VoIP endpoints and call agents such as CUCM and CUCMExpress also have facilities to control and mark packets. These can be used directly if the enterprise markings are the same as the SP UNI markings, and an SBC can be used if markings need to be translated between the enterprise and the SP networks.

Security Considerations

The security concerns of TDM trunking, primarily toll fraud, exist equally on SIP trunking. In addition, SIP trunking exposes your network to IP level threats similar to data WAN or Internet access, such as denial of service (DOS).

For a hacker to gain access to your enterprise IP network via a TDM voice trunk is virtually impossible to do unless the TDM connection is specifically configured for modem dial-up access—and most voice trunks are not. Perpetrating a DOS attack on a TDM trunk is also highly unlikely as it is both expensive to do and requires large-scale auto-dialer equipment the average Internet hacker does not have access to. Launching these same attacks on IP addresses is significantly easier and open to a much larger pool of perpetrators because no sophisticated equipment is necessary, and the attacks can be launched for free from any Internet access connection.

When considering security on SIP trunks, you need to take into account different aspects of security. These aspects call for a series of features and capabilities to mitigate the potential threats. Security is always best deployed in a layered architecture, rather than a single box or feature that strives to protect against all possible attacks. Areas worth exploring for SIP trunk security include

- Determine the level of exposure on the SIP trunk, which depends on how it is deployed and who the provider is.
- Limit the devices that can contact your network via the SIP trunk. Mitigation capabilities include features such as access lists, hostname validation, and voice source group definitions.
- Hide your enterprise network addressing from the outside (which could be Internet-visible) and inspect the validity of traffic that enters your network. Mitigation techniques include network address translation (NAT), topology hiding, firewalls, and intrusion protection services (IPS).
- Determine protocol and session validity. Mitigation techniques include SIP port settings, SIP protocol inspection and termination, registration, and authentication methods.
- Lock down your SIP trunk against toll fraud access using the same methods you used on your TDM gateways.
- Control the privacy of sessions on the SIP trunk. Mitigation techniques involve the control of originator information available outside the enterprise network with the use of SIP privacy headers, SIP normalization, digit manipulation, and encryption

methods of the signaling and the media streams (such as Transport Layer Security [TLS], Secure RTP, and the use of IPSec tunnels or virtual private networks (VPN) on the IP connections).

SIP Trunk Levels of Security Exposure

The level of security exposure depends on the characteristics of how the SIP trunk connects into your network and the strength of security protection your service provider offers.

Figure 7-2 illustrates four increasing levels of exposure depending on the connectivity method of your SIP trunk:

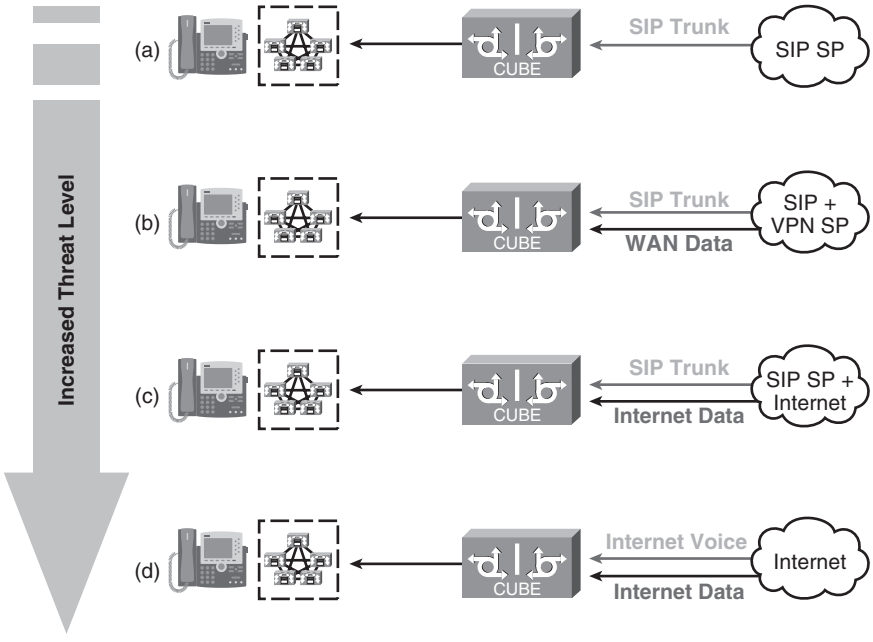

Figure 7-2 *Increasing Levels of Security Exposure*

- In model (a) the SIP trunk connects from a Tier 1 service provider with strong security over a dedicated physical connection into your network. No data traffic traverses this connection. With this model, your security exposure is low, and you can consider not having a firewall in addition to a border element on such a connection.
- In model (b) the SIP trunk connects from a Tier 1 service provider with strong security over a physical connection that carries both your voice and your VPN WAN data connection, such as an MPLS service. No Internet data traffic traverses this connection. With this model, your security exposure is still fairly low, and you might not need a firewall in addition to a border element on such a connection.

- In model (c) the SIP trunk connects from a service provider that offers both SIP trunking and Internet access on the same physical connection. This is often a cost-effective model for smaller businesses with no WAN data service between sites or that have only a single site. Regardless of the strength of security measures in the service provider's network, you are exposed to Internet attacks on this kind of connection, and you have to firewall in addition to deploying a border element to secure this type of connection.
- In model (d) there is no SIP trunk service offering, and you use plain Internet consumer voice access and Internet data from a general Internet service provider. This model is strongly discouraged for business-class voice access because there is no quality control on such a connection, and it is extremely exposed to all kinds of voice and data Internet attacks. Firewalling and border controlling alone are still not sufficient to make this model capable of providing business-quality voice services.

Many security features on both firewalls and border elements protect against attacks on SIP trunks. The following sections discuss these techniques in more detail.

A general best practice for SIP trunk security is always to use a border element to terminate a SIP trunk coming into your network. This can be an appliance function (such as deploying a dedicated CUBE), or it can be an integrated function, such as an IAD or CUCM Express device that acts as a border element and a routing or IP-PBX device in your network.

In addition to a border element, you can choose also to deploy a firewall. Again, this might be a separate appliance, or it might be integrated into a Cisco IOS router providing multiple functions to your business. Separate, dedicated devices tend to be the norm for larger enterprise and higher volume SIP trunks, whereas integrated devices tend to be the cost-effective solution for smaller sites or small business networks.

Access Lists (ACL)

Always strictly limit the devices that can access your SIP trunk, both from internal to your network and external to it. If you terminate your SIP trunk on a border element, you do not need all these security mitigation measures on every enterprise application, only on the border element. The border element itself should be set up to accept connections on the service provider side only from the provider's SBC, and on the enterprise side only from legitimate CUCM, IP-PBX, or other valid applications (for example, SIP proxies and meeting conference servers).

United States federal information reports that hackers are as frequently located inside your enterprise network as on the outside, and for that reason, it is imperative to lock down your border element on both sides so that rogue endpoints and applications inside your network cannot use the SIP trunk service for fraudulent calls. Similarly, rogue endpoints on the Internet should contact your SIP trunk. This configuration is illustrated in Figure 7-3.

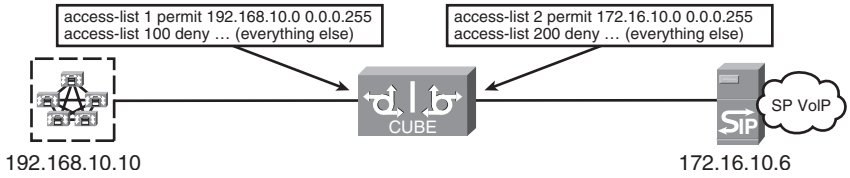

Figure 7-3 *Locking Down a SIP Trunk with ACLs*

Additionally, voice Source IP Groups can be used with the ACLs, as shown in Figure 7-3, to provide further restrictions on the devices that might originate SIP traffic to your border element. On devices in your network that should not run SIP traffic at all, the Control Plane Policing (CoPP) feature can be used to deny all SIP traffic.

CUCM has (by default) a feature that restricts traffic on a SIP trunk to be accepted only from the IP address configured on the SIP trunk.

Hostname Validation

You can use the hostname validation feature of the CUBE to restrict the valid hostnames that are accepted in the host portion of the SIP URI of an incoming SIP INVITE. Example 7-1 illustrates the commands used by this feature to enable calls only from the four hostnames listed.

Example 7-1 *Hostname Validation*

```
sip-ua
  permit hostname dns:example1.sip.com
  permit hostname dns:example2.sip.com
  permit hostname dns:example3.sip.com
  permit hostname dns:example4.sip.com
```

Security features often overlap to some extent, and it is a good practice to deploy these overlapping features because they provide layered security protection. Every layer might protect you against one particular attack that might have skirted around a single layer protection to exploit a weakness in a particular appliance, device, feature operation, or configuration.

NAT and 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 can in some cases be achieved with traditional NAT features. NAT adjusts the IP addressing of IP packet headers and some of the IP addresses appearing elsewhere in SIP packets, but generic NAT devices are Layer 3-capable only. Those that have Application Layer Gateways (ALG) have more sophisticated SIP awareness, but still, generally, might offer only suboptimal capabilities to translate deeply embedded IP addresses in SIP messaging.

It is therefore more secure to use a border element that is a full SIP back-to-back user agent (B2BUA) as the network demarcation offering 100 percent SIP packet inspection and address translation. The CUBE is a full SIP B2BUA and can therefore offer complete network address translation, usually referred to as topology hiding in this context to distinguish this function from appliance NAT devices. Both media and signaling flow through the CUBE and the service provider and off-net endpoints see only the addresses of the border element and never the addresses internal to your enterprise network.

Topology hiding is important to ensure that any attacks that might come from the service provider side can be directed only toward the border element, and the communications and call agents within your enterprise remain unaffected.

Figure 7-4 illustrates how topology hiding can be accomplished by using the CUBE.

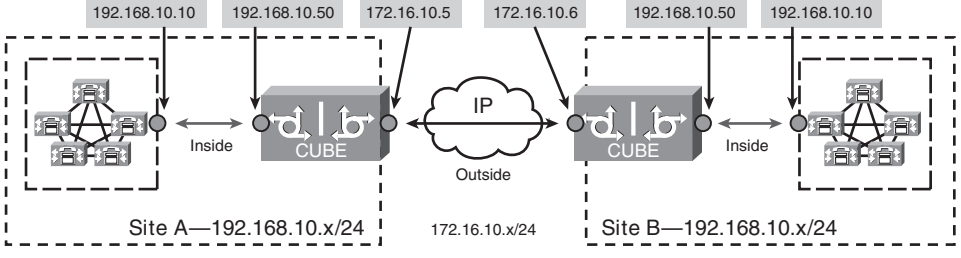

Figure 7-4 *Topology Hiding*

Firewalls

Many security features on both firewalls and border elements protect against attacks on SIP trunks. A certain amount of overlap occurs between the capabilities, especially true for the higher end firewalls with sophisticated SIP ALGs.

Generally you should deploy a firewall to provide generic IP protection against any kind of IP traffic, and your border element as a much more focused, voice-specific session protection function. For the least capable firewall devices, you should simply open pinholes for the traffic destined to the border element and have the border element do all the SIP inspection. For firewalls with SIP ALGs, there is some overlap in the inspection the firewall does and the inspection done by the border element. The border element always

provides the most sophisticated layer of protection because it is a B2BUA whereas the firewall essentially inspects and passes through traffic but does not terminate it.

Functions that firewalls are particularly well suited to mitigate are Layers 2 and 3 inspection functions including:

- General IP DOS attacks
- Black hole routing
- TCP window control and dropping UDP packets
- Access lists, specifying what traffic is correct and allowed
- Optional SIP ALG for cursory SIP rogue and malformed packet inspection
- Optional SIP ALG protection against spikes of SIP calls (SIP-specific DOS)

More sophisticated SIP capabilities that some firewalls can have include

- Whitelist/blacklist filtering of SIP calls based on calling and called numbers
- Rate limiting of specific SIP methods to mitigate against SIP-specific DOS attacks

Firewalls are not as well suited to protecting against attacks launched from inside your network or doing session management at the level of deciding whether packets are arriving for valid sessions only, in valid sequences (or SIP dialogs), and for valid codecs or other negotiated parameters of the session. Some of the more sophisticated firewalls, such as the Cisco ASA product series or the Cisco IOS Firewall, have SIP ALGs that offer some protection services at protocol layers higher than Layer 3.

Specific functions a border element is well suited for include Layers 5 to 7 SIP inspection actions such as:

- Rejecting nonallowed calls and generating CDRs of call attempts for tracking
- Call limiting (only accept a certain number of calls)
- Codec limiting (only accept certain codecs)
- Call admission control to provide bandwidth protection
- Access lists specifying valid source and destination call agents
- Complete rogue and malformed SIP packet protection
- Digest authentication and hostname validation to ensure sessions are set up only between valid endpoints
- SIP registration to authenticate session originations
- SIP listening port configuration

Broadly, firewalls and border elements are deployed in one of two ways:

- Separate devices in series
- Integrated in a Cisco IOS device with collocated functions

Figure 7-5 provides six possible deployment models of firewalls and border elements.

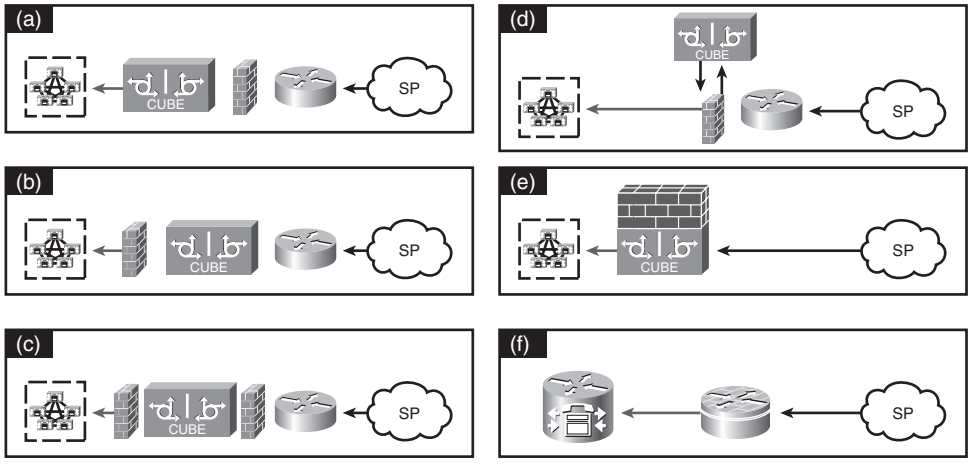

Figure 7-5 *Possible Firewall and Border Element Designs*

Models (a), (b), and (c) shown in Figure 7-5 are better suited to medium-to-large enterprises and high volume contact centers, and models (d), (e), and (f) are better suited to smaller businesses.

- In model (a) the firewall appliance is on the outside of the border element. This is the recommended deployment model if you use separate devices for firewall services and a border element. This deployment generally makes sense for campus and data center locations where there is already a firewall present. This model also makes sense if the firewall is managed by the security team, whereas the border element is managed by the voice team. This is a mandatory model if the physical medium coming into the enterprise premises carries Internet traffic.

  In this model, the firewall provides the first line of defense on all traffic arriving from the outside, passes the voice traffic to the border element for a Layer 7 inspection on the voice traffic. If the firewall has an ALG function, there is bound to be some overlap in functionality between the firewall and the border element. It is nevertheless recommended that you turn on both to get the fullest set of inspection and protection that you can, rather than having potential security holes between the appliances.

- In model (b) the border element is on the outside of the firewall. This deployment model makes sense when the physical medium bringing the SIP trunk into your

premises carries *only* SIP trunk traffic and nothing else. This means your data connections come in on a different physical path, onto different routers, and get firewalled entirely separately from the SIP trunk traffic. This model mandates that you trust your service provider's network to offer only clean SIP traffic to your enterprise.

- In model (c) two firewalls are on either side of the border element. Some refer to this model as the one for the truly paranoid, but this is the classic design of a DMZ (demilitarized zone). It is not an uncommon design, especially in large financial, educational, and government institutions, or any other business particularly attractive to hackers.
- Model (d) is a variation of model (c), where there are two virtual firewalls on either side of the border element, but one physical firewall device is used for the function, routing the unified communications (UC) traffic twice. This is a virtual DMZ design often used in video deployments where the CUBE is not only fronting a SIP trunk, but is also bringing in H.323 Internet video traffic and acting as a Cisco IOS Gatekeeper.
- Model (e) provides a more cost-effective integrated deployment model for smaller sites or businesses where a separate firewall appliance does not already exist, is not desirable, or the cost is not justified. In this model the Cisco IOS router acts as both the CUBE and the firewall. Traffic flowing through this router is inspected first by the firewall and then handed to the border element for further processing. It is therefore conceptually similar to model (a).
- Model (f) provides a lower end offering for commercial or small businesses (without IT departments) that do not want to carry the cost or the management of either a border element or a firewall. In this model, an integrated service from a service provider is purchased, and all security and demarcation issues is handled by the service provider. The service provider puts an IAD at the customer premises to connect to its IP-PBX or key system, such as CUCM Express. The IAD device will likely do NAT, perhaps basic firewalling, but essentially all the service provider's network and security are delivered as a managed service.

Security Protection at the SIP Protocol Level

SIP is a widely used and understood protocol and simple to create because it uses straight text encoding in its messages (unlike H.323 that uses ASN.1 encoding). This makes SIP an easy target for hackers. Many of the protocol attacks can be launched against H.323 as well, but very few incidents of this were in the industry because H.323 is not as accessible as SIP.

Several ways to protect your network against a variety of SIP protocol attacks include

- Setting the SIP listening port
- Using TLS for authentication
- Using a border element B2BUA

- Using SIP normalization techniques to suppress or overwrite information in the SIP message such as the calling phone numbers, hostnames, or descriptive tags before a call enters the public network
- Using digit manipulation techniques to suppress or overwrite phone numbers before a call enters the public network
- Using SIP privacy settings to communicate the information within the SIP message that might or might not be used

Each of these areas is discussed in the following sections.

SIP Listening Port

Every Internet hacker knows the default SIP listen ports and can sweep them from any Internet location to find an open port to launch fraudulent calls, all while your business pays for them. One way to protect against this is to change the SIP listening port to a nondefault setting. It requires the service provider to set the complementary port on the provider edge SBC. This alone can protect you against the majority of hacker attacks launched against SIP port 5060.

Example 7-2 shows the commands needed to set the SIP listening port to a nondefault setting.

Example 7-2 *SIP Listening Port Setting*

```
voice service voip
 sip
   shutdown
voice service voip
 sip
   listen-port non-secure 2000 secure 2050
voice service voip
 sip
   no shutdown
```

Transport Layer Security (TLS)

Another way to protect against this attack is to use TLS (specified in IETF RFC-2246). TLS uses an authentication mechanism that ensures only valid endpoints connect to your SIP trunk, and if the authentication fails, the call is refused.

Although this is a good way to mitigate fraudulent SIP calls, none of the current SIP trunk offerings in the market include TLS as an option. Hopefully this situation will change.

Back-to-Back User Agent (B2BUA)

A B2BUA (such as the CUBE) terminates and reoriginates all calls before they enter your network. All SIP traffic passes through the SIP stack on the B2BUA twice (on ingress and egress) so that all malformed or rogue packets are dropped.

SIP Normalization

There are certain numbers, names, or other internal information you might want to populate informative displays on the endpoints in your network. When these calls exit over the SIP trunk to external destinations, you might not want all this information to remain in the SIP messaging, especially non-DID numbers used by your organization. You can use SIP normalization features to insert, delete, or change this kind of information in the SIP messaging on your border element.

Examples 7-3, 7-4, and 7-5 show how SIP normalization can be used on the CUBE to modify the *From* header in an INVITE to a **gateway@ip-address** format and to add the **phone-context=gateway** field to the *To* header of the INVITE. Example 7-3 shows the commands needed for the configuration; Example 7-4 shows the original SIP INVITE; and Example 7-5 shows the resulting INVITE after normalization has been applied.

Example 7-3 *SIP Normalizations Commands*

```
voice service voip
  sip
    sip-profiles 1
voice class sip-profiles 1
  request INVITE sip-header From modify "(<.*:)(.*@)" "\1gateway@"
  request INVITE sip-header To modify "<(.*)>" "<\1;phone-context=gateway>"
```

Example 7-4 *Original SIP INVITE*

```
INVITE sip:2222000020:5060 SIP/2.0
Via: SIP/2.0/UDP 9.13.24.6:5060;branch=z9hG4bK1AD9E2
Remote-Party-ID: "sipp " <sip:sipp@9.13.24.6>;party=calling;screen=no;privacy=off
From: "sipp "<sip:sipp@9.13.24.6>;tag=23C3F840-99A
To: <sip:2222000020@9.13.24.7>
Date: Thu, 30 Aug 2007 07:04:36 GMT
```

Example 7-5 *Normalized SIP INVITE*

```
INVITE sip:2222000020:5070 SIP/2.0
Via: SIP/2.0/UDP 9.13.24.7:5060;branch=z9hG4bK1191BFD
Remote-Party-ID: "sipp " <sip:sipp@9.13.24.7>;party=calling;screen=no;privacy=off
From: "sipp "<sip:gateway@9.13.24.7>;tag=1EDB2D94-11DD
```

continues

Example 7-5 *Normalized SIP INVITE (continued)*

```
To: <sip:222200002O@9.13.32.240;phone-context=gateway>
Date: Thu, 30 Aug 2007 07:04:36 GMT
```

Digit Manipulation

Another technique to suppress or change nonpublic numbers from exiting your network is to use digit manipulation techniques at the border of your network. For example, a non-DID number can be changed to your organization's basic public PSTN number if the call should go off-net.

SIP Privacy Methods

Various SIP specifications control the privacy of end user information in SIP messaging such that numbers and names can travel in the messaging but still be suppressed from delivery or display to the destination endpoint. Similar methods exist in ISDN when interconnecting to the traditional PSTN.

SIP specifications (and CUBE capabilities) of interest in this area include

- The Privacy SIP header (RFC-3323) provides guidelines for withholding the identity of a person (and related personal information) from one or more parties in an exchange of SIP communications.
- The P-Asserted-Identity (PAI) and P-Preferred-Identity (PPI) (RFC-3325) headers provide extensions that enable the communication of the identity of authenticated users and the application of existing SIP privacy mechanisms to communicating these identities.

If your applications are not SIP-capable, or if they do not insert these headers, you can have your border element insert (or change) the content of these headers as a call leaves your premises over the SIP trunk. The CUBE can also convert between the widely deployed Remote-Party-ID (RPID) header to and from PAI/PPI and Privacy headers.

Registration and Authentication

You can use SIP mechanisms to validate the originator of a SIP call and therefore provide a mechanism to reject SIP INVITEs that come from rogue endpoints. These mechanisms include

- **Registration:** Some service provider SIP trunk offerings include a registration sequence enabling the enterprise edge to register explicitly with the provider's SIP softswitch. Some SIP applications are capable of this; if not you can have your CUBE do the registration on behalf of the endpoints behind it in the enterprise network.
- **Digest Authentication (RFC-2617):** A SIP softswitch can challenge the INVITEs, and the originator must respond with credentials that are then authenticated by the SIP softswitch. Unlike a SIP registration sequence that happens once, the Digest

Authentication happens on every SIP INVITE. The CUBE can respond to Digest Authentication challenges with configured credentials.

Example 7-6 shows sample commands to configure the CUBE to do a SIP registration with credentials, and Example 7-7 shows the configuration for SIP Digest Authentication.

Example 7-6 *SIP Registration*

```
x(config)#sip-ua
x(config-sip-ua)#credentials username 1001 password cisco realm cisco.com

sip-ua
  registrar ipv4:172.16.193.97 expires 3600
  credentials username 1001 password 0822455D0A16 realm cisco.com
```

Example 7-7 *SIP Digest Authentication*

```
sip-ua
  authentication username xxx password yyy
```

Toll Fraud

Toll fraud has existed for as long as telephone networks have been in operation. This constitutes making unauthorized calls that someone else pays for. The perpetrator can be inside your network (for example, an employee making personal international calls) or an external hacker using your SIP trunk to make calls that your company pays for.

Ensure that whatever measures you took to combat toll fraud in your TDM PSTN access network are also implemented on your SIP trunk PSTN access network. Some of the common CUBE tools that enable you to mitigate toll fraud attacks include

- Use ACLs to enable explicit sources of calls and deny all other traffic.
- Apply explicit incoming and outgoing dial-peers to both Border Element interfaces to control the types and parameters of calls allowed through the network border. If an incoming dial-peer is not found for a call, the system default dial-peer 0 is used enabling all calls; to avoid this, specify explicit incoming dial-peers for valid call flows and deny all other calls.
- Use explicit destination-patterns on dial-peers (try to avoid using .T if you can) to block out disallowed off-net call destinations.
- Use translation rules to ensure only valid calling/called numbers are allowed. This allows you to add access codes dialing to gain entry to certain destinations (for example, international destinations). Your employees know these access codes, but off-net hackers do not.
- Use Tool Command Language (Tcl) or Voice Extensible Markup Language (VoiceXML) scripts to do database lookups or require PINs or authorization codes

for additional validity checks to allow/deny call flows. This method protects against internal fraudulent calls.

- Change the SIP listening port to something other than the default of 5060.
- Close unused H.323 or SIP ports—if your Border Element is connected purely to a SIP trunk, there is no need for the H.323 ports to be open.
- The Class of Restriction (COR) feature restricts call attempts based on both the incoming and outgoing dial-peers matched by the call.

Signaling and Media Encryption

Another area of security to consider is the privacy of communications, that is, how to keep hackers from recording calls or hijacking them and inserting or deleting segments. Several encryption features for voice call flows mitigate these types of attacks. Separate features for protection of the signaling traffic (TCP or UDP) and the media traffic (RTP) exist.

- Signaling encryption can be achieved by IPsec tunnels (both TCP and UDP SIP traffic) or TLS (SIP TCP). You can use TLS just for authentication or also for encryption of the signaling stream.
- You can achieve media encryption with Secure RTP (SRTP) (RFC-3711).

As the media encryption keys are exchanged in the signaling stream, there is no point in encrypting media without also encrypting the signaling. Only encrypting signaling is a valid option.

None of the current SIP trunk offerings in the market include TLS or SRTP as an option. Hopefully this situation will change. The CUBE can convert between encrypted communications (TLS/SRTP) on one side and nonencrypted (SIP/RTP) on the other side, so if your business can benefit from (or demands) encryption in the enterprise, you can still connect to a SIP trunk provider.

Session Management, Call Traffic Capacity, Bandwidth Control, and QoS

Managing simultaneous voice call capacity and IP bandwidth use is essential for providing consistent quality in enterprise communications. Areas regarding session management and CAC to be considered in the design of your network include

- Trunk provisioning
- Bandwidth adjustments and consumption
- Call admission control