COMMUNICATIONS TRANSFORMATIONS: Implementation Considerations when Enhancing Enterprise Communications Solutions with SIP Trunks

Contents

Overview

Communications is an essential part of any enterprise. Over the past 10 years Cisco® has been instrumental in transforming the methods enterprises use to enhance communications. Unified communications using voice and video over IP are widely deployed in enterprise networks, both for internal calling in the campus as well as between branch offices of the sites comprising the WAN. Such enterprise networks increasingly use unified communications IP-capable user endpoints such as IP phones or softphones, along with unified communications gateways to provide a point

of interconnection to older time-division multiplexing (TDM) private branch exchanges (PBXs). All these steps work toward implementing a communications system that is entirely IP-based.

One of the points of network interconnection where IP access is now becoming available is to the public switched telephone network (PSTN). Increasingly, service providers are starting to offer an IP method of connecting branches and enterprise communications systems to the PSTN. These systems are generally referred to as “SIP trunk services” because the Session Initial Protocol (SIP) is used as the signaling method between the service provider and the enterprise communications systems. SIP trunk access is being provided for both local and long-distance PSTN access. These SIP trunks allow for unified communications IP access to the PSTN from the enterprise as well as the promise of new applications and services with end-to-end unified communications IP calling between enterprise users and external users.

Most enterprises add unified communications SIP trunk access to their networks in addition to the traditional TDM connectivity already in place. Cisco solutions offer enterprises the unique capability of upgrading their existing equipment that provides TDM connectivity to also support IP connectivity to the service provider. This connectivity provides access to new services for their users while at the same time minimizing reconfiguration in the network by keeping existing call patterns in place as well as providing backup access for emergency calling and redundancy.

When deciding whether, or when, to add a unified communications SIP trunk entry point to your network, be sure to consider more than the cost and services of the offering. This decision also has implications for call routing, high availability, Call Admission Control (CAC), security, and various other aspects of network design. The evaluation steps to consider when adding a unified communications SIP trunk to your network are covered in the white paper “Communications Transformation: Integrating a SIP Trunk into the Enterprise”. This paper discusses the technical considerations of this integration.

SIP Trunk Access Beyond the Enterprise

Traditional PSTN access from the enterprise network is accomplished with a voice gateway located at each site. The voice gateway supports TDM trunk connectivity to the local PSTN central office. This trunk connectivity may be analog or digital, depending on the size of the branch office and the service available from the central office. Calls to and from the PSTN are routed through the gateway and converted between TDM and IP by the gateway.

Call routing from the enterprise to the PSTN is based on two models:

- Distributed model: All calls from the local site to the PSTN use the local PSTN gateway.
- Centralized model: Most calls from all sites are routed to a central site through unified communications IP and use a single, shared PSTN gateway, while local PSTN lines are used for emergency calling.

These two models offer several variations, where local calls can use the local PSTN gateway at a site while long-distance calls from the same site use unified communications IP to a central-site shared PSTN gateway. Considering the two extreme cases, however, facilitates evaluating how unified communications SIP trunk access can change call patterns in your network.

Adding unified communications SIP trunk access to your network may imply changes to both call routing and the placement of SIP trunk access points in the network. When determining where to connect a SIP trunk in your network, consider the following:

- What types of calls are carried by the SIP trunk? Long-distance calls only, or also “local” calls?

  If local calls are included in the unified communications SIP trunk offering from the provider, are the calls “local” only to the site where the trunk physically terminates into your network, or also for “local” calls to geographically distant sites, or a subset of sites, in your network? The answer to this question determines dial-plan changes and high-availability considerations for call routing. Each geographic region has a different definition of local and long-distance calls, and you need to consider local requirements for each site.

- Is the cost of the service offering based on a single or multiple physical entry points?

  If the provider offers a single physical connection, then the best choice is most likely to place this connection at your head office (campus) or data center. If the offering includes multiple physical entry points (within the cost bounds of the service you are willing to consider implementing), then having connectivity from multiple sites may make sense after the initial deployment of the centralized unified communications SIP trunk for both call-routing purposes as well as redundancy and backup purposes.

  The number of multiple entry points may be small, in which case you would likely connect your major sites, or the largest site within each geographic region, to the service. If a large number of remote sites have restricted IP bandwidth, then you should consider connecting a unified communications SIP trunk into each branch office to lower the number of calls and bandwidth crossing your WAN to the head-office site. This determination is similar to the way traditional PSTN connectivity is determined.

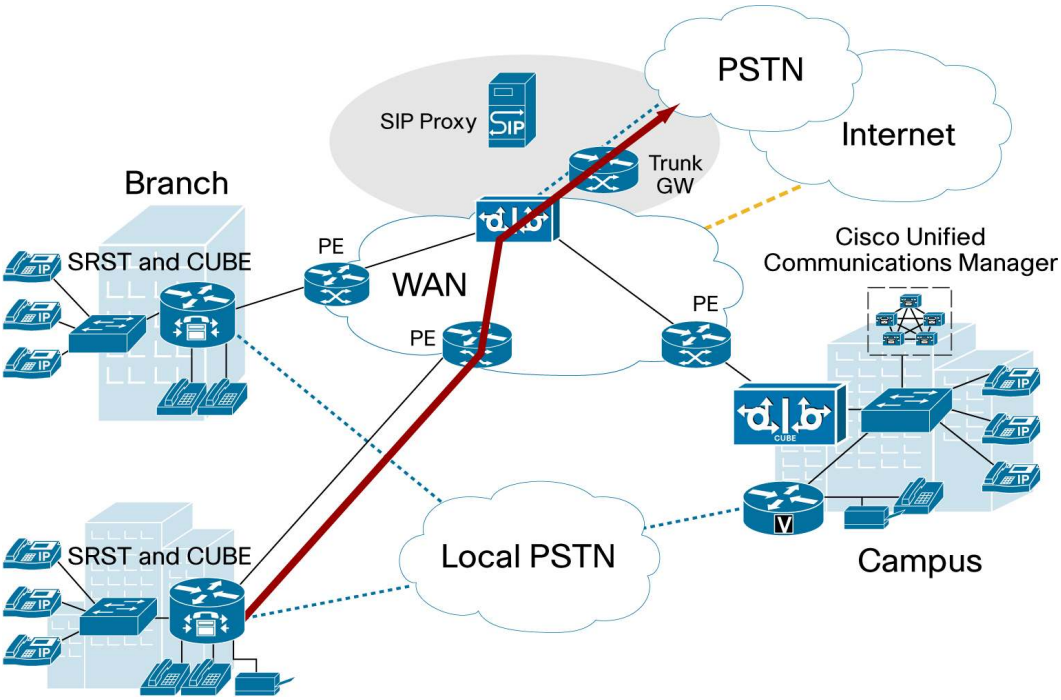
- Does the service offering provide international call access?

  If your network spans multiple countries, you may require a different unified communications SIP trunk entry point, and these entry points possibly from different service providers within each region or country; for example, one in the United States and at least one in Europe. This determination can be important for both regulatory requirements as well as network efficiency.

Distributed Model

In a distributed model, each site has a unified communications SIP trunk to the service provider for local and long-distance access, and TDM trunks are used for high availability, redundancy, and emergency calling. Figure 1 shows an example of distributed access. In this model a Cisco Unified Border Element at each site connects calls directly to the SIP trunk service provider’s border element. If you choose to manage and provision your own internal calling, then the site-to-site calls can be made using VPN or Dynamic Multipoint VPN (DMVPN) connections between sites. The other alternative is to have a service provider manage and provision site-to-site calling as well.

Figure 1. Distributed SIP Trunk Access in the Enterprise



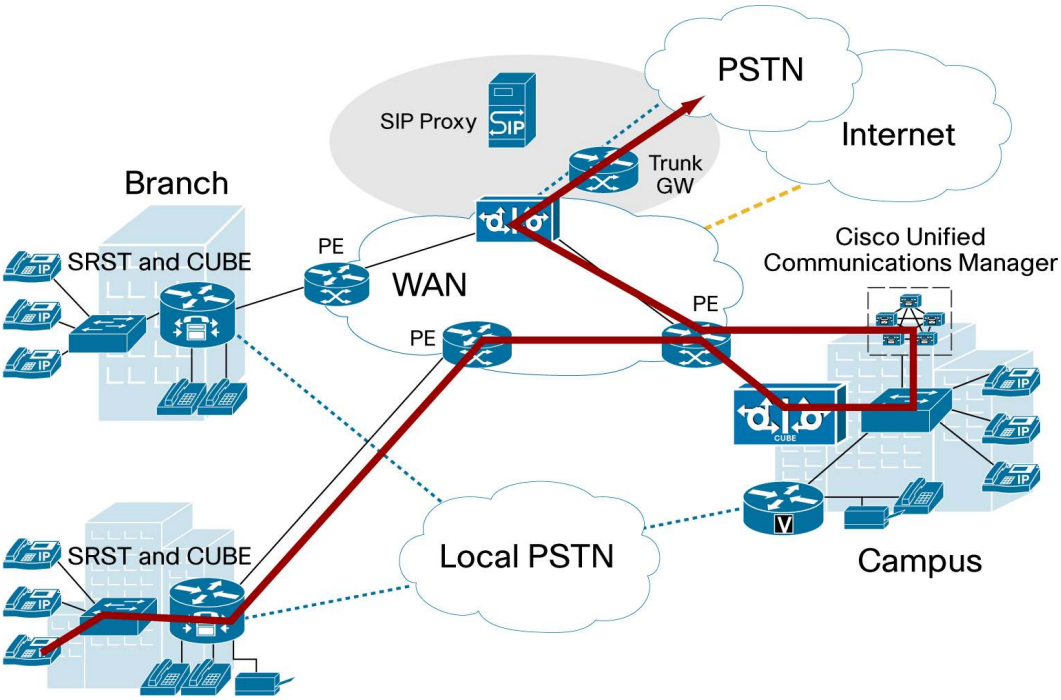
The advantages and disadvantages of the distributed model include:

- Cisco provides a highly available set of hardware for the branch office with the Cisco 2800 and 3800 Series Integrated Services Routers.
- Cisco provides a highly available solution at the branch office with a Cisco integrated services router that integrates components such as Survivable Remote Site Telephony (SRST), Cisco Unified Border Element, and TDM functions in a single chassis.
- Bandwidth requirements between branch sites and the headquarters sites are lower because all calls do not have to egress at the central site.
- The emergency (that is, E911, etc.) location can be tied to the branch-office Cisco Unified Border Element instead of the headquarters border element.
- The solution requires provisioning of a Cisco Unified Border Element at each site and, therefore, the service provider has to accept calls from multiple IP addresses.
- The service provider needs to have a redundancy solution that allows incoming calls to be routed to a PSTN number that is local to the remote site.

Centralized Model

The most common unified communications SIP trunk deployment model currently is the centralized model with a single entry point at the head office or campus site. In this model all extranet voice calls are routed through the central site before they are handed off to the service provider's network, as shown in Figure 2.

Figure 2. Centralized SIP Trunk Access Beyond the Enterpris



The advantages and disadvantages of the centralized model include:

- Provisioning is simplified, with centralized access to the PSTN.
- Dial-plan management is simplified.
- Implementation is simplified: Initially unified communications SIP trunk access is available only to users residing at the campus location, while other sites in the network continue to use traditional PSTN gateways until unified communications SIP trunk access is made available to those sites at a later date.
- Failover must be provided at each site with SRST and TDM PSTN ports.
- Bandwidth consumption increases across the WAN and at the headquarters site because it has to terminate and reroute external calls for all sites in the network.

Much of the decision of where to locate the unified communications SIP trunk(s) from the enterprise network depends on the service offering structure and pricing. These service offerings are still evolving and current models may not prevail, or may mature into different options than are now available.

Cisco recommends that you adopt the solution in stages, as follows:

- Implement a trial at the campus with the addition of a unified communications SIP trunk for a group of users and long-distance traffic.
- Install the trial such that the unified communications SIP trunk is the preferred egress for external calls, and the traditional PSTN gateways provide backup, local call, and emergency call access.
- Phase in branch offices to start routing their long-distance calls through the unified communications SIP trunk.
- Add local calls from the campus location to the traffic using the unified communications SIP trunk.

You should evaluate the benefits and costs at each stage of the deployment.

IP Connectivity Options to the Service Provider

Unified communications SIP trunks are being offered by many different types of service providers. Enterprises will see offers from service providers to transport just their data services or just their voice services. It is important to note that only service providers that have complete control over the quality of service (QoS) can offer business-class voice services. “Over-the-top” voice-over-IP (VoIP) services that ride on non-QoS enabled networks cannot provide guarantees of quality levels. When both voice and data services are acquired from the same provider, this provider may install new physical interfaces dedicated to the VoIP traffic or, more commonly, will share the physical infrastructure of the data traffic to take advantage of using a single pipe to support both voice and data. However, in some cases, for security reasons, the enterprise may prefer a separate physical interface.

The configuration of a unified communications SIP trunk requires coordination between the enterprise to configure its Cisco Unified Border Element and the service provider’s border element before starting to exchange SIP traffic. This configuration is normally done after a data service is established. For the data service, the service provider can offer multiple data-link connections. Table 1 lists some of these options.

Table 1. Network Connectivity Options

Physical Connection	Data Link
Fast Ethernet or Gigabit Ethernet	Metro Ethernet
Broadband interface **Cisco Cable Modem High-Speed WAN Interface Card (WIC; HWIC-CABLE), Cisco Asymmetric DSL WIC (WIC1-ADSL), and Cisco Symmetrical High-Data-Rate WIC (WIC1-SHDSL)**	Cable modem, DSL, and asymmetric DSL (ADSL)
T1/E1 **Cisco 1-Port T1 CSU/DSU WAN Interface Card (WIC-1DSU-T1] and Cisco 1- and 2-Port T1/E1 Multiflex Trunk Voice/WAN Interface Card (VWIC-1MFT-T1, VWIC-2MFT-T1, VWIC-1MFT-E1, VWIC-2MFT-E1, VWIC2-1MFT-T1/E1, and VWIC2-2MFT-T1/E1)**	Point-to-Point Protocol (PPP), Frame Relay, and ATM
Wireless 802.11 **Cisco Third-Generation Wireless WAN High-Speed WIC (HWIC-3G)**	Ethernet

When using a single connection for both voice and data, you should carefully consider congestion management and bandwidth allocation to prevent data traffic from affecting the voice quality (QoS is discussed in a later section).

Some service providers that offer both data and voice over a single IP interface also offer Multiprotocol Label Switching (MPLS) services and require that voice be sent with an MPLS label. This setup enables the service provider to terminate voice traffic, while data traffic marked with a different label can be tunneled through the network; it requires a Virtual Route Forwarding (VRF)-aware voice feature that will soon be available on the integrated service routers.

Unified communications SIP trunks require that the enterprise have a demarcation point for functions such as topology hiding, call accounting, call-quality reporting, and call redirects to multiple providers. The Cisco Unified Border Element offers these functions.

Call Traffic Capacity and Bandwidth Control

Managing simultaneous voice call capacity and IP bandwidth usage is essential for providing consistent quality in enterprise communications. Unified communications IP gateways connecting to the PSTN through a TDM interface provide an implicit form of CAC by virtue of the limited number of channels (or time slots) physically connected between the gateway and the central

office. The number of channels on a TDM gateway is usually carefully engineered to provide just the desired call capacity to the user community at a particular grade of service (blocking factor).

Although TDM capacity engineering is typically targeted at provisioning calls outbound to the PSTN, it also provides implicit traffic management in the reverse direction. No more calls can simultaneously arrive from the PSTN into the enterprise network than there are time slots available on the gateway TDM trunks. This automatic time slot-based CAC function is no longer available on unified communications SIP trunks where there is no concept of time slots, and physical connectivity (typically Ethernet) allows for an unlimited number of sessions. Therefore, you should carefully consider this fact when provisioning unified communications SIP trunk capacity and planning for CAC.

Trunk Provisioning

The capacity of a unified communications SIP trunk is normally defined by the number of simultaneous calls supported and the bandwidth provided for the trunk. An enterprise uses the same Erlang calculations traditionally used in a TDM environment to determine the number of simultaneous calls required on a unified communications SIP trunk.

Generally service providers offer a tiered service based on capacity. One of the major benefits of unified communications SIP trunks is that as an enterprise's needs expand, the number of simultaneous calls can be readily expanded without changing the physical interconnection, or even without an increase in provisioned bandwidth, provided excess bandwidth is already available.

Bandwidth Consumption

Bandwidth consumption for unified communications call traffic inbound from the PSTN on a TDM gateway is easily predicted and controlled because the codec assignment is done by the gateway (or by the enterprise call agent such as Cisco Unified Communications Manager). The use of a Cisco Unified Border Element can ensure that this capability is maintained when an enterprise adds a unified communications SIP trunk to its communications infrastructure.

CAC policies and features are deployed in the enterprise network based on predictable patterns of codec use by calls (that is, G.711 for calls within a site on the LAN, and G.729A for calls that traverse the WAN between sites). The bandwidth consumption of inbound unified communications SIP trunk calls is based on the service provider's configuration, but an enterprise can use a Cisco Unified Border Element to ensure that a service provider does not exceed the expected incoming bandwidth (service-level agreement [SLA]), ensuring that devices internal to the enterprise are not susceptible to problems at the service provider's network.

Another technique you can use to manage bandwidth is transcoding, which should be provided as a network service to translate calls using undesired codecs into an appropriate codec for the enterprise network. A new codec that is used increasingly often is the Internet Low Bitrate Codec (iLBC), which consumes less bandwidth and provides superior quality in networks with packet loss.

Refer to the "Voice over IP – Per Call Bandwidth Consumption" document and the "Voice Codec Bandwidth Calculator" tool for calculations for codec bandwidth requirements.

Call Admission Control

When unified communications IP calls enter an enterprise network on a SIP trunk, there is no direct mechanism for the enterprise network designer to control the codec - and therefore the bandwidth - used by each call. The SIP INVITE that arrives from the service provider includes the codec that the calling endpoint and service provider configuration has chosen. The enterprise can

use a Cisco Unified Border Element with codec filtering as a demarcation point to ensure that only valid codecs are passed toward the enterprise communications network.

One general problem with CAC is that many policies are based on simple “call-counting” mechanisms (such as the Cisco Unified Communications Manager Locations CAC feature) as opposed to bandwidth-based mechanisms (such as Resource Reservation Protocol [RSVP]). It is, therefore, important to control not only the number of calls arriving through the unified communications SIP trunk, but also the codec assigned to the calls.

To control codec selection, and therefore bandwidth consumption, you can also use a transcoding device at the edge of your network. A Cisco Unified Border Element can also provide this function. In addition to transcoding and codec filtering, a Cisco Unified Border Element can support the CAC policy of the enterprise in the following two ways:

- Limiting calls per dial peer: This simple call-counting mechanism controls the maximum number of simultaneous calls entering the enterprise network.
- Limiting calls based on memory and CPU used at the Cisco Unified Border Element: This overload protection mechanism helps ensure your network is not overrun with a burst of unmanageable call traffic.

Limiting Calls per Dial Peer

You can configure the max-conn command on both the inbound and outbound dial peers of the Cisco Unified Border Element to ensure that no more than the configured number of calls is connected at one time. Each call, regardless of codec, is considered a call and is counted.

When a call arrives at a dial peer and the current number of calls in the connected state exceeds the configured amount, the SIP INVITE request is rejected with a result code to indicate that the gateway is out of resources.

Global CAC Based on CPU and Memory

When configured, the average CPU usage or available memory of the Cisco Unified Border Element is checked before it completes the processing of a SIP INVITE request. If the resources exceed the configured amount, then the Cisco Unified Border Element returns a result code in the SIP INVITE request, indicating that the gateway is out of resources.

Quality of Service

Cisco provides many methods of measuring and ensuring QoS in an enterprise IP network. You should always use these methods internally when designing a unified communications system, and you should also extend them to the interconnect point when using IP to connect to service provider services.

Delay and Jitter

The telephone industry standard specified in ITU-T G.114 recommends the maximum desired one-way delay be no more than 150 milliseconds. With a round-trip delay of 300 milliseconds or more, users may experience annoying talk-over effects. In addition to congestion management with proper queuing techniques, you can use fragmentation and interleaving for slower links to ensure that the delay budget for voice packets is met. Variable delay in packet rate results in jitter. The jitter buffer in Cisco IOS Software voice gateways runs in an adaptive mode and de-jitters the packet flow in case of moderate jitter. Please refer to the “Understanding Jitter in Packet Voice

Networks" document for more information about jitter, including troubleshooting. Delay can also cause echo, as described in the next section.

When using unified communications SIP trunks as opposed to TDM trunks for PSTN access, you should now consider the IP delay of both the enterprise and service provider networks. In some cases, centralized unified communications SIP trunk services cannot be effectively deployed because of the resulting increase in latency. A demarcation device at the customer premises is required to ensure that latency in the service provider network and enterprise network can be independently measured and controlled.

Echo

An echo is the audible leak-through of your own voice into your own receive (return) path. The source of echo can be located by understanding the nature of the echo and its parameters, such as loudness, delay, and duration. Acoustic echo can come from improper acoustic insulation on the phone, headset, or speakerphone (all Cisco IP phones have an acoustic echo canceller). A delayed echo can result from a long IP leg, which could be from the PSTN connectivity in the service provider's network. Refer to the "Echo Analysis for Voice over IP" document for details about troubleshooting the source of echo. Cisco Unified voice gateways and endpoints are G.168-compliant for echo cancellation.

A demarcation point at a customer site can help you determine if a problem with echo is occurring at the customer premises or in the service provider's network.

Congestion Management

Unified communications IP signaling and media traffic can be identified and classified as priority traffic by QoS mechanisms available within Cisco IOS Software. You should use Low-Latency Queuing (LLQ) for media traffic streams. To use LLQ you need to know how many simultaneous calls will be made over the unified communications SIP trunk, the codec, and payload size to determine the bandwidth to allocate for voice. Your enterprise can determine the number of calls based on the CAC techniques discussed earlier.

During congestion LLQ traffic is restricted to the allocated bandwidth. Therefore signaling traffic should use Class-Based Weighted Fair Queuing (CBWFQ), because it comes as a burst at the time of call setup and teardown. Refer to the "Quality of Service for Voice over IP" document for further information. When using a conventional Layer 2 connection such as Frame Relay or ATM, you must deploy additional traffic shaping and traffic management mechanisms.

Setting of proper differentiated services code point (DSCP) values on the media and signaling packets leaving the unified communications SIP trunk is important to receive proper service in the provider's network. By default the media packets are marked with DSCP EF (101110) and signaling packets are marked with DSCP AF31 (011010) when they are processed by the VoIP dial peer on the router. These markings can be overridden if needed by the outgoing dial peer on the Cisco Unified Border Element at the demarcation point before entering the service provider's network.

Voice-Quality Monitoring

To ensure acceptable voice quality within the enterprise network and to determine if a service provider is meeting the agreed-upon SLA, your enterprise should monitor some metrics. Each enterprise may choose to monitor different metrics, but an effective method of collecting the

metrics independent from the service provider is important. Table 2 describes some of the important metrics you can monitor.

Table 2. Voice-Quality Monitoring Attributes

Metric	Goal	Definition	Method to Monitor
Round-trip delay (RTD)	100–300 ms	The RTD is the delay for a packet sent from the originating endpoint at the customer location to the terminating endpoint at the service provider and back again.	This metric can be monitored through the RTD metric in Cisco IOS Software. This metric is provided per call and is also available through IP-SLA probes.
Jitter	50–100 ms	Jitter is a measurement of the change in the delay of one packet to another during a call.	Jitter is measured in the per-call statistics; the maximum jitter detected during the call is recorded.
Packet loss	1 percent or lower	Packet loss is the number of packets lost during any given call, including User Datagram Protocol (UDP) and TCP packets.	This metric can be monitored by Simple Network Management Protocol (SNMP) in Cisco IOS Software. It is provided per call and can also be tracked with IP-SLA probes.
Uptime	99.999 percent	Uptime is the percentage of time that a path is available for the customer to complete a call to the PSTN.	When uptime is measured, planned outages should be accounted for, and it should be measured as the number of unplanned minutes of outages and monitored with trouble tickets.
Answer seizure rate (ASR) or call success rate (CSR)	Varies	The ASR can be recorded as the number of calls made divided by the number of calls that complete a voice path. This number varies greatly because of calling numbers that are unassigned or busy. CSR is the percentage of calls successfully completed through a service provider. The CSR rate should be more than 99 percent. The ASR rate is typically around 60 percent.	ASR can be measured by a summary of call activity at the end of the month. The specific value of ASR is not as important as whether there are large swings in the ASR from one month to another that may indicate a problem with end-to-end network connectivity.

High Availability

Unified communications SIP trunk service to an enterprise network is commonly deployed with a single entry point. The physical connection of a unified communications SIP trunk is not traffic-limited to a fixed number of calls in the same way as a TDM PSTN gateway is, so it is easier to have a single SIP trunk serve a much larger user community than it is with a TDM gateway. This characteristic of unified communications SIP trunk access may raise the effect of a failure on your network and the business continuity within your enterprise.

You can deploy several strategies to protect against the business effect of a unified communications SIP trunk failure:

- Continue to use your TDM PSTN gateways in addition to using a unified communications SIP trunk to your network. Configure call routing to use the unified communications SIP trunk as the primary method of access, and the TDM gateways as secondary. You can use the same physical Cisco platform for both functions so that adding a unified communications SIP trunk to your “PSTN” gateway does not mean adding equipment to the network.

- Have multiple physical entry points or unified communications SIP trunks into your network and use load balancing or failover between them as necessary.
- Use the unified communications SIP trunk only for call routing from the site collocated with the trunk, and continue to use TDM gateways (or separate SIP trunks) at other sites so that IP connectivity failure in either the service provider's or your WAN does not affect users at more than one site.
- Deploy redundant hardware (Hot Standby Router Protocol [HSRP]) for the Cisco Unified Border Element routers that terminate the unified communications SIP trunk so that transparent hardware failover is possible while maintaining a single unified communications SIP trunk (that is, a single visible IP address) to the service provider.

Additional high-availability considerations involve Domain Name System (DNS) configuration and registration with the provider's SIP registrar. When the SIP registrar is defined using DNS, then the DNS server provides multiple IP addresses if the primary SIP proxy fails. Post Dial Delay (PDD) can be reduced by setting the number of retries to 2 (the default is 6) with the command retry-invite 2 in the sip-ua configuration on the Cisco Unified Border Element. With the number of retries set to 2 and a default retry interval of 500 msec, you can maintain a PDD lower than 2 seconds during failover from a primary to a backup SIP proxy. You can further reduce this delay by shortening the retry interval (however, use caution when changing this parameter, and consider the response times from a proxy in its busiest hour).

The configuration of the SIP registration interval timer and the number of retry attempts determine the time required for failover. Although tuning these parameters shortens failover time (that is, PDD), it also results in increased bandwidth use because registration messages are sent more frequently. This configuration is done on the SIP registrar, not on the Cisco Unified Border Element.

The Cisco Unified Border Element realizes that the primary SIP proxy has failed when either a registration times out or an outbound SIP message receives no response. At this time it tries to register with the backup SIP proxy.

Protocol and Media Interworking

When all external (PSTN) calls to the enterprise were TDM and only internal calls used unified communications IP, the enterprise network designer was in complete control of:

- Which IP user endpoints communicated with each other, and the compatibility of capabilities among these endpoints
- The H.323 and SIP protocols and vintages used between network elements
- The methods of interworking such as dual tone multifrequency (DTMF), fax, and codec packetization

When external (PSTN) calls arrive as unified communications IP calls, the enterprise no longer has direct control over any of these network design elements. The protocol variations and media encoding methods of externally originated unified communications IP calls (such as the codec or DTMF-relay method chosen) may not be implemented on the enterprise network components, or they may be undesirable because they violate CAC policies or feature interoperability. This situation should be resolved both contractually with the SIP trunk service provider and with technology by having an interworking function at the edge of the enterprise network.

Protocols

Unified communications IP access from a service provider almost always offers a SIP interface, although in rare cases this interface may also be H.323. Many enterprise networks have not yet migrated to SIP internally, and those who have typically have done so first on the “trunk” side of the call agent and not yet on the “line” side, that is, to all user endpoints. The variation of SIP used by the enterprise call agent’s SIP trunk may also not be of the same vintage, or compatible in message use, with the service provider’s SIP offering for all call flows. For example, there are variations in the use of the SIP REFER or REINVITE messages for modifying calls during a transfer. The existing TDM PSTN gateways in the enterprise network are also much more likely to be deployed using H.323 or Media Gateway Control Protocol (MGCP) than SIP, and although many networks have plans to migrate these to SIP, it is not an overnight network transition to accomplish this transition throughout the enterprise network.

Interconnecting these different unified communications IP protocols, vintages, and implementation variations requires a border element at the edge of your network. The Cisco Unified Border Element performs protocol translation (for example, SIP to H.323) as well as dealing with minor SIP message variations between implementations of different providers’ SIP trunk offerings. The Cisco Unified Border Element masks off these incompatibilities from the rest of your enterprise network and affords a one-time compatibility testing or certification cycle, instead of testing or certifying every protocol element (call agents, conference servers, IP phones, softphones, voicemail servers, and TDM gateways) in your network to be interoperable with the service provider’s network.

Using a Cisco Unified Border Element to interconnect to the SIP trunk also allows taking advantage of SIP trunk offerings immediately, without requiring the rest of your network to be upgraded, or to be SIP-capable. The migration to SIP of the entire enterprise network can therefore be done in a timeline suitable to your business needs, while at the same time you can start making available unified communications SIP trunk services and benefits to your user community.

Media

The speech path, or Real-Time Transport Protocol (RTP) stream, of a unified communications IP call typically flows directly between the communicating endpoints, requiring that the endpoints support a common set of capabilities for voice or video encoding such as codecs, DTMF digits, and fax transmission. These streams can be encoded in myriad ways, and it is likely that your enterprise endpoints support only a subset of these methods, or that you wish to deploy only a subset of them to ensure that the interconnect policies of your network are met. For example, you may have standardized on the use of only G.711 and G.729A in your network to ensure CAC is adequately calculated.

The SIP trunk from a provider may not offer the capabilities you have standardized in your network, and the external IP endpoints communicating through the unified communications SIP trunk to enterprise endpoints may use encodings that are neither desirable nor compatible.

A back-to-back user agent (offered by the Cisco Unified Border Element) at the edge of your network terminating the external SIP session and re-originating an internal SIP (or H.323) session can mask these compatibilities from your internal network. DTMF-relay methods (such as in-band DTMF often used on SIP trunks) can be translated by the Cisco Unified Border Element to RFC 2833 DTMF relay common in enterprise SIP networks. Similarly, codec choices can be controlled. The Cisco Unified Border Element can perform transcoding (changing one codec to another), or

codec filtering, ensuring only certain valid codecs based on your enterprise policies are negotiated with external IP endpoints.

Unified Communications IP Endpoints

SIP-capable IP endpoints often have varying capabilities, so even if your network already deploys SIP user endpoints, it is still often undesirable to have these endpoints interact directly with external endpoint across the unified communications SIP trunk. Many SIP messages encode IP addressing inside various SIP-specific header fields that are not changed with traditional NAT services. There are also variations in how SIP messaging is implemented for particular features, and testing or certifying each of your endpoints with the provider's SIP implementation could be an arduous task.

Supplementary Services

Cisco Unified Communication deployments offer a very rich set of supplementary services. With the use of a Cisco Unified Border Element, you can maintain these services even when a unified communications SIP trunk brings external calls into your enterprise.

Voice Calls

Basic supplementary services such as call hold, call transfer, call waiting, three-way conferencing, distinctive alerting, calling line identification (caller ID), calling name, and call toggle can be provided by Cisco Unified Communications Manager Express and Cisco Unified Communications Manager for IP phones and by a voice gateway for analog phones. The analog phones use Hookflash to trigger the supplementary services. Cisco Unified Communications Manager Express and Cisco Unified Communications Manager offer a wide variety of supplementary services for Skinny Client Control Protocol (SCCP) and SIP user endpoints in addition to those listed previously. Additional examples include shared line, hunt groups, music on hold, presence, and so on. These features can be used by external calls arriving over the unified communications SIP trunk.

IP Centrex or Class 5 type features (for example, call forwarding, call screening, call park, call return, and so on) can be provided by central SIP servers resident in the service provider's network. An analog phone in the enterprise can trigger these features with access codes (typically starting with an asterisk) provided by the service provider. Cisco Unified voice gateways send the access codes in the SIP INVITE message over the SIP trunk to the service provider to trigger these features.

Voicemail

You can provide voicemail within the enterprise network with Cisco Unity® Express, the Cisco Unity system, or Cisco Unity Connection. You also can choose to get voicemail services from the service provider. Message waiting indicator (MWI) is visual light on IP phones; it is indicated by a stutter dial tone on analog phones. Cisco Unified voice gateways can relay the MWI over the SIP trunk to the enterprise endpoints. For DTMF interworking needed for voicemail, refer to the "Media" section in the preceding "Interworking" section.

Transcoding

Transcoding can be provided by local digital signal processors (DSPs) for a call that uses a high-bandwidth codec such as G.711 in the LAN and a low-bandwidth codec such as G.729 or iLBC

over the WAN. These transcoding resources can exist in the Cisco Unified Border Element as described in the “Media” section in the preceding “Interworking” section.

Conferencing

Providing conferencing resources within the enterprise consumes less bandwidth than conferencing resources provided by the service provider. In the case where conferencing services are hosted by the service provider, each RTP stream participating in the conference resource consumes a two-way unicast bandwidth to and from the conference bridge.

A local conference bridge can be set up using DSPs within the site or enterprise to save bandwidth. The conference bridge can be controlled by Cisco Unified Communications Manager or Cisco Unified Communications Manager Express and can be co-resident in the Cisco Unified Border Element platform. These local conferencing resources can also be accessed by external callers over unified communications SIP trunks.

Modem and Fax Traffic

The ability to send modem and fax calls over unified communications SIP trunks is an important consideration for both the service provider and enterprise. Enterprises that cannot achieve adequate fax and modem call success over unified communications SIP trunks should maintain these services over their traditional TDM PSTN trunks.

Fax calls terminated by a SIP trunk service provider may have slightly lower speeds than TDM trunks, and they may support fewer native fax protocols. For example, Super Group 3 (SG3) fax and color faxing cannot be completed over unified communications SIP trunks at present. Cisco Unified voice gateways provide mechanisms to transmit SG3 fax calls over SIP trunks using G3. The calls start as voice calls; when the terminating gateway detects the fax or modem tones, the fax or modem method can be communicated with a signaling protocol or in an RTP stream. For details about configuration, refer to the “Fax Configuration Guide”.

T.38 as a Fax Method for SIP Trunks to the PSTN

T.38 is a standard for implementing faxing over an IP network. Cisco IOS Software implements the 1998 T.38 fax standard. For T.38 fax calls to work, both the terminating and originating endpoints must support T.38 faxing.

T.38 fax is more resilient to delay, jitter, and loss on an IP network than a fax passthrough method. The T.38 configuration can replicate both the low-speed V.21 and the high-speed T4 traffic generated with a fax. The replication of packets can aid in providing a higher fax success rate when using unified communications SIP trunks.

A T.38 capability is indicated in the SIP Session Description Protocol (SDP), and the negotiation can occur through a REINVITE request or by using a named signaling event (NSE) in the RTP stream. The NSE option is supported by Cisco Unified voice gateways.

Fax Passthrough as a Fax Method for SIP Trunks to the PSTN

Fax passthrough involves the terminating and originating gateways changing their codec to G.711, and fixing the jitter buffer at a recommended 200 ms to send fax information in-band in the speech path. Fax passthrough consumes more bandwidth than T.38 Fax Relay, and is generally used only to ease interoperability to non-T.38 capable endpoints. After the fax tone is detected, the upspeed request to the G.711 codec for fax passthrough mode can be signaled using a SIP REINVITE request or using NSE. The NSE option is supported by Cisco Unified voice gateways.

Because the recommended best practice is to use T.38 Fax Relay, this feature should be a basic one offered by unified communications SIP trunk providers.

Store-and-Forward Fax Support

T.37 Store-and-Forward Fax is supported on all Cisco Unified voice gateways. On-ramp T.37 is the process of accepting a fax call, encoding that fax into a TIFF file, and sending that TIFF file to an e-mail server as an attachment. Refer to the “T.37 Onramp Fax” document for further information. OffRamp T.37 is the process by which a gateway accepts an e-mail, converts that e-mail to a fax, and then sends it out through a voice port to a TDM fax machine. Refer to the “T.37 Offramp Fax” document for further information.

Modem Passthrough as a Modem Method for SIP Trunks to the PSTN

Modem calls are less and less prevalent as many data communications systems move to using pure IP. However, modems are still used in many retail locations for credit card validation systems (point of sale) and across enterprises for monitoring and triggering security and alarm systems. The main disadvantage of carrying a modem call over an IP network is the absence of a TDM clock, and therefore high-speed 56k connections cannot be achieved. Connection speeds of 28.8k are achievable when the IP network has no jitter or packet loss.

Modem connections send a 2100-KHz tone, which is detected by the voice gateway at the beginning of a call. After it is detected, the gateway sends an NSE as an RTP packet to the other gateway endpoint. This NSE packet causes the gateway to switch into modem passthrough mode, meaning that it fixes the jitter buffer at 200 ms, disables the echo canceller, and switches to the G.711 codec. The call proceeds after this switchover has occurred.

Modem Relay as a Modem Method for SIP Trunks to the PSTN

This method involves interpreting the modem signals by the originating gateway and recreating them on the far side by the terminating gateway. Modem relay provides better resiliency than modem passthrough to network delay, jitter, and loss problems. Cisco Unified voice gateways indicate support for this capability through NSE in the RTP stream to indicate a modem call to the far-end gateway.

Dial Plans and Call Routing

Adding a unified communications SIP trunk service to your network most likely means that there are service changes (numbers accessible and their associated cost) and that you should optimize call routing in your network for the most cost-efficient calling patterns. This optimization, in turn, may affect current CAC and bandwidth-allocation policies implemented in your network.

Some specific items that may affect your dial-plan and call-routing configuration include:

- If you currently have a separate dedicated TDM PSTN voice gateway per Cisco Unified Communications Manager cluster, and now you have a single enterprisewide unified communications SIP trunk shared between them
- If the unified communications SIP trunk offers only long-distance (or certain types of inter-regional) calls, whereas your TDM PSTN gateways offered both local and long-distance calls
- Whether the unified communications SIP trunk is going to be used by all the users in your network (all sites) or only by users co-located at the site where the SIP trunk terminates

- Routing of emergency calls: Discuss these types of calls with your SIP trunk provider, because they may have to continue using traditional PSTN gateways temporarily in order to be routed correctly to emergency centers.

Certain service providers require that a + be added to the front of a phone number sent on a unified communications SIP trunk. Specifically, the From field must be valid, as in From: +14085551212. When interconnecting through Cisco Unified Communications Manager, this configuration can be accomplished by using translation rules on a Cisco Unified Border Element between Cisco Unified Communications Manager and the SIP trunk service provider.

Certain SIP trunk providers require users to complete a registration before they can use unified communications SIP trunk service. This security practice is a good one for service providers to ensure that calls originate only from well-known endpoints. Cisco Unified Communications Manager does not natively support registration on SIP trunks, but this support can also be accomplished by using a Cisco Unified Border Element. The Cisco Unified Border Element registers to the service provider with the phone numbers of the enterprise on behalf of Cisco Unified Communications Manager.

Security

A TDM PSTN gateway provides an explicit demarcation point between your network and the service provider. Because a TDM-to-IP conversion is done at this demarcation point, there are few security concerns surrounding a malicious external user traversing this network interconnection into the enterprise. Getting data access into your IP network through traditional voice TDM trunks on the PSTN is technically virtually impossible, and security concerns about voice access mainly surround protection against toll-fraud calling patterns.

Because unified communications SIP trunks offer direct IP connectivity into your enterprise network, these trunks are intrinsically far more insecure than TDM trunks. Important security considerations for the enterprise using a unified communications SIP trunk include:

- Accepting calls only from the service provider
- Protecting your network from floods of calls
- Protecting your network from denial-of-service (DoS) attacks
- Toll fraud and the general ability to ensure that calls are accounted for correctly

All traditional IP data attacks may also potentially enter on this connection, and traditional security mechanisms such as Network Address Translation (NAT), firewalls, access lists, and intrusion detection and prevention mechanisms should therefore be deployed on the unified communications SIP trunk entry point.

Network Address Topology Hiding

Hiding the IP addresses of enterprise voice endpoints (such as those belonging to IP phones, call agents, and TDM voice gateways) from external view requires more than NAT. NAT adjusts the IP addressing of IP packet headers and some of the IP addresses appearing elsewhere in a SIP packet, but there are additional SIP header fields containing IP addresses that NAT does not adjust. Therefore, you should use a back-to-back SIP user agent at the network demarcation point of the unified communications SIP trunk. The Cisco Unified Border Element can provide this agent, where the media and signaling flow through the Cisco Unified Border Element and the service provider sees only the addresses of this device.

The Cisco Unified Border Element terminates the entire SIP session and re-originates it on the other side, thereby changing IP addresses in all fields of the SIP messaging, ensuring that an endpoint outside the enterprise network never sees an internal enterprise IP address; only the IP address of the Cisco Unified Border Element is visible.

This topology hiding is important to ensure that any attacks that come from the service provider can be directed only toward the demarcation point, and the communications within the enterprise is not disrupted.

Firewalls

The security features in Cisco IOS Software can serve as the first line of defense against outside attacks from outside the enterprise, as well as a checkpoint for the internal traffic exiting to the service provider's network through the router. Infrastructure access control lists (ACLs) are required to keep out unwanted traffic through the physical links to the service provider. These ACLs are used primarily to stop unauthorized access, DoS attacks, or distributed DoS (DDoS) attacks that originate from the service provider or a network connected to the service provider's network, as well as preventing intrusions and data theft.

The configurations used to prevent outside traffic from entering the enterprise include firewall and NAT. A firewall must be application-aware and open up a pinhole for RTP traffic based on reviewing the information in the SDP messages of the SIP setup request. An additional value of a firewall on a unified communications SIP trunk enterprise termination point is that pinholes for RTP traffic can be torn down after specific timeouts to allow for additional protection of the network if an initial SIP INVITE request was received without a corresponding BYE, leaving pinholes open.

Encryption of Media and Signaling

SIP signaling message authentication and encryption can be provided with Transport Layer Security (TLS), which requires the exchange of keys between the service provider and the enterprise customer, as well as the synchronization of the clocks. The media (the RTP stream) of a SIP call can be secured by using Secure RTP (SRTP).

When evaluating a unified communications SIP trunk offering, your enterprise should determine if TLS security can be applied to the traffic to ensure security and protection of call-related information.

Porting Phone Numbers to SIP Trunks

When an enterprise starts using a unified communications SIP trunk for incoming calls, the phone number must be "ported" to this service. When external end users call the number, rather than ringing at the traditional TDM gateway owned by the enterprise, it rings in the service provider's core network and the call is routed to the enterprise with the SIP.

Because of the complexity of porting phone numbers, most SIP deployments start with outbound calls initially. It is important for the enterprise to understand the timelines and transition plans offered by the service provider for porting direct inward dialing (DID) numbers. Enterprises' business users cannot afford to be unreachable on their primary PSTN phone numbers while this porting activity is occurring.

Emergency Calling

Emergency calling is an important consideration to account for when integrating unified communications SIP trunk access into the enterprise. Normally emergency calling is based on the

emergency responder knowing the physical location of the TDM connection from which the call is coming. With a unified communications SIP trunk, that relationship between the physical location and the calling number no longer exists.

Options for handling emergency calling include:

- Continuing to route emergency calls through your TDM PSTN gateways
- Having a small number of TDM trunks dedicated to this function at the physical location of the service provider
- Adopting a SIP-based emergency calling solution

All unified communications SIP trunk providers should provide clear explanations of their solution for providing emergency calling when an IP connection is evaluated.

Billing

Typically, the service provider can do billing without any information from the enterprise. Call detail records (CDRs) from the Cisco Unified Border Element can provide a consolidated aggregate view of calls sent and received on the unified communications SIP trunk and can be used to validate the service provider's billing.

Cisco IOS Software CDRs contain calling and called numbers, local and remote node names, data and time stamp, elapsed time, call failure class fields, and some vendor-specific attribute (VSA) fields. Each call through the Cisco Unified Border Element is considered to have two call legs. The Start and Stop records are generated for each call leg. These records can be sent to a RADIUS server or retrieved with SNMP polling using the dial-control MIB. For information, refer to the "RADIUS VSA Voice Implementation", "CDR Logging with Syslog Server", or "Voice MIB Objects" documents.

Troubleshooting Tools and Methods

In traditional TDM PSTN access, the PSTN gateway terminated the TDM connection from the provider's network and originated a unified communications IP connection inside your enterprise network. If voice quality or connectivity problems existed, this demarcation point was an easy place to conduct testing and isolate whether the problem existed within your enterprise network, or whether it was the service provider's problem. TDM loop testing is common, allowing the service provider to test the TDM loop to the edge of your network to determine if the problem exists on that part of the connection.

Bringing a unified communications SIP trunk into the enterprise removes this demarcation point and therefore also the problem isolation techniques that existed for TDM interconnection. If voice-quality problems occur, it can be very difficult to isolate whether they are caused by something in the service provider's network or by an element in your enterprise network.

Using a Cisco Unified Border Element as an IP demarcation point restores this troubleshooting capability, allowing testing within the enterprise network to the Cisco Unified Border Element, as well as testing from the service provider's side to the Cisco Unified Border Element to determine where a fault may exist. The IP "loop" can be tested in the same conceptual manner (RTP loopback capability, as well as Service Assurance Agent [SAA] responder support) as the TDM loop to allow the service provider to determine if the service is causing a problem, or whether the problem exists in the enterprise.

Summary

Unified communications SIP trunks are becoming an increasingly viable option for enterprises wishing to deploy new IP-based services to their user communities, as well as with customer and vendors external to the enterprise network. As discussed in this paper, it is a network change that should be done with the appropriate planning, and may require several phases of deployment.

Steps to integrate a unified communications SIP trunk into the enterprise should include implementing a pilot deployment at the headquarters site, rerouting certain traffic in a limited manner, and measuring the results.

As the use of SIP continues to expand in the industry, Cisco continues to provide capabilities in Cisco IOS Software products and features to enable rich communications methods.

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
168 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Europe Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: +31 0 800 020 0791
Fax: +31 0 20 357 1100

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-404004-00 06/07

