

CISCO SYSTEMS



Business Ready Branch Solutions for Enterprise and Small Offices—Reference Design Guide

OL-7470-01

April 2005

Corporate Headquarters

Cisco Systems, Inc.

170 West Tasman Drive
San Jose, CA 95134-1706
USA

<http://www.cisco.com>
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100



THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS IN THIS MANUAL ARE SUBJECT TO CHANGE WITHOUT NOTICE. ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS MANUAL ARE BELIEVED TO BE ACCURATE BUT ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. USERS MUST TAKE FULL RESPONSIBILITY FOR THEIR APPLICATION OF ANY PRODUCTS.

THE SOFTWARE LICENSE AND LIMITED WARRANTY FOR THE ACCOMPANYING PRODUCT ARE SET FORTH IN THE INFORMATION PACKET THAT SHIPPED WITH THE PRODUCT AND ARE INCORPORATED HEREIN BY THIS REFERENCE. IF YOU ARE UNABLE TO LOCATE THE SOFTWARE LICENSE OR LIMITED WARRANTY, CONTACT YOUR CISCO REPRESENTATIVE FOR A COPY.

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED "AS IS" WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

CCSP, CCVP, the Cisco Square Bridge logo, Follow Me Browsing, and StackWise are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn, and iQuick Study are service marks of Cisco Systems, Inc.; and Access Registrar, Aironet, BPX, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Enterprise/Solver, EtherChannel, EtherFast, EtherSwitch, Fast Step, FormShare, GigaDrive, GigaStack, HomeLink, Internet Quotient, IOS, IP/TV, iQ Expertise, the iQ logo, iQ Net Readiness Scorecard, LightStream, Linksys, MeetingPlace, MGX, the Networkers logo, Networking Academy, Network Registrar, *Packet*, PIX, Post-Routing, Pre-Routing, ProConnect, RateMUX, ScriptShare, SlideCast, SMARTnet, The Fastest Way to Increase Your Internet Quotient, and TransPath are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or Website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0601R)

Business Ready Branch Solutions for Enterprise and Small Offices—Reference Design Guide
Copyright © 2005 Cisco Systems, Inc. All rights reserved.



CONTENTS

CHAPTER 1

Business Ready Branch Solution Overview 1-1

Introduction 1-1

Understanding the Business Ready Branch Solution 1-2

Service Building Blocks 1-3

 Service Building Blocks Overview 1-3

 WAN Services 1-4

 LAN Services 1-5

 Security 1-8

 Security Overview 1-8

 Securing the WAN 1-9

 Defending the Perimeter 1-12

 IP Communications Services 1-15

 IP Communications Services Overview 1-15

 Call Processing Deployment Models 1-15

Business Ready Branch Solution Summary 1-18

CHAPTER 2

Planning and Designing the Business Ready Branch Solution 2-1

Security 2-1

 Securing the WAN 2-1

 Securing the WAN Overview 2-2

 Direct IPSec Encapsulation 2-2

 IPSec-Protected GRE 2-5

 Static Point-to-Point GRE 2-5

 Dynamic Point-to-Point GRE 2-5

 Dynamic Multipoint GRE 2-6

 WAN Security Summary 2-8

 Defending the Perimeter 2-8

 IP Communications 2-10

 Quality of Service Overview 2-11

 Delay 2-11

 Delay Variation (Jitter) 2-12

 Packet Loss 2-12

 Provisioning the WAN 2-13

 Service Provider QoS 2-14

Call Admission Control	2-15
IP Telephony	2-15
IP Telephony for the Office	2-16
Provisioning for Voice	2-17
Centralized Call Processing with CallManager	2-20
Local Call Processing with CallManager Express	2-26

CHAPTER 3

Choosing a Branch Office Platform 3-1

APPENDIX A

Sample Business Ready Branch Configuration Listings A-1



CHAPTER

1

Business Ready Branch Solution Overview

The Cisco Business Ready Branch or Office solution enables customers to deploy high value network services such as security, IP telephony, business video, and content networking over a variety of WAN technologies. The goal is to make these services fully available to all employees, no matter where they are located.

This chapter provides an overview of the Business Ready Branch Solution, and includes the following sections:

- [Introduction](#)
- [Understanding the Business Ready Branch Solution](#)
- [Service Building Blocks](#)
- [Business Ready Branch Solution Summary](#)

Introduction

This design guide describes how to design a Business Ready Branch or autonomous Business Ready Office where corporate services such as voice, video, and data are converged onto a single office network. This guide is targeted at network professionals and other personnel who assist in the design of branch or commercial office networks.

This guide assists the network designer in successfully designing a branch or an autonomous office. There are numerous combinations of features, platforms, and customer requirements that make up an office design. This design guide focuses on integrated voice, security, and data services within a single access router.

A two-pronged approach was used for testing the access routers: router functionality based on select office profiles (that is, branch offices that contained a specific number of users, PSTN trunks, and a relative amount of WAN bandwidth for that size office); and raw packets-per-second (pps) performance where results were recorded with a graduating number of features being enabled.

The results from this two-pronged approach provide the network designer with the confidence to accurately recommend the specific access router platform that meets customer office network requirements. This document guides the network designer through an example branch office network design, and shows how performance test results are used to select an appropriate office router.

■ Understanding the Business Ready Branch Solution

For more information, see the *Voice and Video Enabled IPSec VPN (V³PN) Solution Reference Network Design* at the following URL:

http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/V3PN_SRND/V3PN_SRND.html.

Various other sources are referenced throughout this document.

Understanding the Business Ready Branch Solution

The Business Ready architecture consists of two deployment models: branch and autonomous office. Although both deployment models are very similar, there are some distinct features and markets that apply to each. Following are some of the attributes that define each deployment model.

The Business Ready Branch has the following attributes:

- An extension of the enterprise campus
- All corporate resources centrally located
- Multiple centrally-managed sites
- Centralized call processing using Cisco CallManager and Cisco Survivable Remote Site Telephony (SRST) for voice
- WAN access—typically T1 to T3
- WAN is primarily a private WAN or Multiprotocol Label Switching (MPLS) virtual private network (VPN) or IP Security (IPSec) VPN over the Internet
- Up to 240 users

The Business Ready Office has the following attributes:

- Mini-campus network
- All corporate resources local
- Single site, or a loose confederation of autonomous offices
- Local call processing using Cisco CallManager Express (CCME) and Cisco Unity Express (CUE) for voice mail
- WAN access—typically DSL up to multiple T1s
- WAN is primarily an IPSec VPN over the Internet
- Remote access VPN is integral for providing mobile worker access to the corporate resources
- Up to 100 users (based on CUE module support of mailboxes)

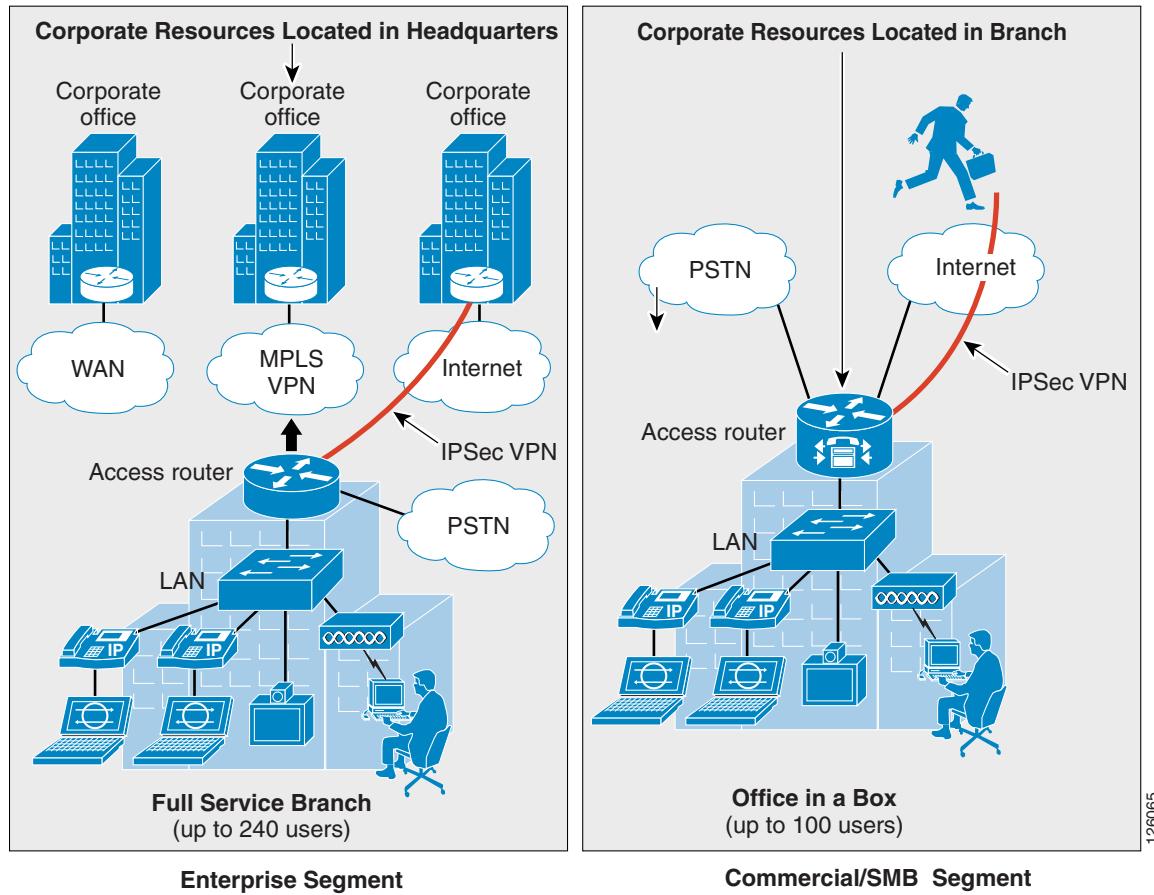
The router currently used in the office as a key component in the Business Ready architecture is no longer simply an access router providing WAN or Internet connectivity, but an integral part of multiple service architectures that are converged onto a single packet-based network. The office network consists of several services integrated into either a single or a small number of networking devices. These devices are typically a modular access router with an integrated Ethernet switch or an access router coupled with an external Ethernet switch.

Wireless access points (APs) may also be used in addition to or in place of the Ethernet switch for end device connectivity. When these offices go beyond the 240 users for the branch or 100 users for the autonomous office, their design resembles that of a campus, so campus design guidelines must be followed. The campus design guidelines are found at the following URL:

http://www.cisco.com/en/US/netsol/ns815/networking_solutions_program_home.html.

Figure 1-1 shows a high level view of these two office deployment models and their associated market segment.

Figure 1-1 Business Ready Branch Overview



Service Building Blocks

This section includes the following topics:

- Service Building Blocks Overview
- WAN Services
- LAN Services
- Security
- IP Communications Services

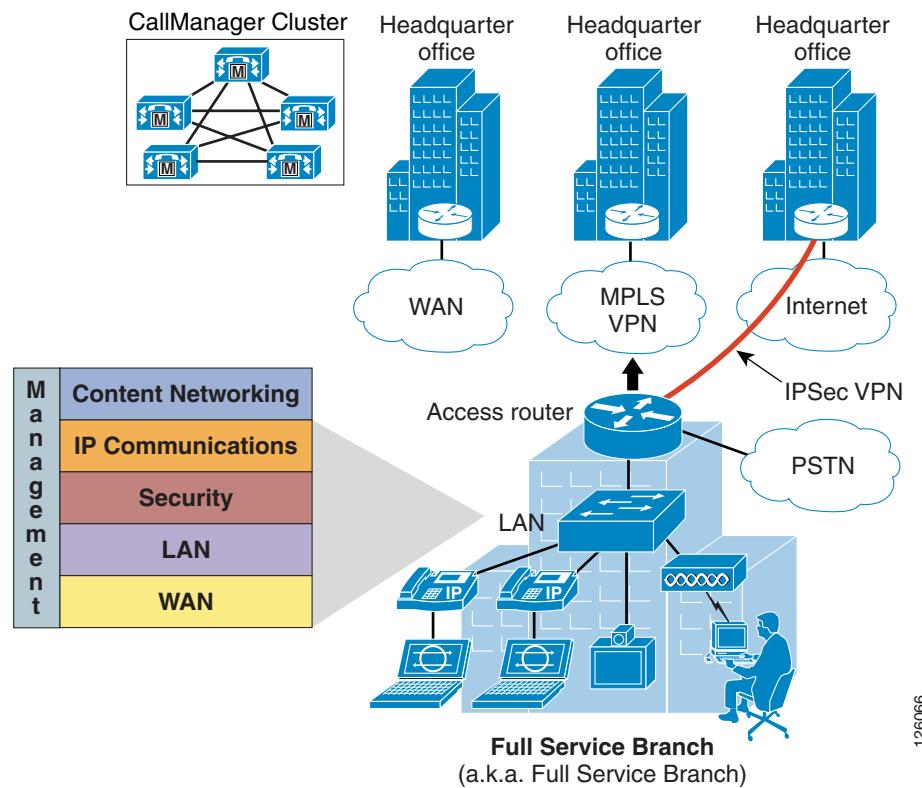
Service Building Blocks Overview

The Business Ready Branch or Office solution uses a layered model in which services are organized into specific categories or building blocks. These building blocks can then be combined to fit specific customer service needs.

The branch and autonomous office have distinct characteristics that influence the combination of building blocks that may be implemented. With the Business Ready Branch, corporate resources such as server farms, IP telephony call processing agents (CallManager), and Internet access are located in a headquarters or regional office and are accessed over the WAN connection. With the autonomous Business Ready Office, all corporate resources and Internet access are located locally within the office. These characteristics as well as the WAN deployment option affect the platform and type of security services that are deployed in the office. The following sections explore each of the service building blocks and describe the choices and guidelines when building the branch.

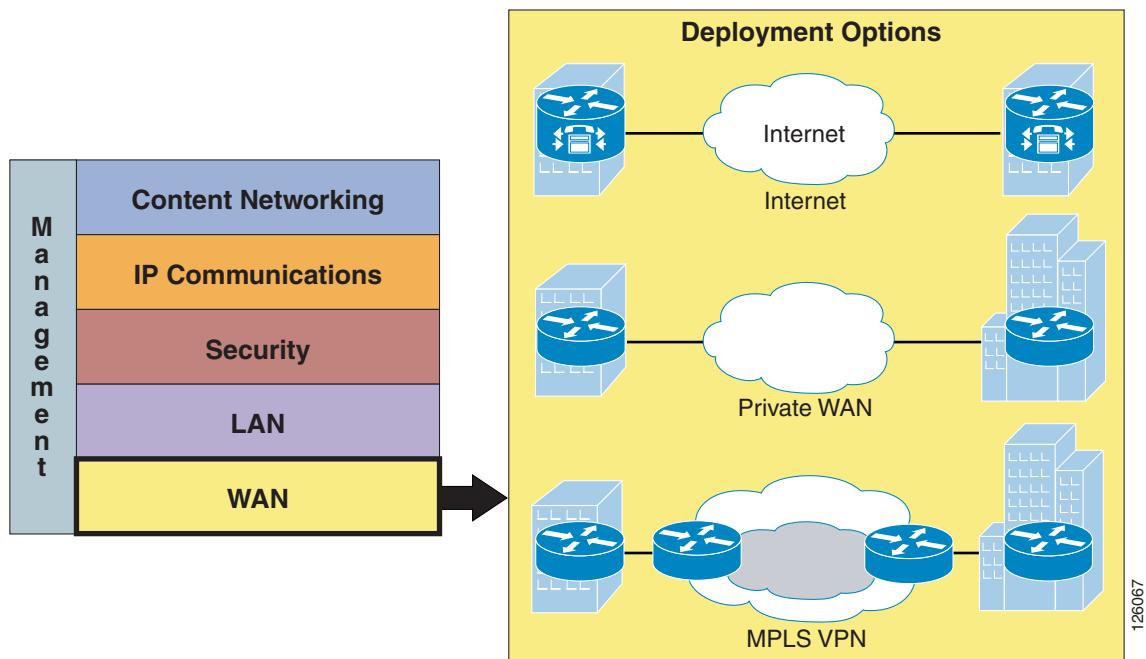
[Figure 1-2](#) shows an exploded view of the service building blocks that make up the office network.

Figure 1-2 Business Ready Branch Building Blocks



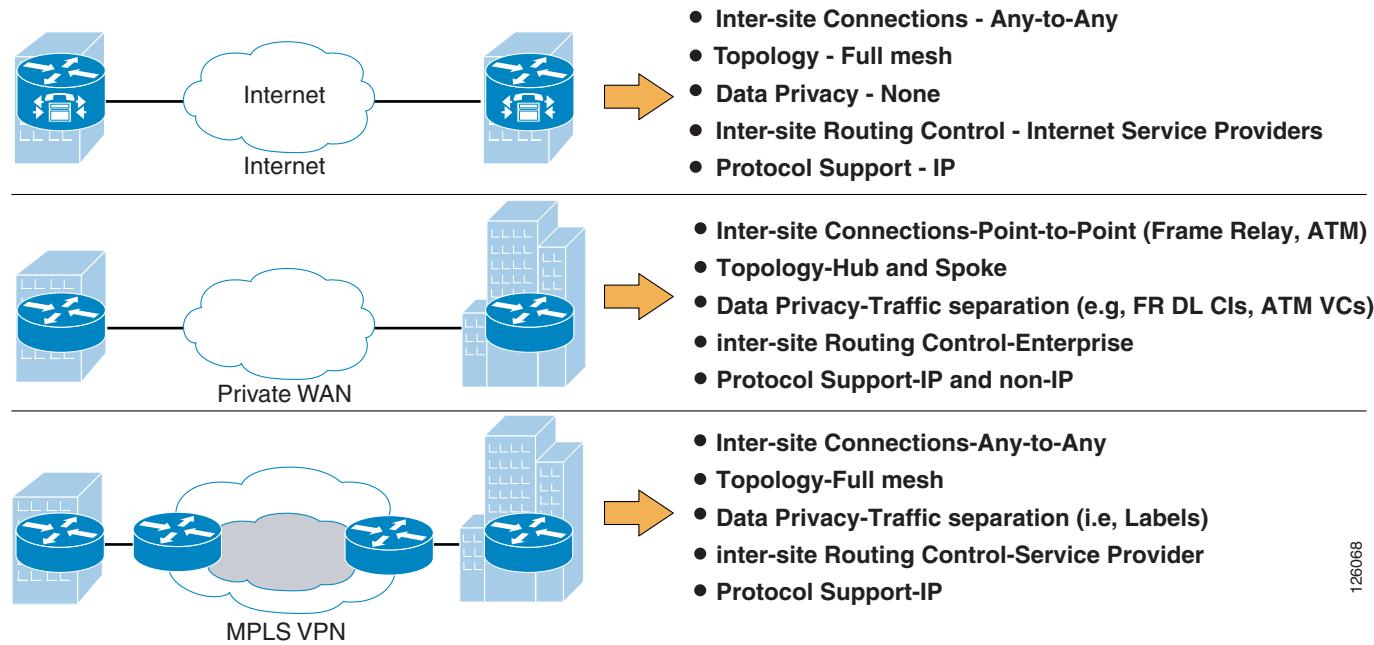
WAN Services

Starting at the bottom of the stack, WAN services provide the foundation for the Business Ready Branch or Office connection to the outside world. The WAN services building block consists of three fundamental deployment options, each with its own set of associated attributes as shown in [Figure 1-3](#).

Figure 1-3 WAN Services

These attributes influence the use of specific features and require special considerations when designing a branch office. For example, if a branch office is connected to the Internet, an IPSec VPN may be required for data privacy between branch and home offices or mobile workers. Another example is Call Admission Control (CAC), which is required for IP telephony or video. These and other examples of services that are influenced by the WAN deployment model are discussed throughout this design guide.

[Figure 1-4](#) lists the WAN deployment options and some of their attributes that influence the design of the branch office.

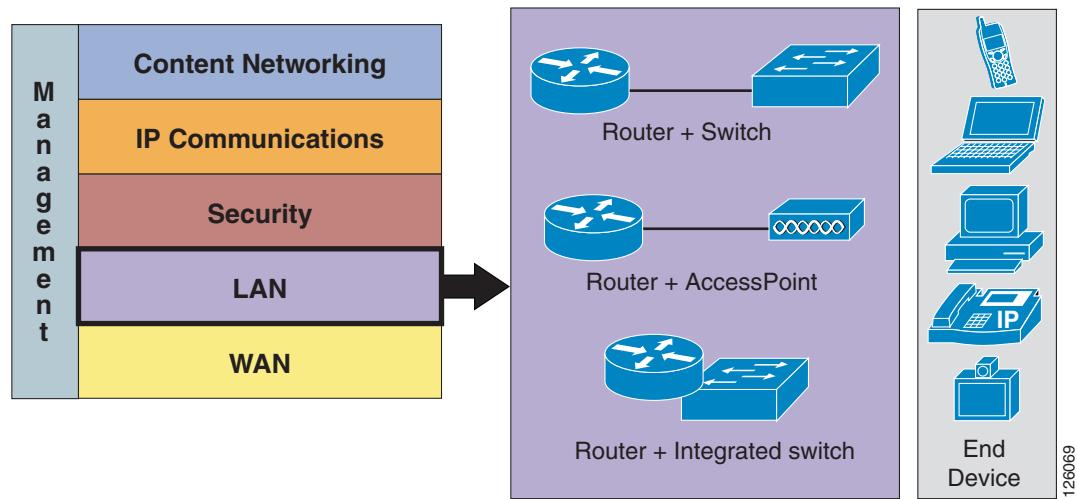
Figure 1-4 WAN Deployment Options

126068

LAN Services

LAN services provide end device connectivity to the corporate network within the office. With the convergence of services onto a single network infrastructure, devices such as computers, telephones, surveillance cameras, cash registers, kiosks, and inventory scanners all require the connection to the corporate network over the LAN. This assortment of devices requires simplified connectivity tailored to the demands of each device. For example, devices such as IP telephones or cameras may be powered using the LAN switch, automatically assigned an IP address, and be placed in a virtual LAN (VLAN) to securely segment them from the other devices.

Wireless APs may be used to provide secure mobile access for laptop computers, scanning devices, wireless IP phones, or kiosks where wiring is difficult to install. These are just a few examples of the LAN services that are used in the Business Ready Branch or Office solution. [Figure 1-5](#) shows the three different physical configurations that may be used in the LAN services building block.

Figure 1-5 LAN Services

The three configurations that are referenced in this document are as follows:

- Access router connected to a physically separate Cisco Catalyst switch
- Access router with an integrated switch
- Access router and an AP

Table 1-1 highlights some of the advantages and disadvantages of each option.

Table 1-1 LAN Equipment Combinations

LAN Service Configuration	Advantages	Disadvantages
Access router with external switch	<ul style="list-style-type: none"> • Good scaling properties. Switches may be stacked or use larger modular chassis. • Extensive feature support. • Typically lower initial per port equipment than using integrated switch. • End devices may be powered inline by connecting to a powered switch. 	<ul style="list-style-type: none"> • Additional device to manage • Per switch recurring maintenance costs

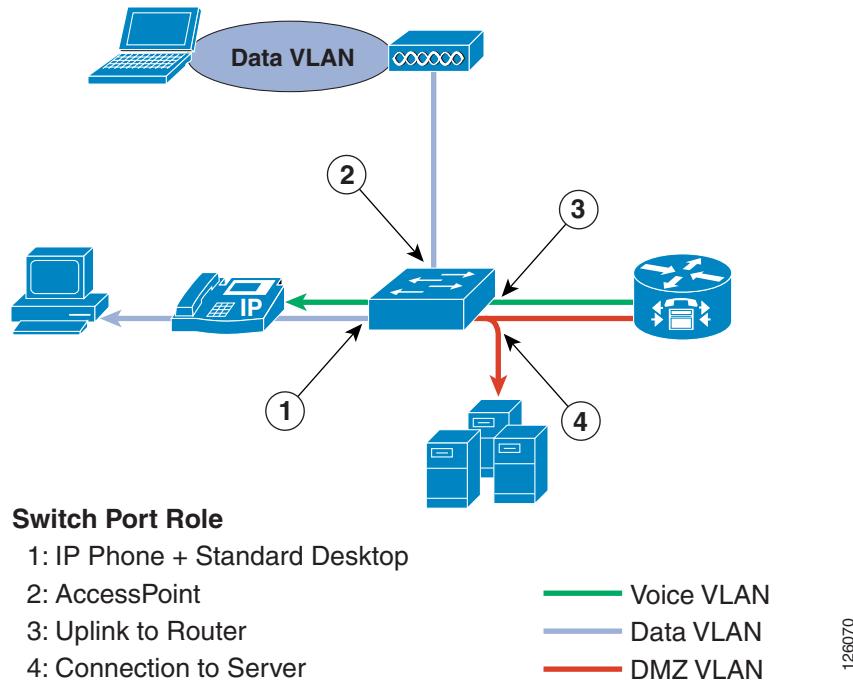
Table 1-1 LAN Equipment Combinations (continued)

LAN Service Configuration	Advantages	Disadvantages
Access router with integrated switch	<ul style="list-style-type: none"> One box solution. Lower TCO than using external switch. Single device with single maintenance contract. Typically higher initial per port equipment costs. End devices may be powered inline by connecting to a powered switch. 	<ul style="list-style-type: none"> Lower port densities. Typically used for small offices especially when deployed with other services (for example, IP telephony). Do not have feature parity with external switches. Depending on the platform, an external power supply may be required for inline powering of end devices.
Access router with AP	<ul style="list-style-type: none"> Flexible endpoint deployment where wiring is not necessary. Quick deployment—no need for wiring. Support for mobile workers. May be deployed as an overlay to a wired LAN. May be powered inline by switch. 	<ul style="list-style-type: none"> Low end point capacity per AP. Typically 10 to 20 devices per 802.11b AP. Special care must be taken to secure a wireless network. Must use Cisco wireless cards to support Basic Security features (for example, TKIP, MIC).

Some of the other considerations when deploying an office LAN are which devices and services must be supported. The following list describes the other considerations of the LAN service building block:

- Quality of service (QoS)—Required to maintain high-quality voice or video within the local LAN or wireless LAN. This includes the defining of trust on ports to prohibit unauthorized use of QoS for preferential treatment of traffic on the office network.
- Virtual LAN (VLAN)—Required to segment the office to provide logical division between services. For example, IP telephony should reside on its own VLAN, separate from that used by the data network.
- 802.1q VLAN tagging—Provides trunking services for IP phones and uplinks to APs or the access router for network routing.
- Inline power—Provides power to the IP phones, APs, or other IP-enabled devices (for example, IP cameras) over the Ethernet cable.
- Port security—Limits the number of MAC addresses allowed on an access port.

For the office network considered in this design guide, the Smartports feature and its canned port templates are used for the LAN switch configuration. [Figure 1-6](#) shows a high level diagram of the devices and port profiles used in the office testing.

Figure 1-6 Office Port Profiles

More information on Smartports can be found at the following URL:
http://www.cisco.com/en/US/docs-switches/lan/catalyst3750/software/release/12.2_40_se/configuration/guide/swmacro.html.

Security

This section includes the following sections:

- Security Overview
- Securing the WAN
- Defending the Perimeter

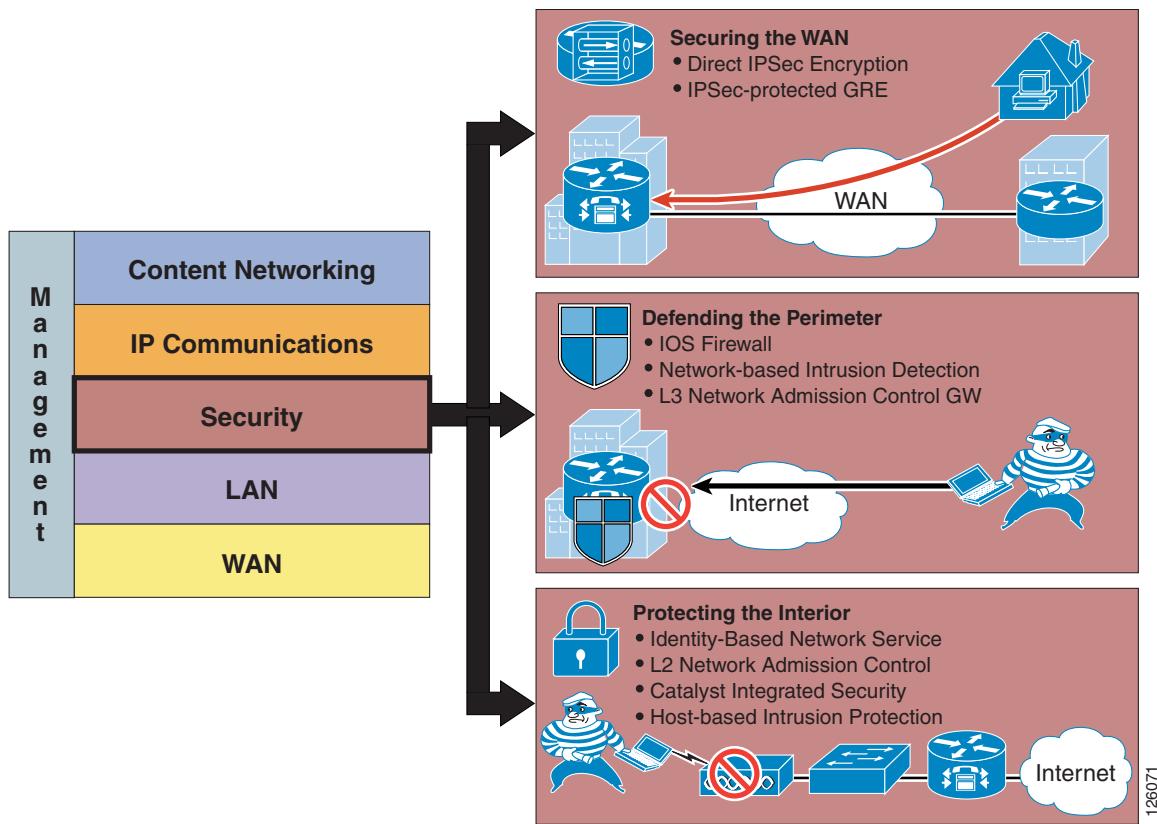
Security Overview

Security is deployed in three places in the office network: on the WAN, on the perimeter between the WAN and the LAN, and on the office LAN.



This document includes only the features integrated in the access router. Therefore, this version of the design guide covers only those integrated features used for securing the WAN and defending the perimeter of the branch office.

Figure 1-7 shows the breakdown of the security building block and the associated technologies used for securing each of these places in the office network.

Figure 1-7 Security Services

Securing the WAN

Securing the WAN consists of using IP Security (IPSec) to secure data traffic traversing the WAN. The IPSec protocol provides data confidentiality through strong encryption, endpoint authentication, and data integrity, and is used as an overlay to the Internet, an enterprise private WAN, or MPLS VPN.

Some of the considerations when securing the WAN are as follows:

- Type of WAN—Internet, private WAN, or MPLS VPN
- Type of traffic to be sent over the VPN, such as IP unicast or IP multicast
- Best VPN deployment option, such as Direct IPSec Encapsulation or IPSec-protected generic routing encapsulation (GRE)
- Configuration complexity or size
- Authentication method—Preshared keys, digital certificates, EZVPN (EZVPN does not support GRE)
- Use of high availability, dual head ends, using routing protocols such as Hot Standby Routing Protocol (HSRP)

Deploying IPSec VPN over the Internet

Using IPSec VPN has become a common method of securing enterprise traffic over the Internet. Each available IPSec VPN option has advantages and disadvantages, which are mentioned in this section and described in more detail in [Chapter 2, “Planning and Designing the Business Ready Branch Solution.”](#)

The following are some of the considerations when deploying IPSec VPN as a means of connecting offices:

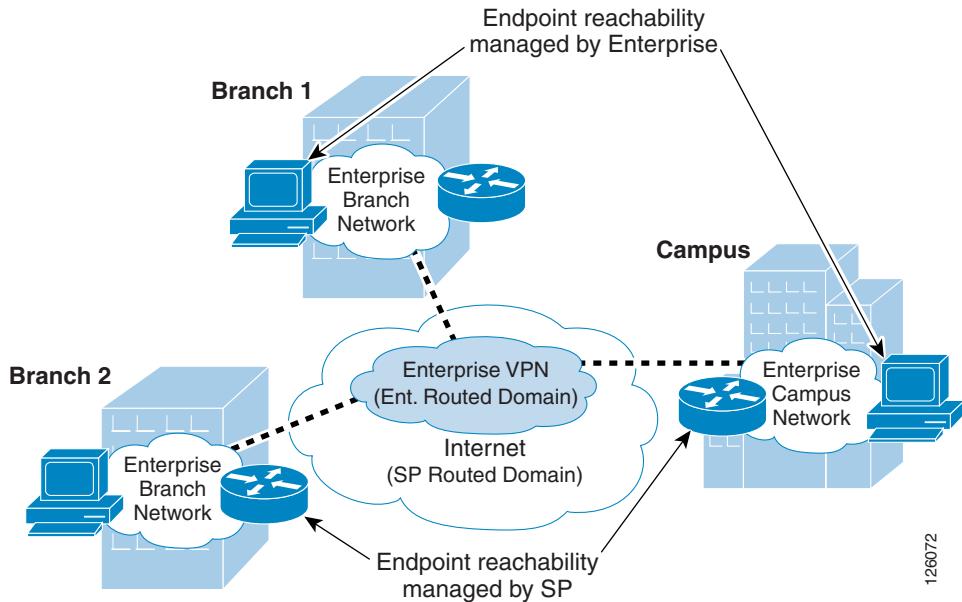
- Dynamic IP addressing—Although branch offices typically have T1 access link to the Internet with fixed IP addresses, cable or DSL are viable alternative access links, and dynamic IP addressing may need to be accommodated by the VPN technology used.
- Level of acceptable quality—if voice or video traverses the WAN, then determining the level of acceptable quality over the Internet must be considered. This may require the negotiation of service level agreements with service providers.
- Higher level of security—Support of a higher level of security may be required for the office network because of the direct connection to the public Internet. Split tunneling of traffic for local Internet used at the branch office requires a firewall for protection.
- Type of authentication—May include EZVPN, digital certificates, or static pre-shared keys. The use of digital certificates is recommended because of its high level of security and ease of key management when deploying several branch offices.

Deploying IPSec VPN over a Private WAN or MPLS VPN

Enterprises are now considering using IPSec VPN technology to provide data privacy over their private WANs because of new privacy laws. This is a viable solution but has the additional challenge of integrating into the enterprise network.

There are two fundamental components that need to be considered when using IPSec VPN for providing data privacy over a private WAN or MPLS VPN: using IPSec for securing the data, and the routing control plane required for establishing endpoint reachability over the VPN. In the traditional IPSec VPN deployment, the enterprise controls the endpoints that send the data to be protected and therefore controls the routing or reachability between the endpoints.

The service provider (SP) has no knowledge of the IP-addressed endpoints of the enterprise. The SP controls the routing between the enterprise VPN routers where the SP owns and controls the reachability of the IP addresses that are assigned to the SP-connected interface of the VPN router. [Figure 1-8](#) shows this relationship.

Figure 1-8 IPSec VPN Overview

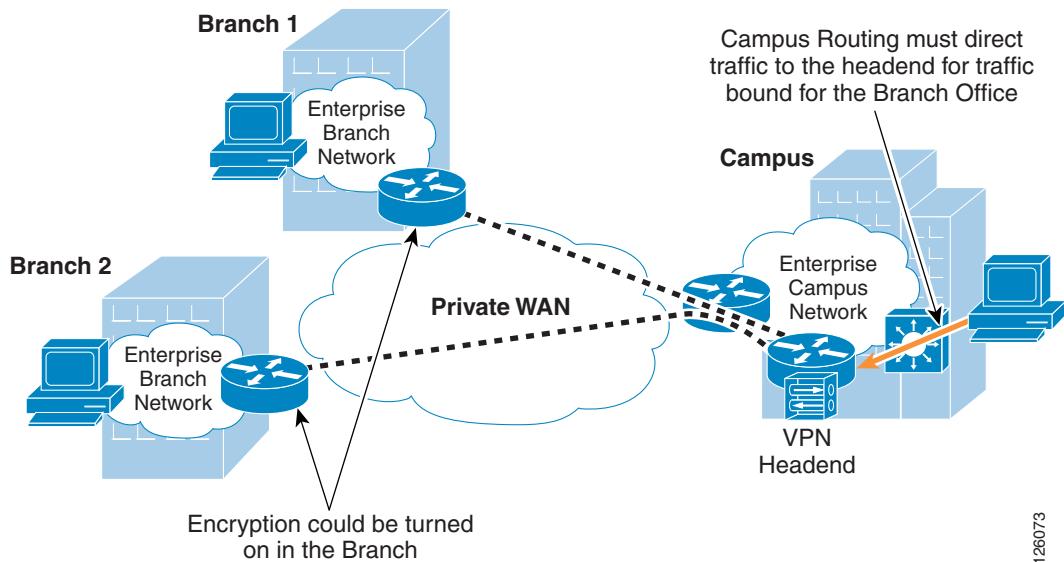
126072

This relationship of the two autonomous routed domains (enterprise and SP) is a fundamental characteristic of a typical IPSec VPN deployment. Because the enterprise does not have control over the WAN routing, routing methods such as static, Reverse Route Injection (RRI), and dynamic are used to establish reachability between the endpoints connected over the VPN.

When deploying VPN as a means of data privacy between branch offices in an existing enterprise private WAN or MPLS VPN, one consideration is how to incorporate this autonomous routing domain. In either of these WAN deployments, the enterprise network already understands how to route between endpoints, so inserting a VPN into the existing network now requires the redirecting of traffic through the local VPN router for encryption. This can be fairly straightforward for the branch office because IPSec can be turned on in the WAN-connected access router.

However, on the campus side of the network, this same approach is probably not permitted because this means turning on IPSec in a WAN-aggregation router. In this case, the installation of a separate VPN headend in the campus is required, and network routing must be modified to steer traffic destined to the branch offices through the VPN headend.

[Figure 1-9](#) shows this private WAN or MPLS scenario.

Figure 1-9 Inserting VPN into an Existing Private WAN or MPLS VPN

126073

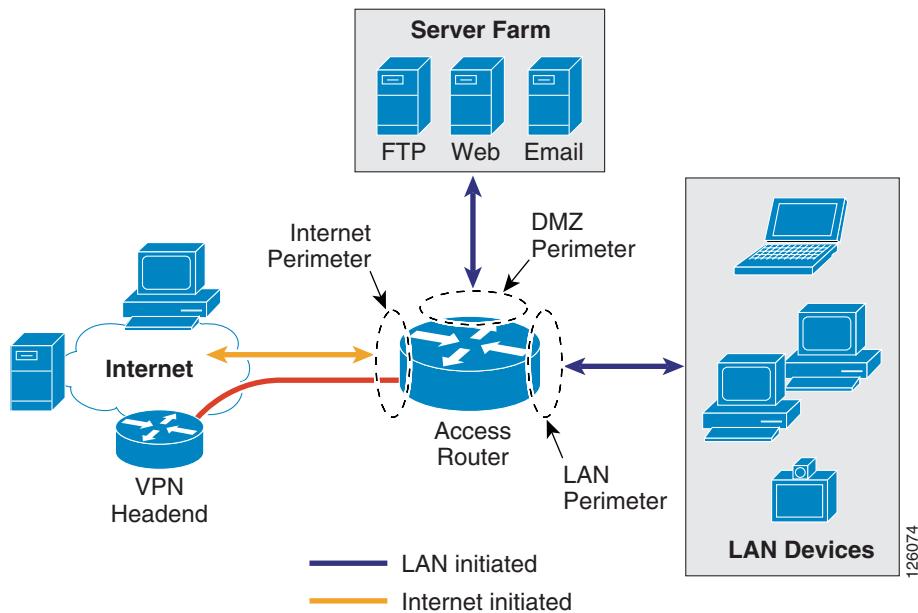
This complexity of deploying IPSec VPN over an existing private WAN or MPLS VPN is one of the primary challenges of securing the WAN, and you must plan comprehensively to ensure a seamless implementation.

[Chapter 2, “Planning and Designing the Business Ready Branch Solution,”](#) provides more detailed information to aid the network designer in choosing the best option for securing the WAN.

Defending the Perimeter

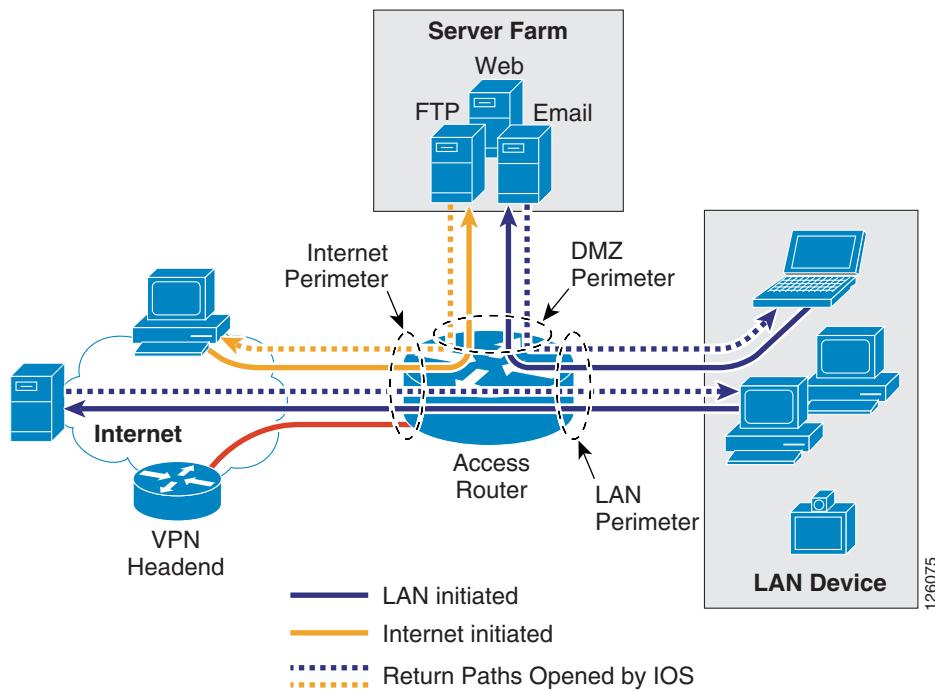
This section provides a high level overview of the Cisco IOS Firewall, access control lists (ACLs), and Cisco Intrusion Detection System (IDS) security features implemented at the perimeters of the office network. This section introduces an overview of these features, with implementation recommendations to follow in [Chapter 2, “Planning and Designing the Business Ready Branch Solution.”](#)

[Figure 1-10](#) shows an example of the perimeter of an office network.

Figure 1-10 Office Network Perimeter Defined

Cisco IOS Firewall and ACLs

The Cisco IOS Firewall provides integrated, inline security services and provides lock-tight, stateful security and control for each protocol traversing the office router. [Figure 1-11](#) shows how traffic flows through the different office perimeters.

Figure 1-11 Traffic Flows through the Office Network Perimeters

ACLs provide strict control of traffic entering the office network (represented by the solid arrows) and the Cisco IOS Firewall opens and inspects the return path for traffic (represented by the dotted arrows) initiated from within the office network.

**Note**

For more information on configuring Cisco IOS Firewall, see the following URL:
<http://www.cisco.com/en/US/products/sw/secursw/ps1018/index.html>.

[Chapter 2, “Planning and Designing the Business Ready Branch Solution,”](#) describes in more detail how the ACLs and IP inspect commands of the Cisco IOS Firewall are configured to defend the perimeters of the office network.

Intrusion Detection System

The Cisco IOS Intrusion Prevention System (IPS) acts as an inline intrusion detection sensor, watching packets and sessions as they flow through the router and scanning each packet to match any of the Cisco IOS IPS signatures. When it detects suspicious activity, it responds before network security can be compromised and logs the event through Cisco IOS syslog or the Cisco Secure Intrusion Detection System Post Office Protocol. The network administrator can configure the Cisco IOS IPS to choose the appropriate response to various threats. When packets in a session match a signature, the Cisco IOS IPS takes any of the following actions, as appropriate:

- Sends an alarm to a syslog server or a centralized management interface
- Drops the packet
- Resets the connection

Cisco developed its Cisco IOS software-based intrusion prevention capabilities and Cisco IOS Firewall to be flexible, so that individual signatures can be disabled in case of false positives. Generally, it is preferable to enable both the firewall and Cisco IOS IPS to support network security policies. Each of these features may be enabled independently and on different router interfaces. [Chapter 2, “Planning and Designing the Business Ready Branch Solution,”](#) provides recommendations on how each of these features is implemented to secure the office network perimeter.

The following are considerations when applying ACLs, configuring Cisco IOS Firewall, and Cisco IDS:

- With Release 12.3.7T and earlier, Cisco IOS Firewall and IDS use a common session-state inspection machine, so router performance impact is nearly the same when using the Cisco IOS Firewall alone or Cisco IOS Firewall and IDS together. This is true for both software-based IDS and hardware-based IDS (NM-CIDS). In fact, a slight decrease in performance is observed when using hardware-based IDS, because of the additional processing required for copying the packet to the IDS module. Even so, the benefit of using hardware-based IDS is the increased number of attack signatures that are monitored.
- Before Release 12.3.8T, ACLs on the Internet perimeter were checked before and after encryption. Release 12.3.8T removed this requirement. See the following URL for more complete information concerning the use of ACLs and IPSec encryption:
http://www.cisco.com/en/US/docs/ios/12_3t/12_3t8/feature/guide/gt_crpk.html.
- With Release 12.3.8T and later, software-based IPS is introduced. IPS moves the packet inspection into the packet path rather than working in a promiscuous manner receiving packet copies for inspection. IPS provides better protection but does impact router performance.

- NM-CIDS that are typically integrated in an office router are limited to 45 Mbps. Cisco recommends that the IDS run on all office perimeter interfaces, but tuning may be required to prevent oversubscribing the IDS monitoring capabilities. Start with the default signatures and filter out select traffic using ACLs and possibly removing IDS from monitoring some interfaces that impose less of a threat to the network (for example, voice VLAN).
- For large office networks, Cisco IOS Firewall default inspection limits must be carefully considered. For example, if the WAN perimeter is configured to deny LAN traffic, and Cisco IOS FireWall IP inspection is responsible for opening the return path from IP phone registration requests, IP phone registration can take an excessive amount of time. This is because of exceeding the default half-open sessions limits of the Cisco IOS Firewall.

For more information on Intrusion Detection Systems, see the following URL:

<http://www.cisco.com/en/US/products/sw/secursw/ps2113/index.html>

Network Admission Control

Network Admission Control (NAC) provides a higher level of protection to network devices by determining the health of the device before allowing it access to the office network. NAC works at Layer 3; when a device attempts to contact another device beyond its own local subnet, the office access router can facilitate a security posture check. This is done by communicating with a software agent on the device, requesting its anti-virus posture, and comparing the received credentials against a database that specifies the minimum requirements for network access. If a PC does not pass the requirements for access, that PC is denied access and the network administrator is notified so that remedial action can be taken.

For additional information on NAC, please see the following URL:

http://www.cisco.com/en/US/netsol/ns617/networking_solutions_sub_solution_home.html

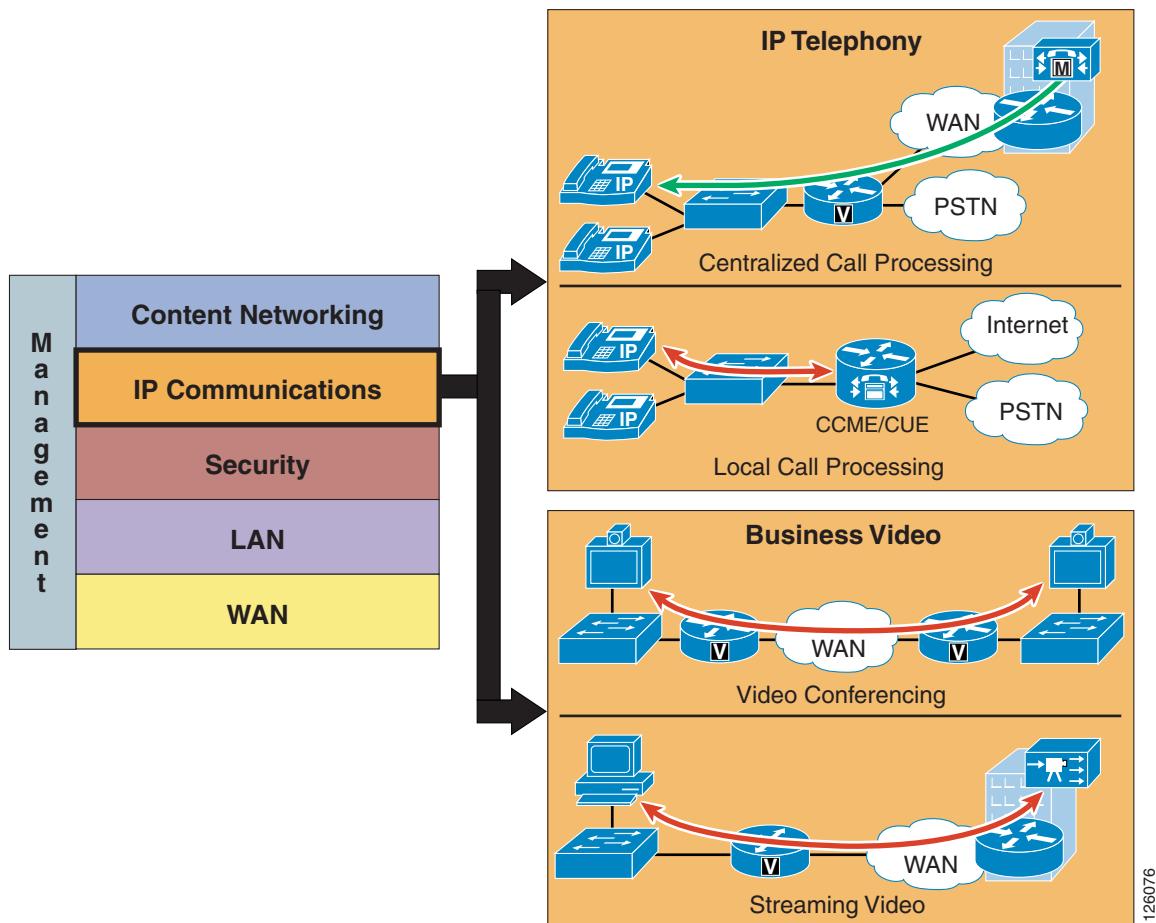
IP Communications Services

This section includes the following topics:

- [IP Communications Services Overview](#)
- [Call Processing Deployment Models](#)

IP Communications Services Overview

IP telephony and business video are IP communications services used between users to carry out day-to-day business, as shown in [Figure 1-12](#).

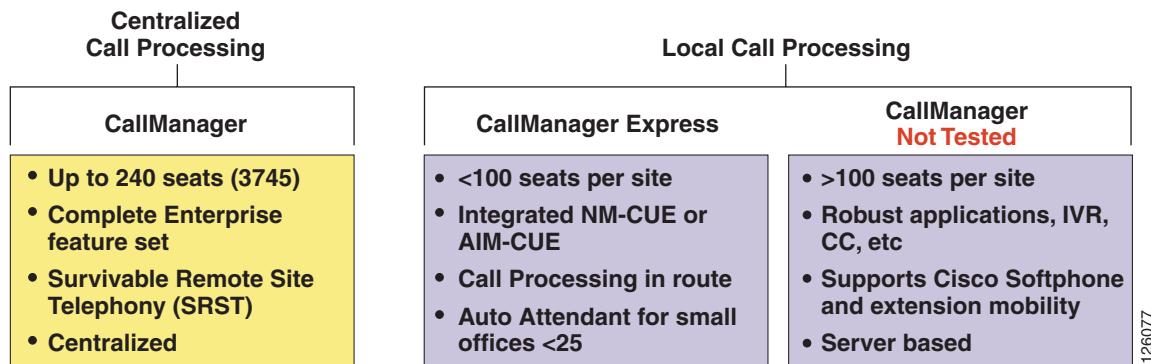
Figure 1-12 IP Communications Services

126076

Call Processing Deployment Models

This design guide examines two deployment models: Centralized Call Processing and Local Call Processing. The call processing models tested depended on the office type: branch or autonomous office. The branch office used CallManager deployed with Centralized Call Processing, and the autonomous office used CallManager Express coupled with Unity Express for voice mail.

[Figure 1-13](#) shows the general positioning of the two call processing methods discussed in this design guide.

Figure 1-13 IP Telephony Call Processing Positioning

The choice of whether to adopt a centralized call processing or distributed local call processing approach for a given site depends on a number of factors such as the following:

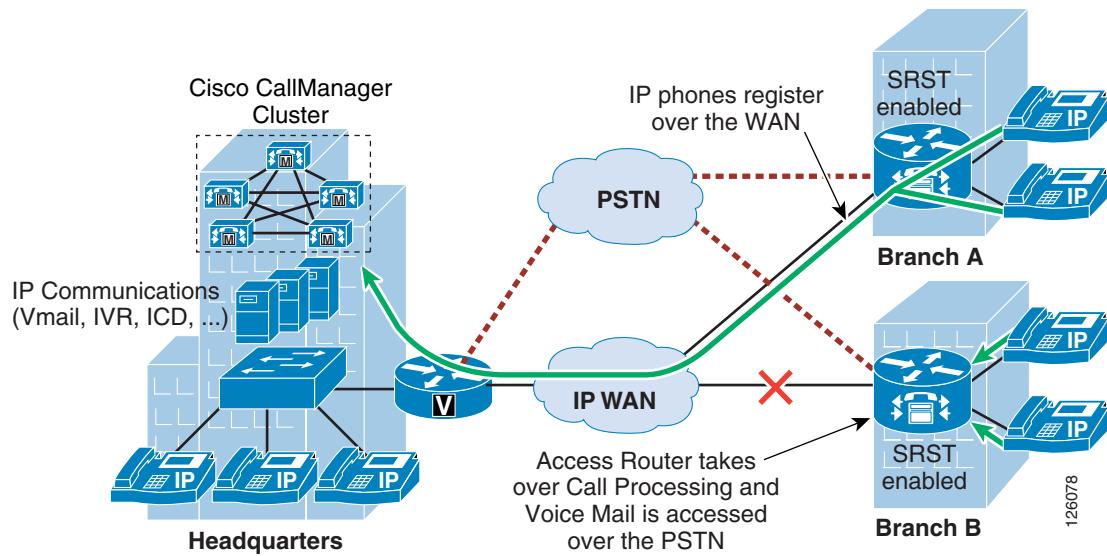
- IP WAN bandwidth
- One-way delay to remote sites
- Criticality of the voice network
- Feature set needs
- Scalability
- Ease of management
- Cost

If a distributed local call processing model is deemed more suitable for customer business needs, the choices include installing a local Cisco CallManager server or running the Cisco CallManager Express on the branch router. This design guide focuses on the use of Cisco CallManager Express for Local Call Processing option for the office.

For more detailed information on designing IP telephony networks, see the following URL:
http://www.cisco.com/en/US/products/ps6788/Products_Sub_Category_Home.html.

Centralized Call Processing

Centralized Call Processing is primarily used to serve branch offices where a centralized CallManager cluster and Unity Voice Mail system resides in the headquarters site, and provides all the call processing and voice mail services for the remote IP phones located in the branch office. (See [Figure 1-14](#).)

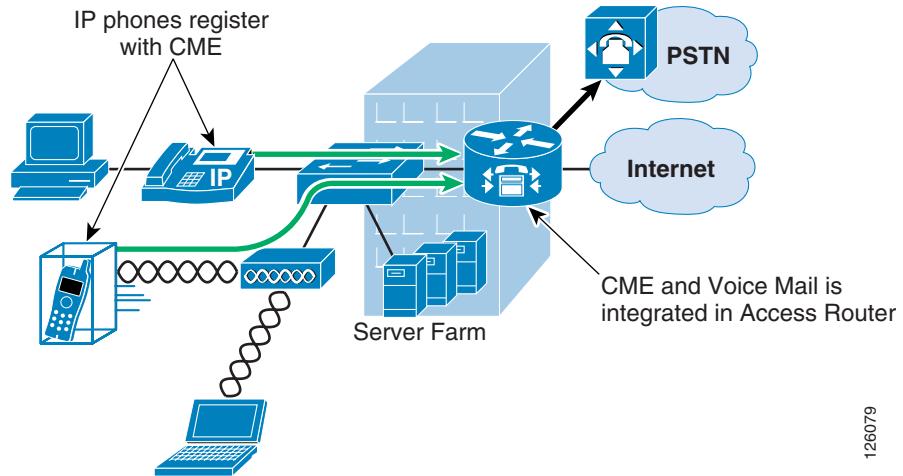
Figure 1-14 Centralized Call Processing

This call processing model eases branch deployment where the enterprise simply connects IP phones to the branch LAN and the phones then register to the CallManager cluster over the WAN. When registered, the IP phone automatically downloads its pre-configured profile and is ready to use. The access router is configured with the Survivable Remote Site Telephony (SRST) feature to provide backup call processing in case contact is lost to the CallManager cluster; for example, during WAN failure.

Another important consideration when deploying Centralized Call Processing is CAC, which limits the number of calls that may be placed over the WAN to maintain consistent high quality voice. This requires the proper provisioning of QoS on the WAN interface, and the configuration of CallManager such that call attempts that exceed the number of calls for which the WAN is provisioned receive a busy signal. CAC and WAN QoS ensure high voice quality for calls placed over the WAN in the Centralized Call Processing deployment model.

Local Call Processing

Local Call Processing is used in the autonomous office where CallManager Express, a software feature in the access router, provides the local call processing and the Unity Express hardware module, NM-CUE, provides the local voice mail and auto-attendant services. (See [Figure 1-15](#)).

Figure 1-15 Local Call Processing

Business Ready Branch Solution Summary

This chapter has presented an overview of the many services that may be deployed in the Business Ready Branch or autonomous Business Ready Office. As mentioned previously, this design guide covers only the integration of IP telephony and security services within the access router. [Chapter 2, “Planning and Designing the Business Ready Branch Solution,”](#) discusses considerations when planning and designing an office network, [Chapter 3, “Choosing a Branch Office Platform,”](#) explains how to choose the right platform for your office network, and [Appendix A, “Sample Business Ready Branch Configuration Listings,”](#) provides a sample configuration listing.



CHAPTER

2

Planning and Designing the Business Ready Branch Solution

This chapter provides more detailed design information on services deployment, and includes actual router configuration. A more complete set of router configuration commands is provided in [Appendix A, “Sample Business Ready Branch Configuration Listings.”](#)

This section includes the following topics:

- [Security](#)
- [IP Communications](#)

Security

The level of network security that is deployed in an office typically depends on the WAN technology, size, and the security policies enforced by the customer. This section describes how to secure the WAN over a private WAN, MPLS VPN, or the Internet, and how to establish secure entry points into the office network by implementing strong perimeter security. This section includes the following topics:

- [Securing the WAN](#)
- [Defending the Perimeter](#)

Securing the WAN

This section includes the following topics:

- [Securing the WAN Overview](#)
- [Direct IPSec Encapsulation](#)
- [IPSec-Protected GRE](#)
- [Static Point-to-Point GRE](#)
- [Dynamic Point-to-Point GRE](#)
- [Dynamic Multipoint GRE](#)
- [WAN Security Summary](#)

Securing the WAN Overview

IPSec is used for securing the WAN, and there are two methods of applying IPSec: direct IPSec encapsulation and IPSec-protected generic route encapsulation (GRE). Both methods support the following:

- Static configurations that typically are used for point-to-point VPN connections between IPSec peering routers that have known IP addresses.
- Multipoint configurations for supporting hub-and-spoke topologies, where the spoke IP addresses may not be known or are dynamically assigned by the service provider.

In both direct IPSec encapsulation and IPSec-protected GRE, there are deployment options for static point-to-point IPSec tunnels and dynamic multipoint IPSec tunnels. As mentioned above, static point-to-point IPSec tunnels are manually configured between each pair of IPSec peering endpoints where endpoint IP addresses are known.

Multipoint dynamic IPSec tunnels use a multipoint configuration, typically on an IPSec VPN head end, which acts as a hub for multiple peering spoke routers in which spoke IP addresses are not known or are dynamically assigned by the DHCP server of the SP. IPSec tunnels are then dynamically created from the spoke (branch) router to the VPN head end when the inter-site traffic requires encryption. This multipoint configuration option saves the network administrator a considerable amount of configuration in the VPN head end but does pose new challenges when troubleshooting.

Although these two IPSec encapsulation methods are somewhat interchangeable, there are some distinct differences that make their use more appropriate to certain applications than others. The next four sections describe how each of these methods work and provide an understanding of which method best fits the application.

Direct IPSec Encapsulation

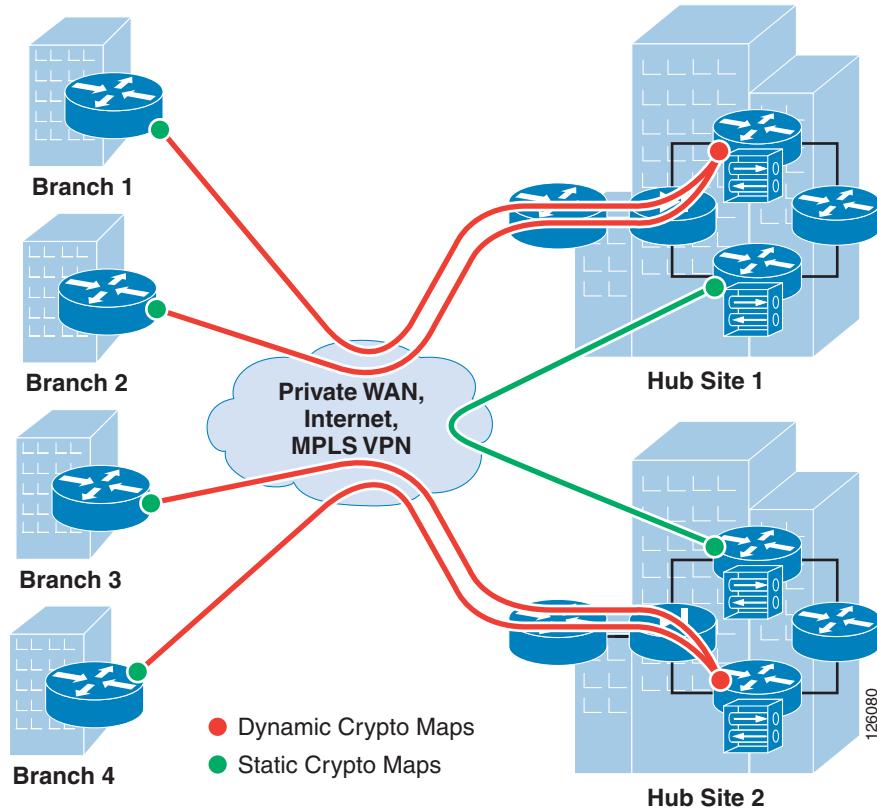
Direct IPSec encapsulation secures IP unicast traffic passing through an ACL configured on the VPN router. This ACL is used to select specific traffic for encryption, and provides the network administrator granular control over the encrypted traffic sent over the IPSec VPN. Direct IPSec encapsulation cannot carry typical routing protocols because it supports only IP unicast traffic, so the three following alternative methods of establishing endpoint reachability over the VPN are used:

- The underlying routing protocol, if the WAN is under enterprise control
- Static routing
- Reverse Route Injection (RRI)

Direct IPSec encapsulation is used primarily for VPN connectivity for mobile users, teleworkers, small branch offices with two or less local subnets, or for overlaying IPSec for IP unicast data encryption over an existing private WAN.

Direct IPSec encapsulation has static or dynamic configuration options, using crypto maps for specifying IPSec parameters for encrypting traffic. Static crypto maps are typically used for point-to-point IPSec tunnels where IP addresses of the peering routers are static. Dynamic crypto maps are typically used on a VPN head end or hub where IPSec tunnels are dynamically created by peering spoke routers whose IP addresses are dynamically assigned.

[Figure 2-1](#) shows where static and dynamic crypto maps are typically used.

Figure 2-1 Static and Dynamic Crypto Maps

Static crypto maps are configured in both of the peering routers, and use ACLs for selecting specific traffic for encryption. This enables tunnel creation to be initialized by either end of the IPSec tunnel. Static crypto maps also require that the IP address of the peering router must be known and configured in the static crypto map. Because a static crypto map is configured for every peer, this option is best suited for point-to-point IPSec connections between hub sites or in spoke routers when connecting to a VPN head end or for small meshed IPSec VPNs.

Another option is to use a dynamic crypto map in a headend router. This option is better suited to routers that act as a hub for multiple spoke sites. The dynamic crypto map option streamlines the head end configuration and automates the creation of IPSec tunnels initiated by the spoke routers. Because the IP address of the hub or VPN head end is static, a static crypto map is configured in the branch router, and an ACL is used to specify the traffic to be encrypted. During the IPSec negotiation between the peering routers, an IPSec Security Association (SA) is automatically created in the head end with the inverse of each line of the ACL of the peering router. This defines the encrypted return path back through the head end to the branch router. Because there are no ACLs configured in the head end, this implies that the branch or spoke router must initialize the tunnel.



Note Although not required, ACLs may be used with dynamic crypto maps to specify which traffic is allowed to initiate a tunnel.

Also, because an IPSec SA is created in the head end for each line in the ACL, the number of ACLs in the branch routers should be kept to a minimum. Even so, using dynamic crypto maps is best suited for peering with branch offices with a small number of subnets or a set of subnets that can be summarized to minimize the number of lines in the local ACL and the resulting SAs required in the headend router.



Note GRE may also be used as a method of minimizing the number of SAs required in the head end, because the ACL has to specify only GRE traffic for encryption. This is discussed further in the following sections.

One significant difference when using dynamic crypto maps rather than static crypto maps is that the tunnel must always be initiated by the branch router, because the head end has no knowledge of its peering routers until the IPSec tunnel is initiated by the remote.

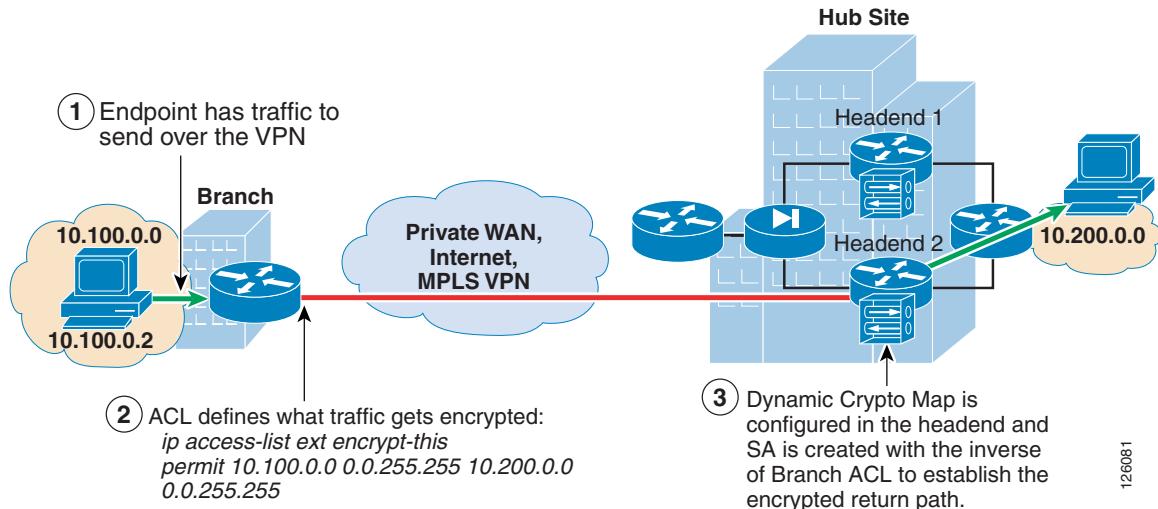
Dynamic crypto maps require a constant flow of traffic to prevent each SA (that is, an SA for each line of the ACL) from timing out because of inactivity. This “always-on” connection is important in a branch office deployment because quite possibly there can be traffic that originates from the hub site to access information at any one of the subnets in the branch office. This required “always-on” connection can easily be accomplished by any periodic source of traffic such as Simple Network Management Protocol (SNMP) polls from a network management station, IP phone keepalives, or the configuration of Service Assurance Agent (SAA) probes to prevent the SA from timing out.



Note SAA can be configured to simply send periodic Internet Control Message Protocol (ICMP) probes through the encryption ACLs in the branch router.

Figure 2-2 shows an example of the use of dynamic crypto maps.

Figure 2-2 Using Dynamic Crypto Maps



IPSec-Protected GRE

IPSec-protected GRE uses a GRE tunnel to encapsulate data traffic before passing through the ACL for encryption. Using GRE allows IP unicast, IP multicast, and other non-IP protocols to be encapsulated and transported over the IPSec VPN. The GRE tunnel is configured in the encryption ACL, and an IP routing protocol is used to steer traffic through the now IPSec-protected GRE tunnel.

Because dynamic routing protocols are typically used for steering traffic for encryption, the appending of the VPN to an existing enterprise-routed domain is fairly straightforward. However, if the VPN is overlaying an existing enterprise private WAN where endpoint reachability is already established, integration can be challenged by overlapping routed domains (VPN and existing WAN) and the need for the redirecting of endpoint traffic through the GRE tunnel for encryption.

Using IPSec-protected GRE provides the following additional benefits:

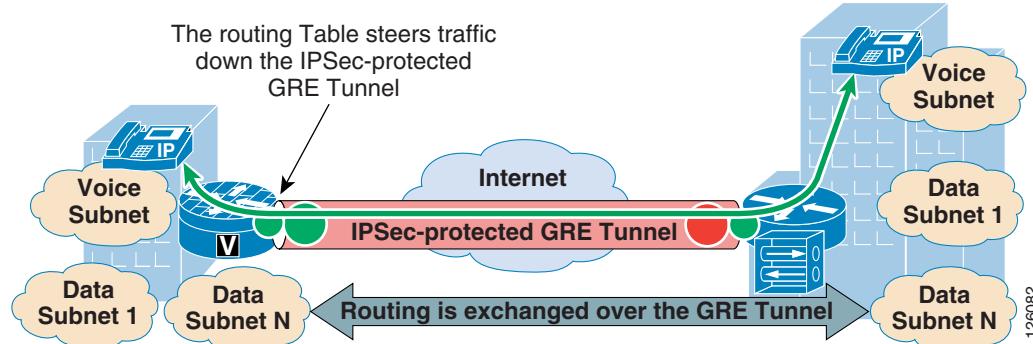
- Only a single SA is required in the head end for each GRE tunnel.
- Branch and headquarters reachability is automatically established using a routing protocol.
- Failover can be easily accommodated by tuning routing metrics.

The following sections describe the two configuration options when using IPSec-protected GRE: point-to-point and multipoint configuration.

Static Point-to-Point GRE

Point-to-point (PTP) GRE tunnels are configured between two peering routers where the static IP addresses for the endpoints of the tunnel (tunnel source and tunnel destination) are typically known and are routable over the WAN. [Figure 2-3](#) shows an IPSec-protected PTP GRE tunnel.

Figure 2-3 IPSec Encryption through GRE Tunneling



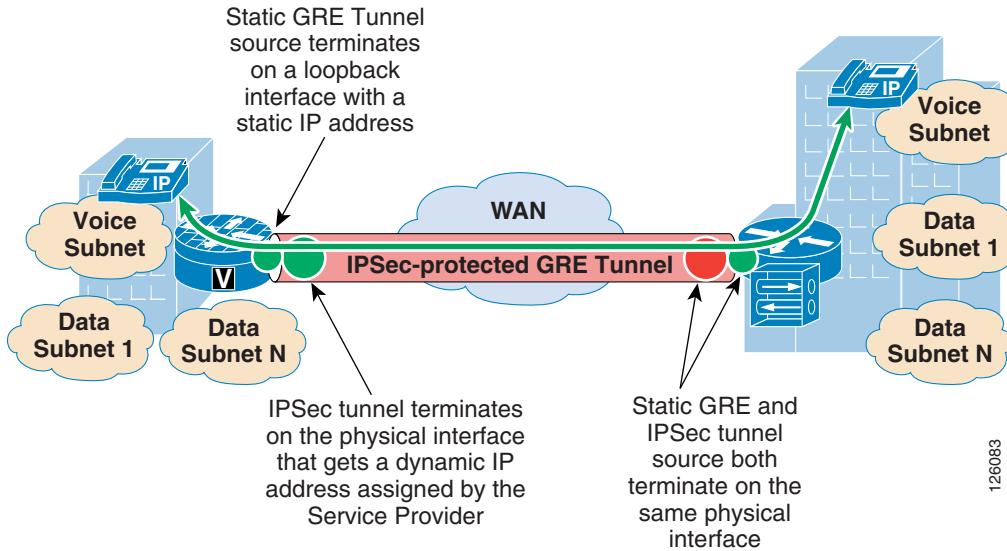
Dynamic Point-to-Point GRE

A misconception is that PTP GRE cannot be used with dynamically-addressed endpoints because of the static nature of the source and destination configuration of the GRE tunnel interface. This section describes how to use static PTP GRE to support dynamically IP-addressed endpoints by using dynamic PTP GRE.

Dynamic PTP GRE is a workaround for the typical static configurations of tunnel source and destination in static PTP GRE. This configuration uses the static address of a loopback interface on the remote endpoint as the GRE tunnel source destination rather than the dynamically-addressed physical interface that would receive an IP address using DHCP. The IPSec source IP address of the remote site is still tied to the physical interface and is dynamically assigned a publicly-routable IP address by the service provider. Traffic from the remote site routing protocol initiates the IPSec tunnel to the hub site.

After this IPSec tunnel is established, the GRE tunnel then comes up and traffic can flow between the remote and the central site. Routing information is exchanged between the remote and central site and reachability is established between the endpoints.

[Figure 2-4](#) shows how dynamic PTP GRE works.

Figure 2-4 Dynamic PTP GRE

Dynamic Multipoint GRE

Dynamic multipoint GRE (mGRE), an integral part of the Dynamic Multipoint VPN (DMVPN) architecture, provides a streamlined GRE configuration option where a single GRE interface can support the dynamic creation of GRE tunnels by peering routers. This is similar in concept to using dynamic crypto maps in the direct IPSec encapsulation option previously mentioned. To protect the mGRE tunnels with IPSec, **tunnel protection interface** commands are used to encrypt the GRE tunnels similar to using the ACL mentioned above. Again, for data traffic to be encrypted, the route table ultimately determines which traffic is sent down the IPSec-protected GRE tunnel.

One other required component when using mGRE is Next Hop Routing Protocol (NHRP). NHRP is used for endpoint IP address resolution for dynamically creating IPSec VPN tunnels between endpoints. This dynamic tunnel capability supports the creation of permanent hub-to-spoke tunnels and optionally on-demand temporary spoke-to-spoke tunnels. This spoke-to-spoke on-demand tunnel creation provides direct IPSec VPN connectivity between branch office routers and does not require the extra encrypt/decrypt cycle required when detouring data traffic through the hub site.

This routing of VPN traffic directly between sites may appear to be the optimal configuration when deploying VPN between branch offices, but careful consideration needs to be made when doing so. For example, the number of tunnels allowed to specific endpoints must be controlled to not exceed device capabilities or available WAN bandwidth. This is similar to using Call Admission Control with IP telephony. However, similar to the dynamic crypto map option, using DMVPN in hub-and-spoke topologies can save considerable VPN head end configuration.

For more detailed information on the known limitations of DMVPN, see the DMVPN FAQ at the following URL:

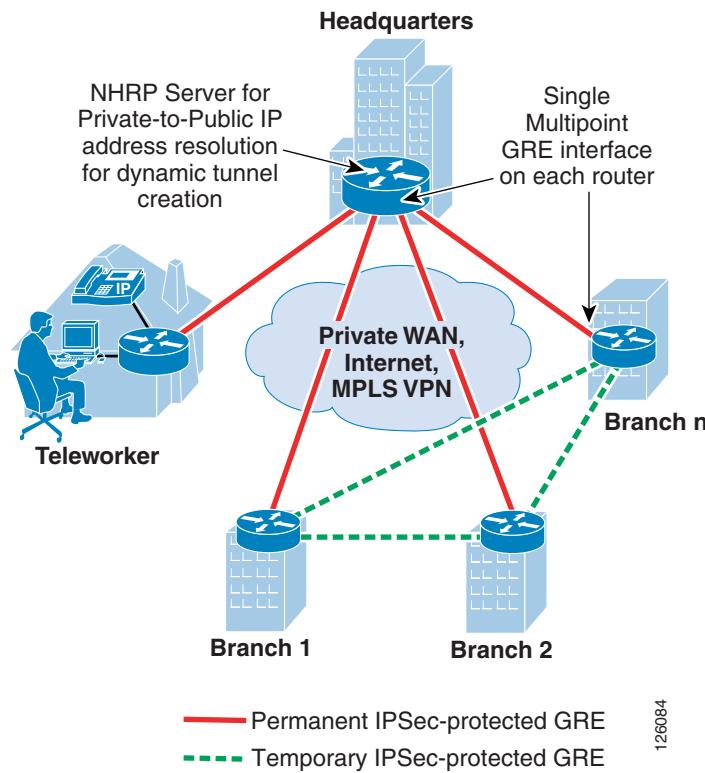
http://www-search.cisco.com/en/US/products/sw/iosswrel/ps1839/products_feature_guide09186a0080110ba1.html.



Note

Spoke-to-spoke dynamic tunnels have not been thoroughly tested and are not recommended at this time.

Figure 2-5 shows a high level view of mGRE in the DMVPN architecture.

Figure 2-5 Using mGRE in the DMVPN Architecture

WAN Security Summary

Customer VPN requirements ultimately drive the method of securing the WAN, whether using direct IPSec encapsulation or IPSec-protected GRE.

The following are considerations when choosing the VPN technology as a WAN for a branch office network:

- Are there other protocols that must be supported besides IP?
GRE can be used to encapsulate other protocols and transport them over IP.
- Is there a requirement to support IP multicast?
If so, then GRE is required, because IPSec does not support IP multicast natively.
- How many networks are behind the branch office router?

If there is a single network or a set of networks that can be summarized into a single routing entry, then direct IPSec encapsulation using dynamic crypto maps may be the preferred solution because of its simplicity. If not, then an IPSec-protected GRE-based VPN may be better, because a routing protocol can be used to advertise the multiple networks behind the branch office router.

- Are there future requirements for spoke-to-spoke dynamic VPN tunnels?
If so, then DMVPN may be an option.

Table 2-1 summarizes the attributes of the different VPN options to provide the network designer help into choosing the best VPN option for the application.

Table 2-1 IPSec VPN Options for the Branch Office

	Protocol Support	Restrictions	Application	VPN Head End Scalability
IPSec static crypto map	IP unicast	No IPmc Music on Hold	Small office	Greater than 1000
IPSec dynamic crypto	IP unicast	No IPmc Music on Hold	Small office	Greater than 1000
IPSec-protected GRE using dynamic crypto	IP unicast, IP multicast, multiprotocol	None	Large office	500–1000
DMVPN	IP unicast, IP multicast	VoIP, hub-and-spoke only	Large office	Less than 500

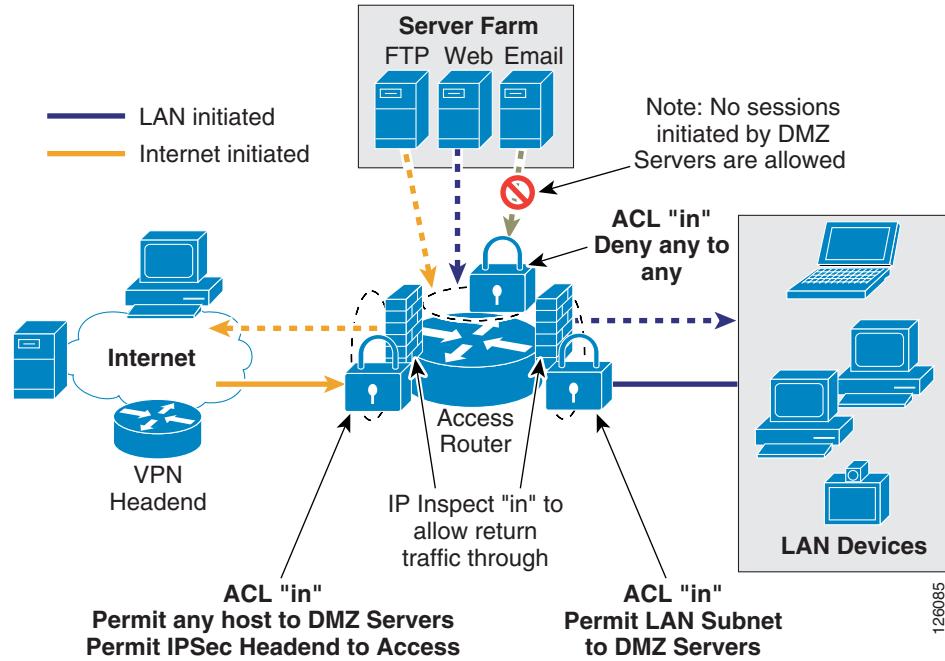
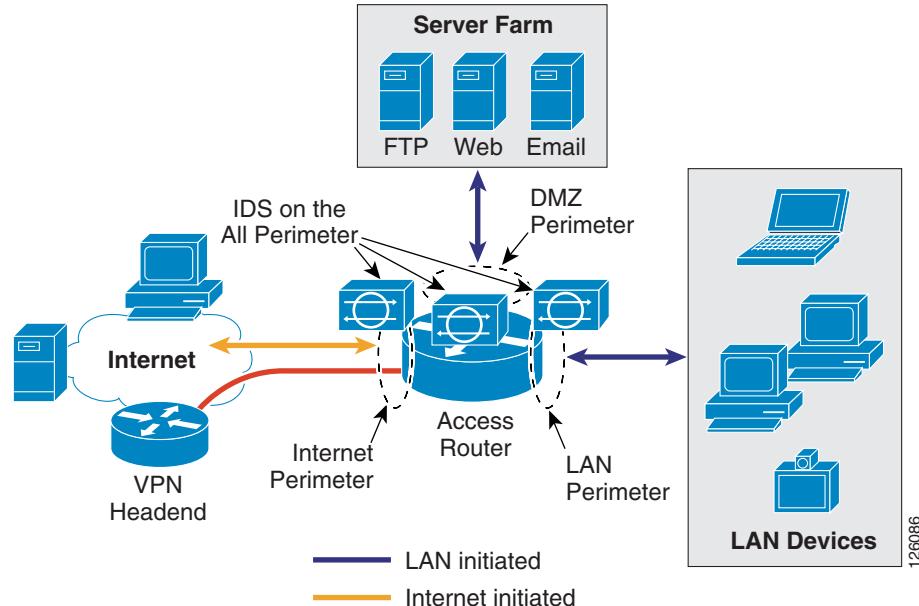
Defending the Perimeter

ACLs, the Cisco IOS Firewall, and the Cisco Intrusion Detection System (IDS) work together to secure the perimeter of the office network. Testing was performed to determine where to deploy these specific features in the office architecture to optimize router performance while providing uncompromised office perimeter security.

Figure 2-6 shows the entry points into the office network where ACLs, Cisco IOS Firewall and/or IDS services could be configured, and the associated traffic flow in and out of these entry points. These entry points may be physical interfaces such as an Ethernet, or logical interfaces such as Frame Relay permanent virtual circuits (PVCs) or Fast Ethernet subinterfaces.

Not all of these entry points and their associated perimeter security may be required in all types of offices. For example, a branch office may not have direct Internet access or a DMZ to secure, so therefore ACLs or a firewall are not required.

Figure 2-6 shows where ACLs and **ip inspect** commands required for Cisco IOS Firewall are configured on the access router.

Figure 2-6 Office Network Perimeter Security with ACLs and Firewall**Figure 2-7 Office Network Perimeter Security with IDS on all Perimeters**

The recommended approach for securing the office network is as follows:

- Use ACLs to tightly restrict traffic inbound to the office network at every entry point.
- Configure firewall inspection (see [Figure 2-6](#)) and IDS protection anywhere external sources of traffic are introduced into the office network (see [Figure 2-7](#)).

- The DMZ is considered a traffic sink, and hosts on the DMZ do not initiate any sessions. Only inbound sessions are allowed to initiate sessions back to the source, whether it is a LAN- or Internet-based host.

Security policies are configured as follows:

- Packets flow from the perimeter (source) to the perimeter (destination).
- ACL policies are applied at the perimeter (source) to filter inbound traffic.
- Firewall inspection and IDS packet monitoring are applied at the source of inbound packet flows.

[Table 2-2](#) summarizes where these three perimeter security features are implemented and how the policies are applied.



Note

Access to the office router (for example, Telnet or SSH) within the office for network management or select ICMP packets may also be permitted, but are not shown for simplicity. These would be additional permit ACLs configured on the Internet perimeter.

Table 2-2 Office Perimeter Security Policies

Packet Flow (From/To)		Security Policies		
Perimeter (source)	Perimeter (destination)	ACL Policy	Firewall Inspection	IDS
Internet	DMZ	Permit any to web, mail, FTP servers	Internet in	Internet in
Internet	VPN	Permit IKE, ESP, GRE from VPN head end Permit HQ subnets to LAN subnet ¹	Internet in	Internet in
Internet	LAN	Deny any	Internet in	Internet in
DMZ	Internet	Deny any	None	DMZ in
DMZ	LAN	Deny any	None	DMZ in
DMZ	VPN	Deny any	None	DMZ in
LAN	Internet	Permit LAN subnet to any	LAN in	LAN in
LAN	DMZ	Permit LAN subnet to any	LAN in	LAN in
LAN	VPN ²	Permit LAN subnet to any	LAN in	LAN in

1. Release 12.3.8T does not require this ACL anymore.

2. If a packet is encapsulated by the router into a GRE tunnel, and the incoming interface is enabled for IDS monitoring, then the packet (before encapsulation) is sent to the IDS Network Module.

IP Communications

This section includes the following topics:

- [Quality of Service Overview](#)
- [IP Telephony](#)

Quality of Service Overview

This section provides a brief overview of quality of service (QoS) factors that provide high quality voice and video transport across the WAN.

This section includes the following topics:

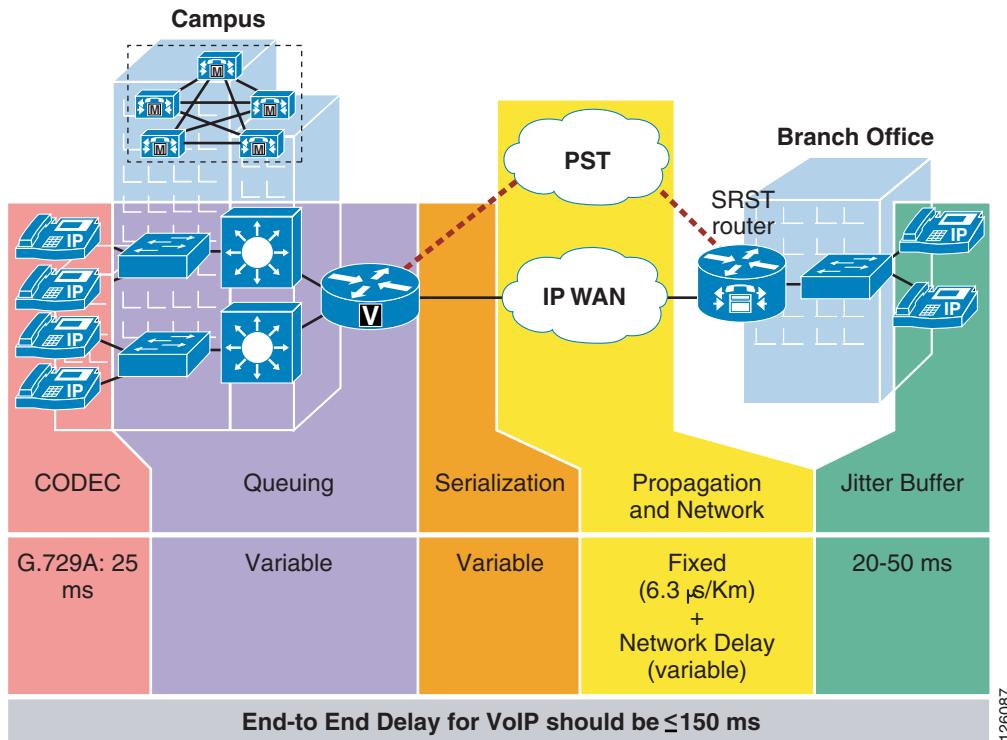
- [Delay](#)
- [Delay Variation \(Jitter\)](#)
- [Packet Loss](#)
- [Provisioning the WAN](#)
- [Service Provider QoS](#)
- [Call Admission Control](#)

Delay

Delay is defined as the finite amount of time it takes a packet to reach the receiving endpoint after being transmitted from the sending endpoint. Limiting delay is critical for high-quality voice or video. In the case of voice, this delay is defined as the amount of time it takes for sound to leave the mouth of the speaker and be heard in the ear of the listener.

If the end-to-end voice delay starts to exceed 200 milliseconds (ms), the conversation begins to sound like two parties talking over a satellite link or a CB radio. The International Telecommunication Union (ITU) standard for Voice over IP (VoIP) G.114 states that a one-way, mouth-to-ear delay budget of 150 ms is acceptable for high voice quality. It has been shown that there is a negligible difference in voice quality scores using networks built with 200 ms delay budgets.

Cisco recommends designing to the ITU standard of 150 ms, but if constraints exist where this delay target cannot be met, then the delay boundary can be extended to 200 ms without significant impact on voice quality. [Figure 2-8](#) shows the various points where delay is incurred when traversing the enterprise network.

Figure 2-8 End-to-End Delay

Delay Variation (Jitter)

Delay variation (jitter) is the difference in the end-to-end delay between packets. Jitter is a factor that must be limited to support high quality voice or video. For example, if one packet requires 100 ms to traverse the network from the source endpoint to the destination endpoint and the following packet requires 125 ms to make the same trip, then the delay variation is calculated as 25 ms. The recommended delay variation or jitter to maintain high-quality voice is less than 30 ms.

Packet Loss

Packet loss is a comparative measure of packets successfully transmitted and received to the total number of packets that were transmitted. Very little packet loss can be tolerated before affecting the quality of voice. Packet loss is expressed as the percentage of packets that were dropped and is typically a function of availability. If the network is highly available, then packet loss during periods of non-congestion is essentially zero. During periods of congestion, however, QoS mechanisms determine which packets are suitable to drop.

Packet loss causes voice clipping and skips. The industry standard codec algorithms used in the Cisco Digital Signal Processor (DSP) can correct for up to 30 ms of lost voice with the use of concealment algorithms. Therefore, the loss of two or more consecutive 20 ms voice samples results in noticeable degradation of voice quality.

Assuming a random distribution of drops within a single voice flow, a drop rate of just 1 percent in a voice stream results in a loss that can not be concealed every three minutes, on average; a 0.25 percent drop rate results in a loss that cannot be concealed once every 53 minutes, on average. The amount of packet loss that can be tolerated is customer-dependent; however, on average, packet loss should be kept to less than 1 percent for high quality voice.

For more detailed information on QoS, see the document at the following URL:
http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND/QoS-SRND-Book.html.

Table 2-3 summarizes the QoS factors that support high quality voice.

Table 2-3 QoS Recommendations

QoS Factor	Less than or Equal to
Delay (latency)	150 ms
Delay variation (jitter)	30 ms
Packet loss	1%

Provisioning the WAN

This section describes the basic requirements for provisioning a WAN link to support real-time services such as IP voice and video.

Critical applications such as VoIP require service guarantees regardless of network conditions. *The only way to provide service guarantees is to enable queuing at any node that has the potential for congestion*, regardless of how rarely this may occur. This principle applies not only to campus-to-WAN/VPN edges, where speed mismatches are most pronounced, but also to campus access-to-distribution or distribution-to-core links, where oversubscription ratios create the potential for congestion. There is simply no other way to guarantee service levels than by enabling queuing wherever a speed mismatch exists.

When provisioning queuing, some best practice rules of thumb also apply. For example, because the best-effort class is the default class for all data traffic, only if an application has been selected for preferential/deferential treatment is it removed from the default class. Because many enterprises have several hundred, if not thousands, of data applications running over their networks, you must provision adequate bandwidth for this class as a whole to handle the sheer volume of applications that default to it. Therefore, it is recommended that you *reserve at least 25 percent of link bandwidth for the default best-effort class*.

The real-time or strict priority class, which is typically used to transport voice or interactive video (and which corresponds to RFC 3246, “An Expedited Forwarding Per-Hop Behavior”), also requires special bandwidth allocation considerations. The amount of bandwidth assigned to the real-time queuing class is variable. However, if you assign too much traffic for strict priority queuing, then the overall effect is a dampening of QoS functionality for non-real-time applications. Remember that the goal of convergence is to enable voice, video, and data to *transparently* co-exist on a single network. When real-time applications such as voice or interactive video dominate a link (especially a WAN/VPN link), then data applications fluctuate significantly in their response times, destroying the transparency of the converged network.

Cisco testing has shown a significant decrease in data application response times when real-time traffic exceeds one-third of link bandwidth capacity. Extensive testing and customer deployments have shown that a general best queuing practice is to *limit the amount of strict priority queuing to 33 percent of link capacity*. This strict priority queuing rule is a conservative and safe design ratio for merging real-time applications with data applications.

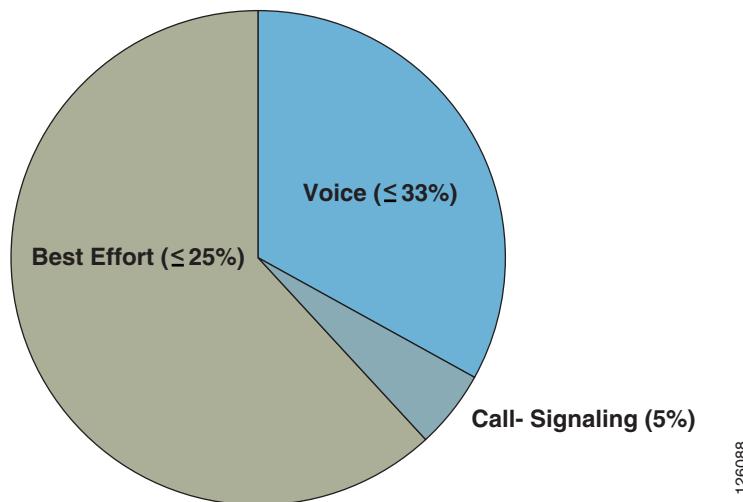
Cisco IOS Software allows the abstraction (and thus configuration) of multiple strict priority low latency queueings (LLQs). In such a multiple LLQ context, this design principle *applies the sum of all LLQs to be within one-third of link capacity*.



Note This strict priority queuing rule (limit to 33 percent) is simply a best practice design recommendation and is not a mandate. There may be cases where specific business objectives cannot be met while holding to this recommendation. In such cases, enterprises must provision according to their detailed requirements and constraints. However, it is important to recognize the tradeoffs involved with over-provisioning strict priority traffic and its negative performance impact on non-real-time-application response times.

A very basic WAN edge bandwidth allocation model is shown in [Figure 2-9](#), in which VoIP is restricted to 33 percent of the link, a moderate but guaranteed allocation is made to protect call-signaling traffic, and best-effort traffic is granted at least 25 percent of the link. Additional classes of traffic may be added, as needs evolve, but these should be provisioned within these best-practice guidelines.

Figure 2-9 Provisioning the WAN Link for Voice



Service Provider QoS

In IP-based service provider networks, the implementation of QoS is required to guarantee high quality voice. This is most prevalent where IPSec VPNs are deployed as the enterprise WAN.

Some of the considerations when deploying over an IP-based service provider WAN are as follows:

- The service provider edge QoS configuration must be consistent and complementary to that of the enterprise office router QoS configuration: class-based weighted fair queuing (CBWFQ), LLQ, Traffic Shaping, and Link Fragmentation and Interleaving (LFI).
- Packets must be delivered end-to-end with minimal delay, jitter, and loss. This can be accomplished by prioritizing packets based on type of service (ToS) IP Precedence/ Differentiated Services Code Point (DSCP) in the core (QoS-enabled core) or over-provisioning the bandwidth in the core.
- The enterprise should be aware of service provider policies for handling high priority traffic that exceeds the rate that was agreed upon with the customer and for traffic that crosses the service provider boundary.
- Negotiating service-level agreements (SLAs) to fulfill end-to-end delay, jitter, and packet loss requirements. Service providers are now offering such SLAs. IP VPN multi-service Cisco Powered Network (CPN) service providers meet the following maximum one-way service levels:

- Latency = 60 ms
- Jitter = 20 ms
- Loss = 0.5%

To find CPN service providers, see the following URL:
http://www.cisco.com/pcgi-bin/cpn/cpn_pub_bassrch.pl

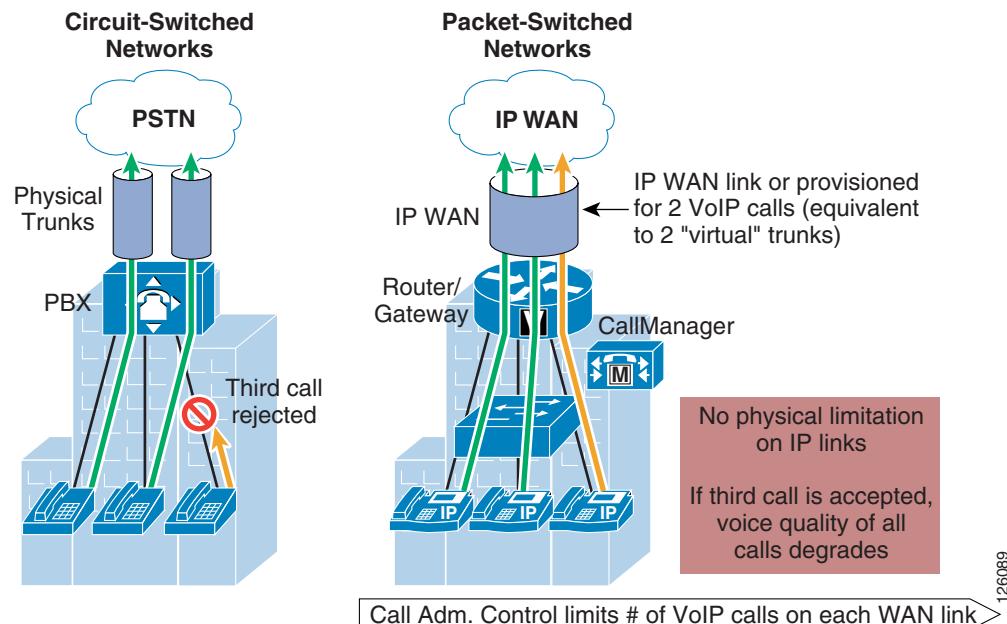
Call Admission Control

Call Admission Control (CAC) maintains high voice quality over an IP WAN by limiting the number of calls that are admitted. This was typically not required with traditional voice networks that use Time Division Multiplexed (TDM) circuits where physical channels limited the number of calls allowed to connect to the PSTN.

This is not the case when VoIP calls traverse an IP WAN where calls are packet streams and there are no physical limitations that control the number of calls admitted to the WAN link. Therefore, an IP WAN link can easily be oversubscribed and the voice quality of all connected calls can be degraded. Because of this problem, some mechanism must be implemented to control the number of admitted calls to maintain high voice quality.

Figure 2-10 compares CAC between traditional TDM-based voice networks and VoIP-based networks.

Figure 2-10 Call Admission Control



IP Telephony

This section includes the following topics:

- IP Telephony for the Office
- Provisioning for Voice

- Centralized Call Processing with CallManager
- Local Call Processing with CallManager Express

IP Telephony for the Office

There are many aspects to designing an IP telephony network, but this design guide covers only those aspects that are specific to an office deployment. The critical aspects that need to be considered are as follows:

- Provisioning QoS on the WAN link to support voice
- Configuring CAC
- Conferencing and transcoding to support the office

Some of the considerations when deploying IP telephony for the office are as follows:

- The call processing deployment model used: Single Site, Centralized Call Processing, or Distributed Call Processing.

This design guide focuses on the Centralized Call Processing deployment model using Cisco CallManager. Single Site and Distributed Call Processing deployment models include Cisco CallManager Express. For more information on Single Site and Distributed Call Processing deployment models using Cisco CallManager, see the following URL:
http://www.cisco.com/en/US/docs/voice_ip_comm/cucme/srnd/design/guide/models.html.

- The number of off-premise voice calls that are made during the busy hour.

Basic oversubscription ratios or Erlang ratios can be used to calculate the number of voice lines required both for the local Public Switched Telephone Network (PSTN) calling and toll-bypass calling over the WAN.

- The voice codec used when calling over the WAN.

Typically, G.729 is used to minimize the amount of bandwidth required, but another consideration is conference calling. The HQ Cisco CallManager may be used for conferencing but only for G.711 calls. Therefore, if G.729 is going to be used for calls over the WAN, then hardware-based digital signal processing (DSP) farms must be used for conferencing and the transcoding between endpoints.

- The number of calls that can be supported over the WAN because of the bandwidth overhead associated with the WAN technology used (for example, Frame Relay, Multi-Link Point-to-Point Protocol (MLPPP), or IPSec VPN).

For example, when sending voice over IPSec VPN, the addition of an IP GRE header and IPSec/Encapsulating Security Payload (ESP) header increases the size of the original voice (or video) IP 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. This does not include the Layer 2 WAN overhead, which adds an additional 4 bytes for PPP, High-level Data Link Control (HDLC), and Frame Relay, or 14 bytes for Ethernet. See the following URL for more information on VoIP over IPSec VPN:

http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/V3PN_SRND/V3PN_SRND.html.

- Do conference call traffic patterns merit the deployment of local media resources (that is, hardware DSP farm)? In other words, are most conference calls initiated by and held between local branch office IP and PSTN phones?

If so, then distributing media resources locally within the branch office conserves WAN bandwidth. In contrast, if the majority of conference call parties are outside the branch office located over the WAN, then locally deploying media resources in the branch probably does not justify the additional costs.

Provisioning for Voice

This section includes the following topics:

- [Voice Lines Requirements for Off-Premise Calling](#)
- [Bandwidth Requirements for Toll-Bypass](#)

Voice Lines Requirements for Off-Premise Calling

There are two types of calls that must be considered when provisioning the branch office for voice: PSTN calls and toll-bypass calls. One of the primary reasons for deploying a branch or an autonomous office in a particular location is to serve that location with whatever product or service the company may offer. This means that some local calls are made to the PSTN, and depending on the business, some calls may traverse the WAN to a connected office. This is commonly referred to as toll-bypass, where the enterprise uses their IP WAN instead of the PSTN to carry the calls to non-local offices to avoid the associated long distance toll charges. Knowing the number of local calls and toll-bypass calls helps to determine the number of voice lines and WAN bandwidth needed for that particular office.

Traditionally, basic oversubscription ratios or Erlang traffic models are used to determine the number of voice lines required for PSTN and toll-bypass calling in a specific size office. Basic oversubscription ratios are typically based on call records collected from other existing offices and applied to the new office. They equate the number of users to the number of PSTN lines required for calling. Typical oversubscription ratios currently used for business offices are four or five users to one voice line. For example, a 20-user office requires four voice lines for PSTN calling if a five to one oversubscription ratio is used.

Using an Erlang traffic model is a more scientific method of determining the number of voice lines required for an office. A few variants of Erlang models can be used, depending on the type of office such as a typical office or a call center. Following is an example of using Extended Erlang B to determine the number of voice lines and WAN bandwidth required for a 23-user office.

An *Erlang* is a unit of measurement of voice traffic. Strictly speaking, an Erlang represents the continuous use of one voice path or line. In practice, it is used to describe the total traffic volume of one hour.

The Extended Erlang B traffic model is used to estimate the number of lines required for PSTN lines, and takes into account the additional traffic load caused by blocked callers that immediately try to call again if their calls are blocked. The four variables involved are recall factor, busy hour traffic (BHT), blocking, and lines:

- *Recall factor* is the percentage of calls that immediately retry if their calls are blocked.
- *Busy hour traffic (in Erlangs)* is the number of hours of call traffic during the busiest hour of operation of a telephone system.
- *Blocking* is the failure of calls because of an insufficient number of lines being available. For example, 0.03 means three calls blocked per 100 calls attempted.
- *Lines* is the number of lines to the PSTN.

Erlang traffic measurements can be used to estimate the number of lines that are required between a voice gateway or IP PBX and a PSTN central office, or between other enterprise offices.

For example, if during the busy hour, each user makes six calls at two minutes per call. This represents .2 Erlangs of BHT per user. The breakdown of this figure is as follows:

- 6 calls @ 2 minutes per call = 12 minutes call traffic per user
- 12 minutes of traffic per user/60 minutes per hour = .2 Erlang per user of BHT



Note .2 Erlang of BHT per user is used to determine the number of voice lines used for Business Ready Branch testing.

For an office of 23 users, a total of 4.6 Erlangs of BHT must be accommodated with an appropriate number of voice lines. Using an Erlang calculator such as the one found at www.erlang.com, an estimate of voice lines may be calculated for the office (see <http://www.erlang.com/calculator/exeb>). For this 23-user office example, nine voice lines are required to accommodate this amount of BHT, assuming only 3 percent of calls would experience blocking.

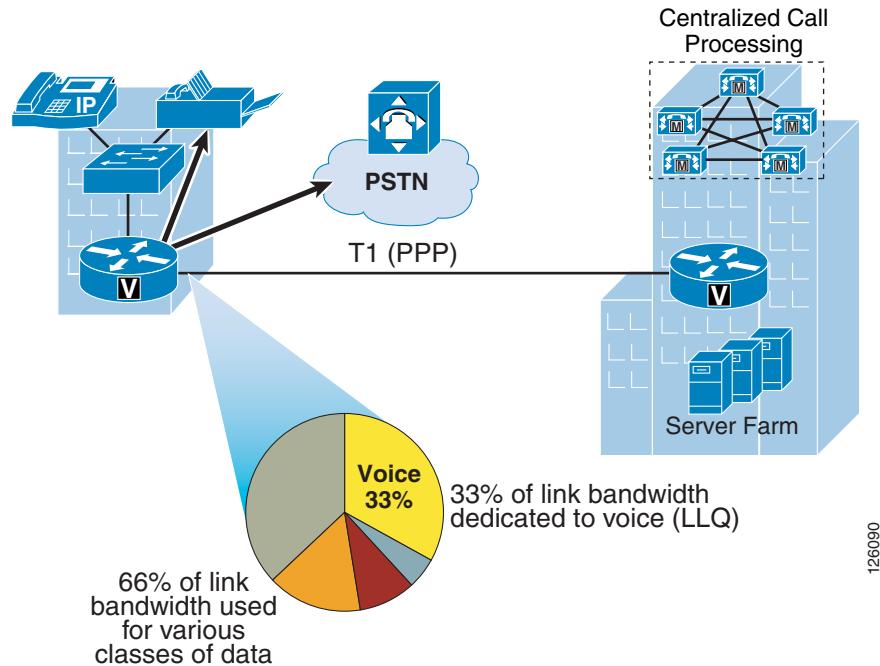
This information is vital for sizing the routing platform and the amount of WAN bandwidth to be provisioned to support this calculated number of voice calls. Referring to the best practices shown in the last section where 33 percent of the WAN link can be used for real-time traffic, using this voice line calculation can be used to determine the WAN link bandwidth required by this office. The next section discusses this further.

Bandwidth Requirements for Toll-Bypass

The amount of WAN bandwidth to carry all types of traffic may be determined either by the available monetary budget allocated for the WAN circuit or optimally by the services that it will carry. This section describes how to determine the amount of WAN bandwidth that is required by voice and how it is provisioned on the WAN link of the branch office.

The previous section described how to determine the number of voice lines required in the branch based on common oversubscription ratios or Erlang model calculations. Using the 23-user office example where each user creates .2 Erlangs of call traffic during the busy hour, we can assume some number of those calls traverses the WAN to other offices, and therefore some percentage of these lines can be dedicated to toll-bypass calling.

For example, three of the nine lines calculated can be considered for provisioning the WAN link LLQs that are dedicated to real-time traffic. This equates to three Real-Time Protocol (RTP) voice streams that must be accommodated by the WAN connection and configured in CallManager for CAC. In a voice and data only deployment where voice traffic is the only real-time traffic type, 33 percent of the WAN link can be dedicated to voice and the remaining 66 percent for data traffic, as shown in [Figure 2-11](#).

Figure 2-11 Provisioning for Voice

Knowing the amount of bandwidth required for voice helps determine the minimum amount of WAN bandwidth required for both voice and data for the branch office.

After the number of voice lines or, in the case of VoIP, RTP streams are determined for toll-bypass, the minimum amount of WAN bandwidth can be determined. The amount of bandwidth that is required for each RTP stream is dependent on the WAN type and its associated encapsulation method (Frame Relay, IPSec, PPP, and so on) as well as the voice sampling rate. [Table 2-4](#) shows sample bandwidth requirements for RTP streams with varying WAN types.

Table 2-4 RTP Stream Bandwidth

CODEC	Sampling rate	Voice payload	Packets per second	Bandwidth per conversion
G.711	20	16	50	80 kbps
G.711	30	24	33	74 kbps
G.729A	20	20	50	24 kbps
G.729A	30	30	33	18 kbps

[Table 2-5](#) shows that a more accurate method for provisioning is to include the Layer 2 headers into the bandwidth calculations.

Table 2-5 Layer 2 Headers Included in Bandwidth Calculations

CODEC	Ethernet 4 bytes of header	PPP 6 bytes of header	ATM 53 bytes cells with 48-byte payload	Frame Relay 4 bytes of header
G.711 at	85.6	82.4	106	81.6
G.711at	77.6	75.5	84	75 kbps
G.729A at	29.6	26.4	42.4	25.6
G.729A at	22.2	20	28	19.5

Assuming G.729A is used over the WAN to conserve bandwidth, 3 RTP streams over a PPP WAN link consume ($3 \times 26.4 \text{ Kbps} = 79.2 \text{ Kbps}$) or approximately 80 Kbps of bandwidth for voice. This requires a minimum total WAN link bandwidth of 240 Kbps, assuming 33 percent of the link is used for voice RTP streams. This method uses the services that the link carries as the way to determine the required WAN bandwidth. Of course, if the WAN bandwidth is known, then this calculation determines the number of calls the WAN link can support.

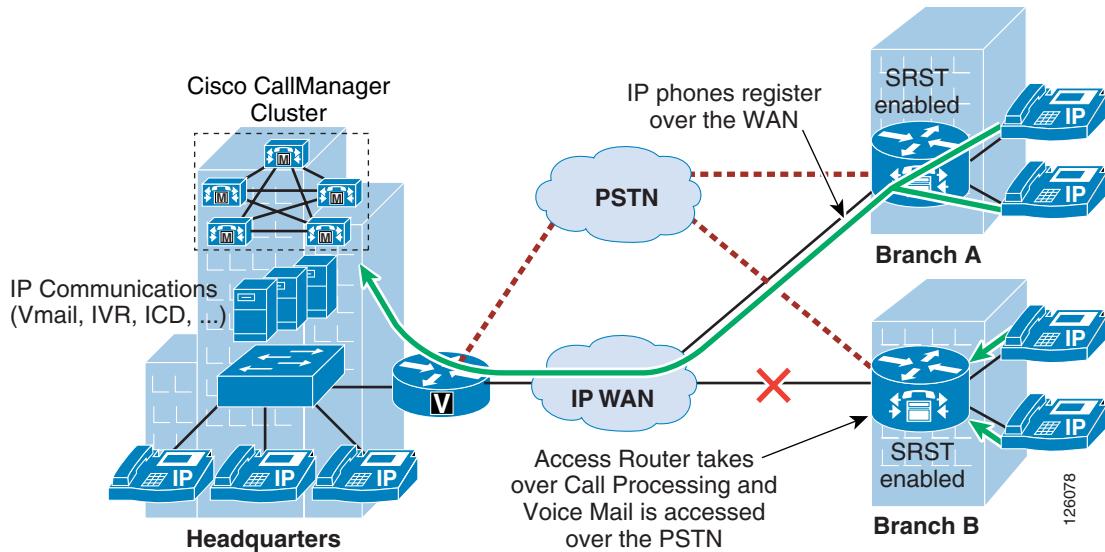
Centralized Call Processing with CallManager

This section includes the following topics:

- [Centralized Call Processing with CallManager Overview](#)
- [Survivable Remote Site Telephony and Fallback Behavior](#)
- [Call Admission Control with CallManager](#)
- [Automated Alternate Routing](#)
- [Conferencing and Transcoding](#)

Centralized Call Processing with CallManager Overview

Centralized Call Processing consists of a single call processing agent that provides services for many sites, and uses the IP WAN to transport IP telephony traffic between the sites. The IP WAN also carries call control signaling between the central site and the remote sites. [Figure 2-12](#) shows a typical Centralized Call Processing deployment, with a Cisco CallManager cluster as the call processing agent at the central site and an IP WAN with QoS enabled to connect all the sites.

Figure 2-12 Centralized Call Processing

The remote sites rely on the centralized Cisco CallManager cluster to handle their call processing. Applications such as voice mail and Interactive Voice Response (IVR) systems are typically centralized as well to reduce the overall costs of administration and maintenance.

Connectivity options for the IP WAN include the following:

- Leased lines
- Frame Relay
- Asynchronous Transfer Mode (ATM)
- ATM and Frame Relay Service Inter-Working (SIW)
- Multiprotocol Label Switching (MPLS) Virtual Private Network (VPN)
- Voice and Video Enabled IP Security Protocol (IPSec) VPN (V3PN)

Routers that reside at the WAN perimeters require QoS mechanisms, such as LLQ and traffic shaping, to protect the voice traffic from the data traffic across the WAN, where bandwidth is typically scarce. In addition, a CAC scheme is needed to avoid oversubscribing the WAN links with voice traffic and deteriorating the quality of established calls.

For Centralized Call Processing deployments, the *locations* mechanism within Cisco CallManager provides CAC. A variety of Cisco gateways can provide the remote sites with PSTN access. When the IP WAN is down, or if all the available bandwidth on the IP WAN has been consumed, users at the remote sites can dial the PSTN access code or use Automated Alternate Routing (AAR) to have their calls sent through the PSTN. The SRST feature, available on Cisco IOS gateways, provides call processing at the branch offices in the event of a WAN failure.

Follow these guidelines and best practices when implementing the centralized call processing:

- Minimize delay between Cisco CallManager and remote locations to reduce voice cut-through delays (also known as clipping).
- For hub-and-spoke topologies, use the locations mechanism in Cisco CallManager for CAC into and out of remote branches.

- If the WAN uses Cisco IOS MPLS or IPSec VPN deployment options where bandwidth allocation cannot be associated with a single physical link per connected site, then CAC must be configured to prevent oversubscription of the physical WAN link capacity of the hub link. Again, the CallManager locations feature is used for specifying the HQ call bandwidth capacity.
- If multicast Music on Hold (MoH) is used to branch offices, this bandwidth must be provisioned in the LLQ in addition to normal voice calls. This is because CallManager does not track multicast MoH audio streams.

Survivable Remote Site Telephony and Fallback Behavior

The SRST feature provides Cisco CallManager with fallback support for Cisco IP phones attached to a Cisco router on your local network. SRST enables routers to provide call-handling support for Cisco IP phones when they lose connection to remote primary, secondary, or tertiary Cisco CallManager installations; or when the WAN connection is down.

Cisco CallManager Release 3.2 supports Cisco IP phones at remote sites attached to Cisco multiservice routers across the WAN. Before SRST, when the WAN connection between a router and Cisco CallManager failed, or connectivity with Cisco CallManager was lost for some reason, Cisco IP phones on the network became unusable for the duration of the failure. SRST overcomes this problem and ensures that the Cisco IP phones offer continuous (yet minimal) service by providing call handling support for Cisco IP phones directly from the SRST router.

The system automatically detects a failure and uses Simple Subnetwork Auto Provisioning technology to autoconfigure the branch office router to provide call processing for Cisco IP phones registered with the router. When the WAN link or connection to the primary Cisco CallManager is restored, call handling reverts back to the primary Cisco CallManager.

When Cisco IP phones lose contact with primary, secondary, and tertiary Cisco CallManagers, they must establish a connection to a local SRST router to ensure call-processing capability necessary to place and receive calls. The Cisco IP phone retains the IP address of the local SRST router as a default router in the Network Configuration area of the Settings menu. When a secondary Cisco CallManager is not available on the network, the local SRST router IP address is retained as the standby connection for Cisco CallManager during normal operation.

When the WAN link fails, calls in progress are sustained for the duration of the call. Calls in transition and calls that have not yet connected are dropped and must be reinitiated after Cisco IP phones reestablish connection to their local SRST router. Telephone service remains unavailable from the time connection to the remote Cisco CallManager is lost until the Cisco IP phone establishes connection to the SRST router.



Note

CallManager fallback mode telephone service is available only to those Cisco IP phones that are supported by an SRST router. Other Cisco IP phones on the network remain out of service until they are able to reestablish a connection with their primary, secondary, or tertiary Cisco CallManager.

The time taken to reestablish connection to a remote Cisco CallManager depends in part on the keepalive period set by Cisco CallManager itself. Typically, three times the keepalive period is required for a phone to discover that its connection to Cisco CallManager has failed. The default keepalive period is 30 seconds. If the phone has an active standby connection established with an SRST router, the fallback process takes 10 to 20 seconds after connection with Cisco CallManager is lost. An active standby connection to an SRST router exists only if the phone has the location of a single Cisco CallManager in its CallManager list. Otherwise, the phone activates a standby connection to its secondary Cisco CallManager.

If a Cisco IP phone has multiple Cisco CallManagers in its CallManager list, it progresses through its list of secondary and tertiary Cisco CallManagers before attempting to connect with its local SRST router. Therefore, the time that passes before the Cisco IP phone eventually establishes a connection with the SRST router increases with each attempt to contact a Cisco CallManager. Assuming that each attempt to connect to Cisco CallManager takes approximately one minute, the Cisco IP phone in question can remain offline for three minutes or more following a WAN link failure.

During a WAN connection failure, when SRST is enabled, Cisco IP phones display a message explaining that they are operating in Cisco CallManager fallback mode. The Cisco IP Phone 7960 and Cisco IP Phone 7940 display a “CM Fallback Service Operating” message and the Cisco IP Phone 7910 displays a “CM Fallback Service” message when operating in Cisco CallManager fallback mode. When the Cisco CallManager is restored, the message disappears and full Cisco IP phone functionality is restored.

While in CallManager fallback mode, Cisco IP phones periodically attempt to reestablish a connection with Cisco CallManager at the central office. When a connection is reestablished with Cisco CallManager, Cisco IP phones automatically cancel their registration with the SRST router. A Cisco IP phone cannot reestablish a connection with the primary Cisco CallManager at the central office if it is currently engaged in an active call.

For more details, see the following URLs:

- Survivable Remote Site Telephony Configuration—
http://www.cisco.com/en/US/products/sw/iosswrel/ps1839/products_feature_guide_chapter09186a0080110b97.html
- MGCP Fallback—
http://www.cisco.com/en/US/tech/tk1077/technologies_tech_note09186a0080144630.shtml

Call Admission Control with CallManager

Multi-site deployments require some form of CAC to ensure the voice quality of calls transmitted across network links that have limited available bandwidth. Cisco CallManager provides a simple mechanism known as locations for implementing CAC in multi-site WAN deployments with Centralized Call Processing. Follow these guidelines when using locations for CAC:

- Locations require a hub-and-spoke network topology.
- Configure a separate location in Cisco CallManager for each site.
- Configure the appropriate bandwidth limit for each site according to the type of codec used at that site. (See [Table 2-6](#) for bandwidth settings.)
- Assign each device configured in Cisco CallManager to a location. If you move a device to another location, change its location configuration as well.
- Cisco CallManager supports up to 500 locations.

Before Cisco CallManager Release 3.1, a cluster supported only one primary (active) Cisco CallManager server when using locations for CAC. With Cisco CallManager Release 3.1 and later, the locations bandwidth is shared among all Cisco CallManager subscriber servers in the cluster, thus enabling you to use the locations mechanism with any size cluster.



The bandwidth values shown in [Table 2-6](#) and configurable in CallManager may differ from the actual bandwidth required for an RTP stream that would be configured in the LLQ within the router.

Table 2-6 Bandwidth Settings by Codec Type

Codec Type	G.729	G.711
Codec bit rate	8 kbps	64 kbps
Parameter Setting		
Cisco CallManager locations	24 kbps	80 kbps
Cisco CallManager gatekeeper	16 kbps	128 kbps
Cisco IOS gateways, before Release 12.2(2)XA	64 kbps	64 kbps
Cisco IOS gateways, Release 12.2(2)XA and later	16 kbps	128 kbps

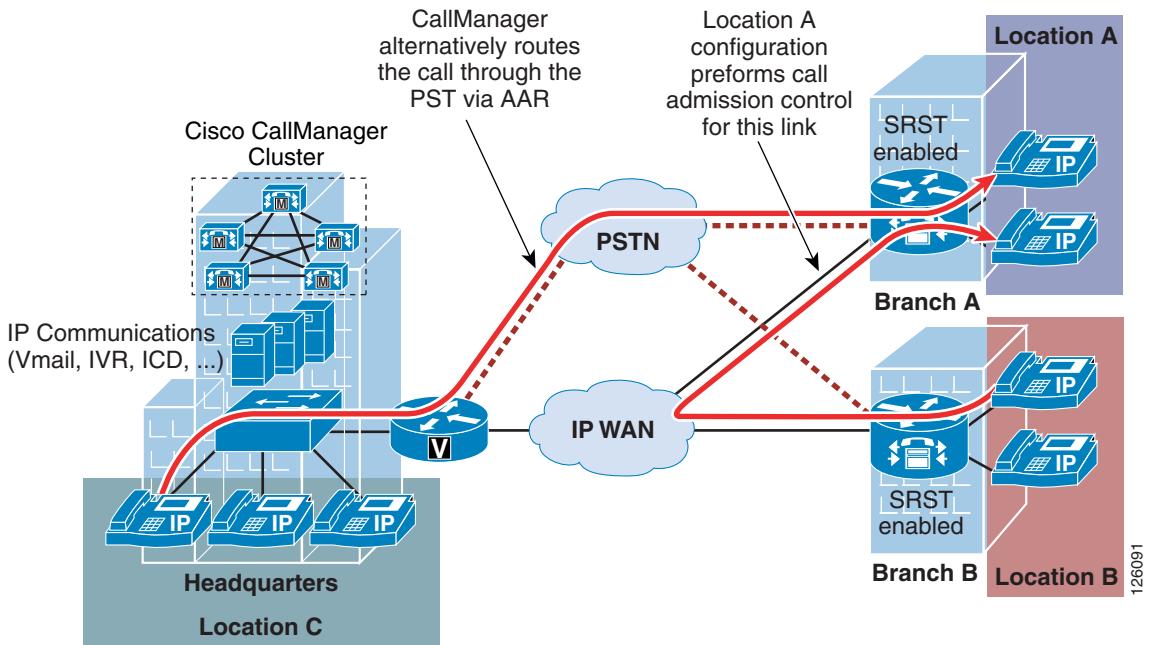
Automated Alternate Routing

The Automated Alternate Routing (AAR) feature enables Cisco CallManager to establish an alternate path for the voice media when the preferred path between two intra-cluster endpoints runs out of available bandwidth, as determined by the locations mechanism for CAC.

The AAR feature applies primarily to Centralized Call Processing deployments. For example, if a phone in the headquarters calls a phone in branch A and the available bandwidth for the WAN link between the branches is insufficient (as computed by the locations mechanism), AAR can reroute the call through the PSTN. The audio path of the call is IP-based from the calling phone to its local PSTN gateway, TDM-based from that gateway through the PSTN to the branch A gateway, and IP-based from the branch A gateway to the destination IP phone.

AAR can be transparent to the users. You can configure AAR so that users dial only the on-net (for example, 4-digit) directory number of the called phone, and no additional user input is required to reach the destination through the alternate network (such as the PSTN).

Figure 2-13 shows an example of AAR configuration.

Figure 2-13 Automated Alternative Routing

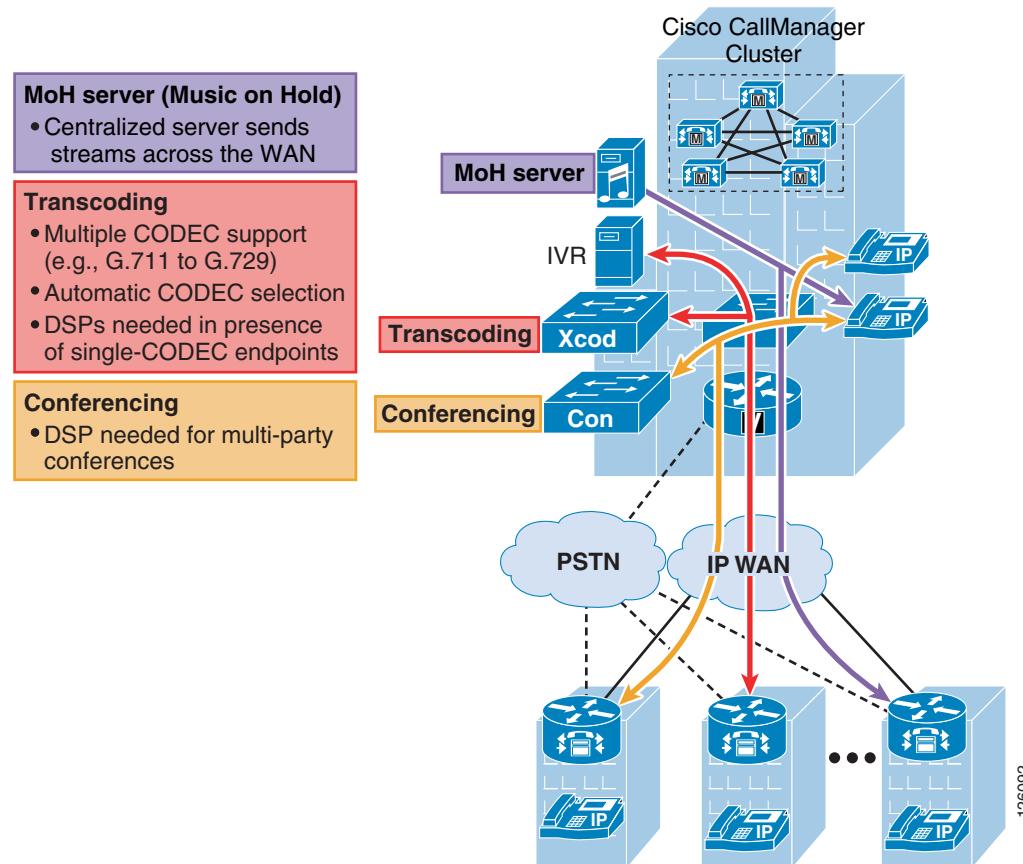
12691

For detailed design and configuration information concerning CAC, see the following URL:
http://www.cisco.com/en/US/docs/voice_ip_comm/cucm/admin/3_0_1/p1admiss.html.

Conferencing and Transcoding

Typically, when call processing is localized at the central site, the MTP, transcoding, and conferencing services may be centralized or distributed, or a combination of both, as shown in [Figure 2-14](#).

Figure 2-14 Centralized Media Resources



If the media resources are centralized, the following results:

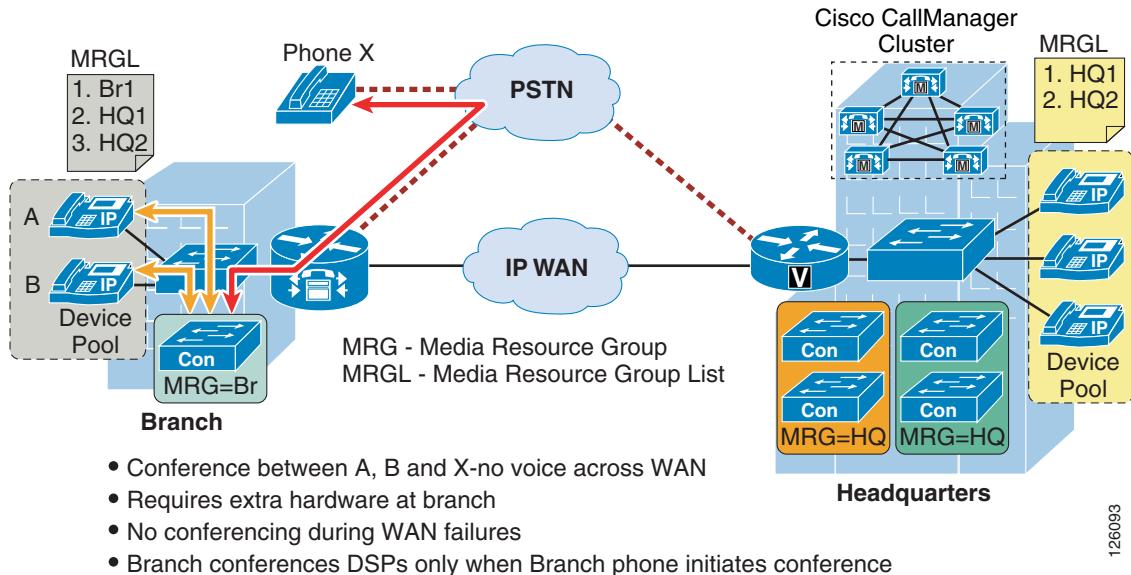
- The WAN is used in every call involving one of these resources.
- Frequently, remote sites use low bit-rate codecs across the WAN; thus, conference calls in a Centralized Call Processing model generally require transcoding resources as well. A hardware conferencing resource is the ideal choice in this scenario because it can eliminate the need for dedicated transcoders.
- Centrally located resources can cause local calls to traverse the WAN, so you must consider the effects on bandwidth consumption.

If the media resources are distributed (see [Figure 2-15](#)), use the following guidelines:

- Group resources into Media Resource Group Lists (MRGL) based on their location to prevent one remote site from using resources located at another remote site. This practice helps you to manage CAC between the sites.

- Use hardware conferencing resources if the cluster contains more than one type of codec (for example, G.711 and G.729).

Figure 2-15 Distributed Conference Resources in the Branch Office



Local Call Processing with CallManager Express

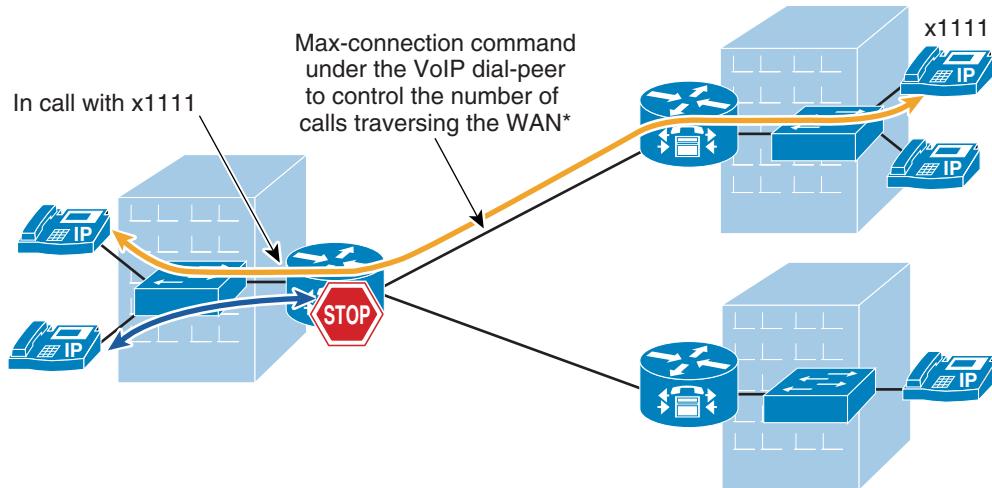
Local Call Processing can be performed by either a local standalone CallManager server in the office or by CallManager Express (CME) integrated in the office router. This section focuses on using the integrated CME for Local Call Processing. CME is primarily used for standalone offices, but can connect to other voice networks such as other CME routers or CallManager using H.323. This section describes some of the considerations that specifically apply to CAC.

This section includes the following topics:

- [Call Admission Control with CME](#)
- [Conferencing and Transcoding](#)

Call Admission Control with CME

A limited CAC capability is supported by CME to ensure that the WAN link does not get oversubscribed and therefore degrade voice quality. The **max-connection** command can be used under the VoIP dial-peer to control the maximum number of call connections that are allowed over that dial-peer (see [Figure 2-16](#)).

Figure 2-16 CAC using max-connection Command

*Maximum-Connection command is only effective across a single dail-peer.
It is not additive across multiple dail-peers. This implies that a single dail-peer be used per physical Wan link.

12694

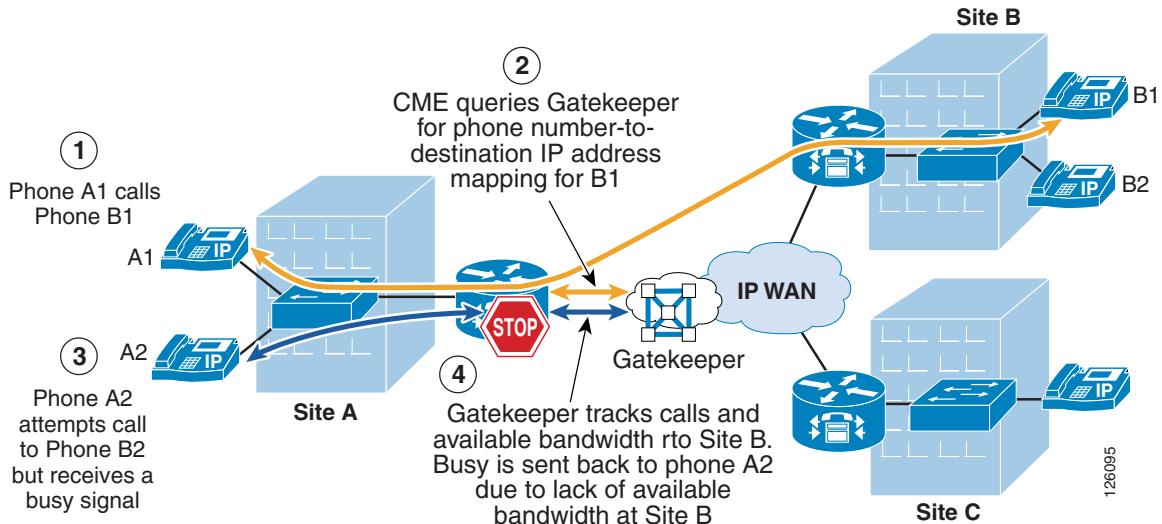
The limitation to this implementation is that the **max-connection** command is associated only with the dial peer under which it is configured. Therefore, call connections are not counted across multiple dial peers. This method of CAC can be used for very limited hub-and-spoke configurations, where a single dial peer per WAN circuit is used to limit the number of calls that a specific circuit can handle.

A gatekeeper may be used for more complex dial plans. The gatekeeper houses the inter-site dial plan and controls the number of calls allowed to each site. This configuration is used for single WAN-connected sites where there are multiple logical circuits connecting to multiple sites.

See the “Voice Routing” section of the following document for more detailed gatekeeper configuration and design documentation:

<http://www-search.cisco.com/en/US/products/sw/voicesw/ps4139/index.html>

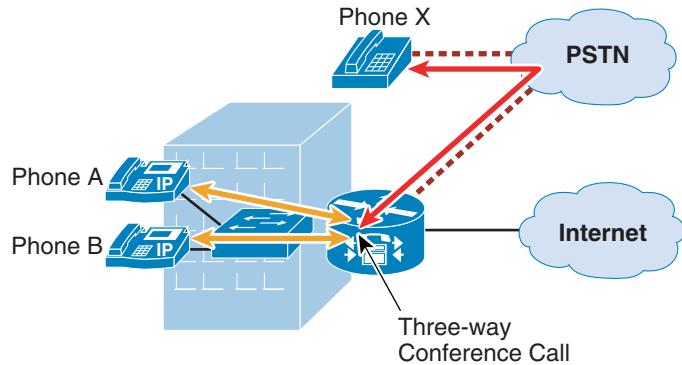
Figure 2-17 shows three sites that are registered with a gatekeeper and how the gatekeeper enforces CAC.

Figure 2-17 Using the Gatekeeper for CAC to Meshed Dial Plans

Sites are registered with the gatekeeper where the inter-site dial plan is configured along with the amount of bandwidth that is available for each site. The gatekeeper provides the phone number-to-IP address mapping of the destination phone and tracks the number of calls made to that particular site. When the maximum number of calls is reached because of the lack of available bandwidth, then the gatekeeper signals this information back to CME and CME presents a busy signal to the calling phone.

Conferencing and Transcoding

Currently, CallManager Express supports only three-way conference calls and no transcoding, as shown in Figure 2-18.

Figure 2-18 Conferencing with CME

- Conference between A, B and X-local to CME
- No extra hardware at branch
- Supports 3 party Conference Calls

126096

For more information on designing with CME and CUE, see the following URL:
<http://www-search.cisco.com/en/US/products/sw/voicesw/ps4625/index.html>



CHAPTER

3

Choosing a Branch Office Platform

This chapter provides the network designer the information needed to successfully choose a routing platform to support a Business Ready office network.

Extensive Cisco testing has resulted in the collection of vital performance information that the network designer can use to determine an appropriate multiservice platform for an office network. Various office profiles that serve a specific number of users, voice lines, and feature combinations were tested and can be used as baselines for customer office networks.

The following exercise guides you through an example branch office design and shows how these performance results can be used to choose an appropriate platform.

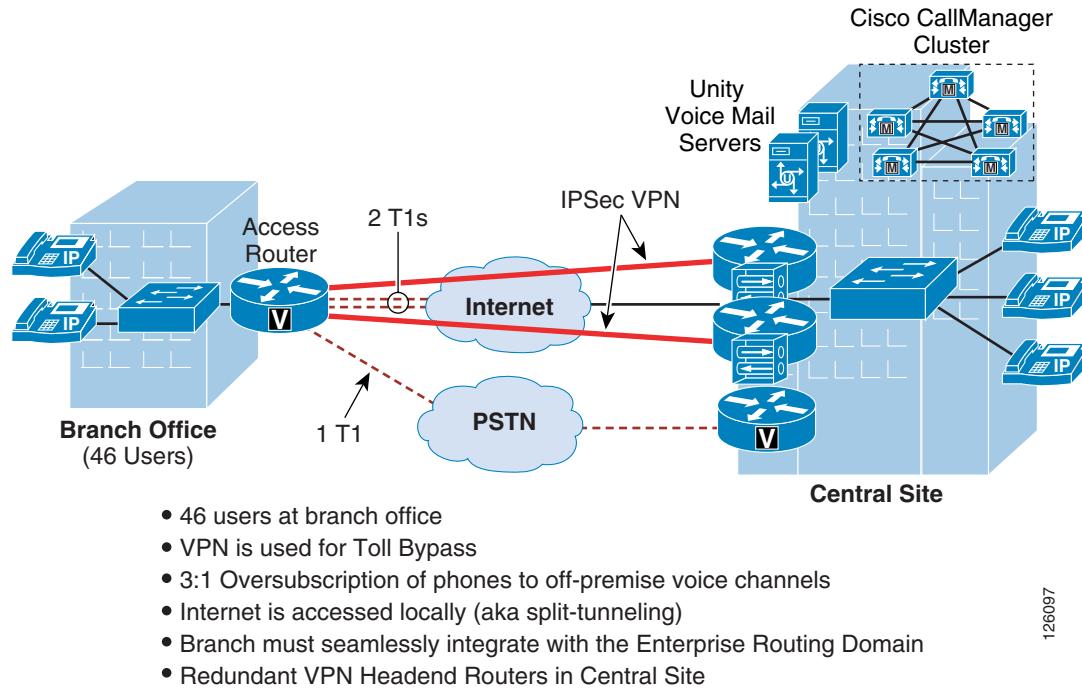
Some of the fundamental questions that need to be answered when choosing a platform for a customer network are as follows:

- What services will the office router be performing? (Cisco IOS Firewall, intrusion detection, WAN connection, voice gateway, SRST, Local Call Processing, voice mail, and so on.)
- Does the routing platform have sufficient slot and performance capacity?
- What “future-proofing” factors need to be considered, such as growth potential, including physical slot and packet processing capacity, new applications on the horizon, and so on?

Choosing the router platform requires the following basic steps:

- Determining the number of slots required to house the appropriate network modules.
- Ensuring that the platform can handle the processing required for that particular size office.

[Figure 3-1](#) shows a diagram of the example branch office network.

Figure 3-1 Example Branch Office Network

Consider a branch office that has 46 users. The requirements at this location are as follows:

- Dual IPSec VPN tunnels over the Internet are the connection back to the HQ site.
- Enhanced Interior Gateway Routing Protocol (EIGRP) is used to provide routing between the HQ and branch offices.
- Internet is accessed locally at the branch office, therefore mandating Cisco IOS Firewall and Port Address Translation (PAT).
- Intrusion detection is implemented at all office perimeters.
- Cisco CallManager provides the call control from the HQ site.
- SRST provides backup call processing in case CallManager is inaccessible.
- A 3:1 ratio of voice lines to IP phones is used to determine the number of voice lines required during the busiest hour to the PSTN and for toll-bypass.
- Half of the voice lines are used for toll-bypass and half of the voice lines access the PSTN locally.

The first step is to choose the set of platforms that can be used for the office network and that have the appropriate slot capacity to handle the required interfaces:

- (2) T1s to the Internet (VWIC-2MFT-T1)
- (1) T1 to the PSTN (AIM-Voice 30, VWIC-1MFT-T1)
- (1) Hardware-based VPN for voice over VPN (AIM-VPN/EP)
- (1) Fast Ethernet to the LAN (built-in)
- (1) Hardware IDS module (NM-CIDS)

[Figure 3-2](#) shows the platform options.

Figure 3-2 Router Slot Chart

	AIM	WIC*	NM
Cisco 1760*	0	2 & 2	0
Cisco 2651XM	1	2	1
Cisco 2691	2	3	1
Cisco 3725	2	3	2
Cisco 3745	2	3	4
Cisco 2801	2	4	0
Cisco 2811	2	4	1
Cisco 2821	2	4	1
Cisco 2851	2	4	1
Cisco 3825	2	4	2
Cisco 3845	2	4	4

126098

*1760 and ISRs
have onboard DSP

Could use but no slots
left, future growth

The platforms shown in green can accommodate the required number of network modules, Advanced Integration Modules (AIMs), and interfaces required for this office network.

Next, the performance must be validated as sufficient to handle the expected traffic in this office. Assuming the customer is cost sensitive and future growth in the platform is not important, you determine whether the performance is adequate in the 2800s as follows:

Two T1s are required for Internet and VPN

- Two T1s amount to $1.536 \text{ Mbps} * 4 = 6.144 \text{ Mbps}$ full duplex or approximately 6 BW points.


Note

This assumes that you have the T1 links fully utilized in both directions. This is probably not typically the case but is considered worst case for this design exercise.

This office network requires two T1s to the Internet, so you can calculate the approximate amount of bandwidth required from the platform. For simplification, the platform performance is defined by an allocation of bandwidth (BW) points. This is similar in concept to populating the 7200 with port adapters based on bandwidth points or the sizing of CallManager based on the number of devices and the associated device weights of the selected devices. BW points used for selecting a routing platform in this document are based on one BW point for every 1 Mbps of actual bandwidth that traverses the router. For example, the formula above shows that 6.144 Mbps or 6 BW points are required if both Internet T1s are loaded to capacity.

The performance tables shown in [Table 3-1](#) and [Table 3-2](#) show the BW points allotted to each platform with a specific combination of features enabled. The tables are divided into two general applications of traffic: data only, and voice and data.

Table 3-1 Bandwidth Points per Routing Platform (1)

	3745	3725	2691	2651XM	1760
Feature Combinations (Data Only)	BW Points				
IDS-SW	119	64	43	13	17
IDS-HW	96	62	42	15	N/A
FW/PAT	80	53	38	9	7
FW/PAT/IDS-SW	72	53	42	8	6
FW/PAT/IDS-HW	58	32	23	6	N/A
FW/PAT/IPSec	41	30	27	7	4
FW/PAT/IPSec/IDS-SW	40	28	21	6	3
FW/PAT/IPSec/IDS-HW	36	19	13	3	N/A
	3745	3725	2691	2651XM	1760
Feature Combinations (Data and Voice)	BW Points				
SRST/QoS (32K)	65	48	39	11	12
SRST/QoS/FW/PAT	49	20	18	5	5
SRST/QoS/FW/PAT/IDS-SW	42	19	17	4	4
SRST/QoS/FW/PAT/IDS-HW	30	16	12	2	N/A
SRST/QoS/FW/PAT/IPSec	20	13	9	2	2
SRST/QoS/FW/PAT/IPSec/IDS-SW	19	12	9	2	2
SRST/QoS/FW/PAT/IPSec/IDS-HW	19	10	7	2	N/A

Table 3-2 Bandwidth Points per Routing Platform (2)

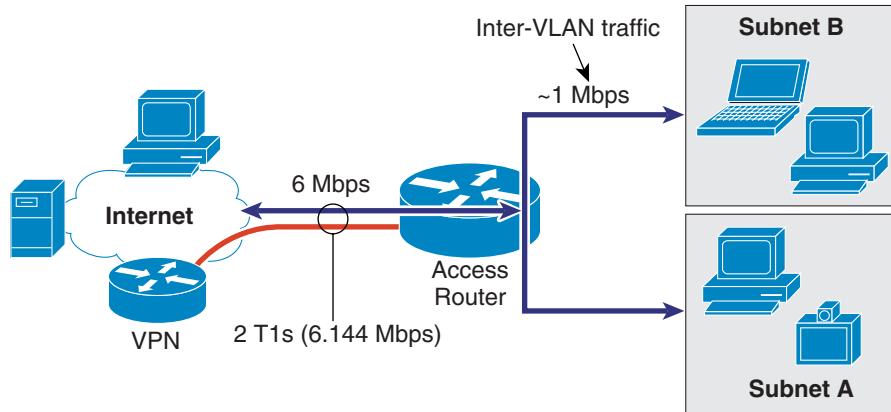
	3845	3825	2851	2821	2811	2801
Feature Combinations (Data Only)	BW Points					
IDS-SW	178	112	94	81	31	25
IDS-HW	148	119	104	81	52	NA
FW/PAT	145	126	98	81	26	34
FW/PAT/IDS-SW	94	92	68	55	16	20
FW/PAT/IDS-HW	88	82	63	50	17	NA
FW/PAT/IPSec	53	42	30	24	7	6

Table 3-2 Bandwidth Points per Routing Platform (2) (continued)

	3845	3825	2851	2821	2811	2801
FW/PAT/IPSec/IDS-SW	43	35	27	20	6	6
FW/PAT/IPSec/IDS-HW	43	35	26	19	6	N/A
	3845	3825	2851	2821	2811	2801
Feature Combinations (Data and Voice)	BW Points					
SRST/QoS (32K)	145	104	85	46	29	26
SRST/QoS/FW/PAT	70	50	47	21	12	13
SRST/QoS/FW/PAT/IDS-SW	39	28	28	25	6	6
SRST/QoS/FW/PAT/IDS-HW	40	30	29	25	6	N/A
SRST/QoS/FW/PAT/IPSec	23	24	16	12	3	3
SRST/QoS/FW/PAT/IPSec/IDS-HW	23	23	21	10	2	N/A

Now that you have validated that the 2821 can support traffic that goes between the LAN and WAN, what capacity is left for LAN-to-LAN traffic? In other words, does the access router have enough capacity to do the Layer 3 routing between LAN subnets, or will a Layer 3 switch have to be deployed?

Determining the amount of capacity left for LAN-to-LAN traffic can be calculated by taking the BW points left over after subtracting out the BW points required for the WAN. [Figure 3-3](#) works through this calculation for determining the LAN-to-LAN bandwidth capacity.

Figure 3-3 Determining the Switching Capacity for LAN-to-LAN Traffic

If there is a low amount of traffic between subnets, then L3 can be done in the router otherwise use L3 in LAN switch.

(The amount of router PPS capacity remaining for L3 switching can be seen in the PPS chart)

SRST/QoS/FW/PAT/IPSec/IDS-	7,730	4,015	3,012
----------------------------	-------	-------	-------

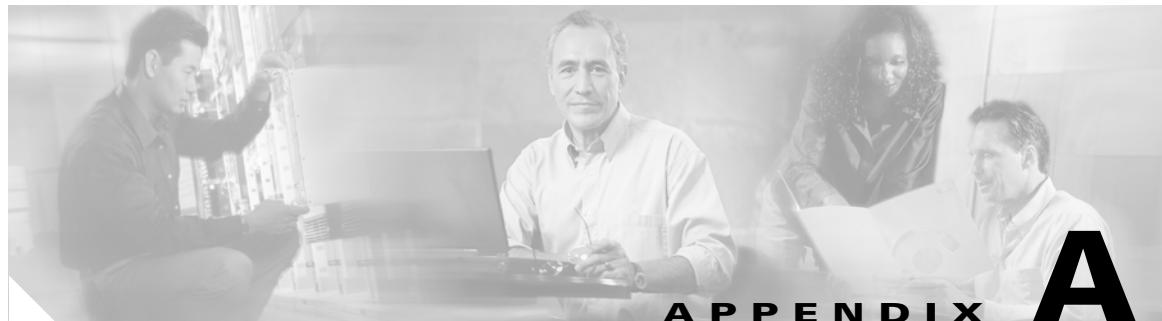
10 BW points - 6 used for the WAN = 4 bw point or 4 Mbps

126099

Subtracting the WAN BW points from the total BW points allocated to the 2821 equals the approximate amount of remaining bandwidth that can be switched by the router; in this case, it is approximately 4 Mbps. This customer expects to have a very small amount of LAN-to-LAN traffic because all corporate resources are located at the HQ site, so the 2821 is a sufficient router platform for this size office.



Note If a Layer 3 switch is required, some features such as IDS that are implemented in the router are not available and standalone appliances may be required.



APPENDIX

A

Sample Business Ready Branch Configuration Listings

The following is a sample configuration of a Business Ready Branch. There are many permutations of feature combinations when setting up the access router for the branch office, but this is one fairly comprehensive example that was used in Cisco testing.

```
!
version 12.3
service timestamps debug datetime msec
service timestamps log datetime msec
service password-encryption
!
hostname Anytown
!
!
logging buffered 4096 debugging
logging rate-limit 20
no logging console
!
clock timezone est -5
clock summer-time edt recurring
no network-clock-participate aim 0
no network-clock-participate aim 1
voice-card 6
dspfarm
!
no aaa new-model
aaa authentication login default local
ip subnet-zero
no ip source-route
ip cef
!
ip dhcp pool phones
    network 10.173.156.0 255.255.255.0
    default-router 10.173.156.1
    option 150 ip 10.59.138.4
    dns-server 10.59.138.4
!
ip dhcp pool pc
network 10.73.26.0 255.255.255.192
default-router 10.73.26.1
dns-server 10.73.26.1
!
ip inspect one-minute high 2000
ip inspect tcp max-incomplete host 100 block-time 0
ip inspect name firewall tcp
ip inspect name firewall udp
```

```

ip inspect name firewall rtsp
ip inspect name firewall netshow
ip inspect name firewall ftp
ip inspect name firewall sqlnet
ip ips po max-events 100
! disables alarming on RFC 1918 addresses detection
ip ips signature 1107 0 disable
! disables alarming ICMP on echo reply
ip ips signature 2000 0 disable
! disables alarming on ICMP host unreachable (this is commonly seen during MTU discovery)
ip ips signature 2001 0 disable
ip ips name softips
no ftp-server write-enable
!
!
class-map match-all VOICE
match ip dscp ef
class-map match-any CALL-SETUP
match ip dscp af31
match ip dscp cs3
class-map match-any INTERNETWORK-CONTROL
match ip dscp cs6
class-map match-all TRANSACTIONAL-DATA
match ip dscp af21
!
!
policy-map reorder
class VOICE
priority percent 33
class CALL-SETUP
bandwidth percent 2
class INTERNETWORK-CONTROL
bandwidth percent 5
class TRANSACTIONAL-DATA
bandwidth percent 22
class class-default
fair-queue
random-detect dscp-based
policy-map shaper
class class-default
shape average 5000000
service-policy reorder
!
!
!
crypto isakmp policy 1
encr 3des
authentication pre-share
crypto isakmp key branch address 0.0.0.0 0.0.0.0
crypto isakmp keepalive 10
!
!
crypto ipsec transform-set brb esp-3des esp-sha-hmac
mode transport
!
crypto ipsec profile dmvpn
set transform-set brb
!
interface Tunnel0
description Hub and Spoke DMVPN link
ip address 10.73.30.2 255.255.255.192
no ip redirects
ip mtu 1400
ip nhrp authentication brb

```

```

ip nhrp map multicast dynamic
ip nhrp map 10.73.30.1 192.168.8.6
ip nhrp map multicast 192.168.8.6
ip nhrp network-id 99
ip nhrp nhs 10.73.30.1
ip route-cache flow
load-interval 30
qos pre-classify
tunnel source Loopback0
tunnel destination 192.168.8.6
tunnel key 10000
tunnel path-mtu-discovery
tunnel protection ipsec profile dmvpn
!
interface Loopback0
  ip address 10.73.1.6 255.255.255.248
!
interface FastEthernet0/0
  no ip address
  no ip proxy-arp
  load-interval 30
  duplex auto
  speed auto
!
interface FastEthernet0/1
  no ip address
  ip route-cache flow
  load-interval 30
  duplex auto
  speed auto
!
interface FastEthernet0/1.16
  description DMZ
  encapsulation dot1Q 16
  ip address 10.57.0.1 255.255.255.0
  ip access-group DMZ in
  ip ips softids in
  ip virtual-reassembly
!
interface FastEthernet0/1.18
  description Voice VLAN for phones
  encapsulation dot1Q 18
  ip address 10.173.156.1 255.255.255.0
  ip access-group voice in
  ip inspect firewall in
    ! note - avoid IPS on Voice LAN
    !           with RPC sigs enabled
!
interface FastEthernet0/1.20
  description Data VLAN for PCs
  encapsulation dot1Q 20
  ip address 10.73.26.1 255.255.255.192
  ip nat inside
  ip access-group LAN in
  ip inspect firewall in
  ip ips softips in
!
interface Hssi3/0
  description ISP 5 Mpbs link
  encapsulation ppp
  load-interval 30
  hssi internal-clock
  serial restart-delay 0
  ip address 192.168.25.30 255.255.255.252

```

```

ip nat outside
ip access-group INPUT_ACL in
ip inspect firewall in
ip ips softids in
ip virtual-reassembly
service-policy output shaper
!
router eigrp 15
  passive-interface FastEthernet0/1.16
  passive-interface FastEthernet0/1.18
  passive-interface FastEthernet0/1.20
  network 10.0.0.0
  no auto-summary
!
ip classless
ip route 0.0.0.0 0.0.0.0 192.168.25.29
!
ip nat inside source list LAN interface Hssi3/0 overload
!
ip access-list extended INPUT_ACL
  remark Allow IKE and ESP from the headend router
  permit udp host 192.168.1.1 any eq isakmp
  permit esp host 192.168.1.1 any
  remark Allow ICMP
  permit icmp any any unreachable
  permit icmp any any echo-reply
  permit icmp any any packet-too-big
  permit icmp any any time-exceeded
  remark Allow DNS name lookup from router
  permit udp any eq domain any
  remark Allow the Internet to DMZ
  permit ip any 10.57.0.0 0.0.0.255
  deny ip any any
!
ip access-list extended LAN
  permit ip 10.73.26.0 0.0.0.63 any
  deny ip any any
!
ip access-list extended DMZ
  deny ip any any
!
ip access-list extended voice
  permit 10.173.156.0 0.0.0.255 any
  deny ip any any
!
control-plane
!----- CCM SRST with MGCP Fallback Voice Section -----
!
isdn switch-type primary-5ess
!
ccm-manager switchback immediate
ccm-manager fallback-mgcp
ccm-manager mgcp
ccm-manager music-on-hold
ccm-manager config server 10.59.138.4
ccm-manager config
!
controller T1 0/0
  framing esf
  linecode b8zs
  pri-group timeslots 1-24 service mgcp
!
interface Serial0/0:23
  no ip address

```

```

isdn switch-type primary-5ess
isdn incoming-voice voice
isdn bind-13 ccm-manager
no cdp enable
!
interface Service-Engine1/0
ip unnumbered Loopback0
service-module ip address 10.73.1.5 255.255.255.248
service-module ip default-gateway 10.73.1.6
!
interface Loopback0
ip address 10.73.1.6 255.255.255.248
!
ip route 10.73.1.5 255.255.255.255 Service-Engine1/0
!
voice-port 0/0:23
!
mgcp
mgcp call-agent VPN2-CM-2 2427 service-type mgcp version 0.1
mgcp rtp unreachable timeout 1000 action notify
mgcp package-capability rtp-package
no mgcp package-capability res-package
mgcp package-capability sst-package
no mgcp package-capability fxr-package
no mgcp timer receive-rtcp
mgcp sdp simple
mgcp fax t38 inhibit
mgcp rtp payload-type g726r16 static
!
mgcp profile default
!
dial-peer cor custom
!
dial-peer voice 29999 voip
description voicemail_cue
destination-pattern 2999.
session protocol sipv2
session target ipv4:10.73.1.5
codec g711ulaw
!
dial-peer voice 25 pots
description PSTN
application mgcpapp
destination-pattern 9T
port 0/0:23
!
dial-peer voice 26 pots
description PSTN
application mgcpapp
destination-pattern 91T
port 0/0:23
!
call-manager-fallback
max-conferences 8
ip source-address 10.73.1.6 port 2000
max-ephones 240
max-dn 250
voicemail 29999
call-forward busy 29999
call-forward noan 29999 timeout 3
!
line con 0
exec-timeout 61 0
password 7 0822455D0A16

```

```
line 33
no activation-character
no exec
transport preferred none
transport input all
transport output all
line aux 0
line vty 0
password 7 00071A150754
login
transport input telnet
line vty 1 4
exec-timeout 61 0
password 7 00071A150754
login
transport input telnet
line vty 5 6
exec-timeout 61 0
login
!
ntp clock-period 17175627
ntp server 172.26.176.10
ntp peer 10.73.30.1
end
```