CHAPTER 3

V3PN Solution Components

Implementation of a site-to-site IPSec VPN design capable of supporting transport of voice and video, requires the combination of three Cisco technologies:

- IP Telephony (Voice over IP), page 3-1)
- Quality of Service (QoS), page 3-2
- IP Security (IPSec), page 3-4

These three technologies have been implemented on many enterprise networks as standalone functions or in some combination—especially IP Telephony and QoS. V3PN combines all three technologies enabled simultaneously on a common network.

This design guide addresses the areas of intersection between the three technologies and provides configuration and verification tips to promote the successful implementation of a similar design—while maintaining the same availability and voice quality demonstrated in Cisco Enterprise Solutions Engineering lab testing.

IP Telephony (Voice over IP)

This design document assumes the target enterprise site includes or will include an IP Telephony deployment—with the expectation of extending this deployment to branch office locations over an IPSec VPN. As with private WAN deployments of IP Telephony, Call Admission Control (CAC) is implemented to limit the number of concurrent calls based on the capabilities and traffic handling capacity of the branch office routers being deployed, as well as the link speed being used.

IP Telephony can be deployed using several different CODEC and sampling rate schemes. Each offers advantages and disadvantages in terms of voice quality, added latency, bandwidth consumption, and router resource consumption. For example, G.711 with 20 msec sampling and a transmission rate of 50 pps is a common deployment. This offers high voice quality, minimal latency, but higher bandwidth consumption. G.729 with 20 msec sampling and a transmission rate of 50 pps is another common deployment, especially for lower speed links, as it offers very good voice quality at a lower bandwidth consumption.

Both G.711 and G.729 voice calls were evaluated as part of this solution. G.729 with 20 msec sampling and a transmission rate of 50 pps is the recommended CODEC/sampling scheme for V3PN deployments. This is primarily due to the bandwidth requirements.

Another possibility would be to deploy G.729 with 30 msec sampling at a transmission rate of 33 pps. This offers an additional bandwidth savings, although there might be a trade-off in voice quality as loss of a single packet can produce an audible *click* or *pop*. With 20 msec sampling, loss of two consecutive packets would be required to cause audible errors.

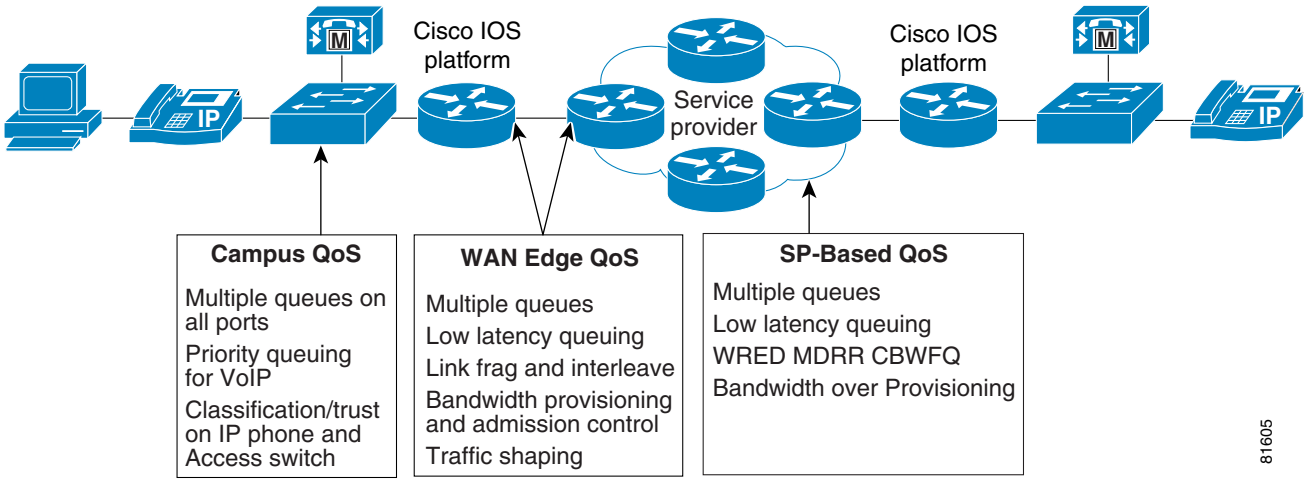
Voice Activity Detection (VAD) is another IP Telephony feature that can be used to achieve bandwidth savings. VAD works by not transmitting the natural periods of silence during a voice call, lowering the transmission rate (in packets per second) during these periods and hence lowering the overall bandwidth consumption (on average). Use of VAD was not evaluated.

Note Use of VAD has the potential to result in perceived voice quality issues and is not recommended. This design guide does not cover IP Telephony deployment in any detail, only the components that affect V^{3}PN deployment.

Quality of Service (QoS)

Cisco QoS is an enabling technology for IP Telephony that prioritizes latency-sensitive packets, such as voice and video, over lower priority traffic (i.e. data) to minimize end-to-end delay. QoS also seeks to insure consistent voice packet delivery which minimizes the arrival variance (jitter). In order for VoIP over IPSec to be successfully implemented, QoS is required at different points in the network. Figure 3-1 shows the different QoS components that are required:

Figure 3-1 Components of QoS for V^{3}PN Deployment



The QoS requirements for the central site (campus) LAN and branch LAN are the same as a typical IP Telephony deployment, including:

- Classification or Class of Service (CoS)/Trust on IP Phone and switch
- Proper speed and duplex between IP Phone and switch
- Multiple queues on IP Phone and switch ports
- WRED within data queue for congestion management
- Mapping ToS to CoS on the LAN edge router

On the enterprise WAN edge, QoS requirements are again the same as for a typical IP Telephony deployment, including:

- CBWFQ with a priority queue (LLQ)
- Traffic Shaping (if applicable)

- Link Fragmentation and Interleaving (LFI) (where appropriate)

One major difference is that now the bandwidth provisioning at the WAN edge must consider the additional overhead of IPSec and IP GRE. Another major difference could be if Compressed Real-Time Protocol (cRTP) is currently being used. IPSec and cRTP are not compatible standards, therefore cRTP is not applicable for V^{3}PN implementations. Encrypted IPSec packets are ignored by cRTP logic, therefore no bandwidth savings will result for the encrypted voice calls.

This Design Guide does not cover campus or WAN edge QoS deployment in any detail, only the components that affect V^{3}PN deployment. Refer to the *Enterprise QoS Design Guidelines* at the following URL: http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND/QoS-SRND-Book.html.

The majority of scalability and performance evaluation was based on Frame Relay and HDLC as the service provider's WAN technology. The natural progression of this solution is to enable V^{3}PN across the Internet. QoS enabled ISP's are in the early adoption phase of deployment. Some ISPs offer SLAs for end-to-end latency and packet drop rates. However, these agreements might treat all data traffic equally, with no distinction between packets with different IP Precedence or differentiated service code point (DSCP) values within the ToS byte.

The requirements for the service provider include:

- Deliver packets end-to-end with minimal delay, jitter, and loss. This can be accomplished by prioritizing packets based on ToS (IP Precedence/DSCP) in the core (QoS enabled core) or over-provisioning the bandwidth in the core
- Provide and meet a SLA to achieve end-to-end delay, jitter, and packet loss requirements. Several SPs offer SLAs with less than 60 msec delay today.
- Implement a policy to handle high priority traffic exceeding the agreed upon rate with the enterprise organization, as well as how traffic will be handled which crosses the service provider boundary.
- Highly preferred that the service provider *mirror* the enterprise WAN edge QoS: CBWFQ, LLQ, Traffic Shaping and LFI.

The service provider QoS component is an important aspect of voice call quality—one over which the enterprise network implementer has the least direct control. For this reason, the ability of a service provider to offer Cisco Powered Network *IP Multi-service VPN* is a competitive advantage and point of service differentiation in the competitive ISP market.

A service provider offering different classes of service within their backbone would need to police the organization's data rate by ToS byte to the agreed upon offered rate, or charge more for transport of the higher priority traffic. Without this policing or tiered billing function, their customer could (intentionally or inadvertently) mark all packets going into the service provider with the highest priority.

For more information regarding service providers for V^{3}PN and SLA requirements, see the "Service Provider Recommendations" section on page 4-24.

IP Security (IPSec)

The IPSec component provides secrecy (confidentially) and integrity of both voice and data over public networks. Government regulations might legislate the use of crypto in financial and health care enterprises, but the driving motivation is ubiquitous low cost network access by Internet Service Providers.

This publication does not cover site-to-site IPSec VPN deployment in absolute detail, only the components that affect V^{3}PN deployment. For more information, see the following URL: http://www.cisco.com/en/US/netsol/ns340/ns394/ns171/ns142/networking_solutions_sub_solution_home.html.

The design principles were proven by scalability testing in the Cisco Enterprise Solutions Engineering lab. The scale testing methods are designed to test worst-case scenarios. From a design standpoint that entails:

- Strong (3DES) encryption for both IKE and IPSec
- IP GRE with IPSec Tunnel mode
- Routing protocol (EIGRP)
- Diffie-Hellman Group 2 (1024 bit) for IKE
- Secure Hash Algorithm (SHA)-HMAC, a 160-bit rather than 128-bit with Message Digest 5 (MD5)-HMAC (both hash algorithms are truncated to 12 bytes in the ESP packet trailer).

Should the organization choose to implement less stringent security parameters, or choose to use IPSec Transport mode rather than Tunnel mode, or not implement IP GRE tunnels, the test results will continue to be applicable from a scalability standpoint.

Advanced Encryption Standard (AES) is a new encryption algorithm, available now in several Cisco platforms. AES will be evaluated in a future revision.

Pre-shared keys were used during this evaluation. Digital Certificates are planned for subsequent scalability, however they are currently deployed in internal Cisco deployments. Whether pre-shared keys or Digital Certificates are being used are not expected to significantly affect V^{3}PN performance.

Issues Specific to V^{3}PN

Combining voice, video, and data traffic on a converged network, encrypting that traffic with IPSec, then prioritizing that traffic with QoS presents unique design considerations for the network designer. The sections that follow address these issues in more detail.

- Packet Header Overhead Increases
- cRTP Not Compatible with IPSec
- Delay Budget
- Spoke-to-Spoke Crypto Delay
- FIFO Queue in Crypto Engine
- Anti-Replay Failures

Packet Header Overhead Increases

The addition of an IP GRE header and IPSec / ESP header increases the size of the original voice (or video) packet. Using Layer 3 packet sizes, a 60-byte G.729 voice packet increases to 136 bytes with IP GRE and IPSec tunnel mode. A 200-byte G.711 voice packet increases to 280 bytes. For low-speed WAN links, G.729 has been the recommended CODEC to conserve bandwidth. This is considerably more important in an IPSec implementation.

See the "Bandwidth Provisioning for WAN Edge QoS" section on page 4-5 for more information regarding IPSec and GRE packet expansion.

cRTP Not Compatible with IPSec

Network managers implement cRTP as a link efficiency mechanism to decrease the overhead of voice traffic on low-speed (less than E1) links. cRTP compresses the IP/UDP/RTP header in an RTP data packet from 40 bytes to approximately 2-to-5 bytes. However, cRTP and IPSec are inherently incompatible standards. The original IP/UDP/RTP header is encrypted by IPSec by the time the RTP compressor is called upon to perform the compression. Therefore cRTP cannot associate the encrypted IP/UDP/RTP packet with a known media stream, compression does not occur, and bandwidth savings are not realized. The encrypted IP/UDP/RTP packet simply by-passes the compression process and continues.

Delay Budget

The delay budget for a typical IP Telephony implementation includes fixed and variable components. The ITU target value is for one-way delay to be 150 msec or less, although up to 250 msec might be acceptable.

In a V^{3}PN deployment, there are two additional delay components added in to the overall delay budget: Encryption at the origination point of the VPN tunnel and decryption at the termination point. Performance and scalability testing results suggest that in most cases the additional delay caused by encryption/decryption is approximately 2-to-5 msec. A conservative planning estimate would be 10 msec for encryption and 10 msec for decryption. Fundamentally, IPSec adds a trivial amount of additional delay to voice deployments versus a clear text IP telephony deployment.

See the "Calculating Delay Budget" section on page 4-2 for more information regarding delay budget.

Spoke-to-Spoke Crypto Delay

In addition to the aforementioned delay components, and the crypto engine encrypt and decrypt delay component, this design guide specifies a hub-and-spoke topology. If one branch places a voice call to another branch, the voice packet must be encrypted at the branch, decrypted at the head-end router, the routing table lookup (path determination) switches the packet out another IP GRE tunnel, which is encrypted and then decrypted at the second branch. The frequency of branch-to-branch calling must be analyzed and if the call frequency warrants, requires creation of a site-to-site VPN tunnel between these particular branch locations (partial meshing).

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 reducing the dependence of VoIP traffic flowing from a spoke (branch) through the central site to another spoke (branch). DMVPN allows 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.

See the "Calculating Delay Budget" section on page 4-2 for more information regarding delay budget.

FIFO Queue in Crypto Engine

In Cisco IOS software releases prior to 12.2(13)T, the crypto engine is a single FIFO queue for admitting packets to the crypto process. Both packets for encryption and packets for decryption share the same queue. In some hardware platforms and traffic profiles, the crypto engine could be the gating throughput factor. LLQ for Crypto Engine is available in Cisco IOS software release 12.2(13)T and higher, to manage any over subscription of the hardware crypto accelerator. On platforms and traffic profiles tested in this guide, voice quality was maintained without LLQ for Crypto Engine.

Software-based crypto engines (with no HW-accelerators installed) have a FIFO queue and will not have the LLQ for Crypto Engine feature. For this reason, as well as the unacceptable latency and jitter performance, software-based crypto is not recommended for V^{3}PN implementations.

See the "Crypto Engine QoS" section on page 4-20 for more information regarding Crypto Engine and QoS.

Anti-Replay Failures

The IPSec ESP (Encapsulating Security Protocol) authentication component provides message integrity. One aspect is a sequence number assigned to packets within each security association. This sequencing of packets prevents a packet from being captured and replayed at a later time to the intended receiver. Fundamentally QoS techniques alter the order of packets between two IPSec peers, voice packets are prioritized over data packets. The message integrity aspect of IPSec runs contrary to the intended re-ordering of packets by QoS. Under the scenarios evaluated for this guide, there was no significant packet loss due to anti-replay/QoS interaction.

See the "Anti-Replay Considerations" section on page 4-16 or more information regarding IPSec anti-replay and its interaction with QoS.