CHAPTER 4

Planning and Design

This chapter addresses planning and design considerations for enabling V^{3}PN. It reviews issues and design considerations specific to IP Telephony, QoS and IPSec. Specifics on product selection for branch and head-end devices are also provided for review. An overview on service provider considerations is also provided. The following specific planning and design sections are presented in this chapter:

This chapter ends with the "Design Checklist" section on page 4-35 to further facilitate V^{3}PN planning.

IP Telephony (Voice over IP)

In this solution, IP Telephony can be thought of as an application transported on a site-to-site IPSec VPN. As such, it is an application with special requirements – more stringent that most data applications – and thus bears special consideration. These requirements include:

- Packets arrive at a constant rate (assuming no VAD)[1]
- Packets arrive in *per call* increments (do not have a portion of a call)
- Quality of the call is a function of jitter, latency and packet loss
- Call Admission Control must be addressed—same as with a Frame Relay deployment

For planning purposes the packet arrival rate is assumed to be 50 packets per second (pps), per call. A call is assumed to be 50 pps transmitted and received. During solution testing, jitter, latency and packet loss are monitored and reported in test results as a gauge of expected voice quality. Like IP Telephony

1. The design throughout this document assumes the Voice Activity Detection (VAD) feature of IP Telephony is disabled. VAD has far-reaching implications on a design and resulting voice quality that are beyond the scope of this solution.

deployments over a private WAN, there is still a requirement to provide Call Admission Control (CAC) as additional voice calls (over-subscription) cannot be permitted to impact the voice quality of established calls.

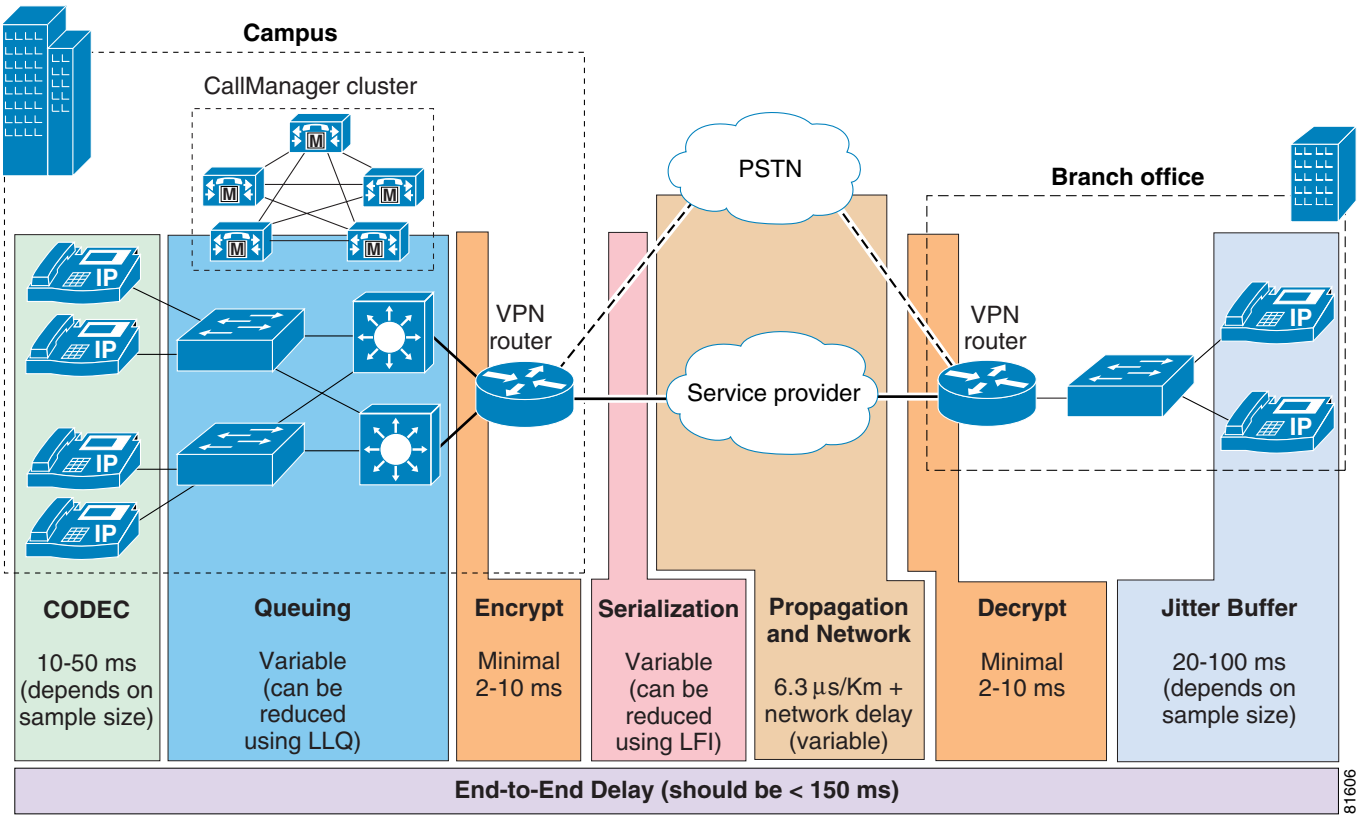
Calculating Delay Budget

One important network design aspect of implementing IP Telephony involves the calculation of the end-to-end, one-way delay budget. Ideally the total budget will fall under 100 msec. At 100 msec, there is the possibility that the phone will be answered but the called party will not hear the caller's greeting. This is referred to as *signal delay*, *voice clipping*, or *voice path cut-through delay.*

The ITU standard G.114 states that a one-way delay budget of 150 msec is acceptable for high voice quality. For most networks, a delay of 200 msec will provide acceptable voice quality with 250 msec as the upper limit.

An additional delay component when running IP Telephony over IPSec VPN is the delay of encrypting and decrypting the voice packets. This is especially important when voice calls are from branch to branch in a hub and spoke environment, as all but the coder and dejitter components are duplicated for each spoke traversed. The specific end-to-end components of the delay budget are shown in Figure 4-1.

Figure 4-1 Calculating End-to-End, One-Way Delay Budget



In most deployments, the addition of a few milliseconds of additional delay for encryption and decryption is insignificant when compared to the total delay budget. For all platforms that support hardware crypto accelerators, they are highly recommended for voice deployments. Software encryption can introduce unacceptable latency and jitter as the CPU becomes fully utilized, which significantly

degrades voice quality. For example, issuing a *copy running start (write memory)* can cause a CPU spike for a few seconds and degrade voice quality. Hardware crypto accelerators help minimize intermittent voice quality issues associated with software crypto and CPU spikes.

Note For planning purposes, 2-to-5 msec of additional delay is added for encryption and decryption, under normal conditions (no over-subscription of the crypto engine). Refer also to the "Crypto Engine QoS" section on page 4-19.

One additional consideration is the codec/sampling scheme being deployed. This delay component can generally be quantified by considering the sampling duration plus any compression delay. For example, G.711 with 20 msec sampling has a delay of approximately 20 msec—only the sampling delay plus less than one msec of encoding and processing delay. In comparison, G.729 with 20 msec sampling has a delay of approximately 25 msec, the sampling delay of 20 msec plus approximately 5 msec for compression, encoding, and processing.

In the Cisco Enterprise Solutions Engineering lab testing, Chariot endpoints report the end-to-end delay of the voice (RTP) streams, as well as jitter and packet loss. Chariot reported values do not include encoding, packetization and dejitter buffer delay. For testing purposes, target threshold values reported by Chariot were as follows:

- Voice jitter—Under 10 msec
- Voice delay—Under 50 msec
- Voice loss—Under 0.5 percent loss

See chapter "Scalability Test Methodology" section on page 5-2 for more information regarding the test methodology and thresholds established for product performance determination.

Note It is also important for Call Signaling packets to experience minimal delay across the network, or call setup issues can result.

Hub-to-Spoke versus Spoke-to-Spoke Calling

Hub and Spoke topologies are prevalent in most enterprise networks. A typical design includes one or two corporate data centers and perhaps a *hot* business recovery site. The remote locations are tied into both data centers and either a third consolidation point connects to the business recovery site or the remote offices have a direct third connection to the business recovery site. Traffic flow, sources and sinks of data, are to and from the data centers and individual remote locations. There is very little traffic volume between the remote offices, little spoke-to-spoke communication. This model is true of hospitality/hotel enterprises, banking, retail or any service oriented business that has large numbers of branch locations which operate in a fairly standalone fashion.

The Cisco Enterprise Solutions Engineering lab testing mimics this design, as a major focus is to identify the scalability of the head-end routers in terms of number of branch locations at various WAN rates.

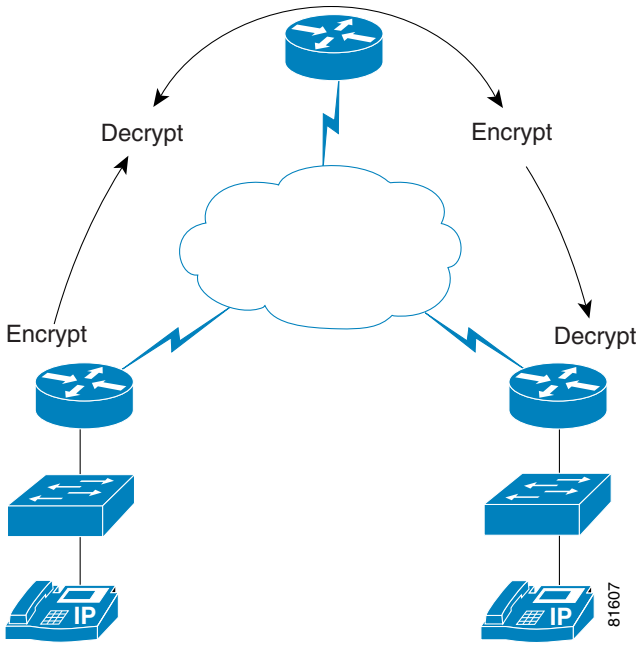
However, the addition of VoIP as an overlay application can change the hub to spoke paradigm. Consider an implementation where branch locations are now individual knowledge worker's home offices and whole workgroups or teams are working remotely. VoIP traffic might be spoke-to-spoke, while the data traffic might continue to be hub-to-spoke.

Figure 4-2 illustrates this situation. In this case, spoke-to-spoke IP Telephony must be considered in regards to the delay budget. The *coder* and *dejitter* values of the delay budget do not change, but the intermediate delay components discussed in the "Calculating Delay Budget" section on page 4-2 are now

duplicated: spoke to head-end, and head-end to spoke. In this example, the VoIP packets are encrypted at the first branch, decrypted at the head-end, routed to another tunnel interface, encrypted again and then decrypted at the receiving spoke. All of the queuing, serialization and network delay is also present on both *legs* of the call.

Both Cisco Enterprise Solutions Engineering lab testing and internal Cisco deployments have demonstrated encrypted spoke-to-spoke voice calls are both feasible and practical provided the overall delay budget is within tolerances.

Figure 4-2 Spoke to Spoke Calling



Dynamic Multipoint VPN (DMVPN) provides an alternative to hub-and-spoke topologies, essentially providing a virtual fully-meshed VPN. This feature can be of great benefit in V^{3}PN designs by allowing a dynamic tunnel to be established directly between the two branch locations. More information regarding DMVPN will be added to this document in a future revision.

Cisco IP Softphone

Cisco IP Softphone is a Windows-based application for the PC. During solution testing a Cisco IP Softphone was included in the test bed and calls were placed between the Softphone and a 7960 phone, from branch to campus. For the purposes of this solution, there are no special considerations which must be addressed if Softphones are deployed in addition to, or in place of, 7960 phones. The Cisco IP Softphone provides the same IP Precedence/DSCP markings as a 7960 for the voice bearer and call setup streams.

Note It should be noted that the Softphone application running on a laptop might be a *best effort* implementation, because the laptop operating system has no provision for QoS. Therefore, it is possible that other applications on the laptop can interfere with voice quality.

Quality of Service (QoS)

IP Telephony deployments have been the catalyst for enhancements in the development and deployment of QoS services in today's networks. Implementing V^{3}PN introduces two new aspects to the traditional QoS implementation to support VoIP:

- Encryption increases the bandwidth requirements of both voice and data, impacting the provisioning of service policies on output interfaces
- Encryption provides confidentiality of portions of the original IP packet that previously were referenced by the output QoS service policy.

These two key QoS issues—and a review of important QoS concepts implemented in this design guide—are addressed in the subsequent sections.

Bandwidth Provisioning for WAN Edge QoS

This section details the bandwidth requirements of encrypted voice, how to provision the enterprise WAN edge to ensure encrypted voice quality.

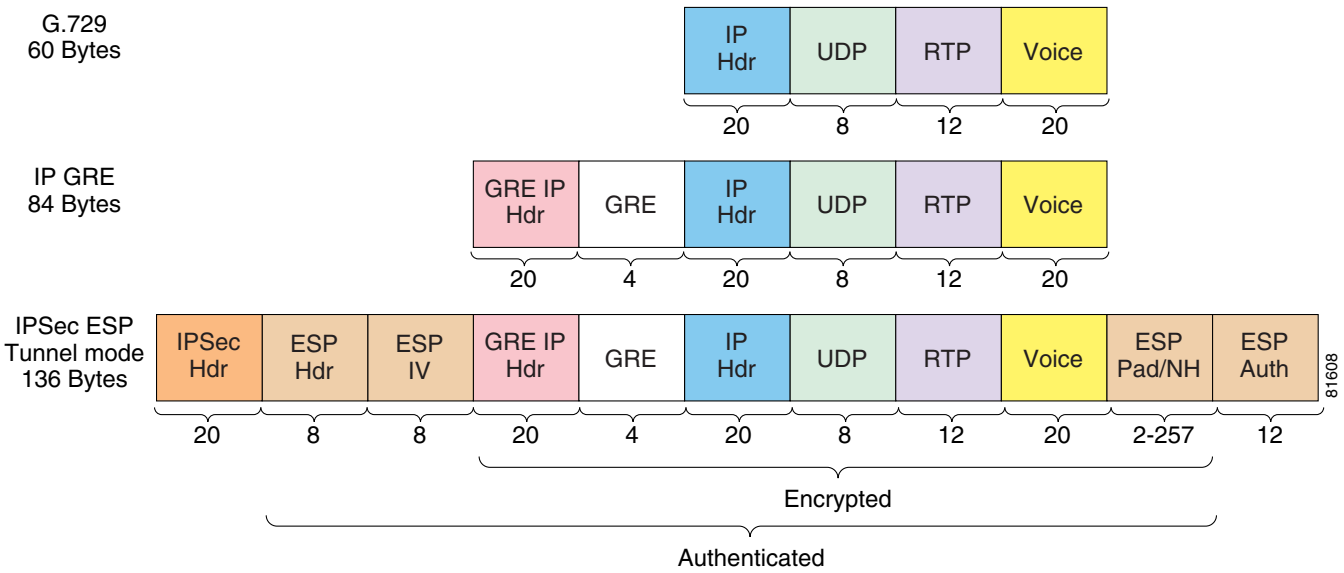
Packet Size—IPSec Encrypted G.729

The Layer 3 data rate for a G.729 call (50 pps) is 24 Kbps. Encrypting that packet using IPSec Tunnel mode for IP GRE increases that rate to approximately 56 Kbps (in each direction). The calculation is as follows:

- **136 bytes** per packet at 50 packets per second = 6,800 bytes or **54,400 bits per second**

The 136-byte packet's header, data and trailer fields are shown below in Figure 4-3.

Figure 4-3 IPSec Encrypted G.729 Packet Anatomy



Start with a 60-byte G.729 IP/UDP/RTP voice packet. IP GRE adds 24 bytes, including a new IP header and GRE encapsulation.

The ESP header contains a 4-byte Security Parameter Index (SPI) field and the 4 byte sequence number. This sequence number is used by ESP authentication—the anti-replay logic.

ESP might add up to 255 bytes of padding. DES is a block cipher, encrypting blocks of 8 bytes (64-bits) at a time, and thus the need for the encryption algorithm to add padding to the plain text. The ESP Authentication Data field must align on a 4-byte boundary. The ESP Pad length field is 1 byte and starts at the third byte of a 4-byte word, and the ESP Next Header field occupies the forth byte. The Next Header field is used to identify the payload's protocol. The ESP Authentication Data field contains either the MD5 16 byte hash or the SHA-1 20 byte hash, both truncated to 12 bytes. See RFC 2104 regarding truncation of the hash value.

The ESP IV (Initialization Vector) ensures the uniqueness cipher text if the same plain text characters are encrypted in different blocks or messages. It is used by block chaining ciphers like DES. The ESP IV byte count can be determined from the command **show crypto ipsec sa** for the security association (SA). The ESP IV field is considered part of the payload and might not be shown by a protocol analyzer. This is also true of the trailer fields: ESP Padding; Pad Length; Next Header; and Authorization.

An increase of the original packet by one byte might result in no increase in the resulting encrypted packet, or it might increase more than one byte. NetFlow can be used to verify the Layer 3 length of the encrypted and decrypted packet sizes. See the "Using NetFlow to Verify Layer-3 Packet Sizes" section on page 7-5 for an example.

In this example, IPSec adds 52 bytes to the IP GRE packet, so the resulting packet combined with IP GRE and IPSec is 136 bytes.

The G.729 codec family is specified by these four specific designations: g729r8, g729ar8, g729br8, and g729abr8. All generate the same format code word, but differ in complexity and support of voice activity detection (VAD):

- g729r8—This is the default codec, it and all forms of G.729 generate 8,000 bps
- g729ar8—Codec is a simplified version of g729r8
- g729br8—Same as g729r8 but includes VAD
- g729abr8—Simplified g729r8 but includes VAD

With VAD, silence is not sent over the network- only speech, and therefore the total bandwidth used by voice traffic might be much less that the 50 pps specified. Studies have shown that 35 percent bandwidth savings can be realized by the use of VAD. Voice quality might be degraded slightly and comfort noise should be considered to provide audible feedback to the listener that the other party is still on the call.

However, for bandwidth capacity planning purposes and for testing, assume VAD is disabled and the RTP streams are 50 packets per second continuously.

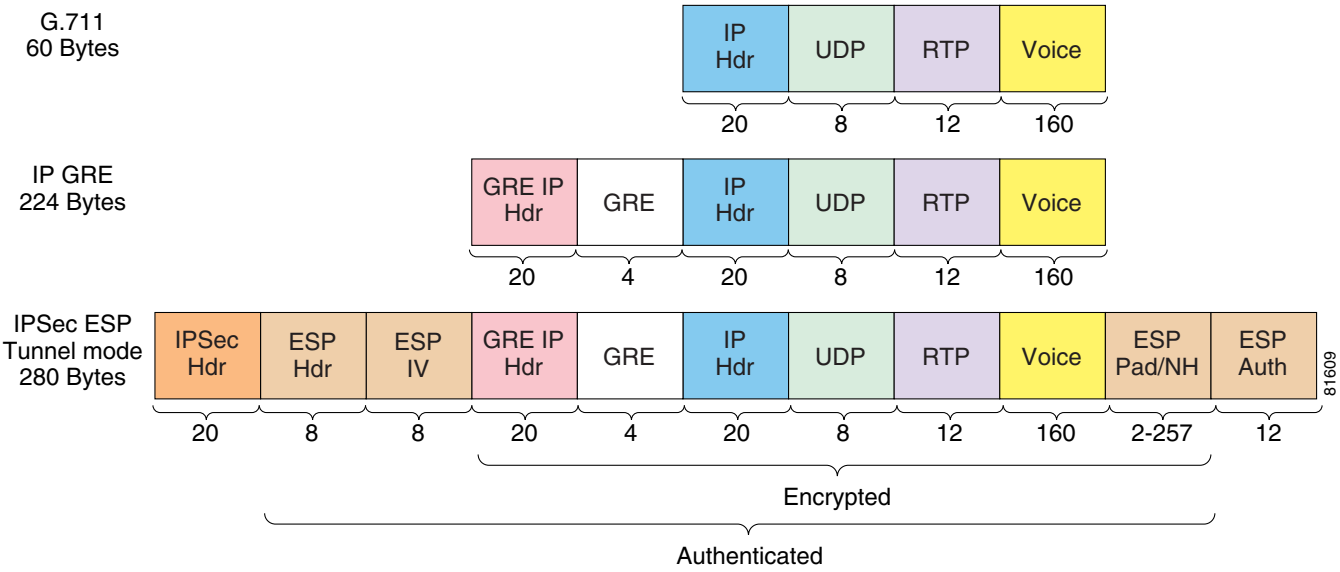
Packet Size—IPSec Encrypted G.711

The Layer 3 data rate for a G.711 call (50 pps) is 80 Kbps. Encrypting that packet using IPSec Tunnel mode for IP GRE increases that rate to approximately 112 Kbps (in each direction). The calculation is as follows:

- **280 bytes** per packet at 50 packets per second = 14,000 bytes or **112,000 bits per second**

The 280 byte packet's header, data and trailer fields are shown in Figure 4-4.

Figure 4-4 IPSec Encrypted G.711 Packet Anatomy



The construction of the G.711 encrypted voice packet is the same as the G.729 example in the"Packet Size—IPSec Encrypted G.729" section on page 4-5. The major difference is that G.711 starts with a 200-byte voice packet. After the addition of IP GRE and IPSec, the packet size becomes 280 bytes.

The codec choices for G711 are g711alaw and g711ulaw and both generate data at 64,000 bits per second, the same data rate as clear-channel. There is no VAD option available with G711.

Packet Size—Layer 2 Overhead

While Layer 2 headers do not increase the original voice packet size to the extent of adding IP GRE and IPSec headers and trailers, they must be included in the calculation for priority (LLQ) queue in the voice class of the CBWFQ policy map. Some common Layer 2 encapsulations are shown in Table 4-1.

Table 4-1 Summary of Common Layer 2 Encapsulations

L2 Encapsulation	Bytes to include bandwidth calculation
Ethernet	14 bytes
Frame Relay	4 bytes[1]
PPP	4 bytes
Multilink PPP	10 bytes
HDLC	4 bytes
ATM	53-byte cell, 48 bytes of payload

1. Frame Relay flags do not need explicit provisioning for the LLQ. For non-fragmented (voice) frames 4 bytes per frame Layer-2 overhead; for fragmented (non-LLQ packets over the frame-relay fragment size) frames, 6 bytes.

To continue the calculation of the G.729 voice packet with IP GRE and IPSec as shown previously, add 4 bytes for the Frame Relay header:

- 136 bytes + 4 bytes = 140 bytes * 50 pps = 7,000 bytes or **56,000** bps

For the G.711 example:

- 280 bytes + 4 byes = 284 bytes * 50 pps = 14,200 bytes or **113,600** bps

These should be considered the minimum values for bandwidth planning.

Jitter in the path of the voice packets can increase or decrease the arrival rate—for short periods of time, the bit per second values can be slightly higher than calculated above. In the organization-specific environment, the service policy should be reviewed for drops in the voice class using this command:

show policy-map interface serial 0/0.100

In the event drops are encountered, increase the **priority** keyword value in the voice class to eliminate voice drops. When the interface is not congested, the priority (voice) class can exceed its bandwidth and not be dropped. When the interface is congested, the offered rate of voice traffic is dropped if it exceeds its allocation.

Special Considerations for Frame Relay Provisioning

The primary configuration used for the scalability and performance evaluation was Frame Relay encapsulation, although HDLC was also evaluated.

To minimize serialization delay on low speed links, Link Fragmentation and Interleaving (LFI) must be implemented by Layer 2. For Frame Relay, this is accomplished by Frame Relay Forum's FRF.12 Implementation Agreement (also known as FRF.11 Annex C).

To reserve bandwidth for, and prioritize voice packets, Class Based Weighted Fair Queuing (CBWFQ) is configured with a priority or low-latency queue (LLQ). The CBWFQ service policy is applied to the Frame Relay Traffic Shaping (FRTS) map-class. FRTS is a prerequisite for FRF.12, but it also provides congestion notification to CBWFQ. CBWFQ allocated bandwidth among the configured classes of traffic during periods of interface congestion.

In this guide, the Cisco IOS Frame Relay Traffic Shaping target CIR values is 95 percent of the carrier's configured CIR. While following this guideline adds an extra step to the planning and configuration process, the conservative approach is to implement this best practice. To make configuration easier, tables have been provided for the link speeds under test.

Bandwidth Allocation by Traffic Category

In order to implement a QoS Service Policy it is necessary to decide how traffic will be sub-divided into various categories and assigned relative priorities for those categories.

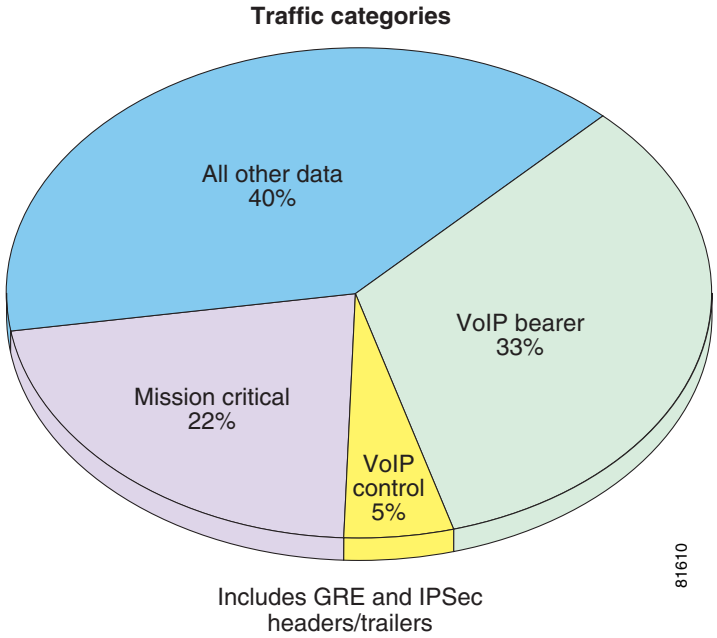
For networks already using QoS, established Service Policies must be considered and might be customized by the enterprise as well.

This design assumes voice and data traffic falls into the categories:

- VoIP control (signaling)
- VoIP bearer (RTP streams)
- Mission critical data traffic
- Other data traffic and overhead

Considerations must be given to the amount of high priority voice traffic allocated on a converged network. This design establishes an upper bound of 33 percent for such traffic, as this places a significant small-packet burden on the VPN. The relative target bandwidth percentages for traffic categories is shown in Figure 4-5**.**

Figure 4-5 Traffic Categorization for Bandwidth Provisioning



Note These percentages (and categories) are not strict design rules, they are provided as a guideline. When the bandwidth of a particular class is not being used, it is available to other traffic categories. These percentages simply call out reserved bandwidth percentages.

These allocations must include:

- Payload (such as voice)
- IP GRE overhead
- IPSec overhead
- Layer 2 encapsulation header overhead

With the addition of IP GRE headers and IPSec headers and trailer, and the Layer 2 header, the per call value using a G.729 codec is 56 Kbps.

Using a 512Kbps link as an example, these traffic categories would be configured as:

```
!
policy-map llq-branch
  class call-setup
   bandwidth percent 5
  class mission-critical
   bandwidth percent 22
  class voice
    priority 168
  class class-default
   fair-queue
!
```

Voice bandwidth must be configured in per call increments (such as 56 Kbps). Therefore, for lower speed links, the allocation exceeds the 33 percent target.

Class-default queues firs-in-first-out (FIFO) unless weighted fair queue (WFQ) is configured. Weighted random early detection (WRED) could also optionally be configured for the mission-critical and class-default. WRED is IP Precedence aware and serves to intelligently drop packets and provide feedback to TCP based applications to reduce their sending data rate. Additionally, WRED configured in a class provides a degree of visibility to the network manager as to the IP Precedence of the traffic that is tail or random dropped.

The key QoS considerations for this configuration are:

- When configuring *percent* in a policy-map, the *percent* value is a percent of the underlying (link or Frame Relay Traffic Shaping MINCIR) bandwidth.
- For serial HDLC-encapsulated interfaces, the sum of the Kbps specified by the classes in the policy map cannot exceed 75 percent of the available bandwidth. This can be manually changed by the **max-reserved-bandwidth** interface command.
- The priority (LLQ) is policed if there is congestion and it exceeds its bandwidth allocation.
- Bandwidth that is not allocated to a class is available to the default class, *class-default*.
- Packets not selected for a class are placed in class-default, regardless of their IP Precedence; class-default is not assumed to be only for packets with IP Precedence of '0'.

It is convenient when configuring a CBWFQ policy-map to specify bandwidth in terms of percentages rather than actual values (in Kbps), since the same **policy-map** can be applied to a range of link speeds without modifications. The sample configuration template for this design guide shows the LLQ specified in Kbps and the *call-setup* and *mission-critical* classes in percent. Using Kbps for the LLQ is easier to illustrate than percentages as bandwidth is allocation in multiples of 56 Kbps per call.

Table 4-2 lists the key QoS and Call Admission Control parameters in this design.

Table 4-2 Bandwidth Allocation Parameters by Link Speed

Line Rate (Kbps)	Maximum Number of G.729 Calls	Max G.729 Calls as Percentage of Line rate	Priority 56k per Call (Kbps)	Call Setup 5 percent (Kbps)	Mission Critical 22 percent (Kbps)	Max-reserved-bandwidth
64	1	**87.5**	56	3	**None**	**100**
128	1	**43.7**	56	6	26	75
256	2	**43.7**	112	12	53	75
512	3	33	168	24	106	75
768	4	29	224	36	160	75
1024	6	33	336	48	213	75
1536	9	33	504	72	320	75
2048	12	33	672	102	450	75

These values should be substituted into the configuration sample to create the branch router configuration, based on the line rate that services the remote location. For branches that connect at 64 Kbps, 128 Kbps and 256 Kbps, the voice traffic exceeds the target of approximately 33 percent voice traffic on the link. These are highlighted, as is the absence of a *mission critical class* and override of the *max-reserved-bandwidth* for 64 Kbps links.

When planning the bandwidth required for a branch office, consider the number of concurrent calls traversing the WAN this branch is expected to make during peak call periods. This varies based on the job function of the employees located at a branch. For example, an office of software engineers would

be expected to make fewer calls than an office of telemarketers. One rule of thumb is one call for every six people (1:6), but this could range from 1:4 to 1:10. Given the 512 Kbps link as an example, with a target of 3 G.729 Calls, that link could theoretically support between 12 and 30 people.

If the maximum number of calls is not active on the link, the bandwidth is available for data traffic. If there is a misconfiguration in the call admission control for the branch and more calls are attempted (the voice class exceeds its calculated data rate), CBWFQ will police (drop) packets during congestion and all calls will exhibit poor voice quality. Either CallManager "Locations" or Gatekeeper Call Admission Control, must be implemented to guarantee voice quality.

Campus QoS

Refer to the recommendations on Campus Switching Designs for Cisco AVVID. Campus QoS techniques have been well documented and are not included in this design guide.

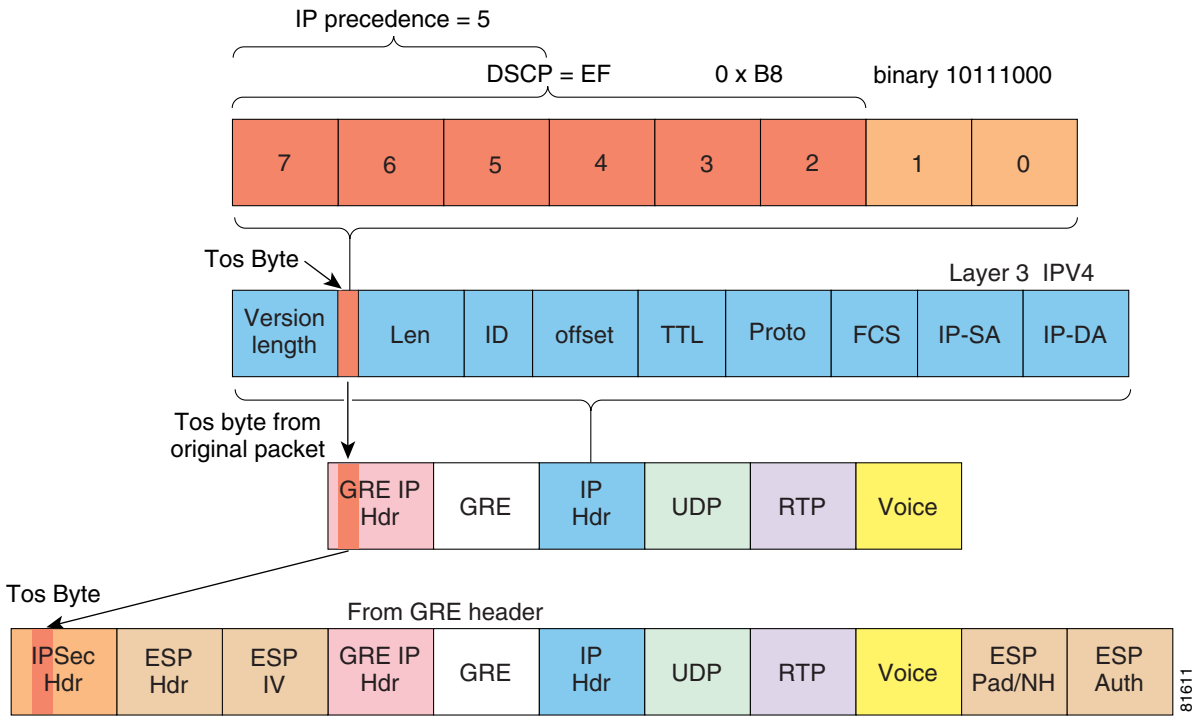
ToS Byte Preservation

In a typical IP Telephony deployment, QoS would be enabled, classifying and marking higher priority traffic (such as voice and H.323 signaling). The ToS byte is commonly used for this purpose.

When encrypting a voice stream the Service Policy can no longer make decisions on the original IP header information, because the original IP header is now encrypted.

However, built into the IPSec protocol standard is the ability to preserve the ToS byte information from the original IP header by automatically copying it to the IP header added by IPSec, so the information is still available for use by Service Policies. Figure 4-6 illustrates this process.

Figure 4-6 IPSec Preserves the ToS Byte



IPSec—RFC 2401 specifies the ToS byte must be copied from the inner header to the outer header. See section *5.1.2.1 IPv4—Header Construction for Tunnel Mode* in the following document: http://www.ietf.org/rfc/rfc2401.txt

Similarly, with IP GRE, as of Cisco IOS software release 11.3, the ToS byte from the original IP header is copied to the IP GRE header. See the following reference for more information regarding IP GRE: http://www.cisco.com/en/US/docs/ios/11_3/feature/guide/greqos.html.

Also see the "QoS Pre-Classify" section on page 4-12 for more information regarding the interaction of IPSec and QoS Service Policies.

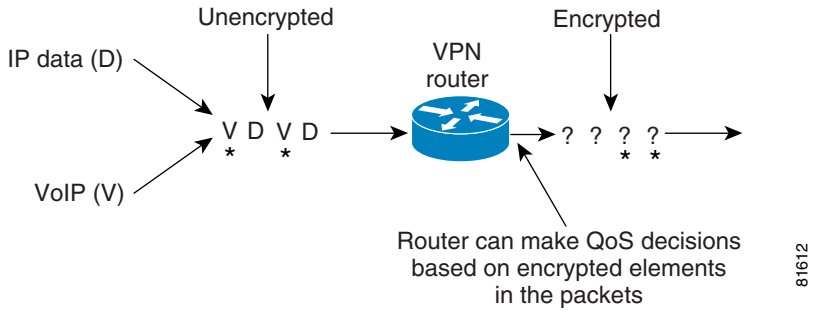
QoS Pre-Classify

QoS Pre-Classify is often confused with the preservation of the ToS byte (see the "ToS Byte Preservation" section on page 4-11) during the packet encryption process. IPSec (and IP GRE) preserve the ToS byte automatically.

In Cisco AVVID solutions, the IP Phone and gateways provide the capability to set the ToS byte so routers can make the appropriate QoS decision. However, most data applications do not set the ToS byte and queuing decisions must be based on other fields of the IP header, including source/destination IP address, port numbers, and protocol.

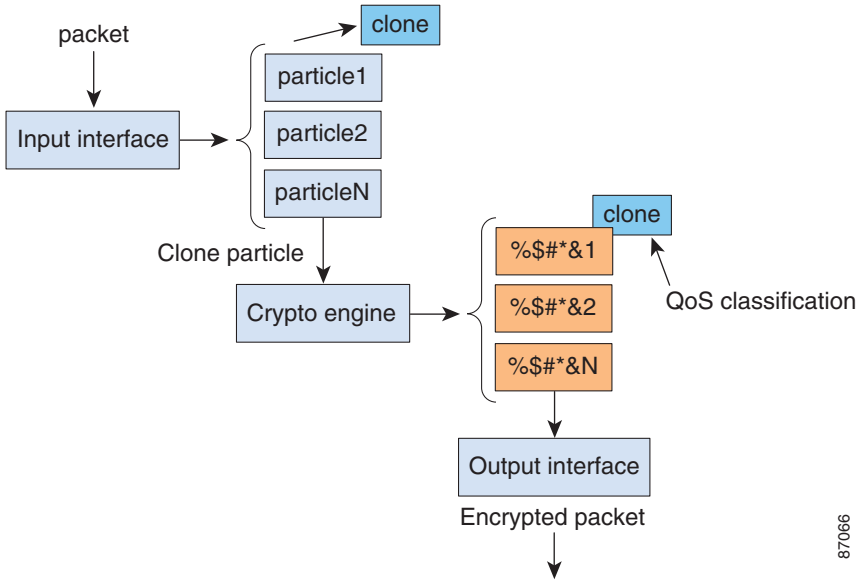
Once the original IP packet is encrypted by IPSec, fields other than ToS byte, such as port numbers, protocol and source/destination IP address fields, are no longer in clear text and cannot match an output service policy. QoS Pre-Classify is an Cisco IOS software feature to allow *fancy queuing*, CBWFQ/WFQ, at the output interface to match on these other fields in the original IP header, even after the original IP header is encrypted. Figure 4-7 illustrates this concept at a high level.

Figure 4-7 QoS Pre-Classify Feature



A key point to remember regarding QoS Pre-Classify is that is only applicable at the encrypting router's output interface. The fields preserved by QoS Pre-Classify are not available to router's beyond the encrypting router. A simplified illustration is shown in Figure 4-8.

Figure 4-8 How QoS Pre-Classify Works



The following stages summarize the QoS Pre-Classify process illustrated in Figure 4-8:

1. A packet enters the input interface and is stored in particles in a given pool.
2. When the packet matches a crypto map on the interface it would be switched out of, it is passed to the Crypto Engine
3. As part of the Crypto Engine's TX ring processing, a *clone* of the particle, including the IP Header, is created and associated with the packet's data structure.
4. The Crypto Engine encrypts the original packet and places the cipher text in new particle(s), associating the *clone* particle with the new particle(s).
5. If the output interface is not congested, the encrypted packet is simply transmitted.

   If the output interface is congested and must be queued for QoS features, the classification can act on the *clone* particle (still in clear text) to match on protocol, source/destination IP address, and port numbers.

Cisco recommends enabling QoS Pre-Classify on all platforms. See the "Configuring QoS Pre-Classify" section on page 6-23 for more information regarding configuration of QoS Pre-Classify.

IP Security (IPSec)

This section outlines the IPSec design options for consideration. It discusses *tunnel* options, and in this design guide, a *tunnel* is specified as an IP GRE tunnel, or an IPSec tunnel. While this document does not go in depth to firewall placement, an example of securing a branch router with access-lists is provided. There is also a discussion on anti-replay and crypto engine QoS.

IPSec and GRE Tunnel Design Considerations

There are currently three recommended design options for a site-to-site IPSec VPN:

- IPSec Tunnel mode—no IP GRE tunnel

- IPSec Transport mode encrypting an IP GRE tunnel
- IPSec Tunnel mode encrypting an IP GRE tunnel (primary recommendation)

This design guide implements IP Tunnel mode encrypting an IP GRE tunnel. The advantages, disadvantages and features and limitations of these options follow.

- **IPSec Tunnel mode—no IP GRE tunnel.** This option does not utilize a IP GRE tunnel. IPSec encrypts IP unicast traffic only, IP Multicast traffic cannot be transported between the IPSec peers without configuring an IP GRE tunnel. This configuration might be sufficient to support the application requirements and its advantage lies in less CPU overhead (primarily at the head-end router) to maintain a IP GRE tunnel to each remote location and a routing protocol's hello and update packets. IPSec security associations are created for each access list line matched. An access list must be specified in the crypto map to designate packets that are to be encrypted. The access list (when encrypting an IP GRE tunnel) is only one line, a match on protocol 47 (GRE) and the source and destination IP address of the GRE endpoints. When not encrypting a GRE tunnel, it is possible to create an access list which has multiple lines, matching on various portions of the five tuples, source/destination IP address, protocol, source/destination port numbers. A separate security association is created for each access list line match. Each security association has its own ESP (or AH) sequence number. Anti-replay drops can be eliminated or minimized by constructing access lists that create a separate security association for each class of traffic being influenced by per-hop QoS policy.

  The *Pre-fragmentation for IPSec VPNs* feature is supported in IPSec Tunnel mode – no IP GRE tunnel, it is first available in Cisco IOS software release 12.1(11)E and is targeted for 12.2(12)T.

- **IPSec Transport mode encrypting an IP GRE tunnel.** This option is commonly implemented; for a G.729 packet it saves 16 bytes per packet over IP GRE tunnels with IPSec Tunnel mode, as an additional IPSec IP header is not required. This byte count would normally be expected to be 20 bytes, the length of an IP header, but 16 bytes as verified by a protocol analyzer, the 4 byte delta would be explained by a different padding length. The IPSec peer IP addresses and the IP GRE peer address must match for transport mode to be negotiated, if they do not match, tunnel mode is negotiated. The *Pre-fragmentation for IPSec VPNs* feature is **not** supported for Transport mode, as the decrypting router cannot determine if the fragmentation was done prior to encryption or post-encryption by downstream router between the encrypting and decrypting router.

  IPSec Transport mode saves link bandwidth, but it does not provide any reduction in packets per second switched by the router. In most instances, packets per second, not packet size, is the limiting factor of a router's main CPU performance.

  IPSec Tunnel mode is the default configuration option. To configure Transport mode, it must be specified under the transform set:

```
!
crypto ipsec transform-set vpn-test esp-3des esp-sha-hmac
 mode transport
!
```

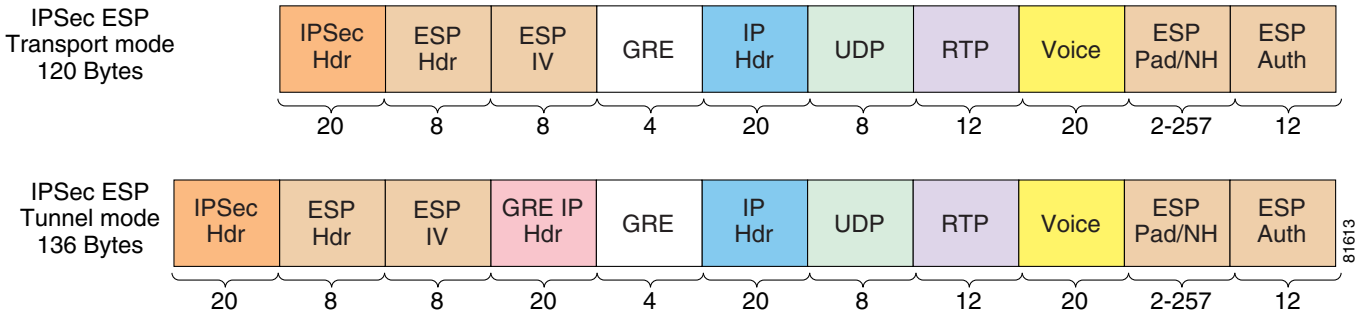
  Configuring IPSec Transport mode to encrypt an IP GRE tunnel provides all the advantages of using IP GRE—it supports IP Multicast, routing protocols and multi-protocol support.

- **IPSec Tunnel mode encrypting an IP GRE tunnel.** This option is implemented in this design guide and in the associated lab testing. It incurs the greatest header overhead of the three options, but it is capable of supporting IP Multicast and the ability to run a dynamic routing protocol within the IP GRE tunnel for failover to an alternative path. It supports Pre-fragmentation for IPSec VPNs. This option was selected for lab testing as it provides the greatest features and flexibility as well as the worst-case scenario in our performance testing in regards to bandwidth consumption.

When configured with a routing protocol running within an IP GRE tunnel, the routing protocol's hello packets maintain the security associations between both (assuming a redundant configuration) head-end routers. There is no need to create a security association to a back-up head-end peer upon failure of the primary peer. Also, routing protocol hello timers (5 seconds by default for EIGRP) can be tuned lower than the hello interval of Internet Security Association and Key Management Protocol (ISAKMP) keepalives—the minimum value is 10 seconds. Detection of a failed head-end peer is quicker when using a routing protocol verses `crypto isakmp keepalive 10`—the dead interval for ISAKMP keepalive is 3 times the keepalive value, or 30 seconds. EIGRP has a default dead interval of three times the hello value of 5 seconds, or 15 seconds.

The diagram in Figure 4-9 illustrates the difference in packet size between IPSec Transport and Tunnel modes.

Figure 4-9 IPSec Transport vs. Tunnel Mode for G.729 Packets



Each option discussed has merit, there is no one design option that is superior to the other alternatives. While this design guide implements IP Tunnel mode encrypting an IP GRE tunnel, internal Cisco deployments of voice over an IPSec Tunnel with no IP GRE tunnel have also been successfully implemented.

Firewall Considerations for Transport of VoIP

Firewall placement at the head-end site depends on the enterprise's security policies. Placing a firewall between the head-end crypto/IP GRE tunnel termination routers and the enterprise network allows the firewall security administrator visibility to the specific port and protocols, as the traffic is unencrypted at that point in the topology.

Placing a firewall between the remote routers and the head-end crypto/IP GRE tunnel termination routers prevents visibility to the specific applications because all traffic is encrypted. IPSec ESP (protocol 50) and UDP port 500 for ISAKMP must be permitted and are the only packets visible to the firewall.

The Cisco IOS Firewall feature set was not tested as part of the Cisco Enterprise Solutions Engineering lab verification of this design as it applies to a split tunneling configuration.

Since this design routes all traffic to the head-end, access to the remote routers from the WAN can be limited by inbound access-lists on the serial interfaces to permit only ISAKMP (UDP port 500) and IPSec ESP (protocol 50) and specific access from the from the head-end routers for management purposes. In this example ICMP is permitted from the upstream router's serial interface but all other ICMP and IP accesses is denied.

```
!
crypto map GRE local-address Loopback0
crypto map GRE 30 ipsec-isakmp
 set peer 192.168.3.1
...
!
interface Loopback0
 ip address 192.168.2.1 255.255.255.252
!
interface Tunnel0
 ip address 10.0.96.1 255.255.255.0
 qos pre-classify
 tunnel source Loopback0
 tunnel destination 192.168.3.1
 crypto map GRE
!
interface Serial0/0.100 point-to-point
 ip address 192.168.1.1 255.255.255.252
 ip access-group 2699 in
 frame-relay interface-dlci 100
  class ts-branch
 crypto map GRE
!
access-list 2699 permit udp host 192.168.3.1 eq isakmp host 192.168.2.1 eq isakmp
access-list 2699 permit esp host 192.168.3.1 host 192.168.2.1
access-list 2699 permit icmp 192.168.1.0 0.0.0.255 192.168.1.0 0.0.0.255
access-list 2699 deny   ip any any log-input
!
```

This is for illustrative purposes only. In the above example, console access would need to be provided by a modem for remote management of the device in the event of a failure, as telnet would be denied to the serial interface's IP address.

Anti-Replay Considerations

IPSec offers message integrity, providing for a means to identify if an individual packet is being replayed at a later time. This concept is called connectionless integrity. It also provides for a partial sequence integrity, preventing the arrival of duplicate packets.

These concepts are outlined in RFC 2401 at ftp://ftp.isi.edu/in-notes/rfc2401.txt.

When ESP Authentication (esp-sha-hmac) is configured in an IPSEC transform set, for each security association, the receiving IPSec peer verifies that packets are received only once. Because two IPSec peers can send millions of packets, a 64-packet sliding window is implemented to bound the amount of memory required to tally the receipt of a peer's packets. Packets can arrive out of order, but they must be received within the scope of the window to be accepted. If they arrive too late (outside of the window), they are dropped.

The operation of the anti-replay window protocol is as follows:

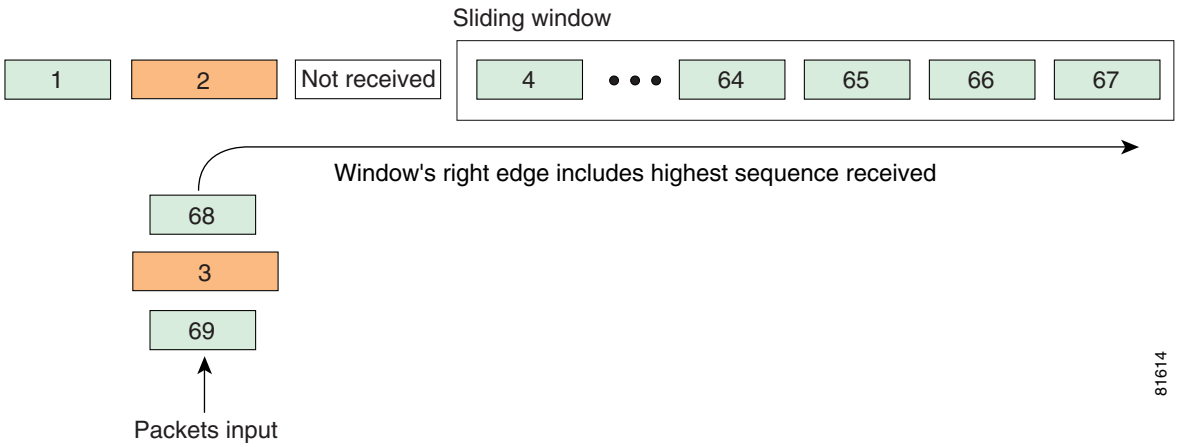
1. The sender assigns a unique sequence number (per security association) to encrypted packets.
2. The receiver maintains a 64-packet sliding window, the right edge of which includes the highest sequence number received. In addition, a boolean variable is maintained to indicate if each packet in the current window was received or not.
3. The receiver evaluates the received packet's sequence number:
   - If a received packet's sequence number falls within the window and it has not been previously received, the packet is accepted and marked as received.

- If the received packet's sequence number falls within the window and was previously received, the packet is dropped and the replay error counter is incremented.
- If the received packet's sequence number is greater than the highest sequence in the window, the packet is accepted, marked as received, the sliding window is moved "to the right."
- If the received packet's sequence number is less than the lowest sequence in the window, the packet is dropped and the replay error counter is incremented.

In a converged network implementation with QoS enabled, lower priority packets are delayed such that higher priority packets receive preferential treatment. This has the side affect of also reordering the packets to be out-of-sequence from an IPSec sequence number perspective. Therefore, there is a concern that through the normal QoS prioritization process, packets will be dropped by the receiver as replay errors, when in fact they are legitimately sent/received packets.

Figure 4-10 a visualization of the process. In this example, voice packets 4 through 67 have been received, our data packet "3" was delayed by the queuing process and was transmitted following voice packet "68". When the anti-replay logic is called to process packet "3", it will be dropped, since it will be outside the left edge of the sliding window. Packets might be received out of order, but they must fall within the window to be accepted.

Figure 4-10 Anti-replay Operation



In a converged network of voice and data, anti-replay drops impact data packets rather than voice packets, the QoS configuration prioritizes voice over data. Voice quality should not be impacted by anti-replay.

Anti-replay drops can be eliminated in a pure IPSec configuration (no GRE) by creating separate security associations for voice and data; voice and data packets must match a separate line in the access-list referenced by the crypto map. This is easily implemented if the IP Phones are addressed by a separate network address (RFC 1918 addresses) than the workstations.

Consider the effect of packet loss on a TCP based application. TCP is connection oriented and incorporates a flow control mechanism. The TCP application has no visibility to the reason a packet was dropped. A packet lost by a service policy on an output interface is not different to the application than a packet lost by an anti-replay drop. From a network perspective it would be more efficient to drop the packet before sending it over the WAN link, but the location or nature of the packet loss is immaterial to the TCP driver. Anti-replay drops can be readily created in a lab environment by applying a QoS policy to an output interface and using a traffic generation tool to persistently congest the link with connectionless traffic.

This traffic profile, however, is not representative of most production networks. Table 4-3 represents the traffic profile of a major hospitality organization's core network during a mid-afternoon weekday. NetFlow was enabled on core routers between the organization's branch locations (hotels) and their data center. Voice is not enabled on this network. The primary function of the network is to provide reservation information between the remote hotel and the headquarters data center. DLSw, TN3270, and in-house developed client/server applications are used to access the reservation data on the mainframe. Email and Web applications are also in use.

Table 4-3 Traffic Profile for Major Hospitality Organization

Protocol	Percent Bytes To Branch	Percent Packets To Branch	Percent Bytes To Head-end	Percent Packets To Head-end
ICMP	1.2	1.4	2.2	1.1
UDP	5.6	7.9	10.8	8.1
TCP	93.2	90.7	87.0	90.8

From Table 4-3, TCP is the predominate protocol on this network. The average number of bytes per packet to the branch was 332 bytes and the average bytes per packet to the data center was 159. The traffic profile implemented in testing this design guide is similar to this network, with the addition of approximately 33 percent voice traffic in the total profile. Adding VoIP increases the percentage of UDP packets in the profile, however the data portion continues to be predominately TCP based.

During testing by Cisco's Enterprise Solutions Engineering lab, using the traffic provide documented in this design guide, voice traffic was not adversely affected by anti-replay drops and data drops were typically less than one percent of the total packets decrypted by the receiving router. This drop rate was determined to not adversely impact the function of the network.

Output drops on the output WAN interface tend to be few, if any, and certainly far less than dropped by anti-replay. Anti-replay triggers packet drops more aggressively than the output service policy. This relates to the default size of the output queues and the number of defined classes. In the sample output service policy below, note that each bandwidth class and the class-default can queue a maximum of 64 packets.

```
vpn18-2600-2#show policy-map
  Policy Map llq-branch
    Class call-setup
      Weighted Fair Queueing
            Bandwidth 5 (%) Max Threshold 64 (packets)
    Class mission-critical
      Weighted Fair Queueing
            Bandwidth 22 (%) Max Threshold 64 (packets)
    Class voice
      Weighted Fair Queueing
            Strict Priority
            Bandwidth 168 (kbps) Burst 4200 (Bytes)
    Class class-default
      Weighted Fair Queueing
            Flow based Fair Queueing
            Bandwidth 0 (kbps) Max Threshold 64 (packets)
```

However, the receiving IPSec peer has a single 64-packet anti-replay window (per IPSec Security Association), with which to process all packets from the above priority class (voice), bandwidth classes (call-setup, mission-critical, and internetwork-control), and the default class (class-default).

So it stands to reason, anti-replay will be more aggressive than the service policy at dropping packets delayed by voice—due to the size mismatch of the queue depth on output verses the width of the anti-replay window. As more bandwidth classes defined in the policy map, this mismatch increases.

By reducing the queue-limit (Max Threshold) of the bandwidth classes the output service policy becomes more aggressive at dropping packets rather than buffering/delaying them—this further reduces the number of anti-replay drops. The default value of 64 packets is designed to absorb bursts of data traffic and delay, rather than drop, those packets. This is optimal behavior in a non-IPSec enabled network.

With IPSec authentication configured (esp-sha-hmac) in the network, decreasing the queue-limit for the bandwidth classes further reduces anti-replay drops from the 0.5-to-1.5 percent range to tenths of a percent. These configuration changes increase the number of packet drops by the sender's output service policy.

In lab testing using NetIQ Chariot™ with voice plus predominately TCP based data traffic profile, decreasing the queue-limit from 64 to 16 for the critical data traffic, and to 6 for the class-default traffic has shown to decrease anti-replay drops to the tenths of a percent range, while of course, increasing drops on the output interface. The **policy-map** command configuration shown below can be used as a starting point and then tuned by the network manager. Decreasing the queue-limit causes the service policy to be more aggressive in dropping rather than delaying packets and decreases the number of anti-replay drops.

Note The queue limits shown in the following example were tuned for a specific traffic profile and a specific link speed (512 Kbps). They might not apply for different traffic profiles or link speed.

```
policy-map llq-branch
  class call-setup
   bandwidth percent 5
  class mission-critical
   bandwidth percent 22
   queue-limit 16
  class voice
    priority 168
  class class-default
   fair-queue
   queue-limit 6
```

As a rule of thumb, the **queue-limit** for **class-default** was set lower than **mission-critical**, since the **mission-critical** class contains data traffic which is more important to the enterprise than **class-default**, which is *best effort* traffic in our profile. In this configuration IP Precedence 6 traffic is included in **mission-critical**. If IP Precedence 6 traffic is separated into its own class (**internetwork-control**) use a queue limit of 16 as a starting value.

Note In most networks, the default **queue-limit** command settings and IPSec anti-replay performance is acceptable. Only in situations where there is a requirement to further reduce the affects of IPSec anti-replay and QoS interactions should **queue-limit** tuning be considered. A modification of these values can have other side effects on the QoS service policy and related performance.

Crypto Engine QoS

This section discusses current and future capabilities of hardware crypto engines as they relate to their interaction with QoS.

Current VoIP over IPSec Crypto Engine Capabilities

A crypto engine within a Cisco VPN router's chassis can be viewed as an internal interface that processes packets for encryption or decryption. In Cisco IOS software releases before 12.2(13)T, the crypto engine operates with a FIFO input queue. Packets received from a serial interface for decryption, are interspersed with packets received from an Ethernet interface, to be encrypted. Figure 4-11 illustrates that is no distinction between a voice packet and data packet.

Figure 4-11 FIFO Crypto Engine Illustration

Consider a Cisco 2651XM router deployed at a branch site. It is configured with a full-duplex Fast Ethernet interface, a Serial E1 interface (also full-duplex) and an AIM-BP encryption accelerator. The Fast Ethernet interface connects to the branch's LAN and the Serial interface to the Internet. Consider the factors limiting the throughput of this configuration:

- Clock rate of the slowest interface—E1 rate transmitted and received (approximately 4 Mbps)
- Packet forwarding rate of the router's main CPU—in packets per second
- Crypto engine encryption/decryption rate—in packets per second

The performance characteristics of the above items, are further influenced by the traffic mix—including the size of the IP packets being switched through the network, the switching path (process, fast, CEF) of the packets and the features present in the configuration. In most hardware platforms, the packet per second capabilities of the router are more important for planning purposes than bits per second switched through the router. If the average packet size switched through the router increases from 128 bytes to 256 bytes, the packet per second capabilities of the main CPU is not necessarily cut in half.

Additionally, the control plane requirements of the internetwork—the number of routes in the routing table, the overall network stability and requirements of the routing protocol in use, the network management (SNMP) requirements, additional features enabled on the router—DLSw, TACACS, NTP, QoS, access control lists, all consume CPU resources. Assume that the amount of available memory for the main CPU, interfaces and crypto engine are adequate, no memory limitations exist.

The ratio of packets switched through, and originated by, the router in relation to those selected by the crypto map's access-list for encryption/decryption must also be considered. If encrypting an IP GRE tunnel, this tends to be a large percentage of encrypted to total packets. If not encrypting a IP GRE tunnel, and selecting only a portion of the total data traffic from the LAN/WAN interface, the ratio could be quite small.

The hardware crypto engine accelerator becomes congested when its packet processing capabilities are less than those of the router's main CPU and interface clock speed.

In cases where congestion occurs in the crypto engine, it is possible for the crypto engine to become over-subscribed—either on a momentary or sustained basis. In these cases, there are three possibilities:

- The over-subscription does not affect VoIP packets to the extent that voice quality issues result.
- Additional processing latency occurs, adding some unnecessary delay and/or jitter into voice streams.
- Voice streams experience some degree of packet loss—affecting voice quality.

Cisco internal testing and evaluation has shown it to be extremely difficult for conditions to arise that cause crypto engine congestion. In nearly all cases, the Cisco VPN Router platform's main CPU is exhausted prior to reaching the limit of the crypto engine's packet processing capabilities.

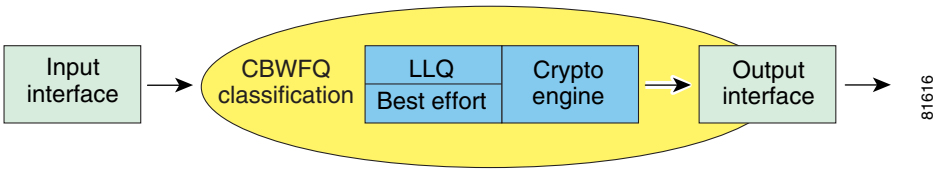
Nevertheless, Cisco provides a solution to the *potential* problem so that networks do not encounter such a situation. Consequently, Cisco developed the *LLQ for Crypto Engine* feature within Cisco IOS software—explained in the "LLQ for Crypto Engine".

LLQ for Crypto Engine

Starting with Cisco IOS software release 12.2(13)T, the crypto engine has been enhanced to provide LLQ for the Crypto Engine. This entails providing a dual-input queuing strategy: a priority or Low Latency Queue; and a best effort queue. The feature is targeted at alleviating any effects of momentary or sustained over-subscription of the hardware crypto engine, which can result in priority traffic (such as voice and video) experiencing quality issues.

The classification component to segregate traffic between the priority (LLQ) and the best effort queue is based on the CBWFQ service policy on the output interface(s). See Figure 4-12.

Figure 4-12 LLQ Crypto Engine Queue Illustration



There is no additional configuration required to enable LLQ for Crypto Engine; it is enabled by the presence of a CBWFQ Service Policy on an output interface of the VPN router.

Traffic specified in the CBWFQ service policy to be included in the priority queue(s) (LLQ) will be sent to the Crypto Engine's LLQ. Traffic included in any bandwidth classes (queues) and default class queue will be put in the Crypto Engine's best effort queue.

It is possible for more than one output interface to be configured, each potentially having a CBWFQ service policy. However, the Crypto Engine acts like a single *interface* inside the VPN router, encrypting/decrypting all outbound/inbound traffic streams for each interface on which Crypto is applied. In the case of multiple CBWFQ service policies (on different interfaces), the Crypto Engine maps all priority queues (LLQ) to its LLQ and all other queues to its best effort queue.

The priority (LLQ) queue for the crypto engine is similar in function to the priority (LLQ) queue for a service policy attached to an output interface. The Crypto Engine is analogous to an interface from Cisco IOS software's perspective.

Further, in the event that the Crypto Engine becomes oversubscribed—for short durations or sustained periods—the LLQ for Crypto Engine feature insures that if packets are dropped by the Crypto Engine, they are of appropriately low priority (not VoIP packets).

Although the feature is enabled by the presence of a CBWFQ Service Policy, like QoS it does not actually engage prioritization via the two-queuing strategy until the crypto engine itself experiences congestion.

As software-based crypto adds unacceptable latency and jitter, there are no plans to incorporate this feature for software crypto. This design guide recommends hardware acceleration of IPSec for V^3PN deployments.

When is LLQ for Crypto Engine Required

The *LLQ for Crypto Engine* feature in Cisco IOS software is not a prerequisite for deploying many V^{3}PN implementations in a high quality manner. As indicated previously, internal Cisco evaluations found it extremely difficult to produce network traffic conditions that resulted in VoIP quality suffering.

However, the feature should be viewed in the same light as Cisco QoS: it is there in Cisco IOS software to safeguard against degradation of high priority traffic delivery in periods of *harsh* network conditions when momentary or sustained over-subscription can occur.

In general, the *LLQ for Crypto Engine* feature offers the most benefit under one of the following conditions:

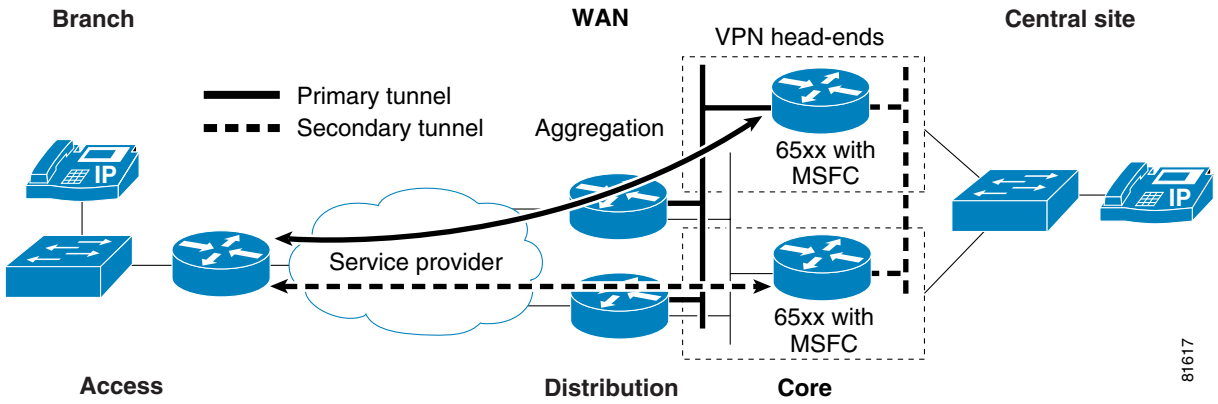
- When implementing Cisco IOS VPN Router platforms that have a relatively high amount of main CPU resources relative to Crypto Engine resources.
- When the network experiences a periodic or sustained *burst* of large packets (for example, video applications).

To summarize, high-quality V^{3}PN deployments are possible today without the *LLQ for Crypto Engine* feature in Cisco IOS software. The addition of this feature in Cisco IOS software further insures that high priority applications such as voice and video can operate in a high-quality manner even under harsh network conditions.

Head-end Topology

This design can be considered an alternative to a typical private WAN deployment where a hub and spoke topology is deployed which transports voice traffic. As such, it is assumed the rules of scalable and redundant network design need to be present in the solution. This Design Guide recommends incorporating the IPSec/GRE head-end devices into a scalable hierarchical network model. The hierarchical network design model is represented in three layers: *core*, *distribution* and *access*. Each layer provides a different functionality. Figure 4-13 illustrates how this design overlays the model presented in this publication.

Figure 4-13 Solution Deployment Topology



In the hub and spoke IPSec VPN topology, the Branch routers represent the ***access*** layer. These routers terminate the WAN interface(s) to the service provider, connect to the local LAN segment, and provide IPSec and GRE tunnel termination. Increasing availability also increases costs, but ideally multiple serial interfaces with PVCs to each WAN Aggregation router would be recommended.

The ***distribution*** layer is implemented with pairs of WAN Aggregation routers. In an implementation over a Layer 3 service provider (Internet Service Provider), these routers would have a high-speed WAN interface(s) and one or more Fast Ethernet or Gigabit Ethernet interfaces. They would typically be eBGP peers with the ISP's edge routers and would peer via iBGP between all WAN Aggregation routers. They would also implement QoS on the WAN interface to prioritize the voice traffic.

In an implementation with a Frame Relay service provider or point-to-point network, the ***distribution*** layer WAN Aggregation routers would have one or more high-speed WAN interfaces and perhaps hundreds of subinterfaces—one for each branch.

The ***core*** layer is implemented with pairs of Layer3/Layer2 switches, ideally Catalyst 6500 with MSFC. These switches provide connectivity to the network core and provide redundant Layer 2 or Layer 3 connectivity between the WAN Aggregation routers and the IPSec/GRE head-end routers.

The IPSec/GRE head-end routers are Cisco 7200VXR or Cisco 3600/3700 series routers. Their LAN connectivity is provided by the Catalyst 6500 switches and is typically implemented as dual Fast Ethernet attached routers. They terminate the IPSec peers as well as the IP GRE tunnels. They advertise the branch subnets learned through the tunnel interfaces to the core switches. Each branch router provides two IP GRE tunnels, one to each pair of IPSec/GRE head-end routers. This provides for an alternate path in the event a head-end router is taken out of service. There is no need for QoS configured on these routers, since the LAN interfaces will not be congested. The main CPU must be monitored and managed as part of the enterprise's capacity planning function.

The design recommendation of terminating the IPSec/GRE traffic on routers dedicated for this purpose is consistent with design principles proven by large scale deployments of SNAsw, TN3270 Server or DLSw. Dedicating routers for a specific function at the network core provides several advantages. The performance characteristics of the IPSec/GRE head-end might be dramatically different than a WAN aggregation router. Separating the two functions allows different ratios between IPSec/GRE head-ends and WAN aggregation: two WAN aggregation routers might be sufficient for four IPSec/GRE head-ends.

WAN aggregation routers might require a lower average CPU busy percentage to accommodate CPU spikes. WAN aggregation routers running BGP might experience considerable CPU spikes during network instability due to link flaps, for example. Also, the separation allows different versions of Cisco IOS software to be run at different locations in the network topology. For example, the WAN aggregation router might be running a General Deployment (GD) release of 12.0 mainline, while the IPSec/GRE head-end routers need an ED (Early Deployment) release to support a new hardware encryption accelerator.

The advantages and flexibility of this design outweigh the additional costs involved by separating the WAN aggregation from the IPSec/GRE head-end.

Head-end Router Locations

The number and geographic locations of the head-end routers require careful consideration. If using a Layer 3 service provider (Internet Service Provider), consider the geographic location of the majority of users/sites and the placement of head-end routers. For example, Cisco has a large concentration of employees in San Jose, California and Research Triangle Park (RTP), North Carolina. It's practical to locate head-end routers at both locations, with West coast sites terminating in San Jose and East coast locations at RTP. Dual head-end routers would be installed at each location, physically diverse within the campus. For global corporations, head-end routers in EMEA and Asia-Pacific should also be considered. A large portion of the voice delay budget will be consumed by the service provider, therefore the goal is to minimize the time spent in the service provider's cloud.

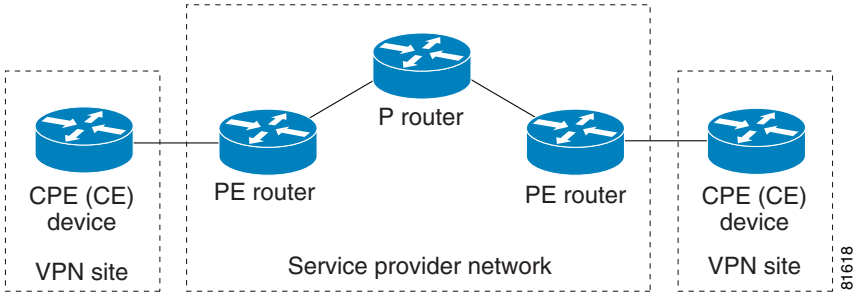
Service Provider Recommendations

For a V[3]PN deployment to operate successfully, the enterprise network designer must address the QoS requirements of encrypted voice traffic if Layer 3 service providers transport traffic between branch and head-end devices.

Boundary Considerations

In a Layer-3 service provider deployment, the Enterprise organization and the service provider must have identical QoS policies implemented on the link between the CPE (Customer Premise Equipment or CE Customer Edge) device and the PE (Provider Edge) router. Figure 4-14 illustrates this simplified topology.

Figure 4-14 Service Provider/Enterprise Boundary

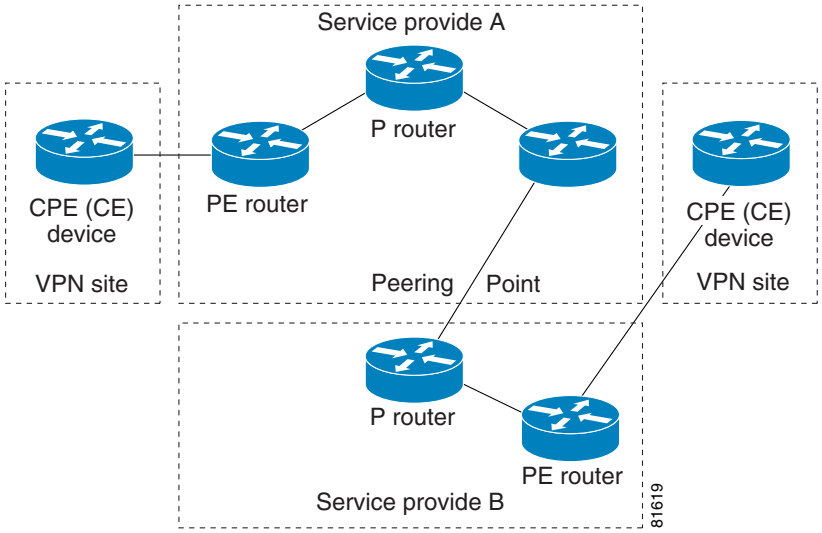


Additionally, for the service provider to offer a QoS enabled network, there needs to be a control point at the PE routers to either limit or police the amount of high priority traffic (voice, based on ToS byte) or provide some accounting function to bill based on the priority level of the traffic. Without this control in place, there is no incentive for the organization to accurately mark packets based on their required priority—they could mark all traffic as highest priority, intentionally or inadvertently.

Cross-Service-Provider Boundaries

If multiple Layer 3 service providers are used to connect the enterprise's sites, the complexity increases. The enterprise must coordinate the CPE to PE QoS with two separate ISP's. This adds time and complexity to the implementation. See Figure 4-15.

Figure 4-15 Multiple Service Providers



To have end-to-end QoS across the IPSec VPN, both service providers must agree on the amount of high priority traffic to be accepted between them. Many ISPs set the IP Precedence/DSCP value to 0 at their peering points and PE routers.

Implementations spanning multiple ISPs are more difficult to implement and manage. A contiguous service provider implementation is recommended, but when not possible, an enterprise is encouraged to seek an agreement with service providers specifying how high priority traffic across boundaries will be handled.

Service Level Agreements (SLA)

In order to support V^{3}PN, Cisco is offering a new Cisco Powered Network Designation called *IP Multi-service VPN* that qualified service providers can attain, signifying that they are capable of offering such VPN services.

Per the Cisco Powered Network Service Provider requirements, they must meet these minimum SLA components:

- Jitter—Less than or equal **20 msec**
- Delay—Less than or equal **60 msec** *one way*
- Packet Loss—Less than or equal **0.5 percent**

They are responsible for meeting the terms of the SLA they provide to the enterprise organization, just as they would if providing service via a Private WAN service offering, such as Frame Relay or ATM.

Cisco Powered Network References

These Cisco documents are useful references for V^{3}PN implementations:

- *IP Multi-service VPN Additional Requirements Summary*—http://www.cisco.com/warp/public/779/servpro/cpn/join/criteria.html
- Search for designated Cisco Powered Network service providers—http://www.cisco.com/pcgi-bin/cpn/cpn_pub_bassrch.pl

Note Please select "VPN/IP–Multi-service" in the selection list on the search page.

- Network Service Definitions—http://www.cisco.com/warp/public/779/servpro/cpn/glossary.html
- Cisco Powered Network program membership application—http://www.cisco.com/pcgi-bin/cpn/cpn_screen_zero.pl

Enterprise organizations can use these links to locate Cisco Powered Network service providers, and service providers can use them to enroll in the program.

Load Sharing

To accommodate higher traffic volumes, some implementations might require multiple physical links—such as two T1 links—for a large branch office. This section addresses several load-sharing topics:

Load Sharing Capabilities

Routing protocols—such as OSPF, EIGRP, or BGP—have the ability to insert multiple equal cost or (in the case of EIGRP) unequal cost routes into the routing table. Path determination is selected from routes in the routing table rather than what might be contained in the routing protocol's topology database. The Cisco IOS fast-switching path will load share on a per destination basis and CEF by default will load share on a per source/destination basis and can be configured for per-packet load sharing. Since CEF includes the source address in the decision process, it provides for a more granular distribution than fast switching. Packets that are process switched will be per-packet load shared; however, process switching is not recommended as it negatively impacts the overall throughput capabilities of the router.

With fast-switching and CEF-switching, load sharing on links tends to be equally balanced as the number of IP flows (source/destination IP addresses communicating) increase. It is advantageous to transmit all the packets for a flow across a single interface because the likelihood of the packets being received out-of-order is less than if the packets are sent over multiple links. If there are only a few IP flows, and one of these flows is a high bandwidth consumer and the remaining flows are low bandwidth consumers, the amount of link utilization will be far from equal. An example of this would be one workstation sending a large file transfer (FTP) and the remaining stations all using Telnet. The link utilization across two links in this situation is far from equal.

A solution to equally balance the two physical links has been to process switch (not recommended due to the performance constraints) or to use CEF's per-packet load sharing. Another approach is to bundle multiple links at Layer 2 so they appear to the routing protocol to be one link and let the interface driver or Layer-2 logic address load sharing. Examples of this are Inverse Multiplexing over ATM (IMA) and Multilink PPP.

Encrypted Traffic Appears as a Few, Large Flows

IPSec tunnels will *hide* the source and destination address of the traffic selected for encryption in its own IP header and the switching decision at the physical interface will be presented with high bandwidth flows from a few sources—the number of IPSec tunnels. Internet Key Exchange (IKE) and any traffic not selected for encryption by the crypto map (split tunneling or other management traffic) will also be seen on the interface—but this traffic will be assumed to be minimal.

Minimize Out-of-Order Packets

Out of order packets are detrimental from TCP's perspective in that they must be re-ordered by the receiving station's TCP stack before being delivered to the application. This consumes memory and decreases throughput. For voice, some out-of-order packets can be tolerated (MAX_MISORDER defines the maximum mis-order of packets allowed by RTP); however, the packets must be correctly reordered by the jitter buffer of the receiving phone. The human ear will not tolerate listening to digitized voice in anything but the correct order. It will sound garbled or unintelligible.

From IPSec's perspective, per packet load sharing for the same IPSec flow after the packets have been encrypted and assigned an ESP sequence number will increase the probability that packets will be dropped due to replay protection checks. At the same data rate, CEF load sharing per-packet on equal cost IPSec-encrypted GRE tunnels (**ip load-sharing per-packet** command on multiple tunnel interfaces) will show less anti-replay drops than routing all the traffic in one IPSec encrypted GRE tunnel and using CEF load sharing per-packet (**ip load-sharing per-packet** command on the WAN interfaces) to switch the encrypted tunnel's packets out multiple interfaces.

However, reducing the anti-replay drops is not the only goal. Per-packet load sharing across two GRE tunnels increases the likelihood the voice packets between any two IP Phones will take different paths through the network—increasing the likelihood that they will arrive out of order.

Load Sharing Design Approach

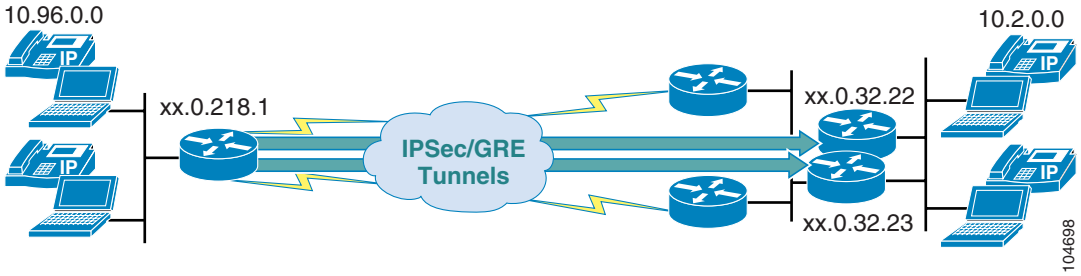
To balance these somewhat conflicting requirements, the best approach when implementing multiple physical links would be to CEF switch (per source/destination load sharing, not per-packet) on two equal-cost GRE tunnels and configure each GRE tunnel so it has an affinity to a particular interface when all interfaces are up. This approach strives to maintain voice packets between any two IP phones in the same IPSec/GRE tunnel and on the same physical interface. With voice and data packets using both IPSec/GRE tunnels, and these tunnels each routed over a separate physical interface, both links will be used while maintaining the same path for any one particular flow.

This approach will not provide a precise distribution of packets over the two links as would per-packet load sharing, but it will allow both links to be used and minimizes the negative aspects of out of order packets for voice and data post decryption and anti-replay drops when those packets are in the IPSec tunnel.

In the event one of the serial links fails, the service policy attached to the physical interface must be configured to provision sufficient bandwidth for the number of calls allows by the sites's Call Admission Control process—CallManager *locations* or use of a Gatekeeper. When both links are up and operational, the site will never consume all the bandwidth provisioned for the voice LLQ, as the Call Admission Control process will limit the number of calls. Bandwidth not used by the LLQ is not wasted, it will be used by the other bandwidth classes and class-default.

Use the topology diagram illustrated in Figure 4-16 as a guide for the examples shown.

Figure 4-16 Load Sharing Topology Example



The router on the left of the diagram represents the branch router and it has two GRE tunnels to the head-end IPSec/GRE peers.

```
vpnjk-2600-18#show run | include tunnel
interface Tunnel0
 tunnel source Loopback0
 tunnel destination xx.0.32.22
interface Tunnel1
 tunnel source Loopback0
 tunnel destination xx.0.32.23

vpnjk-2600-18#show ip interface brief | include Loopback
Loopback0                  xx.0.218.1      YES NVRAM  up    up

vpnjk-2600-18#show run | include ip route xx.0.32
ip route xx.0.32.22 255.255.255.255 Serial0/0.100
ip route xx.0.32.23 255.255.255.255 Serial0/0.101
```

To configure an affinity for the logical (IPSec/GRE) links to the physical links, two host static routes are configured. An alternate approach would be to advertise a dynamic route with a more specific mask (in this example using 255.255.255.255 or a host route) down each WAN interface in addition to a summary, supernet, or less specific route down each WAN link. The remote router will route both IPSec/GRE tunnel interfaces out the remaining link in the event of a WAN link failure with this configuration. The routing table in this example looks like this:

```
vpnjk-2600-18#show ip route  | begin xx.0.0.0
     xx.0.0.0/8 is variably subnetted, 6 subnets, 3 masks
B       xx.0.0.0/8 [20/0] via xx.0.32.6, 00:16:07
                   [20/0] via xx.0.32.2, 00:16:07
S       xx.0.32.23/32 is directly connected, Serial0/0.101
S       xx.0.32.22/32 is directly connected, Serial0/0.100
C       xx.0.32.4/30 is directly connected, Serial0/0.101
C       xx.0.32.0/30 is directly connected, Serial0/0.100
C       xx.0.218.1/32 is directly connected, Loopback0
     10.0.0.0/8 is variably subnetted, 13 subnets, 3 masks
D       10.2.0.0/24 [90/297270016] via 10.96.1.2, 00:16:07, Tunnel0
                    [90/297270016] via 10.96.1.6, 00:16:07, Tunnel1
```

Note that the remote router is learning a route to xx.0.0.0/8 via BGP and with **maximum-paths 2** configured both are inserted in the routing table. In this example the GRE tunnel's interface *delay* has not been changed and EIGRP inserts both routes to 10.2.0.0/24 in the routing table as the are equal cost. The BGP configuration for the remote router is as follows:

```
router bgp 65018
 no synchronization
 bgp log-neighbor-changes
 network xx.0.218.1 mask 255.255.255.255
```

```
 neighbor xx.0.32.2 remote-as 65000
 neighbor xx.0.32.6 remote-as 65000
 maximum-paths 2
 no auto-summary
!
```

CEF is enabled on the branch router and by default will load share per source/destination IP address rather than per-packet:

```
vpnjk-2600-18#show ip cef 10.2.0.0 detail
10.2.0.0/24, version 27, epoch 0, per-destination sharing
```

In this illustration, a traffic generation tool is sending traffic from two different source addresses with three different destination addresses:

```
Summary of IP traffic streams on FastEthernet0/1
  ts#     tos  len   id frag ttl protocol chksm source          destination
    1 TCP  48  576 0000 0000  60     6     67D1 10.96.0.17      10.2.0.45
    2 UDP  B8   60 0000 0000  60    17     68CF 10.96.0.18      10.2.0.43
    3 UDP  40  188 0000 0000  60    17     6967 10.96.0.17      10.2.0.24
    4 UDP  B8   60 0000 0000  60    17     696E 10.96.0.18      10.2.0.24
```

To verify that both IPSec/GRE tunnel interfaces are being used as well as both physical interfaces, the network manager can use the **show ip cef exact-route** command and either a **show interface** command or in this case, with Frame Relay interfaces, **the show frame pvc** command to verify both physical interfaces are being used.

```
vpnjk-2600-18#show ip cef exact-route 10.96.0.17 10.2.0.45
10.96.0.17      -> 10.2.0.45      : Tunnel0 (next hop 10.96.1.2)
vpnjk-2600-18#show ip cef exact-route 10.96.0.17 10.2.0.24
10.96.0.17      -> 10.2.0.24      : Tunnel0 (next hop 10.96.1.2)
vpnjk-2600-18#show ip cef exact-route 10.96.0.18 10.2.0.43
10.96.0.18      -> 10.2.0.43      : Tunnel0 (next hop 10.96.1.2)
vpnjk-2600-18#show ip cef exact-route 10.96.0.18 10.2.0.24
10.96.0.18      -> 10.2.0.24      : Tunnel1 (next hop 10.96.1.6)

vpnjk-2600-18#show frame pvc 101 | include output pkts
  input pkts 37            output pkts 3171         in bytes 4145
vpnjk-2600-18#show frame pvc 100 | include output pkts
  input pkts 154           output pkts 7158         in bytes 20525
```

The **show adjacency** command can also be used to verify the packet distribution.

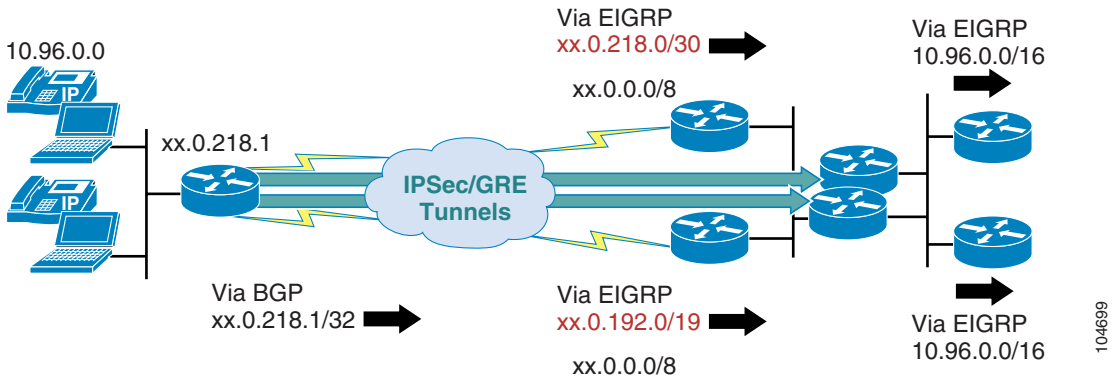
```
vpnjk-2600-18#show adjacency serial 0/0.101 detail | include packet
                                33555 packets, 15210652 bytes
vpnjk-2600-18#show adjacency serial 0/0.100 detail | include packet
                                15427 packets, 2170788 bytes
vpnjk-2600-18#show adjacency tunnel 0 detail | include packet
                                16541 packets, 1389444 bytes
vpnjk-2600-18#show adjacency tunnel 1 detail | include packet
                                37140 packets, 14626480 bytes
```

Load Sharing from Head-end to Branch

The previous load sharing discussions focused on configuration from the branch router's perspective. If the WAN links are Frame Relay or dedicated T1 links between branch router and head-end, the enterprise can control load sharing for the head-end to branch traffic by configuring the head-end WAN aggregation routers to advertise (or learn dynamically from the branch) the appropriate more specific routes to the IPSec/GRE head-end routers.

Review the sample topology shown in Figure 4-17.

Figure 4-17 Load Sharing from Head-end to Branch



The remote router (on the left of the diagram) will advertise its loopback interface IP address to both head-end WAN routers as xx.0.218.1/32. This loopback address is the IPSec/GRE tunnel destination address for both head-end IPSec/GRE routers. The top WAN aggregation router has a static route for xx.0.0.0/8 and xx.0.218.0/30 to the Null0 interface and is redistributing both these routes to the IPSec/GRE head-end routers via EIGRP 23. The lower WAN aggregation router has a static route for xx.0.0.0/8 and xx.0.192.0/19 to the Null0 interface and is also redistributing these routes via EIGRP 23. Note that xx.0.218.0/30 and xx.0.218.1/32 are both more specific (have a longer prefixes) which fall in the address space of xx.0.192.0/19. An example of the WAN router's configuration follows.

```
upperWAN#show run | include Null0
ip route xx.0.0.0 255.0.0.0 Null0
ip route xx.0.218.0 255.255.255.252 Null0

upperWAN#sh ip route | include xx.0.218
B       xx.0.218.1/32 [20/0] via xx.0.32.1, 1d02h
S       xx.0.218.0/30 is directly connected, Null0

upperWAN#show run | begin router eigrp 23
router eigrp 23
 redistribute static
 network xx.0.0.0
 default-metric 1000 100 255 1 1500
 ...


lowerWAN#show run | include Null0
ip route xx.0.0.0 255.0.0.0 Null0
ip route xx.0.192.0 255.255.224.0 Null0

lowerWAN#show ip route | begin xx.0.218
B       xx.0.218.1/32 [20/0] via xx.0.32.5, 6d03h
D       xx.0.218.0/30 [90/2588160] via xx.0.1.20, 1d03h, FastEthernet0/1
```

```
S       xx.0.192.0/19 is directly connected, Null0

lowerWAN#show run | begin router eigrp 23
router eigrp 23
 redistribute static
 network xx.0.0.0
 default-metric 1000 100 255 1 1500
 ....
```

The lower IPSec/GRE termination router will be configured to ignore any routing advertisements learned via EIGRP 23 for the xx.0.192.0/19 network which have a prefix greater than or equal to 30 bits. The upper IPSec/GRE termination router doesn't have this distribute-list filtering based on the IP Prefix-list feature. IP routing is accomplished by using the matching route that has the longest prefix. The net result of this configuration provides for the upper IPSec/GRE router to follow the xx.0.218.0/30 route advertised by the upper WAN aggregation router while the lower IPSec/GRE router will follow the xx.0.192.0/19 advertised by the lower WAN aggregation router. Since both WAN aggregation routers have a matching route for xx.0.218.1 learned via BGP from the remote branch, they will route the tunnel out their respective WAN interfaces.

In the event one of the WAN routers fail, the IPSec/GRE head-end routers will use as a last resort the xx.0.0.0/8 route advertised by the surviving WAN aggregation router to reach the tunnel destination. If the surviving WAN router is the lower router, the xx.0.192.0/19 will continue to be advertised to both IPSec/GRE routers and that route will be the most specific route to reach xx.0.218.1.

The EIGRP configuration for the lower IPSec/GRE termination router follows:

```
router eigrp 23
 network xx.0.0.0
 distribute-list prefix FOLLOWslash19 in
 ... [additional commands removed] ...
!
router eigrp 45
 network 10.0.0.0
 ... [additional commands removed] ...
!
ip prefix-list FOLLOWslash19 seq 5 deny xx.0.192.0/19 ge 30
ip prefix-list FOLLOWslash19 seq 100 permit 0.0.0.0/0 le 32
```

The instance of EIGRP 45 is included for the 10.0.0.0 network which is used in the remote, tunnel and head-end routers. Both IPSec/GRE routers will advertise an equal cost route to the remote subnet, 10.96.0.0/16, so the IP packets from the head-end LANs will use both GRE tunnels to reach the branch subnet. As with the branch configuration, the default of IP CEF per source/destination load sharing will be in effect to send voice and data traffic down each GRE tunnel. Another alternative would be do use multiple HSRP groups on the IPSec/GRE routers so traffic from the head-end arrives on both head-end routers.

Following is an excerpt of the routing table from both IPSec/GRE routers.

```
upperIPSecGRE#show ip route | begin xx.0.0.0
     xx.0.0.0/8 is variably subnetted, 15 subnets, 5 masks
C       xx.0.1.0/24 is directly connected, FastEthernet0/1
D       xx.0.0.0/8 [90/2588160] via xx.0.1.21, 02:07:35, FastEthernet0/1
                   [90/2588160] via xx.0.1.20, 02:07:35, FastEthernet0/1
D       xx.0.32.23/32 [90/156160] via xx.0.1.23, 05:11:16, FastEthernet0/1
C       xx.0.32.22/32 is directly connected, Loopback0
        ... [detail removed] ...

D       xx.0.218.0/30 [90/2588160] via xx.0.1.20, 02:07:35, FastEthernet0/1
D       xx.0.192.0/19 [90/2588160] via xx.0.1.21, 05:09:03, FastEthernet0/1
```

```
lowerIPSecGRE#show ip route | begin xx.0.0.0
     xx.0.0.0/8 is variably subnetted, 14 subnets, 5 masks
C       xx.0.1.0/24 is directly connected, FastEthernet0/1
D       xx.0.0.0/8 [90/2588160] via xx.0.1.21, 02:10:20, FastEthernet0/1
                   [90/2588160] via xx.0.1.20, 02:10:20, FastEthernet0/1
C       xx.0.32.23/32 is directly connected, Loopback0
D       xx.0.32.22/32 [90/156160] via xx.0.1.22, 02:10:17, FastEthernet0/1

        ... [detail removed] ....
D       xx.0.192.0/19 [90/2588160] via xx.0.1.21, 05:14:01, FastEthernet0/1
```

With this configuration, provided both WAN routers are available, the IPSec/GRE tunnels will have an affinity to their particular WAN link.

Service Provider Considerations for Load Sharing

If the enterprise is using two Internet service providers to terminate the branch router and head-end, the head-end configuration can influence the path selection. It can cause the head-end to route packets from the one IPSec/GRE head-end out service provider “A” and from the second IPSec/GRE head-end out service provider “B”—while maintaining the logical to physical affinity concept established in the earlier sections.

However, if the physical links terminate to the same service provider—with either on the same router or a different router at the service provider point-of-presence (POP)—then a more-specific (host) route must be used by the service provider. In this case, the enterprise organization must source the branch router’s tunnel interface off a distinct IP address rather than using the same address (Loopback 0) as shown in Figure 4-17. The service provider might not offer this service to the enterprise organization due to the added configuration complexity and additional routes within its network.

Because service provider offerings vary from service provider to service provider, it is advisable to review the requirements of the enterprise network with proposed service providers prior to implementation.

E911 and 911 Emergency Services

Handling of emergency calls would be implemented on the same model as a centralized CallManager with remote IP Phones on a traditional private Frame Relay deployment. Please refer to the following URL for an explanation of addressing emergency calls in a Cisco AVVID deployment:

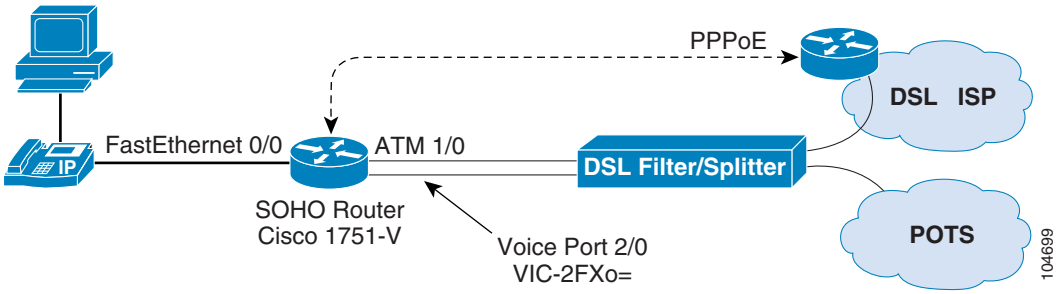
- http://www.cisco.com/en/US/netsol/ns110/ns163/ns165/ns268/networking_solutions_design_guidances_list.html

Survivable Remote Site Telephony

A Survivable Remote Site Telephony (SRST) configuration would be desirable in a V^{3}PN design which utilized a centralized CallManager configuration. In the event the IPSec VPN tunnel is down, remote locations would continue to have limited support and use of their IP Phones. In-depth configuration and testing of SRST is beyond the scope of this document, however proof of concept testing was completed.

The configuration shown in the following diagram uses a Cisco 1751 Virtual Private Network (VPN) Module (MOD1700-VPN), a WIC-ADSL and a VIC-2FXO running Cisco IOS image c1700-k9o3sv8y7-mz.122-13.T1. This router was connected to Cisco and the production CallManager using an ISP and the DSL interface. The analog phone line which is associated with the DSL service was connected to the FXO port of the router. A common DSL filter/splitter was used to split the DSL and analog line to the router. Refer to Figure 4-17.

Figure 4-18 SRST Topology



The following represents the relevant SRST configuration commands for the above topology.

```
!
ip dhcp pool Client
   import all
   network 10.81.2.0 255.255.255.248
   default-router 10.81.2.1
   dns-server 64.102.6.247 171.68.226.120
   domain-name cisco.com
   option 150 ip 64.102.2.93
   netbios-name-server 171.68.235.228 171.68.235.229
!
voice-port 2/0
 connection plar 23685
 description My Home Phone Line
!
voice-port 2/1
!
!
dial-peer voice 45 pots
 destination-pattern 9
 port 2/0
!
call-manager-fallback
 ip source-address 10.81.2.1 port 2000
 max-ephones 2
 max-dn 2
 access-code fxo 9
!
```

The Cisco 1751 router was configured as a DHCP server for the IP Phone, which is a requirement for SRST to function. The phone would have initially needed to register with the production CallManager to function. The phone in question has the five digit extension of 2-3685. The goal of this configuration is to allow the phone to call out the analog POTS line in the event the IPSec tunnel is down and to route incoming calls on the analog POTS line to the IP Phone at extension 2-3685.

To test this configuration, the RJ-11 jack for the DSL line was removed from the router's DSL interface. This loss of network connectivity will result in the IP Phone missing its keepalives with all the phone's configured CallManagers. The 1751 becomes the fallback CallManager and the phone's display will indicate it is in CallManager fall-back mode. This process normally will be completed in approximately two minutes.

At this point, the IP Phone can call an outside phone number by first dialing 9, a second dial tone will be heard, then the area code (if required) and number can be dialed to reach the intended party. Incoming calls will be automatically routed to the extension listed on the **connection plar** [*extension*] command under the voice port. If **connection plar** is not specified, an incoming call will received a dial tone from the Cisco 1751 and the extension can be manually dialed to complete the call. When the router is not functioning as a fallback CallManager, incoming calls will not be completed. It is recommended that no other POTS phones share the analog line connected to the router.

For configurations of SRST based on digital lines and multiple IP Phones, please consult the appropriate configuration and command reference guides.

Design Checklist

This design checklist facilitates pre-implementation planning and the decision process.

Table 4-4 Design Checklist

Design Step	Section References
Identify physical locations for sites to be supported by this design	Organization Specific
Determine IP addressing requirements of branch routers and manual or auto summary scheme	Organization Specific
Decide on location of head-end routers. Will they be in the same rack or across a continent? Will the summary scheme black hole traffic in a failover?	"Head-end Router Locations" section on page 4-23
Determine primary and backup head-end routers. Will branch routers have affinity to the geographically closest head-end?	"Head-end Router Locations" section on page 4-23
Determine number of concurrent voice calls per location. Estimate data bandwidth requirements. Set IP Telephony Call Admission Control requirements appropriately.	"Bandwidth Provisioning for WAN Edge QoS" section on page 4-5
Select appropriate branch site products, based on V^{3}PN link speed and other device requirements.	"Branch Office Product Selection" section on page 5-9
Based on the number of remote sites, bandwidth requirements, determine number and location of head-end routers.	"Head-end Product Selection" section on page 5-6
Consider service provider selection process, consult CCO for Cisco Powered Network designated providers.	"Service Provider Recommendations" section on page 4-24
Plan for traffic load sharing requirements for head-end and large branch offices.	"Load Sharing" section on page 4-26
Review existing emergency services plans.	"E911 and 911 Emergency Services" section on page 4-32
Consider SRST requirements for remote locations.	"Survivable Remote Site Telephony" section on page 4-32