Best Practices When Implementing SIP Trunks for PSTN Access

This document summarizes the best practices for planning, preparing, and deploying Session Initiation Protocol (SIP) trunking for public-switched-telephone-network (PSTN) access in an enterprise network. Several areas of best practices are covered, including:

- Providers
- Deployment
- Network design
- Protocols and codecs
- Cisco® Unified Communications Manager (UCM)
- Session Border Controller (SBC)
- Security
- Redundancy

The text in this white paper is based on an extract from Chapter 11 of the 2010 Cisco Press book “SIP Trunking” (ISBN: 1-58705-944-4). For more comprehensive information about the SIP trunking industry, as well as the planning, design, and implementation of SIP trunks in an enterprise network, you can order the book from: http://www.ciscopress.com/bookstore/product.asp?isbn=1587059444.

Service Providers

Best practices in the area of selecting a SIP trunk service provider include:

- Look for a provider that owns the physical delivery medium (last mile) to your premises along with the service. The last-mile provider is the only one that can guarantee quality of service (QoS).
- If keeping your existing direct-inward-dialing (DID) numbers is important to your business, your most likely provider is the same one with whom you get your current PSTN service. If you are considering a different provider, assess whether you can transfer your DID numbers to the new provider and how long the transfer will take.
- Evaluate SIP trunk service offerings carefully because this service is unregulated and offerings and pricing can vary greatly.
- Always do a proof-of-concept trial before installing a SIP trunk into your production business network. Sample test plans.
- Evaluate SIP trunk offering features against your current time-division multiplexing (TDM) service and make sure you get all the features that are important to you. Features to pay particular attention to because they may not be offered on SIP trunks or they may operate very differently from TDM trunks include: Malicious Call Identification (MCID), Multilevel Priority and Preemption (MLPP), voice-band data (point-of-sale [POS] terminal, alarm systems, etc.), caller ID delivery, recording, and 911 call routing and solutions.
- Discuss SIP trunk status monitoring and troubleshooting methods and responsibilities with your provider.

Deployment

Best practices in the area of deployment include:

- Plan carefully and do not rush into a production deployment until you are certain of what you are getting. Do proof-of-concept testing for the call flows that are critical to your business operation.
- Define the user communities or sites you will deploy with SIP trunking in the different phases of the deployment. For all but the smallest of networks comprising no more than a single or a handful of sites, this plan is for a multi-phased deployment.
- Decide on the call flows that are included in each phase of deployment; for example, inbound, outbound, long distance, contact center, or general business calls. Contact center deployments are often the easiest to show a return on investment (ROI), but are the most complex call flows to validate. General business call flows are easier to validate, but may have a longer ROI. Outbound calls are easiest to implement; inbound calls require DID porting to be resolved.
- SIP trunk provider readiness and cost may vary significantly between regions, countries, and continents. Your deployment plan may require geographical phases.

Network Design

Best practices in the area of network design include:

- Always use an SBC to terminate a service provider SIP trunk into your network, regardless of whether you use Cisco Unified Communications Manager, Cisco Unified Communications Manager Express or any other vendor's call agent. An SBC offers security, demarcation, session management, and interworking features that protect your network from denial-of-service (DoS) and other SIP-based attacks, allow you to resolve troubleshooting problems, provide SIP trunk status monitoring tools, and aid in SIP manipulation to overcome myriad interoperability problems present in the industry. A Primary Rate Interface (PRI) gateway implicitly offers these same services into your network—but when migrating calls from PRI trunking (for PSTN access) to SIP trunking you lose these controls and security protection unless you use an SBC.
- Carefully consider the benefits and challenges of the centralized or distributed SIP trunk designs, even if the choice seems obvious. Centralized designs almost always look more attractive because of their cost savings. However, this choice has network design implications that may increase the overall cost, as well as make redundancy designs a much more serious consideration because the aggregated session counts are much higher than with a TDM gateway. A distributed design may be viable, especially if you already have a distributed Multiprotocol Label Switching (MPLS) network for your data network.
- For redundancy reasons you should consider having at least two SIP trunk entry points into your network. For medium and larger networks, these points should be geographically separated.
- Most SIP trunk providers allow only two IP addresses with the service, either as primary/secondary or in a load-balancing setup. If you have to load balance your calls over more than two devices for either scalability or redundancy reasons, you may have to insert a SIP proxy or load-balancer device in between to distribute the traffic. Some providers use Domain Name System (DNS), which offers more IP address destination choices for a single logical SIP trunk.
- Most SIP trunk providers currently do not use DNS, but instead use absolute IP addressing. This situation may influence your redundancy and load-balancing considerations.
- Load balancing over the two IP addresses the provider allows is generally a more dynamic and flexible call-routing configuration than a primary/secondary algorithm.

- Use a Layer 7 SIP OPTIONS ping to monitor SIP trunk status in addition to one of several Layer 3 mechanisms that may be available, such as Internet Control Message Protocol (ICMP) ping or IP service-level agreements (IP SLAs).
- Investigate troubleshooting tools and methods to determine the source of any voice quality concerns. If one is reported, decide how you will determine if it is present in your network or in the service provider's network. Discuss with your provider how such problems will be investigated and resolved.
- Turn on Call Admission Control (CAC) features on your session border controller whether or not your provider offers an SLA, and especially so if your provider does not offer an SLA. Deploy CAC features to monitor both simultaneously active calls and call arrival rates.

Protocols and Codecs

Best practices in the area of protocol choices include:

- Use SIP end to end in your network if you can. Interoperability is easier and more flexible on SIP-to-SIP connections than translating to other protocols.
- If adhering to the previous guideline is not possible, use your SBC to interoperate H.323 destinations with your SIP trunk.
- Use RFC 2833, Dual-Tone Multifrequency (DTMF) Relay, throughout your network if possible. If not possible, make sure you understand where in your network translations occur between out-of-band signaling methods (such as the traditional H.323 methods) and DTMF, which travels in the media stream. The device (such as a call agent or SBC) that does this conversion must have access to both the signaling and the media streams to do the conversion.
- Investigate DTMF payload type value assignments on your own call agents, endpoints, and applications, as well as the values the service provider uses. Conversion or interworking among these values may be needed on your SBC to ensure proper interoperation of calls.
- Using T.38 fax relay for fax-over-IP transmission is technically a more robust method of faxing and works better than other methods where available. But it is not available on all networks and to all endpoints, so fax passthrough (or fax through G.711) is still widely used because it interoperates easier and with more endpoints. If your provider offers T.38, investigate if failover to G.711 fax is offered as a secondary call negotiation service to connect calls to destinations that may not yet support T.38. Consider keeping fax on TDM trunks for a while longer if faxing is a critical part of your business.
- Similar to the previous point, keep modem, POS, and telecommunications devices for the hearing impaired (TDD) traffic on TDM trunks for the present. SIP trunk technology is not ready to carry these traffic types reliably and predictably.
- Strongly consider getting G.711 service on your SIP trunk. Although it uses more bandwidth than G.729, it does not compromise voice quality (after all, SIP trunk technology is supposed to improve PSTN services, not worsen them). It also eases fax concerns, obviates the need for transcoding at your SBC, and better positions you for new SIP trunk services that will almost certainly include increasingly bandwidth-intensive applications such as high-fidelity wideband codecs and video. Dimensioning your network for G.711 call bandwidth now positions you better for future G.722 service, which offers much better voice quality at the same bandwidth use and is likely the first wideband codec to be offered on SIP trunks in the future.
- Although SIP is a standard protocol, it actually consists of a large number of individual standards (IETF Request for Comments [RFC]), and these RFCs have many optional components, as well as alternative ways of implementing the same call flow. SIP interoperability is not mature enough that all applications predictably interwork with other SIP applications, even though they all may be standards-compliant. This situation means you may need tools to normalize (that is, manipulate) SIP messages as they flow through your network from

your applications across the SIP trunk to the PSTN. Investigate these tools (generally available on call agents, SIP proxies, and SBCs) to see which will meet your needs and where in the network is the best place to do these manipulations to ensure transparent interoperability with all your applications.

Cisco Unified Communications Manager

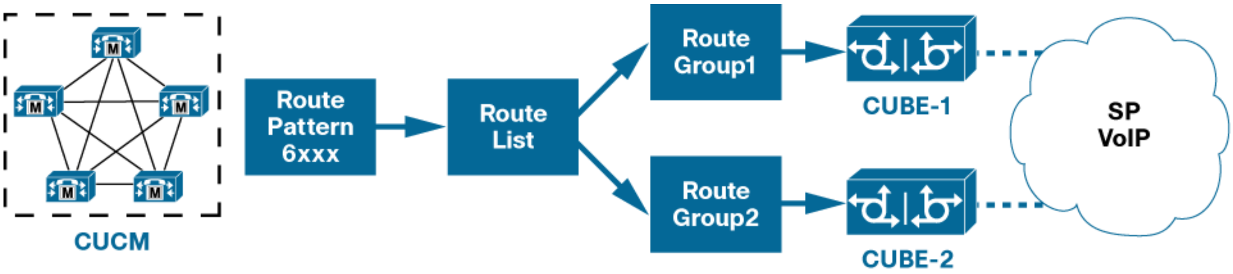
Best practices in the area of Cisco Unified Communications Manager deployment include:

- Generally, the decision between using H.323 and SIP on Cisco Unified Communications Manager depends on the features, your preference, protocol maturity, and the degree of interoperability among the applications you have in your network.
    - A minimum release of Cisco Unified Communications Manager 6.0 is recommended for SIP trunking, using a SIP-to-SIP configuration.
    - If older versions of Cisco Unified Communications Manager are used and they cannot be upgraded to 6.0 or 7.0, use H.323 trunking to the call agent and H.323-to-SIP conversion on your SBC (Cisco Unified Border Element [UBE]).
- The recommended Cisco Unified Communications Manager configuration for SIP trunking includes:
    - Delayed Offer (with no media termination point [MTP]) for Cisco Unified Communications Manager outbound calls
    - Early Offer (with no MTP) for Cisco Unified Communications Manager inbound calls

Note: Use Delayed Offer-to-Early Offer interworking on your SBC (Cisco Unified Border Element).

- Avoid MTP designs if possible; if not, co-locate the MTPs with your SBC (Cisco Unified Border Element) to optimize the media path.
- Older releases of Cisco Unified Communications Manager prefer higher bandwidth codecs to lower ones, so configure codecs carefully to get the optimum negotiation with endpoints in your network and across the SIP trunk. Newer releases may support codec preference ordering configuration options.
- Configure alternate (redundant) PSTN routing if the SIP trunk is down. It is best not to remove TDM PSTN trunking and gateways from your network until after a SIP trunk has been proven. Generally, a Cisco Unified Communications Manager configuration for alternate PSTN routing includes:
    - Trunks contained in route groups
    - Route groups contained in route lists
    - Route lists used to cycle through alternate trunk destinations, as shown in Figure 1
    - Retry timers and counters tuned to optimize failover time to alternate trunks
    - Return code mapping (SIP codes to Q.850) used to stop Cisco Unified Communications Manager trunk selection when needed

Figure 1. Cisco UCM Route Lists Used for Alternate Routing



SBC Best Practices

Best practices in the area of SBC deployment include:

- For small-site (or remote-office) SBC deployments that carry low traffic, use a single Cisco router with integrated services such as Cisco Unified Border Element, MTP, VoiceXML, firewall, and Survivable Remote Site Telephony (SRST) features.
- For larger-site (or campus or data center) SBC deployments that carry high traffic, use a dedicated router for Cisco Unified Border Element, and other dedicated routers for MTP or VoiceXML or any other service you may need at that site.
- Pay careful attention to performance engineering of SIP trunks into your network because this migration most likely changes the call flows as well as the bandwidth allocation to calls.
    - Whether you use H.323-SIP or SIP-SIP on Cisco Unified Border Element makes no difference to its session capacity.
    - DTMF interworking or Delayed Offer to Early Offer adds no significant extra load to Cisco UBE.
    - Using SIP profiles for normalization on Cisco Unified Border Element may have a performance effect depending on the number and complexity of the rules, but is not generally a significant performance factor.
    - Configuring MTPs on the same platform as Cisco Unified Border Element and using transcoding are CPU-intensive tasks and must be carefully engineered.
    - If you are using Cisco Unified Border Element platforms terminating fewer than 500 sessions (calls) per platform, most of the performance engineering needed is likely on the Cisco UBE platform itself and not elsewhere in your network. But if you are using high-end Cisco UBE platforms terminating 2000 or more sessions per Cisco UBE platform, the bottleneck may move to your Cisco Unified Communications Manager servers or IP private branch exchanges (PBXs) connected to Cisco UBE, and these systems must be included in the performance engineering work.
- Define explicit incoming and outgoing dial peers on your Cisco Unified Border Element device. SBC calls have two IP call legs (as opposed to a TDM gateway that has only one), so it is often necessary to have both dial peers to control the characteristics of each individual call leg—the default settings may not be what you need in your network. This practice also gives you additional control points to combat toll-fraud attacks.
- Do not co-locate Network Address Translation (NAT) services on the same router interface that carries Cisco Unified Border Element traffic. Cisco UBE provides topology hiding, so it already performs a “NAT” function on the voice streams.

Security

Best practices in the area of security include:

- In a small single-site or very small network, a firewall or NAT device may be sufficient as the enterprise demarcation and security border device. In all other networks, use a Layer 7 SBC device with additional protection, configuration, and traffic control.
- If you are deploying both a firewall device and an SBC (as most large enterprises do in their campus sites), it is generally recommended to put the firewall on the outside as a first line of Layer 3 security defense for all traffic, and then put Cisco Unified Border Element behind it as a second line of defense optimized for Layer 7 unified communications traffic.
- Use SIP registration on the SIP trunk if your service provider offers it. Also use Digest Authentication on both SIP registrations and INVITEs if available as part of the service.
- Deploy toll-fraud features on your Cisco Unified Border Element or SBC. Hackers target SIP deployments and SIP ports more than other voice-over-IP (VoIP) network architectures. Toll fraud is the single most-often reported security attack on SIP trunks today.
- Consider changing the SIP port from the standard port 5060 for additional protection against Internet sweeps for open SIP ports. Both your border element and the service provider SBC must make this change to be able to interoperate successfully.
- Always put access control lists (ACLs) on your Cisco Unified Border Element to ensure only the service provider SBC can initiate calls to it from the PSTN side, and only your enterprise call agents (Cisco Unified Communications Manager or IP PBXs) can initiate calls from the internal network side.

Redundancy

Best practices in the area of redundancy include:

- Consider centralized and distributed trunking network designs carefully. An advantage of distributed trunking is inherent redundancy and geographic distribution of risk similar to TDM PSTN gateway architectures. If the SIP trunk is down at one site, you can temporarily route calls from that site through the SIP trunk of a different nearby site until service recovers.
- For SIP trunks with large session counts (generally exceeding 500, roughly equivalent to a T3's 672 calls) redundancy is imperative. For SIP trunks with fewer sessions redundancy is generally a good design, but may not always be required. Make an assessment of the business impact of a SIP trunk failure before you decide not to implement redundancy.
- Use redundant (and geographically separate) SIP trunk entry points into your network if the SIP trunk session capacity is 1000 or larger. Consider multiple entry points for 500 sessions or more.
- Use redundant Cisco Unified Border Element hardware deployments (inbox failover or box-to-box failover) for terminating a SIP trunk where a single SBC (Cisco UBE) platform carries 1000 or more sessions. Consider hardware redundancy for 500 sessions or more.
- Generally SIP trunk service providers offer two IP addresses with a SIP trunk, which you can use to terminate onto two different devices. Either one can handle the full load of the SIP trunk if necessary. The service provider can load balance over these two destinations. If you want to load balance across more than two destinations for either scalability or redundancy or both reasons, you may have to put in a SIP proxy such as Cisco Unified SIP Proxy (Cisco USP) or another load balancer in between to help distribute the load.

- If the SIP trunk is down, you can easily route outbound calls from the enterprise using traditional TDM PSTN gateways. However, inbound traffic to your DID numbers is not so easily rerouted to traditional TDM trunks. Discuss inbound traffic alternate routing possibilities with your SIP trunk service provider.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

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 C11-620461-00 09/10