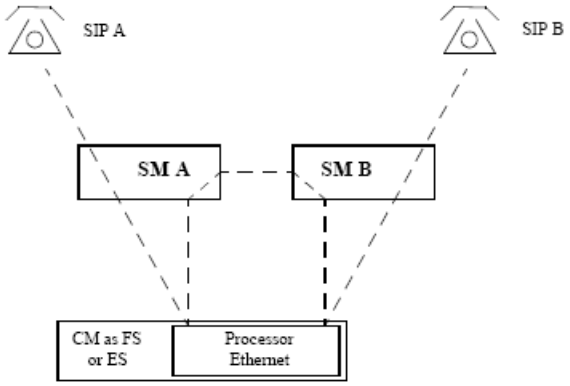


Figure 23: Call between two SIP stations with different Session Manager instances and same Communication Manager

Resources associated with each Session Manager instance for the call depicted in Figure 23: Call between two SIP stations with different Session Manager instances and same Communication Manager on page 209 include:

2 SIP sessions

- SIP A or SIP B - Session Manager - Communication Manager
- Communication Manager - one Session Manager - other Session Manager

Communication Manager resources associated with the call depicted in Figure 23: Call between two SIP stations with different Session Manager instances and same Communication Manager on page 209 include:

- 2 SIP trunk channels if evolution server; 4 SIP trunk channels if feature server
- CPU for 2 SIP trunk call legs if evolution server; CPU for 4 SIP trunk call legs if feature server

Session Manager call types: Example 4

Example 4 describes calls between two SIP stations within different Session Manager instances with different Communication Managers administered as either feature servers or evolution servers.

Call between two SIP stations with different Session Manager instances and same Communication Manager shows the signaling flow associated with a call between two SIP stations registered to different Session Manager instances and using a different feature servers, different evolution servers, or one of each.

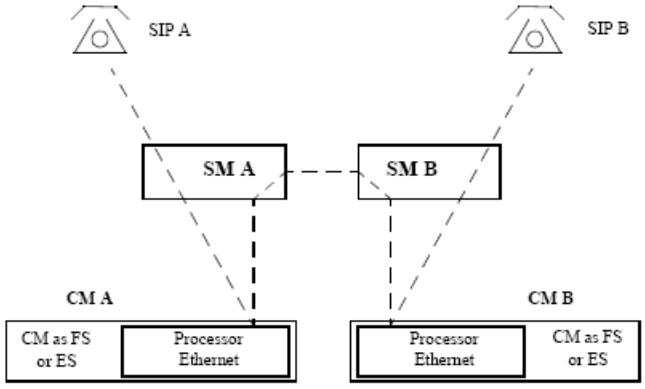


Figure 24: Call between two SIP stations with different Session Manager instances and different Communication Managers

Resources associated with each Session Manager instance for the call depicted in #unique_315 include:

2 SIP sessions

- SIP A or SIP B - Session Manager - Communication Manager
- Communication Manager - one Session Manager - other Session Manager

Communication Manager resources associated with each Communication Manager for the call depicted in #unique_315 include:

- 2 SIP trunk channels (for either feature server or evolution server)
- CPU for 2 SIP trunk call legs (for either feature server or evolution server)

Session Manager call types: Example 5

Example 5 describes calls between a SIP station and a non-IMS SIP element.

Figure 25: Call between a SIP station and a non-IMS SIP element on page 211 shows the signaling flow associated with a call between a SIP station and a non-IMS SIP element. Communication Manager is administered as either feature server or evolution server.

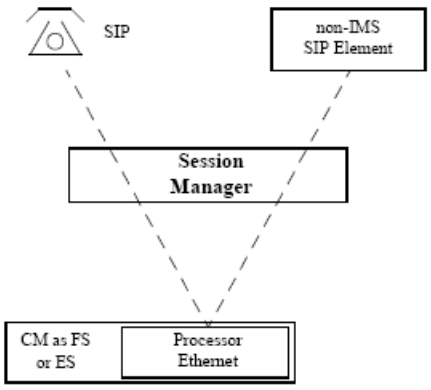


Figure 25: Call between a SIP station and a non-IMS SIP element

Session Manager resources associated with the call depicted in Figure 25: Call between a SIP station and a non-IMS SIP element on page 211 include:

2 SIP sessions

- SIP - Session Manager - Communication Manager
- Communication Manager - Session Manager - non-IMS SIP element

Communication Manager resources associated with the call depicted in Figure 25: Call between a SIP station and a non-IMS SIP element on page 211 include:

- Case 1

  Non-IMS SIP Element is a non-SIP endpoint on the same evolution server that's associated with the SIP station

  - 3 SIP trunk channels
  - CPU for 3 SIP trunk call legs and for 1 non-SIP call leg

- Case 2

  non-IMS SIP Element is a non-SIP endpoint on a different Communication Manager than the one that's associated with the SIP station or is any other type of non-IMS SIP element as defined at the beginning of the Call Types Encountered in an Session Manager Enterprise section.

  - 2 SIP trunk channels
  - CPU for 2 SIP trunk call legs

Session Manager call types: Example 6

Example 6 describes calls between a SIP station and a non-SIP Communication Manager or the PSTN. Communication Manager is administered as either feature server or evolution server.

Figure 26: Call between a SIP station and a non-SIP Communication Manager or the PSTN on page 212 shows the signaling flow associated with a call between a SIP station and a non-SIP Communication Manager or the PSTN.

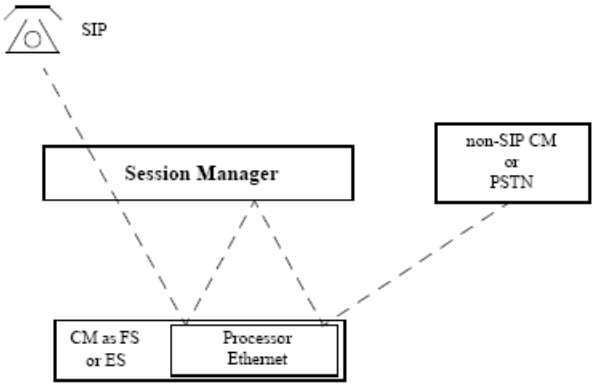


Figure 26: Call between a SIP station and a non-SIP Communication Manager or the PSTN

Session Manager resources associated with the call depicted in Figure 26: Call between a SIP station and a non-SIP Communication Manager or the PSTN on page 212 include:

2 SIP sessions

- SIP - Session Manager - Communication Manager
- Communication Manager - Session Manager - Communication Manager

Communication Manager resources associated with the call depicted in Figure 26: Call between a SIP station and a non-SIP Communication Manager or the PSTN on page 212 include:

- 3 SIP trunk channels
- 1 non-SIP trunk channel
- CPU for 3 SIP trunk call legs and for 1 non-SIP trunk call leg

Note:
Session Manager skips origination processing and application sequencing for emergency calling.

Session Manager call types: Example 7

Example 7 describes calls between two non-IMS SIP elements.

Figure 27: Call between two non-IMS SIP elements on page 213 shows the signaling flow associated with a call between two non-IMS SIP elements.

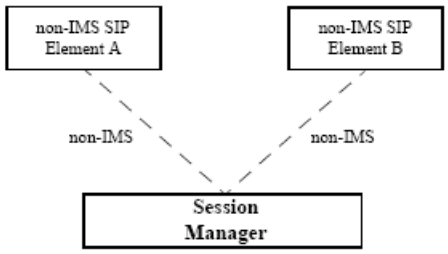


Figure 27: Call between two non-IMS SIP elements

Session Manager resources associated with the call depicted in Figure 27: Call between two non-IMS SIP elements on page 213 include

1 non-IMS - non-IMS SIP session

non-IMS SIP A - Session Manager - non-IMS SIP B

Session Manager call types: Example 8

Example 8 describes calls between a non-IMS SIP element and a non-SIP Communication Manager administered as either a feature server or evolution server.

Figure 28: Call between a non-IMS SIP element and a non-SIP Communication Manager on page 213 shows the signaling flow associated with a call between a non-IMS SIP Element and a non-SIP Communication Manager.

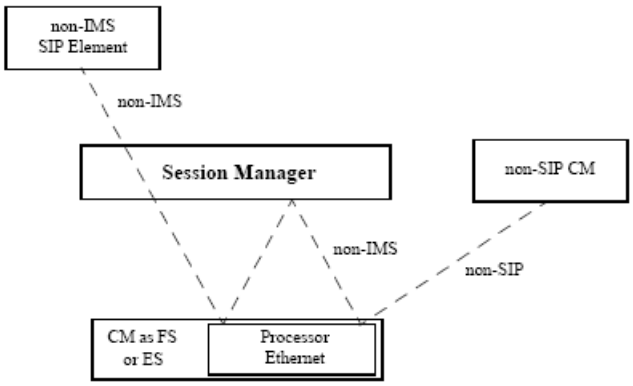


Figure 28: Call between a non-IMS SIP element and a non-SIP Communication Manager

Session Manager resources associated with the call depicted in Figure 28: Call between a non-IMS SIP element and a non-SIP Communication Manager on page 213 include:

2 SIP sessions

- non-IMS SIP Element - Session Manager - Communication Manager
- Communication Manager - Session Manager - Communication Manager

Note:

The signaling path from Session Manager to Communication Manager to Session Manager consists of two IMS SIP legs if Communication Manager is a feature server or two non-IMS legs if Communication Manager is an evolution server.

Communication Manager resources associated with the call depicted in Figure 28: Call between a non-IMS SIP element and a non-SIP Communication Manager on page 213 include:

- 3 SIP trunk channels
- 1 non-SIP trunk channel
- CPU for 3 SIP trunk call legs and for 1 SIP non-SIP trunk call leg

Session Manager call types: Example 9

Example 9 describes calls between two non-SIP Communication Managers administered as either feature servers or evolution servers.

Figure 29: Call between two non-SIP Communication Managers on page 214 shows the signaling flow associated with a call between two non-SIP Communication Manager.

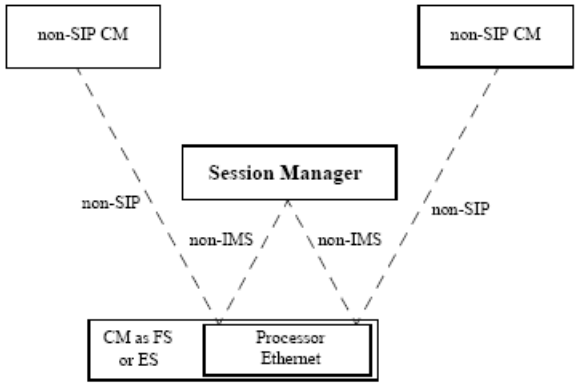


Figure 29: Call between two non-SIP Communication Managers

Session Manager resources associated with the call depicted in Figure 29: Call between two non-SIP Communication Managers on page 214 include:

1 non-IMS - non-IMS SIP session

Communication Manager - Session Manager - Communication Manager

Communication Manager resources associated with the call depicted in Figure 29: Call between two non-SIP Communication Managers on page 214 include:

- 2 SIP trunk channels
- 2 non-SIP trunk channels
- CPU for 2 SIP trunk call legs and for 2 non-SIP trunk call legs

Session Manager call types: Example 10

Example 10 describes calls between Communication Manager administered as an evolution server or a non-SIP Communication Manager and the PSTN.

We assume that the Communication Managers and non-SIP Communication Managers in a Session Manager enterprise supports their own non-SIP trunks directly to the PSTN.

Table 12: Session Manager SIP sessions required per call for various call types on page 215 summarizes the number of SIP sessions involved per Session Manager instance for each of the calls described in the Call Types Encountered in an Session Manager Enterprise section.

Table 12: Session Manager SIP sessions required per call for various call types

Endpoints involved in the call	SIP sessions per SM
Two Avaya SIP stations registered to the same Session Manager instance, using the same Communication Manager administered as either a feature server or evolution server.	3
Two Avaya SIP stations registered to the same Session Manager instance, using different Communication Managers administered as either feature servers or evolution servers	3
Two Avaya SIP stations registered to different Session Manager instances, using the same Communication Manager administered as either a feature server or evolution server.	2
Two Avaya SIP stations registered to different Session Manager instances, using different Communication Managers administered as either feature servers or evolution servers	2
An Avaya SIP station and a non-IMS SIP element	2
An Avaya SIP station and a non-SIP Communication Manager or the PSTN	2
Two non-IMS SIP elements	1
A non-IMS SIP element and a non-SIP Communication Manager	2
Two non-SIP Communication Managers	1
A Communication Manager as an access element or non-SIP Communication Managerand the PSTN via non-SIP trunks	—

Table 13: Communication Manager core resources required per call for various call types on page 216 summarizes the number of SIP trunk channels per Communication Manager administered as either a feature server or evolution server, the number of SIP trunk call legs per Communication Manager administered as either a feature server or evolution server, and the number of non-SIP trunk call legs per Communication Manager administered as either a

feature server or evolution server for each of the calls described in the Call Types Encountered in an Session Manager Enterprise section.

Table 13: Communication Manager core resources required per call for various call types

Endpoints involved in the call	SIP trunk call legs per FS or ES	Non-SIP trunk call legs per FS or ES
Two Avaya SIP stations registered to the same Session Manager instance, using the same core Communication Manager as a feature server	2 or 4[1]	NA
Two Avaya SIP stations registered to the same Session Manager instance, using different core Communication Managers as feature servers	2	NA
Two Avaya SIP stations registered to different Session Manager instances, using the same core Communication Manager as a feature server	2 or 4 [1]	NA
Two Avaya SIP stations registered to different Session Manager instances, using different core Communication Managers as feature servers	2	NA
An Avaya SIP station and a non-IMS SIP element	2 or 3[2]	0 or 1 [2]
An Avaya SIP station and a non-SIP Communication Manager or the PSTN	3	1
Two non-IMS SIP elements	NA	NA
A non-IMS SIP element and a non-SIP Communication Manager	3	1
Two non-SIP Communication Managers	2	2
A Communication Manager as access element or non-SIP Communication Manager and the PSTN via non-SIP trunks	NA	1

[1]SIP trunk channels per call if evolution server; 4 SIP trunk channels per call if feature server
[2]For a call between an Avaya SIP station and a non-IMS SIP element, the only time the larger numbers apply are when the non-IMS SIP element is a non-SIP endpoint on the same evolution server that is associated with the SIP station.

Each non-SIP Communication Manager involved in a call with another element in the Session Manager enterprise requires one non-SIP trunk channel for that call.

Engineering Session Manager instances

The two prominent performance-limiting factors associated with an Session Manager server are Session Manager server processing occupancy and memory constraints.

Session Manager processor occupancy is theoretically directly proportional to the rate at which Session Manager initiates and tears down SIP sessions. A SIP session consists of the signaling associated with a connection between two SIP trunks, communicating via a Session Manager instance. Table 12: Session Manager SIP sessions required per call for various call types on page 215 indicates that there is not generally a one-to-one correspondence between call and SIP session.

As a design criterion, the total occupancy (including the static occupancy) of the Session Manager server complex should not exceed 80%. That number is analogous to the 65% static + call-processing occupancy used in designing Communication Manager systems. The reason the design threshold is higher for Session Manager than Communication Manager is that Session Manager requires no processing cycles to be reserved for hardware maintenance activity.

Memory constraints establish a second, independent constraint, pertaining to the maximum number of simultaneous SIP sessions and the maximum number of TLS sockets supported by a single Session Manager instance.

Communication Manager traffic-engineering rules

This section discusses traffic-engineering rules associated with sizing various Communication Manager resources, including Communication Manager processor occupancy (which is directly related to BHCC capacities), TN799 C-LAN circuit packs, Processor Ethernet interfaces, TN2312 IPSI circuit packs, number of required gateways (for example, from a TDM timeslot perspective), trunk groups, media-processing resources, and TTR resources.

Processor occupancy and BHCC

A system's busy hour call attempt (BHCA) rate is the total number of calls attempted within that system during its busiest hour. This is distinct from a system's busy hour call completion (BHCC) rate, which counts only those calls that were actually completed. A system's call capacity refers to its BHCC rate.

In Communication Manager, processor occupancy (that is, server occupancy) is broken down into three fundamental categories: static occupancy (ST), call processing occupancy (CP), and system management occupancy (SM). Static occupancy refers to the processing required for keep-alive operations. Despite the nomenclature, the value of static occupancy in a List

Measurements report can vary somewhat. Call processing occupancy refers to the processing required for setting up, maintaining, and tearing down calls, and for executing vectoring operations in call centers. System management refers to the processing required for maintaining the sanity of the system, including periodic maintenance and audits.

Due to the bursty nature of system management functions, a fixed portion of the overall processing capacity is allocated to system management for design purposes. For all Communication Manager servers, 27% of the total system processing capacity is earmarked for system management. Note that the 27% occupancy is not actually dedicated solely to system management in practice; that number is only used for traffic configuration calculations.

If the overall processor occupancy (that is, ST + CP + SM) exceeds approximately 92%, all system management operations are temporarily postponed, and subsequent call attempts are disallowed (that is, call throttling is initiated). Therefore, systems should be designed such that the ST + CP occupancy is no more than 65%; that is,

100% - 27% for system management - 8% for the call throttling region.

Figure 30: Processing occupancy budgets for Communication Manager on page 218 shows the various occupancy budgets involved. To illustrate, the relationship between Communication Manager processor occupancy and the call rate is depicted as linear, although that is not always the case.

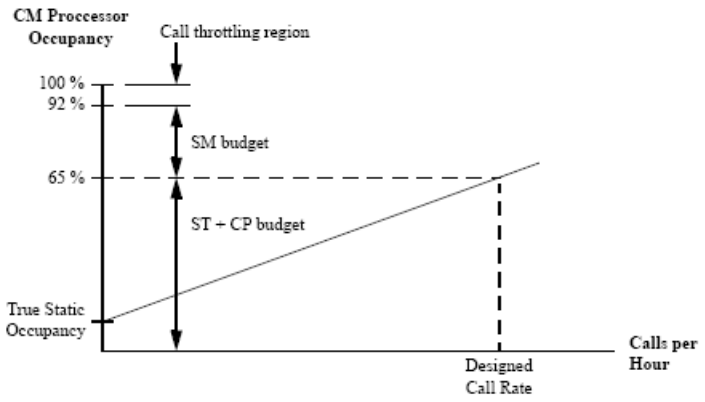


Figure 30: Processing occupancy budgets for Communication Manager

If the value of ST + CP occupancy is between 65% and 92%, some system management functions will be postponed to a quieter traffic period to allow static occupancy and call processing processes to use processor cycles from the system management budget. If the value of ST + CP occupancy exceeds 92%, all system management functions are suppressed and call throttling is initiated.

TN799 C-LAN circuit packs and Processor Ethernet

TN799 C-LAN circuit packs and Processor Ethernet represent the two types of interfaces by which IP endpoints, branch gateways, and adjuncts can register to Communication Manager.

Note that although the Communication Manager templates all come equipped with native Processor Ethernet, TN799 C-LAN circuit packs could be used in addition to the native Processor Ethernet. In fact, such systems can be administered to specify the use of C-LAN circuit packs as the first priority and Processor Ethernet as the second priority.

The two primary considerations when engineering the required number of C-LAN circuit packs or Processor Ethernet interfaces are registration and packet throughput. Registration considerations pertain to the maximum number of entities that can simultaneously reregister upon system failure. Packet throughput refers to the maximum number of simultaneous signaling connections supported by C-LAN circuit packs and Processor Ethernet. For practical purposes, these two concepts can be considered independent of one another. Therefore, the required number of C-LAN circuit packs or Processor Ethernet interfaces is determined based on each of the two criteria, and the maximum of the two results is used to ensure that both criteria are met.

Stable recovery registrations

When an IP endpoint, branch gateway, or adjunct registers to a C-LAN circuit pack, a software object known as a C-LAN socket is allocated to the registering entity, and the socket remains allocated as long as the entity remains registered. From a failure-recovery perspective, each TN799DP C-LAN circuit pack should not support more than 300 C-LAN sockets in simultaneous use.

A single C-LAN socket is capable of supporting either a single IP or SIP telephone, a single IP or SIP trunk signaling group, a single branch (H.248) gateway, or an adjunct application. However, in terms of registration during failure recovery, those sockets are not generally weighted equally. For example, the procedure for registering a branch gateway is far more complex than the procedure for registering a single IP or SIP telephone, because the former is capable of supporting multiple circuit packs and endpoints. Therefore, a branch gateway counts as the equivalent of approximately 15 to 20 IP or SIP telephones when estimating the number of C-LAN circuit packs and Processor Ethernet interfaces required to support a mass reregistration pursuant to an outage.

Registration considerations for adjuncts

Both TN799 C-LAN circuit packs and Processor Ethernet can be used to support adjuncts and services such as Call Management System (CMS), Avaya Aura® Messaging, Call Detail Recording (CDR), DMCC-based recording, and Application Enablement Services (AES). The following adjuncts and services merit special allocation of C-LAN or Processor Ethernet resources.

- If CMS is being used, dedicate one TN799 C-LAN circuit pack to support it. Alternatively, a single Processor Ethernet interface can support up to two CMS applications.
- If Messaging is being used, dedicate one TN799 C-LAN circuit pack to support it.
- If neither CMS nor Messaging is being used, but **Connect Adjuncts: Asynchronous Connections** field is set for **thru C-LAN**, then one TN799 C-LAN circuit pack must be

allocated to support adjuncts such as CDR recording and PMS links. Alternatively, a single Processor Ethernet interface can support up to two CDR applications.

- A single Processor Ethernet interface can support up to 16 AES applications.

The information provided in the preceding bullet items is used to determine the number of C-LAN circuit packs and/or Processor Ethernet interfaces that are required to support just the adjuncts, based on registration considerations.

C-LANs and PE interfaces and packet throughput

In addition to being able to support stable recovery registrations, C-LAN circuit packs must also be engineered to handle the anticipated packet traffic. In particular, a C-LAN circuit pack's on-board processor could be subject to overloading. Note that there is no analogous concern for Processor Ethernet.

The general process for engineering C-LAN circuit packs for packet throughput is similar to the procedure for analyzing the Communication Manager occupancy in terms of call rate (see Processor occupancy and BHCC). Specifically, each call type requires a certain number of processing cycles, whether that is CPU cycles on Communication Manager or CPU cycles on a C-LAN circuit pack's CPU. In the case of the Communication Manager CPU, the static and call processing occupancy is limited to 65% by design. The analogous limit for a C-LAN circuit pack is generally lower because of the bursty nature of packet traffic.

C-LAN and Processor Ethernet summary

The required number of C-LAN circuit packs is the maximum of the number required to satisfy recovery registration considerations and the number required to satisfy packet-throughput considerations. The required number of Processor Ethernet interfaces is based solely on recovery registration considerations for Communication Manager. Once you determine the total number of C-LAN circuit packs required, those circuit packs are typically distributed uniformly among the Avaya G650 Media Gateways within the port networks used in the configuration. However, asymmetric configurations are certainly permitted.

TN2312 IPSI circuit packs requirements

There are two general criteria to consider when determining the required number of active TN2312 IPSI circuit packs in a system. The required number of active IPSI circuit packs must be set equal to the maximum of those two results to satisfy both of them. Once the required number of active IPSI circuit packs is determined, it can optionally be doubled to determine the total number of IPSI circuit packs required for a duplicated IPSI circuit pack configuration; that is, two IPSI circuit packs per port network.

Data Link Connection Identifier considerations

Using TN2312 IPSI circuit packs to support ISDN endpoints, H.323 endpoints, branch gateways, SIP trunks, and certain control functions (for example, EALs, ASAI, asynchronous links, CSCN), requires Data Link Connection Identifier (DLC) resources on those TN2312 IPSI circuit packs. The DLCI resource capacity of each active IPSI circuit pack is 2480. Note that in a system with duplicated IPSI circuit packs, the duplicates *cannot* contribute to the DLCI capacity.

Each C-LAN socket and each ISDN D-channel requires one DLCI resource. Therefore, the minimum number of TN2312 IPSI circuit packs required to support DLCIs is the minimum number that is greater than the result in the following formula:

(total number of C-LAN sockets + total number of ISDN D-channels)/2480

The number of IPSI circuit packs used must be great enough to support all of the DLCI resources required for the system.

IPSI throughput

While the throughput bottleneck on a C-LAN circuit pack is its on-board CPU, the throughout bottleneck on an IPSI circuit pack is its packet interface buffer. To safeguard against buffer overflow, enough IPSI circuit packs must be used to ensure that no single IPSI circuit pack handles more than 15K busy hour call completions (BHCC).

Required number of branch gateways and port networks

To size a given group of branch gateways or port networks, three pieces of information are needed:

- Total call usage (expressed in Erlangs) involving circuit-switched endpoints within the particular group of branch gateways or port networks. Designate this usage by u_{total} for the purposes of this discussion.
- Call usage (expressed in Erlangs) associated with calls between two circuit-switched endpoints within the particular group of branch gateways or port networks. Designate this usage by u_{tdm} for the purposes of this discussion.
- Maximum call usage (expressed in Erlangs) supported by a single branch gateway or port network at the specified grade of service (GOS). Designate this usage by u_{GOS} for the purposes of this discussion. The value of u_{GOS} is determined by applying the Erlang B model to the number of time slot pairs available for bearer traffic and announcements and a specified GOS.
    - For a GOS of P001, use a value of u_{GOS} = 204.5 for port networks, a value of u_{GOS} = 200.8 for G450, and a value of 193.3 for G430 Branch Gateway.

- For a GOS of P000001 (essentially non-blocking), use a value of u_{GOS} = 178.5 for port networks, a value of u_{GOS} = 175.1 for G450, and a value of 168.2 for G430 Branch Gateway.
- The term u_{GOS} represents the maximum *port* usage (expressed in Erlangs) supported by a single branch gateway or port network at the specified GOS.

The number of port networks required for the given group of branch gateways or port networks is the smallest integer that is greater than or equal to the following formula:

$$\frac{(u_{total} + u_{tdm}) + \sqrt{(u_{total} + u_{tdm})^2 - 4u_{tdm}u_{GOS}}}{2u_{GOS}}$$

Sizing of PSTN trunks

To size the PSTN trunks in a general business configuration, apply the Erlang B model to the PSTN call usage (expressed in Erlangs) and a P01 grade of service. For a call center, a reasonable rule of thumb is to multiply the number of agents by 1.4 to derive the number of trunks. However, a more rigorous approach is to apply the M/M/c/k Finite Queue model.

Sizing of media processing resources

For simplicity, the following default assumptions regarding media-processing usage is applied throughout this section:

- SIP–SIP two-party calls always use SIP direct media (that is, no media-processing resources are required)
- SIP–H.323 and H.323–H.323 two-party calls always shuffle (that is, one media-processing channel is required for the first 20 s of the call for each party)
- SIP–TDM and H.323–TDM two-party calls always require one media-processing channel on the gateway to which the TDM endpoint is administered for the entire call duration
- Intergateway TDM–TDM calls always require one media-processing channel on each gateway involved in the call for the entire call duration
- Intragateway TDM–TDM calls require no media-processing resources
- Multiparty (that is, 3-party or greater) calls cannot shuffle and, therefore, require media resources regardless of the types of endpoints involved. Call recorders and service-observing devices count as parties.

A system’s media-processing resources can reside on any port network or branch gateway, and the rules for sizing those resources are contained in subsequent topics.

Port networks - TN2602 IP Media Resource 320

Each port network supports either one (simplex) or two (duplicated) TN2602 IP Media Resource 320 circuit packs, and each circuit pack supplies 320 media-processing channels. Since duplicated TN2602 circuit packs are in an active-active state, a pair of TN2602 circuit packs on a single port network supply a total of 640 channels.

Assuming shuffling is enabled and there is no conference calling and no Music on Hold, a TN2602 circuit pack is nonblocking. This is because the TDM time slots can only support up to 242 simultaneous two-party calls in a single port network. In such cases, each circuit pack effectively has only 242 channels instead of 320.

There are at least three cases in which a TN2602 IP Media Resource 320 circuit pack can be a blocking entity, despite there being only 242 time slot pairs per port network:

- No shuffling

  Suppose there are 160 simultaneous nonshuffled IP station-to-IP station calls using a single TN2602 circuit pack. In this case, all 320 media resource channels are in use (that is, one two-way talk path between each of the 320 IP stations involved to the TDM bus), but only 160 out of 242 time slot pairs (that is, 320 out of 484 time slots) are in use. Therefore, subsequent IP calls are blocked at the circuit pack, despite the fact that there is spare capacity on the TDM bus.

- Conference calls

  Suppose there are 106 simultaneous three-party IP station calls using a single TN2602 circuit pack. Since calls involving more than two IP endpoints can not shuffle, a total of 318 media resource channels are required (that is, one two-way talk path between each of the 318 IP stations involved to the TDM bus), but only 318 out of 484 time slots are in use. Therefore, subsequent three-party IP calls are blocked at the TN2602 circuit pack, despite the fact that there is spare capacity on the TDM bus.

- Music on Hold

  Suppose there are 320 simultaneous IP trunk calls using a single TN2602 circuit pack, and every call is listening to the same Music on Hold. A total of 320 media resource channels are required (that is, one two-way talk path for each of the 320 calls). Even though each caller is only listening and not talking, there is a talk path allocated in advance for an agent's voice, but only 321 out of 484 time slots are in use. Because all parties are listening to the same music in this example, they are all listening to the same time slot. Therefore, subsequent IP calls are blocked at the media resource circuit pack, despite the fact that there is spare capacity on the TDM bus.

Determining the number of TN2602 circuit packs

You use an algorithm to determine the required number of TN2602 circuit packs to support the 320 media-processing channels.

1. Determine the total call usage for media processing as follows:

   total usage = G.711 unencrypted CUR + G.711 encrypted CUR+ G.729 unencrypted CUR + G.729 encrypted CUR+ G.726 CUR + unencrypted T.38 fax and modem over IP CUR

2. Determine the required number of media processing channels:

   Apply an Erlang formula (mixed Erlang B / C is ideal, but pure Erlang B would also work) to the total usage and a suitable grade of service, for example, P001, to obtain the total number of media processing channels required.

3. Determine the required number of TN2602 circuit packs:

   Divide that number by 242 and round up to get the required number of TN2602 circuit packs.

Determining G430 Branch Gateway media resources

A single G430 Branch Gateway can be configured to support up to 105 media-processing channels. You can use the following algorithm to determine the required number of G430 Branch Gateways to support the specified call usage.

1. Determine the total call usage for media processing as follows:

   total usage = G.711 unencrypted CUR + G.711 encrypted CUR + G.729 unencrypted CUR + G.729 encrypted CUR + G.726 CUR + unencrypted T.38 fax and modem over IP CUR

2. Determine the required number of media processing channels:

   Apply an Erlang formula (mixed Erlang B / C is ideal, but pure Erlang B would also work) to the total usage and a suitable Grade of Service (for example, P001) to obtain the total number of media processing channels required.

3. Determine the required number of media processing channels.

4. Divide that number by 105 and round up to get the required number of G430 Branch Gateways.

Determining G450 Branch Gateway media resources

A single G450 Branch Gateway can be configured to support up to 320 media-processing channels. You can use the following algorithm to determine the required number of G450 Branch Gateways to support the specified call usage.

1. Determine the total call usage for media processing as follows:

   total usage = G.711 unencrypted CUR + G.711 encrypted CUR + G.729 unencrypted CUR + G.729 encrypted CUR + G.726 CUR + unencrypted T.38 fax and modem over IP CUR

2. Determine the required number of media processing channels:

   Apply an Erlang formula (mixed Erlang B / C is ideal, but pure Erlang B would also work) to the total usage and a suitable Grade of Service (for example, P001) to obtain the total number of media processing channels required.

3. Determine the required number of media processing channels.
4. Divide that number by 320 and round up to get the required number of G450 Branch Gateways.

Touch tone receivers

When a station user goes off-hook, Communication Manager assigns an available time slot and a touch tone receiver (TTR) that listens to that time slot. The TTR collects the digits, formats a message, and sends the message uplink to the Communication Manager server. Communication Manager sends a downlink message to disconnect the TTR after all the digits have been collected.

For all intercom and auto route selection (ARS) trunk calls, the port on the Tone Detector circuit pack is released immediately when the last digit is dialed. In the case of non-ARS calls (operator-assisted calls, international calls, credit card calls) where the number of digits in a call may be unknown, there is a 10-s time-out period after each digit. If no new digit is generated during this time-out period, the port on the Tone Detector circuit pack is disconnected from the calling station.

If all TTRs in the system are busy, the request is put in a queue. The event of a full queue is treated as an error and results in intercept treatment; that is, a reorder tone is returned to the caller.

TTRs are used to collect digits from the following originating endpoints:

- analog sets
- DCP sets
- DS1 OPS (line-side T1)
- DS1 OPS (line-side E1)
- BRI sets
- analog trunks

- RBS digital trunks (T1)
- CAS digital trunks (E1)

TTRs are *not* used to collect digits from the following originating endpoints:

- IP telephones and trunks
- SIP telephones and trunks
- PRI T1 trunks
- PRI E1 trunks

TTR resources are determined by the originating station or trunk. For an outbound PSTN call, its TTR resource must reside in the same port network or branch gateway as the originating station, which is not necessarily the same port network or branch gateway as the trunk. IP or SIP endpoints do not need the use of a TTR. Incoming DID calls that do not use touch-tone dialing do not require TTRs. Incoming PRI calls that use authorization codes *do* require TTRs.

TTRs are engineered to 0.001 blocking using the blocked calls cleared model. This is conservative in that there is a small (4 entries) buffer for calls who find all TTRs busy.

Default holding time values for the different calls can be obtained by multiplying the number of digits in the call by 0.65 s and adding 3 s, which represents the period from off-hook to the first digit. The TTR usage, expressed in Erlangs, is calculated by multiplying the TTR holding time by the calls per hour, then dividing by 3600. The Erlang B formula with a P001 grade of service is then used to determine the required number of TTR resources. Each G430 branch gateway supports 32 TTR resources, and each G450 branch gateway supports 64 TTR resources. The TTR resources on a port network are scalable through the use of various circuit packs supporting TTR.

IP network bandwidth requirements

There are two general categories of bandwidth requirements: the bandwidth to support the media, and the bandwidth to support the signaling.

Media bandwidth

An IP packet consists of a payload and some amount of overhead, where the payload consists of actual sampled voice, and the overhead represents headers and trailers, which serve to navigate the packet to its proper destination. The overhead due to IP, UDP, and RTP is 40 bytes, while the Ethernet overhead is between 18 and 22 bytes (18 is assumed in this discussion). This represents a total overhead of 58 bytes (464 bits), regardless of the nature of the payload. For this example, Layer 2 (Ethernet) overhead has been included in that total.

At every router boundary, because we have included Ethernet overhead in this example, our calculations are for bandwidth on a LAN. As WAN protocol (for example, ppp) Layer 2 headers are generally smaller than Ethernet headers, WAN bandwidth is typically less than LAN bandwidth.

The size of the payload depends upon certain parameters relating to the codec being used. The two most common codecs used with Communication Manager products are (uncompressed) G.711 and (compressed) G.729. The transmission rates associated with those codecs are 64 kbps for G.711 (this is the Nyquist sampling rate for human voice) and 8 kbps for G.729.

The packet size is sometimes expressed in units of time (specifically, in milliseconds). The following formula yields the packet size, expressed in bits:

number of bits of payload per packet = transmission rate (kbps) x milliseconds per packet

Table 14: Payload size per packet on page 227, which has been populated using this formula, provides the payload size per packet (expressed in bits), as a function of packet size (milliseconds per packet) and codec:

Table 14: Payload size per packet

Packet Size	G.711	G.729
10 ms	640 bits	80 bits
20 ms	1280 bits	160 bits
30 ms	1920 bits	240 bits
60 ms	3840 bits	480 bits

Note that the number of bits of payload per packet depends on the packet size, but it is independent of the sizes of the individual frames contained in that packet. For example, a packet size of 60 ms could be referring to six 10-ms frames per packet, or three 20-ms frames per packet, or two 30-ms frames per packet. Presently, the most commonly-used packet sizes are 20 ms. Both G.711 and G.729 codecs typically utilize two 10-ms frames per packet.

As stated earlier, there is typically an overhead of 464 bits per packet in a LAN scenario. So, the bandwidth (expressed in kbps) associated with a unidirectional media stream (assuming no Silence Suppression is used) is augmented from 64 kbps and 8 kbps (for G.711 and G.729, respectively) to account for this overhead. The results of this exercise are provided in the Table 15: Typical LAN bandwidth requirements for media streams on page 227:

Table 15: Typical LAN bandwidth requirements for media streams

Packet Size	G.711	G.729
10 ms	110.4 kbps	54.4 kbps
20 ms	87.2 kbps	31.2 kbps
30 ms	79.5 kbps	23.5 kbps

Packet Size	G.711	G.729
60 ms	71.7 kbps	15.7 kbps

The kilobits per second values in Table 15: Typical LAN bandwidth requirements for media streams on page 227 were calculated by multiplying the transmission rate by the ratio of the total bits per packet (payload plus overhead) to the payload bits per packet. For example, for the G.711 codec, 20–ms packets, and 58 bytes of overhead per packet, the bandwidth per call is

(64 kbps)[(1280 + 464) / 1280] = 87.2 kbps

Note that the entries in Table 15: Typical LAN bandwidth requirements for media streams on page 227 correspond with *unidirectional* media streams. A *full-duplex* connection with a kilobits per second capacity at least as large as the number in one of the table cells would be sufficient for carrying a two-way voice stream using the corresponding codec, packet size, and packet overhead. In other words, a full-duplex connection with a particular capacity rating would support enough bandwidth to carry that capacity in both directions. Alternatively, two half-duplex connections of the same capacity rating could be used.

99.9th percentile traffic

The call usage (expressed in Erlangs) between two facilities represents the *average* number of simultaneous bidirectional media streams between those facilities. For example, if the call usage between two facilities is 100 Erlangs, then the average number of simultaneous calls is 100. However, since this is only an average, roughly 50% of the time there are more than 100 simultaneous active calls. So, it would be a mistake to simply multiply 100 media streams by the appropriate value for kbps per stream.

The goal is to supply enough bandwidth to adequately support the media streams at least 99.9% of the time. Given a call usage, the Erlang B model is used to estimate the 99.9th percentile value for the number of simultaneous streams. For example, if the call usage rate is 100 Erlangs, the Erlang B model implies that there are at least 128 simultaneous media streams less than 0.1% of the time. So, in that example, it is sufficient to engineer the bandwidth to support 128 simultaneous media streams.

Once you determine the 99.9th percentile for the number of simultaneous media streams, it can be converted to kilobits per second by using the numbers in Table 15: Typical LAN bandwidth requirements for media streams on page 227. For example, for a typical LAN configuration, the G.711 codec, and a packet size of 20 ms, Table 15: Typical LAN bandwidth requirements for media streams on page 227 implies that 87.2 kbps are required per call. In that case, the required bandwidth would be 87.2 kbps x 128 = 11.2 Mbps in each direction.

Call Admission Control

A Call Admission Control (CAC) limit can be administered to any pair of Communication Manager network regions, and it can be specified as either the maximum number of

simultaneous calls between the two network regions or the maximum bandwidth usage between the two network regions. Numbers such as 128 maximum simultaneous calls or 11.2 Mbps derived in the example in the previous section could serve as effective lower bounds for CAC limits.

IGAR and traffic engineering

Inter-Gateway Alternate Routing (IGAR) provides alternative routing over the PSTN for inter gateway calls that would otherwise be precluded from traversing the IP network. Communication Manager offers the capability to use H.323 and SIP trunks in the alternative routes. The reasons for an intergateway call to be rerouted over the PSTN include:

1. The Call Admission Control limit for the link in question was already reached
2. VoIP RTP resources are unavailable
3. The parties on the call are members of incompatible (in the sense of codec) network regions
4. The call was forcibly redirected over the PSTN for testing or debugging purposes

Dial Plan Transparency is somewhat similar to IGAR in that calls whose primary routes are through IP networks are rerouted through the PSTN. However, IGAR applies only to intra-Communication Manager calls, and Dial Plan Transparency applies only to inter-Communication Manager calls. For example, consider a Communication Manager system in which endpoints in two distinct geographic sites can only talk to each other via a particular WAN or via the PSTN. Suppose that the WAN is lost because of a failure, and that the main server complex is coresident with one of the two sites. In that case, the other site must have a survivable core or remote server to keep the endpoints in that site active. In such a scenario, the call in question becomes an inter-Communication Manager call (that is, a call between an endpoint controlled by the main server and an endpoint controlled by a survivable server), and could be rerouted through the PSTN through the use of Dial Plan Transparency. IGAR would not apply to such a scenario.

When engineering a configuration supporting IGAR or Dial Plan Transparency, it is important to engineer the PSTN trunks to be able to support the traffic that would be rerouted if IGAR or Dial Plan Transparency was invoked. For example, if Dial Plan Transparency is being used to provide inter-site connectivity over the PSTN in the event of a WAN failure, the PSTN trunks in both sites should be engineered to an appropriate grade of service, assuming the PSTN call usage includes all of the traffic that would be rerouted pursuant to a WAN failure. For more information see Sizing of PSTN trunks.

Signaling bandwidth

The signaling bandwidth is normally considerably smaller than the corresponding media bandwidth. However, we often must estimate it, especially in SIP configurations and when

separating the bearer and signaling network. Two components typically make up signaling bandwidth:

- Bandwidth supporting keep-alive signaling
- Bandwidth supporting per-call signaling.

The value of the keep-alive signaling and per-call signaling associated with a particular configuration depends on the types of endpoints and gateways involved and must be determined empirically. Once we determine the per-call signaling bandwidth for the various call types involved, those values are multiplied by the corresponding call rates, and those results are then added together.

We can then apply the Erlang B formula with a P001 grade of service to determine the 99.9th percentile bandwidth. See 99.9th percentile traffic on page 228.

Index

Numerics

A

B

C

D

E

F

G

O

P

Q

R

S

T

U

V