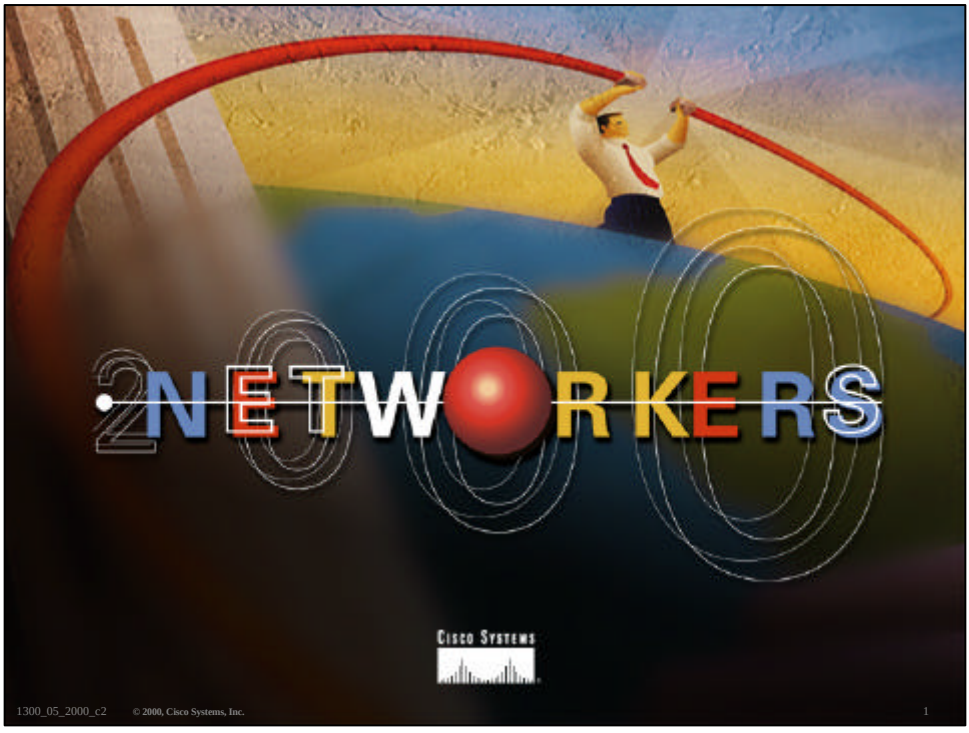


Housekeeping

- Lunch is in the Exhibit Hall D for entire week
- Restrooms and phones
- Audio tapes available for purchase at the ARC Audio Tapes booth near registration
- Please remember to fill out your evaluations and turn them in to a room attendant at the end of the day!

Who Should Attend?

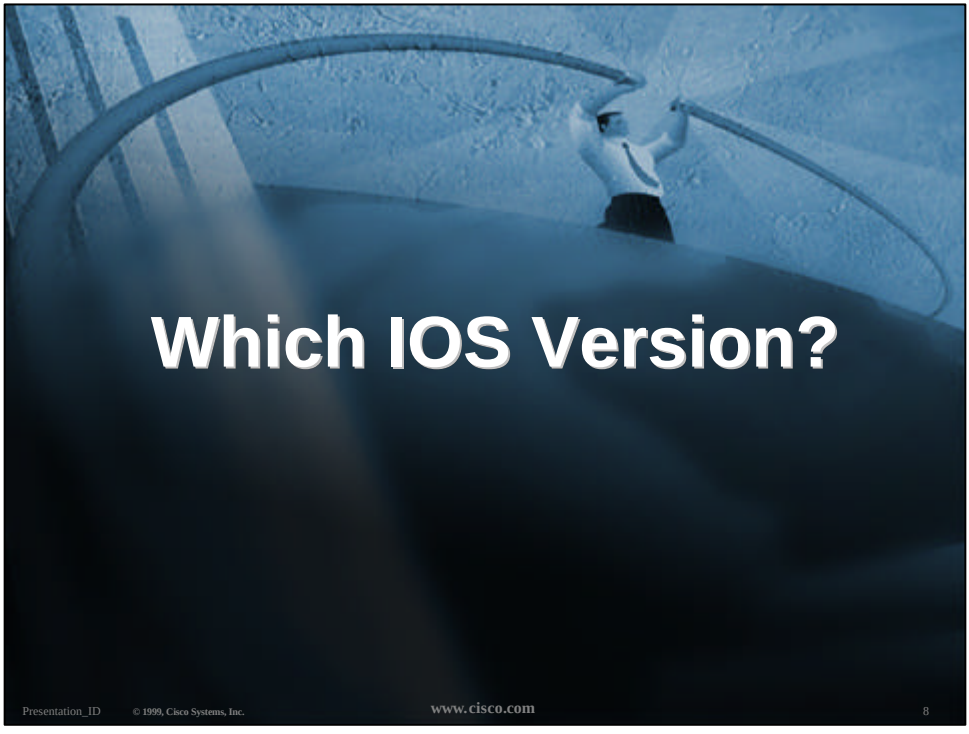
- Engineers from Existing ISPs, ASPs, Telcos, and other Internet based service providers.
- Consultants/CCIEs working with Internet based service providers.
- Anyone else interested in the *gory IOS details.*

Prerequisites

- This is not for people brand new to networking and IOS
- Know a bit about IOS.
- Know a bit about OSPF and BGP
- Know a bit about TCP/IP

Agenda for the Day

- General Features
- ISP Security
- Routing Configuration Guidelines and Updates
- ISP Design Fundamentals



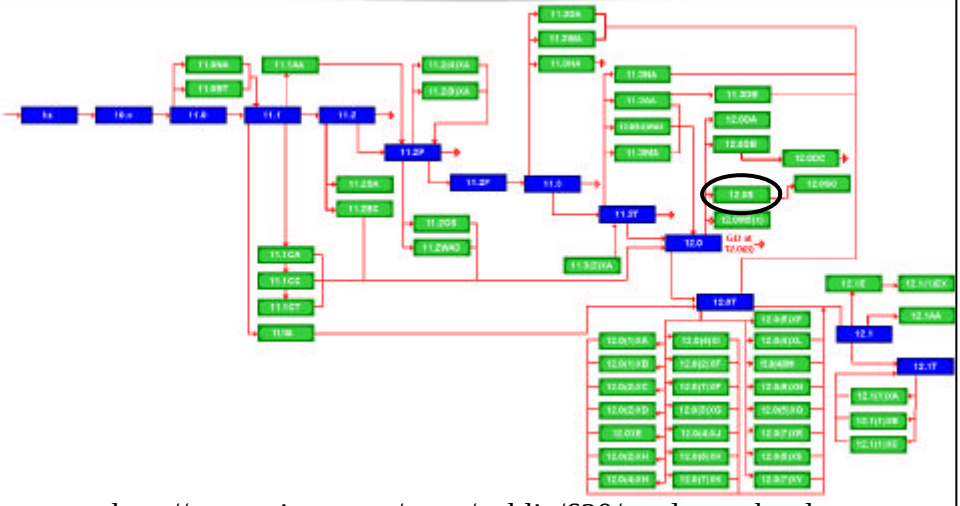
Which IOS version?

- Platforms
  - ✓ GSR, 7500 series, 7200 series
- Recommended release is 12.0S train
  - ✓ Current version is 12.0(10)S (as of May 2000)
  - ✓ Available on CCO
- Has all of latest ISP supported features

Which IOS version?

- Platforms
  - ✓ 4x00, 3600, 2600 and 2500 series
- Recommended release is the 12.0 mainline train
  - ✓ Current version is 12.0(10)
  - ✓ Has many of the features found in 11.1CC, 11.2P and 11.3T
  - ✓ Available on CCO

Cisco IOS Roadmap



http://www.cisco.com/warp/public/620/roadmap.shtml

IOS Software and Router Management

IOS Software Management Flash Memory

- Good practice is to have at least two distinct flash memory volumes
  - ✓ allows backup image(s)
  - ✓ back out path in case of upgrade problems
- Partition the built-in flash
  - ✓ `partition flash 2 8 8`
- Install a PCMCIA flash card in external slot(s) - 20Meg flash cards are worth it.

IOS Software Management Flash Memory

- Ensure that there is a configured backup to selected IOS image
  - ✓ backup image is previous “good” image

```
boot system flash slot0:rsp-pv-mz.120-10.S
boot system flash slot1:rsp-pv-mz.111-32.CC
boot system flash
```

  - ✓ which means “boot quoted image from slot0:. If it isn’t there, boot the quoted image in slot1:. If that isn’t there, try the first image available in flash

IOS Software Management System Memory

- Good practice is to maximise router memory
  - ✓ allows for the rapidly growing Internet
- 128Mbytes needed for full Internet routing table
  - ✓ will (just) work with 64Mbytes, but BGP inefficient
- Recognised that equipment works best when “left alone”

IOS Software Management When to Upgrade

- Upgrades needed when:
  - ✓ bug fixes released
  - ✓ new hardware support
  - ✓ new software features required
- Otherwise:

If it isn’t broken, don’t fix it!

Digression - Loopback Interface

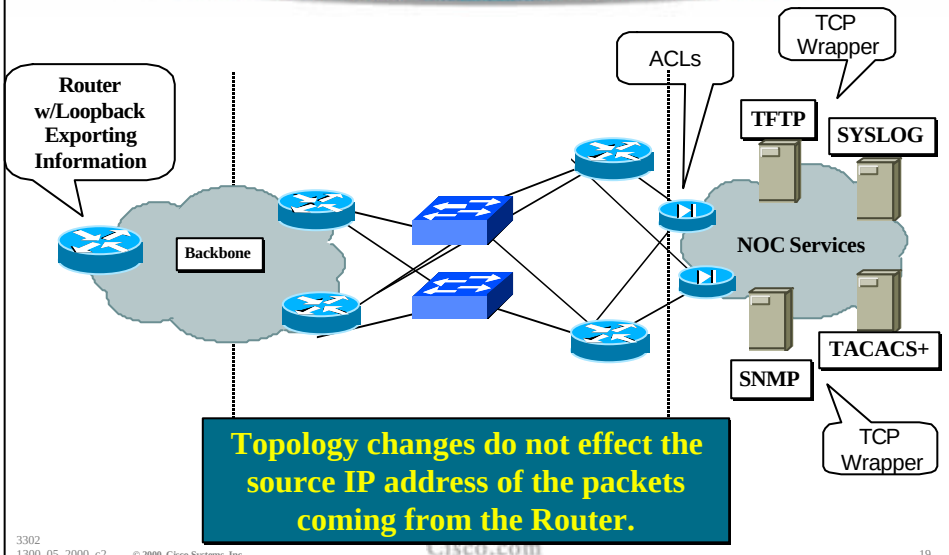
- Most ISPs make use of the router loopback interface.
- IP address configured is a host address
- Configuration example:

```
interface loopback 0
 description Loopback Interface of CORE-GW3
 ip address 215.18.3.34 255.255.255.255
```


Digression - Loopback Interface

- Loopback interfaces on ISP backbone usually numbered:
  - ⇨ out of one contiguous block, or
  - ⇨ using a geographical scheme, or
  - ⇨ using a per PoP scheme
- Aim is to increase stability, aid administration, and improve security

Digression - Loopback Interface



Digression - Loopback Interface

- Loopback interface is not “redundant” or “superfluous”
- Multitude of uses to ease security, access, management, information and scalability of router and network
- Protects the ISP’s Management Systems
- Use the loopback!

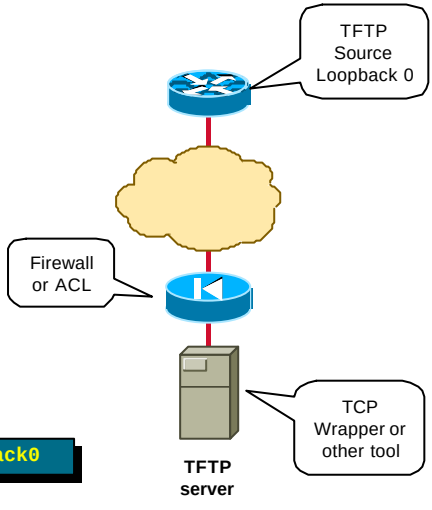
Configuration Management

- **Backup NVRAM configuration off the router:**
  - ✓ **write configuration to TFTP server**
  - ✓ **TFTP server files kept under revision control**
  - ✓ **router configuration built from master database**
- **Allows rapid recovery in case of emergency**

Configuration Management

- **Secure the TFTP Server**
  - ✓ **TFTP Loopback 0 on Router**
  - ✓ **Firewall/ACL**
  - ✓ **Wrapper on TFTP Server which only allows the router's loopback address**

```
ip tftp source-interface Loopback0
```



FTP Client Support

- TFTP has it's security limitations.
- FTP Client support is added in 12.0. This allows for FTP upload/downloads.
- Remember to use the same security/redundancy options with loopback 0:

```
ip ftp source-interface loopback 0
```


FTP Client Support

```
7206-AboveNet-SJ2#copy ftp://bgreene:XXX@ftp.cisco.com slot0:
Source filename []? /cisco/ios/12.0/12.0.9S/7200/c7200-k3p-
mz.120-9.S.bin
Destination filename [c7200-k3p-mz.120-9.S.bin]?
Accessing ftp://bgreene:XXX@ftp.cisco.com
//cisco/ios/12.0/12.0.9S/7200/c7200-k3p-mz.120-
9.S.bin...Translating "ftp.cisco.com"...domain server
(207.126.96.162) [OK]


Loading /cisco/ios/12.0/12.0.9S/7200/c7200-k3p-mz.120-9.S.bin
```

Larger Configurations

- **Compress Configuration**
  - ✓ **Used when configuration required is larger than configuration memory (NVRAM) available.**
  - ✓ `service compress-config`
- **FLASH or remote server**
  - ✓ **Used when NVRAM compression is not enough**

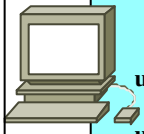
Use Detailed Logging

- **Off load logging information to a logging server.**
- **Use the full detailed logging features to keep exact details of the activities.**

```
service timestamps debug datetime msec localtime show-timezone
service timestamps log datetime msec localtime show-timezone
logging buffered 16384
logging trap debugging
logging facility local7
logging 169.223.32.1
logging source-interface loopback0
```

Use Detailed Logging

- Two Topologies used:
  - ✓ Central Syslog Servers in Operations Center
  - ✓ Syslog Servers in Major POPs



```
unix% tail cisco.log
 Feb 17 21:48:26 [10.1.1.101.9.132] 31: *Mar  2 11:51:55 CST:
   %SYS-5-CONFIG_I: Configured from console by vty0 (10.1.1.2)
unix% date
 Tue Feb 17 21:49:53 CST 1998
unix%
```

Network Time Protocol

- If you want to cross compare logs, you need to synchronize the time on all the devices.
- Use NTP
  - ✓ From external time source

    Upstream ISP, Internet, GPS, atomic clock
  - ✓ From internal time source
  - ✓ Router can act as *stratum 1* time source

Network Time Protocol

- **Set timezone**

```
clock timezone <name> [+/-hours [mins]]
```

- **Router as source**

```
ntp master 1
```

- **External time source (master)**

```
ntp server a.b.c.d
```

- **External time source (equivalent)**

```
ntp peer e.f.g.h
```


Network Time Protocol

- **Example Configuration:**

```
clock timezone SST 8
ntp update-calendar
ntp source loopback0
ntp server <other time source>
ntp peer <other time source>
ntp peer <other time source>
```

Network Time Protocol

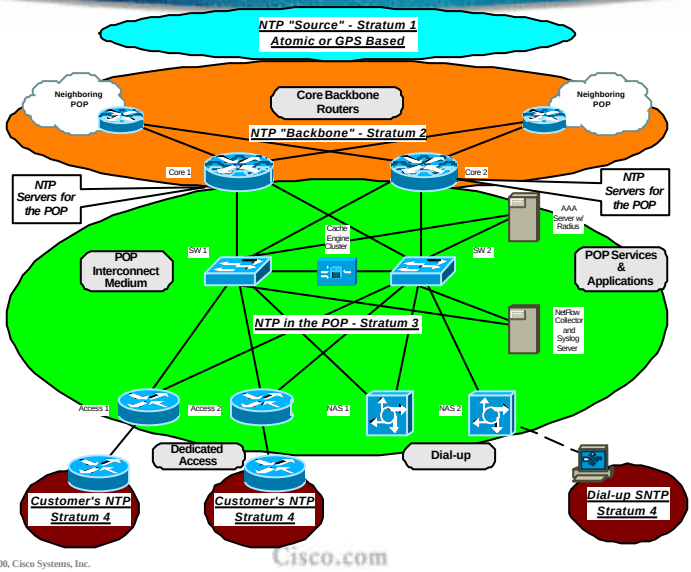
- Network Time Protocol (NTP) used to synchronize the time on all the devices.
- NTP packets leave router with loopback address as source
- Configuration example:

```
ntp source loopback0
ntp server 169.223.1.1 source loopback 1
```


Network Time Protocol

- Motivation - NTP Security:
  - ✓ NTP systems can be protected by filters which only allow the NTP port to be accessed from the loopback address block
- Motivation - Easy to understand NTP peerings:
  - ✓ NTP associations have the loopback address recorded as source address, not the egress interface.

Network Time Protocol



Network Time Protocol

- **Where to get NTP Reference Sources?**
  - ✓ **http://www.eecis.udel.edu/~ntp/hardware.html**
- **Attaching a Telecom Solutions GPS Clock to the Router's AUX port:**

```
Excalabur(config)#line aux 0
Excalabur(config-line)#ntp refclock telecom-solutions pps ?
  cts   PPS on CTS
  none  No PPS signal available
  ri    PPS on RI
```

SNMP

- Remove any SNMP commands if SNMP is not going to be used.
- If SNMP is going to be used:
  - ✓ `access-list 98 permit 169.223.1.1`
  - ✓ `access-list 98 deny   any`
  - ✓ `snmp-server community 5nmc02m RO 98`
  - ✓ `snmp-server trap-source Loopback0`
  - ✓ `snmp-server trap-authentication`
  - ✓ `snmp-server host 169.223.1.1 5nmc02m`

HTTP Server

- HTTP Server in IOS from 11.1CC and 12.0S
  - ✓ router configuration via web interface
- Disable if not going to be used (disabled by default):

```
no ip http server
```

- Configure securely if going to be used:

```
ip http server
ip http port 8765
ip http authentication aaa
ip http access-class <1-99>
```

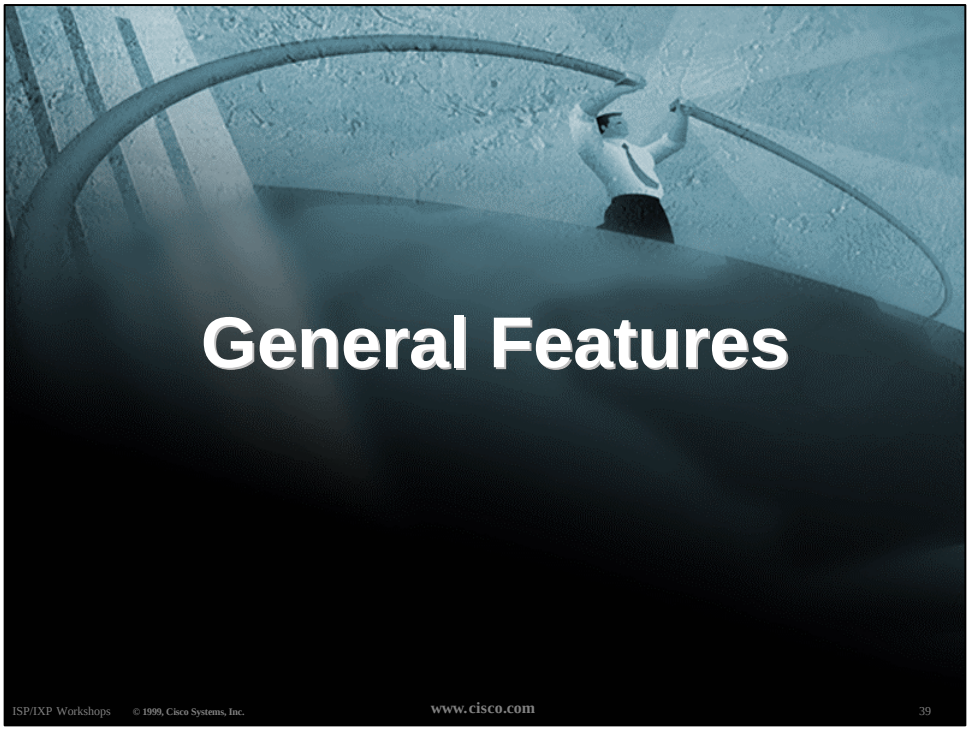
Core Dumps

- Cisco routers have a *core dump* feature that will allow ISPs to transfer a copy of the core dump to a specific FTP server.
- Set up a FTP account on the server the router will send the core dump to.
- The server should NOT be a public server
  - ✓ Use filters and secure accounts
  - ✓ Locate in NOC with NOC Staff access only
  - ✓ Enough Disk Space to handle the dumps

Core Dumps

- Example configuration:

```
ip ftp username cisco
ip ftp password 7 045802150C2E
ip ftp source-interface loopback 0
exception protocol ftp
exception dump 169.223.32.1
```



Command Line Interface Features

- **Some Convenient Editing Keys**
  - ✓ **TAB** command completion
  - ✓ **arrow keys** scroll history buffer
  - ✓ **ctrl A** beginning of line
  - ✓ **ctrl E** end of line
  - ✓ **ctrl K** delete all chars to end of line
  - ✓ **ctrl X** delete all chars to beginning of line
  - ✓ **ctrl W** delete word to left of cursor
  - ✓ **esc B** back one word
  - ✓ **esc F** forward one word

Command Line Interface Features

- **CLI now has string searches**
  - ✓ `show configuration | [begin|include|exclude] <regexp>`
- **Pager "--more--" now has string searches**
  - ✓ `/<regexp>, -<regexp>, +<regexp>`
- **"More" command has string searches**
  - ✓ `more <filename> | [begin|include|exclude] <regexp>`

Command Line Interface Features

- **Example:**

Defiant#show running-config | begin router bgp
router bgp 200
 no synchronization
 neighbor 4.1.2.1 remote-as 300
 neighbor 4.1.2.1 description Link to Excalabur
 neighbor 4.1.2.1 send-community
 neighbor 4.1.2.1 version 4
 neighbor 4.1.2.1 soft-reconfiguration inbound
 neighbor 4.1.2.1 route-map Community1 out
  maximum-paths 2
 --More--

Interface Configuration

- "ip unnumbered"
  - ✓ no need for an IP address on point-to-point links
  - ✓ keeps IGP small
- "description"
  - ✓ customer name, circuit id, cable number, etc
  - ✓ on-line documentation!
- "bandwidth"
  - ✓ used by IGP
  - ✓ documentation!

Interface Configuration - Example

- **ISP router**

```
!
interface loopback 0
description Loopback interface on GW2 Router
ip address 215.17.3.1 255.255.255.255
!
interface Serial 5/0
description 128K HDLC link to Galaxy
Publications Ltd [galpub1] WT50314E  R5-0
bandwidth 128
ip unnumbered loopback 0
!
ip route 215.34.10.0 255.255.252.0 Serial 5/0
```

- **Customer router**

```
!
interface Ethernet 0
description Galaxy Publications LAN
ip address 215.34.10.1 255.255.252.0
!
interface Serial 0
description 128K HDLC link to Galaxy
Internet Inc WT50314E  C0
bandwidth 128
ip unnumbered ethernet 0
!
ip route 0.0.0.0 0.0.0.0 Serial 0
```

Cisco Express Forwarding (CEF)

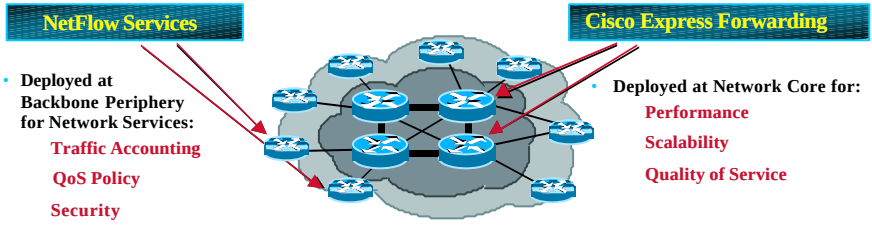
- **Rationale—changing Internet traffic/topology dynamics required optimized L3 switching paradigm for IP:**

Traffic Driven

- Stable traffic patterns
- Performance fluctuations
- Demand caching

Topology Driven

- Dynamic environment
- Predictable, scaleable, performance
- Full topology forwarding



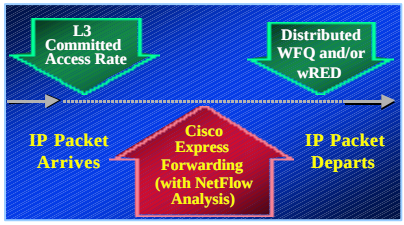
CEF—Benefits

- **Performance**
  - ✓ Implements Cisco patented expedited IP address lookup
- **Scalability**
  - ✓ Full L3 topology distributed, local on-card route processing
- **Resilience**
  - ✓ Consistent switching performance even during major topology changes/network convergence
- **Advanced functionality switching**
  - ✓ E.g. accounting, Class-of-Service, security/DoS prevention via RPF checking, tunneling etc.

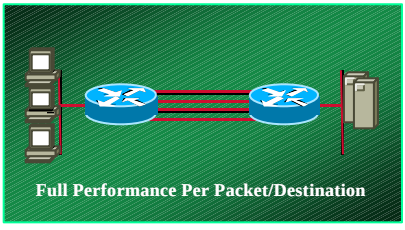
CEF—Significance

Consolidated Switching Path

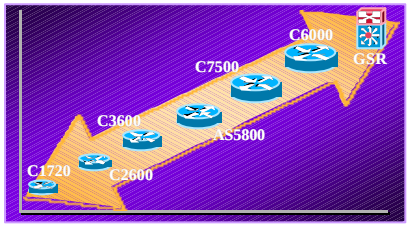
Basis for L3 Quality-of-Service



Load Balancing



Across all Cisco IOS Platforms

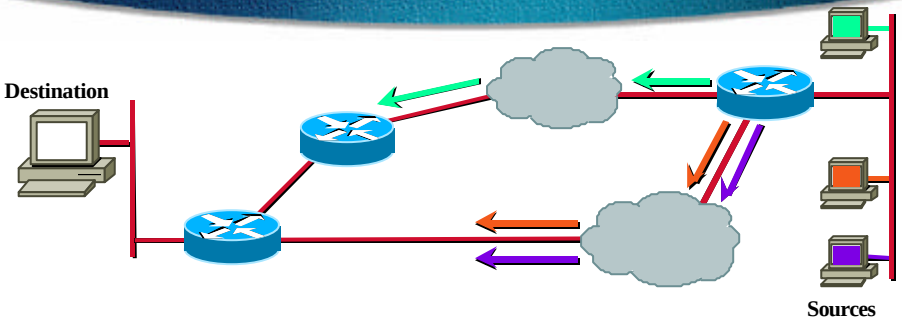


CEF Defaults in 12.0S

On this platform...	The default is...
Cisco 7000 series equipped with RSP7000	CEF is not enabled.
Cisco 7200 series	CEF is not enabled.
Cisco 7500 series	CEF is enabled.
Cisco 12000 series Gigabit Switch Router	Distributed CEF is enabled.

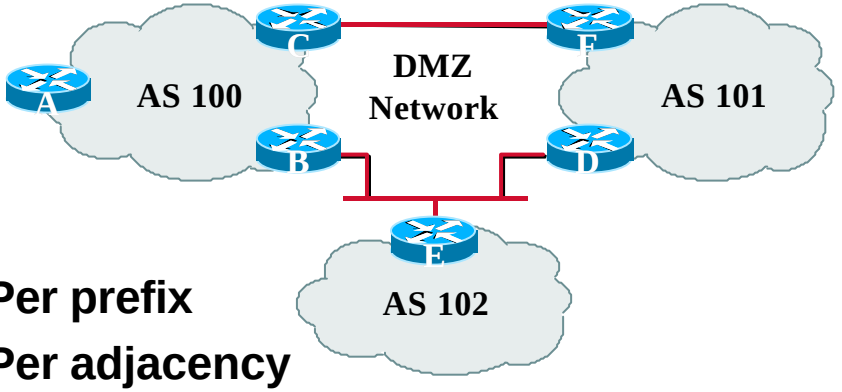
- **7500 with VIP2/4 Cards should have *ip cef distributed* turned on!**
- **7200 should have *ip cef* turned on!**

CEF Load-Sharing



- Per packet and enhanced per destination
- Enhanced per destination is based on source and destination IP addresses
- Each destination flow takes a single, separate path
- Reduces need for per packet load-sharing

CEF Accounting

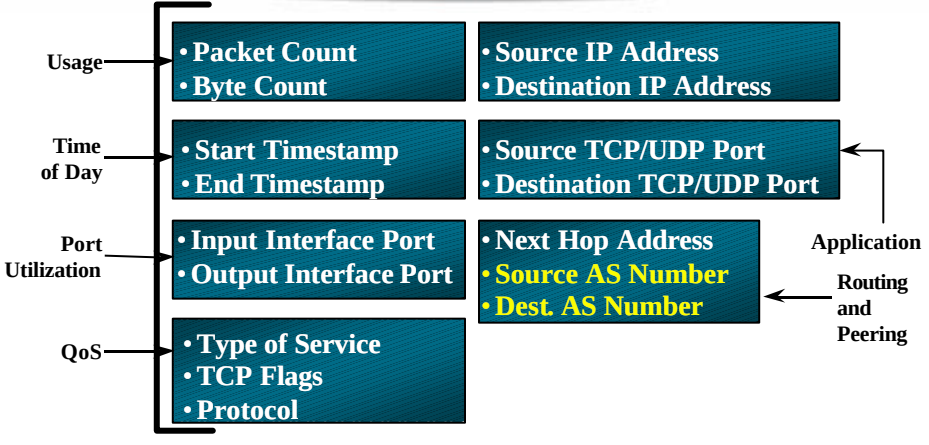


- Per prefix
- Per adjacency
- Per DMZ nexthop accounting

Netflow

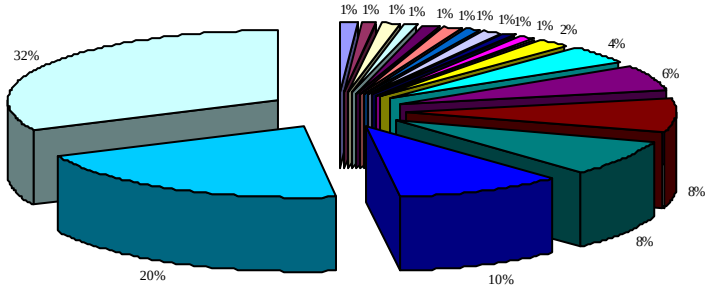
- Providers network administrators with "packet flow" information
- Allows:
  - ✓ Security monitoring
  - ✓ Network management and planning
  - ✓ Customer billing
  - ✓ Traffic flow analysis
- Available from 11.1CC for 7x00 and 12.0 for remaining router platforms

Netflow Data Record (V5)



Netflow - Capacity Planning

Public Routers 1 , 2, 3 Month of September Outbound Traffic



WEC	WebTV	ABSNET	AOL	Compuserve
SURAnet	IBM	OARNet	NIH	PacBell Internet Service
JHU	C&W	UMD	AT&T	BBN
Erols	Digex	Other		

Netflow

- **Configuration example:**

```
interface serial 5/0
 ip route-cache flow
```

- **If CEF not configured, Netflow enhances existing switching path (i.e. optimum switching)**
- **If CEF configured, Netflow becomes a flow information gatherer and feature acceleration tool**

Netflow

- **Information export:**
  - ✓ **router to collector system**

```
ip flow-export version 5 [origin-as|peer-as]
ip flow-export destination x.x.x.x <udp-port>
```

- **Flow aggregation (new in 12.0S):**
  - ✓ **router sends aggregate records to collector system**

```
ip flow-aggregation cache as|prefix|dest|source|proto
  enabled
  export destination x.x.x.x <udp-port>
```


Router-Based Netflow Aggregation (v8)

Key Fields \ Agg Schemes	AS	ProtocolPort	SourcePrefix	DestinationPrefix	Prefix
Source Prefix			-		-
Source Prefix Mask			-		-
Destination Prefix				-	-
Destination Prefix Mask				-	-
Source App Port		-			
Destination App Port		-			
Input Interface	-		-		-
Output Interface	-			-	-
IP Protocol		-			
Source AS	-		-		-
Destination AS	-			-	-
First Timestamp	-	-	-	-	-
Last Timestamp	-	-	-	-	-
# of Flows	-	-	-	-	-
# of Packets	-	-	-	-	-
# of Bytes	-	-	-	-	-

- **Enables router to summarize NetFlow data**
- **Reduces NetFlow Export data volume**
- **Decreases NetFlow Export bandwidth requirements**

Netflow - Simple Traffic Engineering

- **Sample Output on router:**

```
Beta-7200-2>sh ip cache flow
IP packet size distribution (17093 total packets):
   1-32   64   96  128  160  192  224  256  288  320  352  384  416  448  480
   .000 .735 .088 .054 .000 .000 .008 .046 .054 .000 .009 .000 .000 .000 .000

    512  544  576 1024 1536 2048 2560 3072 3584 4096 4608
   .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000

IP Flow Switching Cache, 1257536 bytes
  3 active, 15549 inactive, 12992 added
  210043 ager polls, 0 flow alloc failures
  last clearing of statistics never
Protocol         Total    Flows   Packets Bytes  Packets Active(Sec) Idle(Sec)
--------         Flows     /Sec     /Flow  /Pkt     /Sec     /Flow     /Flow
TCP-Telnet          35      0.0        80    41      0.0      14.5      12.7
UDP-DNS             20      0.0         1    67      0.0       0.0      15.3
UDP-NTP           1223      0.0         1    76      0.0       0.0      15.5
UDP-other        11709      0.0         1    87      0.0       0.1      15.5
ICMP                 2      0.0         1    56      0.0       0.0      15.2
Total:           12989      0.0         1    78      0.0       0.1      15.4

SrcIf         SrcIPaddress    DstIf         DstIPaddress    Pr SrcP DstP  Pkts
Et1/1         144.254.153.10  Null          144.254.153.127 11 008A 008A     1
Et1/1         144.254.153.112 Null          255.255.255.255 11 0208 0208     1
Et1/1         144.254.153.50  Local         144.254.153.51  06 701D 0017    63
```


Netflow - Billing/Accounting Options

NetFlow Statistics

IP NetFlow Switching Cache, 29999 Active, 2769 Inactive, 58411388 added **Statistics Cleared 141949 Seconds Ago**

Protocol	Total Flows	Flows/ Sec.	Packets/ Flow	Bytes/ Pkt	Packets/ Sec.	Active Sec/ Flow	Idle Sec/ Flow
TCP— Telnet	267,034	1.8	233	75	439.3	182.6	36.5
FTP	1,030,837	7.2	10	78	76.6	22.6	43.7
FTPD	554,967	3.9	164	345	641.3	52.7	15.7
WWW	32,107,858	226.2	15	247	3610.6	13.5	28.1
SMTP	3,526,231	24.8	13	159	323.1	10.2	23.6
X	9,600	0.0	121	129	8.2	148.2	55.1
BGP	111,096	0.7	14	77	11.5	229.2	61.1
Other	5,729,172	40.3	70	220	2858.1	71.0	41.3
UDP— TFTP	2,398	0.0	3	62	0.0	13.4	69.5
DNS	12,875,077	90.7	2	110	195.4	5.4	43.6
Other	1,489,072	10.4	30	293	321.8	28.5	68.7
ICMP	665,771	4.6	13	259	62.8	75.7	66.8
IGMP	5,144	0.0	18	278	0.6	82.4	64.3
IPINIP	4,450	0.0	933	377	29.2	166.7	61.0
IP— Other	2,693	0.0	11	136	0.2	80.8	65.7
TOTAL	58,381,400	411.3	20	227	8579.4	0.0	0.0

- **Extensive statistics maintained on L3 device**
- **Snapshot summary traffic characterization**

Netflow Feature Acceleration

- NetFlow Accelerates
  - ✓ NetFlow Policy Routing (NPR)
  - ✓ Router-based network data encryption
  - ✓ Access Control Lists (ACL)
  - ✓ RSVP
  - ✓ IP Accounting
  - ✓ Network Address Translation (NAT)
  - ✓ Committed Access Rate (CAR)
  - ✓ Web Cache Control Protocol (WCCP)
  - ✓ MultiNode Load Balancing (MNLB) *(not in 12.0S)*
- Availability of such acceleration will be announced on a feature-by-feature basis

```
ip flow-cache feature-accelerate
```


IP Switching Path - Hidden Commands

- show interface switching
- show interface <interface> switching
- show interface stat
- show interface <interface> stat

Using DNS

- Map names to addresses
- Descriptive names

```
ip domain-name
ip name-server
```

- Sample trace through network:

```
4:Received echo from sj-wall-2.cisco.com [198.92.1.138] in 440 msec.
5:Received echo from barrnet-gw.cisco.com [192.31.7.37] in 335 msec.
6:Received echo from paloalto-cr1.bbnplanet.net [131.119.26.9] in 335 msec.
7:Received echo from paloalto-br2.bbnplanet.net [131.119.0.194] in 327 msec.
8:Received echo from core6-hssi6-0.SanFrancisco.mci.net [206.157.77.21] in 468 msec.
9:Received echo from bordercore1-loopback.Washington.mci.net [166.48.36.1] in 454 msec.
10:Received 48 bytes from www.getit.org [199.233.200.55] in 466 msec
```


Turn on Nagle

- Telnet was designed to do one character, one packet dialog.
- John Nagle's algorithm (RFC 896) helps alleviate the small-packet problem in TCP.

```
service nagle
```

- Lessens the load on the CPU when using “show XXXX” commands

IP MAC accounting

- **Calculate total packet counts and byte counts for a LAN interface which receives/sends IP packets from/to each unique MAC address**
- **Record a timestamp for the last packet received/sent for each unique MAC address**
- **Available only on ethernet, FastEthernet and FDDI**
- **Available from 11.1(19)CC**

IP MAC Accounting

- **Use command *ip accounting mac* {*input* | *output*} to enable**
- ***show interface* <*interface*> *mac***

Example:

```
Ethernet0/1/3
  Input  (511 free)
    0000.0c04.7ad5(167):  9 packets, 1026 bytes, last: 20512ms ago
                    Total:  9 packets, 1026 bytes
  Output  (510 free)
    ffff.ffff.ffff(0  ):  16 packets, 960 bytes, last: 58108ms ago
    0000.0c04.7ad5(167):  9 packets, 1026 bytes, last: 21060ms ago
                    Total:  25 packets, 1986 bytes
```

IP MAC accounting - the fine print

- Fast Ether Channel supported
- 512 mac address per interface per direction(input or output)
- Support fast/optimum/flow/CEF switching

IP Precedence Accounting

- Calculate the total packet counts and byte counts for an interface which receives/sends IP packets, and sorts out the results based on different IP precedence
- 8 precedence levels
- Supported on any interface and sub-interface
- Switching mode supported: CEF/DCEF/Flow/Optimum

IP Precedence Accounting

- **Use command *ip accounting precedence* {*input* | *output*} to enable**
- ***show interface* <*interface*> *precedence***

Example:

```
Ethernet0/1/3
  Input
    Precedence 0:  9 packets, 1026 bytes
  Output
    Precedence 0:  9 packets, 1026 bytes
    Precedence 6:  16 packets, 960 bytes
```


Command Summary

- **Global Commands**

```
ip cef (-distributed)
ip cef accounting [per-prefix] [non-recursive]
ip flow-cache feature-accelerate
ip domain-name
ip name-server
service nagle
```

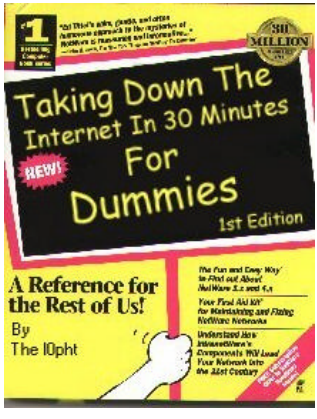
- **Interface Commands**

```
description
bandwidth
ip load-sharing [per-packet] [per-destination]
ip route-cache flow
```

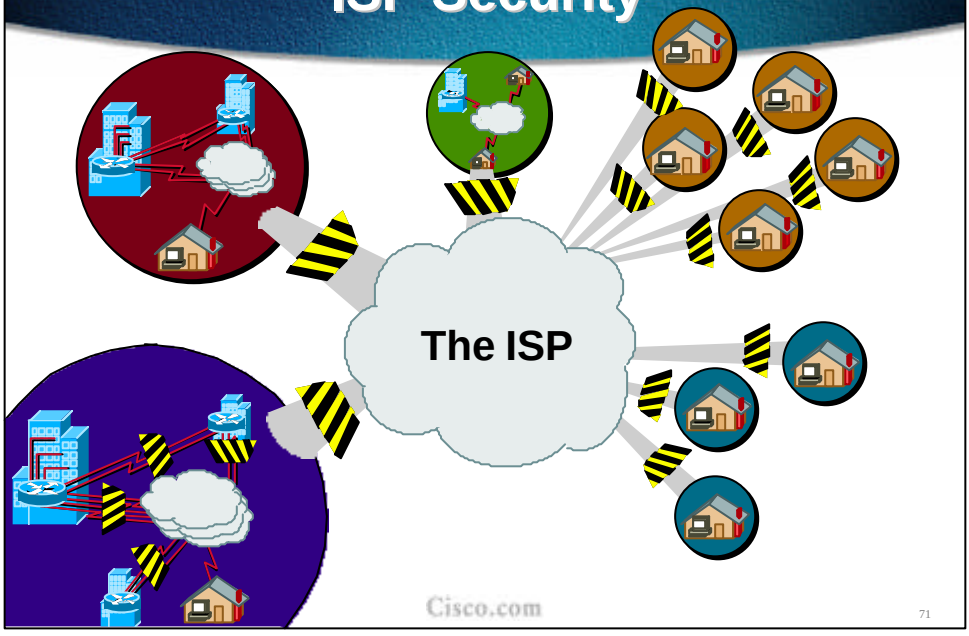
ISP Security

ISP Security

- **ISPs need to:**
  - ✓ **Protect themselves**
  - ✓ **Help protect their customers from the Internet**
  - ✓ **Protect the Internet from their customers**



ISP Security



ISP Security

- **Where to start …..**
  - ✓ **Cisco Internet Security Advisories**
    http://www.cisco.com/warp/public/779/largeent/security/advisory.html
  - ✓ **Cisco IOS documentation for 12.0**
    http://www.cisco.com/univercd/data/doc/software/11_2/2cbook.html
  - ✓ **RFC2196 (Site Security Handbook)**
  - ✓ **Networker's Security Sessions**

ISP Security

- Securing the Router
- Securing the Routing Protocols
- Securing the Network
- Latest Attacks
- Tracking DoS/DDOS Attacks through an ISP's Network

Global Services You Turn OFF

- **Some services turned on by default, should be turned off to save memory and prevent security breaches/attacks**

```
no service finger
no service pad
no service udp-small-servers
no service tcp-small-servers
no ip bootp server
```


Interface Services You Turn OFF

- **Some IP features are great for Campus LANs, but do not make sense on a ISP backbone.**
- **All interfaces on an ISP's backbone router should have the follow as a *default:***

```
no ip redirects
no ip directed-broadcast
no ip proxy-arp
```

Cisco Discovery Protocol

- Lets network administrators discover neighbouring Cisco equipment, model numbers and software versions
- Should not be needed on ISP network

```
no cdp run
```

- Should not be activated on any public facing interface: IXP, customer, upstream ISP
- Disable per interface

```
no cdp enable
```


Cisco Discovery Protocol

```
Defiant#show cdp neighbors detail
-------------------------
Device ID: Excalabur
Entry address(es):
  IP address: 4.1.2.1
Platform: cisco RSP2,  Capabilities: Router
Interface: FastEthernet1/1,  Port ID (outgoing port): FastEthernet4/1/0
Holdtime : 154 sec

Version :
Cisco Internetwork Operating System Software
IOS (tm) RSP Software (RSP-K3PV-M), Version 12.0(9.5)S, EARLY DEPLOYMENT MAINTEN
ANCE INTERIM SOFTWARE
Copyright (c) 1986-2000 by cisco Systems, Inc.
Compiled Fri 03-Mar-00 19:28 by htseng

Defiant#
```

Login Banner

- **Use a good login banner, or nothing at all:**

```
banner login ^
  Authorised access only
  This system is the property of Galactic Internet
  Disconnect IMMEDIATELY if you are not an authorised user!
  Contact noc@net.galaxy +99 876 543210 for help.
^
```


Exec Banner

- **Useful to remind logged in users of local conditions:**

```
banner exec ^
  PLEASE NOTE - THIS ROUTER SHOULD NOT HAVE A DEFAULT ROUTE!
  It is used to connect paying peers. These 'customers'
  should not be able to default to us.
  The config for this router is NON-STANDARD
  Contact Network Engineering +99 876 543234 for more info.
^
```

Use Enable Secret

- Encryption '7' on a Cisco is reversible.
- The “enable secret” password encrypted via a one-way algorithm.

```
enable secret <removed>
no enable password
service password-encryption
```


VTY and Console port timeouts

- Default idle timeout on async ports is 10 minutes 0 seconds

```
exec-timeout 10 0
```

- Timeout of 0 means permanent connection
- TCP keepalives on incoming network connections

```
service tcp-keepalives-in
```

VTY Security

- **Access to VTYs should be controlled, not left open. Consoles should be used for last resort admin only:**

```
access-list 3 permit 215.17.1.0 0.0.0.255
access-list 3 deny   any
line vty 0 4
 access-class 3 in
 exec-timeout 5 0
 transport input telnet ssh
 transport output none
 transport preferred none
 password 7 045802150C2E
```


VTY Security

- **Use more robust ACLs with the logging feature to spot the probes on you network.**

```
access-list 199 permit tcp 1.2.3.0 0.0.0.255 any
access-list 199 permit tcp 1.2.4.0 0.0.0.255 any
access-list 199 deny   tcp any any range 0 65535 log
access-list 199 deny   ip any any log
```

VTY Access and SSHv1

- Secure Shell Supported as from IOS 12.0S
- Obtain, load and run appropriate crypto images on router
- Set up SSH on router

```
Beta7200(config)#crypto key generate rsa
```

- Add it as input transport

```
line vty 0 4
  transport input telnet ssh
```


User Authentication

- Account per user, with passwords

```
aaa new-model
aaa authentication login neteng local
username joe password 7 1104181051B1
username jim password 7 0317B21895FE
line vty 0 4
 login neteng
 access-class 3 in
```

User Authentication

- Use distributed authentication system
  - ✓ RADIUS - Recommended for User Accounting
  - ✓ TACACS+ - Recommended for Securing the Network

```
aaa new-model
aaa authentication login default tacacs+ enable
aaa authentication enable default tacacs+ enable
aaa accounting exec start-stop tacacs+
ip tacacs source-interface Loopback0
tacacs-server host 215.17.1.1
tacacs-server key CKr3t#
line vty 0 4
 access-class 3 in
```


User Authentication

TACACS+ Provides a detailed audit trail of what is happening on the network devices.

User-Name	Group-N	cmd	priv-lvl	service	NAS-Portname	task_id	NAS-IP-A	reason
bgreene	NOC	enable <cr>	0	shell	tty0	4	210.210.51.224	
bgreene	NOC	exit <cr>	0	shell	tty0	5	210.210.51.224	
bgreene	NOC	no aaa accounting exec Worksho	0	shell	tty0	6	210.210.51.224	
bgreene	NOC	exit <cr>	0	shell	tty0	8	210.210.51.224	
pfs	NOC	enable <cr>	0	shell	tty0	11	210.210.51.224	
pfs	NOC	exit <cr>	0	shell	tty0	12	210.210.51.224	
bgreene	NOC	enable <cr>	0	shell	tty0	14	210.210.51.224	
bgreene	NOC	show accounting <cr>	15	shell	tty0	16	210.210.51.224	
bgreene	NOC	write terminal <cr>	15	shell	tty0	17	210.210.51.224	
bgreene	NOC	configure <cr>	15	shell	tty0	18	210.210.51.224	
bgreene	NOC	exit <cr>	0	shell	tty0	20	210.210.51.224	
bgreene	NOC	write terminal <cr>	15	shell	tty0	21	210.210.51.224	
bgreene	NOC	configure <cr>	15	shell	tty0	22	210.210.51.224	
bgreene	NOC	aaa new-model <cr>	15	shell	tty0	23	210.210.51.224	
bgreene	NOC	aaa authorization commands 0 de	15	shell	tty0	24	210.210.51.224	
bgreene	NOC	exit <cr>	0	shell	tty0	25	210.210.51.224	
bgreene	NOC	ping <cr>	15	shell	tty0	32	210.210.51.224	
bgreene	NOC	show running-config <cr>	15	shell	tty66	35	210.210.51.224	
bgreene	NOC	router ospf 210 <cr>	15	shell	tty66	45	210.210.51.224	
bgreene	NOC	debug ip ospf events <cr>	15	shell	tty66	46	210.210.51.224	

Source Routing

- IP has provision to allow source IP host to specify route through Internet
- ISPs should turn this off, unless it is specifically required:
  - ✓ `no ip source-route`
- *traceroute -s* to investigate network failures - valuable tool.

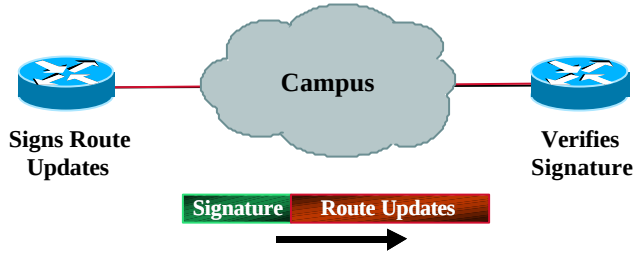
Routing Protocol Security

- Routing protocol can be attacked
  - ✓ Denial of Service
  - ✓ Smoke Screens
  - ✓ False information
  - ✓ Reroute packets

May be accidental or intentional

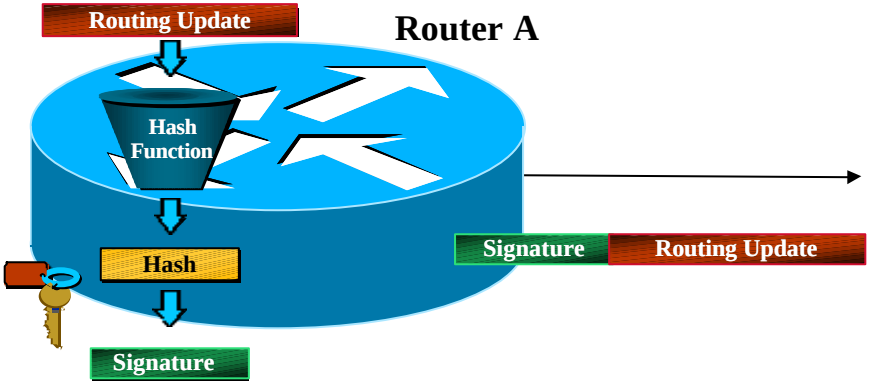
Secure Routing Route Authentication

Configure: Key and Hash Function



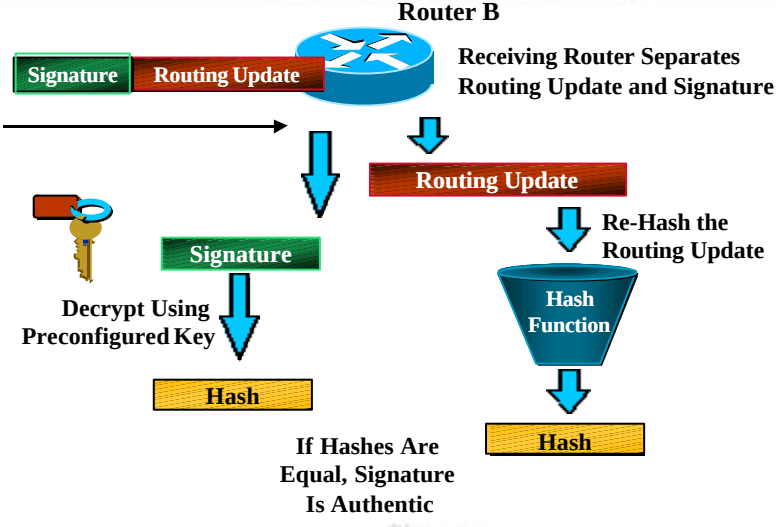
- Certifies **authenticity** of neighbor and **integrity** of route updates

Signature Generation



Signature = Encrypted Hash of Routing Update

Signature Verification



Route Authentication

- Authenticates routing update packets
- Shared key included in routing updates
  - ✓ Plain text—protects against accidental problems only
  - ✓ Message Digest 5 (MD5)—protects against accidental and intential problems

Route Authentication

- Multiple keys supported
  - ✓ Key lifetimes based on time of day
  - ✓ Only first valid key sent with each packet
- Supported in: BGP, IS-IS, OSPF, RIPv2, and EIGRP(11.2(4)F)
- Syntax differs depending on routing protocol

OSPF Route Authentication

- **OSPF Area Authentication**
  - ✓ **Two Types**
    - **Simple Password**
    - **Message Digest (MD5)**

ip ospf authentication-key *key* (this goes under the specific interface)
area *area-id* **authentication** (this goes under "router ospf <process-id>")

ip ospf message-digest-key *keyid* **md5** *key* (used under the interface)
area *area-id* **authentication message-digest** (used under "router ospf <process-id>")

OSPF & ISIS Authentication Example

- **OSPF**
  - ✓ **interface ethernet1**
  - ✓ **ip address 10.1.1.1 255.255.255.0**
  - ✓ **ip ospf message-digest-key 100 md5 cisco**
  - ✓ **!**
  - ✓ **router ospf 1**
  - ✓ **network 10.1.1.0 0.0.0.255 area 0**
  - ✓ **area 0 authentication message-digest**
- **ISIS**
  - ✓ **interface ethernet0**
  - ✓ **ip address 10.1.1.1 255.255.255.0**
  - ✓ **ip router isis**
  - ✓ **isis password cisco level-2**

BGP Route Authentication

```
router bgp 200
 no synchronization
 neighbor 4.1.2.1 remote-as 300
 neighbor 4.1.2.1 description Link to Excalabur
 neighbor 4.1.2.1 send-community
 neighbor 4.1.2.1 version 4
 neighbor 4.1.2.1 soft-reconfiguration inbound
 neighbor 4.1.2.1 route-map Community1 out
 neighbor 4.1.2.1 password 7 cisco
```


BGP Route Authentication

- **Works per neighbor or for an entire peer-group**
- **Two routers with password mis-match:**

%TCP-6-BADAUTH: Invalid MD5 digest from [peer's IP address]:11004 to [local router's IP address]:179

- **One router has a password and the other does not:**

%TCP-6-BADAUTH: No MD5 digest from [peer's IP address]:11003 to [local router's IP address]:179

Selective Packet Discard

- When a link goes to a saturated state, you will drop packets. The problem is that you will drop any type of packets - including your routing protocols.
- Selective Packet Discard (SPD) will attempt to drop non-routing packets instead of routing packets when the link is overloaded.
  - ✓ `ip spd enable (11.1 CA & CC)`
- Enabled by default from 11.2(5)P and later releases, available option in 11.1CA/CC.

Selective Packet Discard

- Attack of IP packets with bad TTL are processed switched with ICMP reply - crippling the router
  - ✓ `ip spd mode aggressive`
- `show ip spd`

  Current mode: normal.

  Queue min/max thresholds: 73/74, Headroom: 100

  IP normal queue: 0, priority queue: 0.

  SPD special drop mode: aggressively drop bad packets

What Ports Are Open on the Router?

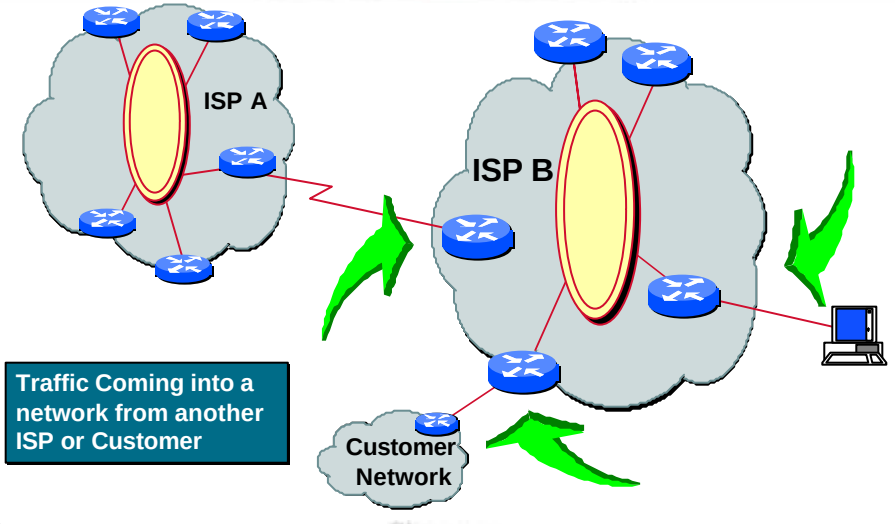
- It may be useful to see what sockets/ports are open on the router.
- *Show ip sockets*

```
7206-UUNET-SJ#show ip sockets
Proto    Remote      Port      Local       Port  In Out Stat TTY
OutputIF
 17 192.190.224.195   162 204.178.123.178  2168   0   0    0   0
 17   --listen--          204.178.123.178    67   0   0    9   0
 17 0.0.0.0           123 204.178.123.178   123   0   0    1   0
 17 0.0.0.0             0 204.178.123.178   161   0   0    1   0
```

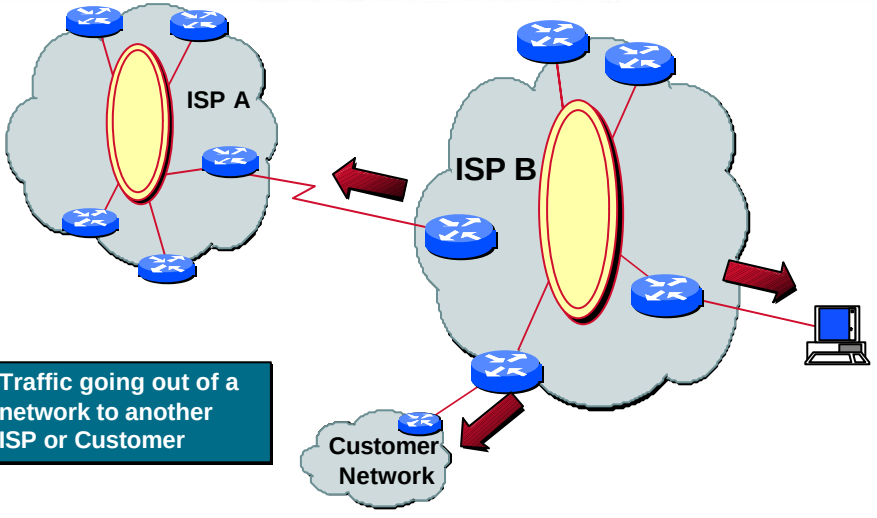

Securing the Network

- Route Filtering
- Packet Filtering
- Rate Limits

Ingress Filters - Inbound Traffic



Egress Filters - Outbound Traffic



Traffic going out of a network to another ISP or Customer

Route Filtering

Ingress and Egress Route Filtering

- There are routes that should NOT be routed on the Internet.
  - ✓ RFC 1918 and “Martian” Networks
  - ✓ 127.0.0.0/8 and Multicast blocks
  - ✓ See Bill Manning’s ID for background information:

    ftp://ftp.ietf.org/internet-drafts/draft-manning-dsua-03.txt
- BGP should have filters applied so that these routes are not advertised to or propagated through the Internet.

Ingress and Egress Route Filtering

- Quick Review
  - ✓ 0.0.0.0/8 & 0.0.0.0/32 - Default and Broadcast
  - ✓ 127.0.0.0/8 - Host Loopback
  - ✓ 192.0.2.0/24 - TEST-NET for documentation
  - ✓ 10.0.0.0/8, 172.16.0.0/12, and 192.168.0.0/16 - RFC 1918 Private Addresses
  - ✓ 169.254.0.0/16 - End node auto-config for DHCP

Ingress and Egress Route Filtering

- **Two *flavors*** of route filtering:
  - ✓ **Distribute List - widely used**
  - ✓ **Prefix List - increasingly used**
- **Both work fine - engineering preference.**

Ingress and Egress Route Filtering

Extended ACL for a BGP Distribute List

```
access-list 150 deny ip host 0.0.0.0 any
access-list 150 deny ip 10.0.0.0 0.255.255.255 255.0.0.0 0.255.255.255
access-list 150 deny ip 127.0.0.0 0.255.255.255 255.0.0.0 0.255.255.255
access-list 150 deny ip 169.254.0.0 0.0.255.255 255.255.0.0 0.0.255.255
access-list 150 deny ip 172.16.0.0 0.15.255.255 255.240.0.0 0.15.255.255
access-list 150 deny ip 192.0.2.0 0.0.0.255 255.255.255.0 0.0.0.255
access-list 150 deny ip 192.168.0.0 0.0.255.255 255.255.0.0 0.0.255.255
access-list 150 deny ip 224.0.0.0 31.255.255.255 224.0.0.0 31.255.255.255
access-list 150 permit ip any any
```

Ingress and Egress Route Filtering

BGP w/ Distribute List Flavor of Route Filtering

```
router bgp 200
no synchronization
bgp dampening
 neighbor 220.220.4.1 remote-as 210
 neighbor 220.220.4.1 version 4
 neighbor 220.220.4.1 distribute-list 150 in
 neighbor 220.220.4.1 distribute-list 150 out
 neighbor 222.222.8.1 remote-as 220
 neighbor 222.222.8.1 version 4
 neighbor 222.222.8.1 distribute-list 150 in
 neighbor 222.222.8.1 distribute-list 150 out
 no auto-summary
!
```


Ingress and Egress Route Filtering

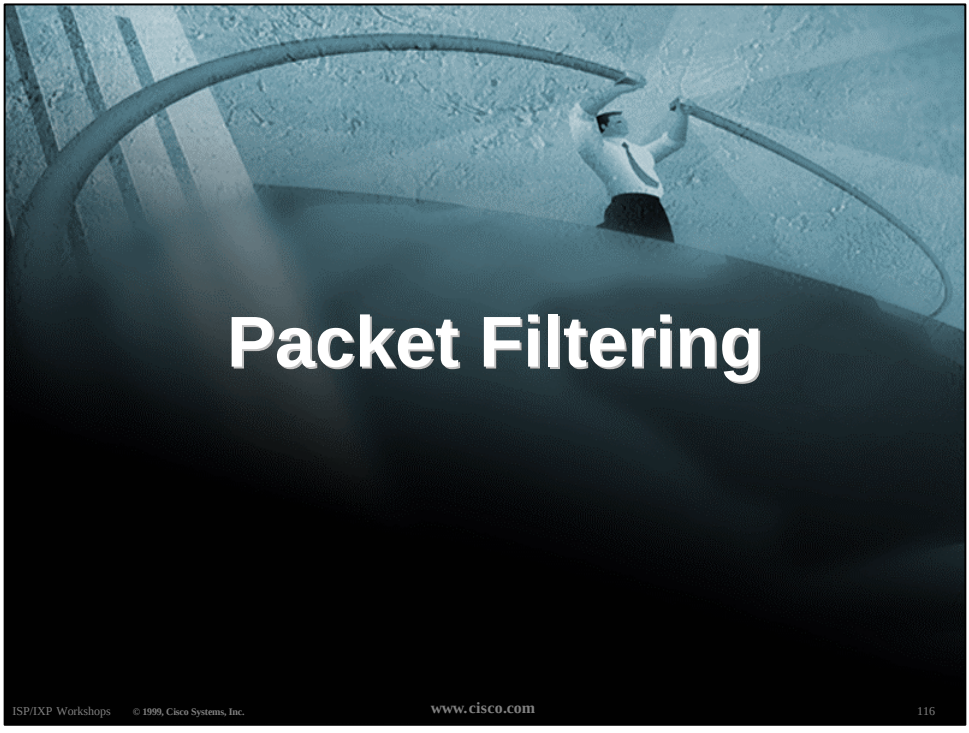
Prefix-List

```
ip prefix-list rfc1918-dsua deny   0.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny   10.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny   127.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny   169.254.0.0/16 le 32
ip prefix-list rfc1918-dsua deny   172.16.0.0/12 le 32
ip prefix-list rfc1918-dsua deny   192.0.2.0.0/24 le 32
ip prefix-list rfc1918-dsua deny   192.168.0.0/16 le 32
ip prefix-list rfc1918-dsua deny   224.0.0.0/3 le 32
ip prefix-list rfc1918-dsua permit 0.0.0.0/0 le 32
```

Ingress and Egress Route Filtering

BGP w/ Prefix-List Flavour of Route Filtering

```
router bgp 200
no synchronization
bgp dampening
 neighbor 220.220.4.1 remote-as 210
 neighbor 220.220.4.1 version 4
 neighbor 220.220.4.1 prefix-list rfc1918-dsua in
 neighbor 220.220.4.1 prefix-list rfc1918-dsua out
 neighbor 222.222.8.1 remote-as 220
 neighbor 222.222.8.1 version 4
 neighbor 222.222.8.1 prefix-list rfc1918-dsua in
 neighbor 222.222.8.1 prefix-list rfc1918-dsua out
 no auto-summary
!
```

Ingress & Egress Packet Filtering

Your customers should not be sending *any* IP packets out to the Internet with a source address other then the address you have allocated to them!

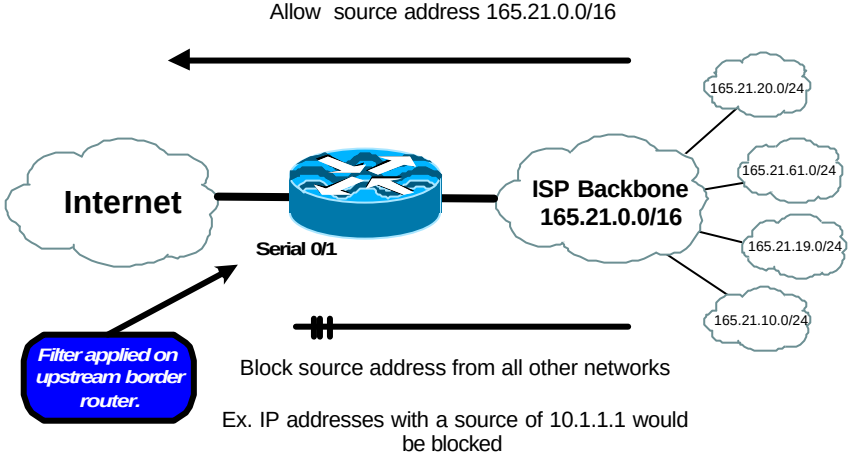
Ingress & Egress Packet Filtering

- **BCP 38/ RFC 2827**
- **Title: Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing**
- **Author(s): P. Ferguson, D. Senie**

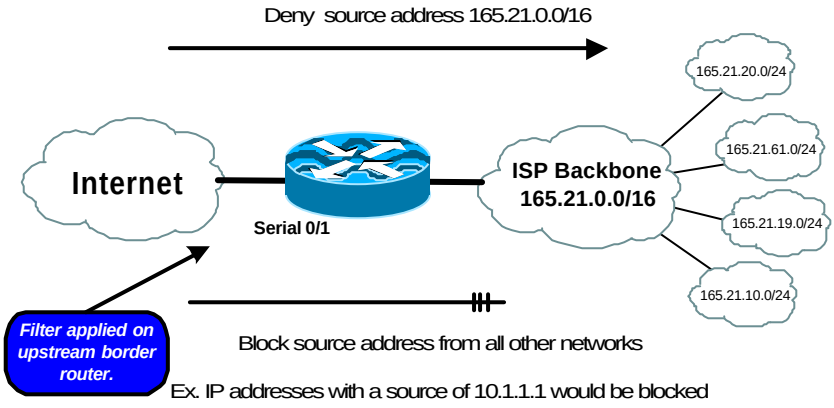
Packet Filtering

- **Static Access List on the edge of the Network.**
- **Dynamic Access List with AAA Profiles**
- **Unicast RPF**

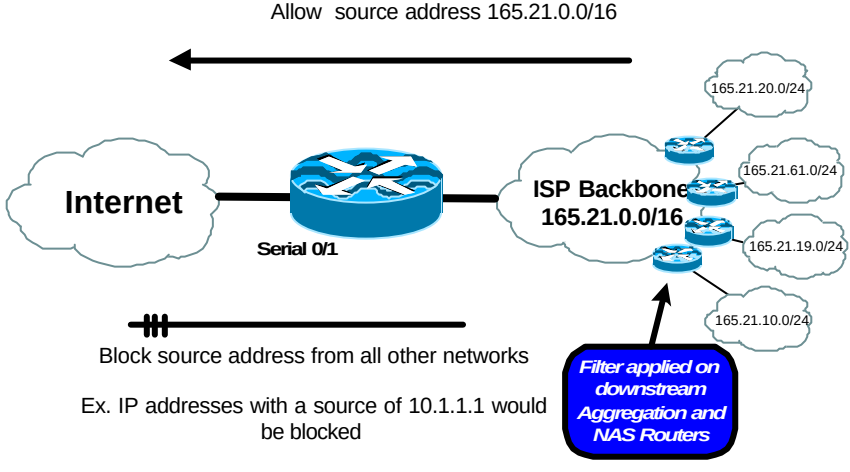
Egress Packet Filtering Upstream Border



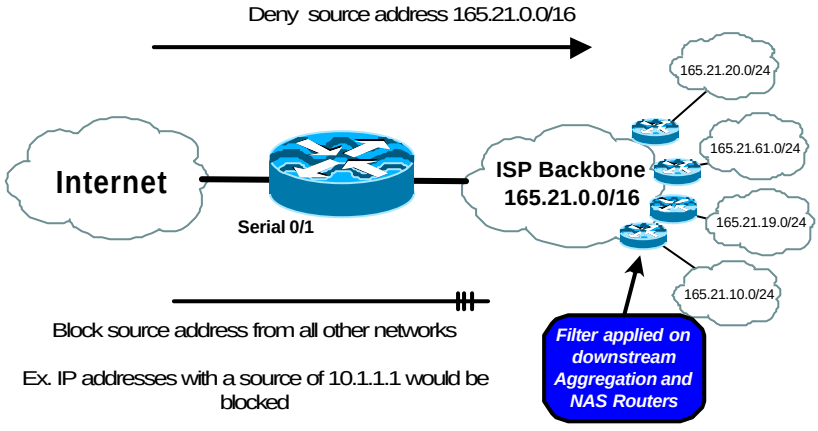
Ingress Packet Filtering Upstream Border



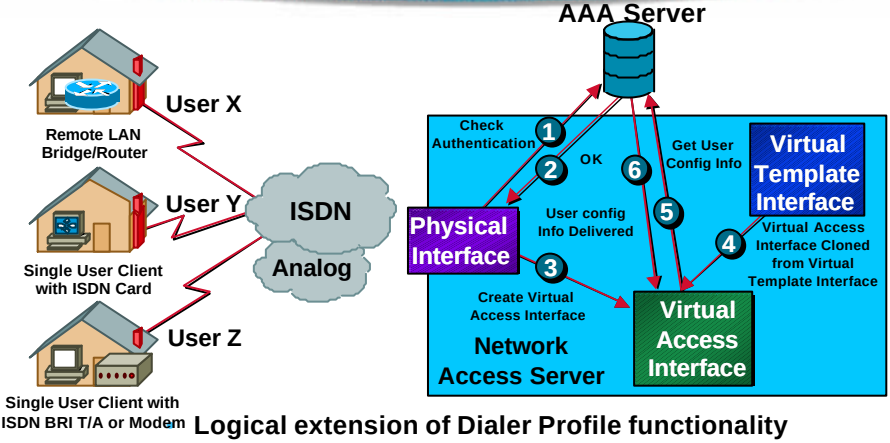
Egress Packet Filtering Customer Edge



Ingress Packet Filtering Customer Edge



Dynamic ACLs with AAA Virtual Profiles

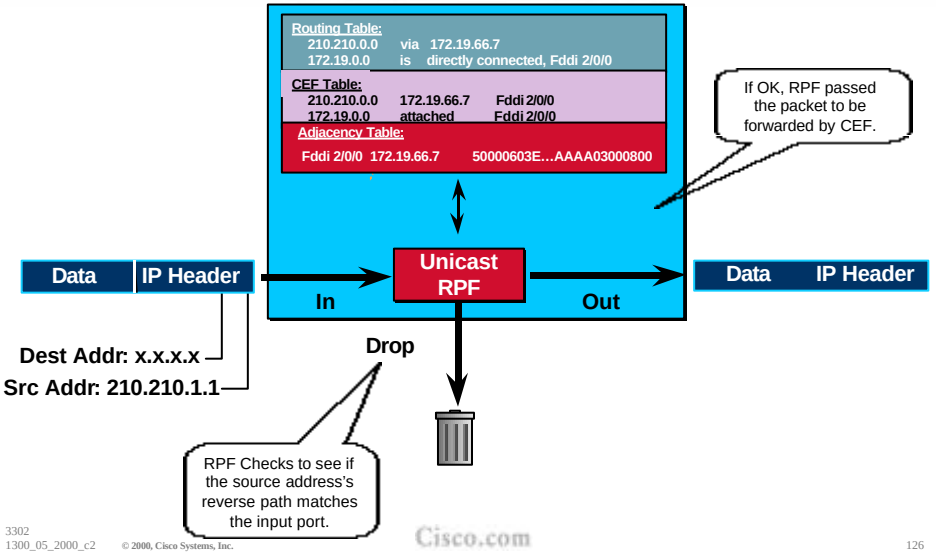


- **Logical extension of Dialer Profile functionality**
  - **ACLs stored in the Central AAA Server**
  - **Supports both Radius and Tacacs+**

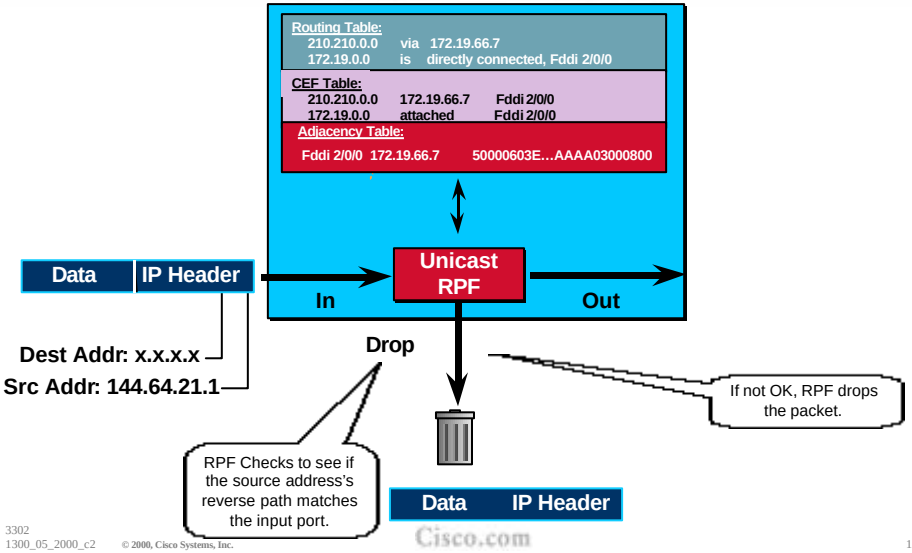
Reverse Path Forwarding

- Supported from 11.1(17)CC images
- CEF switching must be enabled
- Source IP packets are checked to ensure that the route back to the source uses the same interface
- Care required in multihoming situations

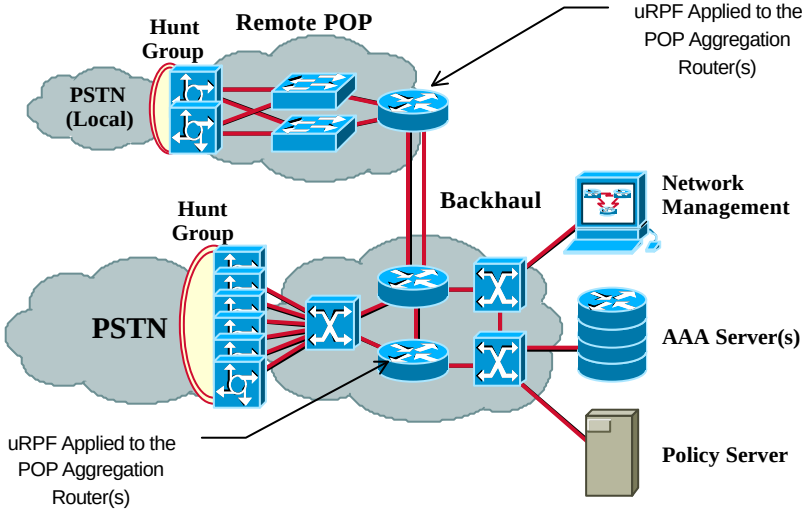
CEF Unicast RPF



CEF Unicast RPF



CEF Unicast RPF



CEF Unicast RPF

- **Configure RPF on the interface using the following interface command syntax:**

```
[no] ip verify unicast reverse-path [<ACL>]
```

- **For example on a leased line aggregation router:**

```
ip cef ! or "ip cef distributed" for an RSP+VIP based box
!
interface serial 5/0/0
  ip verify unicast reverse-path
```

- ***interface Group-Async* command for dial-up ports.:**

```
ip cef
!
interface Group-Async1
  ip verify unicast reverse-path
```

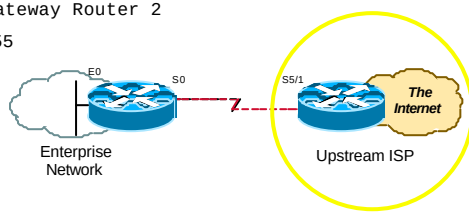

CEF Unicast RPF

- **Exceptions to RPF**

```
lookup source address in forwarding database
 if the source address is reachable via the source interface
       pass the packet
    else
 if the source is 0.0.0.0 and destination is a 255.255.255.255
       /* BOOTP and DHCP */
          pass the packet
       else if destination is multicast
          pass the packet
       else
          drop the packet
```

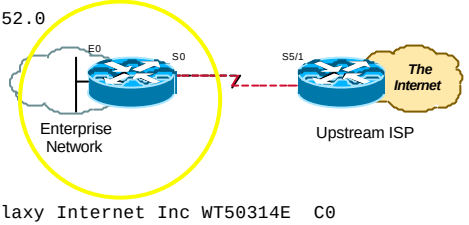
CEF Unicast RPF

```
interface loopback 0
  description Loopback interface on Gateway Router 2
  ip address 215.17.3.1 255.255.255.255
  no ip redirects
  no ip directed-broadcast
  no ip proxy-arp
!
interface Serial 5/0
  description 128K HDLC link to Galaxy Publications Ltd [galpub1]  R5-0
  bandwidth 128
  ip unnumbered loopback 0
  ip verify unicast reverse-path  ! Unicast RPF activated here
  no ip redirects
  no ip directed-broadcast
  no ip proxy-arp
!
ip route 215.34.10.0 255.255.252.0 Serial 5/0
```



CEF Unicast RPF

```
interface Ethernet 0
  description Galaxy Publications LAN
  ip address 215.34.10.1 255.255.252.0
  no ip redirects
  no ip directed-broadcast
  no ip proxy-arp
!
interface Serial 0
  description 128K HDLC link to Galaxy Internet Inc WT50314E  C0
  bandwidth 128
  ip unnumbered ethernet 0
  ip verify unicast reverse-path  ! Unicast RPF activated here
  no ip redirects
  no ip directed-broadcast
  no ip proxy-arp
!
ip route 0.0.0.0 0.0.0.0 Serial 0
```



CEF Unicast RPF

- Unicast RPF provides
  - ✓ Automatic Ingress Filtering based on routing information.
  - ✓ Can be part of the default configuration
  - ✓ Packet Drops at CEF - before the router processes spoofed packets
- If this feature is so great - why is it not used?

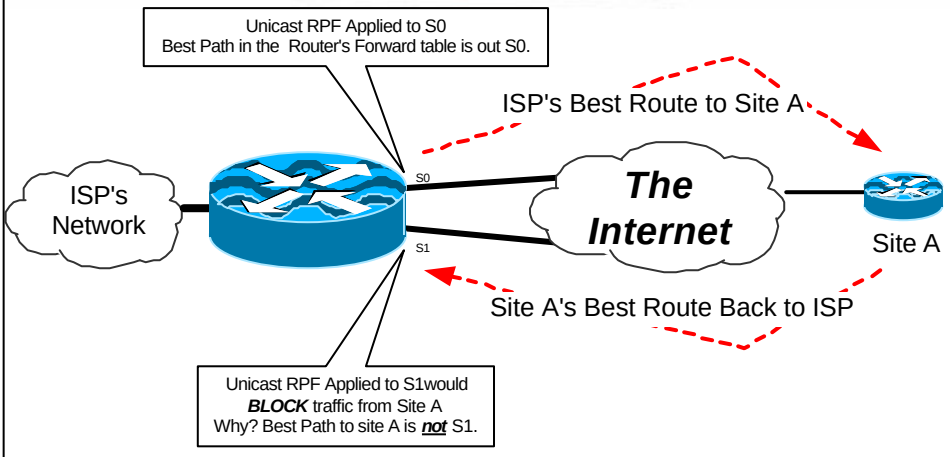
CEF Unicast RPF

- The *Myth*
  - ✓ What people say:

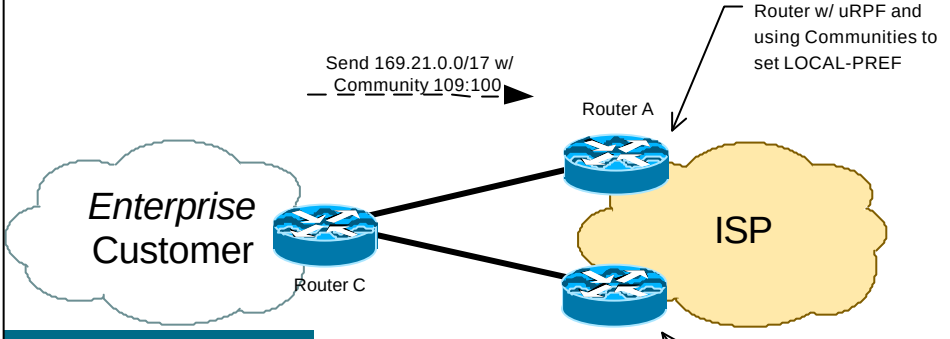
    Unicast RPF will not work with asymmetrical routing. Since the Internet has a lot of asymmetrical routing, it will not work.
  - ✓ The Real Reason:

    ISP Network Engineers have not given the feature enough thought!

CEF Unicast RPF



Unicast RPF - Dual Homed Enterprise - One ISP



Apply BGP Weights on Routers A & B to always prefer the prefix directly from Router C

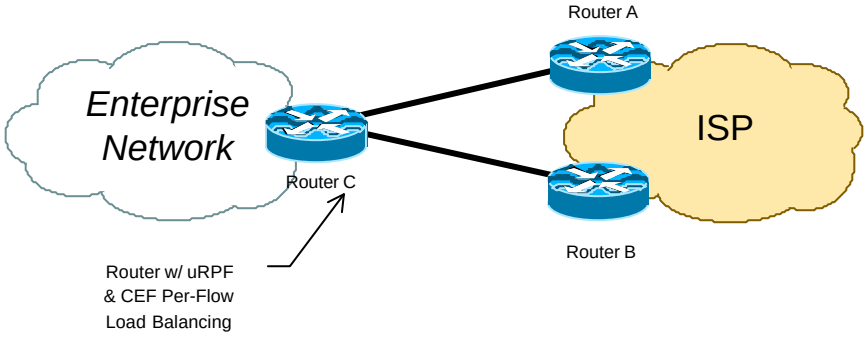
Unicast RPF - Dual Homed Enterprise - One ISP

```
Router A - Link to Router C
interface serial 1/0/1
 description Link to Acme Computer's Router C
 ip address 192.168.3.2 255.255.255.252
 ip verify unicast reverse-path
 no ip redirects
 no ip directed-broadcast
 no ip proxy-arp
 ip route-cache distributed
```


Unicast RPF - Dual Homed Enterprise - One ISP

```
router bgp 109
 neighbor 192.168.10.3 remote-as 65000
 neighbor 192.168.10.3 description Multihomed Customer –
Acme Computers
 neighbor 192.168.10.3 update-source Loopback0
 neighbor 192.168.10.3 send-community
 neighbor 192.168.10.3 soft-reconfiguration inbound
 neighbor 192.168.10.3 route-map set-customer-local-pref in
 neighbor 192.168.10.3 weight 255
 .
ip route 192.168.10.3 255.255.255.255 serial 1/0/1
ip bgp-community new-format
```

Unicast RPF - Dual Homed Enterprise - One ISP



- ✓ **Used to protect against spoof attacks**
- ✓ **Some attacks get around the RFC1918 filters by using un-allocated IP address space.**

Unicast RPF - Dual Homed Enterprise - One ISP

```
router bgp 65000
 no synchronization
 network 169.21.0.0
 network 169.21.0.0 mask 255.255.128.0
 network 169.21.128.0 mask 255.255.128.0
 neighbor 171.70.18.100 remote-as 109
 neighbor 171.70.18.100 description Upstream Connection #1
 neighbor 171.70.18.100 update-source Loopback0
 neighbor 171.70.10.100 send-community
 neighbor 171.70.18.100 soft-reconfiguration inbound
 neighbor 171.70.18.100 route-map Router-A-Community out
 neighbor 171.70.18.200 remote-as 109
 neighbor 171.70.18.200 description Upstream Connection #2
 neighbor 171.70.18.200 update-source Loopback0
 neighbor 171.70.18.200 send-community
 neighbor 171.70.18.200 soft-reconfiguration inbound
 neighbor 171.70.18.200 route-map Router-B-Community out
 maximum-paths 2
 no auto-summary
```

```
route-map Router-A-Community permit 10
 match ip address 51
 set community 109:70
!
route-map Router-A-Community permit 20
 match ip address 50
 set community 109:100
!
route-map Router-B-Community permit 10
 match ip address 50
 set community 109:70
!
route-map Router-B-Community permit 20
 match ip address 51
 set community 109:100
!
access-list 50 permit 169.21.0.0 0.0.127.255
access-list 51 permit 169.21.128.0 0.0.127.255
```

Unicast RPF - Dual Homed Enterprise - One ISP

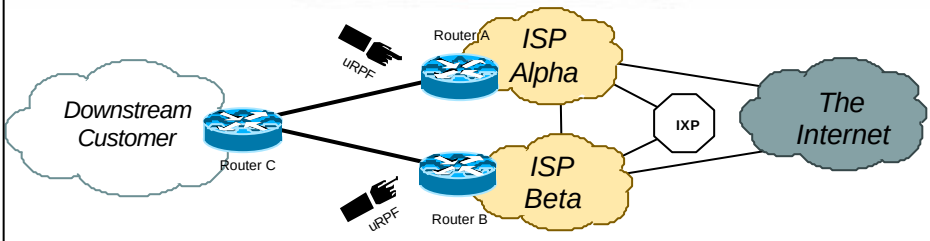
```
ip route 169.21.0.0 0.0.255.255 Null 0
ip route 169.21.0.0 0.0.127.255 Null 0
ip route 169.21.128.0 0.0.127.255 Null 0
ip route 171.70.18.100 255.255.255.255 S 1/0
ip route 171.70.18.200 255.255.255.255 S 1/1
ip bgp-community new-format
!
```

```
interface serial 1/0/
 description Link to Upstream Router A
 ip address 192.168.3.1 255.255.255.252
 ip verify unicast reverse-path
 no ip redirects
 no ip directed-broadcast
 no ip proxy-arp
 ip load-sharing per-destination
 ip route-cache distributed
!
interface serial 1/0
 description Link to Upstream ISP Router B
 ip address 192.168.3.5 255.255.255.252
 ip verify unicast reverse-path
 no ip redirects
 no ip directed-broadcast
 no ip proxy-arp
 ip load-sharing per-destination
 ip route-cache distributed
```


Unicast RPF - Dual Homed Enterprise - Two ISPs

- ISP Configuration for both ISPs are similar.
  - ✓ BGP Weight is used to over ride AS Path Prepends
- Enterprise Configuration cannot us *maximum-paths*
  - ✓ Need Equal AS paths for Maximum-paths to work

Unicast RPF - Dual Homed Enterprise - Two ISPs



- BGP Weight Override an AS Path Prepend
  - ✓ BGP Weight on Router A will keep the preferred path to be C↔A
  - ✓ BGP Weight on Router B will keep the preferred path to be C↔B

Unicast RPF - ACL

- ACLs can now be used with Unicast RPF:
  - ✓ ip verify unicast reverse-path 171
- ACLs are used to:
  - ✓ Allow exceptions to the Unicast RPF check.
  - ✓ Identify characteristics of spoofed packets being dropped by Unicast RPF

Unicast RPF - ACL

- **Cisco 7206 with Bypass ACL**

```
interface ethernet 1/1
 ip address 192.168.200.1 255.255.255.0
 ip verify unicast reverse-path 197
!
access-list 197 permit ip 192.168.201.0 0.0.0.255 any log-input

show ip interface ethernet 1/1 | include RPF
 Unicast RPF ACL 197
 1 unicast RPF drop
 1 unicast RPF suppressed drop
```


Unicast RPF - ACL

- **Cisco 7500 with a classification filter:**

```
interface ethernet 0/1/1
 ip address 192.168.200.1 255.255.255.0
 ip verify unicast reverse-path 171
!
access-list 171 deny icmp any any echo log-input
access-list 171 deny icmp any any echo-reply log-input
access-list 171 deny udp any any eq echo log-input
access-list 171 deny udp any eq echo any log-input
access-list 171 deny tcp any any established log-input
access-list 171 deny tcp any any log-input
access-list 171 deny ip any any log-input
```

Unicast RPF - ACL

- **Show the "log-input" results:**
  - ✓ **7200 - Logging done in the RP**

    show logging
  - ✓ **7500 - Logging done on the VIP**

Excalabur#sh controllers vip 4 logging

show logging from Slot 4:

.

4d00h: %SEC-6-IPACCESSLOGNP: list 171 denied 0 20.1.1.1 -> 255.255.255.255, 1 packet

.

Unicast RPF - Operations Tools

```
Excalabur#sh cef inter serial 2/0/0
Serial2/0/0 is up (if_number 8)
  Internet address is 169.223.10.2/30
  ICMP redirects are never sent
  Per packet loadbalancing is disabled
  IP unicast RPF check is enabled
  Inbound access list is not set
```

Unicast RPF - Operations Tools

- **Other Commands:**
  - ✓ **show ip traffic | include RPF**
  - ✓ **show ip interface ethernet 0/1/1 | include RPF**
  - ✓ debug ip cef drops rpf <ACL>

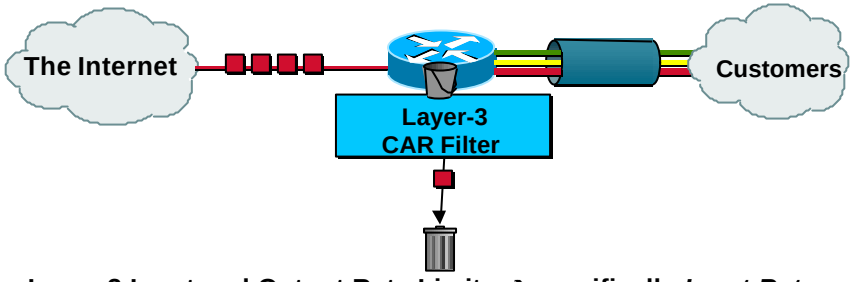
Unicast RPF - Bottomline

- **Unicast RFP is another tool to help defend the Internet**
- **Unicast RPF works when it is deployed within it's operational envelop**
- **Unicast RPF does not work when *just thrown into the network.* Give it some thought.**

Rate Limiting as a Security Tool

- **Why would anyone want to send over 45 Mbps of ICMP Traffic?**
  - ✓ **If they did, how would you stop it?**
  - ✓ **Answer - Rate Limit the *bad* traffic**
- **Committed Access Rate (CAR)**

CAR as a Security Tool



- **Layer-3 Input and Output Rate Limits ⇨ specifically *Input Rate Limits***
- **Security Filters use the Input Rate Limit to drop packets before there are forwarded through the network.**
- **Aggregate and Granular Limits**
  Port, MAC address, IP address, application, precedence
- **Excess Burst Policies**

Rate Limiting as a Security Tool

- Limit all ICMP echo and echo-reply traffic received at the borders to 256 Kbps with a small amount of burst:

```
! traffic we want to limit
access-list 102 permit icmp any any echo
access-list 102 permit icmp any any echo-reply
! interface configurations for borders
interface Serial3/0/0
 rate-limit input access-group 102 256000 8000 8000 conform-
  action transmit exceed-action drop
```

- Multiple “rate-limit” commands can be added to an interface in order to control other kinds of traffic as well.

Rate Limiting as a Security Tool

- Use CAR to limit TCP SYN floods to particular hosts -- without impeding existing connections. Some attackers have started using very high streams of TCP SYN packets in order to harm systems.
- This example limits TCP SYN packets directed at host 10.0.0.1 to 8 kbps or so:

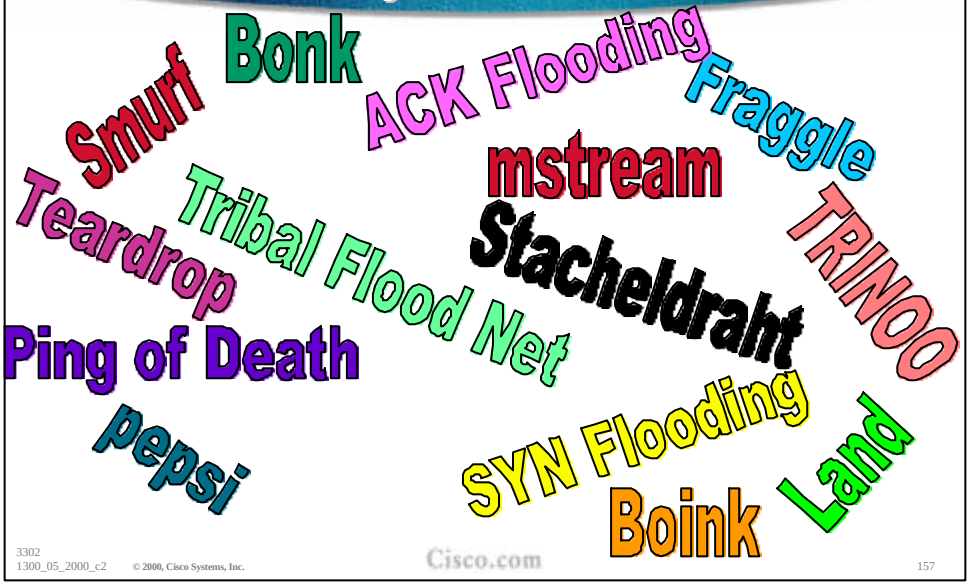
```
! We don't want to limit established TCP sessions -- non-SYN packets
access-list 103 deny tcp any host 10.0.0.1 established
! We do want to limit the rest of TCP (this really only includes SYNs)
access-list 103 permit tcp any host 10.0.0.1
! interface configurations for network borders
interface Serial3/0/0
 rate-limit input access-group 103 8000 8000 8000 conform-action transmit
exceed-action drop
```

Filtering Fragments

- **Cisco ACLs can now filter identifiable fragments.**
  - ✓ **Fragment have been used to bypass ACLs**
  - ✓ ***fragments* open is now added on ACLs to drop fragments**

```
Extended IP access list 199
    deny ip any host 169.132.32.242 (120 matches)
    deny ip any host 169.132.32.242 fragments (4506 matches)
    permit ip any any
```

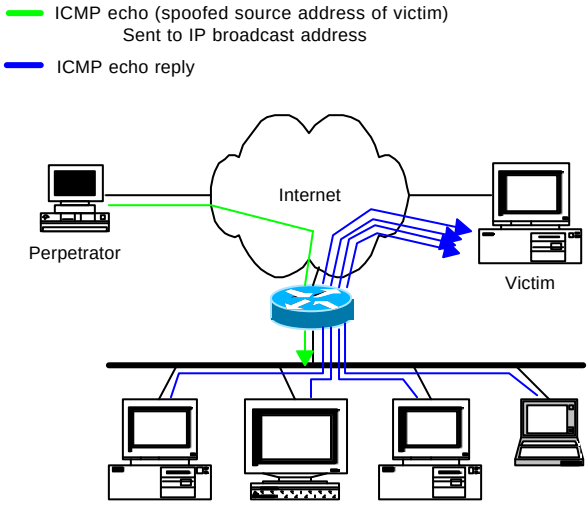

The Internet is not a nice place anymore ……



Description of "Smurfing"

- Smurf is Denial of Service attack
  - ✓ Network-based, fills access pipes
  - ✓ Uses ICMP echo/reply packets with broadcast networks to multiply traffic
  - ✓ Requires the ability to send spoofed packets
- Abuses "bounce-sites" to attack victims
  - ✓ Traffic multiplied by a factor of 50 to 200

Description of "Smurfing"



Multiplied Bandwidth - Example

- **Perpetrator has T1 bandwidth available (typically a cracked account), and uses half of it (768 Kbps) to send spoofed packets, half to bounce site 1, half to bounce site 2**
- **Bounce site 1 has a switched co-location network of 80 hosts and T3 connection to net**
- **Bounce site 2 has a switched co-location network of 100 hosts and T3 connection to net**

Multiplied Bandwidth - Consequences

- (384 Kbps * 80 hosts) = 30 Mbps outbound traffic for bounce site 1
- (384 Kbps * 100 hosts) = 37.5 Mbps outbound traffic for bounce site 2
- Victim is pounded with 67.5 Mbps (!) from half a T1!

Profiles of Participants

- Typical Perpetrators
  - ✓ Cracked superuser account on well-connected enterprise network
  - ✓ Superuser account on university residence hall network (Ethernet)
  - ✓ Typical PPP dial-up account (for smaller targets)
- Typical Bounce Sites
  - ✓ Large co-location subnets
  - ✓ Large switched enterprise subnets
  - ✓ Typically scanned for large numbers of responding hosts
- Typical Victims
  - ✓ IRC Users, Operators, and Servers
  - ✓ Providers who eliminate troublesome users' accounts

Prevention Techniques

- How to prevent your network from being the source of the attack:
  - ✓ Apply filters to each customer network
    - Ingress: Allow only those packets with source addresses within the customer's assigned netblocks
  - ✓ Apply filters to your upstreams
    - Egress: Allow only those packets with source addresses within your netblocks to protect others
    - Ingress: Deny those packets with source addresses within your netblocks to protect yourself

Prevention Techniques

- Filters will also prevent other forms of attacks as well
- If you do become a bounce site:
  - ✓ Trace the traffic streams to the edge of your network, and work with your upstream or peer in order to track the stream further
    - MCI's DoSTracker tool
    - Manual tracing/logging tips

Prevention Techniques

- How to suppress an attack if you're the victim:
  - ✓ Implement ACL's at network edges to block ICMP echo responses to your high-visibility hosts, such as IRC servers
    - Will impair troubleshooting -- "ping" breaks
    - Will still allow your access pipes to fill
  - ✓ Work with upstream providers to determine the help they can provide to you
    - Blocking ICMP echoes for high-visibility hosts from coming through your access pipes
    - Tracing attacks

Prevention Techniques

- Technical help tips for Cisco routers - One:
  - ✓ BugID CSCdj35407 - "fast drop" ACL code
    - This bug fix optimizes the way that packets denied by an ACL are dropped within IOS, reducing CPU utilization for large amounts of denied traffic.
    - First major release of integration is 11.1(14)CA
    - Not available in 11.2 yet, but coming

Prevention Techniques

- **Technical help tips for Cisco routers - Two:**
  - ✓ **BugID CSCdj35856 - ACL logging throttles**

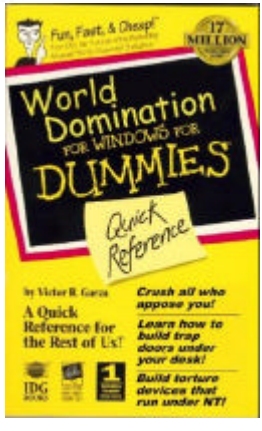
    This bug fix places a throttle in IOS which will allow a user to specify the rate at which logging will take place of packets which match a condition in an ACL where “log” or “log-input” is specified.

    First maintenance release of integration is 11.1(14.1)CA

    Not available in 11.2 yet, but coming

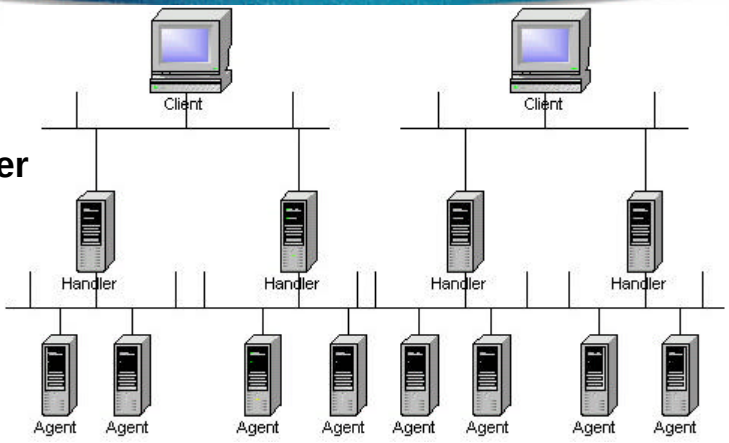
DDoS versus DoS

- **Same methods and tools as DoS**
- **Much larger scale attacks - Elephant hunting**
- **Uses hundreds or even thousands of attacking points to overwhelm target**
- **Very difficult to determine difference between DDoS and normal network outage**



DDoS Method

- Client
- Handler
- Agent



Client / Master

- End User
- Using Generic tools and Root Kits
- Sweeps sites for Handler
- Plants Handler, not Agents
- Only talks to Handler through strong encryption
- Use Free dial up services, Internet Cafes, Wireless Airports, and Internet Dialtone

Handler / Daemon

- Planted by Client
- Sweeps for Agent machines
- Infects systems
- Coordinates agents to start attack
- Uses strong encryption
- Works with Root Kits

Agent / Zombie

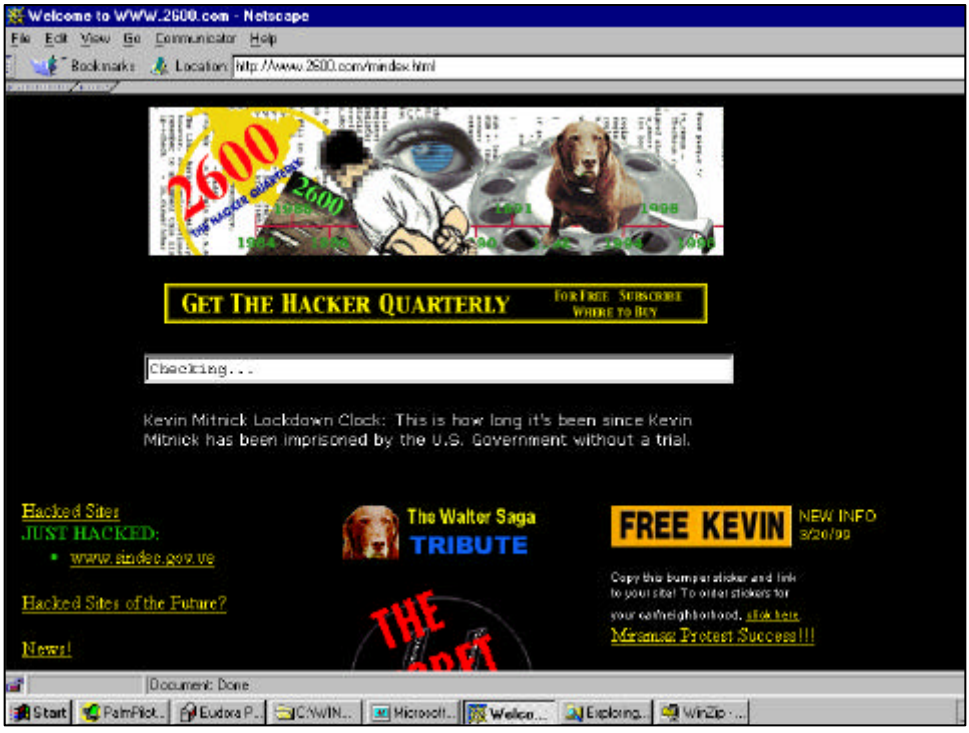
- Infected with attack software
- Uses canned attack or attack scripts
- Actually launches the attack
- Receives instructions only from Handler - trigged by time or instruction
- Uses Strong Encryption
- Usually lightly administrated system - University
- Unknowing participant - think xDSL and Cable Customers

Discovering Zombies

- Infected users unaware
- User needs education
- Slight traffic increase
- Needs Host based detection tools
- Zombie Detection tools
  - ✓ www.fbi.gov/nipc/trinoo.htm
  - ✓ staff.washington.edu/dittrich/misc/ddos_scan.tar

DDoS Tools

- Trinoo
- Tribe Flood Network
- TFN2K
- Stacheldraht
- MStream

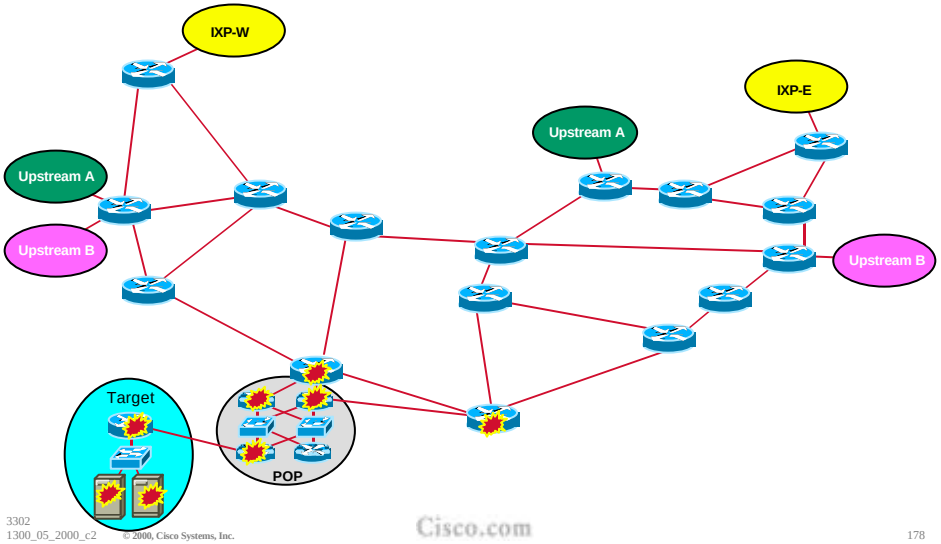


Tracking DoS/DDoS Attacks through an ISP's Network

Tracking Attacks - ISP POV

- **Situation in the NOC**
  - ✓ Alarms go off in the NOC - circuits are dropping packets.
  - ✓ Major Content Customer calls - their site is being hit by a DoS/DDoS Attack
  - ✓ Management calls, they want to know what is going on.
  - ✓ Other customers call, slow network performance.
  - ✓ Reporter calls - not sure how they got the NOC's number, they are looking for a quote
  - ✓ It's been 5 minutes since the first alarm went off, what do you do?!?!?!?!

The Network

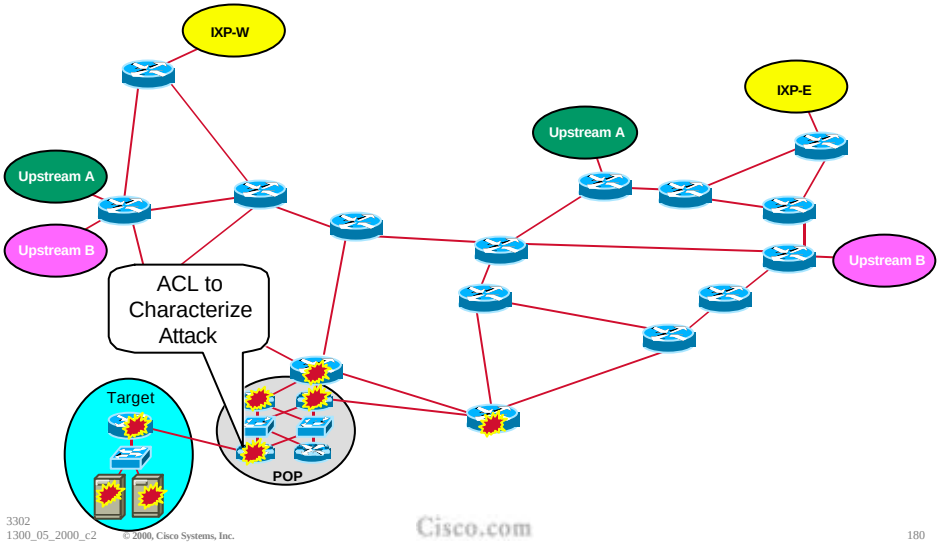


Step 1 - Classifying the Attack

- **Use ACL to find out the characteristics of the attack.**

```
access-list 169 permit icmp any any echo
access-list 169 permit icmp any any echo-reply
access-list 169 permit udp any any eq echo
access-list 169 permit udp any eq echo any
access-list 169 permit tcp any any established
access-list 169 permit tcp any any
access-list 169 permit ip any any

interface serial 0
ip access-group 169 out
```


Step 1 - Classifying the Attack



Step 1 - Classifying the Attack

- **Use the show access-list 169 to see which protocol is the source of the attack:**

```
Extended IP access list 169
        permit icmp any any echo (2 matches)
        permit icmp any any echo-reply (21374 matches)
        permit udp any any eq echo
        permit udp any eq echo any
        permit tcp any any established (150 matches)
        permit tcp any any (15 matches)
        permit ip any any (45 matches)
```


Step 2 - Capture a Source IP

- **Tracing spoofed source IP addresses are a challenge.**
- **Tracing needs to happen hop by hop.**
- **The first step is to use the ACL "log-input" function to grab a few packets.**
- **Quick in and out is needed to keep the router for overloading with logging interrupts to the CPU.**

Step 2 - Capture a Source IP

- **Preparation**
  - ✓ **Make sure your logging buffer on the router is large.**
  - ✓ **Create the ACL**
  - ✓ **Turn off any notices/logging messages to the console or vty (so you can type the command *no access-group 170***

Step 2 - Capture a Source IP

```
access-list 170 permit icmp any any echo
access-list 170 permit icmp any any echo-reply log-input
access-list 170 permit udp any any eq echo
access-list 170 permit udp any eq echo any
access-list 170 permit tcp any any established
access-list 170 permit tcp any any
access-list 170 permit ip any any

interface serial 0
  ip access-group 170 out
! Wait a short time - (i.e 10 seconds)
  no ip access-group 170 out
```

Step 2 - Capture a Source IP

- **Validate the capture with *show access-list 170*. Make sure it the packets we counted.**
- **Check the log with *show logging* for addresses:**

%SEC-6-IPACCESSLOGDP: list 170 permit icmp 192.168.212.72 (Serial0 *HDLC*) -> 198.133.219.25 (0/0), 1 packet

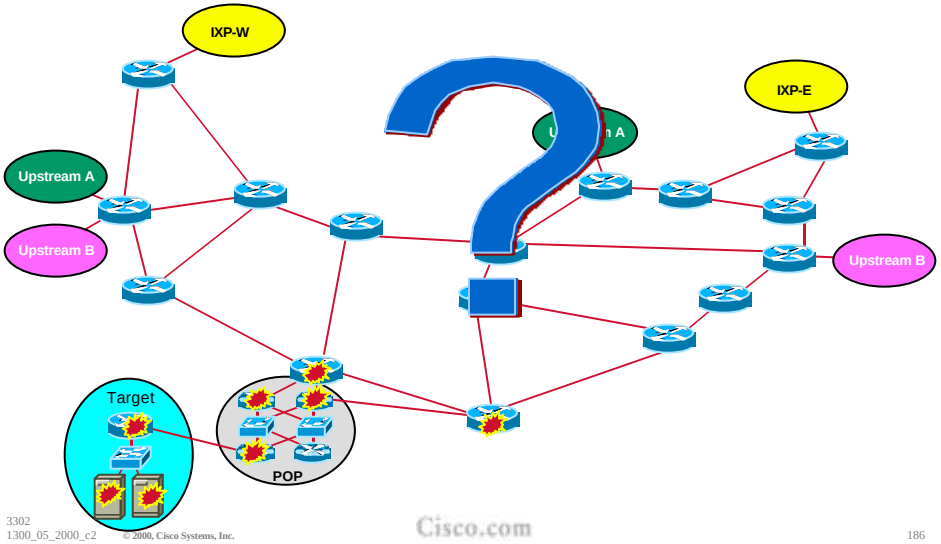
%SEC-6-IPACCESSLOGDP: list 170 permit icmp 172.16.132.154 (Serial0 *HDLC*) -> 198.133.219.25 (0/0), 1 packet

%SEC-6-IPACCESSLOGDP: list 170 permit icmp 192.168.45.15 (Serial0 *HDLC*) -> 198.133.219.25 (0/0), 1 packet

%SEC-6-IPACCESSLOGDP: list 170 permit icmp 192.168.45.142 (Serial0 *HDLC*) -> 198.133.219.25 (0/0), 1 packet

%SEC-6-IPACCESSLOGDP: list 170 permit icmp 172.16.132.47 (Serial0 *HDLC*) -> 198.133.219.25 (0/0), 1 packet

Step 3 - Tracing the Source



Step 3 - Tracing the Source

- **Two Techniques**
  - ✓ **Apply temporary ACLs with *log-input* and examine the logs (like step 2).**
  - ✓ **Query Netflow's Flow Table (if *show ip cache-flow* is turned on).**

Step 3 - Tracing the Source

- **Using Netflow for hop-by-hop traceback:**

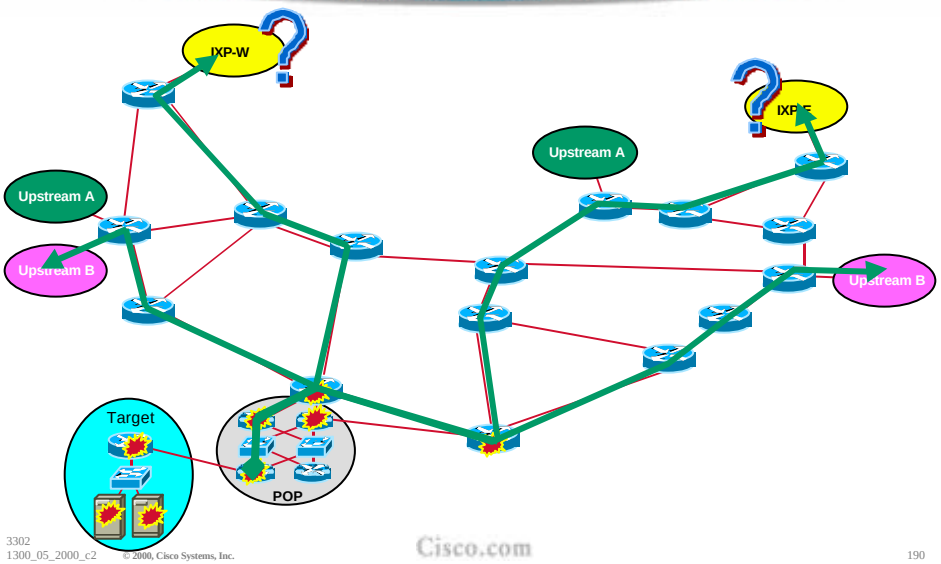
```
Beta-7200-2>sh ip cache 198.133.219.0 255.255.255.0 verbose flow
IP packet size distribution (17093 total packets):
   1-32   64   96  128  160  192  224  256  288  320  352  384  416  448  480
   .000 .735 .088 .054 .000 .000 .008 .046 .054 .000 .009 .000 .000 .000 .000

    512  544  576 1024 1536 2048 2560 3072 3584 4096 4608
   .000 .000 .000 .000 .000 .000 .000 .000 .000 .000 .000

IP Flow Switching Cache, 1257536 bytes
  3 active, 15549 inactive, 12992 added
  210043 ager polls, 0 flow alloc failures
  last clearing of statistics never
Protocol         Total    Flows   Packets Bytes  Packets Active(Sec) Idle(Sec)
--------         Flows     /Sec     /Flow  /Pkt     /Sec     /Flow     /Flow
TCP-Telnet          35      0.0        80    41      0.0      14.5      12.7
UDP-DNS             20      0.0         1    67      0.0       0.0      15.3
UDP-NTP           1223      0.0         1    76      0.0       0.0      15.5
UDP-other        11709      0.0         1    87      0.0       0.1      15.5
ICMP                 2      0.0         1    56      0.0       0.0      15.2
Total:           12989      0.0         1    78      0.0       0.1      15.4

SrcIf         SrcIPaddress    DstIf         DstIPaddress    Pr SrcP DstP  Pkts
Fa1/1         192.168.45.142  POS1/0        198.133.219.25  11 008A 008A     1
Fa1/1         192.168.45.113  POS1/0        198.133.219.25  11 0208 0208     1
Fa1/1         172.16.132.154  POS1/0        198.133.219.25  06 701D 0017    63
```

Step 3 - Tracing the Source

- **Ways to use the Netflow Command:**
  - ✓ **show ip cache <addr> <mask> verbose flow**
  - ✓ **show ip cache flow | include <addr>**
  - ✓ **Proactive approach - create scripts ……**

    ssh -x -t -c [des|3des] -l <username> <IPAddr> "show ip cache <addr> <mask> verbose flow"

Step 3 - Tracing the Source

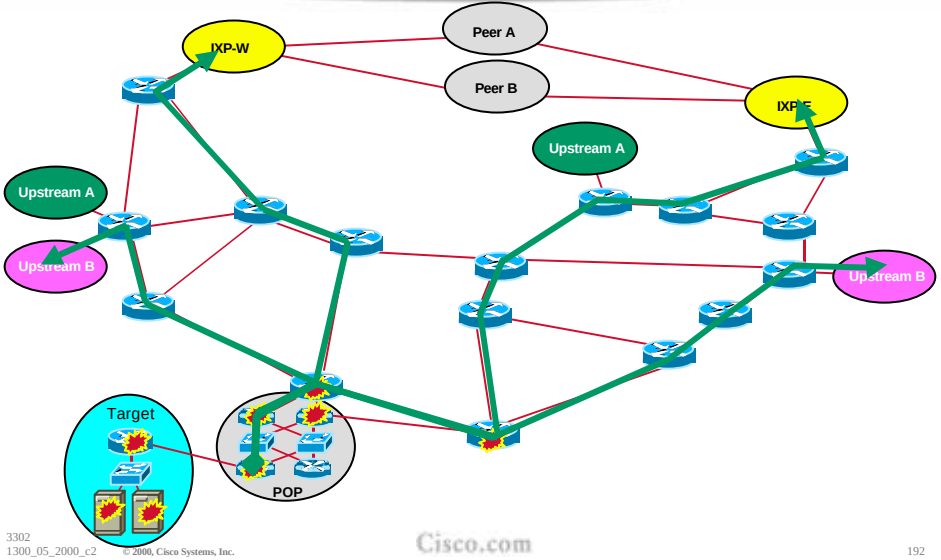


Step 3 - Tracing the Source

- **Tracing across a shared access medium (I.e. like IXPs) require that ACL technique.**

```
May 23 4:30:04.379: %SEC-6-IPACCESSLOGP: list 170 permitted
  icmp 192.168.45.142(0)(FastEthernet3/0/0 00d0.bc83.58a0)
  -> 198.133.219.25 (0), 1 packet
May 23 4:30:05.379: %SEC-6-IPACCESSLOGP: list 170 permitted
  icmp 192.168.45.142(0)(FastEthernet3/0/0 00d0.bc83.58a0)
  -> 198.133.219.25 (0), 1 packet
May 23 4:30:06.379: %SEC-6-IPACCESSLOGP: list 170 permitted
  icmp 192.168.45.142 (0)(FastEthernet3/0/0 00d0.bc83.58a0)
  -> 198.133.219.25 (0), 1 packet
```


Step 3 - Tracing the Source



Troubleshooting Split

- Split in the Security Reaction Team's Flow:
  - ✓ One Team Starts Calling NOCs

    Upstream 2, Peer A, & Peer B
  - ✓ Other Team Drops Filters in to push the packet drops to the edge of the network.

Step 4 - Pushing the Packet Drops to the Edge

- Options:
  - ✓ Rate Limit the attack with CAR (input feature)
  - ✓ ACL to Drop the packets
  - ✓ uRPF (perhaps)
  - ✓ Drop the connection to the peer/upstream

Step 4 - Pushing the Packet Drops to the Edge

- **Select Rate Limiting Option. Limit ICMP Echo-Reply for everyone and limit the Peer's traffic.**

interface FastEthernet3/0/0

rate-limit output access-group 2020 256000 16000 24000 conform-action transmit exceed-action drop

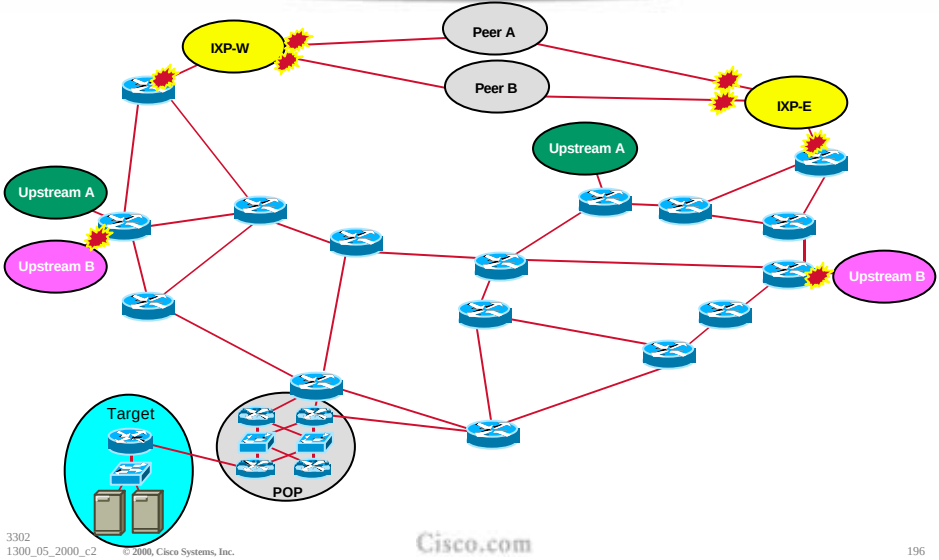
rate-limit input access-group rate-limit 100 8000000 64000 80000 conform-action transmit exceed-action drop

!

access-list 2020 permit icmp any any echo-reply

access-list rate-limit 100 00d0.bc83.58a0

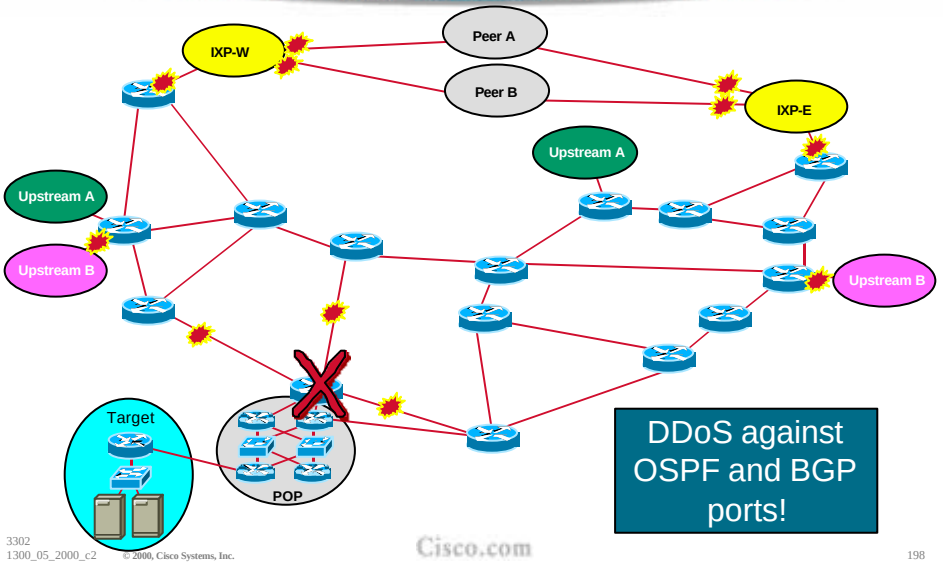
Step 4 - Pushing the Packet Drops to the Edge



Check Point

- SitRep - Attack Still in progress - packets being dropped at the ISP Edge
- Work with Upstream and Peer ISP NOCs to continue the trace back to the sources
- Collect Evidence - work with customer and call your legal team

Alert!



Next Phase of the Attack

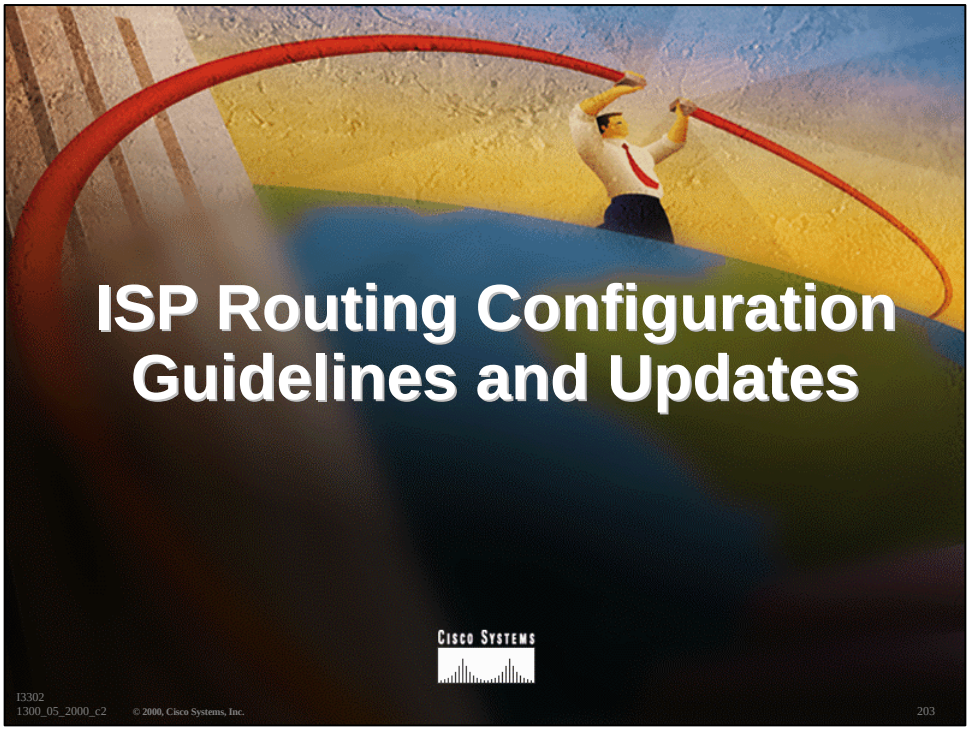
- **The attackers have shifted the attack to their target's infrastructure.**
  - ✓ **ISPs and IXPs <u>have and will be</u> directly attacked to get at the target!**

In case you wondering ...

- How to work a DoS attack against the routing protocol?
  - ✓ Out of Band Access to the Router!
  - ✓ Rate Limits on traffic to the routing protocol
  - ✓ ACLs to block outside traffic to the routing protocol ports

DDoS Links

- http://www.denialinfo.com/
- http://www.staff.washington.edu/dittrich
- http://www.fbi.gov/nipc/trinoo.htm
- http://www.sans.org/y2k/DDoS.htm
- http://www.nanog.org/mtg-9910/robert.html
- http://cve.mitre.org/
- http://packetstorm.securify.com/distributed/



Agenda

- General ISP Routing Principles and Features
- OSPF Best Practices and Updates
- BGP Best Practices and Updates

ISP Routing - Quick Review

What Is an IGP?

- Interior Gateway Protocol
- Within an Autonomous System
- Carries information about internal infrastructure prefixes
- Examples - OSPF, ISIS, EIGRP…

Why Do We Need an IGP?

- ISP backbone scaling
  - ✓ Hierarchy
  - ✓ Modular infrastructure construction
  - ✓ Limiting scope of failure
  - ✓ Healing of infrastructure faults using dynamic routing with fast convergence

What Is an EGP?

- Exterior Gateway Protocol
- Used to convey routing information between Autonomous Systems
- De-coupled from the IGP
- Current EGP is BGP

Why Do We Need an EGP?

- Scaling to large network
  - ✓ Hierarchy
  - ✓ Limit scope of failure
- Policy
  - ✓ Control reachability to prefixes
  - ✓ Merge separate organizations
  - ✓ Connect multiple IGPs

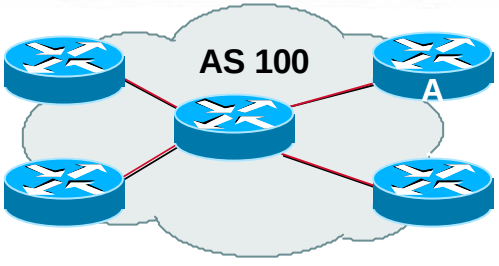
Interior versus Exterior Routing Protocols

- Interior
  - ✓ automatic neighbour discovery
  - ✓ generally trust your IGP routers
  - ✓ prefixes go to all IGP routers
  - ✓ binds routers in one AS together
- Exterior
  - ✓ specifically configured peers
  - ✓ connecting with outside networks
  - ✓ set administrative boundaries
  - ✓ binds AS's together

Interior versus Exterior Routing Protocols

- **Interior**
  - ✓ Carries ISP infrastructure addresses only
  - ✓ ISPs aim to keep the IGP small for efficiency and scalability
- **Exterior**
  - ✓ Carries customer prefixes
  - ✓ Carries Internet prefixes
  - ✓ EGPs are independent of ISP network topology

Autonomous System (AS)

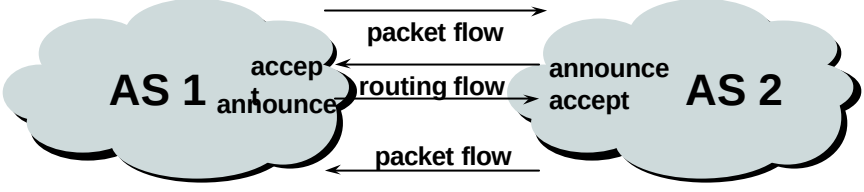


- **Collection of networks with same routing policy**
- **Single routing protocol**
- **Usually under single ownership, trust and administrative control**

Definition of terms

- **Neighbours - AS's which directly exchange routing information**
- **Announce - send routing information to a neighbour**
- **Accept - receive and use routing information sent by a neighbour**
- **Originate - insert routing information into external announcements (usually as a result of the IGP)**
- **Peers - routers in neighbouring AS's or within one AS which exchange routing and policy information**

Routing flow and packet flow



For networks in AS1 and AS2 to communicate:

AS1 must announce to AS2

AS2 must accept from AS1

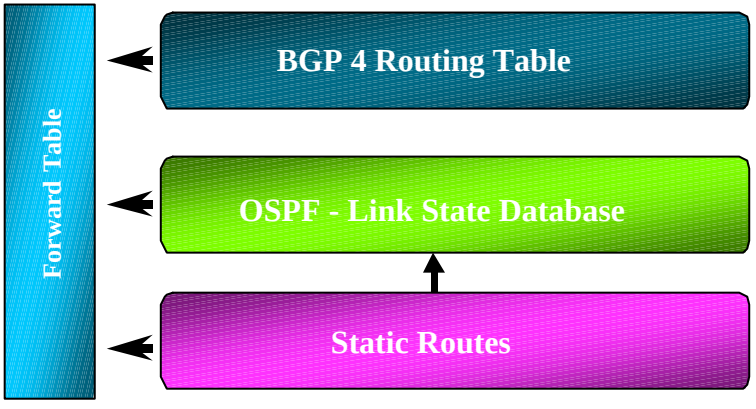
AS2 must announce to AS1

AS1 must accept from AS2

Routing flow and Traffic flow

- Traffic flow is always in the opposite direction of the flow of routing information
  - ✓ filtering outgoing routing information inhibits traffic flowing in
  - ✓ filtering incoming routing information inhibits traffic flowing out

Routing Tables Feed the Forwarding Table

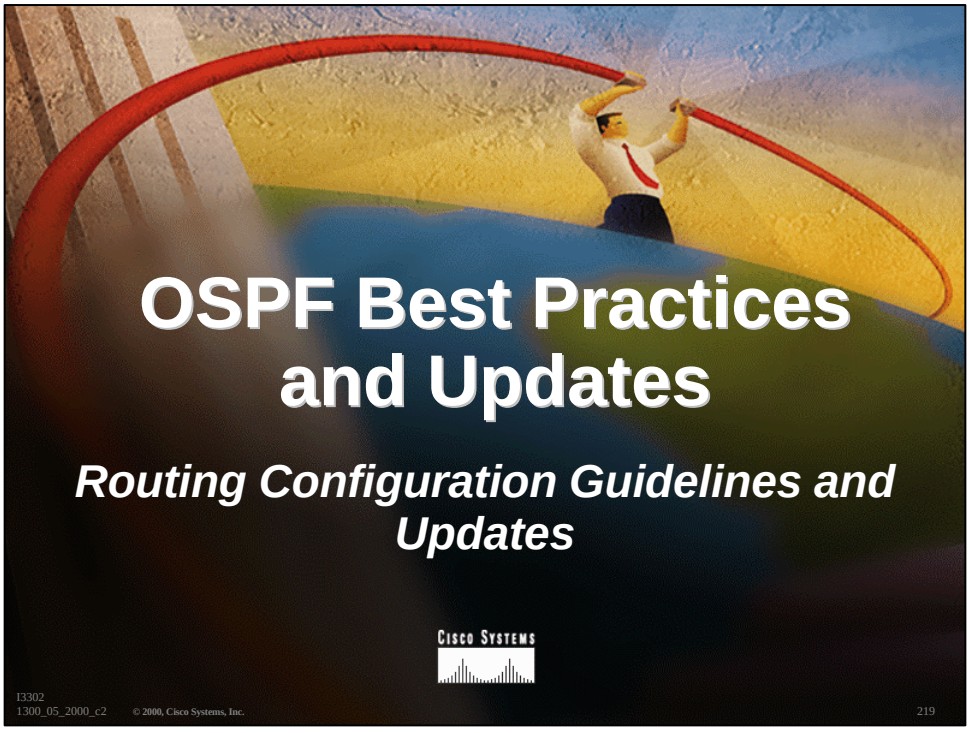


Default Administrative Distances

Route Source	Default Distance
Connected Interface	0
Static Route	1
Enhanced IGRP Summary Route	5
External BGP	20
Internal Enhanced IGRP	90
IGRP	100
OSPF	110
IS-IS	115
RIP	120
EGP	140
External Enhanced IGRP	170
Internal BGP	200
Unknown	255

CIDR Features

- The Internet is a classless world. All routers connect to the Internet must be CIDR compliant, else there will be problems with the network connection to the Internet.
- All Cisco routers should have the following commands configured for CIDR:
  - ✓ `ip subnet-zero`
  - ✓ `ip classless`
- These are default from IOS 12.0 onwards



OSPF Best Practices and Updates

Routing Configuration Guidelines and Updates

Cisco Systems

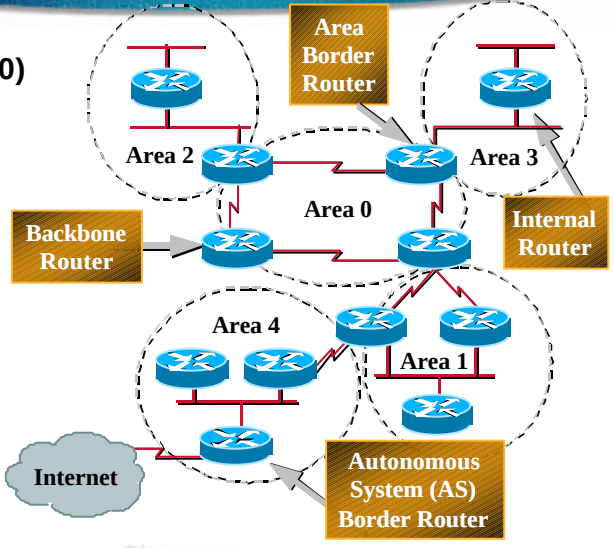
OSPF Quick Review

OSPF

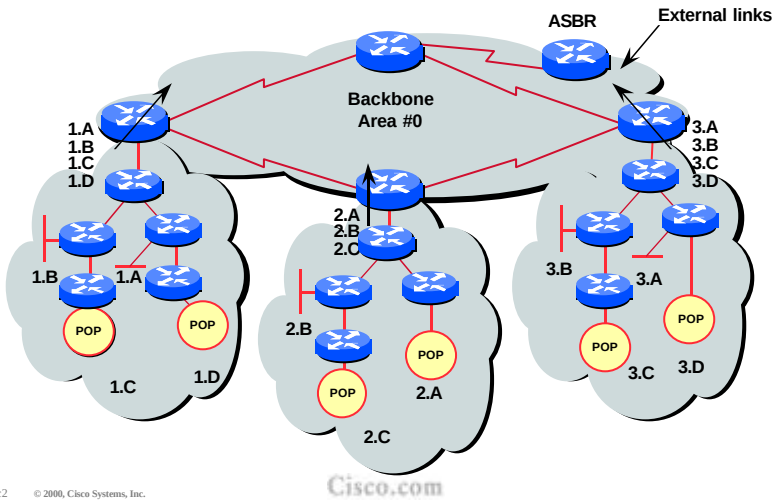
- Open Shortest Path First
- Link state or SPF technology
- Developed by OSPF working group of IETF (RFC 2328 - STD54)
- Designed expressly for TCP/IP Internet environment
- Fast convergence
- Variable-length subnet masks
- Discontiguous subnets
- No periodic updates
- Route authentication
- Delivered two years after IGRP

OSPF Areas and Rules

- Backbone area (0) must be present
- All other areas must have connection to backbone
- Backbone must be contiguous
- Do not partition area (0)



OSPF Hierarchy



OSPF Design

- **Attack addressing first - OSPF and Addressing go together.**
  - ✓ **Objective is to keep the Link Stata Data Base *lean.***
  - ✓ **Create address hierarchy to match topology**
  - ✓ **Separate Blocks for infrastructure, customer interfaces, customers, etc.**

OSPF Design

- Examine physical topology
  - ✓ Is it meshed or hub-and-spoke?
- Try to use as Stubby an area as possible
  - ✓ It reduces overhead and LSA counts
- Push the creation of a backbone
  - ✓ Reduces mesh and promotes hierarchy

OSPF Design

- One SPF per area, flooding done per area
  - ✓ Watch out for overloading ABRs
- Different types of areas do different flooding
  - ✓ Normal areas
  - ✓ Stub areas
  - ✓ Totally stubby (stub no-summary)
  - ✓ Not so stubby areas (NSSA)

OSPF Design

- Redundancy
  - ✓ Dual Links out of each area - using metrics (cost) for traffic enginering
  - ✓ Too much redundancy…

    Dual links to backbone in stub areas must be the same - other wise sub-optimal routing will result

    Too Redundancy in the backbone area without good summarization will effect convergence in the area 0

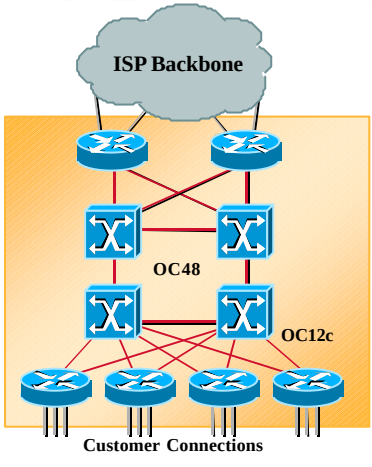
OSPF for ISPs

- OSPF features should consider.
  - ✓ OSPF logging neighbour changes
  - ✓ OSPF reference cost
  - ✓ OSPF Router ID Command
  - ✓ OSPF Process Clear/Restart

OSPF BCP
Adding Networks

OSPF - Adding Networks

- **BCP - Individual OSPF Network statement for each infrastructure link.**
  - ✓ **Have separate IP address blocks for *infrastructure* and *customer links.***
  - ✓ **Use *IP Unnumbered* Interfaces or BGP to carry /30s to customers**
  - ✓ **OSPF should only carry infrastructure routes in an ISP's network.**



OSPF - Adding Networks

- **Three Techniques**
  - ✓ **redistributed connect subnets**

    Works for all interfaces on the router but sends networks a E2s - which are not summarized.

```
router ospf 100
 redistributed connected subnets
```


OSPF - Adding Networks

- **Three Techniques (cont.)**
  - ✓ **network statements - specific**

    Every interface needs a OSPF network statement. Interface that should not be broadcasting OSPF Hello packets need *ospf passive-interface.*

```
Router ospf 100
 network 192.168.1.4 0.0.0.3 area 51
 network 192.168.1.6 0.0.0.3 area 51
 passive interface Serial 1/0/1.2
```

OSPF - Adding Networks

- **Three Techniques (cont.)**
  - ✓ **network statements - wildcard mask**

    Every interface needs a OSPF network statement. Interface that should not be broadcasting OSPF Hello packets need *ospf passive-interface* or *default passive-interface.*

```
Router ospf 100
 network 192.168.1.0 0.0.0.255 area 51
 default passive-interface default
 no passive interface POS 4/0
```


OSPF - Adding Networks

- **Key Theme when selecting a technique: Keep the Link State Database Lean**
  - ✓ **Increases Stability**
  - ✓ **Reduces the amount of information in the Link State Advertisements (LSAs)**
  - ✓ **Speeds Convergence Time**

OSPF - New and Useful Features

OSPF Logging Neighbour Changes

- The router will generate a log message whenever an OSPF neighbour changes state
- Syntax:

  [no] ospf log-adjacency-changes

- Example of a typical log message:

  %OSPF-5-ADJCHG: Process 1, Nbr 223.127.255.223 on Ethernet0 from LOADING to FULL, Loading Done

Number of State Changes

- **The number of state transitions is available via SNMP (ospfNbrEvents) and the CLI:**
  - ✓ **show ip ospf neighbor [type number] [neighbor-id] [detail]**
    - **Detail—(Optional) Displays all neighbours given in detail (list all neighbours). When specified, neighbour state transition counters are displayed per interface or neighbour ID**

State Changes (Continued)

- **To reset OSPF-related statistics, use the clear ip ospf counters EXEC command. At this point neighbor is the only available option; it will reset neighbour state transition counters per interface or neighbour id**
  - ✓ **clear ip ospf counters [neighbor [<type number>] [neighbor-id]]**

OSPF Cost: Reference Bandwidth

- Bandwidth used in Metric calculation
  - ✓Cost = 10^8/BW
  - ✓Not useful for BW > 100 Mbps
- Syntax:
  - ✓ospf auto-cost reference-bandwidth <reference-bandwidth>
- Default reference bandwidth still 100 Mbps for backward compatibility

OSPF Router ID

- If the loopback interface exists and has an IP address, that is used as the router ID in routing protocols - stability!
- If the loopback interface does not exist, or has no IP address, the router ID is the highest IP address configured - danger!
- New sub command to manually set the OSPF Router ID:

```
router-id <ip address>
```

OSPF Clear/Restart

- **clear ip ospf [pid] redistribution**

  This command can now clear redistribution based on OSPF routing process ID. If no pid is given, it assumes all OSPF processes.

- **clear ip ospf [pid] counters**

  This command can now clear counters based on OSPF routing process ID. If no pid is given, it assumes all OSPF processes.

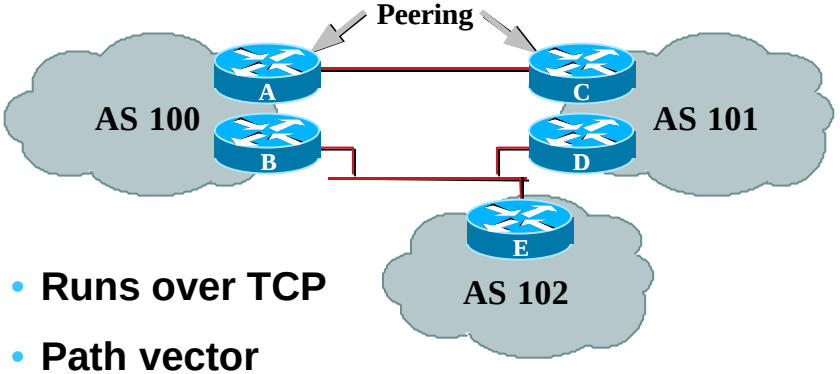
- **clear ip ospf [pid] process**

  This command will restart the specified OSPF process. If no pid is given, it assumes all OSPF processes. It attempts to keep the old router-id, except in cases, where a new router-id was configured, or an old user configured router-id was removed. Since this command can potentially cause a network churn, a user confirmation is required before performing any action.

BGP

- RFC 1771
- Border Gateway Protocol
- Version 4 is current
- Exterior routing protocol (vs. interior)
- Uses TCP for transport
- Many options for policy enforcement
- Classless Inter Domain Routing (CIDR)
- Widely used for Internet backbone
- Autonomous systems

BGP Basics

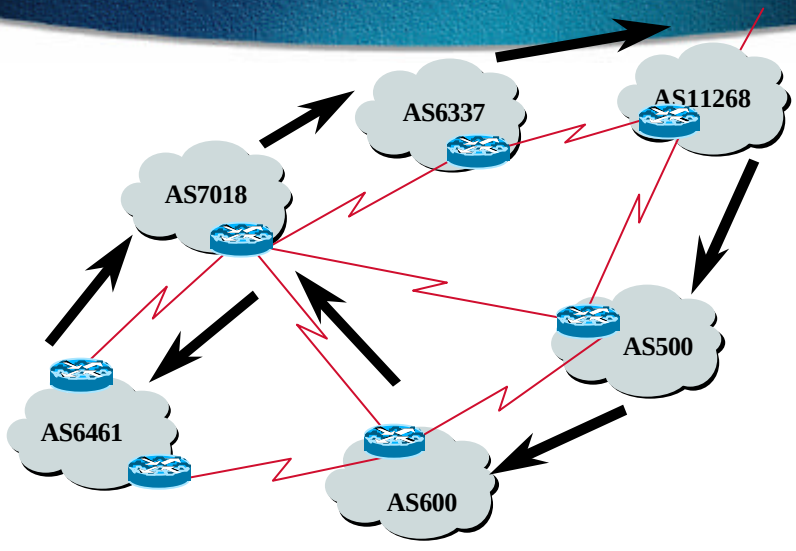


- Runs over TCP
- Path vector protocol
- Incremental update

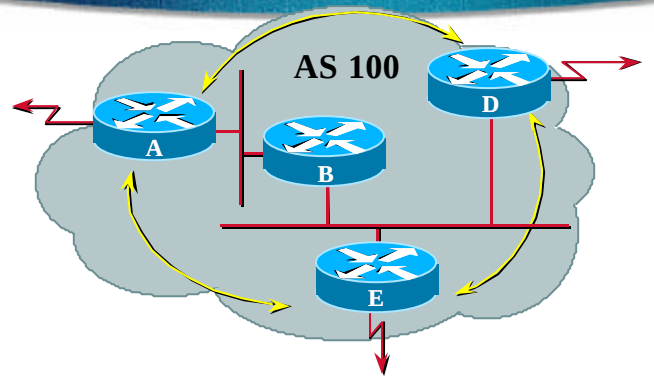
Path Vector Protocol

- **BGP is classified as a *path vector* routing protocol** (see RFC 1322)
  - ✓ **A path vector protocol defines a route as a pairing between a destination and the attributes of the path to that destination.**

Path Vector Protocol

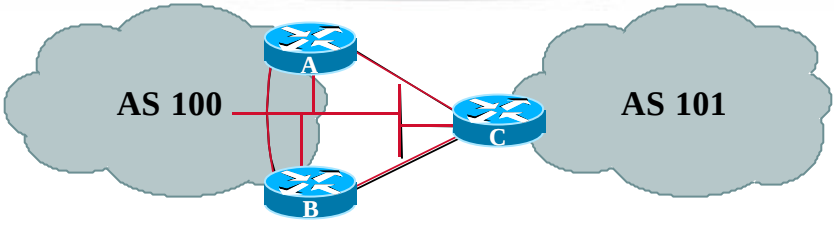


Internal BGP (iBGP) Peering



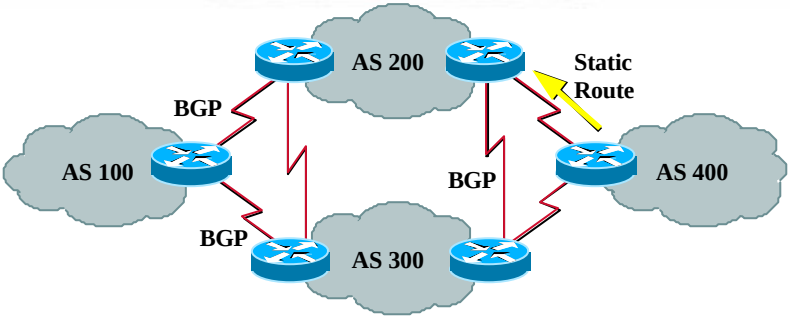
- BGP peer within the same AS
- Not required to be directly connected
- iBGP neighbors should be fully meshed
- Few BGP speakers in corporate network

External BGP (eBGP) Peering



- Between BGP speakers in different AS
- Should be directly connected
- Don’t run an IGP between eBGP peers

Policy Drives BGP Requirements



- Policy for AS 100: Always use AS 300 path to reach AS 400

BGP versus OSPF/ISIS

- Internal Routing Protocols (IGPs)
  - ✓ examples are ISIS and OSPF
  - ✓ used for carrying infrastructure addresses
  - ✓ NOT used for carrying Internet prefixes or customer prefixes

BGP versus OSPF/ISIS

- BGP used internally (iBGP) and externally (eBGP)
- iBGP used to carry
  - ✓ some/all Internet prefixes across backbone
  - ✓ customer prefixes
- eBGP used to
  - ✓ exchange prefixes with other ASes
  - ✓ implement routing policy

BGP versus OSPF/ISIS

- DO NOT:
  - ✓ distribute BGP prefixes into an IGP
  - ✓ distribute IGP routes into BGP
  - ✓ use an IGP to carry customer prefixes
- YOUR NETWORK WILL NOT SCALE

BGP

- There are key BGP features that should be configured by ISPs:
  - ✓ `update-source loopback 0`
  - ✓ `ip bgp-community new-format`
  - ✓ `no synchronization`
  - ✓ `bgp dampening`
  - ✓ `no auto-summary`
  - ✓ `bgp neighbor authentication`
  - ✓ `bgp neighbor maximum-prefix`

iBGP configuration

- Use loopback interface
  - ✓ it never goes away
  - ✓ routers have multiple external paths
  - ✓ has multiple uses

```
interface loopback 0
 ip address 215.17.1.34 255.255.255.255
router bgp 200
 neighbor 215.17.1.35 remote-as 200
 neighbor update-source loopback 0
 neighbor 215.17.1.36 remote-as 200
 neighbor update-source loopback 0
```


BGP Community Format

- Communities are used extensively
- Cisco IOS supports two formats
  - ✓ One 32 bit integer eg 13107210
  - ✓ Two 16 bit integers eg 200:10
- RFC1998 recommends 16:16 format
  - ✓ Format AS:xxxx
  - ✓ `ip bgp-community new-format`

BGP Synchronization

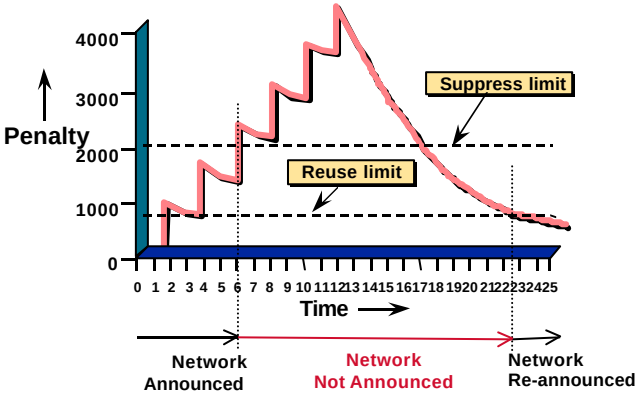
- **BGP does not advertise a route before all routers in the AS have learned it via an IGP**
- **Disable synchronization if:**
  - ✓ AS doesn't pass traffic from one AS to another
  - ✓ All transit routers in AS run BGP
  - ✓ iBGP is used across backbone
  - ✓ `no synchronization`

BGP Neighbour Shutdown

- **Shutdown BGP peering**
  - ✓ **previously required to delete configuration**
  - ✓ **now can simply "shutdown" the peering**
- **Configuration example:**
  - ✓ `router bgp 200`
  - ✓ `neighbor 215.7.1.1 remote-as 210`
  - ✓ `neighbor 215.7.1.1 shutdown`
- **Can be reactivated with**
  - ✓ `no neighbor 215.7.1.1 shutdown`

BGP Dampening

- **Route flap dampening to minimise instability in local network and Internet**



BGP Dampening

- **Recommended values and sample configurations for ISPs at:**
  - ✓ http://www.ripe.net/docs/ripe-178.html
- **Example techniques:**
  - ✓ Internet Routing Architecture - Bassam Halabi
  - ✓ **`bgp dampening`**

BGP Auto Summarisation

- Automatically summarises subprefixes to the classful network.
- Must be turned off for any Internet connected site using BGP.
- Internet is classless - class A, class B and class C are no more.

✓ `no auto-summary`

BGP Neighbour Authentication

- MD5 authentication between two peers
  - ✓ password must be known to both peers
- peer-group can be used to apply to multiple peerings

✓ `neighbor 169.222.10.1 password v61ne0qkel33&`

BGP Maximum Prefix Tracking

- Allow configuration of the maximum number of prefixes a BGP router will receive from a peer
- Two level control
  - ✓ Warning threshold: log warning message
  - ✓ Maximum: tear down the BGP peering, manual intervention required to restart

BGP Maximum Prefix Tracking

`neighbor <x.x.x.x> maximum-prefix <max> [<threshold>] [warning-only]`

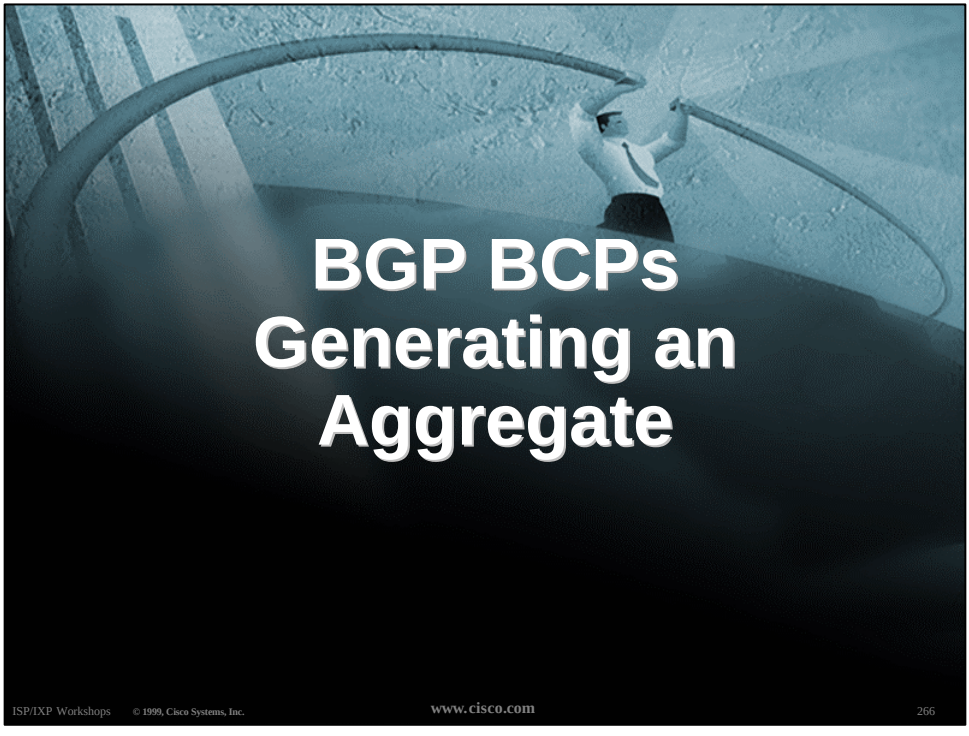
- Threshold is an optional parameter between 1 to 100 percent
  - ✓ Specify the percentage of <max> that a warning message will be generated. Default is 75%.
- Warning-only is an optional keyword which allows log messages to be generated but peering session will not be torn down

BGP Maximum Prefix Tracking

- **Sample logs:**
  - ✓ **The number of prefixes received from a peer reaches 75% of the maximum configured:**

    %BGP-4-MAXPFX: No. of prefix received from 44.1.1.2 reaches 3, max 4
  - ✓ **The number of prefix exceeds the maximum number of prefixes configured:**

    %BGP-3-MAXPFXEXCEED: No. of prefix received from 44.1.1.2: 4 exceed limit 3



Aggregation

- ISPs receive address block from Regional Registry or upstream provider
- Aggregation means announcing the address block only, not subprefixes
- Aggregate should be generated internally

Configuring Aggregation - Cisco IOS

- ISP has 221.10.0.0/19 address block
- To put into BGP as an aggregate:
  - `router bgp 100`
  - `network 221.10.0.0 mask 255.255.224.0`
  - `ip route 221.10.0.0 255.255.224.0 null0 250`
- The static route is a “pull up” route
  - ✓ more specific prefixes within this address block ensure connectivity to ISP’s customers
  - ✓ “longest match lookup”

BGP BCPs Announcing Aggregate

Aggregation

- Address block should be announced to the Internet as an aggregate
- Subprefixes of address block should NOT be announced to Internet unless special circumstances (more later)

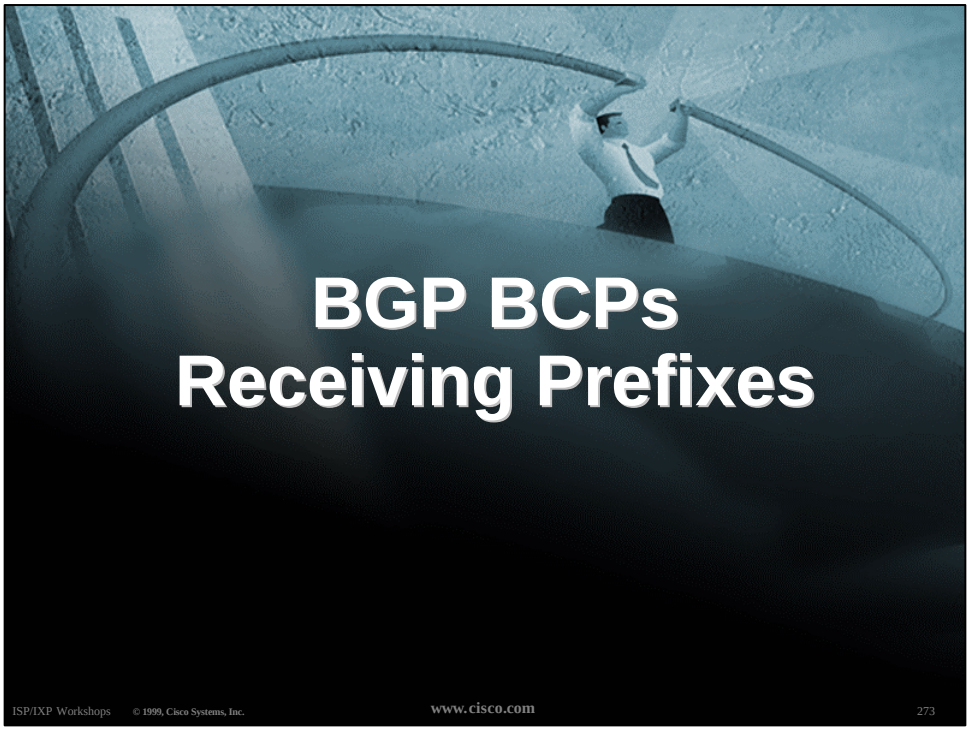
Announcing Aggregate - Cisco IOS

- **Configuration Example**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 222.222.10.1 remote-as 101
 neighbor 222.222.10.1 prefix-list out-filter out
!
ip route 221.10.0.0 255.255.224.0 null0
!
ip prefix-list out-filter permit 221.10.0.0/19
ip prefix-list out-filter deny 0.0.0.0/0 le 32
```


Announcing an Aggregate

- **ISPs who don't and won't aggregate are held in poor regard by community**
- **Registries minimum allocation sizes are /19s or /20s now**
  - ✓ **no real reason to see anything longer than a /21 or /22 prefix in the Internet**
  - ✓ **BUT there are currently >44000 /24s!**



Receiving Prefixes from downstream peers

- ISPs should only accept prefixes which have been assigned or allocated to their downstream peer
- For example
  - ✓ downstream has 220.50.0.0/20 block
  - ✓ should only announce this to peers
  - ✓ peers should only accept this from them

Receiving Prefixes - Cisco IOS

- **Configuration Example on upstream**

```
router bgp 100
 neighbor 222.222.10.1 remote-as 101
 neighbor 222.222.10.1 prefix-list customer in
!
ip prefix-list customer permit 220.50.0.0/20
ip prefix-list customer deny 0.0.0.0/0 le 32
```


Receiving Prefixes from upstream peers

- **Not desirable unless really necessary**
  - ✓ **special circumstances**
- **Ask upstream to either:**
  - ✓ **originate a default-route**
  - ✓ **announce one prefix you can use as default**

Receiving Prefixes from upstream peers

- **Downstream Router Configuration**

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 221.5.7.1 remote-as 101
 neighbor 221.5.7.1 prefix-list infilt in
 neighbor 221.5.7.1 prefix-list outfilt out
!
ip prefix-list infilt permit 0.0.0.0/0
ip prefix-list infilt deny 0.0.0.0/0 le 32
!
ip prefix-list outfilt permit 221.10.0.0/19
ip prefix-list outfilt deny 0.0.0.0/0 le 32
```


Receiving Prefixes from upstream peers

- **Upstream Router Configuration**

```
router bgp 101
 neighbor 221.5.7.2 remote-as 100
 neighbor 221.5.7.2 default-originate
 neighbor 221.5.7.2 prefix-list cust-in in
 neighbor 221.5.7.2 prefix-list cust-out out
!
ip prefix-list cust-in permit 221.10.0.0/19
ip prefix-list cust-in deny 0.0.0.0/0 le 32
!
ip prefix-list cust-out permit 0.0.0.0/0
ip prefix-list cust-out deny 0.0.0.0/0 le 32
```

Receiving Prefixes from upstream peers

- If necessary to receive prefixes from upstream provider, care is required
  - ✓ don't accept RFC1918 etc prefixes
  - ✓ don't accept your own prefix
  - ✓ don't accept default (unless you need it)
  - ✓ don't accept prefixes longer than /24

Receiving Prefixes - Cisco IOS

```
router bgp 100
 network 221.10.0.0 mask 255.255.224.0
 neighbor 221.5.7.1 remote-as 101
 neighbor 221.5.7.1 prefix-list in-filter in
!
ip prefix-list in-filter deny 0.0.0.0/0            ! Block default
ip prefix-list in-filter deny 0.0.0.0/8 le 32
ip prefix-list in-filter deny 10.0.0.0/8 le 32
ip prefix-list in-filter deny 127.0.0.0/8 le 32
ip prefix-list in-filter deny 169.254.0.0/16 le 32
ip prefix-list in-filter deny 172.16.0.0/12 le 32
ip prefix-list in-filter deny 192.0.2.0/24 le 32
ip prefix-list in-filter deny 192.168.0.0/16 le 32
ip prefix-list in-filter deny 221.10.0.0/19 le 32  ! Block local prefix
ip prefix-list in-filter deny 224.0.0.0/3 le 32
ip prefix-list in-filter deny 0.0.0.0/0 ge 25              ! Block prefixes >/24
ip prefix-list in-filter permit 0.0.0.0/0 le 32
```

"Documenting Special Use Addresses" - DSUA

- **This prefix-list MUST be applied to all external BGP peerings, in and out!**
  - ✓ **http://www.ietf.org/internet-drafts/draft-manning-dsua-01.txt**

```
ip prefix-list rfc1918-dsua deny 0.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny 10.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny 127.0.0.0/8 le 32
ip prefix-list rfc1918-dsua deny 169.254.0.0/16 le 32
ip prefix-list rfc1918-dsua deny 172.16.0.0/12 le 32
ip prefix-list rfc1918-dsua deny 192.0.2.0/24 le 32
ip prefix-list rfc1918-dsua deny 192.168.0.0/16 le 32
ip prefix-list rfc1918-dsua deny 224.0.0.0/3 le 32
ip prefix-list rfc1918-dsua deny 0.0.0.0/0 ge 25
ip prefix-list rfc1918-dsua permit 0.0.0.0/0 le 32
```


Injecting prefixes into iBGP

- Use iBGP to carry customer prefixes
  - ✓ don't use IGP
- Point static route to customer interface
- Use BGP network statement
- As long as static route exists (interface active), prefix will be in BGP

Router Configuration

- Example:

```
interface loopback 0
 ip address 215.17.3.1 255.255.255.255
!
interface Serial 5/0
 ip unnumbered loopback 0
 ip verify unicast reverse-path
!
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 network 215.34.10.0 mask 255.255.252.0
```

Injecting prefixes into iBGP

- 200 network statement limit removed
- interface flap will result in prefix withdraw and reannounce
  - ✓ use “ip route…permanent”
- many ISPs use redistribute static rather than network statement
  - ✓ only use this if you understand why

Router Configuration

- Example:

```
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
 redistribute static route-map static-to-bgp
<snip>
!
route-map static-to-bgp permit 10
 match ip address prefix-list ISP-block
 set origin igp
<snip>
!
ip prefix-list ISP-block permit 215.34.10.0/22 le 30
!
```

Injecting prefixes into iBGP

- Route-map ISP-block can be used for many things:
  - ✓ setting communities and other attributes
  - ✓ setting origin code to IGP, etc
- Be careful with prefix-lists and route-maps
  - ✓ absence of either/both means all statically routed prefixes go into iBGP

BGP

- More helpful features:
  - ✓ BGP Log-Neighbor-Changes
  - ✓ BGP Peer Groups
  - ✓ IP Prefix-Lists
  - ✓ BGP Conditional Advertisement
  - ✓ BGP Policy Propagation
  - ✓ Smooth AS Transistion
  - ✓ Third Party Next-Hop Override
  - ✓ MED Comparison

BGP log-neighbor-changes

- Log neighbour up/down events, and the reason for the last neighbour peering reset
- In 11.1 CC and 12.0 releases
- Syntax (router subcommand):

  [no] log-neighbor-changes

- Typical log messages:
  - ✓ %BGP-6-ADJCHANGE: neighbor x.x.x.x Up
  - ✓ %BGP-6-RESET: neighbor x.x.x.x reset (User reset request)

Reason for Last Peer Reset

- Router keeps reason for the last BGP peer reset for each of its peers. Useful to analyse BGP session resets
- Available as part of the show ip bgp neighbor command output
- Accessible also through SNMP

Current Reset Reasons

- ✓ "Error during connection collision"
- ✓ "Peer closing down the session"
- ✓ "Peer exceeding maximum prefix limit"
- ✓ "Interface flap"
- ✓ "Router ID changed"
- ✓ "Neighbor deleted"
- ✓ "Member added to peergroup"
- ✓ "Administratively shutdown"
- ✓ "Remote AS changed"
- ✓ "RR client configuration modification"
- ✓ "Soft reconfiguration modification"

BGP Peering

- By default, peerings are reset immediately a peer fails to respond
  - → bad for high latency, long distance, or congested links
  - → this is the default action
- IOS option to disable this
  - ✓ `no bgp fast-external-fallover`

BGP peer groups

- Reduces CPU load and memory
  - ✓ update generation processed once
  - ✓ BGP configuration simplified

```
router bgp 109
 neighbor internal peer-group
 neighbor internal remote-as 109
 neighbor internal update-source loopback 0
 neighbor 131.108.10.1 peer-group internal
 neighbor 131.108.20.1 peer-group internal
```

Prefix Lists

- High performing access-list
- Faster loading of large lists
- Incremental configuration
  - ✓ sequence numbers optional
  - ✓ `no ip prefix-list sequence-number`
- Available from 11.1(17)CC and 12.0
- Configured by:
  - ✓ `ip prefix-list <list-name>`

Prefix-list Command

[no] ip prefix-list <*list-name*> [seq <*seq-value*>] deny | permit <*network*>/<*len*> [ge <*ge-value*>] [le <*le-value*>]

<*network*>/<*len*>: The prefix and its length

ge <*ge-value*>: "greater than or equal to"

le <*le-value*>: "less than or equal to"

Both "ge" and "le" are optional. Used to specify the range of the prefix length to be matched for prefixes that are more specific than <*network*>/<*len*>

Prefix Lists - Examples

- Deny default route
  - ✓ `ip prefix-list EG deny 0.0.0.0/0`
- Permit the prefix 35.0.0.0/8
  - ✓ `ip prefix-list EG permit 35.0.0.0/8`
- In 192/8 allow up to /24
  - ✓ `ip prefix-list EG permit 192.0.0.0/8 le 24`
- In 192/8 deny /25 and above
  - ✓ `ip prefix-list EG deny 192.0.0.0/8 ge 25`
- Permit all
  - ✓ `ip prefix-list EG permit 0.0.0.0/0 le 32`

Prefix Lists in BGP

- Prefix-list can be used as alternative to distribute-list
  - ✓ `router bgp 200`
  - ✓ `neighbor 169.222.1.1 remote-as 200`
  - ✓ `neighbor 169.222.1.1 prefix-list FILTER-IN in`
  - ✓ `neighbor 169.222.1.1 prefix-list FILTER-OUT out`
- Prefix-lists and access-lists are mutually exclusive

Prefix-list route-map command

```
route-map <name> permit|deny <seq-num>
  match ip address | prefix-list <name>
  [<name> ...]
```

- Used for route filtering, originating default, and redistribution in other routing protocols as well
- Not for packet filtering

BGP Conditional Advertisement

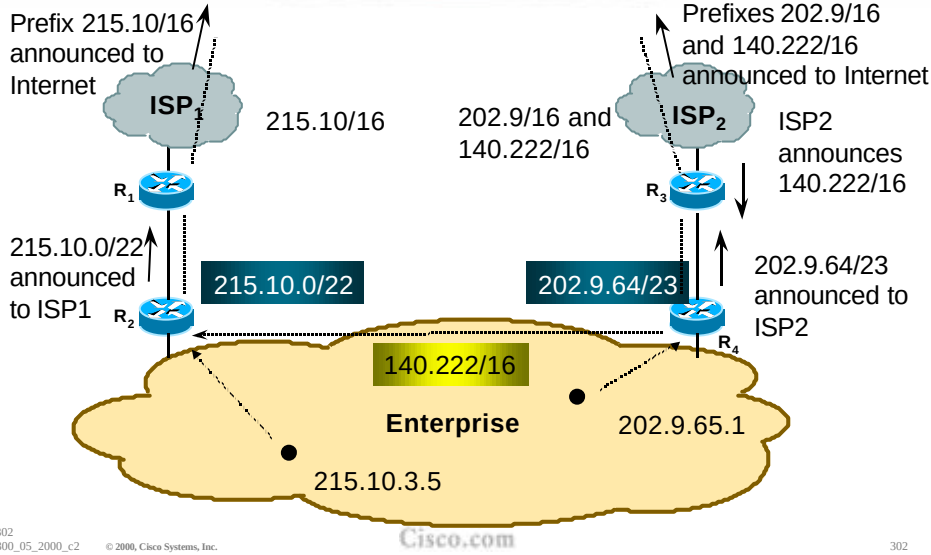
- Reduce the number of prefixes advertised when there is no failure
- Prefix injected when there is a failure to restore connectivity
  - ✓ For multihoming customers or backup scenario
- Help scale the Internet backbone
  - ✓ It is in everybody's best interest…

BGP Conditional Advertisement: configuration

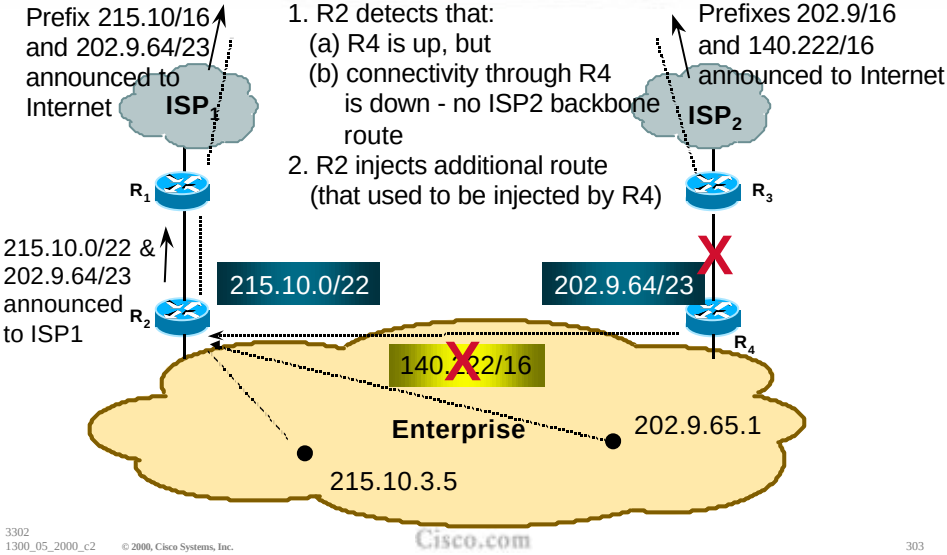
```
neighbor <x.x.x.x> advertise-map <route-map>
  non-exist-map <route-map>
```

- **<*route-map*> is a standard route-map**
- **non-exist-map specifies prefix that BGP speaker will track**
- **advertise-map specifies prefix that will be advertised when prefix in non-exist-map no longer exist**

Example - steady state



Example - link failure



Example Configuration

On router R2:

```
router bgp 100
  neighbor <R1> advertise-map ISP2-subblock non-exist-map ISP2-backbone
route-map ISP2-subblock permit 10
   match ip address prefix-list ISP2-sub       ! <ISP2-subblock-prefix>
route-map ISP2-backbone permit 10
   match ip address prefix-list ISP2-bb        ! <ISP2-backbone-prefix>
ip prefix-list ISP2-sub permit 202.9.64.0/23   ! <ISP2-subblock-prefix>
ip prefix-list ISP2-bb permit 140.222.0.0/16   ! <ISP2-backbone-prefix>
```

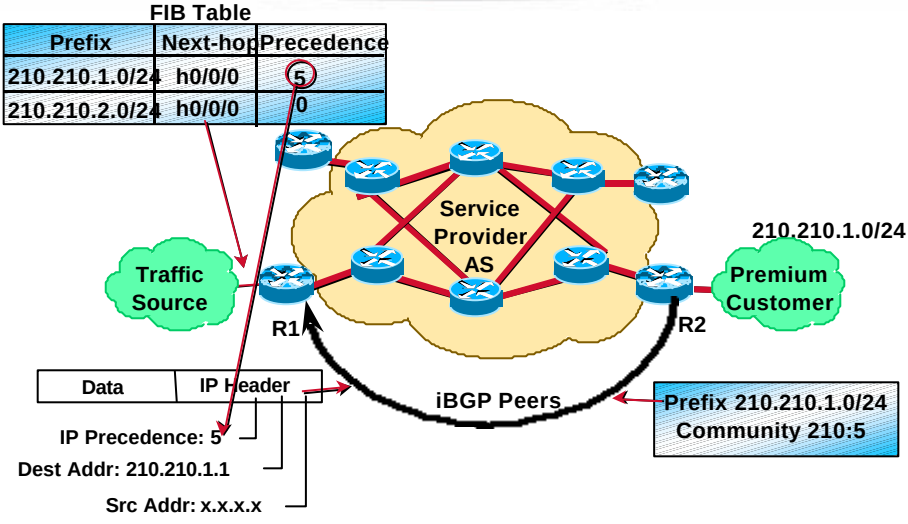
BGP Policy Propagation

- Conveys IP precedence to be used in forwarding to specified destination prefix via BGP community tag
- Allows ingress routers to prioritise incoming traffic
- Also allows IP precedence setting based on AS-path attribute or access list
- Inter-ISP Service Level Agreements (SLAs)

BGP Policy Propagation (Continued)

- Mapping a BGP prefix/community/as-path into a precedence value
  - ✓ It is done when a prefix is added from BGP table into IP routing table
  - ✓ This precedence value is moved from routing table to CEF table
  - ✓ The precedence value in the CEF table is used to set on the packet and executing other functions in the core

BGP Policy Propagation (Continued)



BGP Policy Propagation—Sample Configuration

R2#write term

```
!
router bgp 210
 neighbor 210.210.14.1 remote-as 210
 neighbor 210.210.14.1 route-map comm-relay-prec out
 neighbor 210.210.14.1 send-community
!
ip bgp-community new-format
!
access-list 1 permit 210.210.1.0 0.0.0.255
!
route-map comm-relay-prec permit 10
 match ip address 1
 set community 210:5
!
route-map comm-relay-prec permit 20
 set community 210:0
!
```

BGP Policy Propagation—Sample Configuration

R1#write term

```
!
router bgp 210
 table-map precedence-map
 neighbor 200.200.14.4 remote-as 210
 neighbor 200.200.14.4 update-source Loopback0
!
ip bgp-community new-format
!
ip community-list 1 permit 210:5
!
route-map precedence-map permit 10
 match community 1
 set ip precedence 5
!
route-map precedence-map permit 20
 set ip precedence 0
!
```

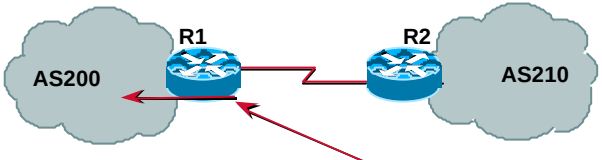

Configuring BGP Policy Propagation

- **Configuring BGP policy propagation**
  - ✓**[no] bgp-policy ip-prec-map**

BGP Policy Propagation—Sample Configuration

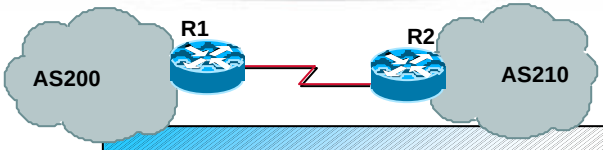
```
!
int hssi0/0/0
 ip address 210.210.2.1 255.255.255.252
 bgp-policy ip-prec-map out
!
```


BGP Policy Propagation Inter-AS



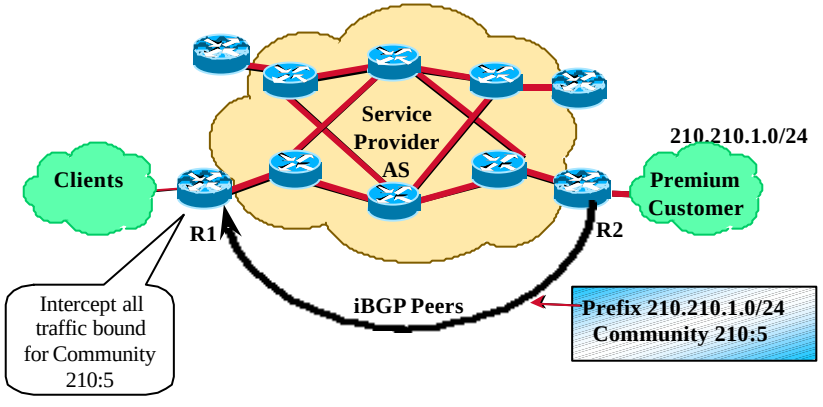
Prefix	Community
210.210.1.0/24	200:5
210.210.2.0/24	200:4
210.210.3.0/24	200:0

BGP Policy Propagation AS-Path

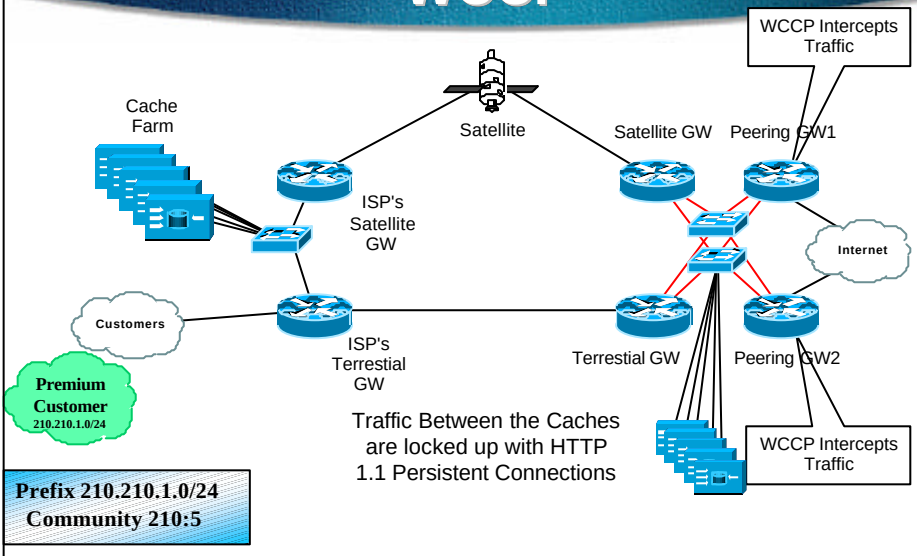



```
!
router bgp 210
 table-map as-path-precedence-map
 neighbor "R1" remote-as 200
!
ip as-path access-list 101 permit
$200^
!
route-map as-path-precedence-map
 match ip as-path 101
 set precedence 3
!
interface hssi/0/0/0
 bgp-policy ip-prec-map
!
```


BGP Policy Propagation for WCCP



BGP Policy Propagation for WCCP



BGP Policy Propagation for WCCP

- **The following example shows only "premium" traffic being cached.**
  - ✓ **"Premium" traffic is defined as traffic which has:**
  - ✓ **The policy defined below is:**
    any traffic with community 4433:1050 set,
    any traffic with community 4433:1055 set,
    any traffic originating from directly-connected AS 65521,
    any traffic passing thru directly-connected AS 65522,
    any traffic passing thru AS 65523
  - ✓ **is eligible for intercept.**
  - ✓ **Standard "web-cache" service is used -- which is a standard assignment of 'match tcp destination port 80', distribute traffic among participating caches as hashed by destination ip address.**

BGP Policy Propagation for WCCP

```
!
ip cef distributed          # ensure Distributed CEF is enabled
!
ip wccp version 2           # enable WCCPv2
ip wccp web-cache password <pass> policy source 50
                            # enable WCCP standard web-cache
                            # service, apply policy "source"-
                            # match on WCCP route-tag 50
!
interface <xyz>                                 # incoming i/face
  ip wccp web-cache redirect in                 # redirect on input traffic
!
ip bgp-community new-format
ip community-list 3 permit 4433:1050            # AS4433 community 1050 is premium
ip community-list 3 permit 4433:1055            # AS4433 community 1055 is premium
!
ip as-path access-list 121 permit ^65521$       # only traffic from AS65521 is premium
ip as-path access-list 121 permit ^65522        # any traffic thru AS65522 premium
!
route-map neighbor-xyz-in permit 10             # incoming route filter on
  match as-path 121
  set ip wccp 50
!
route-map neighbor-xyz-in permit 15
  match community 3
  set ip wccp 50
```


Clear BGP Sessions per AS

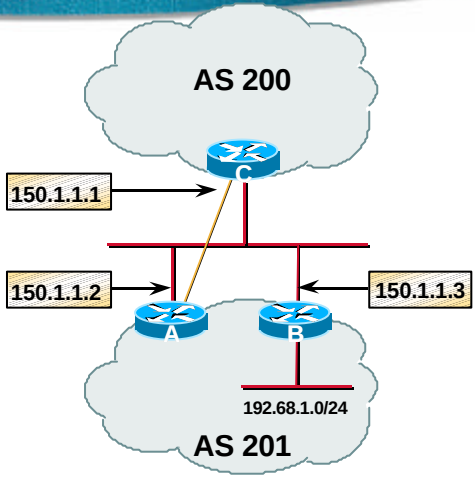
- Ability to clear the BGP sessions of all the neighbors configured with a specific AS number
- Syntax:
  - ✓ clear ip bgp <as number>
- Availability — 11.1(14)CA, 11.1CC, 11.2(9), 11.3(2)

Smooth AS Transition

- Currently, synchronization and coordination between providers/customers is needed when a change in ASN is required
- Syntax:
  - ✓ neighbor x.x.x.x remote-as a or b or c
  - ✓ Accept BGP sessions from any of the listed ASNs
- Availability 11.1CC (later this year), 12.0

Override Third-Party Next-Hop

- Example:
  - ✓ A and B are in the same AS
  - ✓ Router A will advertise 192.68.1.0/24 with a NEXT_HOP of 150.1.1.3.



- More efficient!

Override Third-Party Next-Hop

- Alternative to configuring a specific IP address to be the next-hop for BGP routes
- Syntax (route-map command):

  set ip next-hop peer-address

- Availability 11.1CC, 11.2(12), 11.3(2)

Override Third-Party Next-Hop

- If used in an **inbound** route-map, the next-hop of the received (matching) routes will be set to be the **neighbor peering address**, thus overriding any third-party next-hops. The same route-map can be applied to multiple BGP peers to override third-party next-hops
- If used in an **outbound** route-map, the next-hop of the advertised (matching) routes will be set to be the peering address of the local router, thus disabling the next-hop calculation. This command has finer granularity than the per-neighbor "**next-hop-self**" command

MED Comparison

- Currently, MED is compared ONLY for prefixes with the same AS_PATH
  - ✓ (unless bgp always-compare-med is enabled)
- If the AS_PATH is made up of only confederation sub-ASNs, its length is not considered
  - ✓ also, the MED is not compared
- If an update is received with no MED, the router assigns it a value of 0
  - ✓ may the the preferred path

MED Comparison

- New command allows the user to change the default best path selection algorithm
  - ✓ bgp bestpath med [confed | missing-as-worst | {confed missing-as-worst}]
- Availability
  - ✓ confed 11.1(20)CC, 12.0
  - ✓ missing-as-worst 12.0

Multi-Protocol BGP

- Extension to the BGP protocol in order to carry routing information about other protocols
  - Multicast
  - MPLS
  - IPv6
  - CLNS
  - IPX
  - …
- Exchange of Multi-Protocol NLRI must be negotiated at session set up

  BGP Capabilities negotiation

Multi-Protocol BGP - RFC2283

- New non-transitive and optional BGP attributes
  - MP_REACH_NLRI

    "Carry the set of reachable destinations together with the next-hop information to be used for forwarding to these destinations" (RFC2283)
  - MP_UNREACH_NLRI

    Carry the set of unreachable destinations

BGP Capabilities Negotiation

- BGP routers establish BGP sessions through the OPEN message
- OPEN message contains optional parameters
- BGP session is terminated if OPEN parameters are not recognised
- A new optional parameter: CAPABILITIES

Capabilities Negotiation

- Allows for the advertisement of capabilities (type 2)
- Backwards Compatible
  - ✓ new error subcode introduced to indicate which capabilities are not supported - the session must be reset

```
+------------------------------+
| Capability Code (1 octet)    |
+------------------------------+
| Capability Length (1 octet)  |
+------------------------------+
| Capability Value (variable)  |
+------------------------------+
```

draft-ietf-idr-bgp4-cap-neg-06, March 2000

Capabilities Negotiation

- Current Capabilities
  - ✓ 1 - Multiprotocol
  - ✓ 128 - Route Refresh
  - ✓ 129 - Outbound Route Filter

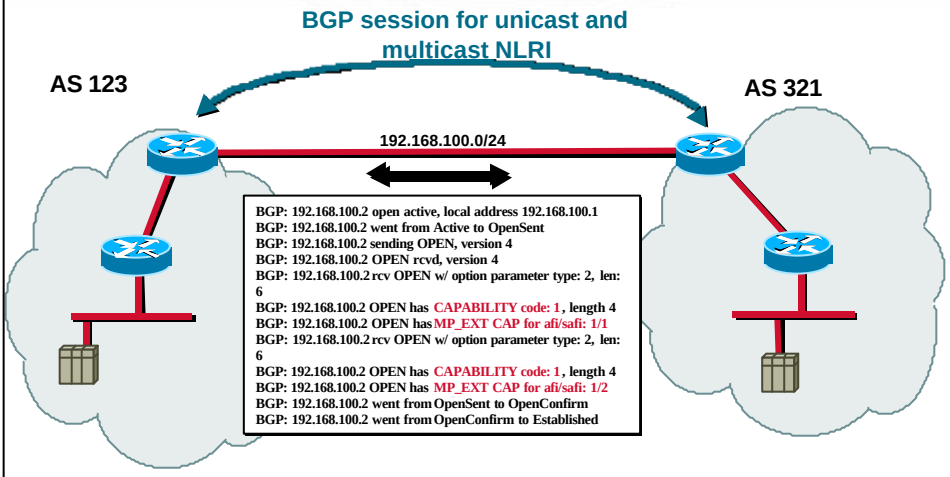
MBGP **= Multi-Protocol BGP not Multicast BGP. Multicast BGP is a capability of MBGP**

BGP Capabilities Negotiation

- **BGP routers determine capabilities of their neighbors by looking at the capabilities parameters in the open message**
- **Unknown or unsupported capabilities may trigger the transmission of a NOTIFICATION message**

> "The decision to send the NOTIFICATION message and terminate peering is local to the speaker. Such peering should not be re established automatically"
>
> draft-ietf-idr-bgp4-cap-neg-02

BGP Capabilities Negotiation



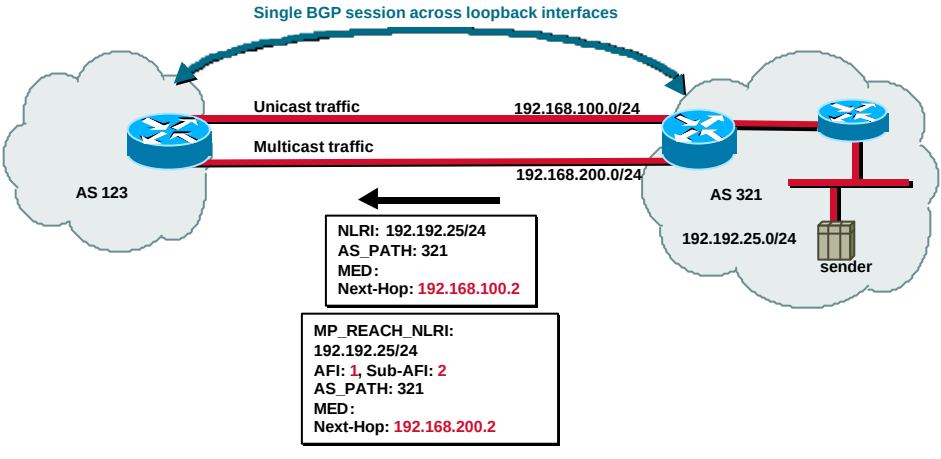
BGP Capabilities Negotiation

- BGP routers use BGP-4 Multiprotocol Extension to carry label (tag) mapping information
  - Multiprotocol Extension capability
  - Used to negotiate the Address Family Identifier

    AFI and Sub-AFI
  - Multiple routes to destination capability

    BGP routers may advertise multiple routes to same destination (unicast and multicast)

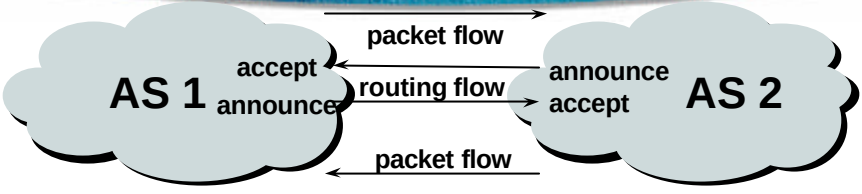
BGP Capabilities Negotiation



Route Refresh Capability

- **Facilitates non-disruptive policy changes**
- **No configuration is needed**
- **No additional memory is used**
- **clear ip bgp x.x.x.x [soft] in**
- **draft-ietf-idr-bgp-route-refresh-01.txt**

Route Refresh Capability



- **clear ip bgp <addr> -** Hard reset of the peer. Clears tables on both sides - traffic flow stops.
- **clear ip bgp <addr> [soft] out -** Resents the outbound advertisements. Traffic flow does not stop.
- **clear ip bgp <addr> [soft] in -** The *soft-reconfiguration* is required - keeps a copy of all inbound advertisements - takes up more memory. Traffic flow does not stop.
- **clear ip bgp <addr> soft -** Tells peer to resend data - both peers resend. Traffic flow does not stop. If capability not negotiated, then it is ignored.

Route Refresh Capability

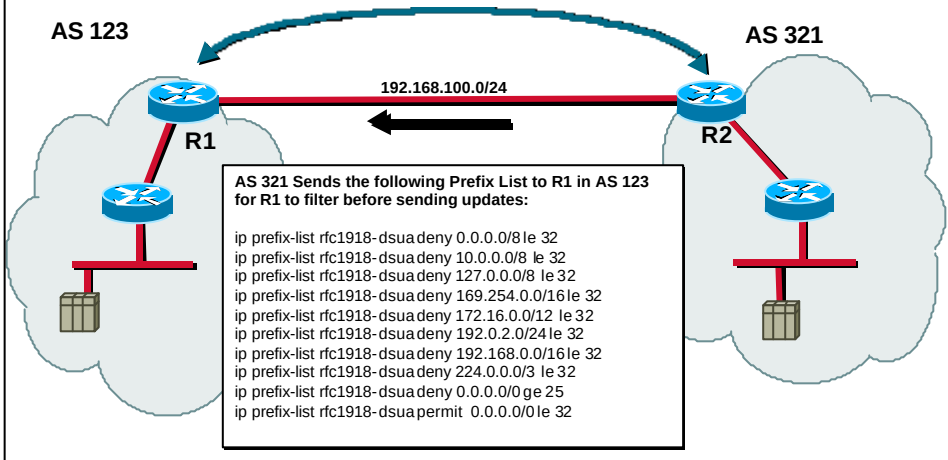
```
7206-AboveNet-SJ2#sh ip bgp neighbor 207.126.96.42
BGP neighbor is 207.126.96.42, remote AS 6461, external link
.
  Neighbor NLRI negotiation:
    Configured for unicast routes only
    Peer negotiated unicast routes only
    Exchanging unicast routes only
  Received route refresh capability from peer
.
  Route refresh request: received 0, sent 0
.
  Number of unicast/multicast prefixes received 77249/0
  Number of prefix received but not used 0
```


Outbound Route Filter Capability

- Allows for the use of the neighbor's inbound prefix-list as part of the local outbound policy (currently only for IPv4 unicast NLRI)
  - ✓ reduces the number of updates
  - ✓ 5 sec. delay after session is established before updates are sent
  - ✓ incremental updates not currently supported

Outbound Route Filter Capability

BGP ORF Capability Negotiated



Outbound Route Filter Capability

- **AS 321 Router 2 Configuration**

```
router bgp 321
 network 221.10.0.0 mask 255.255.224.0
 neighbor 192.168.100.1 remote-as 123
 neighbor 192.168.100.1 capability prefex-filter
 neighbor 192.168.100.1 prefix-list rfc1918-dusa in
 neighbor 192.168.100.1 send prefix-filter
 neighbor 192.168.100.1 prefix-list outfilter out
!
```

Outbound Route Filter Capability

- Works with Peer Groups (i.e. eBGP Peer groups on a IXP)
- Commands:
  - ✓ clear ip bgp x.x.x.x in prefix-filter
  - ✓ show ip bgp neighbor x.x.x.x received prefix-filter
  - ✓ show ip bgp neighbor x.x.x.x

    Prefix ORF:

    Capability: advertised, received

    Filter: sent; received (25 entries)

Multiprotocol Extensions I

- Address Family Identifier - rfc1700
  - ✓ 1 IPv4
  - ✓ 2 IPv6
  - ✓ 8 E.164
- Sub-AFI (for IPv4)
  - ✓ 1 Unicast
  - ✓ 2 Multicast
  - ✓ 3 Unicast + Multicast

Multiprotocol Extensions II

- MPLS VPN
  - ✓ used to carry both intra and inter VPN routing information
- New AFI == VPN-IPv4
- NLRI format for VPN addresses
  - ✓ Tag
  - ✓ VPNID (32 bits)
  - ✓ Prefix (variable length, 0 - 32 bits)

ISP Network Design

- Routed Backbone
- Switched Backbone
- Leased point-to-point circuits
  - ✓ nx64K, T1/E1, T3/E3, OC3, OC12,...
- ATM/Frame Relay service from telco
  - ✓ T3, OC3, OC12,… delivery
  - ✓ easily upgradeable bandwidth (CIR)

Routing Protocols

- IGP - Interior Gateway Protocol
  - ✓ carries infrastructure addresses, point-to-point links
  - ✓ examples are OSPF, ISIS, EIGRP...
- EGP - Exterior Gateway Protocol
  - ✓ carries customer prefixes and Internet routes
  - ✓ current EGP is BGP version 4
- No link between IGP and EGP

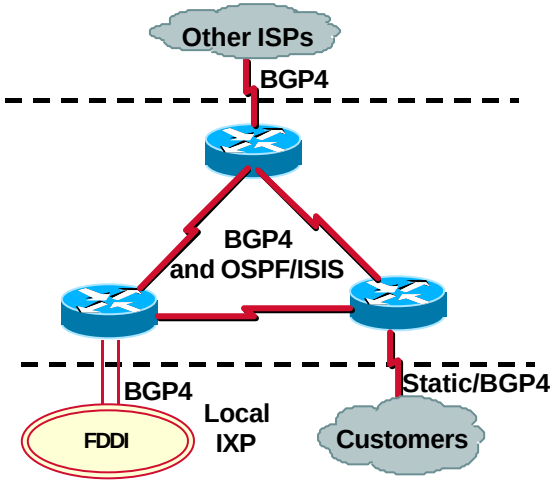
Why Do We Need an IGP?

- ISP backbone scaling
  - ✓ Hierarchy
  - ✓ Modular infrastructure construction
  - ✓ Limiting scope of failure
  - ✓ Healing of infrastructure faults using dynamic routing with fast convergence

Why Do We Need an EGP?

- Scaling to large network
  - ✓ Hierarchy
  - ✓ Limit scope of failure
- Policy
  - ✓ Control reachability to prefixes
  - ✓ Merge separate organizations
  - ✓ Connect multiple IGPs

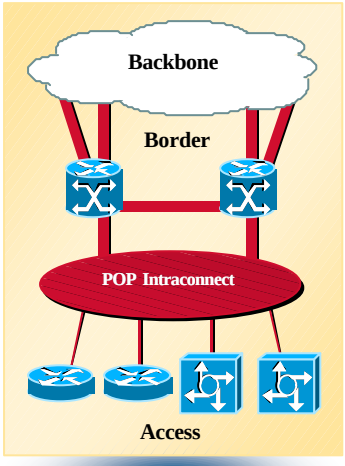
Hierarchy of Routing Protocols



Point of Presence Topologies

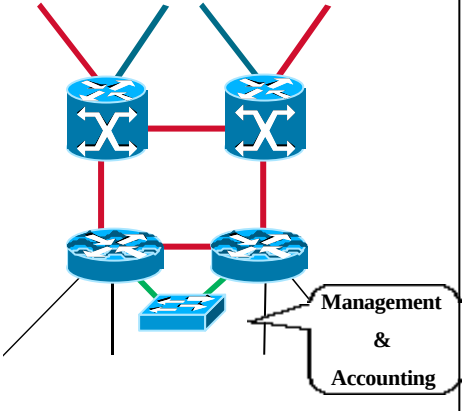
PoP Design

- **Triple Layered POP Redundancy**
  - ✓ **Two connection to the backbone from any border router**
  - ✓ **Two border routers, load balanced with one able to take the full load**
  - ✓ **Two POP interconnect devices and/or a physical failover medium (FE/GE, POS, DTP)**



Key Design Principles

- **Interconnection for Management, Security, and Accounting services**
  - ✓ Netflow Devices - FlowCollector
  - ✓ Syslog collector for all network devices
  - ✓ SNMP collector (PC Based UNIX)
  - ✓ Security Auditing Tools (NetSonar)



PoP Topologies

- Core routers - high speed trunk connections
- Distribution routers and Access routers - high port density
- Border routers - connections to other providers
- Service routers - hosting and servers
- Some functions might be handled by a single router

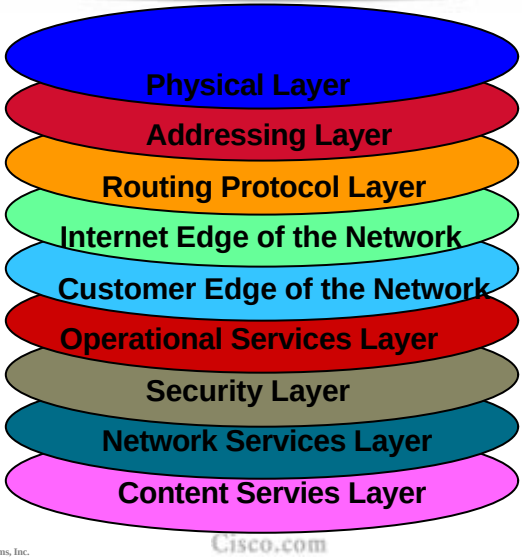
PoP Design

- Modular Design
- Aggregation Services separated according to
  - ✓ connection speed
  - ✓ customer service
  - ✓ contention ratio
  - ✓ security considerations

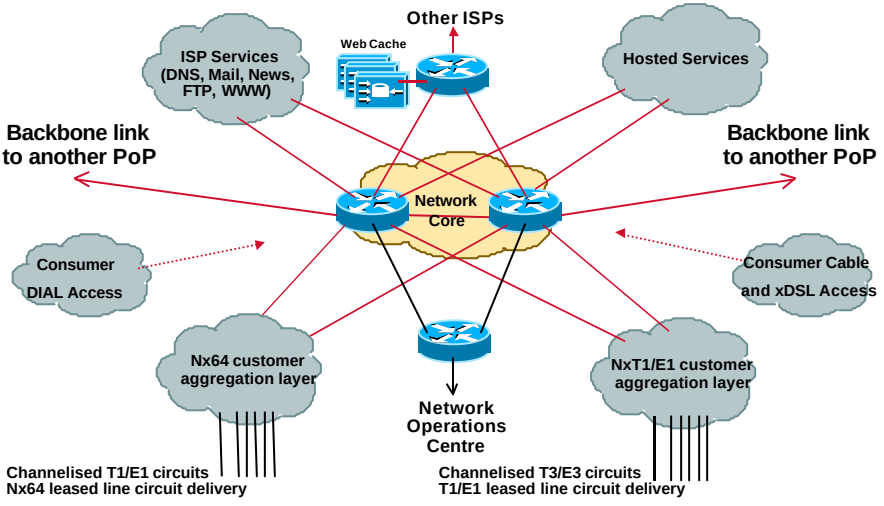
Use a Conceptualization Model in Your ISP Design

- ISP Design involves a lot of interrelated and interactive factors.
- It is use full to find and use conceptualization models to insure you are covering all aspects of an ISP's design.

Think Interrelated Layers

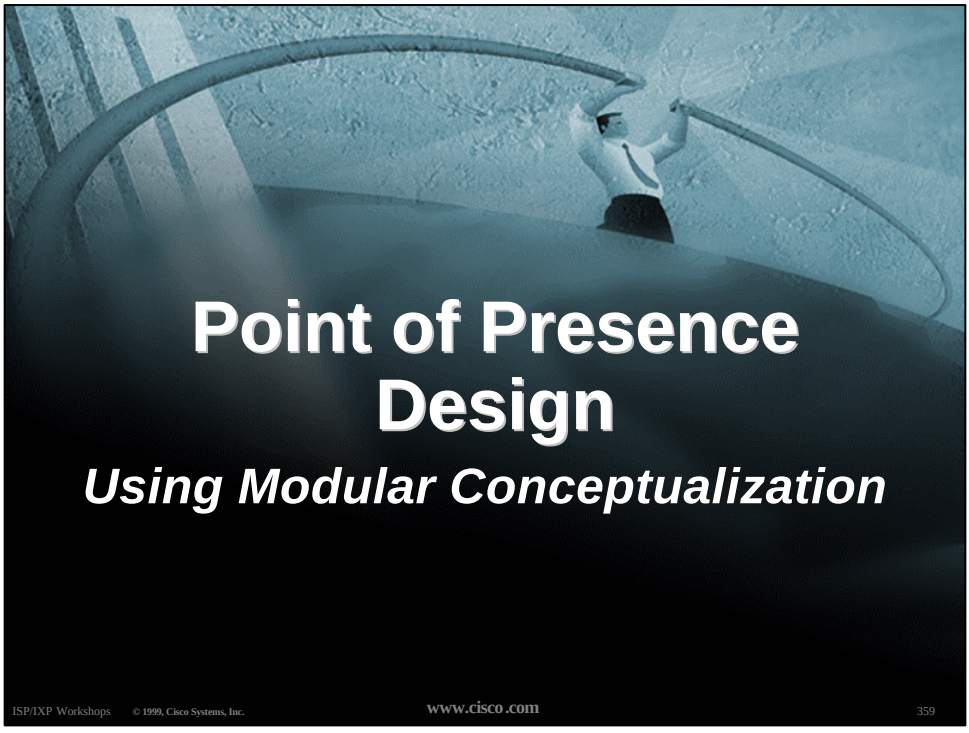


Think Modular



Modular Routing Protocol Design

- Modular IGP implementation
  - ✓ IGP “area” per module
  - ✓ aggregation/summarisation into the core
- Modular iBGP implementation
  - ✓ BGP route reflector cluster per module
  - ✓ core routers are route-reflectors
  - ✓ clients peer with core only



Point of Presence Design

Using Modular Conceptualization

PoP Modules

- **Low Speed customer connections**
  - ✓ **PSTN/ISDN dialup**
  - ✓ **low bandwidth needs**
  - ✓ **low revenue, large numbers**
- **Medium Speed customer connections**
  - ✓ **56/64K to sub-T1/E1 speeds**
  - ✓ **low bandwidth needs**
  - ✓ **medium revenue, medium numbers**

PoP Modules

- High Speed customer connections
  - ✓ E1++ speeds
  - ✓ medium bandwidth needs
  - ✓ high revenue, low numbers
- Broad Band customer connections
  - ✓ xDSL and Cable
  - ✓ high bandwidth needs
  - ✓ low revenue, large numbers

PoP Modules

- PoP Core
  - ✓ Two dedicated routers
  - ✓ High Speed interconnect
  - ✓ Backbone Links ***ONLY***
  - ✓ ***Do not touch them!***
- Border Network
  - ✓ dedicated border router to other ISPs
  - ✓ the ISP's "front" door
  - ✓ transparent web caching

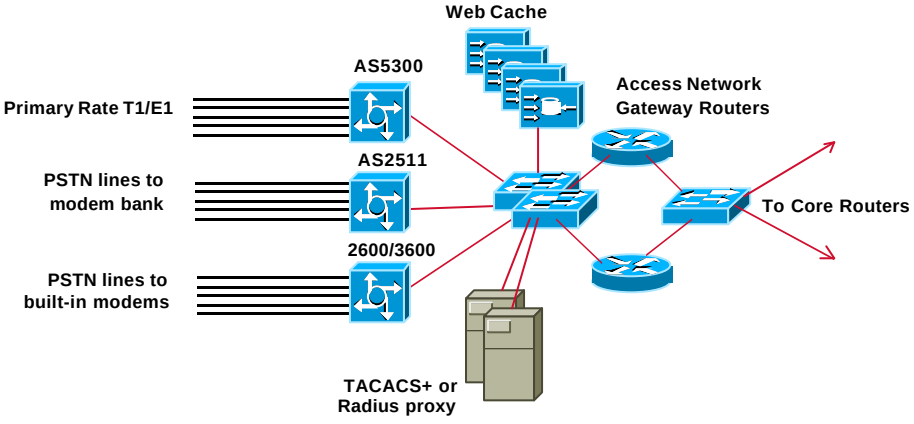
PoP Modules

- ISP Services
  - ✓ DNS (cache, secondary)
  - ✓ News, Mail (POP3, Relay)
  - ✓ WWW (server, proxy, cache)
- Hosted Services
  - ✓ Virtual Web, WWW (server, proxy, cache)
  - ✓ Information/Content Services
  - ✓ Electronic Commerce

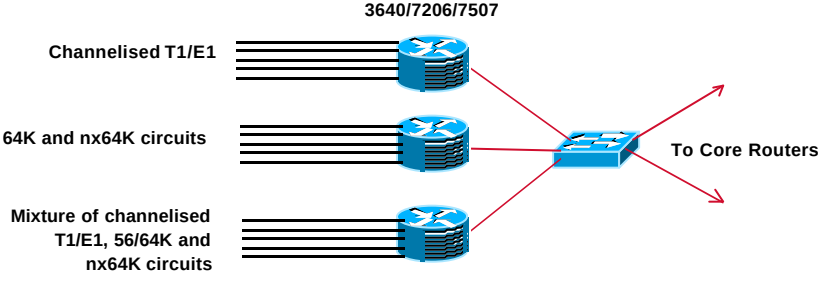
PoP Modules

- Network Operations Centre
  - ✓ primary and backup locations
  - ✓ network monitoring
  - ✓ statistics and log gathering
  - ✓ direct but secure access
- Out of Band Management Network
  - ✓ The ISP Network “Safety Belt”

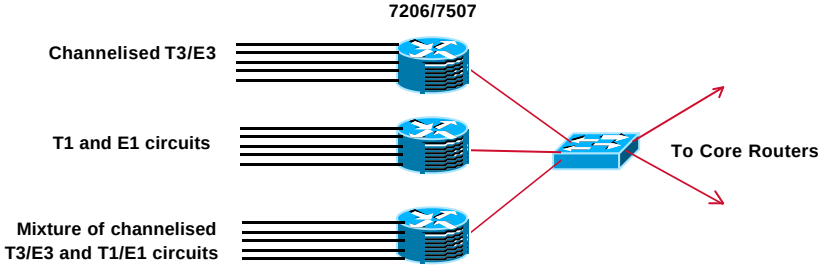
Low Speed Access Module



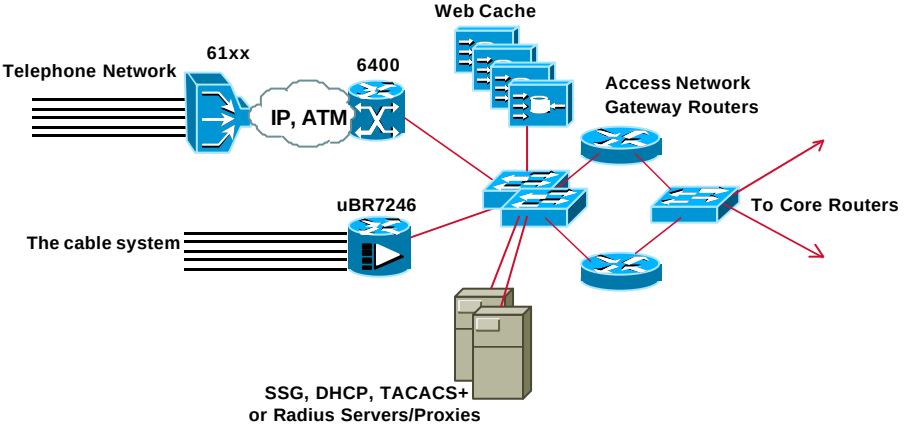
Medium Speed Access Module



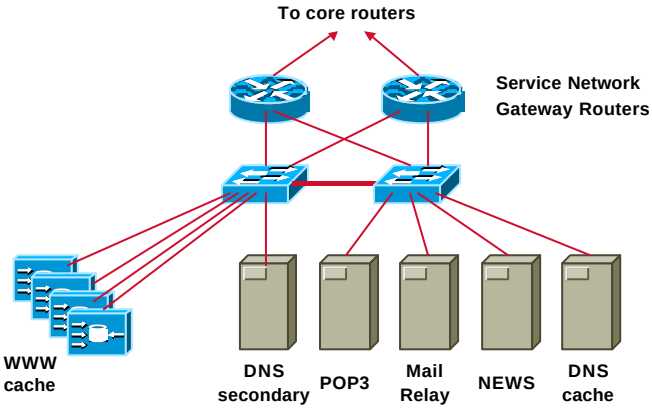
High Speed Access Module



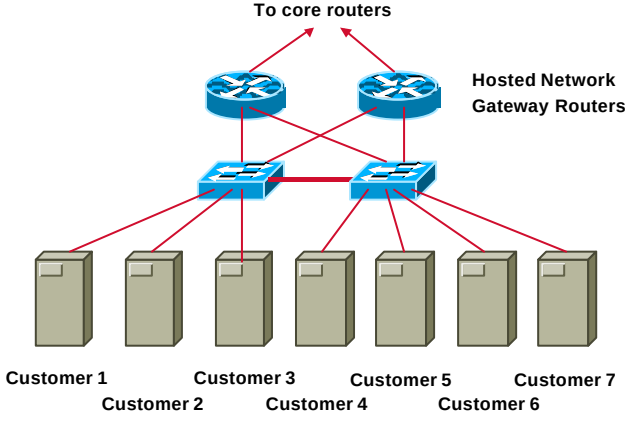
Broad Band Access Module



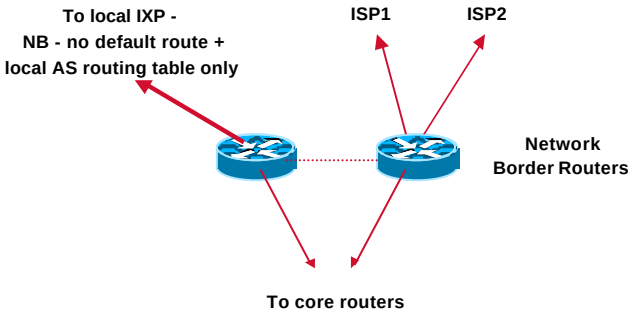
ISP Services Module



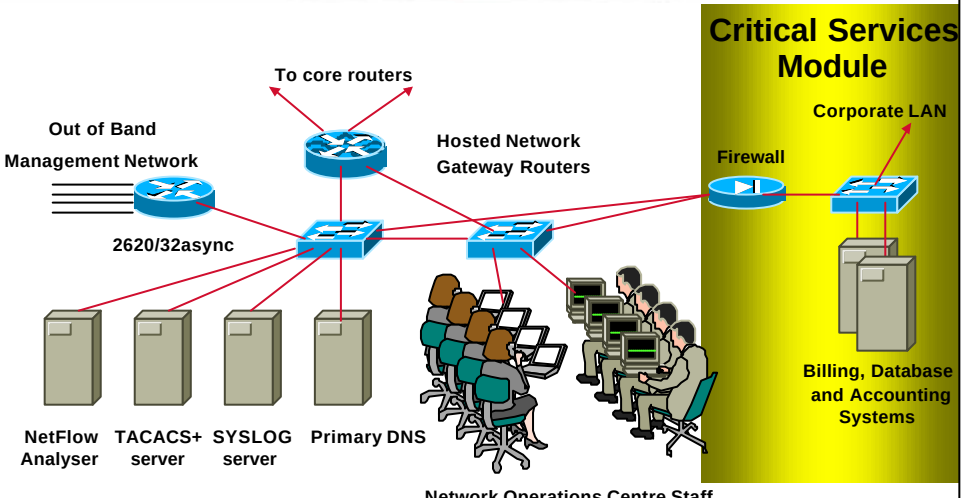
Hosted Services Module



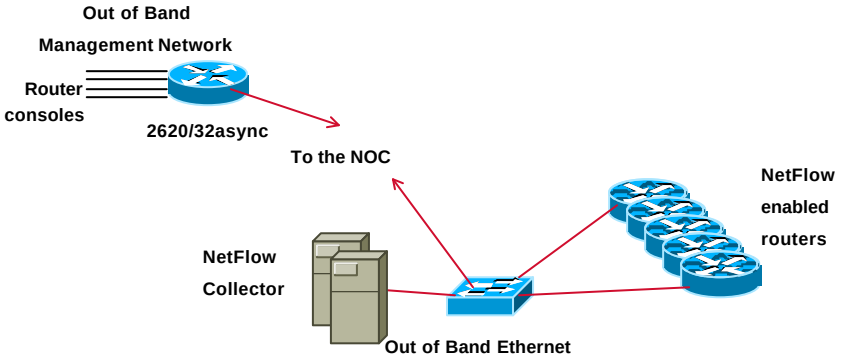
Border Module



NOC Module



Out of Band Network

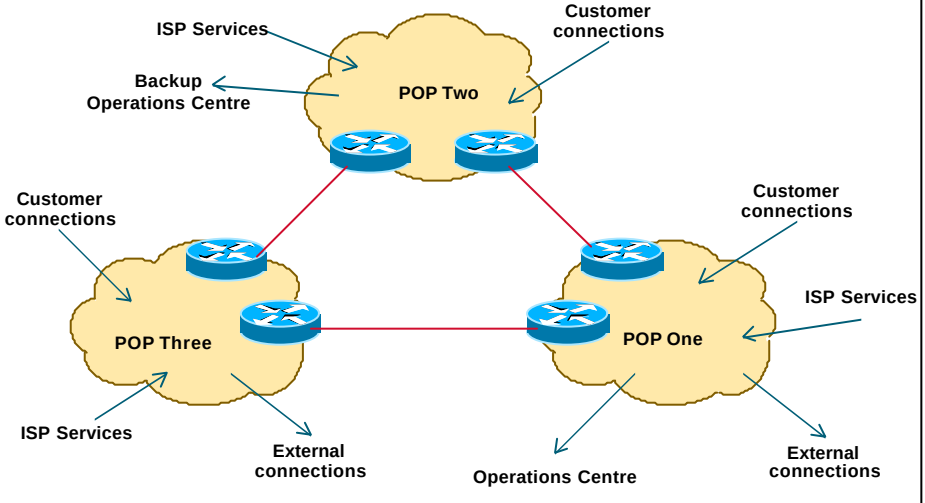


Backbone Network Design

Distributed Network Design

- PoP design “standardised”
  - ✓ operational scalability and simplicity
- ISP essential services distributed around backbone
- NOC and “backup” NOC
- Redundant backbone links

Distributed Network Design



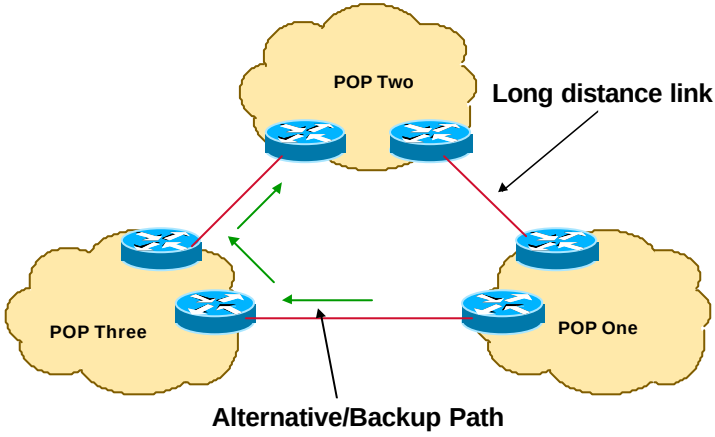
Backbone Links

- ATM/Frame Relay
  - ✓ now less popular due to overhead, extra equipment, and shared with other customers of the telco
- Leased Line
  - ✓ more popular with backbone providers
  - ✓ IP over Optics and MPLS coming into the mainstream

Long Distance Backbone Links

- Tend to cost more
- Plan for the future (at least two years ahead) but stay in budget
  - ✓ Unplanned “emergency” upgrades can be disruptive without redundancy
- Allow sufficient capacity on alternative paths for failure situations
  - ✓ sufficient can be 20% to 50%

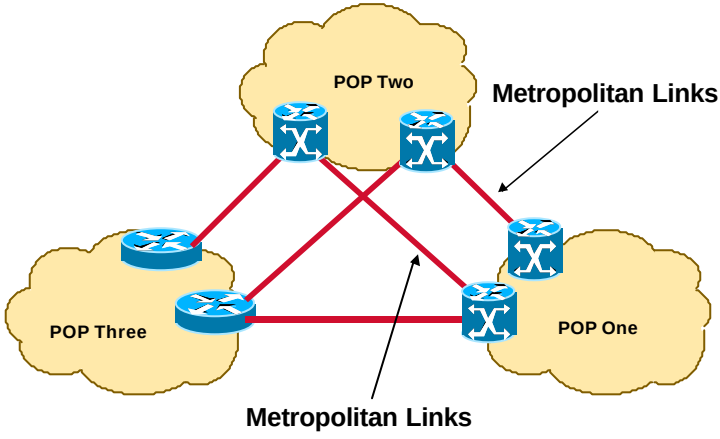
Long Distance Links



Metropolitan Area Backbone Links

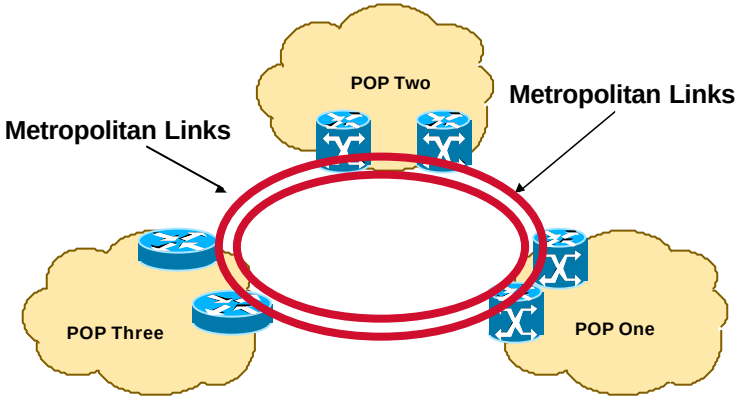
- Tend to be cheaper
  - ✓ Circuit concentration
  - ✓ Choose from multiple suppliers
- Think big
  - ✓ More redundancy
  - ✓ Less impact of upgrades
  - ✓ Less impact of failures

Metropolitan Area Backbone Links - Option One



Traditional Point to Point Links

Metropolitan Area Backbone Links - Option Two



DPT - Dynamic Packet Transport

DPT

- Dual counter rotating ring, scalable bandwidth
- Supports multicast, traffic prioritization
- Multiple nodes can transmit simultaneously - “spatial reuse protocol”
- Uses the SRP fairness algorithm to control access to the ring
  - ✓ No token—unlike Token Ring or FDDI
- Scalable to large number of nodes
  - ✓ Unlike SONET/SDH

IP Registries

- www.apnic.net
- www.ripe.net
- www.arin.net
- In formation
  - ✓ www.afrinic.org
  - ✓ www.lacnic.org

Addressing Plans - ISP Infrastructure

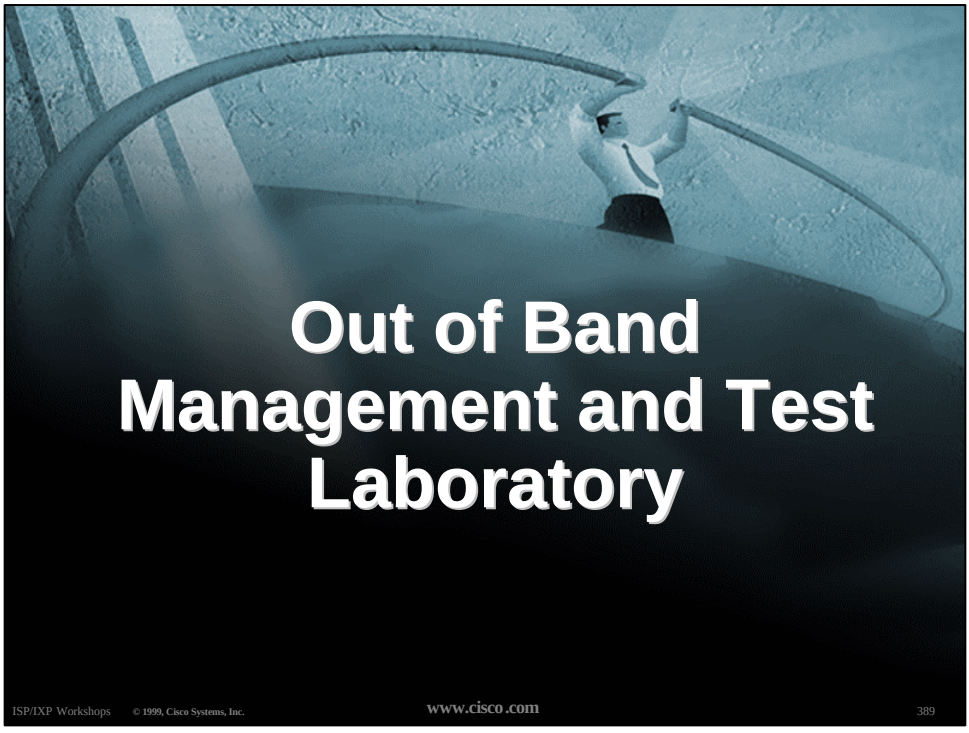
- Address block for router loop-back interfaces
- Address block per PoP for infrastructure
  - ✓ summarise between sites
  - ✓ allocate according to genuine requirements, not historic classful boundaries
- Address block for links to customers

Addressing Plans - Customer

- Customers assigned address space according to need
- Should not be reserved or assigned on a per PoP basis
  - ✓ ISP iBGP carries customer nets
  - ✓ aggregation not required and usually not desirable

Addressing Plans (contd)

- Document infrastructure allocation
  - ✓ eases operation, debugging and management
- Document customer allocation
  - ✓ contained in iBGP
  - ✓ eases operation, debugging and management
  - ✓ submit network object to APNIC Database

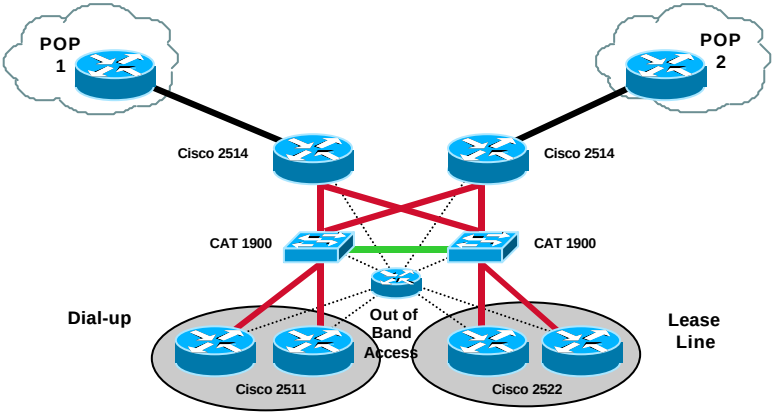


Out of Band Management and Test Laboratory

Other Design Considerations

- Out of Band Management
  - ✓ how to get to equipment when “the network is down”
- Test Laboratory
  - ✓ how to test new services and features
  - ✓ how to debug network problems

Out of Band Management



Out of Band Management

- Not optional!
- Allows access to network equipment in times of failure
- Ensures quality of service to customers
  - ✓ minimises downtime
  - ✓ minimises repair time
  - ✓ eases diagnostics and debugging

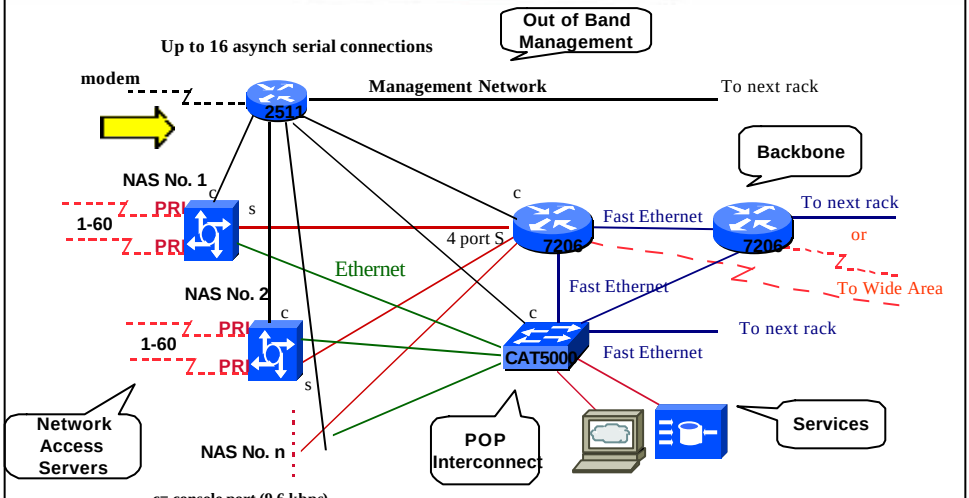
Out of Band Management

- OoB Example - Access server:
  - ✓ modem attached to allow NOC dial in
  - ✓ console ports of all network equipment connected to serial ports
  - ✓ LAN and/or WAN link connects to network core, or via separate management link to NOC
- Full remote control access under all circumstances

Out of Band Management

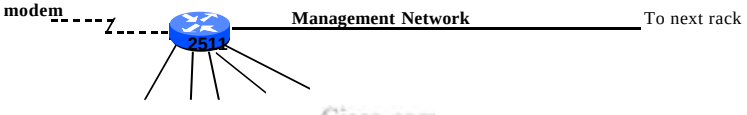
- OoB Example - Statistics gathering:
  - ✓ Routers are NetFlow and syslog enabled
  - ✓ Management data is congestion/failure sensitive
  - ✓ Ensures management data integrity in case of failure
- Full remote information under all circumstances

Out of Band Management



Out of Band Management

- **Out of Band Access's Objective: Manage your POP <u>remotely</u>**
  - ✓ **Use a Cisco 2511, 2600, or 3600 to connect to the console ports of all you equipment in the POP (routers, hubs, switches, PBXs, workstations, SDH equipment, modems banks, UPS equipment, etc.)**
  - ✓ **Engineer in the NOC gets to the out of band access via modem, IDSN, Frame Relay, Lease Line or ethernet.**
  - ✓ **NOC Team can access the entire POP comfort**



Out of Band Management

- **Sample Config:**

```
line 1 16
 no exec
 transport input telnet
```

- **telnet 192.168.1.1 2001 (line 1)**
- **When connecting to the console port of a SUN Workstation - be careful:**
  - ✓ **http://www.cisco.com/warp/public/770/fn-tsbreak.html**

Test Laboratory

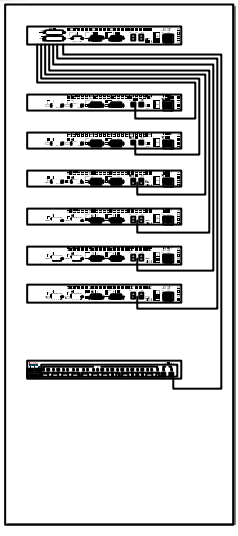
- **Looks like a typical PoP**
- **Used to trial new services or new software under realistic conditions**
- **Allows discovery of potential problems before they are introduced to the network**
- **Every major ISP in the US and Europe has a test lab**

Test Laboratory

- Some ISPs dedicate equipment to the lab
- Other ISPs “purchase ahead” so that today’s lab equipment becomes tomorrow’s PoP equipment
- Other ISPs use lab equipment for “hot spares” in the event of hardware failure

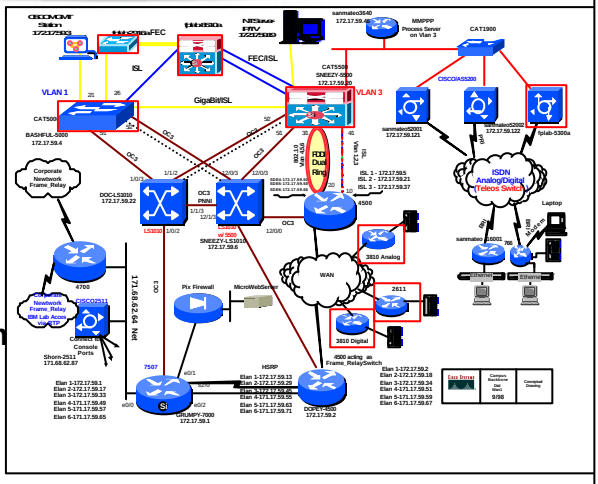
Test Laboratory

- Minimum Configuration for Routing Protocol Work:
  - ✓ 6 Cisco 2514s (two serial and two ethernet)
  - ✓ 1 CAT 1900XL
  - ✓ V.35 DTE & DCE cables to have back to back connections.
  - ✓ 1 Cisco 2511 to connect to all the console ports (do the work from your desk).



Test Laboratory

- Cisco's SE Field Lab
  - ✓ Example of how detailed the labs can become.
  - ✓ SE Field Lab designed to provide maximum flexibility so SE's can simulate network proposals.



ISP Design Summary

- KEEP IT SIMPLE !
- Simple is elegant is scalable
- Use Redundancy, Security, and Technology to make life easier for yourself
- Above all, ensure quality of service for your customers

Where to get more information

- Supporting *IOS Essentials* WhitePaper
  - ✓ `http://www.cisco.com/public/cons/isp/documents/IOSEssentialsPDF.zip`
- Check the CTO Consulting Engineering ISP Resources page:
  - ✓ `http://www.cisco.com/public/cons/isp/`
- Join the cisco-nsp mailing list - set up by ISPs for ISPs
  - ✓ `send e-mail to majordomo@puck.nether.net with the words "subscribe cisco-nsp" in the body`

For Further Reference…

- **Computer Networks, Third Edition**
  by Andrew Tanenbaum (ISBN: 0-13349-945-6)

- **Interconnections : Bridges and Routers (second Ed)**
  by Radia Perlman (ISBN: 0-20163-448-1)

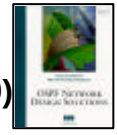
- **Internetworking with TCP / IP, Volume 1: Principles, Protocols, and Architecture**
  by Douglas Comer (ISBN: 0-13216-987-8)

- **IP Routing Fundamentals**
  by Mark Sportack (ISBN: 1-57870-071-x)

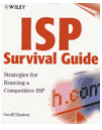
- **IP Routing Primer**
  by Robert Wright (ISBN: 1-57870-108-2)

For Further Reference…

- **Routing in the Internet**
  by Christian Huitema (ISBN: 0-13132-192-7)



- **OSPF Network Design Solutions**
  by Thomas, Thomas M. (ISBN: 1-57870-046-9)



- **ISP Survival Guide : Strategies for Running a Competitive ISP**
  by Geoff Huston (ISBN:0-47131-499-4)

- **Internet Routing Architectures**
  by Bassam Halabi (ISBN: 1-56205-652-2)

IOS Essentials — Best Practice Cisco IOS Techniques to Scale the Internet

Session 3302

Please Complete Your Evaluation Form

Session 3302

CISCO SYSTEMS

EMPOWERING THE INTERNET GENERATION℠