Collaboration Edge

Revised: November 20, 2015

This chapter describes the Collaboration Edge preferred architecture, which includes a series of servers and gateways defining access to services at the perimeter of a collaboration network. The Collaboration Edge preferred architecture provides access to public networks, including the Internet and PSTN.

The chapter presents a detailed Architecture description of Collaboration Edge, followed by a Deployment Overview section that describes how to deploy Cisco Expressway for Internet access and Cisco Unified Border Element for PSTN access. The chapter also covers High Availability for Collaboration Edge, Security for Collaboration Edge, and Scaling the Collaboration Edge Solution. Then the section on the Collaboration Edge Deployment Process presents more detailed information on deploying Cisco Expressway, Cisco Unified Border Element, Cisco voice gateways, and Cisco ISDN video gateways.

What's New in This Chapter

Table 4-1 lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 4-1 New or Changed Information Since the Previous Release of This Document

New or Revised Topic	Described in:	Revision Date
A number of minor changes to clarify some of the content	Various sections of this chapter	November 20, 2015

Core Components

The core components of the Collaboration Edge architecture are:

- Cisco Expressway-C and Expressway-E, for Internet connectivity and firewall traversal for voice and video
- Cisco Unified Border Element, for audio PSTN connectivity via IP trunks
- PSTN voice gateway, for direct audio PSTN connectivity
- ISDN video gateway, for direct video ISDN connectivity

Key Benefits

- Connect to customers and partners, independent of the technology they are implementing and the public network they are using.
- Provide for a resilient, flexible and extendable architecture.
- Provide any hardware and software client with the ability to access any public network (Internet and PSTN).
- Provide secure VPN-less access to collaboration services for Cisco mobile and remote clients and endpoints.

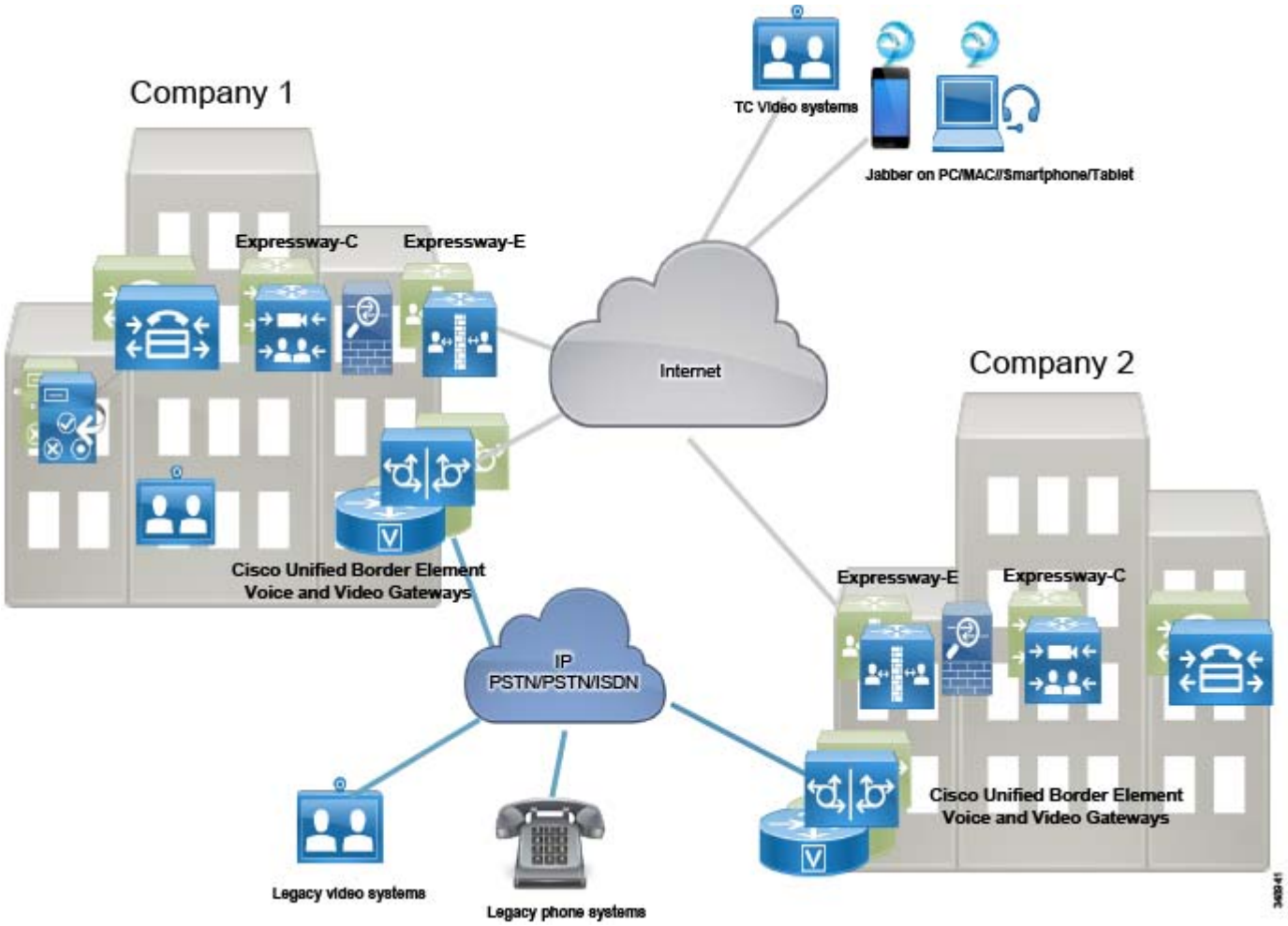
Architecture

The architecture for Collaboration Edge interfaces with two major networks: Internet and PSTN.

Internet connectivity enables VPN-less Mobile and Remote Access (MRA) and business-to-business communications. These services allow Jabber users and hardware video endpoints to access corporate collaboration services securely outside the organization's network boundaries, and they provide business-to-business audio and video communications with external organizations.

Cisco Expressway-C and Expressway-E should be deployed as a pair and in almost all cases where a firewall boundary needs to be traversed. Expressway-C sits in the internal network and Expressway-E in the demilitarized zone (DMZ), one for each side of the firewall, in order to enable firewall traversal capabilities. In addition, each Expressway-C and Expressway-E can be clustered. (See Figure 4-1.) In most cases the firewall boundary crossed is an Internet connection, but it could also be a separate corporate WiFi network for Bring Your Own Device (BYOD) connections.

Figure 4-1 High-Level View of the Architecture



PSTN connectivity enables audio and video communications to Telecom carrier networks. The PSTN connection can be achieved in multiple ways:

- Through an IP trunk to a Telecom carrier, usually for voice-only services. This connectivity is provided by the Cisco Unified Border Element (CUBE) on an Integrated Service Router (ISR) G2/G3 or Aggregation Services Routers (ASR). Cisco Unified Border Element should be deployed in a central site where the Telecom carrier's network communicates with the enterprise network.
- Through voice gateways. Gateways include analog and ISDN interfaces on a variety of router platform, such as Cisco Integrated Service Routers (ISR) G2/G3. In this document only ISDN voice interfaces are considered. Voice gateways should be deployed locally in the sites where a PSTN connection is required
- Through Cisco TelePresence ISDN Gateways 3241 or MSE 8321, which enable legacy H.320 video access to PSTN. TelePresence ISDN gateways should be centralized anywhere an ISDN video connection is required. Due to the nature and cost of TelePresence ISDN gateways, they can be shared through multiple locations.

There are cost savings associated with deploying Internet communications for video calls (Expressway) and IP PSTN connections for audio-only calls (CUBE). However, it is worth noting that:

- Not all companies have enabled Internet communications for video systems (business-to-business). If some of the partners and customers are using ISDN only for video communications, video gateways are still recommended.
- Although IP network reliability is increasing over time, network connectivity problems might prevent remote sites from accessing centralized IP PSTN services. If such sites are heavily relying on PSTN connectivity to run daily business, a local PSTN connection used as backup for the centralized access is recommended.

The recommendations for PSTN are:

- Centralize PSTN, which will help reducing operational costs and expenses.
- Local PSTN connections maintained only for those sites highly relying on PSTN to run daily business. In these cases, the number of ISDN channels should be reduced because they will be used only in those situations where central PSTN access is not available. This would help save money by reducing hardware costs and simplifying the management.

Based on the above considerations, IP trunk connections to the PSTN for voice, with local PSTN breakout used as backup and Internet for video, satisfy the vast majority of connectivity requirements. However, to provide fully connectivity, ISDN video gateways are also recommended to reach partners and customers that are still not reachable on the Internet.

Cisco Collaboration Edge includes scenarios where users have access to the following options:

- Mobile and Remote Access (MRA) for teleworkers and mobile connectivity
- Business-to-business video communications between organizations
- PSTN for cellphones and access to landlines
- ISDN video access for communications to existing H.320 video systems

Under these scenarios any corporate user inside the company or on the Internet has access to PSTN voice calls, ISDN video calls, and business-to-business communications as if they were inside the enterprise. Services such as hold, transfer, and conference are also available in most cases. Independently from who is calling whom, the Collaboration Edge solution enables interconnectivity between mobile and remote access, business-to-business, PSTN voice, and video services.

Role of Expressway-C and Expressway-E for Internet Access

Use of the Internet for collaboration services continues to increase in popularity and is quickly replacing existing legacy ISDN video systems. The two primary protocols leveraged for Internet based collaboration services are SIP and H.323.

Moreover, the Internet is also used to connect remote and mobile users to voice, video, IM and presence, and content sharing services without the use of a virtual private network (VPN).

Mobile and remote access, as well as business-to-business services, can be enabled as part of the same Expressway-C and Expressway-E solution pair. Expressway-C is deployed inside the corporate network, while Expressway-E is deployed in the DMZ.

The Expressway-C and Expressway-E pair performs the following functions:

- Interworking — The capability to interconnect H.323-to-SIP calls for voice, video, and content sharing.
- Boundary communications services — While Expressway-C sits in the corporate network, Expressway-E is in the enterprise DMZ and provides a distinct connection point for communication services between the enterprise network and the Internet.
- Security — The capability to provide authentication and encryption for both mobile and remote access and business-to-business communications.

Mobile and remote access and business-to-business calls flow through Expressway-E and Expressway-C, which handle both call signaling and media as well as other collaboration data flows, including XMPP and HTTP.

Mobile and Remote Access

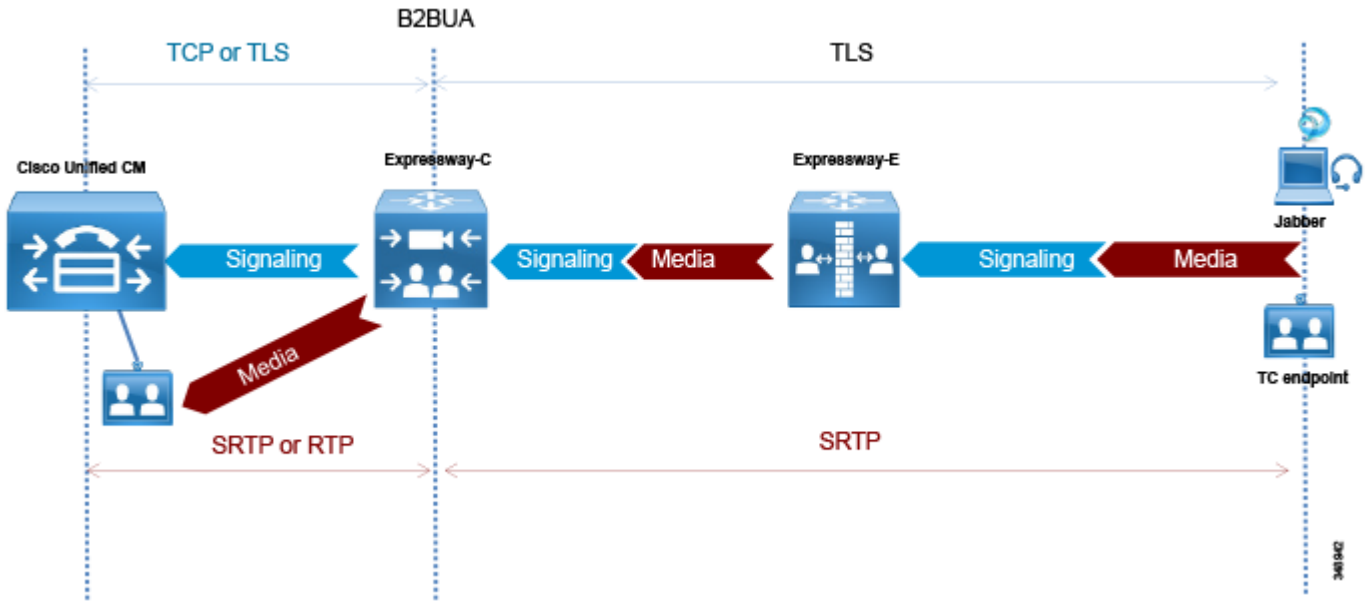
The mobile and remote access feature of the Cisco Expressway solution provides secure reverse proxy firewall traversal connectivity, which enables remote users and their devices to access and consume enterprise collaboration applications and services.

As shown in Figure 4-2, the Cisco Expressway solution encompasses two main components: the Expressway-E node and the Expressway-C node. These two components work in combination with Cisco Unified Communications Manager (Unified CM) to enable secure mobile and remote access. The Expressway-E node provides the secure edge interface to mobile and remote devices.

Expressway-C creates a secure TLS connection with the Expressway-E node. The Expressway-C node provides proxy registration to Unified CM for remote secure endpoint registration. The Expressway-C node includes a back-to-back user agent (B2BUA), which provides media termination capabilities.

Figure 4-2 shows that both signaling and media traverse Expressway-C and Expressway-E for all mobile and remote access calls.

Figure 4-2 B2BUA and Call Legs on Expressway

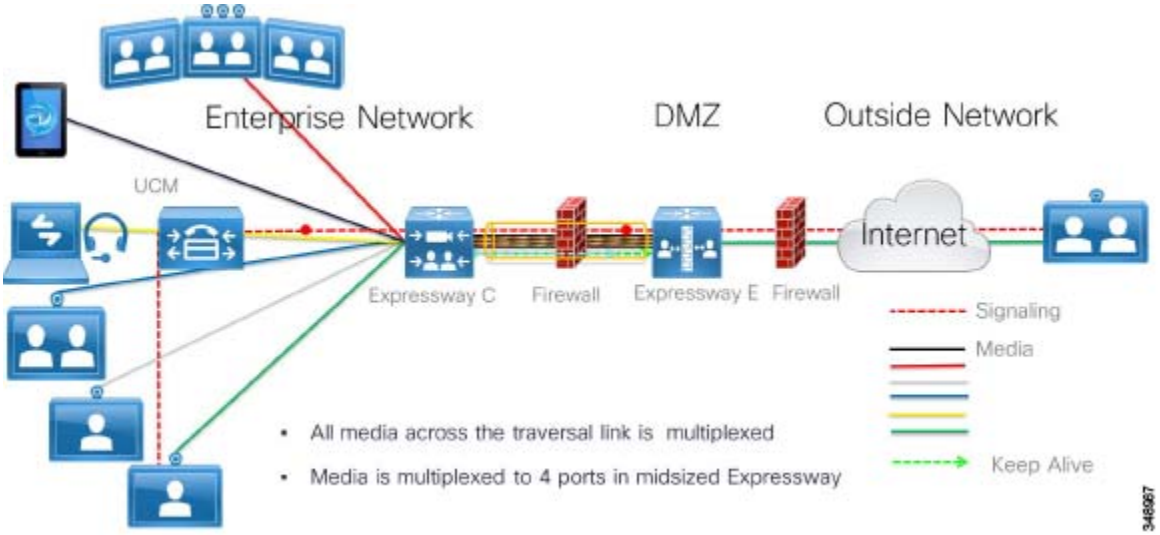


Business-to-Business Communications

Expressway-C and Expressway-E are designed to work together to form a firewall traversal solution that is the core component for business-to-business communications over the Internet.

Expressway-C sits on the inside (trusted side) of the enterprise network and serves the role of providing a secure, trusted, and standards-based way of connecting to Expressway-E. It acts as a traversal client to all devices behind it. This solves the problem for devices using a large number of media ports by multiplexing all of the media to a very small number of ports opened for outbound communications. It provides an authenticated and trusted connection from inside the enterprise to outside by sending a keep-alive for the traversal zone from Expressway-C to Expressway-E. Additionally, it provides a single point of contact for all Internet communications, thus minimizing the security risk. (See Figure 4-3.)

Figure 4-3 Expressway-C Multiplexing and Keep-Alive



Real-time and near real-time communication protocols such as SIP, H.323, and XMPP do not address the need to communicate with devices that might be behind a firewall. Typical communications using these protocols include the device IP address in the signaling and media, which becomes the payload of the TCP and UDP packets, respectively. When these devices are on the same internally routable network, they can successfully communicate directly with each other. The signaling IP address carried in the payload of the TCP packet is routable back to the initiating device, and vice versa. However, when the initiating device is on a different network behind a public or network edge firewall, two problems are encountered. The first problem is that the receiving device, after decoding the packet, will respond to the internal IP address carried in the payload. This IP address is typically a non-routable RFC 1918 address and will never reach the return destination. The second problem encountered is that, even if the return IP address is routable, the media (which is RTP/UDP) is blocked by the external firewall. This applies to both business-to-business and mobile and remote access communications.

Expressway-E sits at the network edge in the DMZ. It serves the role of solving both the signaling and media routing problems for SIP, H323, and XMPP, while maintaining standards interoperability. It changes the appropriate headers and IP addresses to process the media and signaling on behalf of the endpoints, devices, and application servers that are inside the network.

Instant Messaging and Presence Federation

Instant messaging and presence federation involves allowing users to send XMPP traffic through an organization's external firewall for chat and presence status information to and from users in another organization.

Prior Cisco architectures involved using the Cisco Adaptive Security Appliance (ASA) firewall as a TLS proxy and allowing inbound ports to be opened through the external firewall to directly access the internal IM and Presence servers. This is still the recommended solution for SIP federation.

XMPP federation uses the same Expressway-C and Expressway-E paired architecture for a trusted secure firewall traversal solution for XMPP traffic to and from external destinations. Expressway-E provides a secure DMZ-based termination point for XMPP to the Internet. Expressway-C provides a TLS-based authenticated secure connection to Expressway-E for firewall traversal, and as such does not require any port to be opened on the firewall.

Expressway-C also provides an AXL API connection to the IM and Presence server. The AXL API sends XMPP server-to-server information collected from Expressway-E to the IM and Presence database. This provides the IM and Presence server with the necessary connection information to initiate a federated connection to the other organization through Expressway-E without opening any other ports on the firewall. XMPP federation allows voice and video escalation. The same organization might implement both XMPP and SIP federation at the same time.

PSTN Access

This section describes the architecture for PSTN access using Cisco Unified Border Element as the session border controller (SBC).

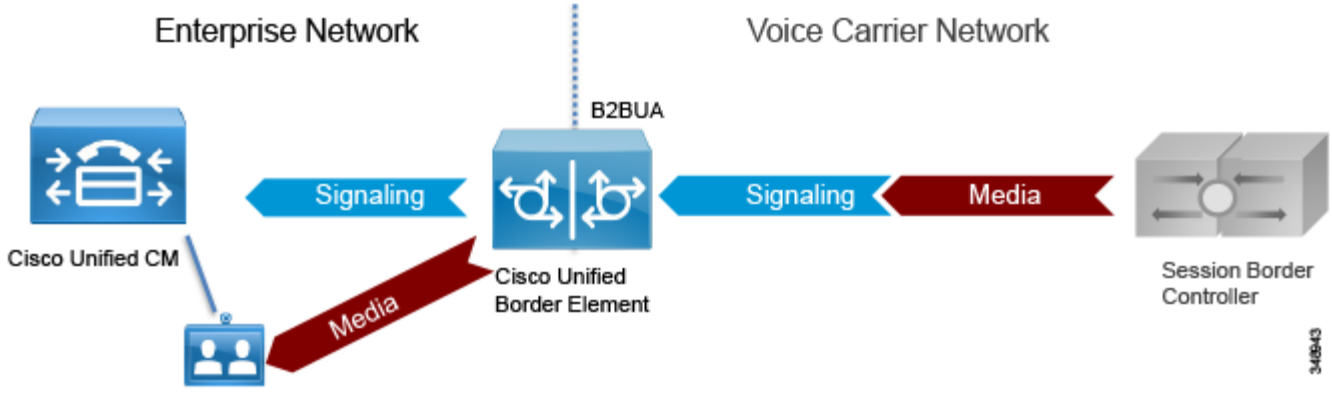
Role of the Cisco Unified Border Element

Voice connectivity using IP trunks to Telecom carriers, instead of traditional PSTN connections, is increasing in popularity and gradually replacing existing TDM-based PSTN access. SIP is commonly used as the access protocol for connectivity into provider networks, and today many Telecom carriers offer a voice-only service to the PSTN through a session border controller such as Cisco Unified Border Element. Session border controllers are SIP back-to-back user agents (B2BUAs) and are typically used in flow-through mode, where both the voice media and SIP signaling for each call flow through the Cisco Unified Border Element. (See Figure 4-4.)

Cisco Unified Border Element is a licensed Cisco IOS application available on a wide range of Cisco router and gateway platforms, and it is the recommended platform to connect to the PSTN through a SIP trunk to the Telecom carrier's border element.

Cisco Unified Border Element enables enterprise voice networks based on Cisco Unified Communications Manager (Unified CM) to connect to and interoperate with Telecom carriers through SIP trunk services. Cisco Unified Border Element terminates and re-originates both signaling and media streams to provide secure border interconnection services between IP networks. Using Cisco Unified Border Element, customers can save on their current network services, simplify their network architectures, and position their networks for ongoing enhancements in collaboration services.

Figure 4-4 Cisco Unified Border Element as B2BUA



Cisco Unified Border Element performs the following functions between the enterprise and Telecom carrier networks:

- Session control — The capability to offer flexible trunk routing, call admission control, resiliency, and call accounting for the SIP sessions.
- Interworking — The capability to offer media transcoding services for voice, and interoperability between SIP Delayed Offer and Early Offer.
- Demarcation — The capability to act as a distinct demarcation point between two networks for address and port translation and to facilitate troubleshooting.
- Security — The capability to intelligently allow or disallow real-time traffic between networks, and to encrypt the real-time traffic as appropriate for the application.

Role of Voice Gateways

We recommend using TDM gateways to connect to the PSTN if centralized PSTN access is not available. Cisco offers a full range of TDM gateways for analog and digital connections to the PSTN on ISR G2/G3 routers with appropriate interface cards enabling: low-density digital (BRI), high-density digital (T1, E1, and T3), and analog (FXS, FXO, and E&M) interfaces.

For more information on voice gateways, refer to the *Cisco 3900 Series, 2900 Series, and 1900 Series Software Configuration Guide*.

Role of Video ISDN Gateways

Video communications has a long history with ISDN. Videoconferencing first became commercially viable with ISDN as the communications protocol. Because of this, there is still the need to communicate inbound and outbound via ISDN with legacy video systems. The Cisco TelePresence ISDN Gateway performs the role of converting ISDN to SIP, and vice versa, for videoconferencing calls. The Cisco Preferred Architecture for Enterprise Collaboration includes ISDN gateway access trunked to Unified CM for the purpose of communicating with legacy videoconferencing systems.

Although Cisco ISDN video gateways manage both H.323 and SIP, only SIP is considered in the Preferred Architecture. While H.323 can still be used on networks based on the Cisco TelePresence Video Communication Server (VCS), networks based on Cisco Unified Communications Manager require SIP.

More information on the Cisco TelePresence ISDN Gateway, refer to the documentation available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-isdn-gateway/tsd-products-support-series-home.html

Deployment Overview

This section presents a general description of how to deploy Cisco Expressway for Internet connectivity and Cisco Unified Border Element for PSTN access.

Deployment of Expressway-C and Expressway-E for Internet Connectivity

The standard deployment of the Cisco Collaboration Edge architecture involves deploying at least one Expressway-C and Expressway-E pair for secure mobile device and remote VPN-less access back to enterprise collaboration services.

Both Expressway-C and Expressway-E should be deployed in a cluster to provide better resiliency. The number of servers for each cluster depends on the number of the concurrent proxied registrations to Unified CM and the number of concurrent calls. While the first takes into consideration mobile and remote users who register through Expressway to Unified CM, the second accounts for concurrent calls for business-to-business and for mobile and remote access (MRA). (See the Sizing chapter for details.)

This service is provided to Jabber clients and Cisco TelePresence System endpoints (C, EX, MX, DX, and SX Series models). Frequently, multiple pairs of Expressway-C and Expressway-E are deployed for geographic coverage and scale, providing access to multiple instances of collaboration services. GeoDNS should be used to balance remote client and endpoint access based on a variety of metrics from the Internet service provider.

This same Expressway can be leveraged for business-to-business communications as well. When the volume of calls exceeds the capacity of the Expressway cluster (600 concurrent calls for the Medium OVA template or 2,000 for the Large OVA template), business-to-business and MRA services have to be split on different boxes. See the Sizing chapter for further details.

When Expressway is used for both services, Unified CM is connected to Expressway-C through a SIP trunk for unified business communications access over the Internet. Expressway-C sits on the trusted side of the network, providing secure firewall traversal services to Expressway-E.

Based on the enterprise security policy, a number of different deployment models can be implemented. In this document we focus on a DMZ deployment with a dual interface because it is the most common and secure deployment model. For additional deployment models, refer to the *Cisco Expressway Basic Configuration Deployment Guide*.

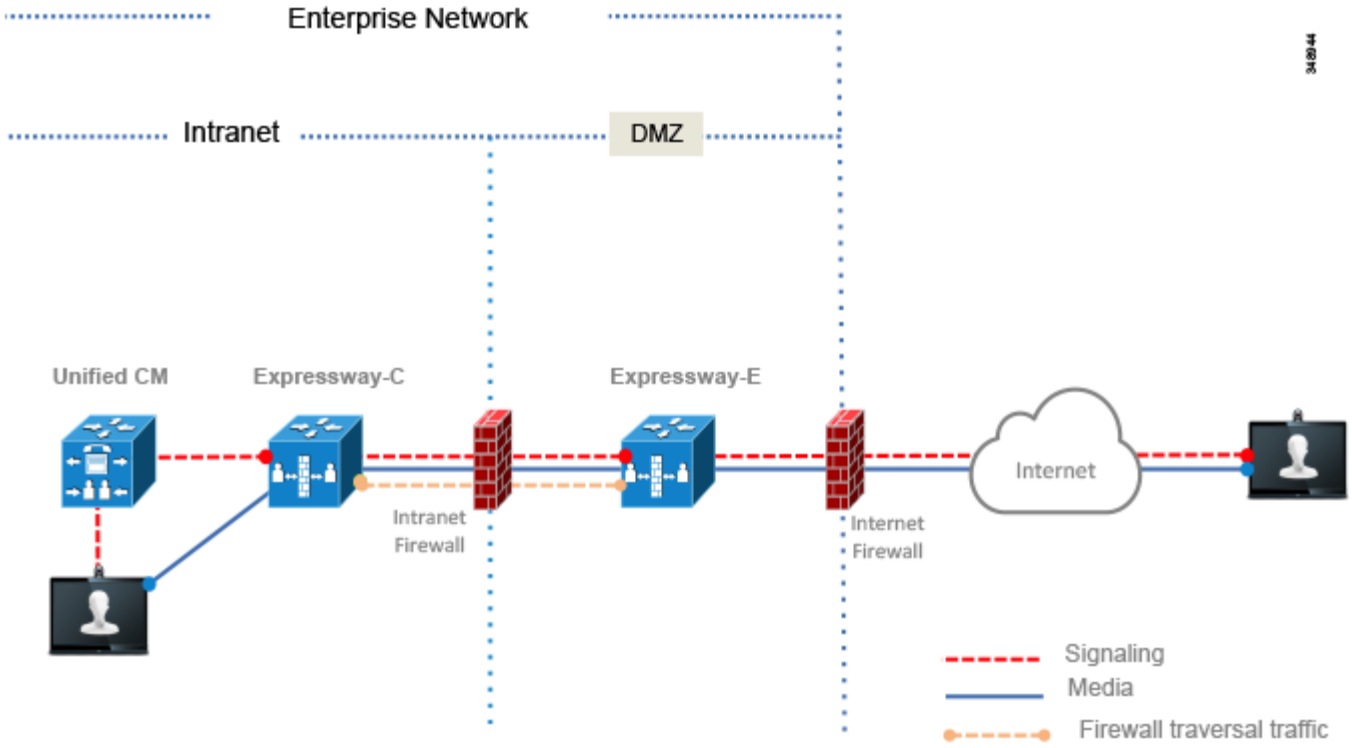
Expressway-C and Expressway-E provide firewall traversal capabilities. Firewall traversal works as follows:

1. Expressway-E is the traversal server installed within the enterprise DMZ. Expressway-C is the traversal client installed inside the enterprise network.
2. Expressway-C initiates traversal connections outbound through the firewall to specific ports on Expressway-E, with secure login credentials. If the firewall allows outbound connections, as it does in the vast majority of cases, no additional ports are required to be opened in the enterprise firewall. For ports details, refer to the *Unified Communications Mobile and Remote Access via Cisco Expressway Deployment Guide*.

   Mobile and remote access requires a separate traversal zones, called the Unified Communications traversal zone. The Unified Communications traversal zone works with SIP and requires TLS and media encryption, while the business-to-business traversal zone allows SIP and H.323 as voice and video signaling protocols. The Unified Communications traversal zone also allows XMPP and HTTPs, which are used to connect to IM and Presence servers and for provisioning purposes.
3. Once the connection has been established, Expressway-C sends periodic keep-alive packets to Expressway-E to maintain the connection.
4. When Expressway-E receives an incoming call or other collaboration service request, it issues an incoming request to Expressway-C.
5. Expressway-C then routes the request to Unified CM or other collaboration service applications.
6. The connection is established, and application traffic (including voice and video media) traverses the firewall securely over an existing traversal connection.

In order for firewall traversal to work, a traversal client zone has to be configured on Expressway-C and a traversal server zone has to be configured on Expressway-E. Figure 4-5 summarizes the firewall traversal process.

Figure 4-5 Expressway-C and Expressway-E Firewall Traversal Process



In the dual-interface deployment scenario, Expressway-E sits in the DMZ between two firewalls: the Internet firewall provides for NAT services toward the Internet, and the intranet firewall provides access to the corporate trusted network.

Expressway-E has two LAN interfaces: one toward the Internet firewall (also called the *external interface*), and the other toward the intranet firewall (also called the *internal interface*).

There is no need for the external interface to be assigned a public IP address because the address can be translated statically by NAT. In this case, the public IP address has to be configured on Expressway-E itself.

Expressway-C has an embedded B2BUA to terminate mobile and remote access as well as business-to-business calls. Expressway-E has an embedded B2BUA, used to terminate business-to-business calls. Expressway-C and Expressway-E have other B2BUAs dedicated to different services, such as Microsoft and H.323-to-SIP protocol interworking; however, the B2BUA term used in this chapter identifies the B2BUA used only in mobile and remote access and business-to-business call scenarios.

The B2BUA terminates collaboration application traffic. A connection from the Internet to Expressway-C via Expressway-E is always encrypted for mobile and remote access, while the connection between Expressway-C and the Unified Communications Manager endpoint can be encrypted or not based on the configuration. A connection from the Internet for business-to-business communications may or may not be encrypted, based on the configuration and dictated by the corporate policies. Note that in this case the communication will be encrypted on the Internet only if the remote business-to-business party supports encryption with public certificates. In all other cases, the video call will be sent unencrypted. This document focuses on encryption between the Internet and Expressway-C for mobile and remote access services only, while communications between Expressway-C and the internal back-end servers and clients are sent unencrypted. Business-to-business encryption capabilities

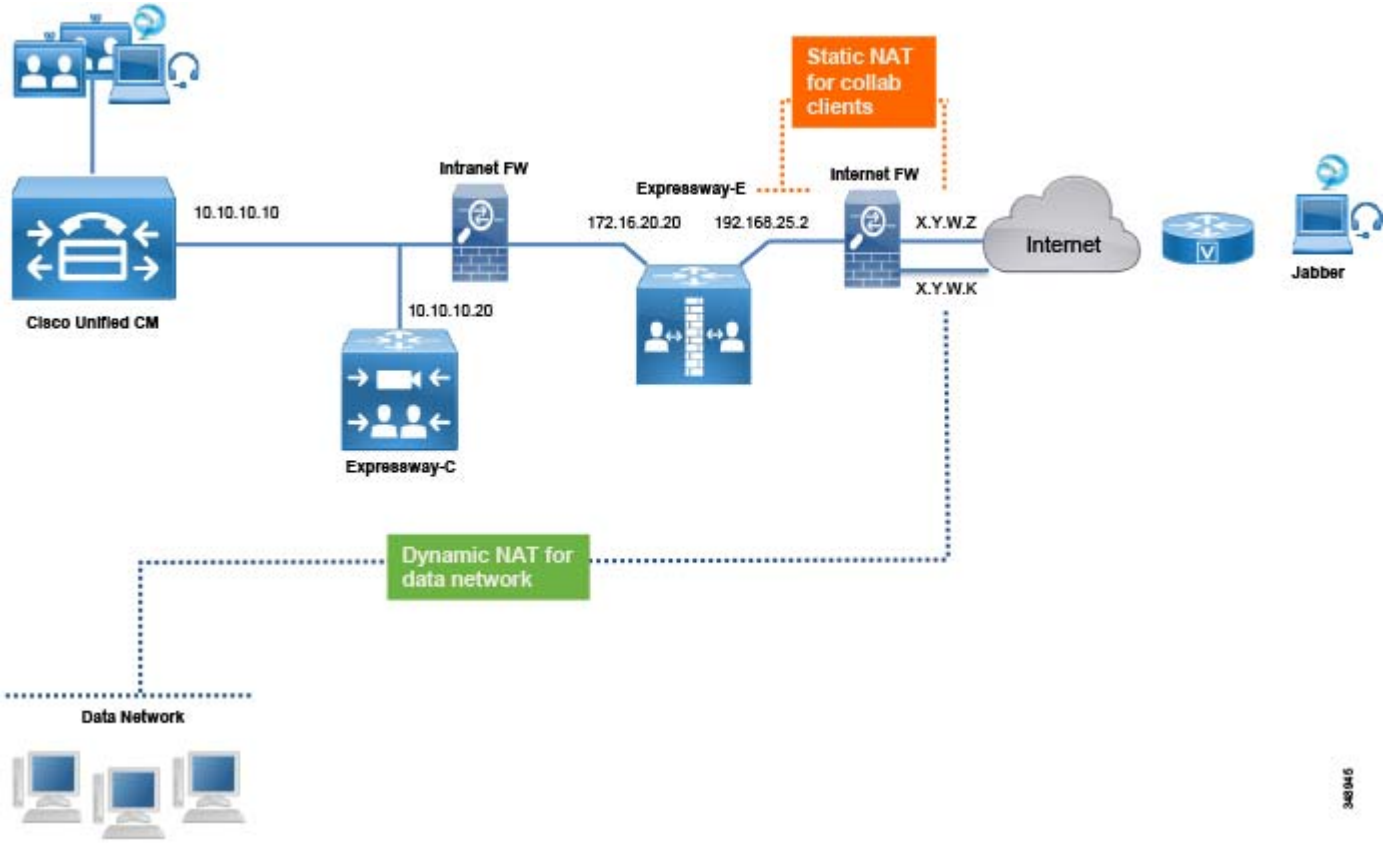
are discussed further in the section on Security for Collaboration Edge.

Expressway-C proxies the registration of mobile and remote access Jabber clients or Cisco TelePresence System endpoints to Unified CM, which lists them as registered devices with the IP address of Expressway-C.

Figure 4-6 shows the deployment described above. The relevant IP addresses are shown in the figure. Public IP addresses, which will vary based on location and Internet service provider, are shown with letters instead of digits.

Expressway-E has two interfaces; the internal interface has IP address 172.16.20.20, while the external interface has IP address 192.168.25.2. The external interface IP address is statically translated to X.Y.W.Z. This address is also configured on Expressway-E. When Expressway-E sends an INVITE, it creates the Session Description Protocol (SDP) message with the IP address set to the translated interface address instead of using its own address, so that the called party can use the public routable address instead of the private one.

Figure 4-6 NAT Interfaces on the Internet Firewall



When an endpoint on the Internet connects to Unified CM or other collaboration application through Expressway, its IP address is first translated to a public IP address. On Expressway-E, the source IP address is replaced by the address of the internal IP LAN interface of Expressway-E. When the packet enters Expressway-C, Expressway-C replaces the source IP address of the packet with its own IP address before forwarding the packet to the collaboration service applications.

In the other direction, when traffic from internal endpoints traverses the Expressway toward the Internet, their source IP addresses are replaced by the Expressway-E external LAN interface address, which is later statically translated by NAT on the Internet firewall. Source IP addresses of data devices are dynamically translated to X.Y.W.K by using another interface of the Internet firewall.

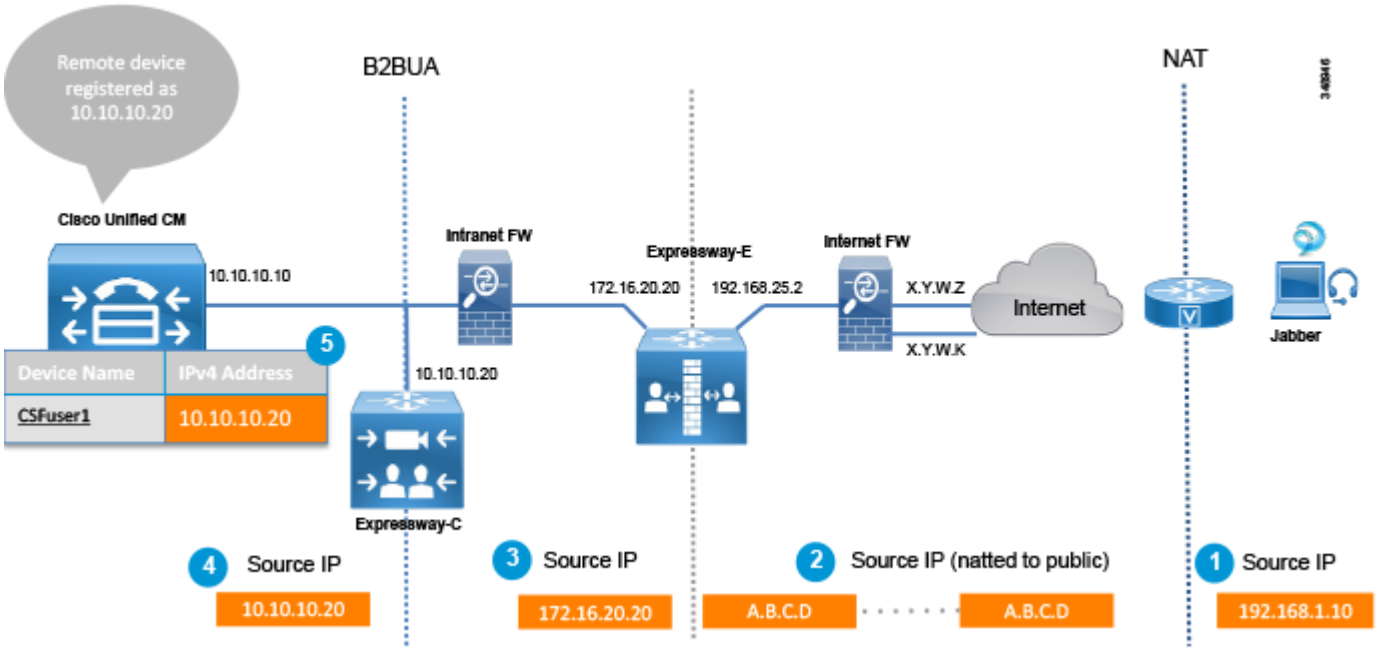
For a PC with data and a communication application, such as Jabber and a browser, the Jabber application address would be statically translated by NAT and the browser application address would be dynamically translated by NAT.

Even if the static NAT translation occurs in the firewall, the packet source IP address is transformed during its travel: it is translated to the IP address of Expressway-C when the packet goes from Expressway-C to Expressway-E, and it is translated to the IP address of Expressway-E when the packet goes from Expressway-E to the firewall. In the firewall, the packet is statically translated by NAT and sent to the Internet.

Mobile and Remote Access

In the case of call control services, Expressway-C proxy registers the endpoint to Unified CM using its own IP address, as shown in Figure 4-7.

Figure 4-7 NAT on the Life of a Packet



The address translation process shown in Figure 4-7 involves the following steps:

1. The source IP address of the endpoint is translated by NAT at the router that gives access to the Internet (192.168.1.10 to A.B.C.D.) if the endpoint does not have a public IP address.
2. The packet arrives at Expressway-E.
3. Expressway-E sends the packet to Expressway-C by using its own internal LAN interface address (A.B.C.D to 172.16.20.20).

4. Expressway-C receives the packet and terminates the connection. It re-originates another connection toward Unified CM by using its own IP address (172.16.20.20 to 10.10.10.20).
5. The endpoint is registered on Unified CM with the IP address of Expressway-C (10.10.10.20).

Registering the device to Unified CM with the IP address of the Expressway-C has some inherent benefits. For example, it is possible to limit the video bandwidth of remote devices when they are not connected directly to the corporate network, and assign them a different value when they are on-premises. Although we do not discuss it here, this can easily be achieved through the use of mobility features on Unified CM, which allow for definition of specific policies based on the IP address range.

When an endpoint is registered through the Internet, it cannot be managed remotely by the Cisco Collaboration architecture. This is because the endpoint IP address is dynamically translated and is behind a firewall. If remote management is required, deploy the endpoint through a VPN.

VPN technologies are not part of this architecture, but can be added as required.

Mobile and remote access has to be enabled on Expressway-E and Expressway-C. Expressway-C can then be configured to discover Unified CM and IM and Presence clusters by specifying the DNS name of the Unified CM and IM and Presence publisher nodes.

Expressway-E, deployed in the DMZ, provides a trusted point of entry for Jabber clients and TelePresence endpoints that use the mobile and remote access service. It also provides authentication, provisioning, registration, calling services, IM and presence, voice messaging, and directory services for remote Jabber clients and TelePresence endpoints, as well as business-to-business connectivity over the Internet.

Expressway-C connects to Unified CM and IM and Presence clusters and Cisco Unity Connection using HTTPs, SIP, and XMPP. (See Figure 4-8.)

Moreover, there are a number of cases where Jabber has to connect via HTTP to a specific server; for example, for Visual Voicemail, Jabber Update Server, custom HTML tabs and icons, and directory photo host. In these cases Jabber would connect directly to these servers without passing through Unified CM, and Expressway-C would need an HTTP allow list that specifies which servers the Jabber client is allowed to connect to.

Figure 4-8 Expressway Connection to Unified CM, IM and Presence Service, and Unity Connection

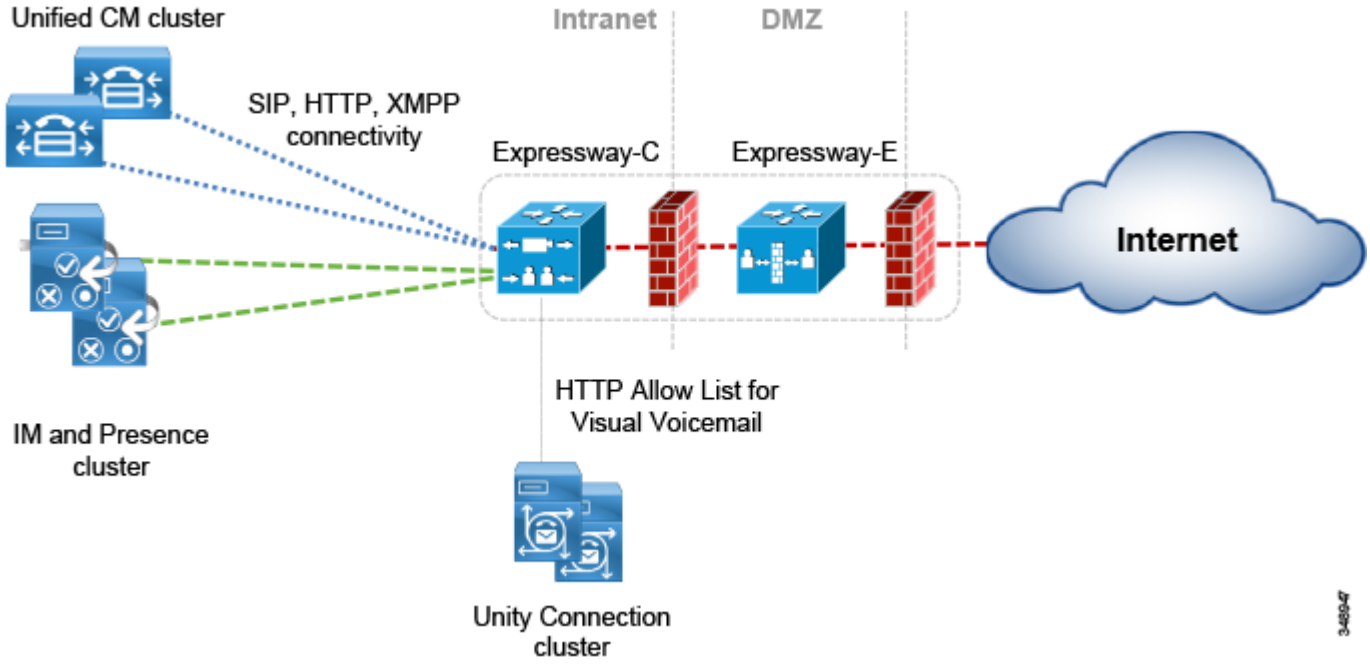


Table 4-2 summarizes the protocols used by Expressway for mobile and remote access.

Table 4-2 Expressway Protocols for Mobile and Remote Access

Protocol	Security	Service
SIP	TLS	Session establishment – register, invite, and so forth
HTTPS	TLS	Login, provisioning, configuration, contact search, visual voicemail
XMPP	TLS	Instant messaging, presence
RTP	SRTP	Audio, video, content sharing, advanced control

When a Jabber or TelePresence endpoint user logs in, they specify their fully qualified name (for example, user1@ent-pa.com). The client queries the public DNS server for specific SRV records:

- _cisco-uds._tcp.ent-pa.com, which is configured only on the corporate DNS server.
- _collab-edge._tls.ent-pa.com, which is configured only on the public DNS server and resolves to the public interfaces of the Expressway-E cluster. Note that this record always specifies TLS.

If the client is connected over the Internet, no answer will be provided by the public DNS server for _cisco-uds, but the client will receive an answer for the _collab-edge SRV record.

The DNS server will then send the A-record for Expressway-E (or multiple records if Expressway-E is clustered) to the client. Once the client knows the DNS name of Expressway-E, it can start the provisioning and registration procedure.

While provisioning takes place by using HTTPs, registration uses SIP and XMPP.

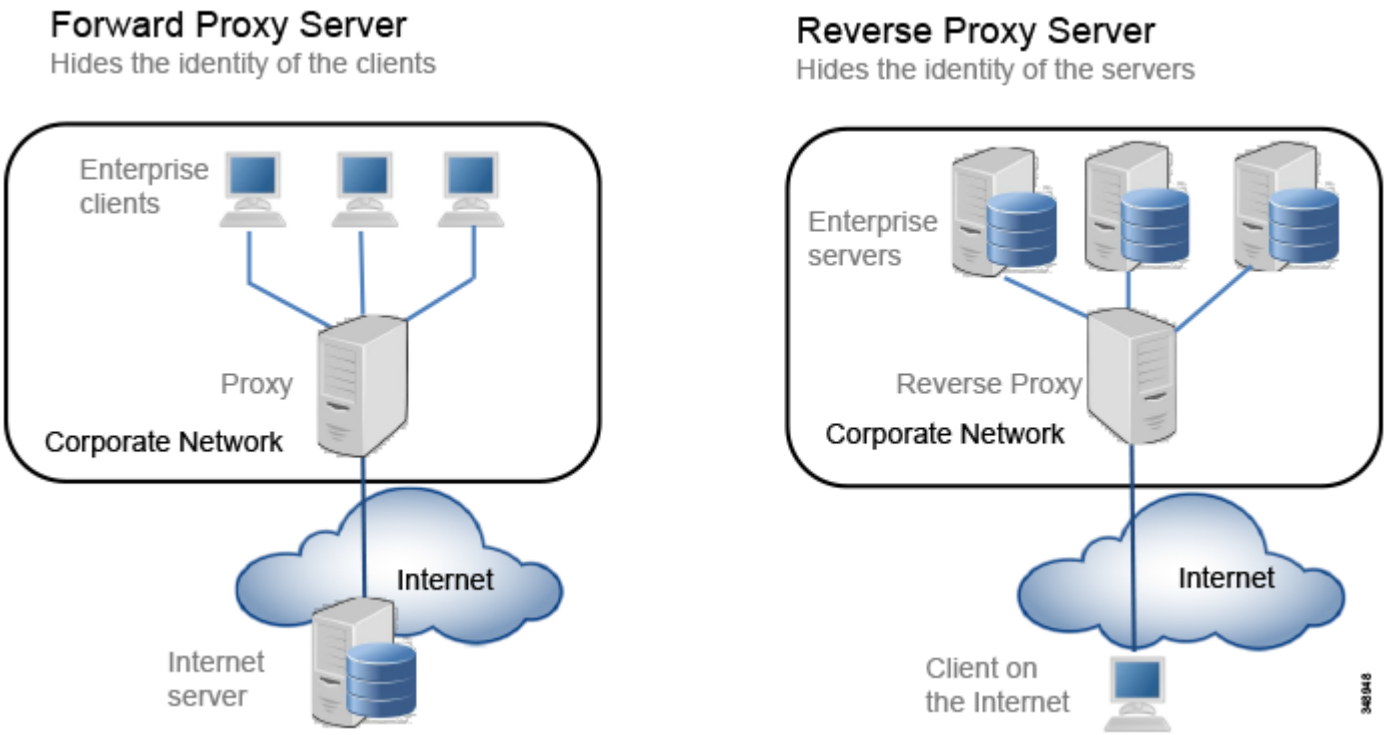
Expressway-C has an HTTPs reverse proxy server feature to manage the provisioning process. A reverse proxy is the opposite of the most common forward proxy server, also referred to as the *proxy server*.

As shown in Figure 4-9, while a forward proxy server provides services information for on-premises clients by hiding client details when connecting to Internet servers, a reverse proxy server provides information for off-premises clients by hiding on-premises server information. Clients in the corporate network, connecting through a forward proxy to an Internet server, know the identity of the server they are connecting to, but the servers do not know the identity of the clients.

On the other side, clients on the Internet connecting through a reverse proxy do not know the identity of the on-premises servers because they are connecting through the reverse proxy server, but the on-premises servers know the identity of the clients they are connecting to. This information is then returned to the client as though it originated from the on-premises servers themselves.

Expressway-C has a reverse proxy feature that provides provisioning, registration, and service details to the clients on the Internet, on behalf of collaboration application servers such as Cisco Unified CM, IM and Presence, and Unity Connection.

Figure 4-9 Forward Proxy vs. Reverse Proxy Server



Also consider that for services like Visual Voicemail, Jabber Update Server, custom HTML tabs and icons, directory photo host, Expressway-C will allow these connections if these services are specified under the *HTTP allow list*, which is a type of access list for HTTP services.

Provisioning and registration are multi-step processes that involve the client, Expressway-C, Expressway-E, Unified CM, and IM and Presence server.

The following is an overview of the major steps involved when a client registers through the Collaboration Edge.

1. Provisioning starts with the **get_edge_config** request issued from the client. For example:

   https://expressway_e.ent-pa.com:8443/ZeW50LXBhLmNvbQ/get_edge_config?service_name=_cisco-uds&service_name=_cuplogin

   Along with the request, the client sends the credentials of the user (For example, username "user1", password "user1"). The query is sent to Expressway-E, which forwards it to Expressway-C.

2. Expressway-C performs a UDS query to Unified CM to determine the home cluster for user1. This is essential for multi-cluster scenarios:

   GET cucm.ent-pa.com:8443/cucm-uds/clusterUser?username=user1

3. Once the home cluster is found, a response is sent to Expressway-C. This response includes all servers in the cluster.

4. Expressway-C asks the home cluster for provisioning information by making the following queries for user1 on behalf of the client:

   GET /cucm-uds/user/user1/devices retrieves the devices association list.

   GET /cucm-uds/servers retrieves the list of servers for the cluster.

   GET /cucm-uds/user/user1 retrieves the user and line configuration for user1.

   In response to the queries, the TFTP servers are also returned.

Subsequent queries, such as **http://us_cucm1.ent-pa.com:6970/SPDefault.cnf.xml**, are TFTP queries over HTTP. Thus, the provisioning process is done by queries to the UDS and to the TFTP server. As a result of these queries, provisioning information is forwarded to the client, and the client is able to start the registration process.

The registration process consists of two actions:

1. IM and Presence login, which is achieved via XCP router functionality on Expressway-C. The XCP router queries the IM and Presence clusters configured on Expressway-C in order to find the IM and Presence cluster where the user is configured, and the Jabber client is able to login for IM and Presence services.

2. Unified CM registration using SIP REGISTER messages, which are proxied by the Expressway SIP Proxy function.

Business-to-Business Communications

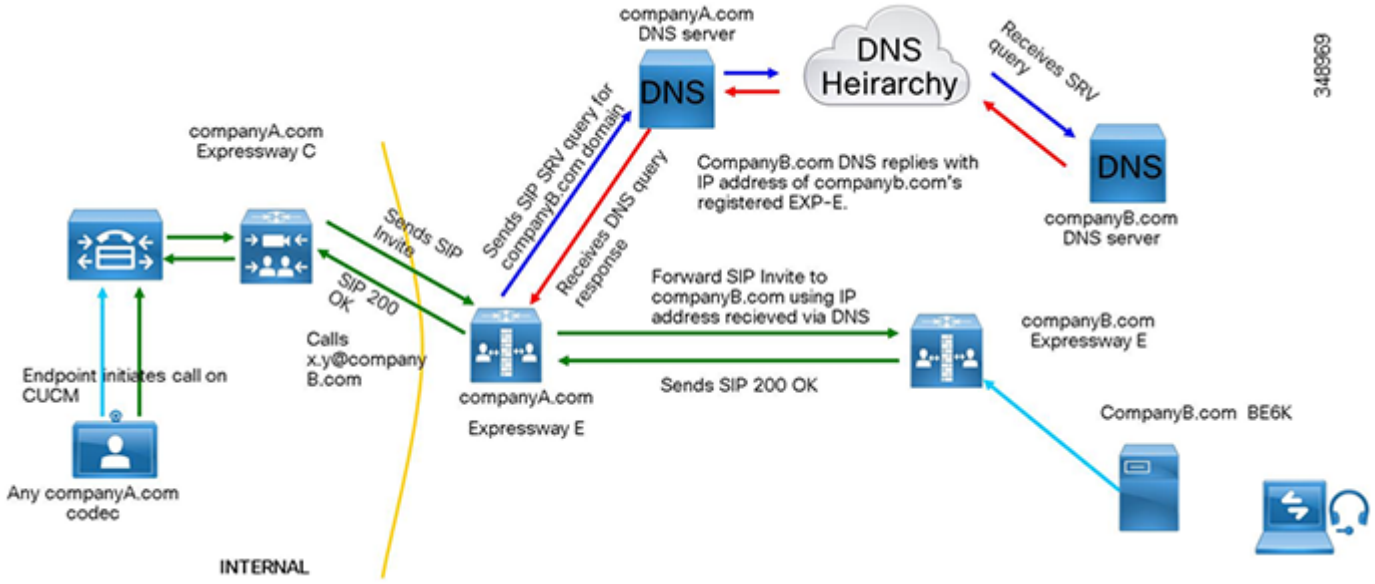
Business-to-business communications require the ability to look up the domains of remote organizations for the purpose of URI routing. This is done by creating a DNS zone on Expressway-E. This zone should be configured with the default settings. Both SIP and H.323 are set by default. This allows Expressway-E to automatically re-initiate a DNS query using the other protocol not used by the initiating call, thereby giving the call the best chance of success. Expressway-C and Expressway-E use the protocol that was used to initiate the call, and they automatically try the other protocol when SIP-to-H.323 gateway interworking is enabled on the Expressway.

SIP-to-H.323 interworking should be set to **On** for Expressway-E. If a call is receive as an H.323 call, this allows Expressway-E to interwork the call to SIP and use native SIP for the rest of the call legs to Unified CM. Likewise, an outbound call to an H.323 system will remain a SIP call until it reaches Expressway-E, where it will be interworked to H.323.

In order to receive business-to-business communications over the Internet, External SIP and H.323 DNS records are required. These records allow other organization to resolve the domain of the URI to the Expressway-E that is offering that call service. Cisco' s validated design included the SIP and SIPS SRV records and the H.323cs SRV record for business-to-business communications. The H.323ls SRV record is not necessary for Expressway-E because this record is used by an endpoint to find its gatekeeper for registration.

Figure 4-10 shows the DNS process for resolving the domain of the URI, and Example 4-1 shows an SRV lookup example.

Figure 4-10 URI Dialing with DNS



Example 4-1 SRV Record Examples for the Domain ent-pa.com

```
>nslookup
set type=srv
_sips._tcp.ent-pa.com

Non-authoritative answer:
_sips._tcp.ent-pa.comSRV service location
    priority= 1
    weight  = 10
    port    = 5061
    srv hostname= expe.ent-pa.com.
```

For more information on configuring a DNS zone on Expressway-E, refer to the *Cisco Expressway Basic Configuration Deployment Guide*.

IP-based Dialing for Business-to-Business Calls

IP-based dialing is a feature well known and used in most scenarios, when dealing with H.323 endpoints. The Cisco Collaboration Architecture uses SIP URIs and does not need IP-based dialing. However, when interacting with endpoints in other organizations that are capable of making and receiving calls using IP addresses only, the Cisco Collaboration Architecture allows IP-based dialing for both inbound and outbound calls.

Outbound Calls

Outbound IP dialing is supported on Expressway-E and Expressway-C, but it does not have full native support on Cisco Unified Communications Manager. However, it is possible to set up Unified CM to have IP-based dialing, as described below.

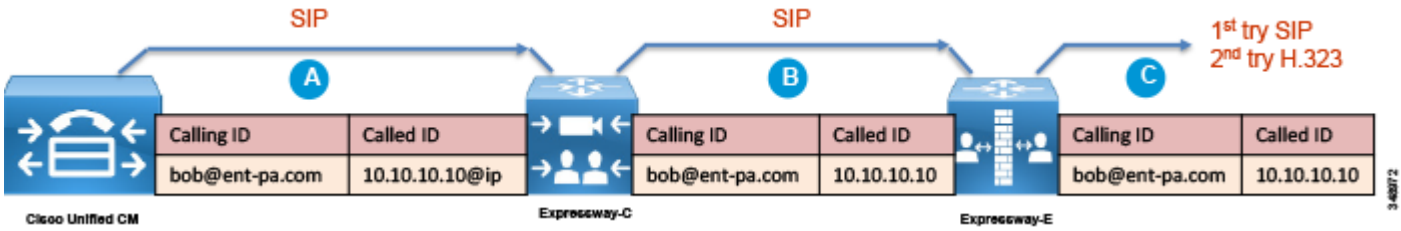
Instead of dialing the IP address alone, users on Cisco Unified CM can dial a SIP URI-based IP address as shown in this example: 10.10.10.10@ip, where "@ip" is literal and could be replaced with "external", "offsite" or other meaningful terms.

Unified CM will match a SIP route pattern configured to route the "ip" fictional domain to Expressway-C. Expressway-C strips off the domain "@ip" and sends the call to the Expressway-E, which is also configured for IP address dialing.

Calls to unknown IP addresses on Expressway -E should be set to **Direct**. Since IP-based address dialing is mostly configured in H.323 endpoints when no call control is deployed, this allows Expressway-E to send H.323 calls directly to an endpoint at a public IP address. The call will remain a SIP call until interworked on Expressway-E, as shown in Figure 4-11.

There are other options to provide for IP address dialing. One option is to replace the "." used in the IP address field with the symbol "*", as in the following example: "10*10*10*10". Cisco Unified Communications Manager will match it against a route pattern, and Expressway will replace the "*" with the "." using regular expression (regex) search rules.

Figure 4-11 Example of Outbound IP-based Dialing



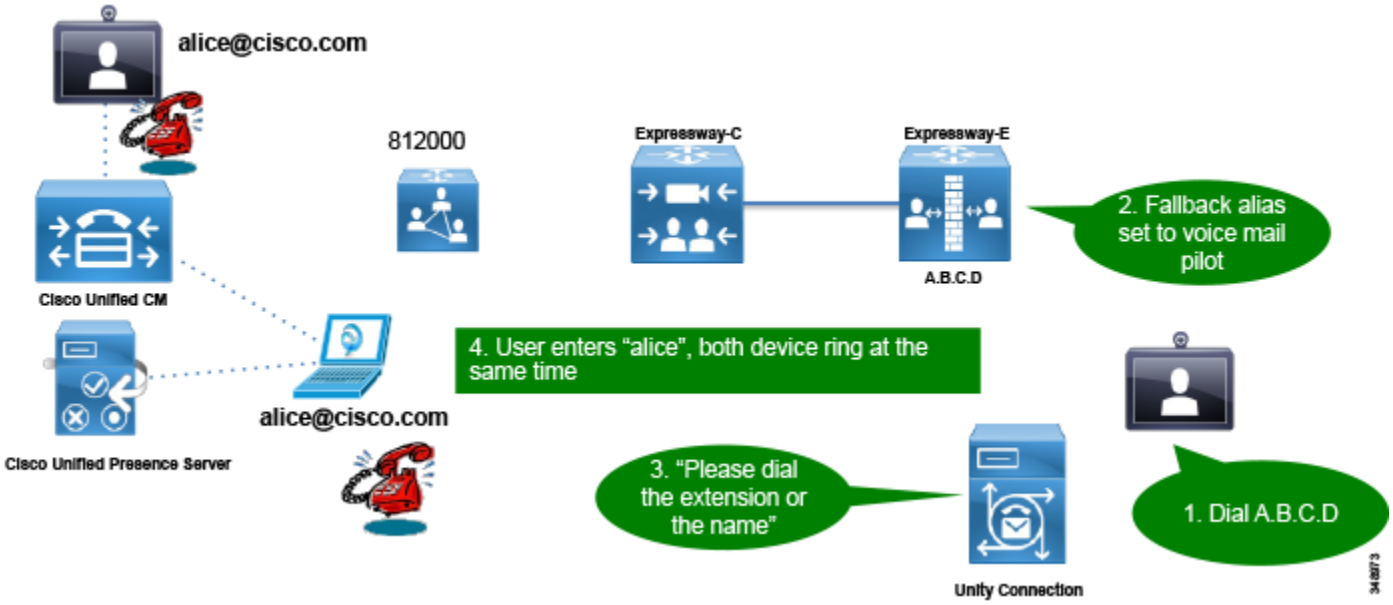
Inbound Calls

IP-based inbound calls make use of a *fallback alias* configured in Expressway-E. When a user on the Internet dials the IP address of the Expressway-E external LAN interface, Expressway-E receives the call and sends the call to the alias configured in the fallback alias setting. As an example, if the fallback alias is configured to send the call to conference number 80044123 or to the conference alias meet@ent-pa.com, the inbound call will be sent to the TelePresence Server in charge of such conferences.

If the static mapping between the IP address and the fallback alias is too limited, it is possible to set the fallback alias to the pilot number of Cisco Unity Connection. In this way it is possible to use the Unity Connection auto-attendant feature to specify the final destination through DTMF, or by speech recognition if Unity Connection is enabled to support this feature.

If Unity Connection is used as an auto-attendant feature for external endpoints dialing the IP address of the Expressway-E, remember to set the **Rerouting Calling Search Space** on the Unified CM trunk configuration for Unity Connection. Figure 4-12 shows the setup.

Figure 4-12 Example of Inbound IP-based Dialing



Deployment of External XMPP Federation through Expressway-C and Expressway-E

XMPP federation utilizes the same type of traversal connection – Unified Communications traversal – as does mobile and remote access. XMPP federation can be deployed as a standalone service. It can also be deployed on the same Expressway-C and Expressway-E pair with mobile and remote access, utilizing the same Unified Communications traversal link.

Perform the following typical tasks to deploy instant messaging and presence federation:

1. Validated email addresses for federation.

   XMPP federation through Expressway does not support translation of email addresses to XMPP addresses. Translation of email addresses to Jabber IDs is a feature of the IM and Presence server federation model. This feature is typically used to improve user experience and simplify communication for XMPP federation when the email URI convention and JID URI convention are different. When deploying XMPP federation through Expressway, the same goals of improved user experience and simplified communications apply. We recommends setting the IM and Presence domain to the same domain as the email domain. We also recommend using the LDAP sAMaccount name for UserID, email address convention, and Jabber ID. In the overall collaboration architecture, we recommend having a comprehensive and consistent strategy for URI convention that is repeatable and scalable.

2. Ensure that the IM and Presence service is operational and has XMPP federation turned off.

   XMPP federation on the IM and Presence server must be turned off so that it does not interfere with the federation configured on the Expressway.

3. Complete server certificate requirements.

   Plan ahead when setting up certificates for Expressway-C and Expressway-E. If you plan to use chat node aliases as a part of XMPP federation, the chat nodes alias FQDN must be included in the subject alternate name (SAN) field of the certificate. Doing this ahead of time avoids having to generate new certificates and possibly incurring greater expense for the public certificates on your Expressway-Es.

4. Configure the local domains for XMPP federation on Expressway-C.

5. Configure Expressway-E for XMPP federation and security.

   This step enables federation and the level of security desired for external federation. Authentication is required and is set up via the dialback secret. Securing the communications via TLS is the recommended configuration. Authorization of which foreign domains and external chat node aliases are allowed or denied, is configured in this section as well.

6. Configure how XMPP servers for federated domains and chat node aliases are located using either DNS lookups or static routes.

   The Expressway series supports federation via DNS SRV records and federation via static routes. Static routes define a path to reach external domains without having to do a DNS query. Public XMPP SRV records are used to resolve external domains that support federation. These records are required for other organizations to reach your organization when deploying an open federation model.

7. Ensure that the correct firewall ports are open.

8. Check the status of XMPP federation.

Deployment of Cisco Unified Border Element for PSTN Voice Connection Through a SIP Trunk

Cisco Unified Border Element is the recommended session border controller for PSTN centralized access. It is deployed as a demarcation point between the enterprise network and the Telecom carrier network. It gives access to the IP PSTN through its external interface and to the enterprise network through its internal interface. It enables centralized PSTN service and therefore has to be deployed where the enterprise network connects to the Telecom carrier's network.

Because all remote sites leverage central PSTN connectivity, Cisco Unified Border Element has to be highly redundant. If the PSTN central service is not available, only those offices with local PSTN access would be able to make external calls. Therefore, we recommend deploying Cisco Unified Border Element in pairs to provide redundancy.

Unified Border Element is a Cisco IOS feature set supported on the Cisco IOS Integrated Services Router (ISR) and Aggregation Services Router (ASR) platforms. For information on how to choose the correct platform, see the Sizing chapter.

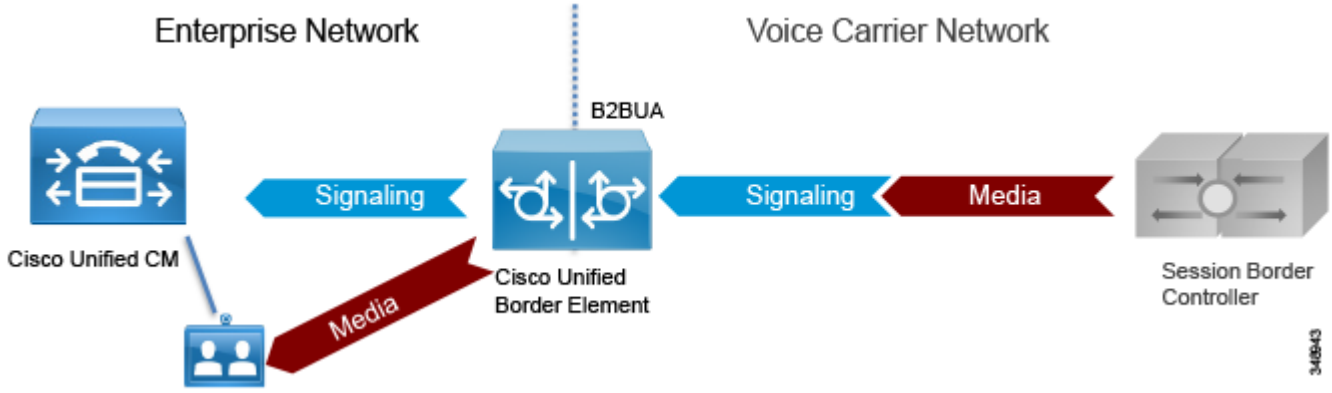
Cisco Unified Border Element is a session border control that terminates sessions from Unified CM and re-originates them toward the Telecom carrier network, and vice-versa. Note that in contrast to Expressway-E, which is exposed on the Internet, Cisco Unified Border Element is deployed between private networks: the corporate network and the carrier's network. From the carrier’s perspective, the

traffic to the centralized PSTN originates from the Cisco Unified Border Element external interface. From the enterprise's perspective, the traffic from the carrier originates from the internal Cisco Unified Border Element interface. In this sense, the Cisco Unified Border Element performs topology hiding.

Deployment of Cisco Unified Border Element is different from that of the Expressway. While the former gives access to the carrier network - a private, controlled and secured network - the latter gives access to the Internet. For this reason, deployment of Cisco Unified Border Element does not require a DMZ.

For this Preferred Architecture, as shown in Figure 4-13, the Unified Border Element has a WAN interface on the Telecom carrier network and a LAN interface on the enterprise network.

Figure 4-13 IP PSTN Architecture

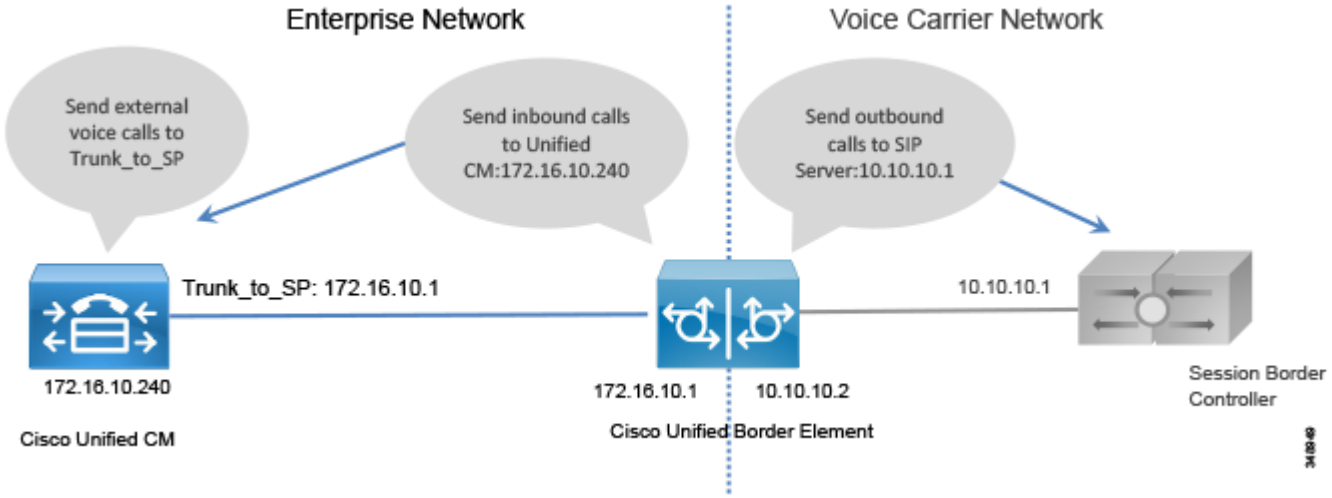


Cisco Unified Border Element performs the following functions:

- Topology hiding, as shown in Figure 4-14, including address and port translations. All traffic from Unified CM is sent to the Unified Border Element internal interface, and all traffic from the Telecom carrier soft-switch is sent to the Unified Border Element external interface. There is no direct connection between them. Figure 4-14 details the trunking configuration on Cisco Unified CM and the voice routes on the Unified Border Element.
- Delayed Offer to Early Offer conversion, and vice versa
- Media interworking — In-band and out-of-band DTMF support, DTMF conversion, fax passthrough and T.38 fax relay, volume and gain control
- Call admission control (CAC) — CAC can be performed by Unified Border Element based on resource consumption such as CPU, memory, and call arrival spike detection. CAC can be implemented at interface level or globally. While CAC configured on Unified CM is location-based, CAC configured on the Unified Border Element is resource-based. Resources-based CAC is recommended to avoid over-subscription of the Unified Border Element and for security reasons (see the section on Security for Collaboration Edge).
- Security capabilities, including RTP-to-SRTP interworking, SIP malformed packet detection, non-dialog RTP packet drops, SIP listening port configuration, digest authentication, simultaneous call limits, call rate limits, toll fraud protection, and a number of signaling and media encryption options
- Mid-call supplementary services, including hold, transfer, and conference
- PPI/PAI/Privacy and RPID — Identity Header Interworking with Telecom carriers
- Simultaneous connectivity to SIP trunks from multiple Telecom carriers

- Conversion of multicast music on hold (MoH) to unicast MoH
- Billing statistics and call detail record (CDR) collection

Figure 4-14 Trunking Considerations for Cisco Unified Border Element



PSTN Gateways

Legacy PSTN gateways are deployed in a distributed architecture, where each site has its own PSTN connection. We recommend using Cisco Unified Border Element for centralized PSTN access, but PSTN gateways can still be used as a backup for those sites heavily relying on external calls to run their daily business.

In this case the number of concurrent ISDN channels can be much lower than the number of concurrent calls to the centralized PSTN because they are used just in backup scenarios. As an example, if the normal situation allows for 30 concurrent calls to the centralized PSTN, it would be possible to size the backup ISDN gateway to support only two BRI channels, since they would be used in backup scenarios only.

Cisco voice gateways support:

- DTMF relay capabilities
- Supplementary services support — Supplementary services are basic telephony functions such as hold, transfer, and conferencing.
- Fax passthrough and T.38 fax relay

PSTN gateways support many protocols (SCCP, MGCP, H.323, SIP). SIP is the recommended protocol because it aligns with the overall Cisco collaboration solution and is the protocol of choice for new voice and video products.

Voice gateway functionality is enabled on any Cisco ISR with appropriate PVDMs and service modules or cards.

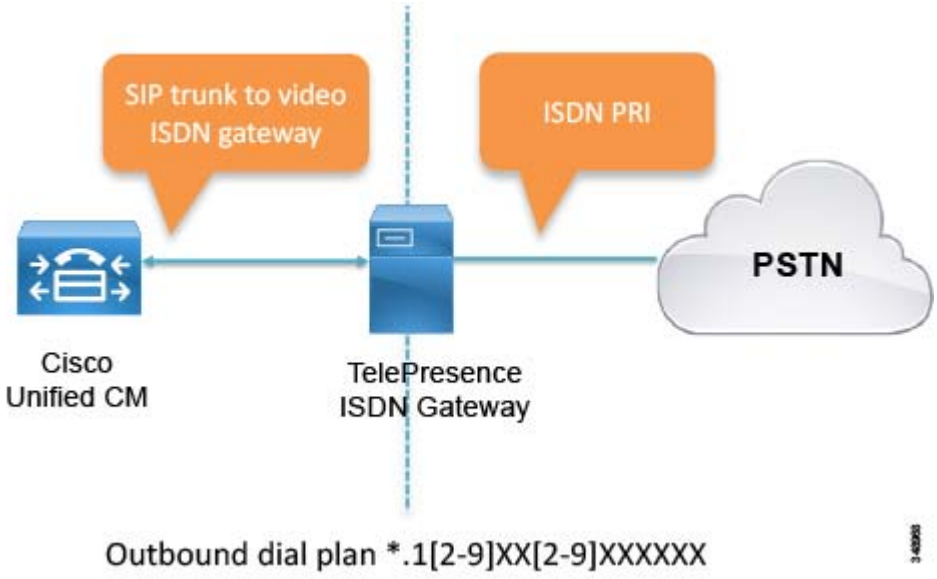
Video ISDN Gateways

The Preferred Architecture for Enterprise Collaboration incorporates the standard deployment for Cisco TelePresence ISDN Gateways with ISDN PRI trunks, and SIP trunks to Unified CM. The Cisco TelePresence ISDN Gateways include the MSE 8321 gateway and the GW 3241 gateway. The TelePresence ISDN Gateway is an optional item and is required only if there is a need to provide the ability to send and receive calls from legacy ISDN videoconferencing systems.

Requirements and Recommendations

- Use SIP instead of H.323 as the IP protocol for communicating with Unified CM.
- Make the dial plan as simple as possible on the ISDN gateway, and perform all dial string manipulations on Unified CM.
- Leave the dial plan setup for last. The ISDN gateways block all calls by default until the dial plan settings are configured. This secures the gateway until it is ready to be used.
- Use of a * in front of the ISDN number as a prefix to route calls to the TelePresence ISDN gateway. The * allows video ISDN calls to be differentiated from voice PSTN calls and does not conflict with existing international numbering plans. This also minimizes changes to the user experience when dialing. The Unified CM dial plan removes the * and routes the remaining digits through to the TelePresence ISDN gateway. (See Figure 4-15.)

Figure 4-15 Video ISDN Gateway Dial Plan



Limitations and Restrictions for ISDN Video Gateways

- Only the hold and resume features are supported during ISDN video calls.
- Video call transfer is not supported.
- ISDN calls into CMR Cloud are not supported.

Limitations and Restrictions for Mobile and Remote Access

The following limitations and restrictions apply to mobile and remote access (MRA) connectivity:

- CTI is not supported.
- Jabber desktop phone control is not supported.
- MRA does not support peer-to-peer file transfers using the IM and Presence Service or Jabber. MRA supports only managed file transfer (MFT) with IM and Presence Service 10.5.2 (and later) and Jabber 10.6 (and later) clients.
- Jabber mobile features dial-via-office reverse (DVO-R), dual-mode handoff, and session persistency are not supported.
- One-Button-To-Push for TelePresence Conductor-based TelePresence is not supported.
- TelePresence Conductor endpoint management capabilities are not supported.

Limitation and Restrictions for Cisco Unified Border Element

- For Cisco ISR, the transport protocol can be TCP and UDP only. For Cisco ASR the transport protocol can be TLS as well.
- Transcoding, DTMF interworking, IVR, SIP-to-TLS and RTP-to-SRTP conversions, and fax and modem features are preserved in failover scenarios. For Cisco ISRs, no DSP-related functions are preserved.

High Availability for Collaboration Edge

High availability is a critical aspect of designing and deploying collaboration systems. Collaboration Edge allows for redundancy, load-sharing, and call license sharing.

High Availability for Expressway-C and Expressway-E

We recommend deploying Expressway-C and Expressway-E in clusters. Each cluster can have up to six Expressway nodes and a maximum of N+2 physical redundancy. All nodes are active in the cluster. For details about cluster configuration, refer to the *Cisco Expressway Cluster Creation and Maintenance Deployment Guide*.

Expressway clusters provide configuration redundancy.The first node configured in the cluster is the *publisher*, while all other nodes are *subscribers*. Configuration is done in the publisher and automatically replicated to the other nodes.

Expressway clusters provide call license sharing and resilience. All rich media sessions and TURN licenses are shared equally across nodes in the cluster. Call licenses are contributed by the licenses configured on each node.

Expressway-C and Expressway-E deployed as virtual machines support VMware VMotion. VMware VMotion enables the live migration of running virtual machines from one physical server to another. When moving the virtual machine, Expressway-C and Expressway-E servers will maintain actives calls when handling signaling only or when handling both signaling and media. This provides high availability for the Expressway nodes as well as call resilience across Cisco Unified Computing System (UCS) hosts.

The following rules apply to Expressway clustering:

- Expressway-C and Expressway-E node types cannot be mixed in the same cluster.
- All nodes in a cluster must have identical configurations for zones, authentication, and call policy.
- Configuration changes should be made only on the master node, and this will overwrite the configuration on the other peers in the cluster when replication occurs.
- If a node becomes unavailable, the licenses it contributed to the cluster will become unavailable after 2 weeks.
- Deploy an equal number of nodes in Expressway-C and Expressway-E clusters.
- Deploy the same OVA template throughout the cluster.
- All nodes in a cluster need to be within 30 ms maximum round-trip time to all other cluster nodes. Clustering over the WAN is thus not recommended due to latency constraints.
- You must use the same cluster preshared key for all nodes within the same cluster.
- H.323 must be enabled on all nodes in a cluster for database replication. At the same time, if you also want to block H.323 calls coming from the Internet, you can configure Expressway-E with firewall rules to drop H.323 traffic on the external LAN interface.
- If mobile and remote access and business-to-business communications are enabled on the same Expressway-C and Expressway-E pairs, the SIP port number used on the SIP trunk between Unified CM and Expressway-C needs to be changed from the default 5060 or 5061.
- A DNS SRV record must be available for the cluster and must contain A or AAAA records for each node of the cluster.

Since Expressway-C is deployed in the internal network and Expressway-E in the DMZ, Expressway-C has to be connected to Expressway-E through a Unified Communications *traversal zone* for mobile and remote access. Business-to-business calls require a separate traversal zone, which retains the name of traversal client zone for Expressway-C and traversal server zone for Expressway-E. The traversal server, traversal client, and Unified Communications traversal zones include all the nodes of Expressway-C and Expressway-E, so that if one of the nodes is not reachable, another node of the cluster will be reached instead.

As shown in Figure 4-16, Expressway-C connects to all servers of the Cisco Unified CM, IM and Presence, and Unity Connection clusters, so high availability and redundancy are preserved across the entire connection path.

Figure 4-16 Expressway Services Connection

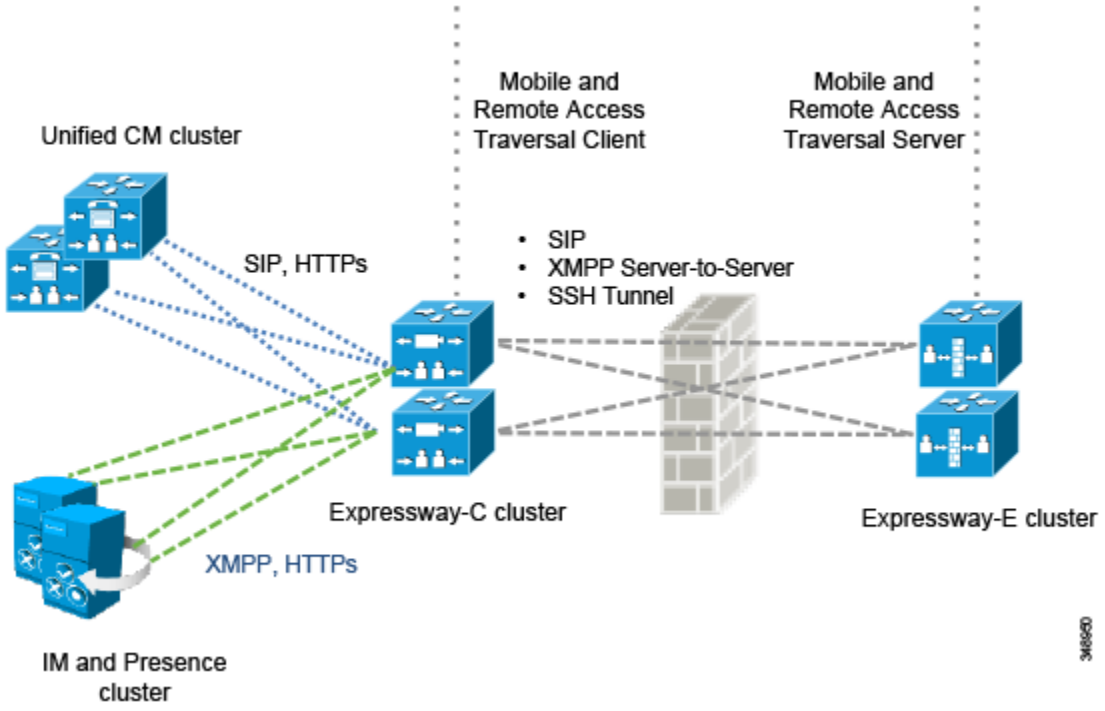


Figure 4-16 shows the high availability built into the Unified Communications traversal zone and into mobile and remote access. However, the following description applies to both Unified Communications traversal zones and standard (client and server) traversal zones.

The traversal client zone configured on Expressway-C should contain the fully qualified domain names of all of the cluster nodes of the corresponding Expressway-E cluster. Likewise, the traversal server zone should connect to all Expressway-C cluster nodes. This is achieved by including in the subject alternative names of the Expressway-C certificate the FQDNs of the Expressway-C cluster nodes and by setting the **TLS verify subject name** equal to the FQDN of the Expressway-C cluster. This creates a mesh configuration of cluster nodes across the traversal zone and provides continuous and high availability of the traversal zone until the last cluster node is unavailable.

Expressway-C connects to Unified CM via trunks for routing inbound and outbound business-to-business calls. Unified CM also trunks to Expressway-C. For high availability, the fully qualified domain names of each Expressway-C cluster node should be listed in the trunk configuration on Unified CM. If Unified CM is clustered, the fully qualified domain name (FQDN) of each member of the cluster should be listed in the neighbor zone profile of Expressway-C.

A meshed trunk configuration is created here as well. Unified CM will check the status of the nodes in the trunk configuration via a SIP OPTIONS Ping. If a node is not available, Unified CM will take that node out of service and will not route calls to it. Expressway-C will also check the status of the trunk from Unified CM via a SIP OPTIONS Ping. Calls will be routed only to nodes that are shown as active and available. This provides high availability for both sides of the trunk configuration.

DNS SRV records can add to availability of Expressway-E for inbound business-to-business traffic. For high availability all nodes in the cluster should be listed with the same priority in the SRV record. This allows all nodes to be returned in the DNS query. A DNS SRV record helps to minimize the time spent

by a client on lookups since a DNS response can contain all of the nodes listed in the SRV record. The far-end server or far-end endpoint will typically cache the DNS response and will try all nodes returned in the DNS query until a response is received. This provides the best chance for a successful call.

In addition, Expressway clusters support rich media license sharing across clusters. If a node is lost from the cluster, its call licenses will continue to be shared for the next 2 weeks. Any one Expressway cannot process any more rich media licenses than its physical capacity, even though it can carry more licenses than its physical capacity.

High Availability for Cisco Unified Border Element

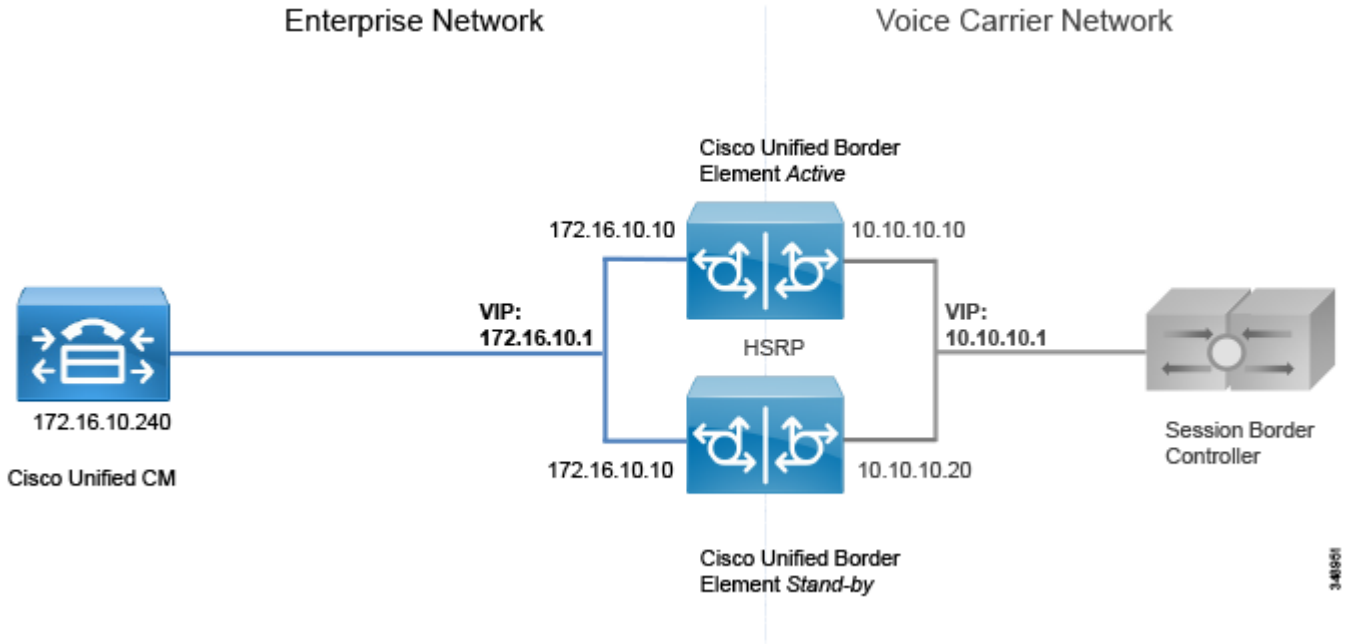
High availability for Cisco Unified Border Element can be achieved in more than one way. For the Preferred Architecture, box-to-box redundancy with call preservation is recommended because it provides both signaling and media call preservation if the Unified Border Element fails.

Unified Border Element servers are deployed in pairs, following the active/standby model. If the active Unified Border Element goes down, the standby Unified Border Element is engaged and all active sessions are transferred. This provides high availability for both signaling and media. (See Figure 4-17.)

Hot Standby Routing Protocol (HSRP) technology provides high network availability by routing IP traffic from hosts on networks without relying on the availability of any single router. HSRP is used in a group of routers for selecting an active router and a standby router. HSRP monitors both the inside and outside interfaces; if any interface goes down, the whole device is considered down, and the standby device becomes active and takes over the responsibilities of the active router.

Box-to-box redundancy uses the HSRP protocol to form an HSRP active/standby pair of routers. The active and standby servers share the same virtual IP address and continually exchange status messages. Unified Border Element session information is shared across the active/standby pair of routers, as seen in Figure 4-17, where 172.16.0.1 and 10.10.10.10 are the virtual IP addresses of the Cisco Unified Border Element pairs. This enables the standby router to immediately take over all Unified Border Element call processing responsibilities if the active router goes out of service for planned or unplanned reasons.

Figure 4-17 Cisco Unified Border Element Box-to-Box Redundancy



High Availability for Voice Gateways

PSTN gateways directly connect via physical interfaces to the PSTN network. If a gateway goes down, all communications with the PSTN are cleared. Mechanisms such as HSRP would not be of any benefit in this case, as they would in the case of PSTN access through IP trunks to a Telecom carrier. Unlike the Unified Border Element, a TDM-based PSTN gateway deployment is by nature distributed, although there are cases where a centralized PSTN with gateway interconnection is deployed. Also, a PSTN voice gateway manages a smaller amount of calls than a Unified Border Element does. Due to the nature of PSTN, media preservation is not possible in this scenario.

However, it is possible to provide signaling resilience by configuring multiple gateways in the same Unified CM route group, so that load balancing of calls will occur. If one of the gateways in the group goes down, all calls will be dropped, but new calls will be established using one of the remaining available gateways.

Security for Collaboration Edge

This section explains how to implement security in the Collaboration Edge.

Security for Expressway-C and Expressway-E

Security on Expressway-C and Expressway-E can be further partitioned into network level and application level. Network level security includes feature such as firewall rules and intrusion protection, while application level security includes authorization, authentication, and encryption.

Network Level Protection

Network level protection on Expressway-C and Expressway-E consists of two main components: firewall rules and intrusion protection.

Firewall rules enable the ability to:

- Specify the source IP address subnet from which to allow or deny traffic.
- Choose whether to drop or reject denied traffic.
- Configure well known services such as SSH and HTTP/HTTPS, or specify customized rules based on transport protocols and port ranges.
- Configure different rules for the LAN 1 and LAN 2 interfaces on Expressway-E.

The Automated Intrusion Protection feature should be used to detect and block malicious traffic and to help protect the Expressway from dictionary-based attempts to breach login security.

Automated Intrusion Protection works by parsing the system log files to detect repeated failures to access specific service categories, such as SIP, SSH, and web/HTTPS. When the number of failures within a specified time reaches the configured threshold, the source host IP address (the intruder) and destination port are blocked for a specified period of time. The host address is automatically unblocked after that time period so as not to lock out any genuine hosts that might have been temporarily misconfigured.

Mobile and Remote Access

TLS, SRTP, HTTPS, and XMPP are the only configuration options between the client on the Internet and Expressway-C for mobile and remote access.

The connection between Unified CM and Expressway-C may be encrypted and authenticated, depending on the configuration. If Unified CM is in mixed-mode, we recommend end-to-end encryption of media and signaling.

Security certificates are needed on these connections. Certificates provide the identity of servers and clients and must be deployed on Expressway-C, Expressway-E, Unified CM, and the Unified CM IM and Presence Service. The recommended configuration is to use a certificate authority (CA) to sign certificates.

CAs can be private or public. Private CA deployments have the benefit of being cost-effective, but these certificates are valid only inside the organization. Public CAs increase the security and are trusted by every organization; thus, they are commonly used for communications between different companies.

If Expressway-C and Expressway-E are used only for mobile and remote access, the company can choose to deploy a private CA or to rely on a public one. However, if Expressway-C and Expressway-E are also used for business-to-business communications, a certificate signed by public CA has to be deployed on Expressway-E. In this case, Expressway-E certificates have to be signed by a public CA such as VeriSign/Symantec, GeoTrust, GoDaddy, or others. If Expressway-E receives a business-to-business call from an entity whose certificate is signed by a CA that is present in the Trusted CA certificate list, the call is accepted. If the CA that has signed the certificate is not in the list, the call will be rejected. It is therefore important to pre-populate Expressway-E with a trust list of major CA certificates if the Expressway is enabled for business-to-business too.

For cost reduction, Expressway-C certificates may be signed by an internal CA not recognized outside the company. In this case, it is important that the internal CA certificate be included in the Trusted CA certificate list of Expressway-E in order for Expressway-C and Expressway-E to establish a connection. Table 4-3 summarizes the public and private approach for certificate deployment.

Table 4-3 Public and Private Certification Authority And Certificates

	Unified CM	IM and Presence Service	Expressway-C	Expressway-E
Certificate signed by	Internal CA	Internal CA	Internal CA	Public CA
Trust List includes	Internal CA certificate	Internal CA certificate	Internal CA and public CA certificates	Internal CA and public CA certificates

Business-to-Business Communications

Securing business-to-business communications include authentication, encryption, and authorization. Business-to-business communications use an authenticated traversal link by default. The traversal link can also benefit from the use of a Public Key Infrastructure (PKI) verified by a mutually authenticated transport layer security (MTLS) connection between Expressway-C and Expressway-E. If the business-to-business traversal link is deployed on the same Expressway-C and Expressway-E infrastructure as mobile and remote access, make sure that the traversal zone uses the FQDNs of the cluster nodes of Expressway-C and Expressway-E. This makes it straightforward to use certificates for each server to validate the offered certificate against its certificate trust for the traversal connection.

Inbound calls can be differentiated by whether they are authenticated or unauthenticated. This differentiation can be used to authorize access to protected resources. Unknown remote business-to-business calls should be treated as unauthenticated and restricted from access to protected resources such as IP voice and video gateways. This is accomplished by configuring Call Processing Language (CPL) rules with regular expressions to block access to prefixes used for gateway access. (See Figure 4-18.)

Figure 4-18 CPL Rules for Unauthenticated Callers

Signaling and media encryption is important for business-to-business calls, but it needs to be deployed carefully so as not to restrict or limit the ability to receive calls. There is a variety of older SIP and H.323 systems that you may be communicating with that do not support signaling or media encryption.

Requirements and Recommendations

- Set media encryption on the traversal client side (Expressway-C) of the business-to-business traversal zone to **best effort**. This means encryption for calls will always be tried first and DMZ traffic will be encrypted. It also allows fallback to unencrypted calls. If strong encryption policy is required, set the media encryption to **force encrypted**.
- Use TLS for signaling encryption for the SIP trunk between Unified CM and Expressway-C.

If mobile and remote access (MRA) is also deployed in the organization, both MRA to business-to-business calls can potentially go out unencrypted. Best-effort media encryption means that the outbound calls will be tried with encryption first.

Certificate settings for mobile and remote access and business-to-business call scenarios:

- General requirements for certificates are that the fully qualified DNS name (FQDN) of Expressway-C and Expressway-E must match the hostname in the certificate.
- While business-to-business communication does not have any other requirements for certificates, MRA has more. For a detailed explanation on how to set up certificates on Expressway for mobile and remote access, refer to the *Unified Communications Mobile and Remote Access via Cisco Expressway Deployment Guide*.

Security for Cisco Unified Border Element

Unlike Internet connections, PSTN connectivity over IP trunks is delivered through a private network offered by the Telecom carrier. As such, it is a controlled network. Security deployed for the Internet Edge is thus different from that deployed for IP PSTN access. Between the Cisco Unified Border Element and the carrier, there are no firewalls; however, in specific cases, companies and Telecom providers require the use of an enterprise DMZ.

Between the carrier and the enterprise network, the traffic is sent unencrypted. Depending on the corporate policies, internal enterprise traffic can be encrypted or not. In such cases, the Unified Border Element is able to perform TLS-to-TCP and SRTP-to-RTP conversion. Usage of the internal CA to sign Unified Border Element certificates is recommended when multiple gateways are deployed.

Because the Unified Border Element is deployed without a firewall, it is protected at various layers. As an example, it is possible to create an access control list to allow only the Telecom carrier's session border control to initiate calls from the PSTN side, and to allow only Unified CM to initiate calls from the internal network side.

The Unified Border Element is also protected against toll fraud and telephony denial of service (TDoS) attacks. Large packet arrival rates can also be mitigated through call admission control mechanisms based on CPU, memory, bandwidth utilization, and call arrival spike detection.

Security for Voice Gateways

PSTN gateways have an interface on the customer network and a second interface on the PSTN. They are deployed inside the corporate network and they are not reachable from the Internet. The PSTN is inherently secure; thus, there are no specific tools to use to protect the gateway, unless the gateway is

deployed on a router that also gives access to the Internet. In this case Cisco IOS features on the gateway can be used to perform firewall and intrusion protection. In all other cases, no specific tools are required to protect the gateway (for example, denial of service (DoS) prevention and so on).

However, it is always possible to encrypt media from the endpoint to the gateway. In such cases the gateway will use TLS and SRTP. Use of CA-signed certificates is recommended in this case.

Security for Video ISDN Gateways

Video ISDN gateways have an IP interface on the customer network and another interface on the PSTN. Two common security threats that need to be guarded against with gateways are toll fraud from IP-to-ISDN and ISDN hairpinning of calls.

Basic CPL rules on Expressway-E can be used to block access to ISDN gateway resources from the Internet. Calling search spaces should be used to block access from Unified CM registered devices.

Even though the entire dial plan and permissions are configured on Cisco Unified Communications Manager, it is good security practice to use CPL rules on Expressway-E to block fraudulent attempts to access voice and video gateways before they are routed to Unified Communications Manager.

Scaling the Collaboration Edge Solution

The number of Collaboration Edge clusters deployed does not depend on the number of call control clusters, but rather on the number of connection points to the Internet. A customer with multiple Unified CM and IM and Presence clusters, and multiple TelePresence Conductor clusters, will have a single Internet Edge if that customer has a single Internet breakout point. The same environment might have multiple PSTN hop-offs if the Telecom carrier offers more than one connection point to the PSTN network. The same considerations apply for video ISDN access.

Scaling the Internet Edge Solution

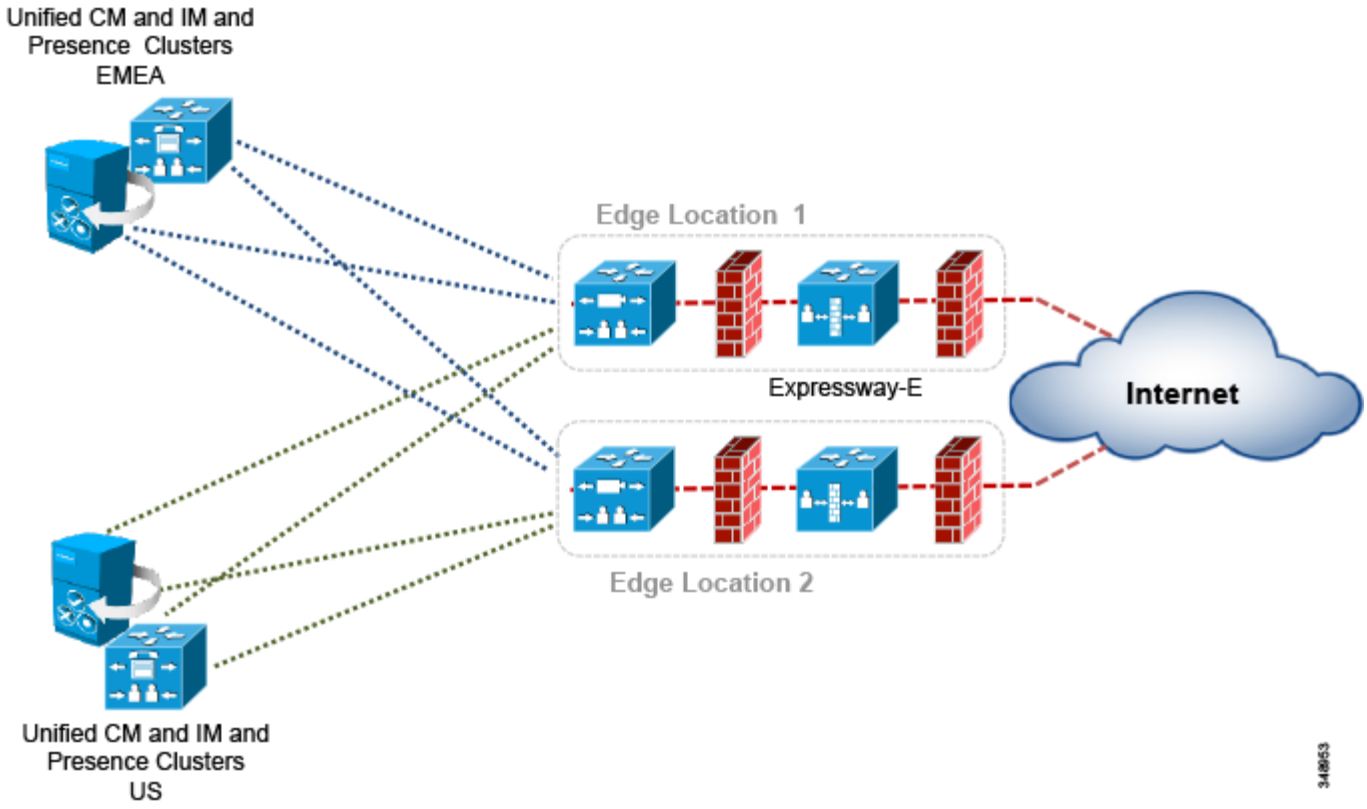
When multiple Internet Edges are deployed, it is important to set routing rules properly in order to send collaboration traffic to the nearest Internet Edge.

Mobile and Remote Access

If multiple Unified CM and IM and Presence clusters are deployed, every Expressway-C must discover all Unified CM clusters. If Expressway-C discovers only some of the clusters, it will be able to proxy registration only for those users belonging to the clusters that have already been discovered.

If a registration request is made from a client belonging to a Unified CM and IM an Presence cluster that has not been discovered by Expressway-C, that client will not be able to log in. This is the reason why it is important for each Expressway-C to discovers all Unified CM and IM and Presence clusters if users are enabled for mobility, as shown in Figure 4-19.

Figure 4-19 Service Discovery of Multiple Unified CM and IM and Presence Clusters



When two or more Internet Edges are deployed, it is important to understand how to split the load between them. If the Internet Edges are deployed in the same datacenter or in the same area, load balancing can occur at the DNS SRV level. As an example, if the enterprise network includes three Internet Edges used for mobile and remote access, each one consisting of a cluster of two Expressway-E and Expressway-C nodes, the _collab-edge._tls.ent-pa.com will include all six Expressway-E records at the same priority and weight. This distributes the registrations and calls equally across the various Expressway-E and Expressway-C clusters.

Once a mobile-and-remote-access connected endpoint is registered through a specific Expressway cluster pair, it will stay connected until the client is disconnected or it has been switched off.

However, if the Expressway clusters are deployed across geographical regions, some intelligent mechanisms on top of the DNS SRV priority and weight record are needed to ensure that the endpoint uses the nearest Expressway-E cluster.

As an example, if an enterprise has two Expressway cluster, one in the United States (US) and the other in Europe (EMEA), it is desirable for users located in the US to be directed to the Expressway-E cluster in the US while users in Europe are directed to the Expressway-E cluster in Europe. This is facilitated by implementing GeoDNS services. GeoDNS services are cost effective and easy to configure. To show how GeoDNS services work, the example below uses an Amazon Route 53 Geo DNS server. There are many GeoDNS services available in the market, including Amazon Route 53, Edge Director, GeoScaling, Max Mind GeoIP2, and others.

With GeoDNS it is possible to route traffic based on different policies such as location (IP address routing), latency (minimum latency), and others. Amazon Route 53 allows routing by both latency and geographical location. We have chosen to configure latency-based routing, but the configuration steps are identical for geographical location routing based on IP addresses.

With latency-based routing, a client in the same site might access different datacenters over time if latency on the Internet changes. However, this does not happen instantly as soon as latency changes, since it is measured as a mean value over a period of time. Spikes due to instant congestion of the Internet are thus absorbed by the mean value.

In our scenario, two Internet Edge Expressway clusters are deployed, one in the US and one in Europe, each composed of two Expressway-C and Expressway-E servers. If the measured latency between the endpoint and the European Edge is less than the latency between the endpoint and the US Edge, the endpoint will be directed to the European Edge for registration.

Although some GeoDNS providers support GeoDNS services on SRV records, many others allow CNAME or A-records only. The recommendation is to implement GeoDNS services on SRV records. The following example shows how to configure the Geo DNS if only CNAME is supported for Geo DNS services.

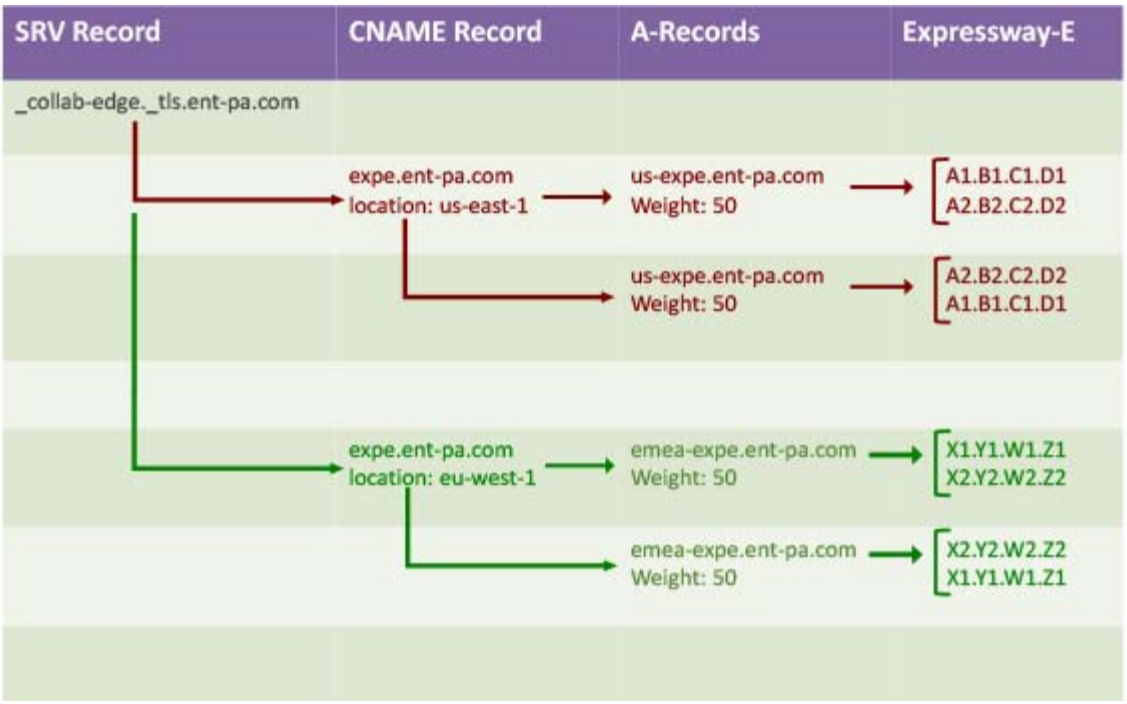
Following this scenario, a single SRV record _collab-edge._tls.ent-pa.com is configured for mobile and remote access. This record resolves into expe.ent-pa.com, a CNAME record, which is an alias that resolves into the real A-records for that resource. There are two records for expe.ent-pa.com; the first resolving into us-expe.ent-pa.com (DNS name for the US Edge) and the second resolves into emea-expe.ent-pa.com (DNS name for the EMEA edge). A-records us.expe.ent-pa.com and emea-expe.ent-pa.com resolve to the IP addresses of Expressway-E server nodes for US and Europe.

While _collab-edge._tls.ent-pa.com is configured for standard routing, expe.ent-pa.com records have a routing policy set to "latency." As a result, the locations of the Expressway-E clusters have to be specified. If the latency between the client and emea-expe.ent-pa.com is less than the latency between the client and us-expe.ent-pa.com, the registration request will be sent to the European Expressway-E. If for any reason latency changes over time and goes above the latency to the US, the us-expe.ent-pa.com will be selected instead.

Both emea-expe.ent-pa.com and us-expe.ent-pa.com are A-records, and the Jabber client or TelePresence Conductor system performs a subsequent query to emea-expe.ent-pa.com or us-expe.ent-pa.com, based on the answer of the _collab-edge._tls.ent-pa.com SRV record. However, since standard A-records cannot set priority and weight like SRV records can, another load-balancing and redundancy mechanism is needed in order to specify which server of the Expressway-E cluster the client has to connect to. This can be done by using a round-robin mechanism. As an example, two emea-expe.ent-pa.com records are created, each one with the routing policy set to "weighted." Specifying the same weight for the two records assures that an equal load-balancing process takes place between the servers of the cluster. The first record resolves into multiple Expressway-E servers of the same cluster (in this case, two servers). The second record resolves into the same set of servers, but in reverse order.

Figure 4-20 shows the DNS record structure for GeoDNS with latency-based routing between regional Expressway-E clusters and round-robin inside the same cluster. As shown in the figure, both records emea-expe.ent-pa.com resolve into the same set of Expressway-E nodes, but in different orders. This provides both redundancy and load balancing.

Figure 4-20 Route 53 DNS Record Structure for Latency-Based Routing



For each Expressway cluster node, an A-record has to be created, as shown in Figure 4-21.

Figure 4-21 DNS A-Records for Expressway Nodes

A-Records for Expressway-E	IP addresses for Expressway-E
us-expe1.ent-pa.com	X1.Y1.W1.Z1
us-expe2.ent-pa.com	X2.Y2.W2.Z2
emea-expe1.ent-pa.com	A1.B1.C1.D1
emea-expe2.ent-pa.com	A2.B2.C2.D2

Each new Expressway location deployment will require a new CNAME record and as many A-records as the number of nodes in the Expressway cluster. In addition, an A-record for each individual Expressway-E node is also needed.

Business-to-Business Communications

Scalability for business-to-business communications can be addressed by adding multiple Expressway-C and Expressway-E clusters, either in the same physical location or geographically dispersed.

When multiple Expressway-C and Expressway-E pairs are deployed, Unified CM can direct an outbound call to the edge server that is nearest to the calling endpoint, thus minimizing internal WAN traffic. Additionally, when utilizing multiple edge clusters, the Expressway-Cs should form a meshed trunk configuration with the Unified CM clusters. This adds more scalability and resiliency by allowing additional outbound traversal paths if the geographically located traversal is full or not available.

For large deployments it might be preferable to host business-to-business communications on Expressway-C and Expressway-E pairs separate from mobile and remote access. This allows the server resources to be dedicated to external Internet communications.

Considerations for Inbound Calls

DNS SRV records are used to determine which Expressway-E clusters are authorized for the SIP and H.323 ent-pa.com domain. SRV records with the same weight and priority are used to balance calls across Expressway-E cluster nodes.

When scaling inbound calls across multiple geographically dispersed Expressway-E clusters, load balancing traffic becomes the primary consideration. Expressway-C and Expressway-E do not support load balancing of SIP or H.323 traffic. Therefore load balancing of the response to the DNS query becomes an important means of scaling the solution.

Much like with the mobile and remote access service, GeoDNS is used to direct different DNS responses to the same queries. Different metrics such as network latency and geographical location should be used to provide the correct Expressway-E cluster in the DNS response. Depending on the Internet service provider providing the GeoDNS service, status monitoring of the Expressway-E servers should also be included. This allows for a more efficient DNS response, for example, that does not include an out-of-service Expressway-E.

GeoDNS is a very good method of providing the best edge, Expressway-E, for the other server or endpoint to connect to, based on the metrics chosen by the customer. The response here is typically based on the edge physically closest to the server making the query. The mechanism is the same as described in the previous section, except that the SRV records are different. As an example, a SRV record for SIP TLS would be: _sips._tcp.ent-pa.com. Figure 4-20 can be used in order to set up GeoDNS service, where _collab-edge._tls.ent-pa.com is replaced by _sips._tcp.ent-pa.com

An alternative solution is designed to return the edge that is closest to the destination endpoint or device. This requires finding or knowing where the destination endpoint is located and then returning the appropriate edge. The benefit of this solution is to minimize the use of bandwidth on the customer network by delivering the shortest internal path to the endpoint.

This can be achieved by configuring Expressway-E to direct the call to the Expressway-E in another region if the called endpoint belongs to the other region.

As an example, consider two Expressway-C and Expressway-E clusters in EMEA, and another two Expressway-C and Expressway-E clusters in APJC. The Unified CM inbound calling search space on the Expressway-C trunk in EMEA will contain the partition of the EMEA phones but not the partition of the APJC phones. Analogously the inbound calling search space on the Expressway-C trunk in APJC will contain the partition of the APJC phones but not the partition of the EMEA phones. If a user on the Internet in EMEA calls a corporate endpoint in APJC, the call will be sent by DNS to the EMEA Expressway-E cluster, the default for business-to-business calls. The EMEA Expressway-E and Expressway-C will try to send the call to the destination, but the inbound calling search space of the

Expressway-C trunk will block the call. The EMEA Expressway-E will then forward the call to the APJC Expressway-E. This time the call will be delivered to the destination because the inbound calling search space of APJC Expressway-C contains the APJC endpoints partition.

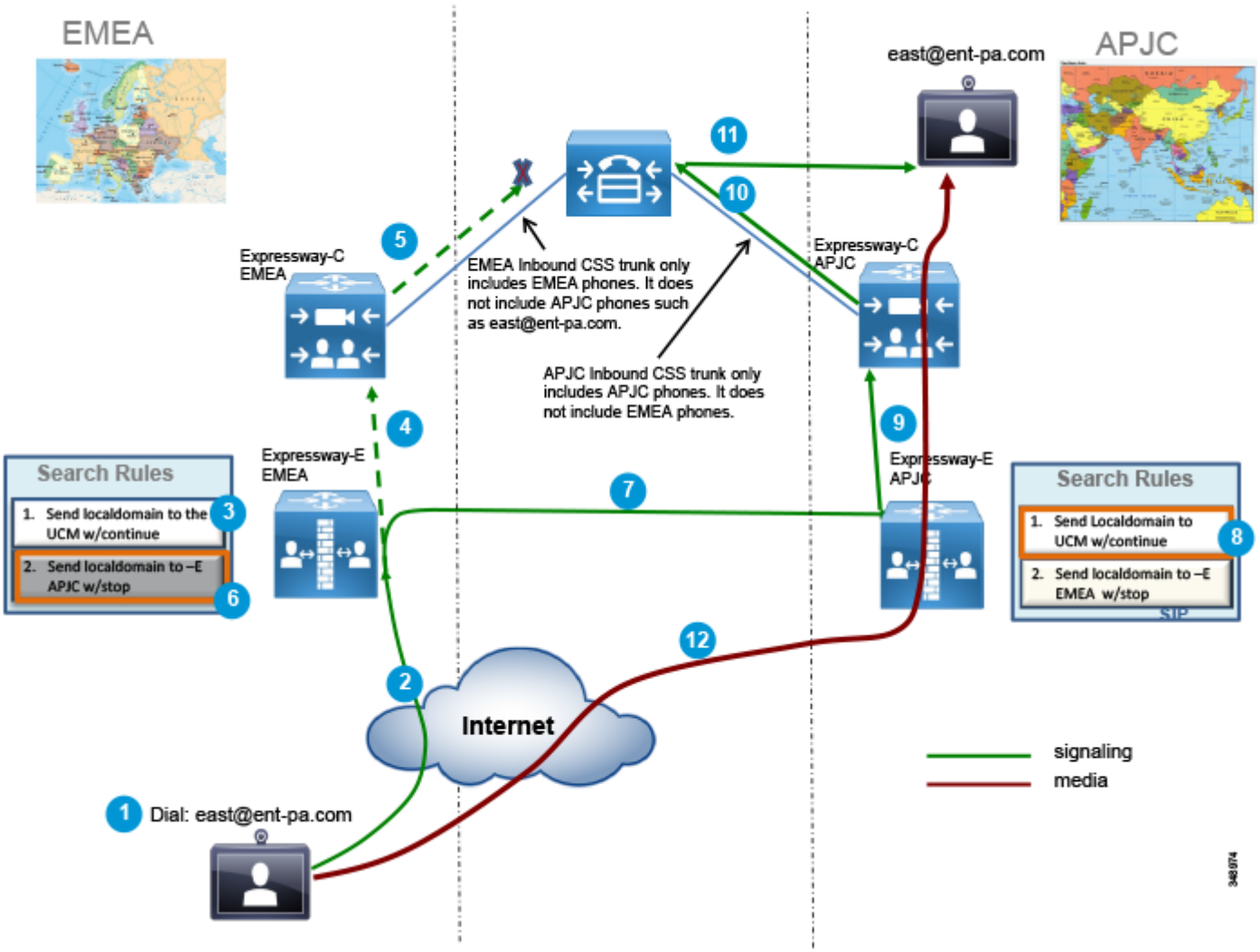
In order to allow the Expressway-E in EMEA to remove itself from the signaling and media path, it is important to make sure that there is no TCP-to-TLS or RTP-to-SRTP conversion on Expressway-E EMEA clusters, and to make sure that the call signaling optimization parameter is set to **on** in all Expressway-C and Expressway-E.

Because this is not a deterministic process, in the case of three or more Expressway edges the searching mechanism would require too much time. Therefore, this configuration is recommended for no more than two Expressway edges.

To scale to more that two edges, a different architecture called Directory Expressway can be deployed. Directory Expressway architecture is not part of the Preferred Architecture.

Figure 4-22 shows the Expressway edge design that enables selection of the edge closest to the destination endpoint.

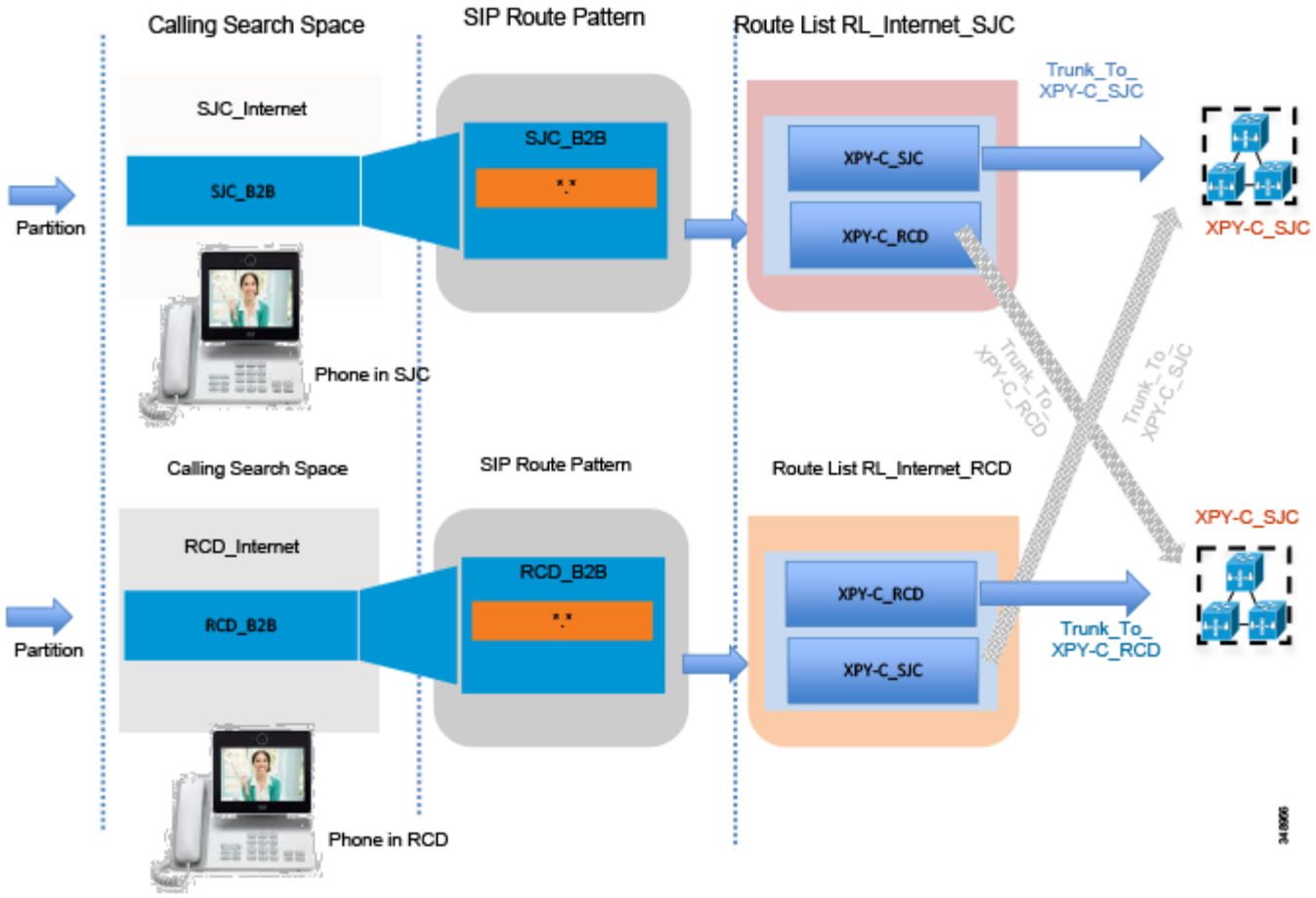
Figure 4-22 Selection of the Expressway Cluster Closest to the Destination



Considerations for Outbound Calls

Outbound calls should be directed to the Expressway-C that is nearest to the calling endpoint. This can be achieved by using Cisco Unified CM mechanisms such as calling search spaces and partitions. Figure 4-23 shows the Unified CM configuration.

Figure 4-23 Partitions and Calling Search Spaces Configured in Unified CM



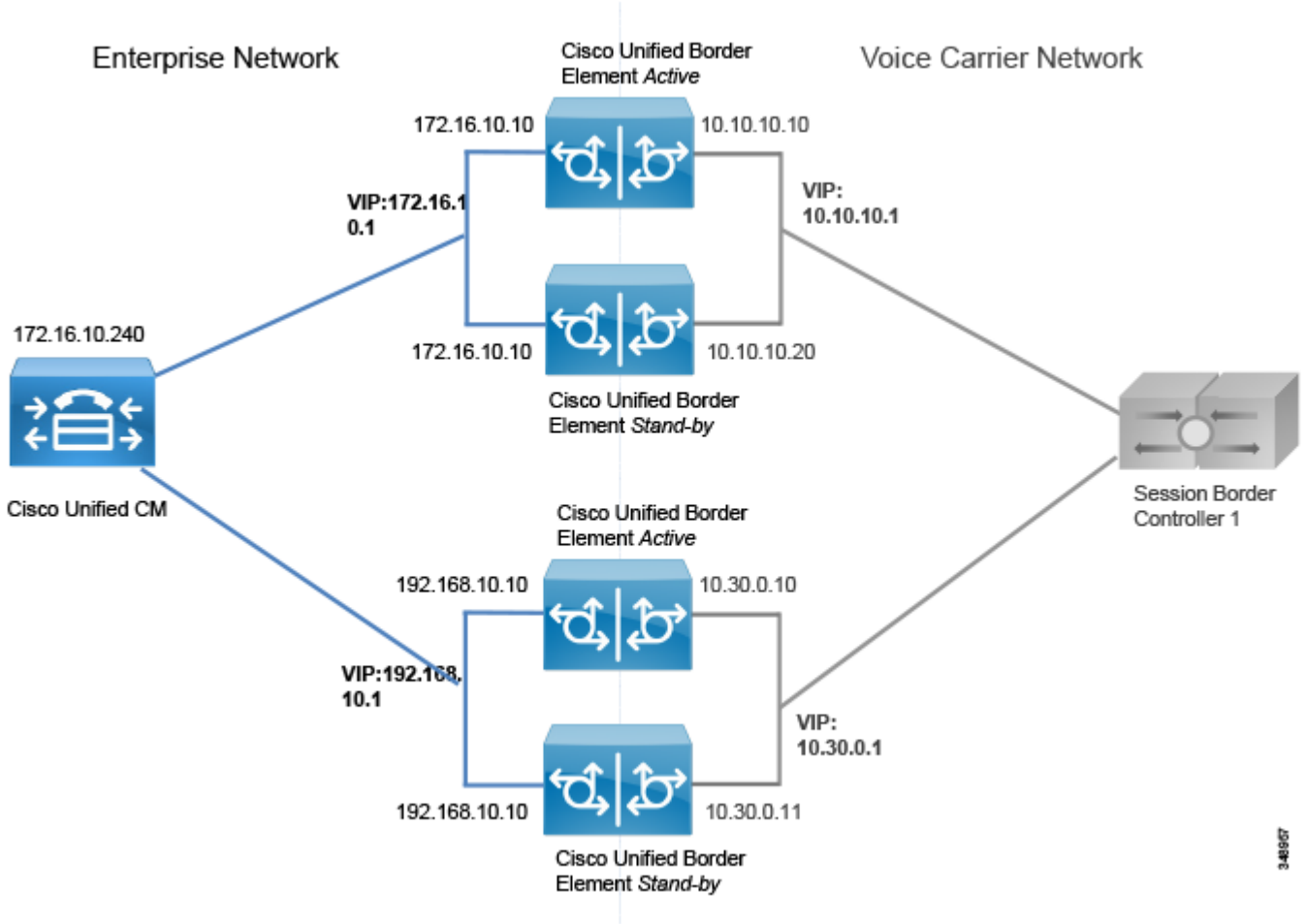
The Unified CM Local Route Group feature helps scale this solution when multiple sites access two or more Expressway-C clusters. This mechanism is also applied on ISDN gateways and Cisco Unified Border Element, and it is further described in the next section. A full description of the configuration is documented in the next two sections, since it also applies to Cisco Unified Border Element and voice gateways.

Scaling the Cisco Unified Border Element

For session capacities per platform, see the Sizing chapter.

If more than one datacenter is deployed, Cisco Unified Border Element can be deployed in each datacenter. This might happen for many reasons; for example, if it is required for a disaster recovery architecture, as shown in Figure 4-24.

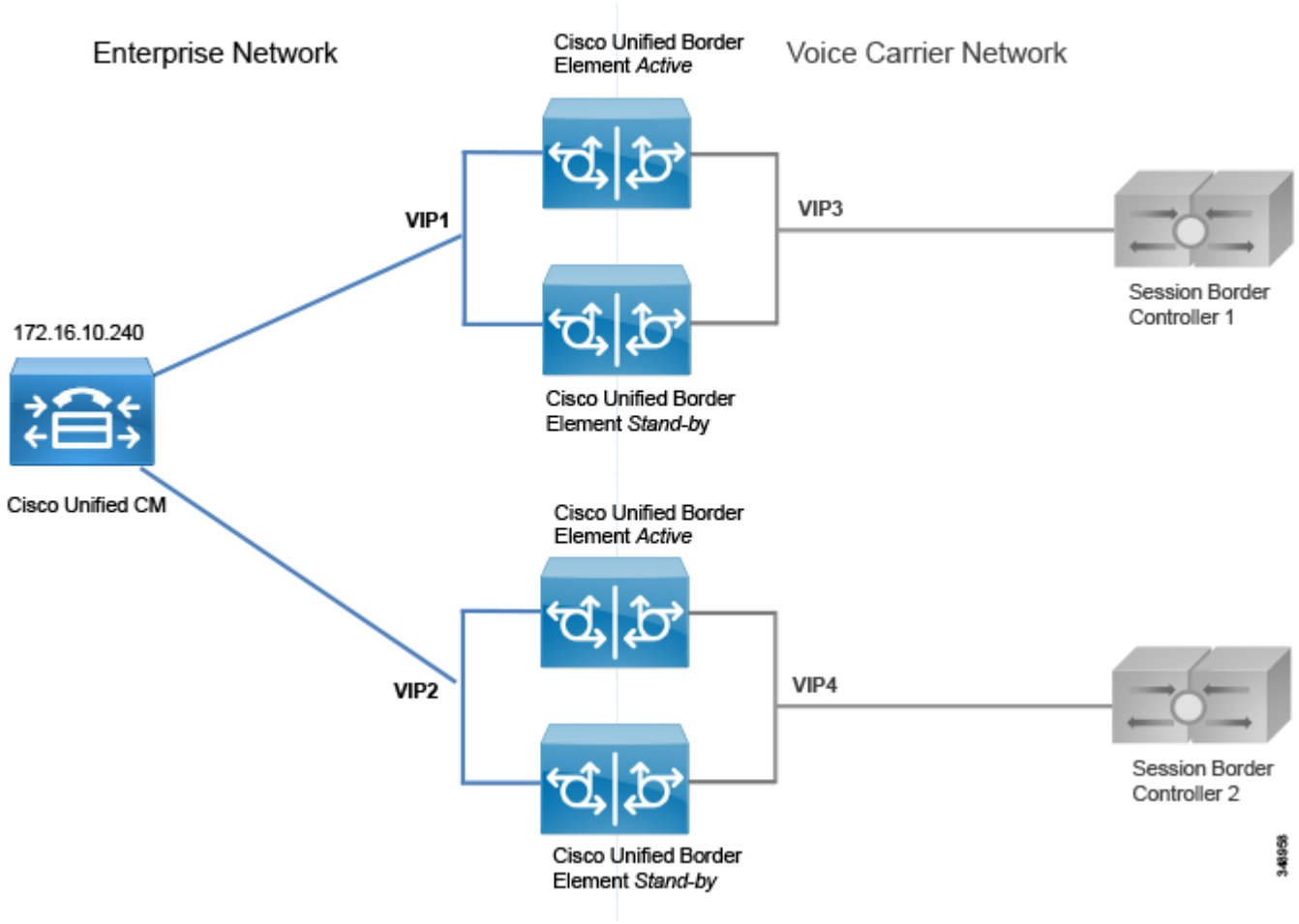
Figure 4-24 Multiple Cisco Unified Border Elements



All trunks to the Unified Border Element (usually two or three) can be inside the same route group. This would provide load balancing between datacenters. If the active router in the datacenter breaks, active calls will be preserved. If a datacenter becomes unreachable, call requests will be sent to the remaining datacenters. In this case, active calls would be dropped and users would have to reestablish them manually.

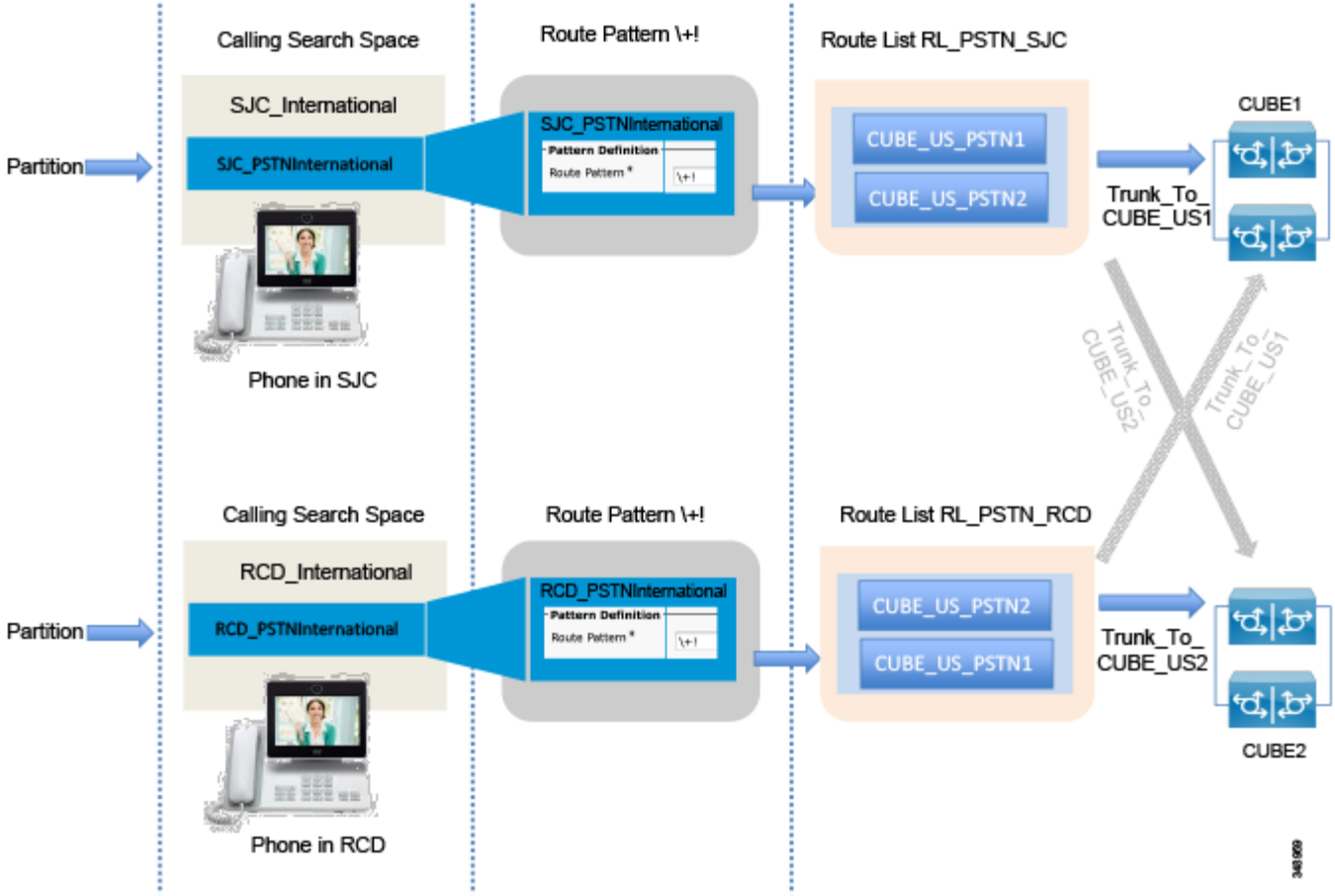
As shown in Figure 4-25, if the enterprise voice network is spread across a wide area, more than one session border controller (SBC) from the Telecom carrier is used. For each SBC, a Cisco Unified Border Element might be deployed, based on the carrier's recommendations.

Figure 4-25 Multiple Cisco Unified Border Element Connected to Different SBCs



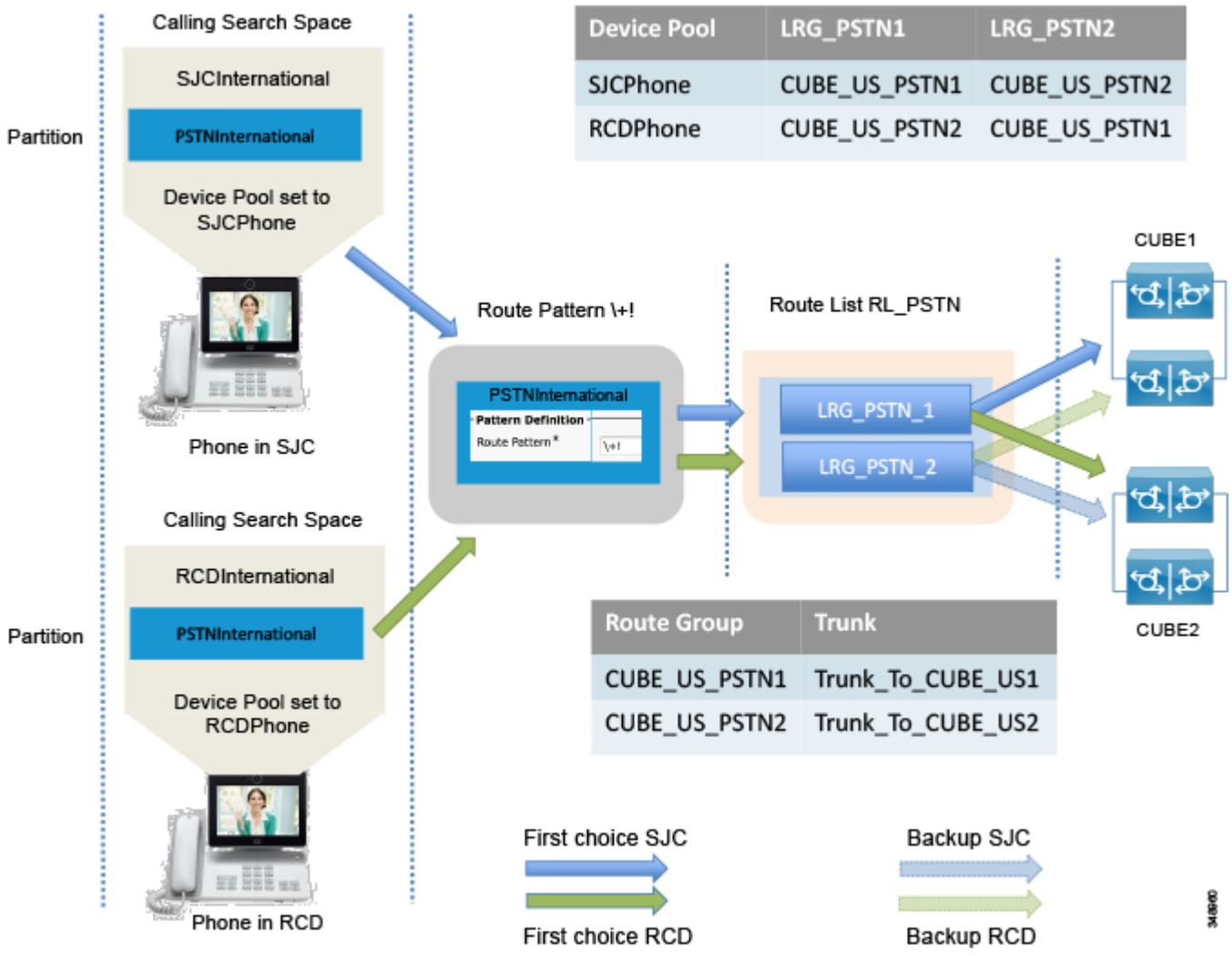
As an example, assume that another Unified Border Element is needed in the US besides the one already deployed. A new trunk called Trunk_to_CUBE_US2 is added. Figure 4-26 shows the configuration based on standard 1:1 mapping between calling search space and route pattern. This configuration has some limitations because, as the number of Unified Border Elements increases, it has a big impact on Unified CM resources. It is shown in Figure 4-26 in order to contrast this approach with the Local Route Group approach shown in Figure 4-27.

Figure 4-26 Unified CM Configuration for Cisco Unified Border Element Connection



The same route pattern, \+!, is repeated for every physical destination, and it resides in different partitions. The original partition PSTNInternational needs to be split into two, SJC_PSTNInternational and RCD_PSTNInternational, and the route pattern \+! has to be deleted and moved into the two newly created partitions. This approach works if the number of sites is not high, no more than two or three. A much better approach is to use the Local Route Group concept, as shown in Figure 4-27.

Figure 4-27 Unified CM Configuration for Cisco Unified Border Element Connection by Using the Local Route Group Approach



In this case, the device pool SCJPhone has LRG_PSTN1 set equal to the route group CUBE_US_PSTN1, while device pool RCDPhone has LRG_PSTN1 set equal to the route group CUBE_US_PSTN2. LRG_PSTN2 is set equal to CUBE_US_PSTN2 for SJC phones and equal to CUBE_US_PSTN1 for RCD phones. This approach is recommended because new partitions and route patterns are not required, and this approach is much more scalable than the approach shown in Figure 4-26.

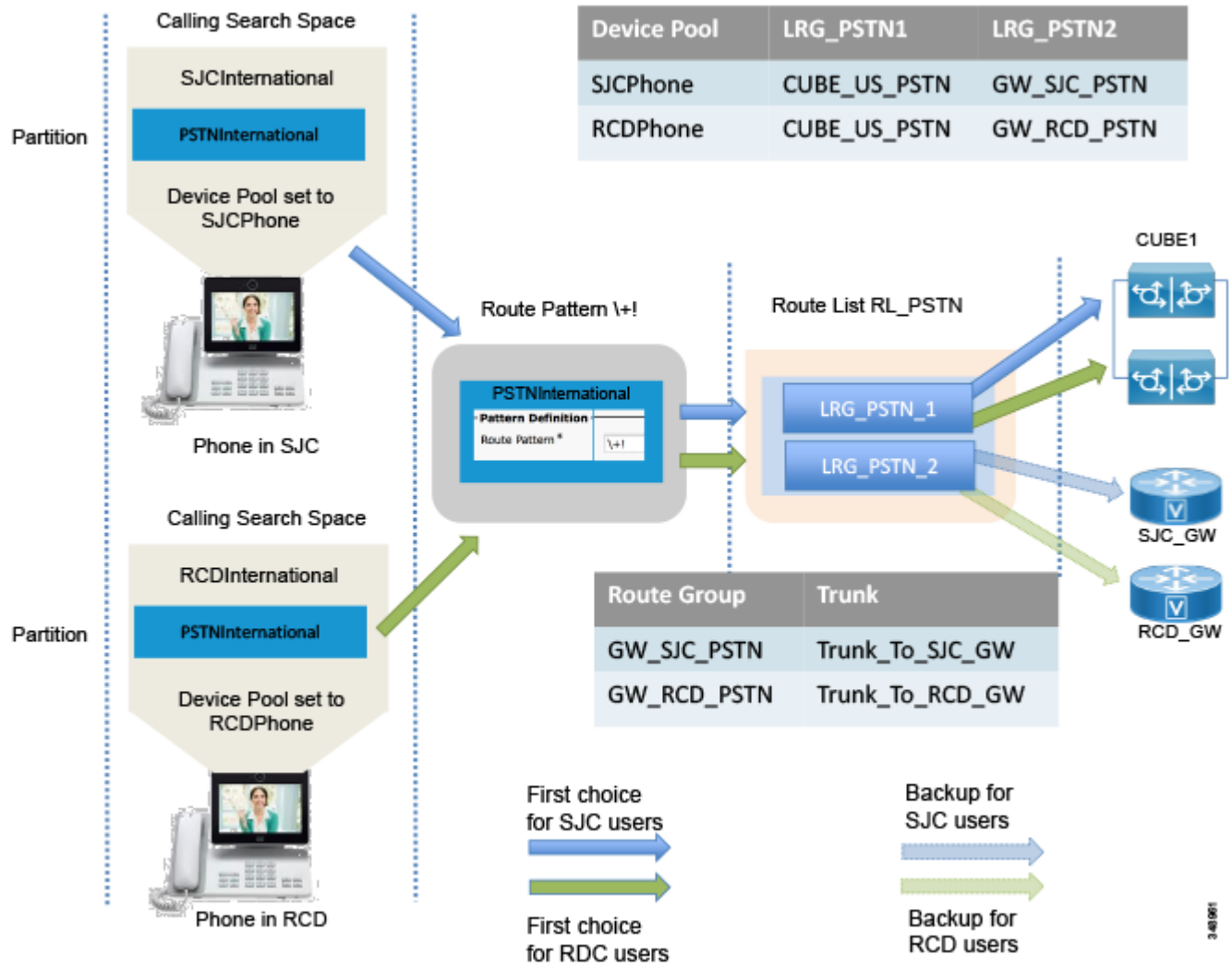
Scaling the PSTN Solution

Distributed gateways providing local PSTN access are deployed in branch offices and used as backup services.

If the number of branches is high, the route group and route list configuration construct within Unified CM does not scale well. For this deployment we recommend using the Local Route Group feature, so that route patterns to the PSTN do not have to be replicated for each site.

The configuration presented in the previous section is easily adapted to cover this scenario. What is needed is to assign the device profile LRG_PSTN1 to route group CUBE_US_PSTN, and to assign LRG_PSTN2 to the route group corresponding to the local gateway for that device pool, as shown in Figure 4-28.

Figure 4-28 Configuration for Centralized PSTN Access with Local ISDN Gateways



Scaling the Video ISDN Solution

Scaling the video ISDN solution is very similar to scaling the PSTN solution. If more than a few video ISDN gateways are required, using local route groups in Unified CM is the recommended method for routing calls out the PSTN. Having geographically dispersed ISDN gateways does aid in reducing toll charges for both inbound and outbound calls.

Collaboration Edge Deployment Process

This section summarizes the Collaboration Edge deployment process. Each component of Collaboration Edge is treated separately since each deployment may not require all access methods. As an example, a company might have only PSTN. Another company might use PSTN as a local backup for IP PSTN at specific local sites, have a Internet Edge deployment, and use ISDN video gateways to call those users who are not enabled for business-to-business Internet calls.

The Collaboration Edge components should be deployed in the following order:

- Deploy Expressway-C and Expressway-E
- Deploy Cisco Unified Border Element
- Deploy Cisco Voice Gateways
- Deploy Cisco ISDN Video Gateways

Deploy Expressway-C and Expressway-E

This section provides an overview of the tasks required to install and deploy Expressway-C and Expressway-E. The task should be performed in the following order:

1. Download and deploy Expressway-C and Expressway-E OVA templates and install the Expressway software. If the appliance model is used (Cisco Expressway CE500 or CE1000), there is no need to download and install OVA templates and the Expressway software.
2. Configure network interfaces and settings, including DNS and NTP, and system host name and domain name. Expressway-E has two LAN interfaces. If the external interface IP address is to be translated statically, the IP address of the translated interface has to be configured. Expressway-E will use the public IP address in payload references.
3. Configure clustering.

Deploy Mobile and Remote Access

1. Enable mobile and remote access by setting the Unified Communications mode to **Mobile and remote access**.
2. Select the domains for which mobile and remote access is enabled. Turn on **SIP registration and provisioning on Unified CM**, **IM and Presence service on Unified CM**, and **XMPP federation if inter-company federation**.
3. Upload the CA certificate to Expressway-C and Expressway-E. This is needed to discover Unified CM and IM and Presence clusters if **TLS verify mode** is **on** (recommended). This way Expressway-C will verify the identity of the cluster servers by checking the certificate.
4. Discover Unified CM and IM and Presence servers by configuring the publisher for each cluster.

5. Install certificates on both Expressway-C and Expressway-E. Both Expressway node types are able to generate a Certificate Signing Request (CSR) which is then signed by a CA. If an internal CA is use, the CSRs have to be signed by it. If Expressway-C and Expressway-E are also used for business-to-business communications, a public CA has to sign the certificate of Expressway-E, as mentioned previously. The signed certificates then need to be uploaded on Expressway-C and Expressway-E.
6. Configure a Unified Communication traversal zone between Expressway-C and Expressway-E, and allow for proxy registration to Cisco Unified CM.
7. To ensure that everything has been set up properly, check the Unified Communications status.

Note

- This configuration enables mobile and remote access. Business-to-business requires an additional configuration.
- The configuration above is done entirely on Expressway-C and Expressway-E.
- These steps are required for TCP/RTP connection to Unified CM (TLS/SRTP is not shown).

Deploy Business-to-Business Communications

This section provides an overview of the additional steps necessary to setup business-to-business communications.

1. Configure the basic Layer 3 configuration, including NTP, DNS, and system name, on both Expressway-C and Expressway-E.
2. Set up NAT configuration on Expressway-E, including IP routes necessary for routing traffic.
3. Ensure that the external firewall is set to block all traffic to Expressway-E before placing it in the DMZ.
4. Configure an administrative access policy, including local and/or remote authentication for both Expressway-C and Expressway-E.
5. Configure DNS A records in the appropriate DNS servers to be able to resolve the FQDN of each server.
6. Set up local authentication credentials in Expressway-E for the purpose of authenticating the traversal client connection coming from Expressway-C.
7. Set up the traversal server zone on Expressway-E for SIP only.
8. Set interworking on Expressway-E to **On**. This allows Expressway-E to send and receive H.323 calls and interwork them to SIP at the edge of the network, thus maintaining a single protocol inside the enterprise.
9. Set up the traversal client zone on Expressway-C for SIP only.
10. Use the FQDN of Expressway-E to enable the traversal link to allow for the possible use of PKI.
11. Configure the external DNS zone for outbound domain resolution for business-to-business communications.
12. Place basic CPL rules in place on Expressway-E to restrict access to protected resources such as video, voice, and IP PSTN gateways.
13. Set up domains for which Expressway-C and Expressway-E will have authorization.
14. Set up the dial plan on Expressway-C and Expressway-E with presearch transforms, search rules, DNS search rules, and external IP address routing.

15. Configure the SIP neighbor zone to Unified CM on Expressway-C.
16. Configure the SIP trunk on Unified CM to communicate with Expressway-C.

Deploy Cisco Unified Border Element

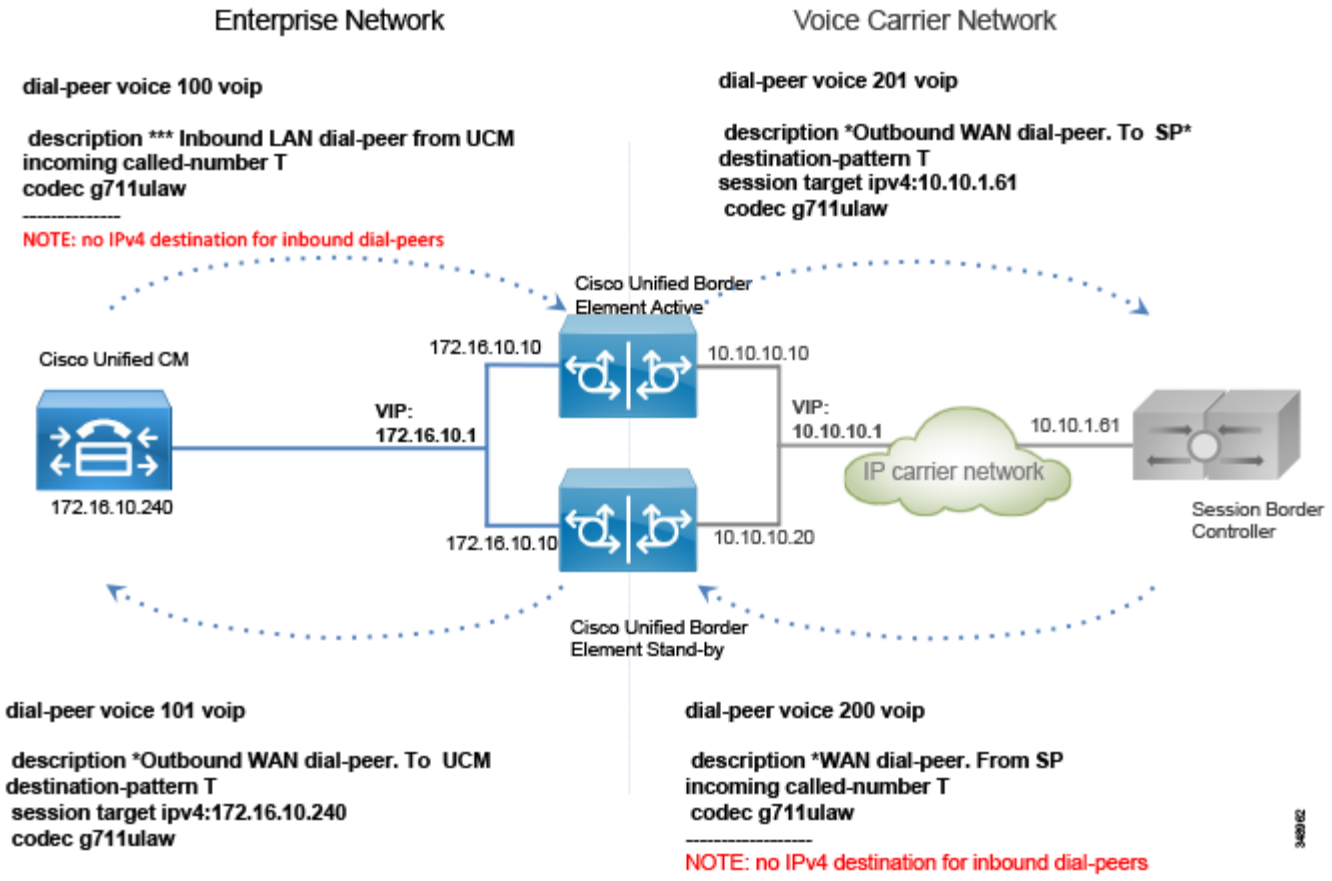
This section provides an overview of the process for deploying Cisco Unified Border Elements with box-to-box redundancy. Box-to-box redundancy has to be configured on both Unified Border Element routers, and the configuration is the same on both. It is possible to copy and paste the configuration from the active to the standby Unified Border Element.

1. Configure the network settings: the two Ethernet interfaces (one toward the LAN and the other facing the WAN) on both active and standby Unified Border Elements, as well as IP routing.
2. Enable the Unified Border Element on both routers for SIP-to-SIP calls, fax relay or passthrough, calling ID treatments as privacy headers, and enforcement of Early Offer. We recommend enabling this feature on Unified Border Element because Unified CM is configured for Best Effort Early Offer only. Although in new deployments only Early Offer will be sent from endpoints, there might be some cases involving old Cisco devices where Delayed Offer is sent instead. Even though these cases are not covered in this document, it is good practice to enforce Early Offer on Cisco Unified Border Element.
3. Enable box-to-box redundancy, and configure HSRP globally and on both the LAN and WAN interfaces for the active and standby routers.
4. Configure the voice codecs preference (in case the voice codec can be negotiated and it is not enforced by Unified CM or the Telecom carrier soft switch).
5. Configure music on hold.
6. Configure the dial-peers. Dial-peers are associated with call legs and can be matched inbound or outbound. As an example, an inbound call from Unified CM will be matched by an inbound dial-peer (corresponding to the inbound call leg). Another call leg will be generated by Cisco Unified Border Element (CUBE) toward the session border controller (SBC) of the Telecom carrier, and thus will be matched against another dial-peer. although the same dial-peer can match inbound or outbound calls, we recommend having each dial-peer match a specific call leg. Following this recommendation, there will be 4 distinct dial-peers: inbound dial-peer from Unified CM to CUBE, outbound dial-peer from CUBE to SBC, inbound dial-peer from SBC to CUBE, and outbound dial-peer from CUBE to Unified CM. Dial-peers can be matched against calling or called numbers or patterns. A dial-peer can force a single codec or can negotiate the list of codecs configured in step 4. The **incoming called-number** command makes a dial-peer inbound only.

   Inbound dial-peers do not have an associated target, while outbound dial-peers have Unified CM or the carrier's SBC defined as targets.

   Since calls to external destinations match generic patterns, dial-peer configuration on Unified Border Elements might lead to errors. As an example, in Figure 4-29 an outbound call matches both dial-peer 201 and 101, and therefore the routing does not work properly.

Figure 4-29 Inbound and Outbound Dial-Peer Configuration on Cisco Unified Border Element



The variable T in Figure 4-30 indicates any numeric string of any length, since calls from Unified CM might be sent to any destination in the world. A closest match might help, but when the Unified Border Element is centralized and provides the service for multiple locations, it might not be practical to list all the possible destinations in the "destination pattern" configuration. To overcome this limitation, and in order to simplify the routing process and make it more responsive, the following additional configuration is implemented:

a. Server groups in outbound dial-peers — If a server group is set as the destination in a dial-peer, and a round-robin algorithm is selected, the Unified Border Element will share the load between multiple servers:

```
voice class server-group 1
    ipv4 172.16.10.240
    ipv4 172.16.10.241
    ipv4 172.16.10.242
    ipv4 172.16.10.243
    ipv4 172.16.10.244
    hunt-scheme round-robin
```

b. SIP Out-of-Dialog OPTIONS Ping — It is possible to configure many parameters, such as the ping interval when a server is up and running, and the interval when it is down (set to 30 and 60 seconds respectively in this example):

```
voice class sip-options-keepalive 171
    transport tcp
    sip-profile 100
    down-interval 30
    up-interval 60
    retry 5
    description Target Unified CM
```

This way, the outbound dial-peer to Unified CM will be as follows:

```
dial-peer voice 101 voip
    description *Outbound WAN dial-peer. ToUnified CM
    destination-pattern T
    session protocol sipv2
    session server-group 1
    voice-class sip options-keepalive profile 171
    codec g711ulaw
```

c. The outbound call leg to the Telecom carrier will be matched by an outbound dial-peer:

```
dial-peer voice 201 voip
description *Outbound WAN dial-peer. To  SP*
destination-pattern T
session target ipv4:10.10.1.61
 codec g711ulaw
```

d. A leading "*" is sent by Unified CM on outbound calls (inbound calls from Unified Border Element's perspective), which enables the router to distinguish the direction of the call. This character must be eliminated before the call goes out to the IP PSTN. Further, according to the configured dial plan, the calling number has to be normalized with the "+". Rule 2 prefixes a "+" and is applied to the calling number, while rule 1 replaces the leading "*" with "+". The rules are applied to the called number. Two rules might be created for this, one for the called number and one for the calling number. However, since the called number always matches the first rule, and the calling number always matches the second rule, it is possible to use a single voice translation rule. This is configured on the inbound dial-peer.

The outbound call leg (dial-peer) is bound to the inbound dial-peer via the **dpg** command, so that if any call is received with a leading "*", it is sent to the dial-peer facing the SBC and not to the one going to Cisco Unified CM:

```
voice class dpg 201
    dial-peer 201

voice translation-rule 2
    rule 1 /^\*/ /+/
    rule 2 // /+/
voice translation-profile SIPtoE164
    translate called 2
    translate calling 2
dial-peer voice 100 voip
    translation-profile outgoing SIPtoE164
    incoming called-number *T
destination dpg 201
    codec g711
```

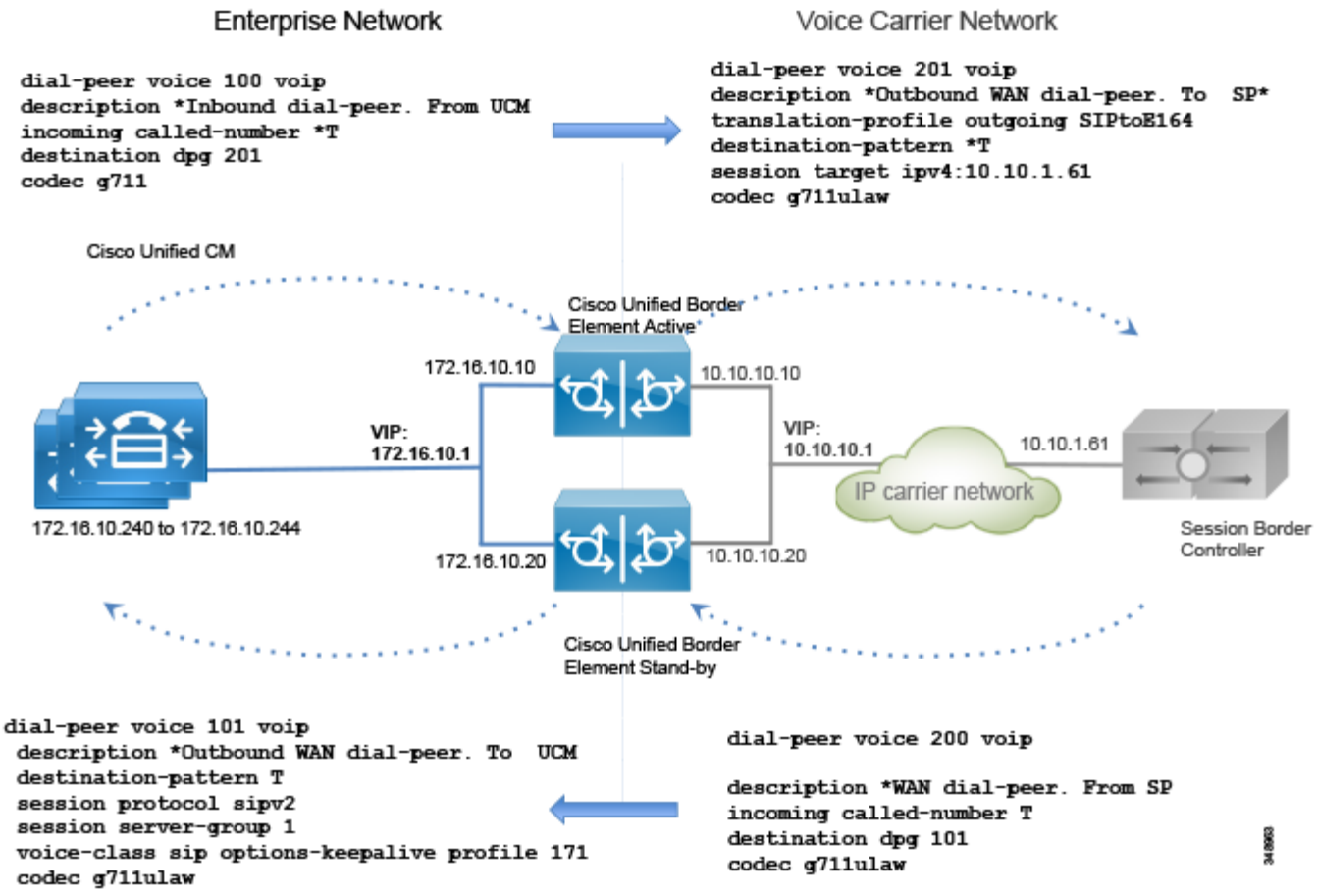
Dial-peer 200 needs also to be bound to dial-peer 101:

```
voice class dpg 101
    dial-peer 101

dial-peer voice 200 voip
    description *WAN dial-peer. From SP
    incoming called-number T
    destination dpg 101
    codec g711ulaw
```

Figure 4-30 shows this configuration.

Figure 4-30 Dial-Peer Configuration for Cisco Unified Border Element



If the call leg comes from Unified CM, it will hit the Unified Border Element with a leading "*", thus matching dial-peer 100. The call is then sent to dial-peer 200 by using the outbound dial-peer group as an inbound dial-peer destination. Dial-peer 200 removes the leading "*" and sends the call to the PSTN. Note that without this feature, dial-peer 201 would also be matched, resulting in routing errors.

If the call leg comes from the SBC, it might match dial-peers 201, 101, and 200. But since the "incoming called-number" takes precedence over the "destination pattern," dial-peer 200 will be matched; and since dial-peer 200 is linked to the dial-peer 101, the call is correctly routed to the destination.

7. Configure transcoding if required. Remember that transcoding requires dedicated hardware resources (DSPs).

Perform the following configuration tasks on Unified CM:

1. Configure a Best Effort Early Offer trunk for each Unified Border Element, as specified in the Call Control chapter.
2. Configure route group CUBE_US_PSTN and add the Unified Border Element trunk as a member.
3. Configure local route group LRG_PSTN1.
4. Configure a route list that includes the default local route group and the route group LRG_PSTN1.
5. For each device pool set LRG_PSTN1 to CUBE_US_PSTN.

Deploy Cisco Voice Gateways

PSTN interfaces are available across a wide range of routers, such as Cisco ISR G2/G3 and ASR routers. PSTN interfaces include analog, BRI, and PRI ISDN voice cards. Analog interfaces are used mostly to connect fax machines and analog telephones.

Perform the following tasks to configure a PSTN gateway with ISDN voice interfaces:

1. Configure network settings and routing on the router.
2. Activate the ISDN interface.
3. Set the ISDN parameters for user side, switch-type, framing, and linecode, based on the Telecom carrier's requirements.
4. Configure the dial-peers.

The dial-peer logic is the same as for the IP PSTN and Unified Border Element, but in this case besides the "voip" dial-peers, a voice gateway also has "pots" dial-peers toward the PSTN.

If there are analog devices such as fax machines, they can be connected to the router through an analog interface.

If the router is used only for analog fax interconnection and with the PSTN interfaces attached to another router, T.38 fax-relay can be configured since it provides for better resiliency, especially if the path to the PSTN gateway traverses the WAN.

The dial-peer configuration is different from IP PSTN and the Unified Border Element configuration. Since the gateway is deployed within a specific location and serves phones for that location, the pattern destination is well known, as for example +14085554XXX.

On the other side, an incoming PSTN call has an address that is composed of plan, type, and number. Plan and type are not supported in SIP, and based on the Telecom carrier, the call can reach the gateway with a different plan and type. As an example, for a call to E.164 destination 4961007739764 on a trunk in Germany in the same area code 6100, the called party number in the outgoing ISDN SETUP message could be sent as (plan/type/number) ISDN/national/61007739764, ISDN/subscriber/7739764, or unknown/unknown/061007739764.

Based on the plan/type, the number changes, and thus the dial-peers might not match. For this reason it is necessary to force the plan/type to unknown/unknown. This way the full E164 number will be released to the destination. Dial-peer structure is described in detail in the Call Control chapter and is referenced here for consistency.

For outbound dial-peers, this rule transforms any calling party number to plan “unknown” and type “unknown”, and it transforms the called party number with the leading "*" to the +E.164 number.

```
voice translation-rule 1
        rule 1 /^\*/ // type any unknown plan any unknown
        rule 2 // // type any unknown plan any unknown
voice translation-profile ISDNunknown
        translate called 1
translate calling 1
dial-peer voice 1 pots
        translation-profile outgoing ISDNunknown
```

For inbound dial-peers, if the calling party information has a 10-digit number with type “national” (and does not include the country code "1" for US), the call will be transformed correctly to the +E.164 number, prepending "+1". If it is "unknown" the following rules will not be matched.

If the called number comes from an international destination, and thus it has the country code and is in the E.164 format, then rule 2 will add the leading "+".

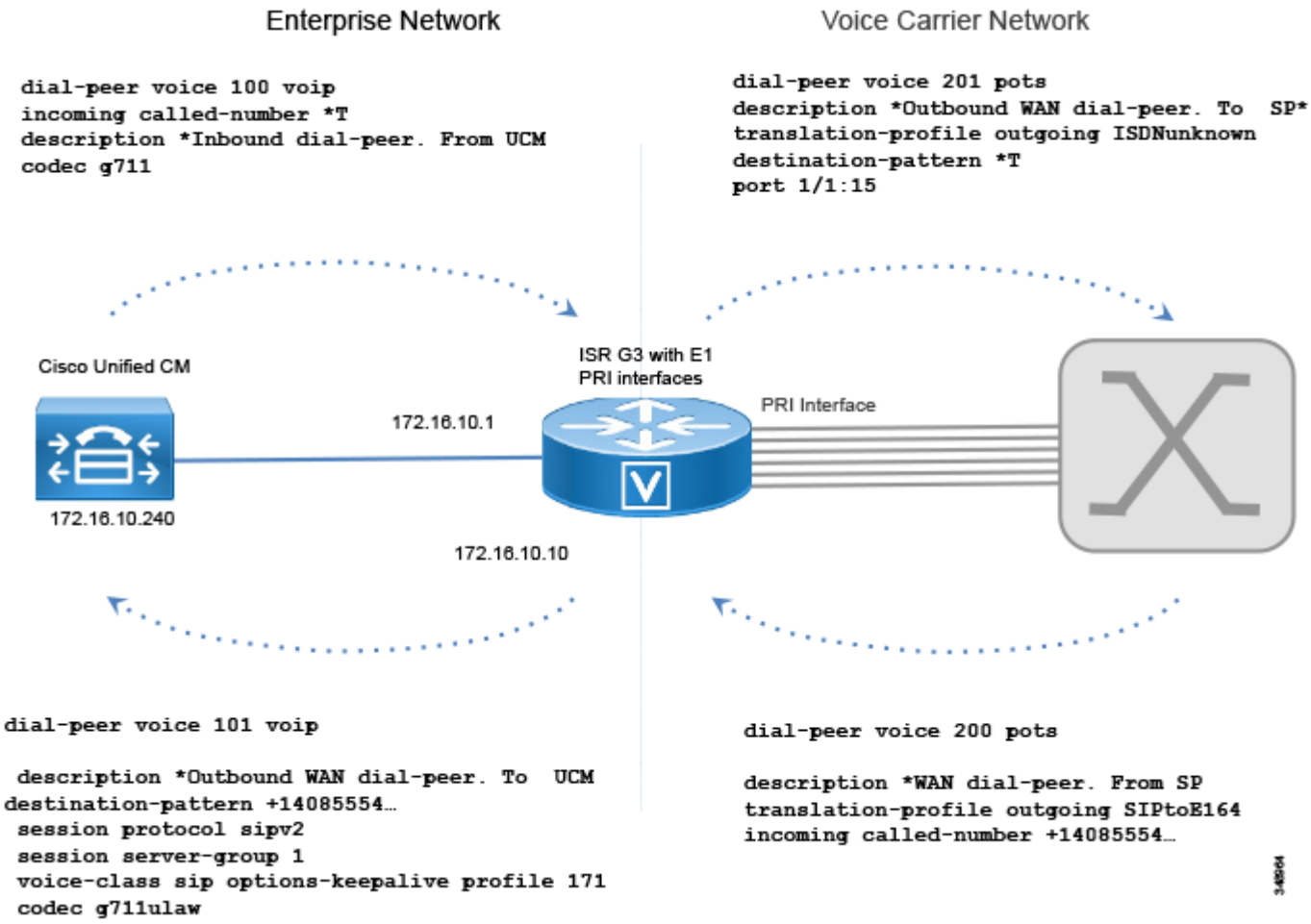
However, since ISDN setup is hop-by-hop, we are not expecting to see many calls with type "national" since the latest switch might force it to "national". In any case, these rules normalize the calling and called party numbers correctly.

```
voice translation-rule 3
    rule 1 /^\(.+\)$/ /+1\1/ type national unknown plan any unknown
    rule 2 /^\(.+\)$/ /+\1/ type international unknown plan any unknown
voice translation-profile ISDNtoE164
    translate called 3
    translate calling 3

dial-peer voice 1 pots
    translation-profile incoming ISDNtoE164
```

Figure 4-31 shows a dial-peer configuration for G.711 and the E1 PRI interface.

Figure 4-31 Dial-Peer Configuration for Voice Gateways



Perform the following configuration tasks on Unified CM:

1. Configure a Best Effort Early Offer trunk for each gateway (Trunk_to_SiteID_GW, SiteID is a variable that identifies the location).
2. Configure route group LRG_PSTN1 and include the gateway trunk as member.
3. Configure local route group LRG_PSTN1.
4. Configure a route list that includes the default local route group and LRG_PSTN1.
5. For each device pool, set LRG_PSTN1 to Trunk_to_SiteID_GW. This configuration assumes, as recommended, that for each site there is a device pool SiteIDPhone.

By using the local route group configuration, it is easy to reconfigure PSTN access. As an example, it is possible to use the Unified Border Element for centralized access to the PSTN and to use the local PSTN connection as backup. In this case, the device pool would specify the Unified Border Element route group as LRG_PSTN1, and LRG_PSTN2 will include the trunk to the local gateway (Trunk_to_SiteID_GW).

Deploy Cisco ISDN Video Gateways

Deployment of a Cisco TelePresence ISDN GW 3241 or a Cisco TelePresence ISDN MSE 8321 is a fairly straightforward process:

1. Log into the web interface.
2. Allocate port licenses. Each port license activates a PRI interface. Port licenses are configured on the supervisor MSE 8050 for the 8321 ISDN gateway, and port licenses are configured on-box for the ISDN GW 3241.
3. Set up the ISDN interface. This is accomplished under **Settings > ISDN**. These settings are the typical settings received from the service provider for the type and form of ISDN being delivered.
4. Configure the ISDN ports. This is accomplished under **Settings > ISDN ports**. Directory number, channel range, and channel search order are all configured here. The **Enabled** box must be checked here for each port to enable the ISDN port for use.
5. Configure call control. This is accomplished under **Settings > SIP**. These are the SIP settings on the ISDN gateway that include the hostname and SIP domain used for Unified CM.
6. Configure the dial plan. This is accomplished on two tabs under the dial plan heading: **IP to ISDN** and **ISDN to IP**. Ensure that the incoming ISDN number range is translated correctly to the IP number range on the **ISDN to IP** dial plan tab.

For more information on the gateway installation and initial configuration, refer to the Cisco TelePresence ISDN Gateway installation and upgrade guides, available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-isdn-gateway/products-installation-guides-list.html