Designing a Long-Distance VoIP Network

Version History

Version Number	Date	Notes
1	1/26/2001	This document was created.

The long-distance Voice over IP (VoIP) network solution is a set of network design and configuration strategies that provide trunk-level transport of global switched telephone traffic distributed over VoIP. Calls originate in the Public Switched Telephone Network (PSTN), are routed through interexchange carriers (IXCs), and are handed off to a wholesale VoIP carrier for transport. To the subscriber, the long-distaance service seems like any other inexpensive long-distance service. To the originating long-distance carrier, the wholesale carrier is only one of a number of termination options.

The long-distance VoIP network solution offers service providers the required architecture design, network components, software features, functional groups, and provisioning methodologies needed to run a VoIP wholesale service. This solution enables the service provider to design a wholesale network and sell unbranded voice services to retailers, such as internet telephony service providers (ITSPs), application service providers (ASPs), IXCs, and post, telephone, and telegraph (PTTs) administrations.

This document describes the fundamentals of long-distance VoIP and provides a nine-step methodology for designing and implementing a long-distance VoIP network solution. This document is intended to cover the high-level design of a long-distance VoIP network; therefore, it does not discuss specific configuration information.

This document includes the following sections:

- Long-Distance VoIP Network Overview
- Long-Distance VoIP Design Methodology

Long-Distance VoIP Network Overview

The long-distance VoIP network solution includes multiple components in various combinations from both Cisco and third-party vendors. Voice points of presence (POPs) that are connected to other service providers are a central component in the delivery of wholesale voice services. The types of interconnections or call topologies service providers support will determine the specific components and design methods we recommend. Service providers use the call topologies to build a set of *deployment scenarios* that enable wholesale applications. Figure 1 shows some of the interconnection possibilities.

Figure 1   *Long-Distance VoIP Network Solution Interconnection Possibilities*
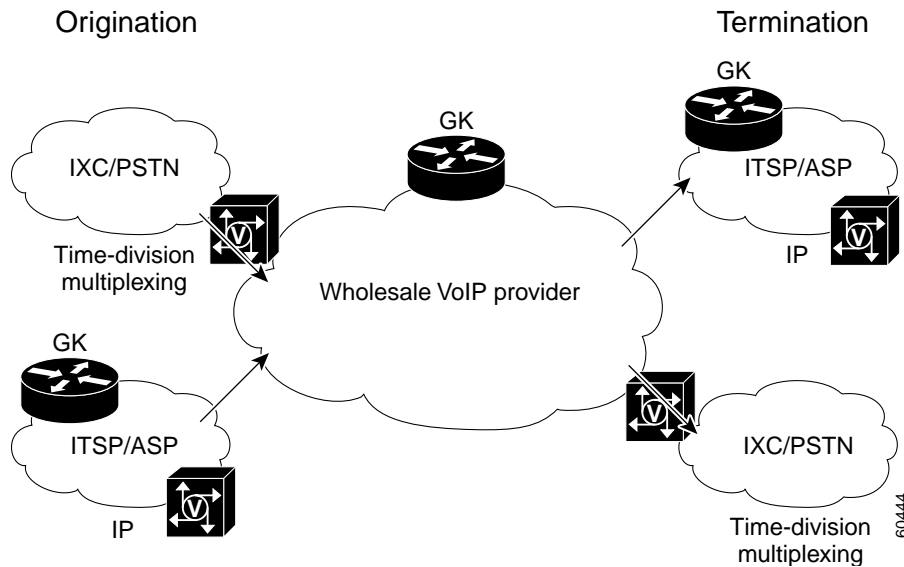
Long-Distance VoIP Network Benefits

The long-distance VoIP network solution provides the following benefits:

- Voice quality that is indistinguishable from that of the PSTN

- A cost-effective, reliable VoIP network infrastructure

- Support for least-cost routing and other enhanced call-routing methods

- Intercarrier call authorization and accounting (peer-to-peer)

- Support for intercarrier clearing and settlement services

- Support for local, national, and international dial plans

- Connectivity with the PSTN over carrier interfaces

- Connectivity with other VoIP service providers and the VoIP equipment of other vendors

- A worldwide network of other VoIP service providers interested in interconnecting

Long-Distance VoIP Design Methodology

To design your own personalized long-distance VoIP solution, systematically perform the following Cisco-recommended steps:

Step 1	Identify the services you plan to sell.
Step 2	Identify the type of carriers or providers with which you plan to interconnect.
Step 3	Determine the interconnection types you plan to use.
Step 4	Determine the call topologies you plan to use.
Step 5	Identify the appropriate deployment scenario.

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

Step 6   Identify the functional areas you require.

Step 7   Identify the required hardware and software components.

Step 8   Identify design and scalability issues.

**Step 9**   Configure and provision components.

Step 1: Identify Services

A key feature of the Cisco long-distance VoIP solution is its ability to support various mixes of services to suit the needs of a single service provider or multiple partnering service providers. Supported services are described in the following sections:

- Minutes Aggregation and Termination (Including ASP Termination)
- Calling Card Services
- Clearinghouse Services
- Service Options

Figure 2 depicts all of the components that may be needed to provide these services. These components include gatekeepers (GKs), gateways (GWs), signaling link termination (SLTs) equipment, signaling controllers (SCs), and intermachine trunks (IMTs).

Figure 2   High-Level View of End-to-End Service Possibilities
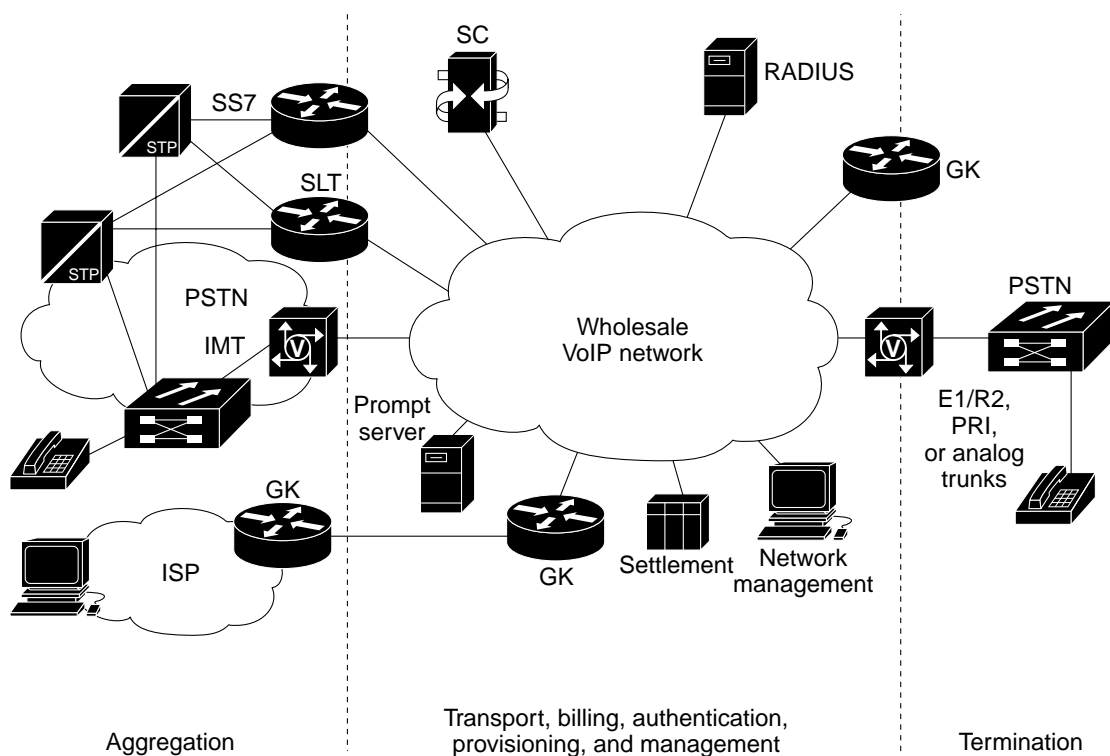
Minutes Aggregation and Termination (Including ASP Termination)

The Cisco wholesale voice solution supports the originating carrier that hands calls over to a VoIP wholesaler at a profit. Termination settlement rates are generally lower than PSTN termination rates—the key reason why long-distance carriers will choose a VoIP carrier for termination. Furthermore, termination bandwidth is often available over VoIP to countries where PSTN termination is unavailable because of congested international gateway facilities or other reasons. The average call success rate is as good as or better than that provided by PSTN carriers, and voice quality, including echo cancellation, is uncompromised.

Key features of this service include the following:

- H.323 VoIP interconnect using standards-based H.323 implementation
- Gatekeeper LRQ forwarding for call routing and accurate call accounting
- Support for voice, modem, and fax calls
- Support for SS7, T1/E1 E&M, E1 R2, and E1 PRI interfaces

As part of this service, ASP carrier to carrier termination services are supported. The ASP originates the call, often over an Internet-enabled PC-telephony application, or through a PSTN portal for cellular phone callers. The ASP provides pre-call services, such as content delivery (prerecorded messages, voice mail, private number dialing) or supervision-related services, such as "find me/follow me." The ASP then hands off any long-distance calls to a wholesale carrier for termination by the PSTN. This service requires accurate call accounting.

Calling Card Services

The Cisco wholesale voice solution supports the following calling card services:

- Prepaid—A wholesale VoIP carrier can host prepaid services for multiple service providers on its infrastructure. In addition, most prepaid service providers use VoIP wholesalers to terminate long-distance calls that are placed by prepaid subscribers. Using the integrated voice response (IVR) feature in Cisco wholesale VoIP gateways, and the real-time authorization and call accounting systems provided by Cisco Ecosystem Partners, you can offer this service over a VoIP network and lower the cost and deployment time of calling card services.

- Postpaid—Like the prepaid service, this service can be hosted by a wholesale VoIP carrier. An example is basic calling that is accessed by the 800 prefix, a calling card number, or a personal identification number (PIN). Postpaid service is similar to the prepaid service, except that with postpaid service the authorization is not tied to call rating. Consequently, call rating need not happen in real time, and there may be more partner billing-system options that perform adequately at scale. After calls are made, a billing system contracted by the company charges the carrier.

Figure 3 illustrates a variety of calling card services, including those provided by third parties.

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

Figure 3   Calling Card Services
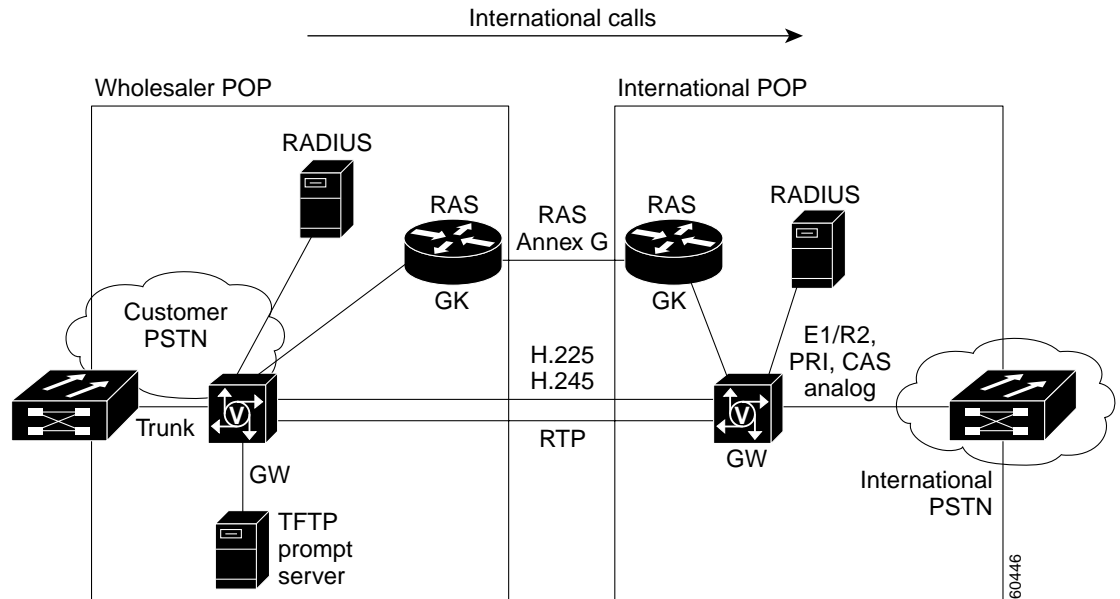
Clearinghouse Services

When multiple partners join to provide wholesale voice services, the services described in the preceding sections may require the assistance of clearinghouse services for billing and settlement. The Cisco wholesale voice solution supports call termination agreements through Open Settlement Protocol (OSP) in Cisco devices.

OSP relies upon Cisco Open Packet Telephony (OPT) framework at the call control layer. Service providers that use OSP (the only standard IP interface for VoIP clearinghouse functions), must do business with only one settlements provider. As a result, there is no need to negotiate separate agreements with carriers in multiple countries, meet varied technical requirements for interconnection, make repeated arrangements for call accounting, or establish multiple credit accounts. The OSP clearinghouse solution virtually eliminates the risk of doing business with new service providers that have a limited credit history—or with carriers in countries subject to currency fluctuations. In addition, it gives virtually every VoIP provider the worldwide calling reach it requires.

OSP uses a standard protocol approved by the Internet Protocol Harmonization over Networks organization of the European Telecommunications Standards Institute (ETSI TIPHON).   By allowing gateways to transfer accounting and routing information securely, this protocol provides common ground between VoIP service providers. Consequently, third-party clearinghouses with an OSP server can offer call authorization, call accounting, and settlement—including all the complex rating and routing tables necessary for efficient and cost-effective interconnections.

In most cases, a wholesale provider will subcontract with a clearinghouse partner to provide wholesale voice services with proper settlement. However, a clearinghouse solutions vendor can also independently take advantage of the Cisco wholesale voice solution to achieve market objectives.

Service Options

In addition to the services previously listed, the other two service options are described in the following sections:

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

- Limited Egress Carrier-Sensitive Routing
- Interconnection to Clarent-Based Clearinghouses

Limited Egress Carrier-Sensitive Routing

As an enhancement to simple carrier-interconnection applications, the Cisco wholesale voice solution makes it possible to route a call to different destination carriers. You have the same considerations as with simple carrier-interconnection models, but with slightly increased call-routing responsibilities. The directory gatekeeper can make limited egress carrier-sensitive routing (CSR) decisions by using the sequential LRQ feature, which is available to the applications using directory gatekeeper routing. Generally speaking, this means any TDM partners and directory gatekeeper peering partners, but also includes any OSP partners in which an OSP interconnection zone is used, as opposed to a direct implementation on your gateways.

In this CSR application, the sequential LRQ feature is used to route a call to different carriers, each of which supports a different destination. For example, you may provision your gatekeepers to route certain destination patterns to carrier A first. If carrier A (an Internet telephony service provider, or ITSP) is unavailable as a result of a location request reject (LRJ) or LRQ timeout, you may decide to route the call to carrier B (an interexchange carrier, or IXC), then to carrier C, and so on. Figure 4 illustrates this application.

Figure 4   Limited Egress Carrier-Sensitive Routing
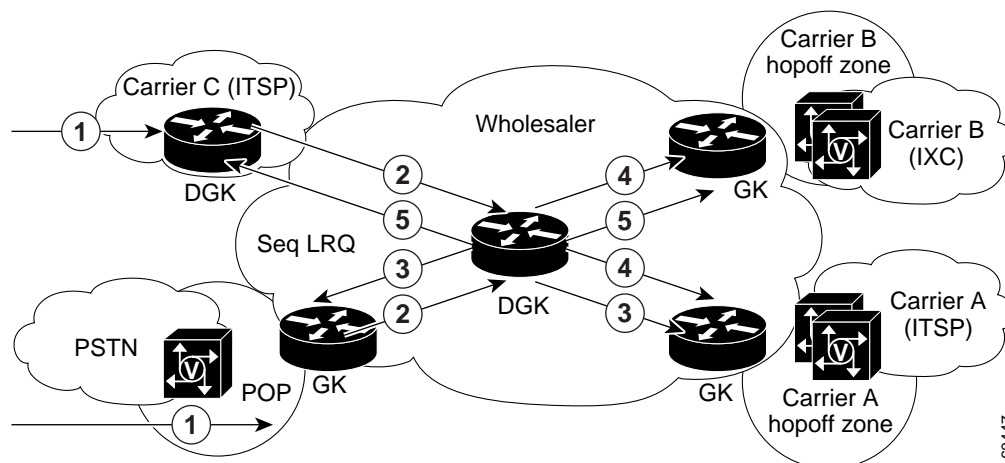
There are three restrictions to remember with limited egress CSR:

- Independence of ingress and egress carriers. The egress carrier is selected independently of the source carrier. The gatekeeper routes calls on the basis of dialed number identification service (DNIS). The list of possible egress carriers that you statically configure are tried in order. Routing decisions are not based on which carrier sourced the call. For example, the fact that carrier A sourced the call does not influence the carrier on which the call will be terminated.

- Independence of destination carriers. Each destination carrier must be contained in its own zone. For ITSP carriers, this is fairly simple. Interconnected ITSPs are seen as single remote zones to which your directory gatekeeper sends LRQ messages. For interconnected TDM carriers, this implies that the gateways that can send calls to the carrier are grouped into their own *hopoff* zone that is managed by a gatekeeper, and that multiple carriers are never supported by a single gateway.

- Static versus dynamic routing. Dynamic routing decisions are not supported; you configure the order of sequential LRQs statically. Consequently, there is no provision for percentage-based routing, maximum minute cutoffs, and so on. Egress carriers are always chosen from a statically

configured list of routes. If the directory gatekeeper determines that an OSP interconnection zone handles a route, it is possible that the OSP server returns a terminating gateway on the basis of advanced routing logic (if so provisioned). For example, the OSP server may dynamically select a least-cost, terminating carrier on the basis of time of day or best voice quality.

Interconnection to Clarent-Based Clearinghouses

You can interconnect with a Clarent-based service provider (URL www.clarent.com) provided that the gateways register to a Clarent gatekeeper; however, this would mean dedicating specific gateways to be part of the Clarent zone. Back-to-back gateways may be used to provide a "transit" zone between the Cisco and the Clarent-based network. One of the back-to-back gateways registers to a Clarent gatekeeper in the Clarent-based service provider network; the other registers to a Cisco gatekeeper in your network. This architecture is very similar to using back-to-back gateways to interconnect OSP partners, except that here the relationship is H.323 gateway to gatekeeper instead of OSP.

There are two limitations to using Clarent-based interconnection as follows:

- IP-to-IP interconnection. The use of back-to-back gateways enables Clarent-based interconnection partners to exchange traffic not only with wholesaler TDM-based interconnections, but also with other IP-based interconnection partners. Those partners may be either directory gatekeeper- or OSP-based. It may be necessary to modify the dial plan architecture to support directory gatekeeper-based IP carrier interconnections.

- Interoperability considerations. Before interconnection is possible with Clarent-based networks, H.323 interoperability must be sustained between Cisco gateways and Clarent gatekeepers. Currently, only voice-bearer interoperability is supported for G.711, G.723.1, and G.729 codec types. Because of tandem compression, back-to-back gateways impair voice quality.

Step 2: Identify Carriers or Providers

As a wholesale voice service provider, you need to interconnect with other service providers (ITSPs and ASPs) and carriers (IXCs and PSTNs) in order to offer the services you selected in Step 1. This interconnection method is referred to as a *call topology*. Because each call topology is specific to the carrier or service provider with which you plan to connect, you need to first identify the targeted carriers and service providers.

Step 3: Determine Interconnection Types

There are two application interconnection types you can use to interconnect with other service providers: IP and TDM. Your call topology is determined by the application interconnection type you use. The line of demarcation between you and the other service providers determines whether the interconnection type is IP or TDM.

Step 4: Determine Call Topologies

The call topology influences the ultimate configuration requirements of the functional areas within the network to support a given application. For example, if you enable simple carrier interconnection between an ASP and an IXC, then you would use an IP-to-TDM call topology. You would then need to address the configuration requirements for that application (such as call routing and shared support services needed for billing, settlement, and security options) as influenced by that topology type.
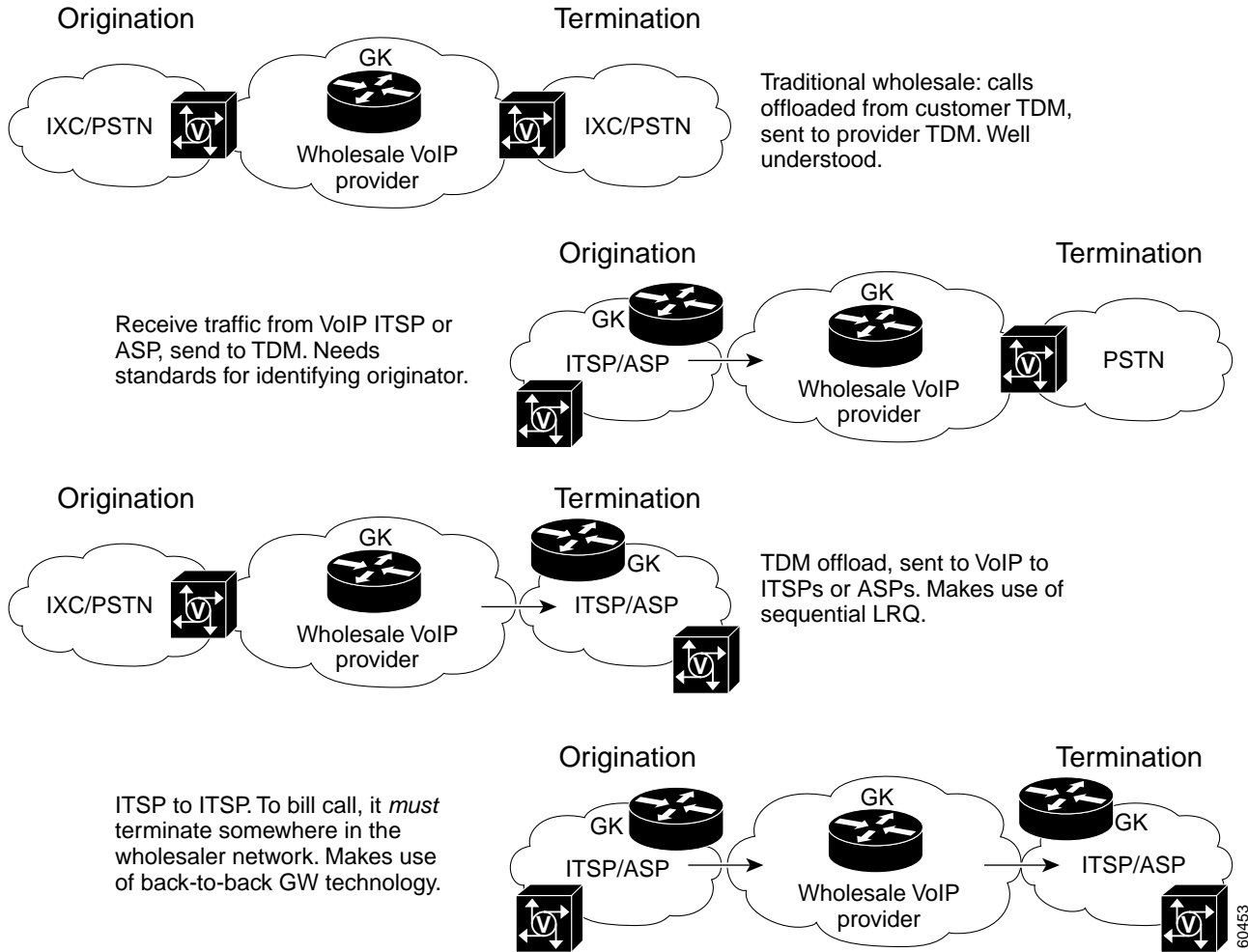
Four call topologies or interconnection methods are described in the following sections:

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

- Originating TDM/Terminating TDM Call Topology
- Originating TDM/Terminating IP Call Topology
- Originating IP/Terminating TDM Call Topology
- Originating IP/Terminating IP (Transit VoIP Network) Call Topology

Figure 5 summarizes each of these topologies.

Figure 5   *Summary of Call Topologies*

Originating TDM/Terminating TDM Call Topology

The originating TDM/terminating TDM call topology is a single administrative domain and the most fundamental call topology. With this topology, you receive and terminate traffic from other service providers via TDM interfaces. Figure 6 illustrates this topology.

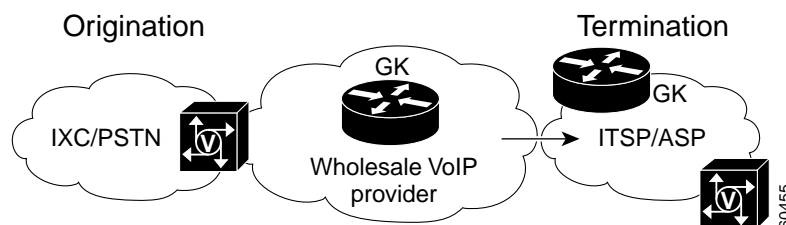*Figure 6   Topology 1: Originating TDM/Terminating TDM*
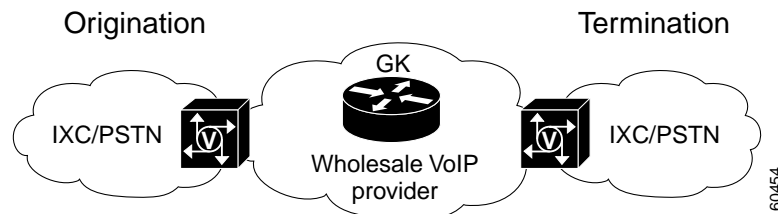
Because interconnection is confined to TDM interfaces on gateways you administer, deployment considerations in the areas of routing, security, billing, and settlement are fairly straightforward. Limited-egress CSR applications demand additional call routing provisioning tasks. Your concerns are primarily confined to supporting the proper TDM signaling and the transparency of bearer traffic, such as voice, fax, or modem pass-through.

The originating TDM/terminating TDM call topology is appropriate for the following applications:

- Card services
- IXC-to-IXC interconnection
- IXC offload
- LEC-to-LEC interconnection (simple toll bypass)
- LEC-to-IXC interconnection

Originating TDM/Terminating IP Call Topology

If you want to increase call volume or coverage area by adding interconnections with other IP-based service providers, use the originating TDM/terminating IP call topology. With this topology, you receive traffic from IXC or PSTN providers over TDM interfaces. If the provider cannot terminate the call within its own network POPs, it can send traffic to other service providers such as ITSPs or ASPs over IP. Figure 7 illustrates this topology.

Figure 7   Topology 2: Originating TDM/Terminating IP

In addition to the TDM-related issues described in the originating TDM/terminating TDM call topology, this topology has the added considerations of IP interconnection. You must consider issues pertaining to call routing, interoperable bearer transport, billing, settlement, and security.

The originating TDM/terminating IP call topology is appropriate for the following applications:

- Card services
- LEC-to-ASP interconnection
- LEC-to-ITSP interconnection (simple toll bypass)
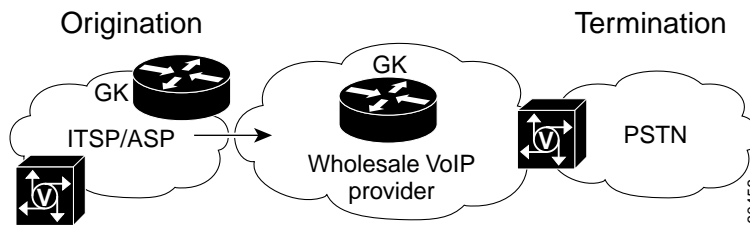- IXC-to-ASP interconnection

- IXC-to-ITSP interconnection

Originating IP/Terminating TDM Call Topology

The originating IP/terminating IP call topology is essentially the same as the originating TDM/terminating IP call topology, but the call direction is reversed. With this topology, you receive traffic from other service providers via IP and terminate traffic at your POPs to IXC or LEC providers through TDM interfaces. Figure 8 illustrates this topology.

Figure 8   Topology 3: Originating IP/Terminating TDM

Because you are now receiving traffic from other providers through IP interconnect, you must be concerned with call routing, originating carrier identification for billing and settlement, interoperable bearer transport, and security.
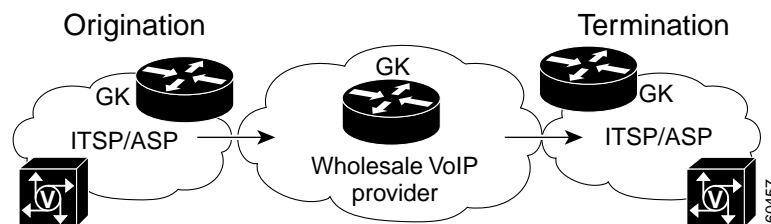
The originating IP/terminating TDM call topology is appropriate for the following applications:

- TSP-to-LEC interconnection (toll bypass)
- ASP-to-LEC interconnection (toll bypass)
- ITSP-to-IXC interconnection
- ASP-to-IXC interconnection

Originating IP/Terminating IP (Transit VoIP Network) Call Topology

If you want to provide transit between different IP-based interconnection partners, use the originating IP/terminating IP call topology. With this topology, you exchange traffic between other service providers using only IP connections. Figure 9 illustrates this topology.

Figure 9   Topology 4: Originating IP/Terminating IP

Typically, you receive traffic from an ITSP or ASP, and if you cannot terminate the call at one of your own POPs, you send the call to another service provider.

When sending and receiving traffic between two IP interconnects, you have increased challenges in the areas of call routing, carrier identification, billing, settlement, security, and potentially masking originating carrier information from the terminating carrier.

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

The originating IP/terminating IP call topology is appropriate for the following applications:

- ASP-to-ITSP interconnection
- ASP-to-ASP interconnection
- ITSP-to-ITSP interconnection
- ITSP-to-ASP interconnection

IP Interconnection Variations

In addition to using the call topologies described, you can interconnect with other IP-based service providers (ITSPs and ASPs) using one of the methods described in the following sections:

- Directory Gatekeeper-Based Interconnection Method
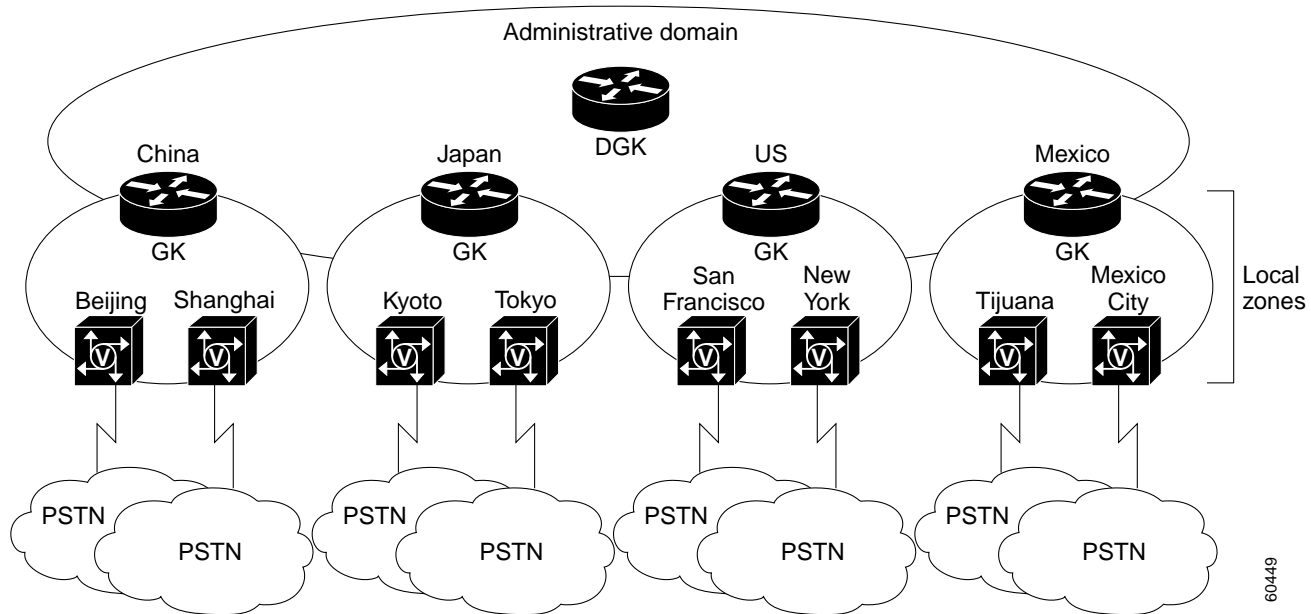- OSP-Based Interconnection Method

Each method has its own provisioning requirements.

Directory Gatekeeper-Based Interconnection Method

With the directory gatekeeper-based interconnection method, you provision call routing between your IP interconnect partners by peering directory gatekeepers to which you send LRQ RAS messages. You can direct certain destination patterns to specific interconnect partners. These destination patterns could have been modified upon ingress into your network to provide limited ingress carrier-sensitive routing applications. Additionally, you can use sequential LRQ features to provide limited egress carrier-sensitive routing applications.

With directory gatekeeper-based interconnect, you benefit from centralizing route provisioning in the directory gatekeeper rather than pushing it to the edge gateways as with OSP. However, billing/settlement functions and security options are processes external to call routing that require some configuration in the gateways, gatekeepers, and related shared-services components.

If you are a large service provider with many POP gateways, provisioning complexities may determine that this is the best option for interconnection. Figure 10 illustrates a directory gatekeeper-based interconnection with other ITSP/ASP partners.

Figure 10   *Directory Gatekeeper-Based Interconnection with Other Service Providers*

OSP-Based Interconnection Method

With the OSP-based interconnection method, an OSP server performs call routing, billing/settlement, and security functions; however, additional provisioning is required. All edge gateways must be registered with the OSP server, and rotary dial-peer failover must be provisioned to route calls through the OSP interconnection.

OSP may be an attractive interconnection option if you want to combine call routing, security, and billing/settlement into one architecture. However, in current Cisco implementations, limitations with OSP deployments require extensive provisioning in the gateways so that they can interact with the required shared services, support the dial plan architecture, and cover termination caveats.

Figure 11 illustrates an OSP-based interconnection with other ITSP/ASP service partners.

Figure 11   OSP-Based Interconnection with Other Service Partners
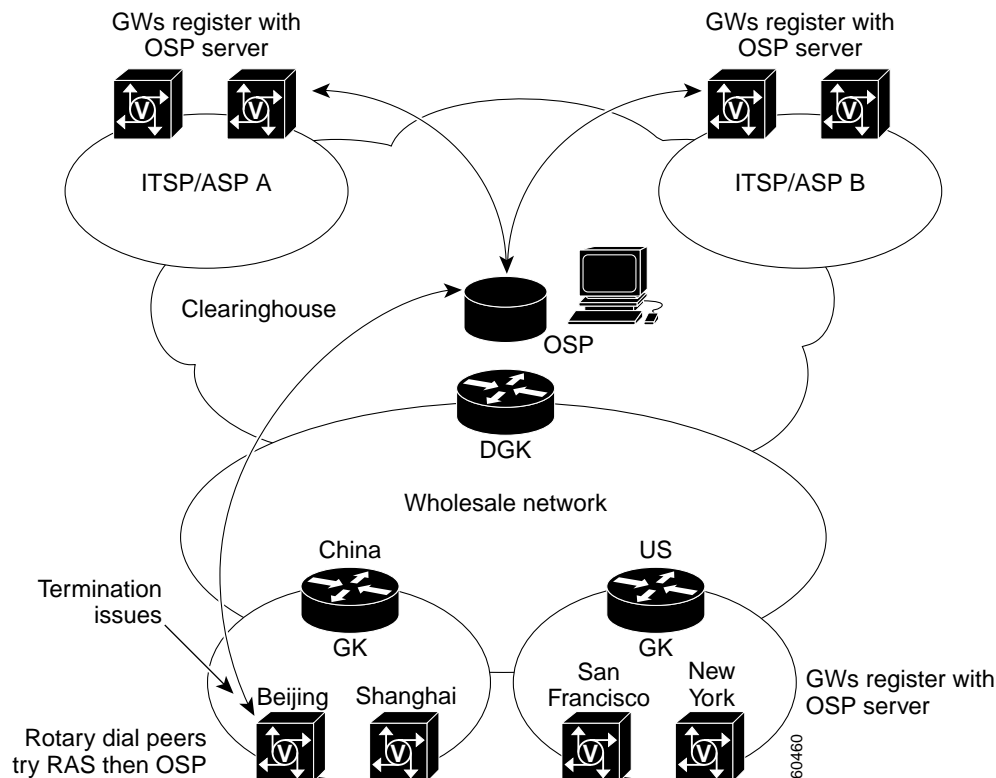
Step 5: Identify Deployment Scenario

Select the appropriate deployment scenario based on functional areas (described below) and call topologies.

The Cisco wholesale voice solution supports the deployment scenarios described in the following sections:

- TDM to TDM
- TDM to IP
- TDM to IP with OSP
- IP to IP with directory gatekeeper
- IP to IP with OSP

Step 6: Identify Functional Areas

The Cisco wholesale voice solution encompasses the primary functional areas described in the following sections:

- Gatekeeper Core
- Shared Services
- Non-SS7-Based POP

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

- SS7-Based POP

- Back-to-Back Gateway System

Figure 12 shows each of the functional areas.

Figure 12   Functional Areas of the Cisco Wholesale Voice Solution
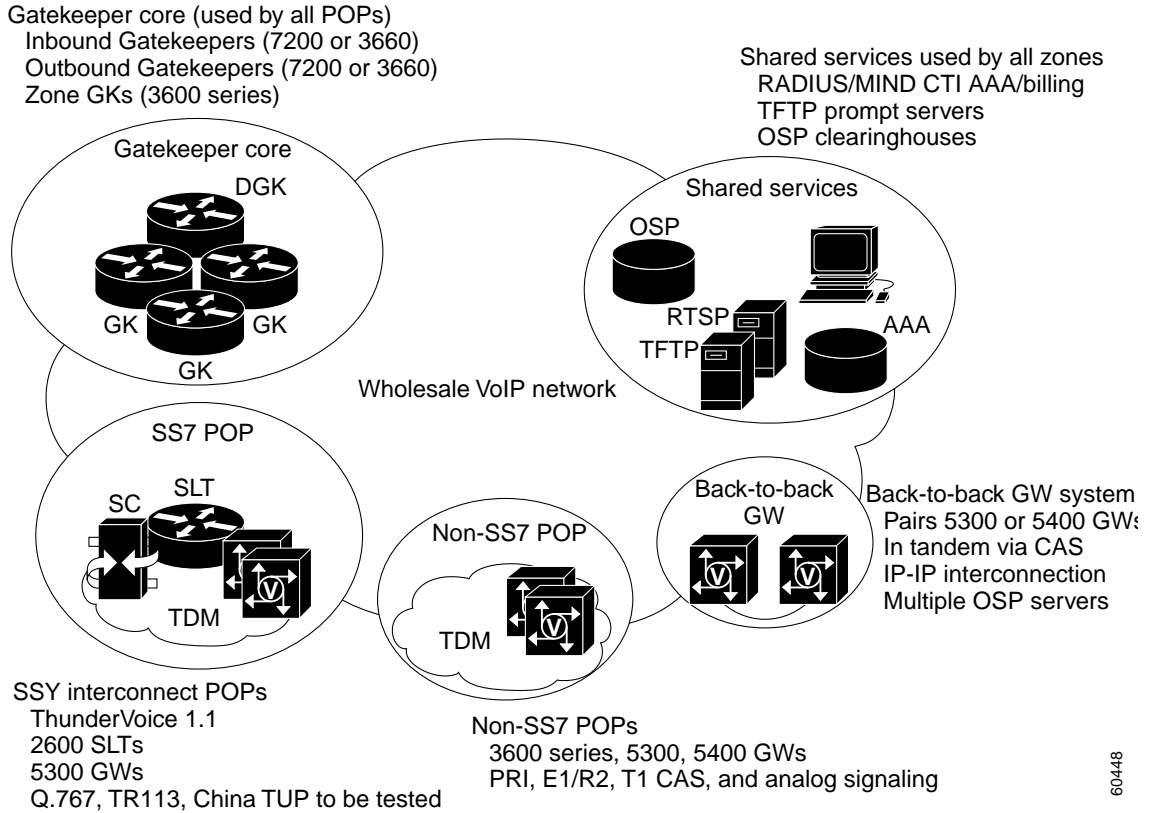
Your wholesale VoIP cloud may include some or all of the areas depicted in Figure 12, depending on the issues specific to your interconnection methods, billing services, call control, settlement, IVR options, and network management.

Gatekeeper Core

The gatekeeper core functional area, illustrated in Figure 13, is used by all POPs and is the foundation of a large-scale H.323 network design. It consists of Cisco gatekeepers (GKs), Cisco directory gatekeepers (DGKs), and optionally Ecosystem Partner gatekeeper platforms.

Gatekeepers enable a network to scale in growth, performance, and dial-plan administration. Gatekeepers and directory gatekeepers provide for resource management, call routing, security, fault tolerance, external Gatekeeper Transaction Message Protocol (GKTMP) applications, and call detail record (CDR) generation. Gatekeepers support interactions with shared services and provide gatekeeper-based interconnect with other providers if the application demands.

Figure 13   Role of Gatekeepers and Directory Gatekeepers in the Gatekeeper Core
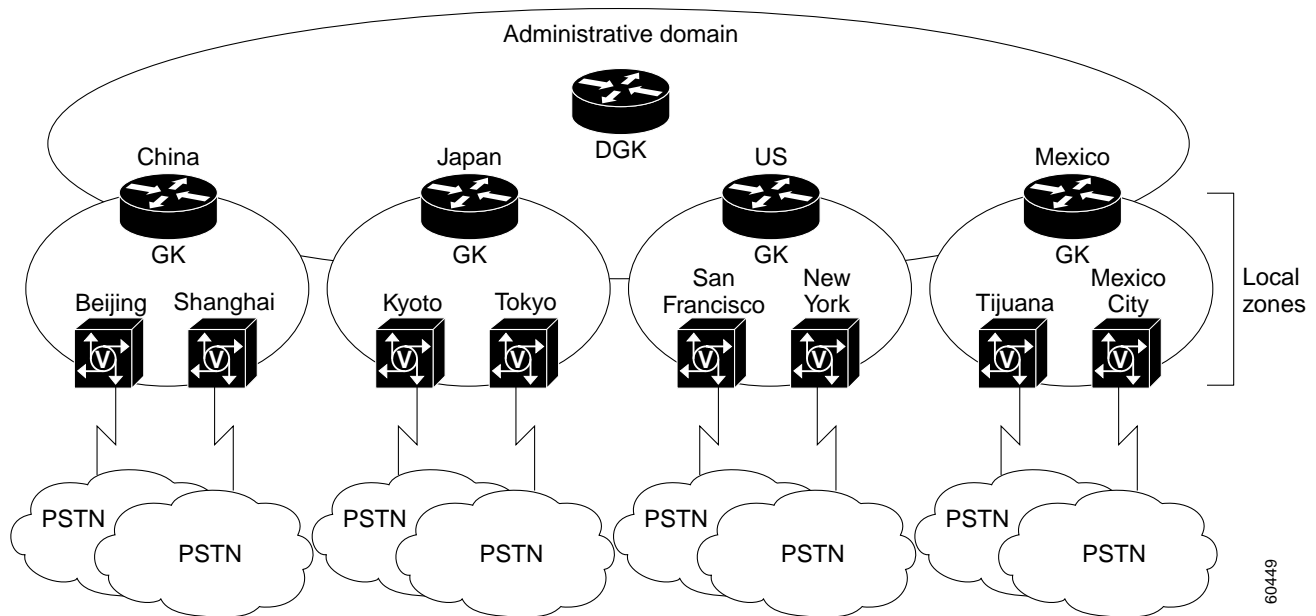
Inbound directory gatekeepers are Cisco 7200 series routers or Cisco 3660 routers. Zone gatekeepers are Cisco 3600 series routers. Cisco 3640s and 3660s routers, and AS5300s and AS5400 universal access servers are examples of gateway platforms.

Shared Services

Shared support services are central resources that enable network applications in the areas of card services, call routing, billing, settlement, security, and network management. The primary elements that enable these applications are OSP servers, TFTP servers, AAA servers, billing systems, NMS platforms, and EMS platforms.

Non-SS7-Based POP

Wholesale service provider networks consist of POPs that house gateways to transport voice traffic between TDM and IP networks. POPs are active components in the originating TDM/terminating TDM, originating TDM/terminating IP, and originating IP/terminating TDM call topologies. Non-SS7-based POPs receive signaling from the TDM network on the same physical interface that supports bearer traffic. There may be a logical separation of signaling and bearer traffic within the interface, such as with ISDN. Actual gateway platforms used at these POPs will depend upon the signaling type offered by the TDM interconnect. Figure 14 shows non-SS7-based POP signaling. The following interfaces are supported through in-band signaling:

- FXO/FXS
- E&M
- BRI/PRI
- DS1/DS3
- E1/R2
- T1 CAS

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

Figure 14   Non-SS7-Based POP Signaling
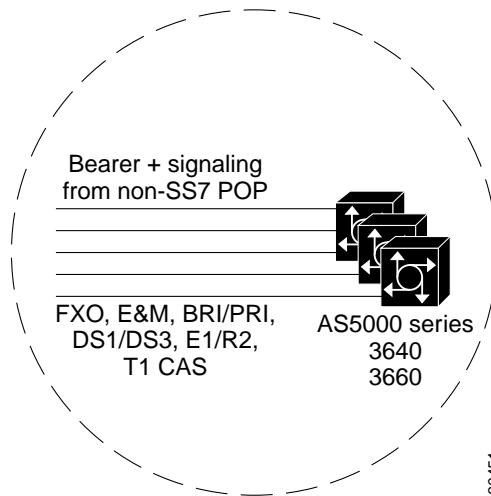
Gateway components include the Cisco 3600 series routers and the Cisco AS5300 universal access servers.

In addition to the physical interface and signaling variations, a number of platform-independent software features and functions must be enabled on the POP gateways to support an application, such as POP size, dial plan, fault-tolerance, security, billing, network management, and bearer transport responsibilities.

SS7-Based POP

These PoPs generally have the same deployment considerations with billing, security, network management, transparent bearer transport, and TFTP servers as the non-SS7 based PoP. However, this PoP has additional considerations related to SS7 interconnect signaling, which is required to conform to the PSTN TDM network. Added considerations also appear in PoP size, dial-plan responsibilities, and fault tolerance.

Gateway components include Cisco 2600 Signaling Link Terminals (SLTs) and Cisco AS5300 universal access servers. Support is provided for Q.767 and TR-113 signaling.

Back-to-Back Gateway System

The back-to-back gateway system is a special component used to provide a variety of functions for a variety of applications. Gateways are deployed as a pair in a back-to-back TDM trunk, single-stage-dialing configuration.

Depending on the application, back-to-back gateways may function as unidirectional or bidirectional call devices. For example, in an IVR application, half of the back-to-back gateway is dedicated receiving calls and the other half is dedicated to originating calls. In contrast, for an OSP interconnect zone application, the back-to-back gateway may process calls in both directions, although each gateway is responsible for separate protocols. For added clarity when discussing back-to-back gateway pairs, we refer to the individual gateways in a pair as an inbound VoIP gateway and an outbound VoIP gateway with respect to the call direction for unidirectional applications. For bidirectional applications, we refer to the gateway by the protocol it supports, where possible.

Figure 15 shows the relationship of the back-to-back gateway to an ingress and egress carrier and to your wholesale VoIP cloud.

Figure 15   Relationship of Back-to-Back Gateways to Wholesalers and Carriers
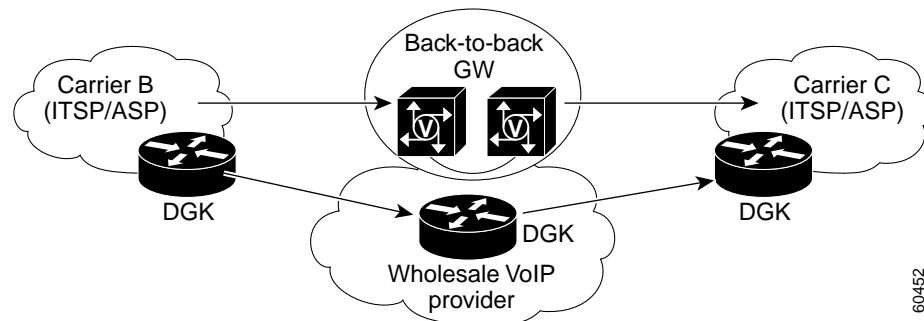
In many ways, the back-to-back gateway system functions just like a normal non-SS7-based POP. The gateway pair helps with applications that use different bearer transport options (such as codec type or security options) on the two interconnecting networks for which you are providing transit. It allows you to have a presence in the call-signaling path in IP-to-IP interconnect call topologies so that you can generate usage records through AAA, interconnect with Clarent-based and OSP-based environments, and front-end PC-to-phone applications for IP-based interconnect partners. It also provides a way to obscure information about interconnection partners.

The platforms that might be used as back-to-back gateways are the Cisco 3600 series routers, the Cisco AS5300 universal access server, and the Cisco AS5400 universal gateway.

Step 7: Identify Required Hardware and Software Components

This section describes the actual hardware and software components, both Cisco and third-party vendor equipment, that can be used to implement a wholesale voice solution.

Major Components

The major components described in the following sections are used in implementing a wholesale voice solution:

- Cisco Voice Gateways
- Cisco H.323 Gatekeepers and Directory Gatekeepers
- Cisco Signaling Controllers
- Cisco SLT Systems

Cisco Voice Gateways

Wholesale solutions require a range of small to large-scale PSTN interconnections with the wholesaler TDM-based customers (typically IXCs, PTTs, or other wholesalers), depending on anticipated call volumes. Similar interconnections may be required to offload traffic. Gateways may handle their own signaling, or they may provide intermachine trunks (IMTs) and receive external SS7 signaling through a Cisco SC2200 signal controller running Cisco SS7 Interconnect for Voice Gateways Solution software with Q.931 signaling backhaul.

Possible gateway platforms include the Cisco 3640 and 3660 routers, and AS5300 and AS5400 universal access servers, along with various supporting network modules.

Note The Cisco wholesale voice solution does not support gateway platforms that use MGCP call signaling. Cisco AS5800 gateways cannot be used in SS7 POPs that are using the Cisco SS7 Interconnect for Voice Gateways Solution software.

Cisco H.323 Gatekeepers and Directory Gatekeepers

Gatekeepers and directory gatekeepers are mandatory network elements used to scale a wholesale network to large sizes. They consist of specialized Cisco IOS software images running on a dedicated Cisco 3660 or 7200 series router.

Directory gatekeepers further supplement network scalability and are mandatory if gatekeeper-based carrier interconnection is desired. Cisco gatekeepers perform the following tasks:

- Resource management. Cisco gatekeepers determine the health of H.323 gateways by monitoring registration and unregistration (RRQ/URQ) messages and resource availability indicators (RAIs).

- Call routing. Cisco gatekeepers provide call routing based on destination E.164 addresses. They may use their knowledge of local gateway health levels to make routing decisions in order to increase network availability of the gateways. Cisco gatekeepers may also route calls between remote gatekeepers within the same administrative domain, using intergatekeeper LRQ RAS messages. Similarly, Cisco directory gatekeepers may also route calls to other carrier administrative domains using LRQ RAS messages.

- Security. In conjunction with an external server (such as RADIUS), Cisco gatekeepers may be used for secure call admission in intradomain call scenarios (calls within the same service provider). Cisco gatekeepers also have limited applications in implementing interdomain security functions for calls sent between carriers through IP interconnection.

- GKTMP applications. Cisco gatekeepers may act as a control point from which an application server can affect call routing, number translation, call admission/blocking, and so on. These application servers interface with a Cisco gatekeeper or directory gatekeeper using GKTMP.

- CDR Generation. Cisco gatekeepers have limited abilities to generate CDR records for calls. This is an option if you do not own the gateways at a POP, or if you simply wish to reduce the amount of messaging overhead associated with AAA in your smaller POPs. Billing in this manner has limitations.

Cisco Signaling Controllers

Cisco Signaling Controllers are optional components, but are required in SS7 interconnection solutions. The supported platform is the Cisco SC2200.

Cisco SLT Systems

Cisco SLY systems are optional Cisco 2600 series routers capable of terminating Message Transfer Part (MTP) Levels 1 and 2 SS7 layers and backhauling Level 3 and higher SS7 layers to the Cisco SC2200 in an SS7 interconnection solution.

Additional Components for Shared Services

Additional components, provided by third parties, support shared services and are described in the following sections:

- RADIUS/OSS Servers

- Ecosystem Partner H.323 Gatekeepers

- Gatekeeper Application Servers

- OSP Servers

- Prompt Servers

- TFTP Servers

- Network Management Systems (NMS)

- Element Management Systems (EMS)

RADIUS/OSS Servers

Ecosystem partner OSS servers interface with Cisco gateway and gatekeeper components through AAA RADIUS vendor specific attributes (VSAs) and are mandatory elements of the wholesale network. Current examples include Cisco Secure and Cisco ecosystem partners such as MIND/CTI and Belle Systems billing platforms. Cisco has defined a set of VSAs in the document, *RADIUS Vendor-Specific Attributes Voice Implementation Guide*. VSAs can be used to achieve the following functions:

- CDR collection and billing system front-ending. Cisco gateways send call start/stop records to a RADIUS server using AAA. The billing application can extract these records to generate CDRs. CDRs may be shared between carriers as a method of settlement through billing system mediation applications.

- User authentication and authorization. For card services, a AAA RADIUS server may validate end users based upon ANI or username and password combinations. AAA interaction occurs directly on the Cisco gateway.

- Application hosting. A Cisco gateway may run a call script that interacts with an application mounted on the RADIUS server. The server is capable of manipulating call information through VSAs in AAA. An example would be a debit card application. The AAA server interacts with a debit card billing application to determine account balances, call rates, and time remaining for an individual user. This information is sent to the gateway script in AAA VSAs.

> **Note** Cisco Secure does not support applications that are dependent upon VSAs, such as debit card.

- Security. Gatekeepers can administer security options to perform secure endpoint registrations and also to verify that incoming calls are from authorized users or endpoints. Access control lists are the recommended solution for security. H.235-based intradomain security (access tokens) are not supported.

- Settlement. Some billing system vendors support interdomain settlement based on CDRs collected from each local domain. This offers a viable alternative to OSP in some cases. Mediation vendors such as XACCT also provide servers dedicated to settling CDRs between the billing systems of different vendors. These are known as mediation servers and are optional components in a wholesale network.

Ecosystem Partner H.323 Gatekeepers

These optional gatekeepers may be used on the network fringe to compliment the Cisco gatekeeper/directory gatekeeper infrastructure and to host a variety of applications. Individual applications will vary between ecosystem partners.

> **Note** The Cisco wholesale voice solution does not require or specify the use of these gatekeepers, but the architecture does not exclude them from being inserted into your network.

Gatekeeper Application Servers

Enhanced call routing applications may optionally reside on an external server and interface with a Cisco wholesale VoIP network through the Cisco gatekeepers or directory gatekeepers using the GKTMP interface specification.

> **Note** The Cisco wholesale voice solution does not require or specify the use of specific GKTMP applications, but the architecture does not prohibit you from adding them to your network.

OSP Servers

To support carrier interconnect, you may choose to use OSP servers. Using OSP for secure settlement transactions requires a clearinghouse entity, or at least a dominant carrier in the interconnect relationship that administers the OSP server. GRIC and TransNexus currently provide OSP-based clearinghouse services. OSP servers perform the following functions:

- Authentication of gateways or carriers. An OSP server can verify whether or not an originating or terminating carrier's gateway is a valid participant in the OSP interconnect by using a secure exchange of certificates.

- Call authorization. An OSP server generates an access token for each call sent from an originating gateway into the OSP-based interconnect. The originating gateway includes this token in the SETUP message to the terminating gateway. Upon receiving SETUP, the terminating gateway may either send the token back to the OSP server for validation or perform the validation locally.

- Call routing. The OSP server provides the originating gateway with a terminating gateway selected from among registered OSP endpoints.

- CDR collection. OSP usage indications are sent to the OSP server from both the originating and terminating endpoints after a call has ended. The OSP server uses this information to generate a CDR.

- CDR correlation and settlement. Once CDRs are collected, the OSP server may interface with a billing application to generate settlement billing between the two interconnecting carriers.

Prompt Servers

A prompt server is an optional component that maintains a prompt database for gateways running interactive voice response (IVR) functionality for applications such as card services. Prompt databases may be stored locally on the gateway in flash memory if they are not too large. Larger prompt databases, such as those needed when there are many branded retailers or when many languages must be supported, may be dynamically downloaded as needed from a prompt server using TFTP. TFTP servers are generic third-party devices that can be hosted on a wide variety of platforms.

TFTP Servers

TFTP servers are used to store audio (IVR) files, IOS files, configuration files, dial plans, and other files that do not need to reside on a local machine. These files can be downloaded as needed.

Network Management Systems (NMS)

NMS systems are optional components that are used for network monitoring, fault management, trap correlation, and reporting. Any NMS system can extract this information from wholesale components using SNMP. The Cisco wholesale voice solution recognizes CiscoWorks Internet Protocol Manager (IPM) to monitor network QoS and Cisco InfoCenter (CIC) for fault management and trap correlation. For reporting, it is possible for third-party vendors such as Trinagy to provide reports by interfacing with Cisco Voice Manager (CVM).

Element Management Systems (EMS)

EMS systems are optional components that are used for managing or provisioning other components in the solution. Cisco Voice Manager (CVM) provides limited provisioning support and is the only EMS currently supported in the Cisco wholesale voice solution.

Detailed Component Inventory

The following component hardware and software products and subordinate solutions are relevant to the Cisco wholesale voice solution:

- VoIP Gateways
- H.323 Gatekeepers
- SS7 Elements
- Shared Services Components

VoIP Gateways

The following Cisco devices can be used as VoIP gateways:

- Cisco 3620 router
- Cisco 3640 router
- Cisco 3660 router
- Cisco AS5300 access server
- Cisco AS5350 access server
- Cisco AS5400 access server
- Cisco 7200 series routers

These platforms support a variety of analog and digital interfaces. For more information about supported interfaces for a specific platform, please refer to the documentation for that specific platform at http://www.cisco.com.

H.323 Gatekeepers

Candidate gatekeepers are as follows:

- Cisco 3660
- Cisco 7200 series

SS7 Elements

Candidate SS7 elements are as follows:

- Cisco SC2200
- Cisco 2600 SLT

Shared Services Components

Candidate shared-services components are as follows:

- Cisco Voice Manager (CVM)
- Trinagy Trend Reporting Application
- Cisco Info Center (CIC)
- Internet Performance Module (IPM)
- AAA RADIUS Security Server (various vendors)
- MIND/CTI Billing System
- OSP server (various venders)
- Generic TFTP server

Step 8: Identify Design and Scalability Issues

Some of the design issues associated with the Cisco wholesale voice solution have already been mentioned in previous steps. The following paragraphs look at these issues in detail and organize them into the following groups:

- General Design Issues
- Functional Areas Design Issues
- Service Design Issues

General Design Issues

Because of the many ways in which multifunctional groups interact, there are general design issues associated with the following topics:

- Call Routing
- Billing and Settlement
- Basic Dial Plans
- Fault Tolerance in Dial Plans
- Security Considerations Associated with Dial Plans

Call Routing

Call routing between IP service providers can be either directory gatekeeper-based or OSP-based. The billing and call routing functions you desire will determine whether your network will be directory gatekeeper-based or OSP-based.

Directory gatekeeper (DGK)-based call routing uses Location Request (LRQ) RAS messages to resolve call routing for desired prefixes. An LRQ is sent from the originating service provider's directory gatekeeper to the terminating service provider's directory gatekeeper to request the terminating gateway IP address. The directory gatekeeper method of call routing can be used when the originating and terminating service providers are trusted peers.

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

OSP-based call routing uses a separate OSP clearinghouse entity that maintains OSP servers. The OSP servers contain the prefix call-routing tables of all service providers that subscribe to the OSP clearinghouse. The originating gateway sends an OSP Authorization Request to the OSP server; the OSP server responds with an Authorization Response containing a list of possible IP addresses of the terminating gateway plus a security token. This token is included in the setup message to provide security validation at the terminating gateway. The OSP method of call routing is used when carriers want a third party to provide the billing and settlement.

Billing and Settlement

You must accurately identify the originating carrier and terminating carrier for calls in order to properly bill for service. The degree of difficulty of this varies depending upon the call topology used. Furthermore, the usage indication must be extracted from a reliable source. This implies that the devices supplying call usage indications are somewhere within the H.225 call-signaling path. Therefore, if billing is desired, you must own at least one gateway in any given conversation.

Billing and settlement functionality can be AAA/RADIUS-based or OSP-based. These methods may be used either individually or in conjunction with each other and will directly depend upon the method of interconnect. Though differing in protocol, each method addresses the same basic needs for call accounting.

AAA billing must be used for any intradomain calls because OSP is designed to bill for interdomain calls only. AAA may also be used for interdomain calls if interconnect is handled by a peering directory gatekeeper relationship rather than by an OSP server. In this scenario, the billing application correlates the usage records to generate CDRs. The CDRs are then distributed to customers in the form of a periodic bill. Customers can verify this bill against their own records before exchanging money or settling the call. Various mediation vendors exist that help automate the verification and settlement stages.

For interconnect using OSP, you may either own an OSP server or depend upon a third-party clearinghouse OSP server to provide accounting services. The OSP server receives accounting information from your gateway in much the same manner as with AAA. Because usage indications are received from both gateways across administrative domains, the OSP server gets accurate terminating and originating carrier information. The usage records are then correlated to generate CDRs. The CDRs may be distributed as periodic bills to customers. Customers can verify this bill against their own records before exchanging money or settling the call. To provide personal accounting records for verification, parallel AAA accounting records can be used.

It is possible that a third party could manage an interconnecting TDM PoP. If this is the case, you cannot depend upon gateways to send them CDR information. You may therefore choose to do billing from the terminating gateways only (if you own them) or from the gatekeeper.

Billing from the gatekeeper has limitations. Cisco gatekeepers can send call start/stop records to a AAA RADIUS server based upon receipt of ARQ and DRQ RAS messages from gateways. However, RAS messages are sent over UDP and are not guaranteed to arrive at the gatekeeper. Furthermore, this method of billing lacks answer supervision. Also, firewalls between gatekeepers and AAA servers can cause problems because certain ports need to be opened up for these messages to be received.

Therefore, billing is most reliable and accurate if performed at the gateway.

Basic Dial Plans

Dial plan responsibilities are distributed among gateways, gatekeepers, and directory gatekeepers. Because SS7 deployments leverage NI-2 type Q.931 backhaul signaling, the basic H.323 dial plan architecture is the same regardless of whether the PoPs in the network are SS7 based, non-SS7 based, or a mixture of both. Figure 16 depicts a typical large-scale H.323 network design.

Figure 16   Typical Large-Scale H.323 Network Design
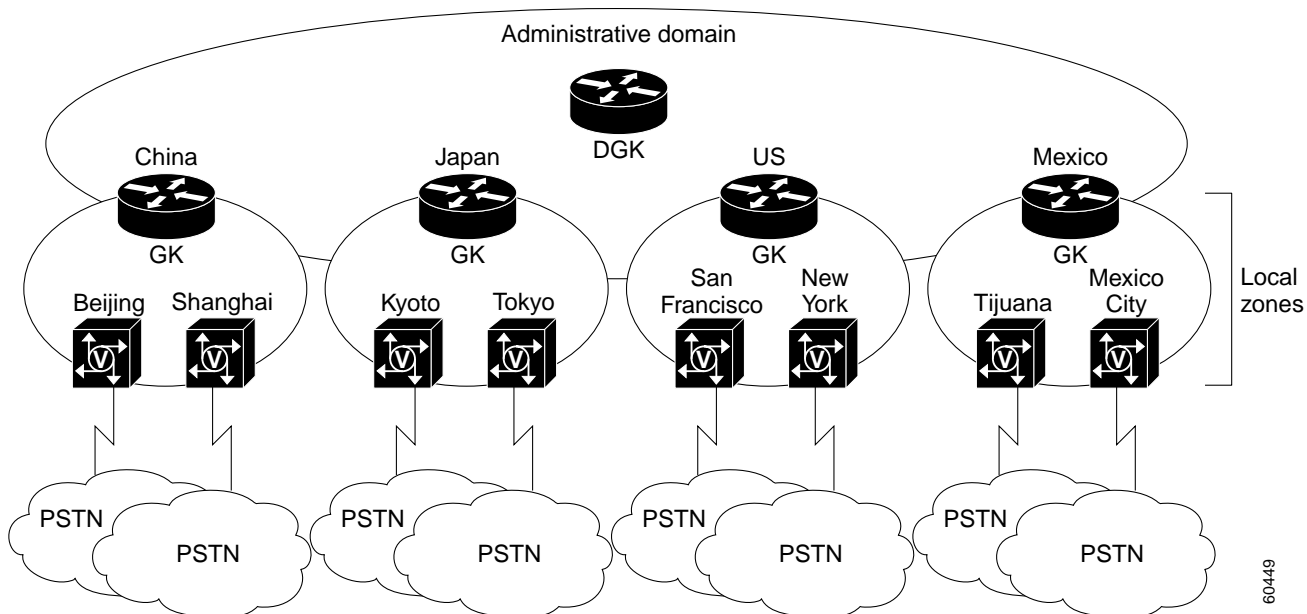
Gateways deal with the local PoP portion of the dial plan. This encompasses any digit manipulation needed to normalize numbers or implement local PSTN access rules. It also includes notifying a gatekeeper when resource availability indicator (RAI) thresholds are crossed in order to increase call-completion rates. Furthermore, the gateway may implement rotary dial peers to handle call failover routing options such as trying OSP if normal gatekeeper RAS call routing offers no possible termination.

For example, you may want the gateway to notify the gatekeeper when its resource limits are nearly exhausted, thereby prompting the gatekeeper to select a different gateway. Additionally, to simplify route provisioning in the gatekeepers and directory gatekeepers, you may wish to normalize numbers into a standard format (for example, country code + area code + local number) before sending calls into the VoIP network. Or, you may need to prepend or strip digits such as area codes or access codes, as PSTN access rules require, before sending calls out the TDM interfaces.

Local gatekeepers monitor gateway health levels and maintain detailed routing tables, mapping destination patterns to specific terminating gateways within one or more local zones. The local gatekeepers can use features such as lightweight registration, RAI, and static gateway-priority assignments to influence gateway selection. For all other nonlocally supported destination patterns, the local gatekeeper configures a wild-card route to the directory gatekeeper.

The directory gatekeeper maintains an overall routing table of destination patterns and the corresponding local gatekeepers that support them. The directory gatekeeper simply forwards LRQ requests to the local gatekeeper that handles that destination pattern.
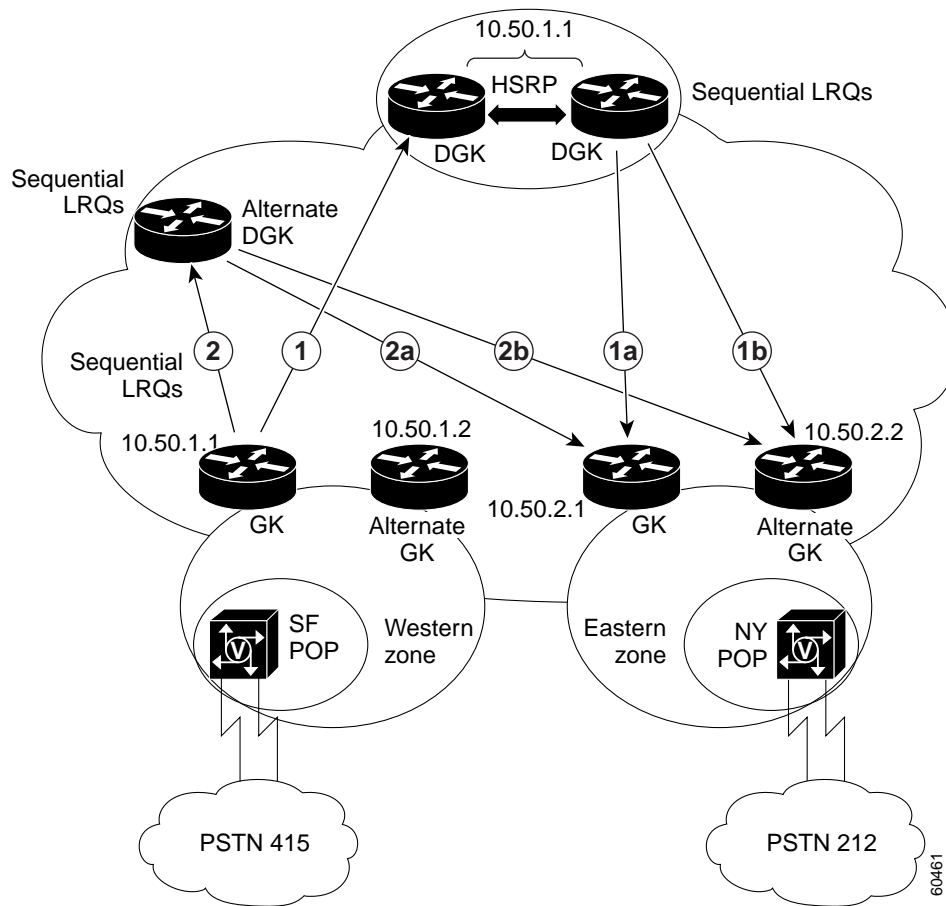
(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

This use of gatekeepers and directory gatekeepers allows the addition of new gatekeeper zones, PoPs, and certain types of IP interconnect partners with minimal impact to dial plan provisioning. Changes are isolated to the local gatekeeper and the directory gatekeeper. The rest of the elements in the network are untouched. Often, the level of dial plan resolution at the directory gatekeeper level can be simplified. For example, a directory gatekeeper may know to route all calls beginning with a country code of 1 to the local U.S. gatekeeper. The local U.S. gatekeeper can then expand selection to more digits to route the call to the proper terminating gateway.

Fault Tolerance in Dial Plans

For intradomain calls and directory gatekeeper-based IP interconnects, you have the option of overlaying fault tolerance onto the basic H.323 VoIP network dial plan design. This is accomplished by using a combination of Cisco IOS software features such as alternate gatekeepers on the gateway, Hot Standby Router Protocol (HSRP) on the directory gatekeeper, and sequential LRQs on the gatekeepers and directory gatekeepers. Figure 17 illustrates a fault-tolerant architecture using alternate gatekeepers.

Figure 17   Fault-Tolerant Architecture Using Alternate Gatekeepers

Gateways may be configured to register to a primary gatekeeper and an alternate gatekeeper in the event that the primary gatekeeper fails. This implies that at any given time, gateways may be registered to either a primary or alternate gatekeeper. Since Cisco gatekeepers do not communicate registration states to each other, sequential LRQs must be configured on the gatekeepers and directory gatekeepers to accommodate zone fragmentation.

For example, a gatekeeper in the Western Zone supports gateways in San Jose (408) and San Francisco (415). Under normal circumstances, when San Jose calls San Francisco, the route is resolved in the local primary gatekeeper. However, let's say that San Jose fails over to the alternate gatekeeper while San Francisco remains on the primary gatekeeper. To continue to support regional call completion within the Western Zone, the primary and alternate gatekeepers must be provisioned to send local prefixes to each other if no local resource exists—that is, if the terminating gateway has failed over to the other gatekeeper. In this case, in order for San Francisco to complete calls to San Jose, the primary gatekeeper must know to send LRQs for the San Jose prefix to the alternate gatekeeper. Similar provisioning is required on both primary and alternate gatekeepers to support calls in both directions.

Provisioning is also required on the directory gatekeeper to prevent zone fragmentation when calls are originated from other zones. For example, if San Francisco sends a call to New York, the directory gatekeeper does not know with which gatekeeper (primary or alternate) the NY gateway is registered. The directory gatekeeper must be provisioned to send sequential LRQ requests to both primary and alternate terminating local gatekeepers for all Eastern Zone supported prefixes (messages 1a and 1b in Figure 17). Similar provisioning is required for the Western Zone prefixes to support calls in the other direction.

HSRP is used to provide fault tolerance for the directory gatekeeper. However, HSRP failover detection may take some time during which no calls will be processed. To cover this possibility, local gatekeepers may be configured to point to more than one directory gatekeeper (that is, an ALTdirectory gatekeeper) for their wild-card routes using sequential LRQs.

For example, the gatekeeper may point to an HSRP directory gatekeeper pair as its primary option (message 1). If no response is received because HSRP failover has not yet been detected, the gatekeeper may initiate another LRQ (message 2) to an ALTdirectory gatekeeper after a configurable timeout of from 100 to 1000 ms. During this time calls will still be completed, but with additional post-dial delay. The ALTdirectory gatekeeper is configured exactly the same as the primary directory gatekeeper HSRP pair (messages 2a and 2b).

Security Considerations Associated with Dial Plans

You can implement various security mechanisms throughout your H.323 VoIP network. The security mechanism you select may have different provisioning needs within multiple functional areas. For intradomain calls, you may use complex access-lists. For interdomain calls, you may use either complex access-lists or, where OSP is used, OSP access tokens.

Note   The Cisco wholesale voice solution does not support Cisco H.235 access tokens.

You may provision your gateways with complex access-lists to accept calls from only known entities; however, this is neither scalable nor friendly to network changes or elements that use DHCP.

Functional Areas Design Issues

There are design issues to consider for each of the following functional areas:

- Gatekeeper Core
- Shared Services

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

- SS7 PoP
- Non-SS7 PoP
- Back-to-Back Gateways

Gatekeeper Core

Consider the following issues when designing the gatekeeper core:

- **Network Size Scaling**—Large H.323 VoIP networks are segmented into different regional zones, each managed by a gatekeeper. Segmentation is based upon several factors, such as desired call throughput, dial plan, and the number of active endpoints. As network coverage and capacity grow, you can expand by adding new gateways or PoPs to gatekeepers until performance limitations for the gatekeeper platform are met. At that point, you can expand by simply adding new gatekeepers. Traffic is routed between gatekeeper zones using LRQ/LCF RAS messages.

- **Dial Plan Scaling**—As more gatekeepers are added to the network, inter-gatekeeper routing configurations increase dramatically. The smallest change to the dial plan requires configuration changes to all gatekeepers in the network. When the number of zones is relatively small, these changes can be managed by having a single dial plan that is downloaded through TFTP to all the gatekeepers within your administrative domain. As scale increases, the number of zones and the rate of dial plan updating increases. At this point, rather than burdening every gatekeeper with routing information for the entire network, a directory gatekeeper should be used to isolate and alleviate dial plan provisioning. For information on dial plan provisioning, refer to the document *Designing Static Dial Plans for Large VoIP Networks*.

- **Fault Tolerance**—Cisco gatekeepers and directory gatekeepers can be designed to enable redundancy in the dial plan. At the edge, gateways at each PoP are configured to support registration with an alternate gatekeeper in case the primary gatekeeper fails. In the core, gatekeepers are configured to support sequential LRQ messages to provide redundant paths to alternate directory gatekeepers and also to accommodate local zone fragmentation conditions. At the directory gatekeeper level, both sequential LRQs to accommodate zone fragmentation and HSRP are configured to provide redundancy at the highest layer.

- **Directory Gatekeeper-Based IP Interconnect**—If you choose to interconnect routes with other service providers using a directory gatekeeper, configure the directory gatekeepers to exchange LRQ RAS messages between their administrative domains to resolve call routing for desired prefixes. Sequential LRQs may be implemented on the directory gatekeeper to support limited egress CSR applications. Back-to-back gateways may be used to support IP-to-IP call topologies.

- **Security**—To validate whether or not a call was originated from a valid endpoint, Cisco gateways and gatekeepers can implement access lists to provide secure gateway registration and admission. To support this, gatekeepers must be configured to interact with a AAA server.

- **Network Management**—Gatekeepers must be enabled to support SNMP community strings so that external management platforms such as CVM and CIC can provision, access reporting information, and receive traps using SNMP.

- **TFTP Server Access**—If you desire, the gatekeeper can be configured to support the remote downloading of software images and configurations through a TFTP server.

Shared Services

Consider the following issues when designing shared services:

- **Call Routing**—For OSP-based interconnect scenarios, an OSP server handles call routing functions along with some complimentary provisioning on the OSP gateway dial peers. The impact on the dial plan is discussed in more detail in the document *Designing Static Dial Plans for Large VoIP Networks*. Additionally, it is possible for an external server to provide enhanced call routing functions by interfacing with Cisco gatekeepers and directory gatekeepers via GKTMP.

- **Billing**—A AAA server collects usage records directly from the gateways. Alternatively, an OSP server may also collect usage records for interdomain calls. Details on billing implementations vary, depending on the application enabled.

- **Security**—You can provision complex access lists on the gateways to implement security functions. In the case where IOS configurations exceed the NVRAM capacity of the router, a TFTP server may be employed to centrally store, administer, and upload gateway configurations. Cisco H.235 access tokens are not currently supported. An OSP server supplies security functions for OSP interconnect methods.

- **Network Management**—Standard SNMP NMS platforms can be deployed to provide generic SNMP management functions. CVM provides SNMP-based statistics collection along with a very limited dial plan and component-provisioning tool. Reports can be generated by using ecosystem partner reporting engines that integrate with CVM. Cisco recognizes Trinagy as one of these vendors. CIC can be used if fault management is desired. Additionally, Cisco IPM can be used to provide monitoring of network QoS.

- **Remote Download**—A TFTP server can be used to remotely store IVR prompts, TCL scripts, software images, and configurations for download.

SS7 PoP

Consider the following issues when designing SS7 PoPs:

- **Signaling**—SS7 PoPs are large in size and consist of DS1 and DS3 IMTs to the gateways. PSTN-side call control is provided using Q.931 backhaul from the Cisco SC2200 to Cisco AS5300 and AS5400 gateways. PoPs may optionally support Cisco 2600 SLT gateways to terminate SS7 layers MTP1 and MTP2 on behalf of the SC2200 signaling controller.

Figure 18 shows the signaling used in an SS7 PoP, and the relationship among Cisco SC2200 nodes and hosts, Cisco AS5x00 gateways, and Cisco SC26xx SLTs.

(DRAFT LABEL) ALPHA DRAFT - CISCO CONFIDENTIAL

Figure 18   SS7 PoP Signaling
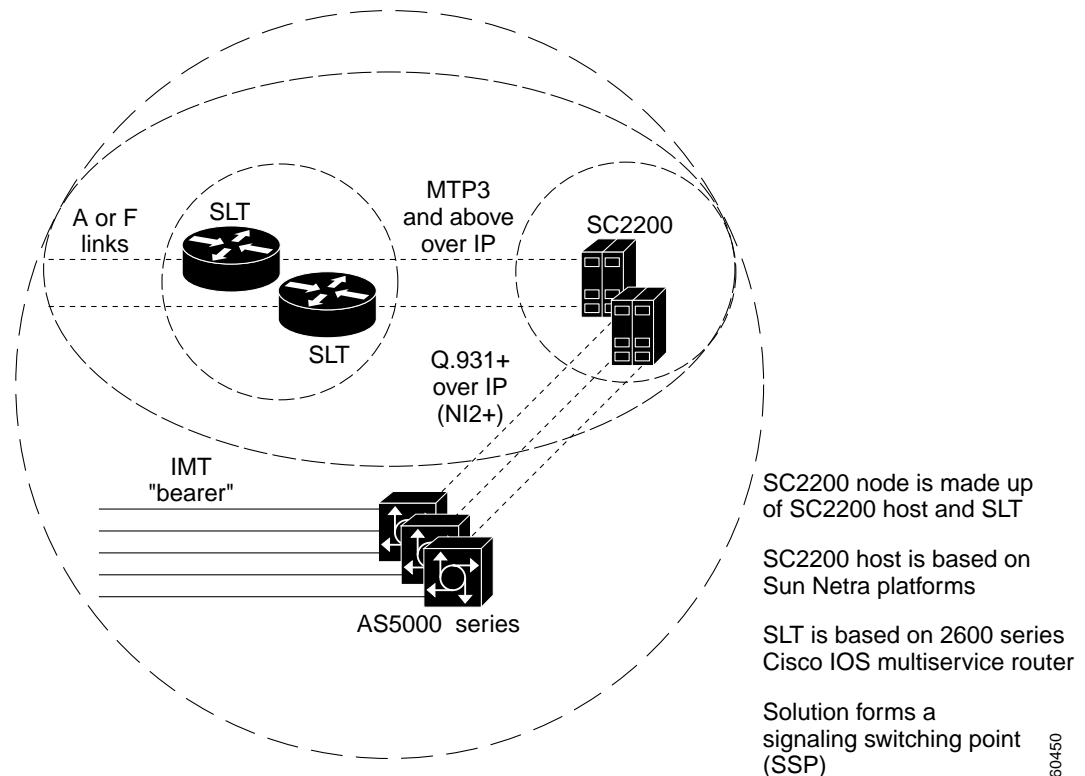
- **Dial Plan**—For SS7-based PoPs, you can perform number modification in either the gateway, the Cisco SC2200 signaling controller, or both. The Cisco SC2200 allows digits in the called-party number or calling-party number fields to be added, deleted, or modified. It is also possible to modify the nature of address (NOA), perform black-listing and white-listing, and AIN triggering. The gateway must be provisioned with an RLM group to interface with the Cisco SC2200 in addition to normal H.323 configurations. Once the Cisco SC2200 and gateway are provisioned to interface with each other, the rest of the H.323 dial plan remains the same as discussed in the document *Designing Statuc Dial Plans for Large VoIP Networks*.

- **Fault Tolerance**—Gateways can support a backup Cisco SC2200 in the event the primary SC2200 fails. It may take up to three seconds for the gateway to detect and fail over to the new SC2200. During this time, any new calls will not be processed. Furthermore, any calls that were in the process of being set up will be lost. Active calls at the point of failover, however, remain in progress.

Non-SS7 PoP

Consider the following issues when designing Non-SS7 PoPs:

- **Signaling Types**—Signaling types can vary greatly and include analog FXO, analog E&M, BRI, DS1 interfaces (E1/R2 variations, T1 CAS variations, PRI), and perhaps DS3 interfaces on the upper boundary.

  Low-density analog interfaces generally discourage carrier interconnects, so calls that ingress the PoP will almost always be for card services and calls that egress the PoP are reoriginated into the PSTN, usually to bypass PTT interconnect tariffs. DS1 and DS3 interfaces generally provide either card services or interconnect wholesale systems to their customers.

- **Size**—Additional considerations surface at small-scale PoPs. The hardware footprint of the equipment must be minimized as well as the amount of non-bearer traffic, because the IP network bandwidth coming into the PoP is likely to be sub-E1 bandwidth.

- **Dial Plan**—Dial plan responsibilities are distributed among gateways, gatekeepers, and directory gatekeepers. The gateways have to deal with the local PoP portion of the dial plan. This includes provisioning needed dial peers, translation rules, and RAI thresholds. Dial plans encompass more than one functional area and are discussed in greater detail in the document *Designing Static Dial Plans for Large VoIP Networks*.

- **Billing**—For performance and accuracy reasons, it is recommended that billing be done from the gateway whenever possible. You must configure the Cisco gateways to interact with shared AAA services to support billing and optional debit card applications.

- **Fault Tolerance**—If you desire, you can configure a gateway to support an alternate gatekeeper with which it will register should the primary gatekeeper fail. This requires a related configuration in the gatekeeper functional area.

- **Security**—To support security, gateways can be configured with complex access lists. For OSP-based interconnect scenarios, the gateways must be provisioned to interact with the OSP server to support OSP security options.

- **Network Management**—Gateways must be enabled to support SNMP community strings so that external management platforms such as CVM and CIC can provision, access reporting information, and receive traps using SNMP.

- **Transparent Bearer Transport**—Unless you have previously agreed to limit the types of calls exchanged between other carriers, you may receive traffic of any bearer type. Your gateways must be able to transparently pass voice, real-time fax, and modem traffic across the VoIP network.

- **TFTP Server**—If you desire, you can configure a gateway to support remote downloading of prompts, software images, and configurations through a TFTP server.

Back-to-Back Gateways

Consider the following issues when designing back-to-back gateways:

- **Signaling**—Back-to-back gateways need to be configured with similar TDM signaling types.

- **Voice Quality and Bearer Issues**—Voice quality suffers especially in the case of tandem compression. The addition of back-to-back gateways introduces additional post-dial delay as well as added latency for all calls. There is even greater impact if more than one back-to-back zone is traversed. Fax relay may also suffer. Modem passthrough is highly unreliable, and as a result is not supported in scenarios that employ back-to-back gateways.

- **Dial Plan**—The back-to-back gateway is responsible for manipulating digits and tech prefixes to fit into the general gatekeeper and directory gatekeeper dial plan. This also includes separating ingress and egress gateways in the gatekeeper call routing table. The extent of these considerations depends upon the application and the DGK/GK dial plan design. Dial plan responsibilities are discussed in greater detail in the document *Designing Static Dial Plans for Large VoIP Networks*.

- **Billing**—One of the main benefits of the back-to-back gateway is establishing a point in the call-signaling path from which to bill for IP-to-IP call topologies. The back-to-back gateway largely functions as a normal PoP gateway. Billing options vary by application type.

- **Fault Tolerance**—If you desire, a back-to-back gateway system may be configured just like a normal TDM PoP gateway to support an alternate gatekeeper with which it will register should the primary gatekeeper fail.

- **Security**—Back-to-back gateways have the same security options and implications as normal PoP gateways.

- **Network Management**—Back-to-back gateways have the same network responsibilities responsibilities as in a normal TDM PoP.

Service Design Issues

This section describes the issues you should consider for service design. We will consider solutions for the following two kinds of services and discuss the issues associated with each, depending on the call topology used:

- Minutes Aggregation and Termination
- Card Services (Prepaid and Postpaid)

Minutes Aggregation and Termination

This solution enables you to collect traffic from multiple originating providers, then aggregate and deliver it to the termination providers you select. This may include target greenfields, resellers, dial-around callback operators, and international ISPs.

TDM-to-TDM Call Topology

If you select the TDM-to-TDM call topology for this service, consider the following issues:

- **Dial plan—Gatekeeper core**—This application utilizes the basic large-scale H.323 dial plan concept as previously discussed in this document.

- **Shared services—Billing and Settlement**—Dedicate separate gateways for each TDM interconnect partner. Provision the billing system to identify carriers by using originating and terminating gateway IP addresses. This allows you to generate appropriate CDRs to settle with customers.

- **Security**—Calls in this template type are all intradomain calls.

TDM-to-IP Call Topology Using Directory Gatekeeper-Based IP Interconnect

If you select the TDM-to-IP call topology using directory gatekeeper-based IP interconnect for this service, consider the following issues:

- **Dial plan**—The basic large-scale H.323 dial plan concept is still used. To interconnect your PoPs with your IP interconnect partners, you must add additional LRQ route statements to the peering directory gatekeepers to direct certain destination patterns between you and the interconnect partners. While these routes are added and modified in the directory gatekeepers, the rest of the network remains untouched.

- **Billing and Settlement**—In this scenario, you own only one of the gateways in the conversation, either the originating or terminating gateway, depending on the call direction. Your billing application must be able to extract enough information from one side of the call to generate a CDR.

  This requires correlating either source or destination IP addresses with a particular IP interconnecting carrier, depending on the call direction. Your billing system must maintain a database of this information to bill the interconnecting customer accurately. For calls sourced from ASPs, the list of possible originating IP addresses is typically limited to a few call-signaling proxy servers. However, for ITSPs with many gateways or PC clients, this list can be quite extensive. The list may be reduced if the ITSP forgoes performance and uses gatekeeper RCS. Once carrier identification issues are solved, AAA billing and settlement is done on the gateways.

Alternatively, the originating ITSP or ASP can include a mutually recognized carrier ID (for example, prepend ANI) in the H.323 SETUP message. The terminating gateway will then include this information in the AAA record. You can provision the billing application to recognize this carrier ID and associate it with an originating carrier. This implies, however, a trusting relationship between service providers.

- **Security**—Security can be accomplished by using Cisco H.235 access tokens. However, this means you must share a database of all gateway user IDs and passwords with all IP-based interconnecting partners.

TDM-to-IP-Based Interconnect with OSP Call Topology

If you select the TDM-to-IP-based interconnect with OSP call topology for this service, consider the following issues:

- **Dial plan**—An OSP-based interconnect partner can connect to your network by implementing OSP directly on the gateway, or through a back-to-back OSP interconnection zone.

  From a call routing perspective, OSP is most readily accepted into the network if an OSP interconnection zone consisting of back-to-back gateways is used. One gateway handles the RAS side of the call; the other the OSP side of the call. From the perspective of the directory gatekeeper, this looks like another TDM zone managed by a local gatekeeper. The directory gatekeeper simply adds LRQ routes to the OSP interconnect zone gatekeeper for specific destination patterns serviced by that OSP interconnect partner.

  Provisioning requirements for the gateways within this OSP interconnection zone are only slightly different from the requirements for a normal wholesaler TDM PoP. The OSP-side gateway is configured to interface with the OSP server. The RAS-side gateway is provisioned like a normal PoP RAS gateway. The back-to-back gateways are then configured to send all calls received through IP out TDM interfaces to the opposite gateway, using single-stage dialing. This method of OSP interconnect isolates provisioning tasks to the back-to-back gateway pair, the local hopoff gatekeeper configuration, and an added LRQ route in the directory gatekeeper. The rest of the network is unaffected.

  If OSP is implemented without using the interconnect zone, dial-peer provisioning increases dramatically in order to support OSP directly on the gateways. Separate dial peers are needed on all PoP gateways to send calls to the OSP server for route resolution instead of through RAS. You may provision dial peers on the gateways to send calls to OSP for specific destination patterns.

  For example, if an interconnect partner knows that all calls to Australia need to be terminated by OSP, you may insert a dial peer into your gateways that sends all calls beginning with a "61" country code to an OSP session target. However, any changes to the OSP dial plan require modification to the dial peers on all gateways in the network.

  You may choose to configure the gateway with rotary dial peers to handle OSP-based interconnects instead of explicit patterns. Although this may reduce the dial plan's sensitivity to changes, it still requires additional dial-peer provisioning to support failover. In this case, gateways are configured to try to terminate the call within their own administrative domain, first through RAS. If RAS offers no termination possibilities, either by explicit ARJ or RAS timeout, the gateways may fall back to a secondary dial peer to reoriginate the VoIP call through OSP.

  Consider a gateway provisioned with two dial peers having identical generic destination patterns. One dial peer points to session target RAS; the other points to session target settlement. The RAS dial peer is given a higher priority than the settlement dial peer, so that it is always attempted first. In the event that the RAS dial peer fails, the gateway then attempts to send the call to an OSP server through the secondary dial peer.

  This reduces the amount of maintenance of OSP dial peers to accommodate dial plan changes, but adds post-dial delay to all OSP-based interconnect calls.

- **Billing and Settlement**—In any OSP implementation, the OSP server collects usage information and generates CDRs. This usage information is extracted directly from the gateways registered to the OSP server, regardless of whether they are functioning as back-to-back gateways or as normal PoP gateways.

  You can also send duplicate records to a AAA server for internal accounting. These CDRs may be used to cross-check any settlement issues with the OSP provider. You may optionally employ a mediation application to automate this process.

- **Security**—If OSP is performed directly on the terminating gateway, intradomain security continues to use (optionally) Cisco access lists. Interdomain security uses OSP H.235 tokens, with the noted caveats to the dial plan. If a back-to-back gateway zone is used, the OSP token management is offloaded from your PoP gateways and is instead handled by the OSP gateway in the back-to-back zone. The OSP gateway in the back-to-back pair supports the H.235 OSP tokens, whereas the RAS gateway optionally implements Cisco access lists. This use of the back-to-back OSP transit zone allows security caveats previously mentioned in the direct method to be sidestepped.

IP-to-IP-Based Interconnect (Transit Network) with DGK Call Topology

If you select the IP-to-IP-based interconnect (transit network) with directory gatekeeper call topology for this service, consider the following issues:

- **Dial plan**—Interconnections between IP-based service providers are sent to a back-to-back gateway transit zone. Each IP interconnecting partner has a dedicated transit zone. If both interconnecting partners are made through a directory gatekeeper peering relationship, this adds complexity to the large-scale H.323 dial plan architecture. The dial plan must be altered to provide dedicated ingress and egress directory gatekeepers to route calls properly through your network. IP interconnect from one carrier using directory gatekeeper peering and an OSP-based interconnection partner using a back-to-back OSP interconnection zone is accomplished in essentially the same way as discussed for the TDM-to-IP call topology using directory gatekeeper-based IP interconnect.

- **Billing and Settlement**—The back-to-back gateway provides a point in the call-signaling path from which you may gather accounting information. Billing can be done from the back-to-back gateway in the same manner as described in the simple interconnect method of the TDM-to-TDM solution.

- **Security**—The back-to-back gateway zone also allows you to obscure originating ITSP carrier information from the terminating ITSP carrier, if desired. Calls sent into the terminating ITSP B look as if you sourced them. The terminating ITSP B has no idea that ITSP A originated the call.

  You must still share gateway-IDs and passwords with your interconnecting partners. However, the back-to-back gateway allows you to isolate interdomain security information between service providers. That is, ITSP A does not need to know ITSP B's security information, and vice-versa, for the two to complete calls between each other.

IP-to-IP-Based Interconnect (Transit Network) with OSP Call Topology

If you select the IP-to-IP-based interconnect (transit network) with OSP call topology for this service, consider the following issues:

- **Dial plan**—This extends the method described in the TDM-to-IP-based interconnect with OSP solution to include sending calls to another OSP provider through another back-to-back gateway zone or another directory gatekeeper-based service provider, depending on LRQ routing entries in the directory gatekeeper.

- **Billing and Settlement**—Billing between OSP providers is done just as discussed in the TDM-to-IP-based interconnect with OSP solution, but for two OSP back-to-back gateway zones. The originating zone provides settlement CDRs for the originating carrier; the terminating zone

provides settlement CDRs for the terminating carrier. If the call is instead sent to a directory gatekeeper interconnect, then AAA RADIUS records are used on that side. The AAA may be reconciled with the OSP usage records by means of a mediation application.

- **Security**—Security is accomplished as described in the TDM-to-IP-based interconnect with OSP solution.

Card Services (Prepaid and Postpaid)

You can host prepaid services for multiple service providers on their infrastructure. In addition, most prepaid service providers use VoIP wholesalers to terminate long-distance calls that are placed by prepaid subscribers. Using the integrated voice response (IVR) feature in the Cisco VoIP gateways, and real-time authorization and call accounting systems provided by Cisco ecosystem partners, service providers can offer this service over a VoIP network and lower the cost and deployment time of calling-card services.

Like prepaid services, you can also host postpaid services. An example is basic calling that is accessed by the 800 prefix, a calling card number, or a PIN. With postpaid service, the authorization is not tied to call rating. Consequently, call rating does not have to happen in real time, and there may be more partner billing-system options that perform adequately at scale. After calls are made, a billing system contracted by the company charges the carrier.

TDM-to-TDM Call Topology

If you select the TDM-to-TDM call topology for this service, consider the following issues:

- **Dial plan**—Card services typically affect dialing habits by employing two-stage dialing. Aside from this, dial plans remain basic. Once inside your network, the call may either be terminated at one of your PoPs or sent to another service provider through a TDM hopoff, using the basic large-scale H.323 dial plan architecture.

- **Billing and Settlement**—Your originating gateway supports card services for TDM-based interconnecting partners. AAA-based billing is done on the gateways and settled as discussed in the TDM-to-TDM solution. However, the billing server must interact in real time with the AAA server in order to offer prepaid services.

- **Fault Tolerance**—Basic H.323 fault tolerance is used.

- **Security**—An IVR script running on the originating gateway performs user authentication. This IVR script interacts with a AAA RADIUS security server. On top of this, either user-level or gateway-level security may be implemented for registration and call admission.

- **Prompting**—In order to support branding requirements, you must be able to identify the necessary IVR script for the carrier. Different call scripts may be invoked, depending on the supplied DNIS. Prompts may be stored remotely on a TFTP server if desired.

TDM-to-IP Call Topology Using Directory Gatekeeper-Based IP Interconnect

If you select the TDM-to-IP call topology using directory gatekeeper-based IP interconnect for this service, consider the following issues:

- **Dial plan**—For card services provided to TDM interconnect partners, this has the same considerations as outlined in the TDM-to-TDM template. However, you may wish to provide card services for IP interconnecting partners. In this case, you may route incoming VoIP calls directly to the terminating gateway as normal and then implement the IVR.

   Alternatively, you can provision the gatekeepers and directory gatekeepers to first route the call to a back-to-back gateway for IVR services, based on the end user dialing a specific access number. The directory gatekeeper knows to send calls destined to this access number to a particular IVR zone consisting of back-to-back gateways. The local gatekeeper is provisioned to send calls

destined to this access number to a designated ingress-only gateway of the back-to-back pair. The egress gateway is explicitly given a gateway priority of **0** to avoid sending calls through the back-to-back gateway in the reverse direction.

The ingress back-to-back gateway is provisioned to pass this call through TDM to the egress gateway. The egress gateway then applies the required IVR script, based upon the DNIS received. The egress gateway collects the desired destination pattern and reoriginates the call into the H.323 network as if it were a normal TDM PoP.

- **Billing and Settlement**—AAA-based billing is done on the gateways. However, the billing server must interact in real time with the AAA server in order to offer prepaid services. For back-to-back gateway scenarios, billing is done on one of the gateways as if it were a normal TDM PoP.

- **Fault Tolerance**—Basic H.323 fault tolerance is used.

- **Security**—Security is accomplished as described in the simple interconnect application above. Added security is provided by the IVR script in authenticating IP-based users either before the call enters your network (as with the back-to-back implementation), or at least before the call is completed through your network (as with the terminating gateway implementation).

- **Prompting**—Prompting for TDM interconnects is the same as in the TDM-to-TDM solution. In order to support the proper welcome announcements and local languages that are required for branding in IP interconnections, you must be able to identify the source carrier before authenticating the user.

  In the case where IVR is implemented directly on the terminating gateway, the called number is supplied by the end user and is routed to the destination. It is unreliable to identify the originating carrier based upon dialed DNIS. Modifications may be made to ANI, but this is also unreliably enforced on originating PC endpoints. Therefore, multiple branding is not supported in this implementation for IP interconnect partners.

  For IP interconnects front-ended with a back-to-back gateway, you may support branding services to individual carriers by providing separate access numbers which PC users dial to reach various back-to-back gateway zones. For example, carrier A is given a special destination number to dial into a back-to-back gateway IVR pool.

TDM-to-IP-Based Interconnect with OSP Call Topology

If you select the TDM-to-IP-based interconnect with OSP call topology for this service, consider the following issues:

- **Dial plan**—Dial plans may be administered in a similar fashion as discussed in the card services application in the TDM-to-TDM solution. However, in this case, front-ending IVR calls do not require routing to separate back-to-back gateway IVR zones. IVR services may be performed directly on the interconnecting OSP back-to-back gateway pair.

- **Billing and Settlement**—Billing is done as discussed in the card services application in the TDM-to-TDM solution.

- **Fault Tolerance**—Basic H.323 fault tolerance is used.

- **Security**—Security is implemented as discussed in the simple carrier-interconnect application above. Added security is provided by the IVR script in authenticating IP-based users either before the call enters your network (as with the back-to-back gateway implementation), or at least before the call is completed through your network (as with the terminating gateway implementation).

- **Prompting**—Prompting is implemented in the same manner as discussed in the card services application in the TDM-to-TDM solution. For OSP interconnects using a back-to-back gateway zone, the IVR services may be implemented on the RAS-side gateway as if it were a normal PoP gateway.

IP-to-IP-Based Interconnect (Transit Network) with Directory Gatekeeper Call Topology

If you select the IP-to-IP-based interconnect (transit network) with directory gatekeeper call topology for this service, consider the following issues:

- **Dial plan**—You may wish to provide card services for IP interconnecting partners by using a back-to-back gateway IVR zone as the front-ending application. This is done in the same way as in the TDM-to-IP call topologies using directory gatekeeper-based IP interconnect solution.

- **Billing and Settlement**—Billing is done on one of the gateways as if it were a normal TDM PoP. AAA-based billing is done on the gateways as previously discussed.

- **Security**—Security is accomplished as in the IP-to-IP-based interconnect (transit network) with OSP solution. The IVR script provides additional security by authenticating IP-based users before the call traverses the network in the back-to-back gateway.

- **Prompting**—Prompting is done as in the TDM-to-IP call topologies using directory gatekeeper-based IP interconnect solution. The back-to-back gateway essentially operates as the front-end application.

IP-to-IP-Based Interconnect (Transit Network) with OSP Call Topology

If you select the IP-to-IP-based interconnect (transit network) with OSP call topology for this service, consider the following issues:

- **Dial plan**—You may wish to provide card services for OSP-based IP interconnecting partners by using a back-to-back gateway zone, as discussed in the TDM-to-IP call topologies using directory gatekeeper-based IP interconnect solution.

- **Billing and Settlement**—Billing is done on one of the gateways as if it were a normal TDM PoP, as in the TDM-to-IP call topologies using directory gatekeeper-based IP interconnect solution.

- **Security**—Security is accomplished as in the IP-to-IP-based interconnect (transit network) with OSP solution. The IVR script provides additional security by authenticating IP-based users before the call traverses the network in the back-to-back gateway.

- **Prompting**—Prompting is done as in the TDM-to-IP call topologies using directory gatekeeper-based IP interconnect solution. The back-to-back gateway essentially operates as the front-end application.

Step 9: Configure and Provision Components

Describing how to configure and provision the components associated with your wholesale voice solution is beyond the scope of this document. For more information about configuring specific devices, please refer to the configuration material that shipped with your network devices, or, for Cisco products, please refer to http://www.ciscco.com.

Related Documents

- TBD