ISP Edge design

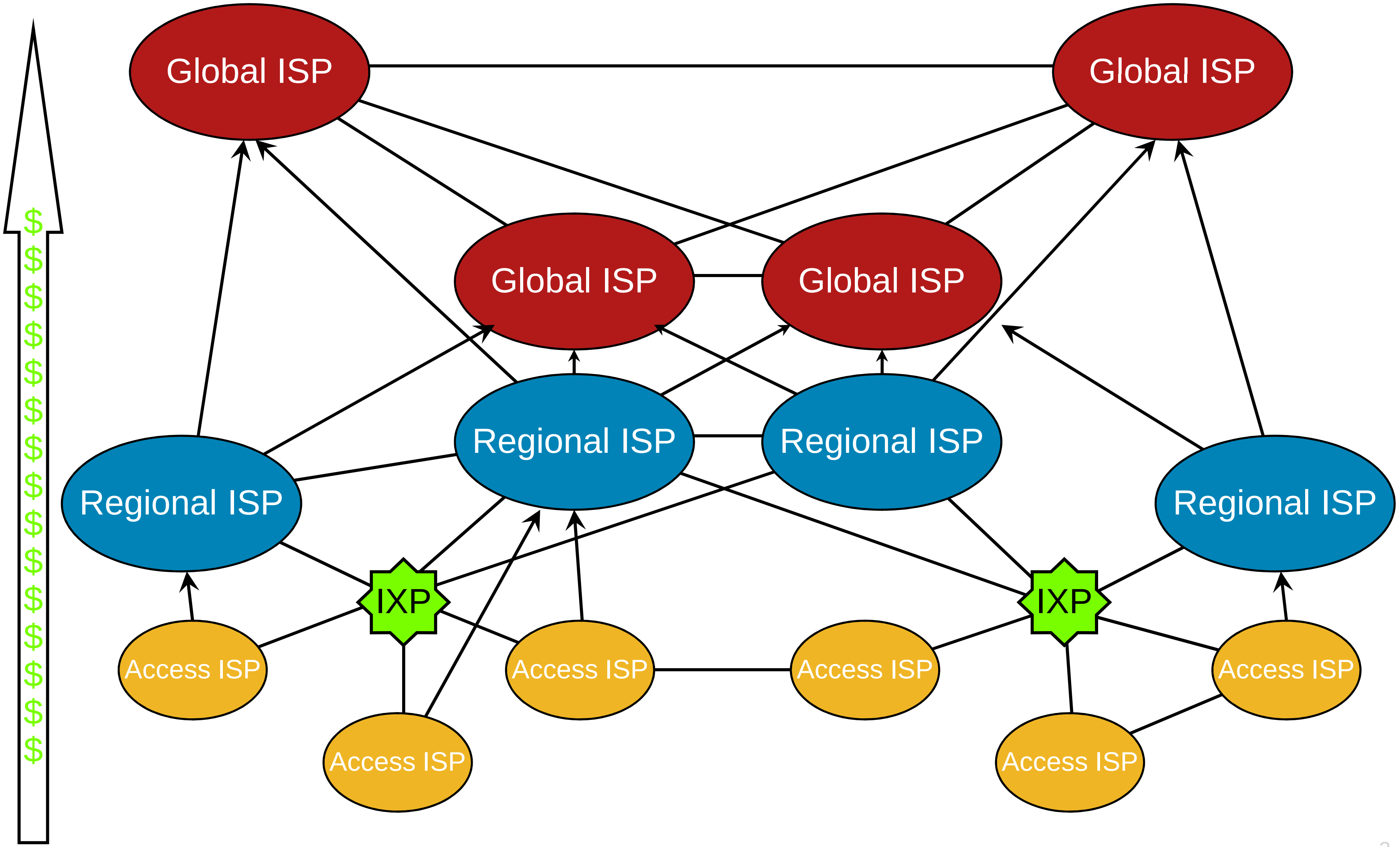
Josef Ungerman

CCIE #6167

Agenda

- The Internet
- IXP Intro
- Euro-IX
- Technical Details
- Live Examples
- OTT, Video and IXP
- Summary & Resources

Categorising ISPs

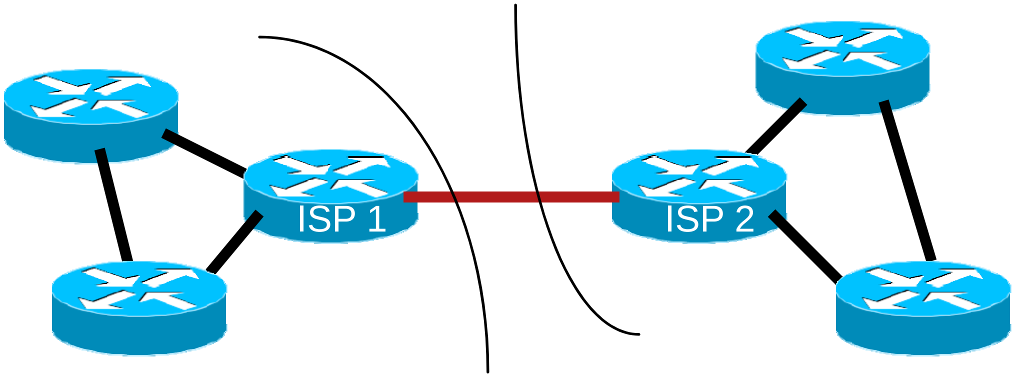


Peering and Transit

- **Transit**
  - Carrying traffic across a network
  - **Usually for a fee**
  - Example: Access provider connects to a regional provider
- **Peering**
  - Exchanging routing information and traffic
  - **Usually for no fee**
  - Sometimes called **settlement free peering**
  - Example: Regional provider connects to another regional provider

Private Interconnect

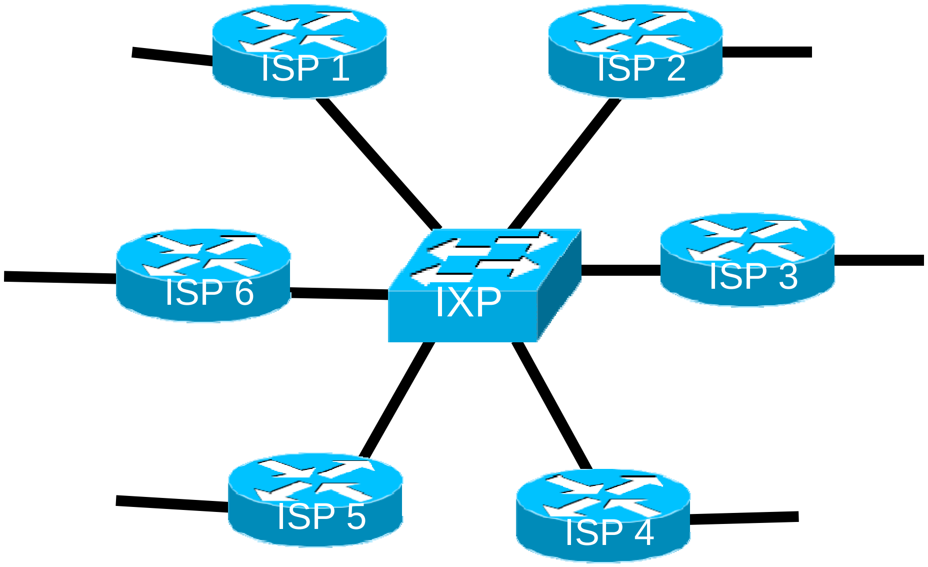
- Two ISPs connect their networks over a **private link**
  - Can be peering arrangement
    - No charge for traffic
    - Share cost of the link
  - Can be transit arrangement
    - One ISP charges the other for traffic
    - One ISP (the customer) pays for the link



Public Interconnect

- Several ISPs meeting in a common neutral location and interconnect their networks

  Usually is a peering arrangement between their networks



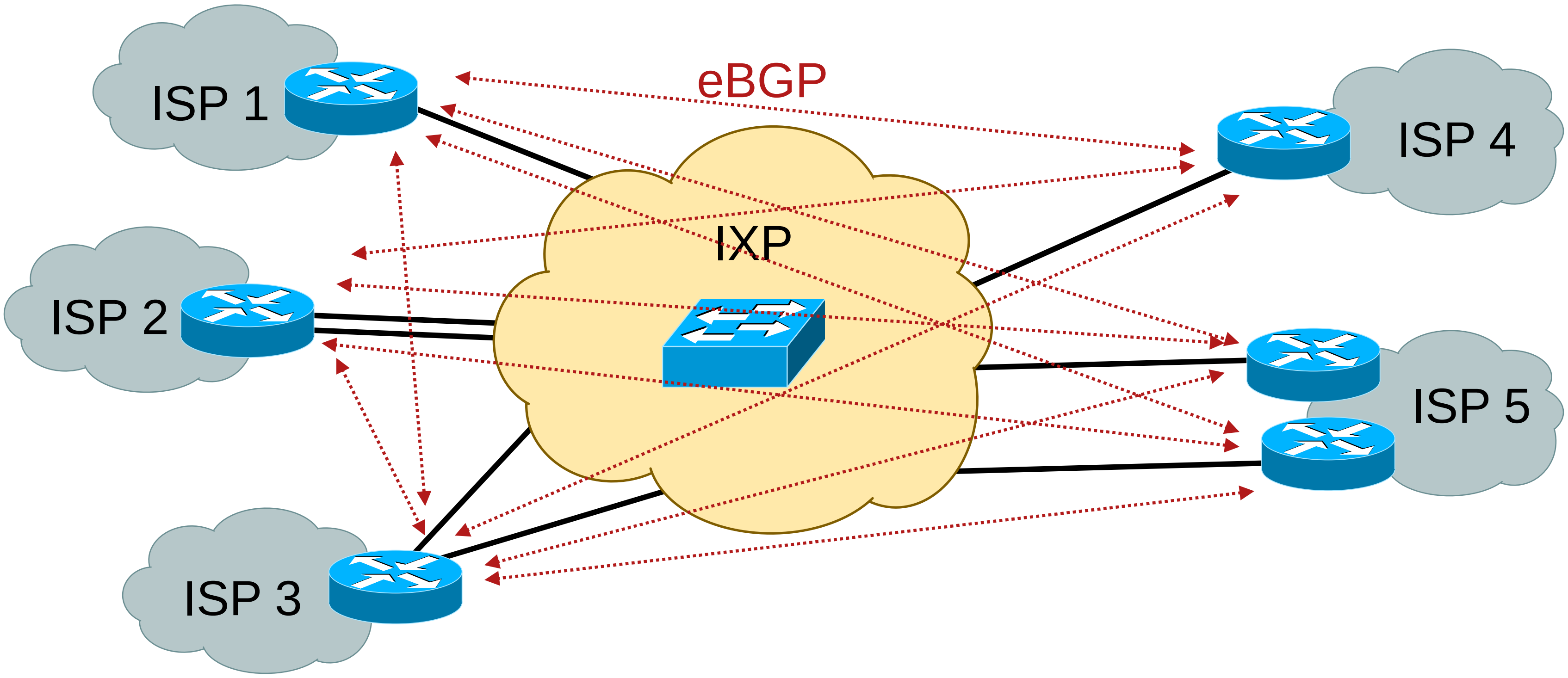
IXP
(Internet Exchange Points)

IXP (Internet eXchange Point)

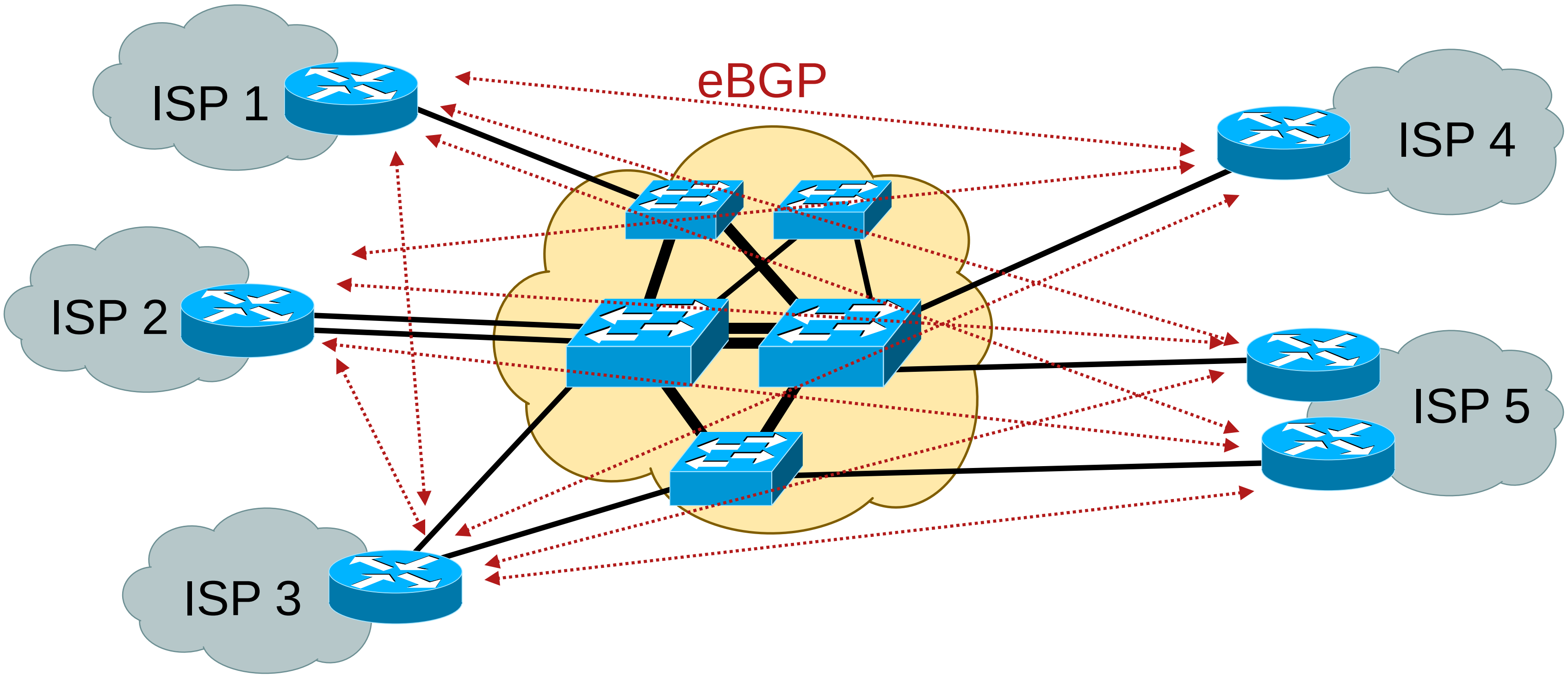
A physical network infrastructure operated by a single entity with the purpose to facilitate the exchange of Internet traffic between Autonomous Systems. The number of Autonomous Systems connected should at least be three and there must be a clear and open policy for others to join.

- High-speed/Low-cost Internet Traffic Exchange
- A.k.a. Public Peering or Settlement-Free Peering
- Non-Profit Associations or Commercial Datacenters
- Around 300 big IXPs in the world

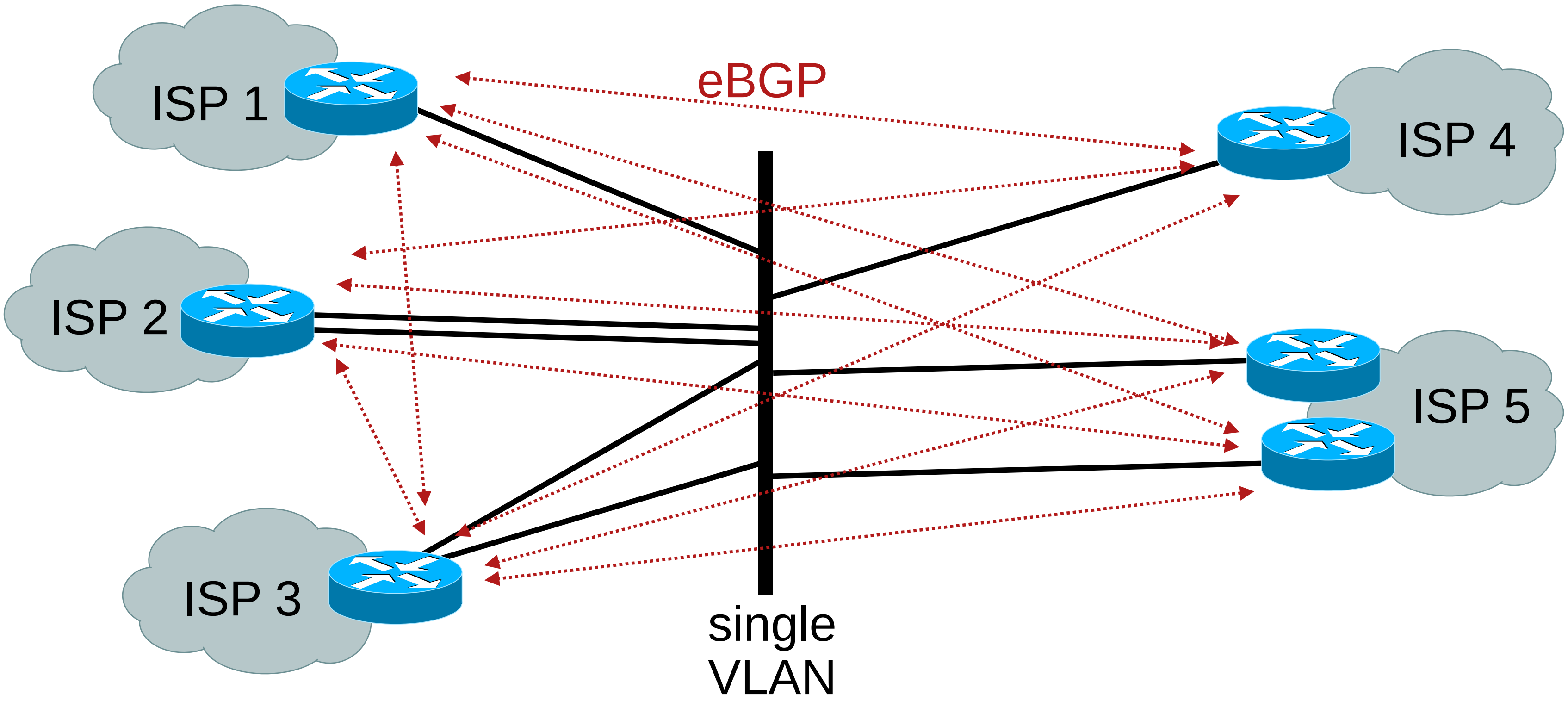
IXP (Internet eXchange Point)



IXP (Internet eXchange Point)



IXP (Internet eXchange Point)

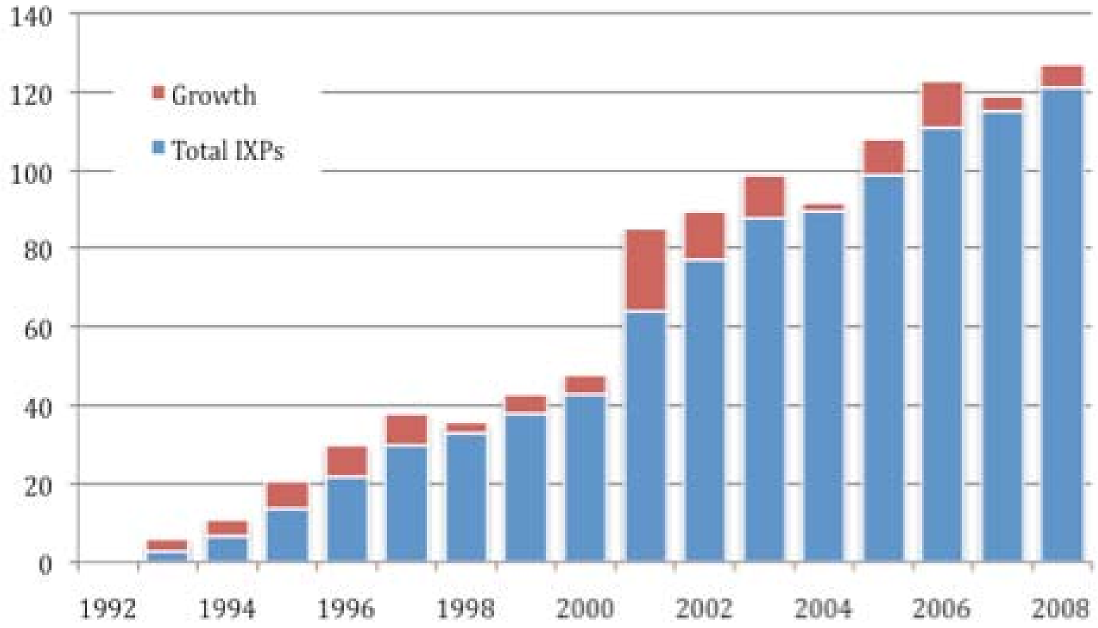


Euro-IX

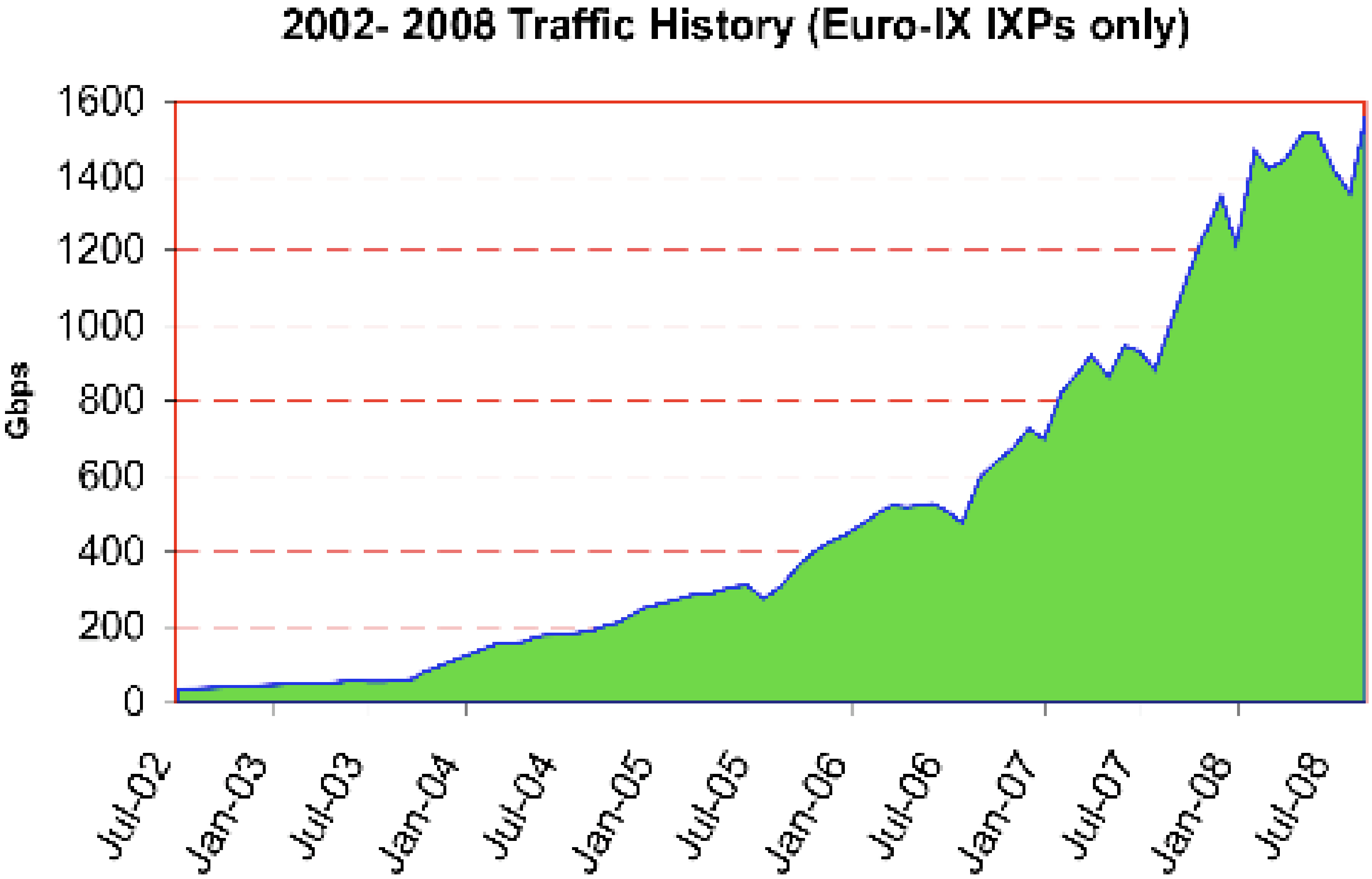
Euro-IX (European Internet Exchange Association) was formed in May 2001 with the intention to further develop, strengthen and improve the Internet Exchange Point (IXP) community

- **105 IXPs in 102 cities in 31 countries**
- 9 non-european members
- www.euro-ix.net

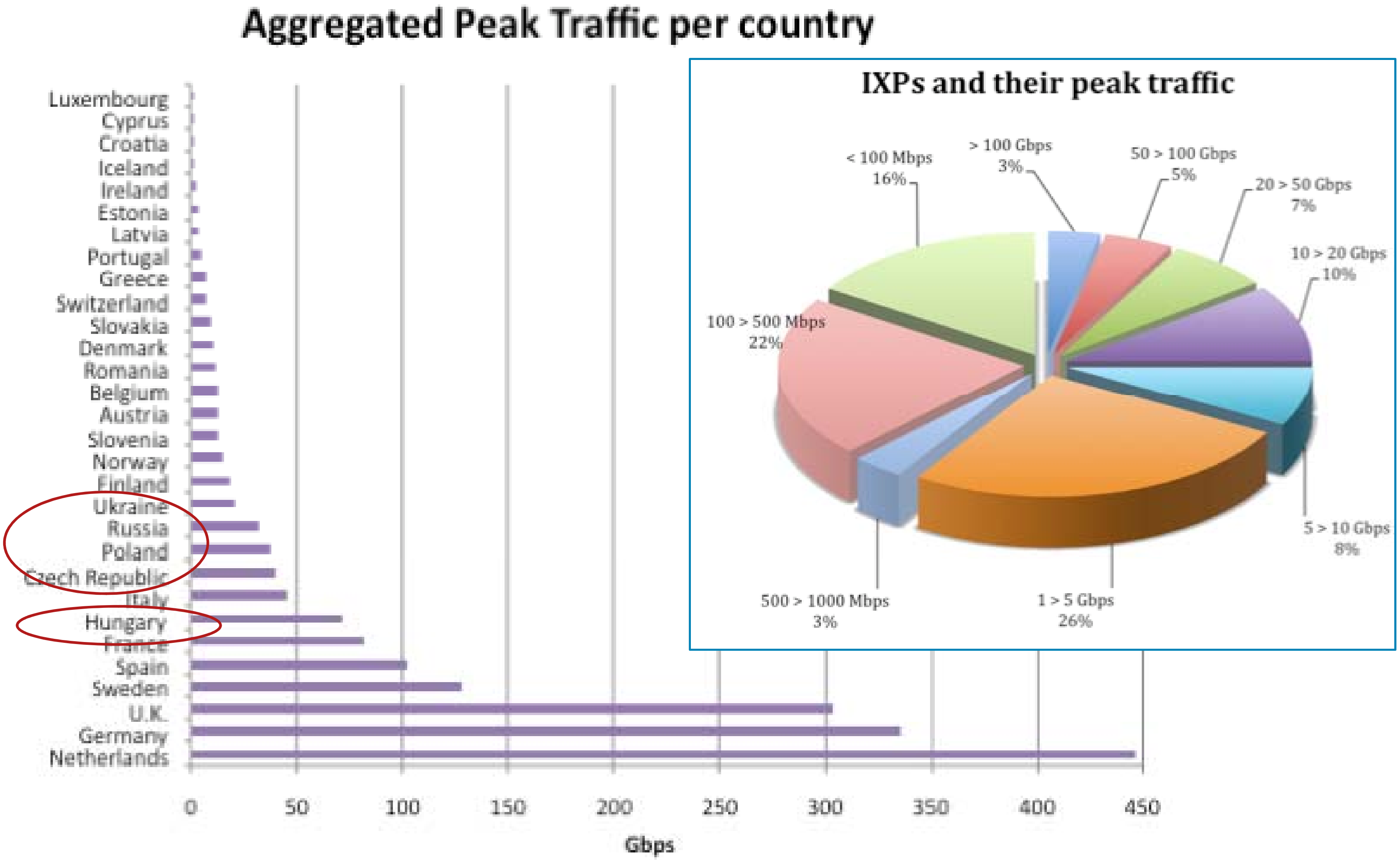
European IXP growth



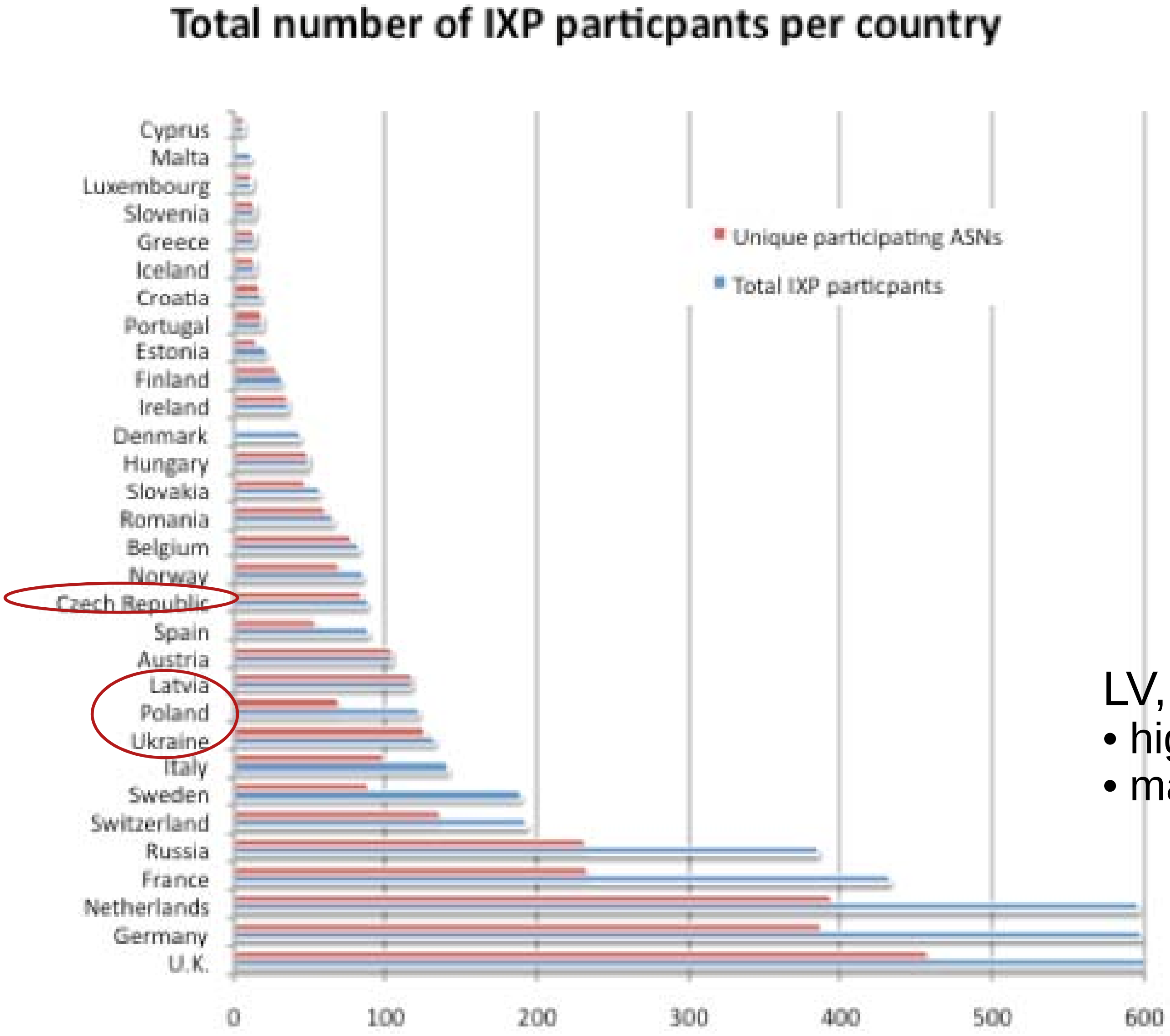
Euro-IX Report 2008



Euro-IX Report 2008



Euro-IX Report 2008

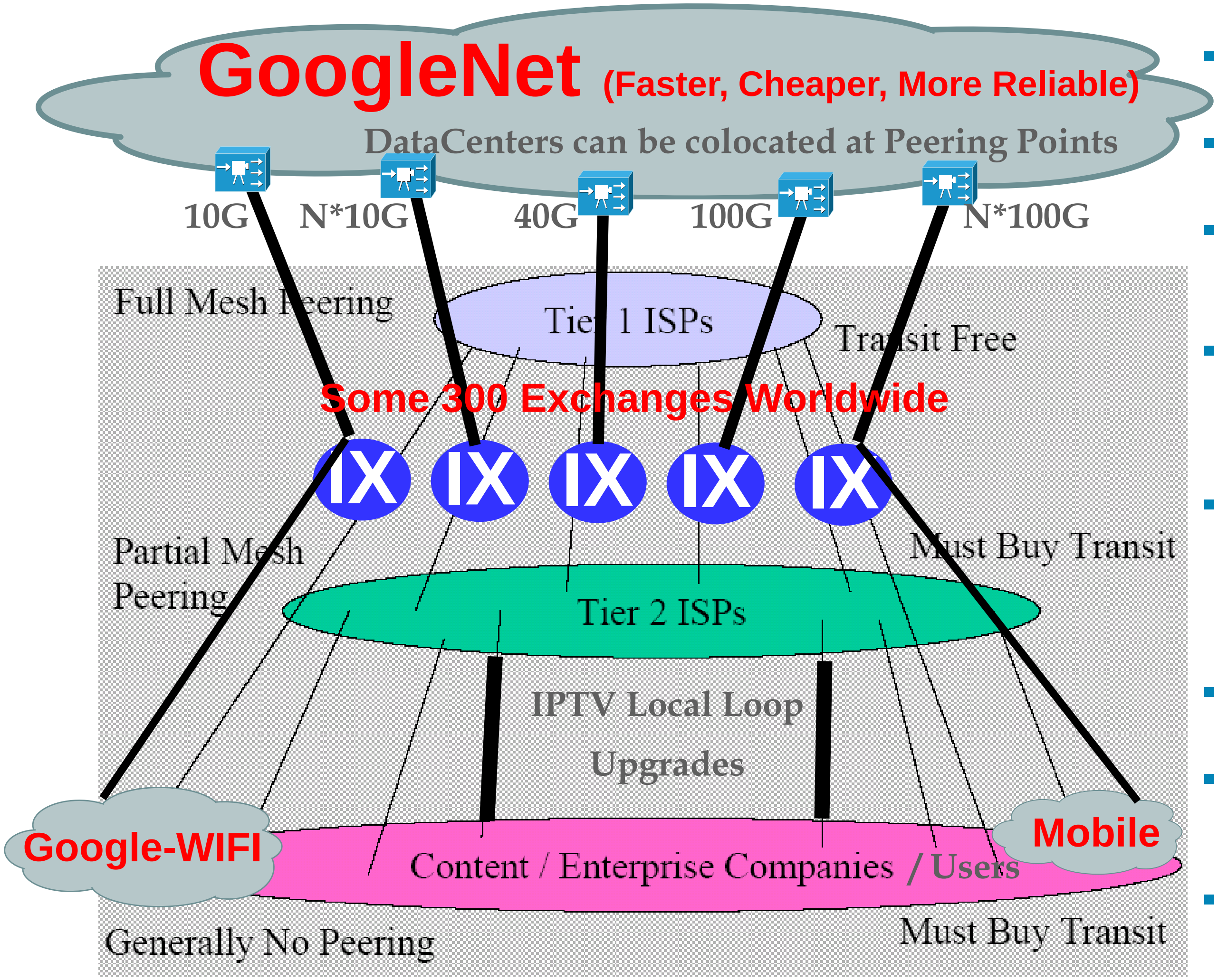


LV, PL, UA –

- highly fragmented ISP market
- maybe a lot of Hosting DC's

Example: GoogleNet…

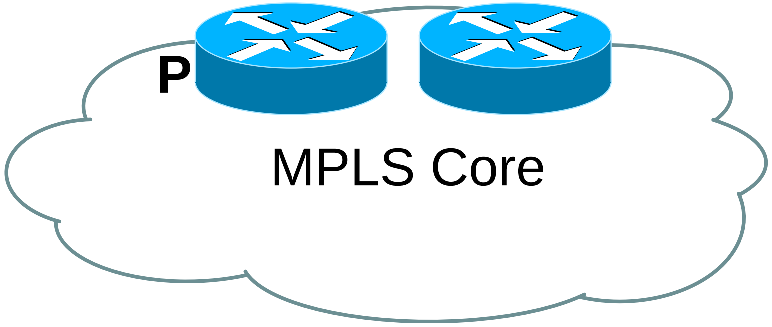
A PortalNet… Dedicated CDN… Parallel Internet BackBone



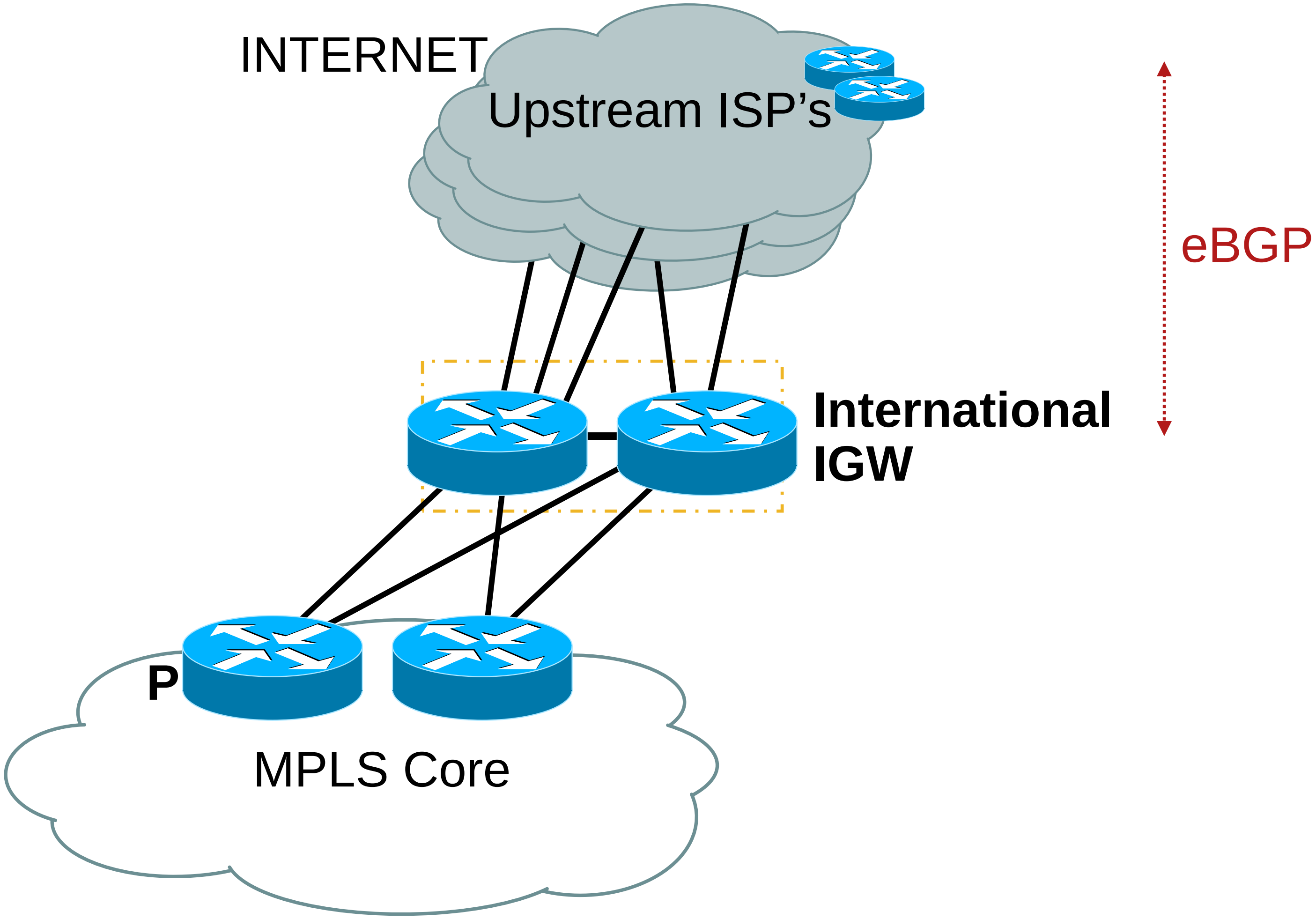
- Google has been buying Fiber on a Worldwide basis
- Google builds it's own worldwide IP Backbone.
- Google peers locally, often on a Settlement Free Basis, with Eyeball Carriers.
- Google can send any amount of traffic into the Internet without paying anyone, they are Nobody's Customer.
- Google distributes it's DataCenters to be virtually ONnet to Eyeball networks. Google is now only a few Hops away from Any User on the Internet.
- Tier2 ISP's invest in massive Local Loop upgrades to support IPTV.
- Google drives Net Neutrality so that whatever Traffic they send, can't be impaired.
- Google can now addresses Service Substitution (Google TV, Voice…)

Internet Edge

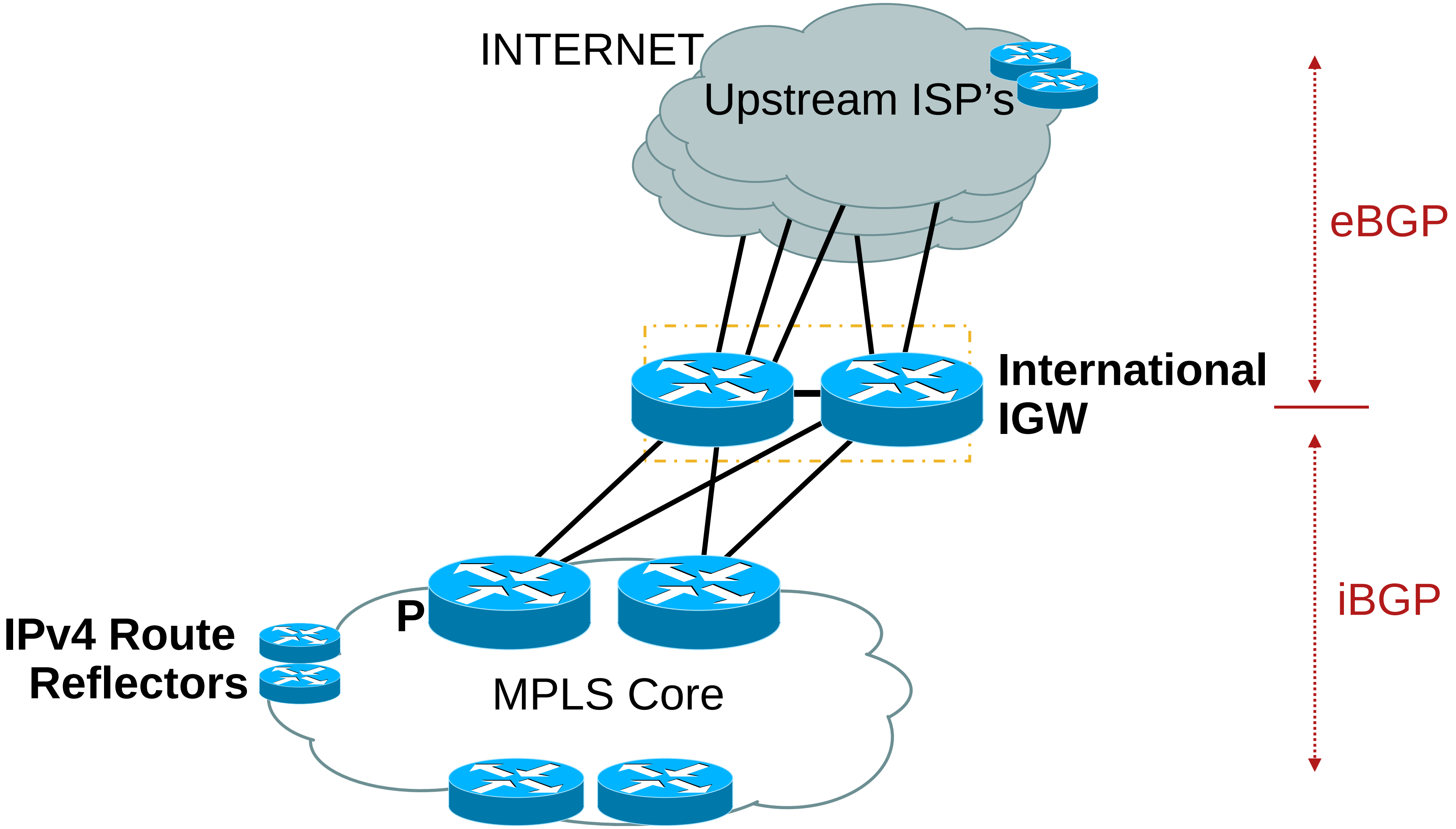
ISP design – peering layer



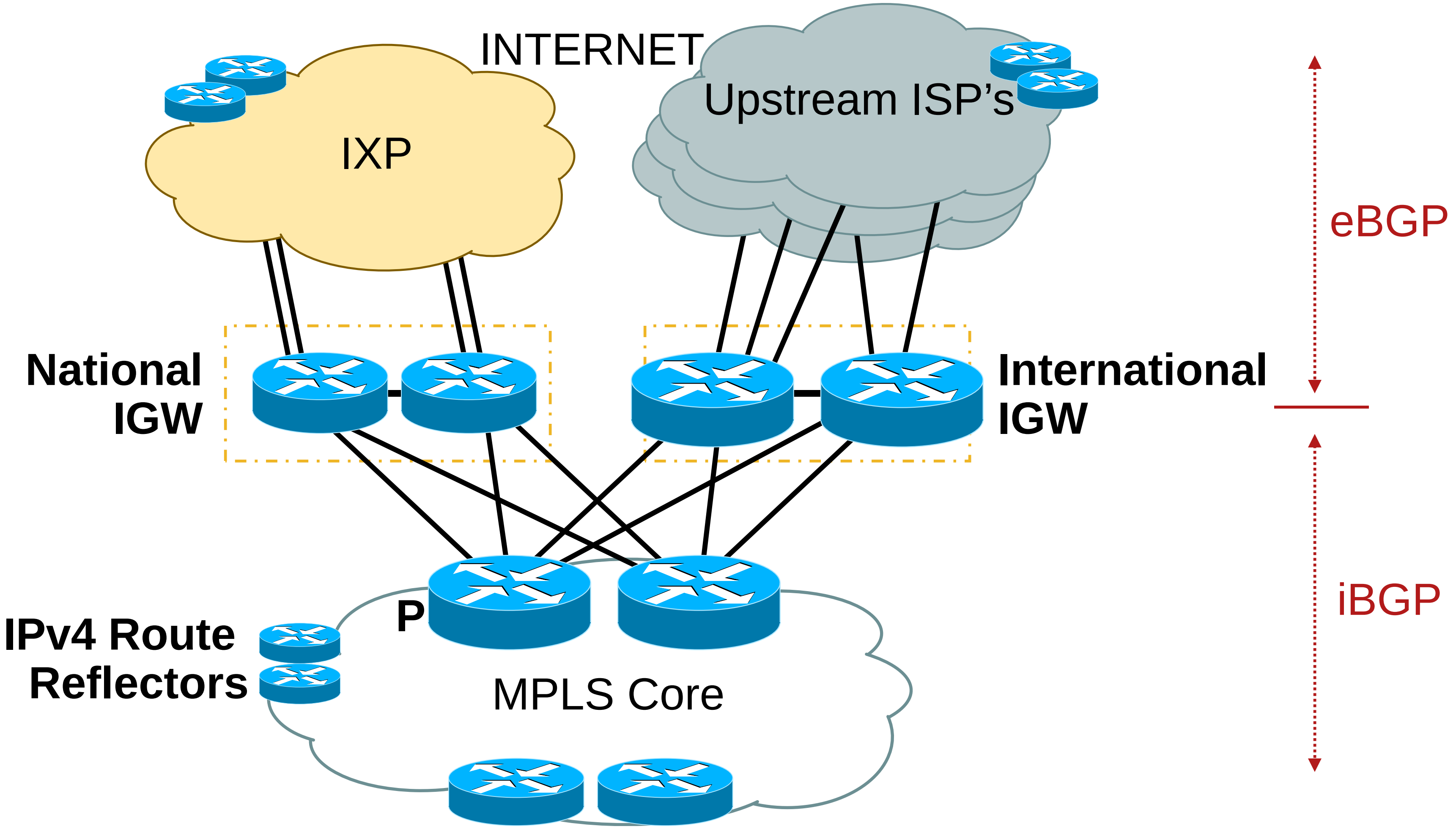
ISP design – peering layer



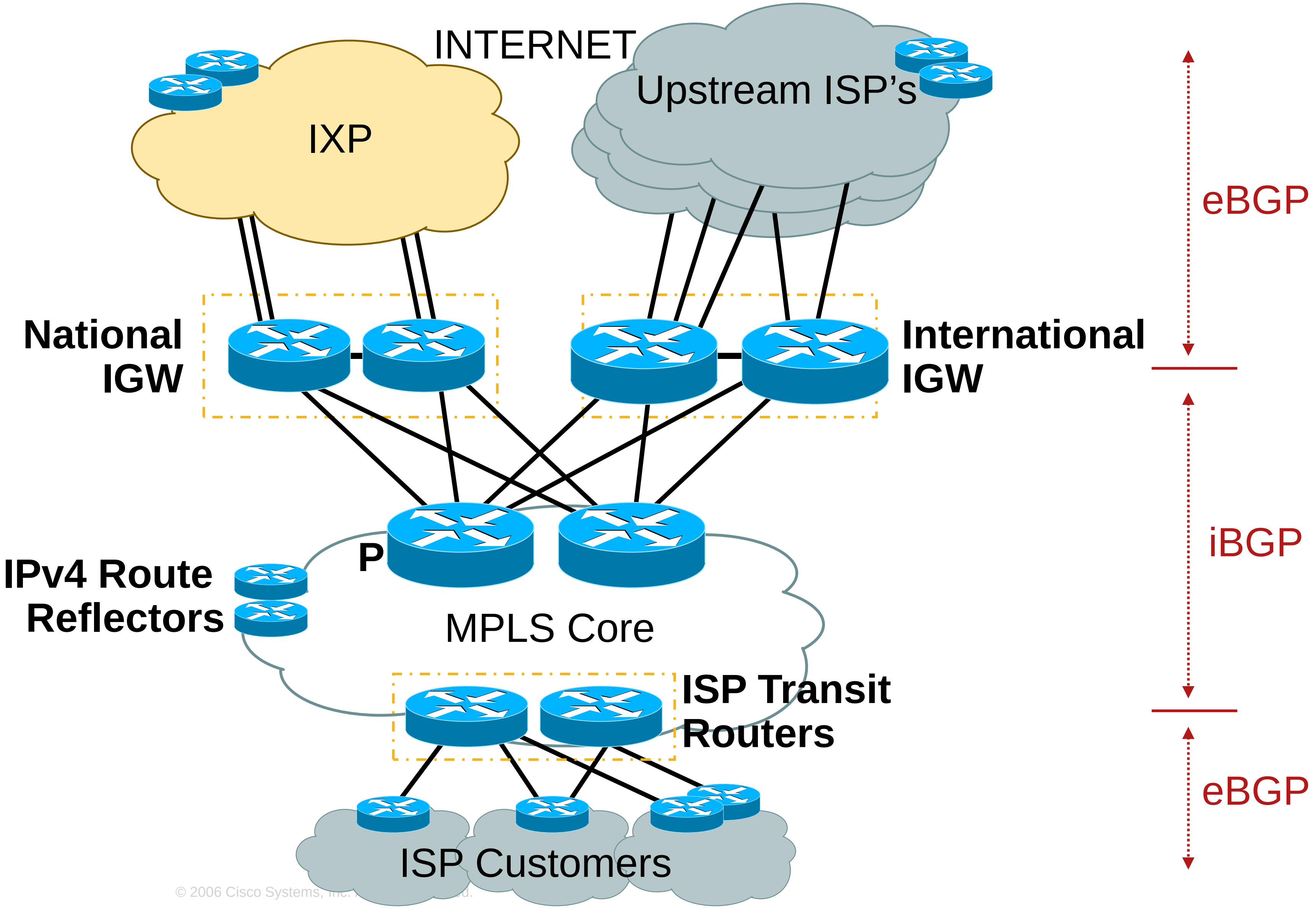
ISP design – peering layer



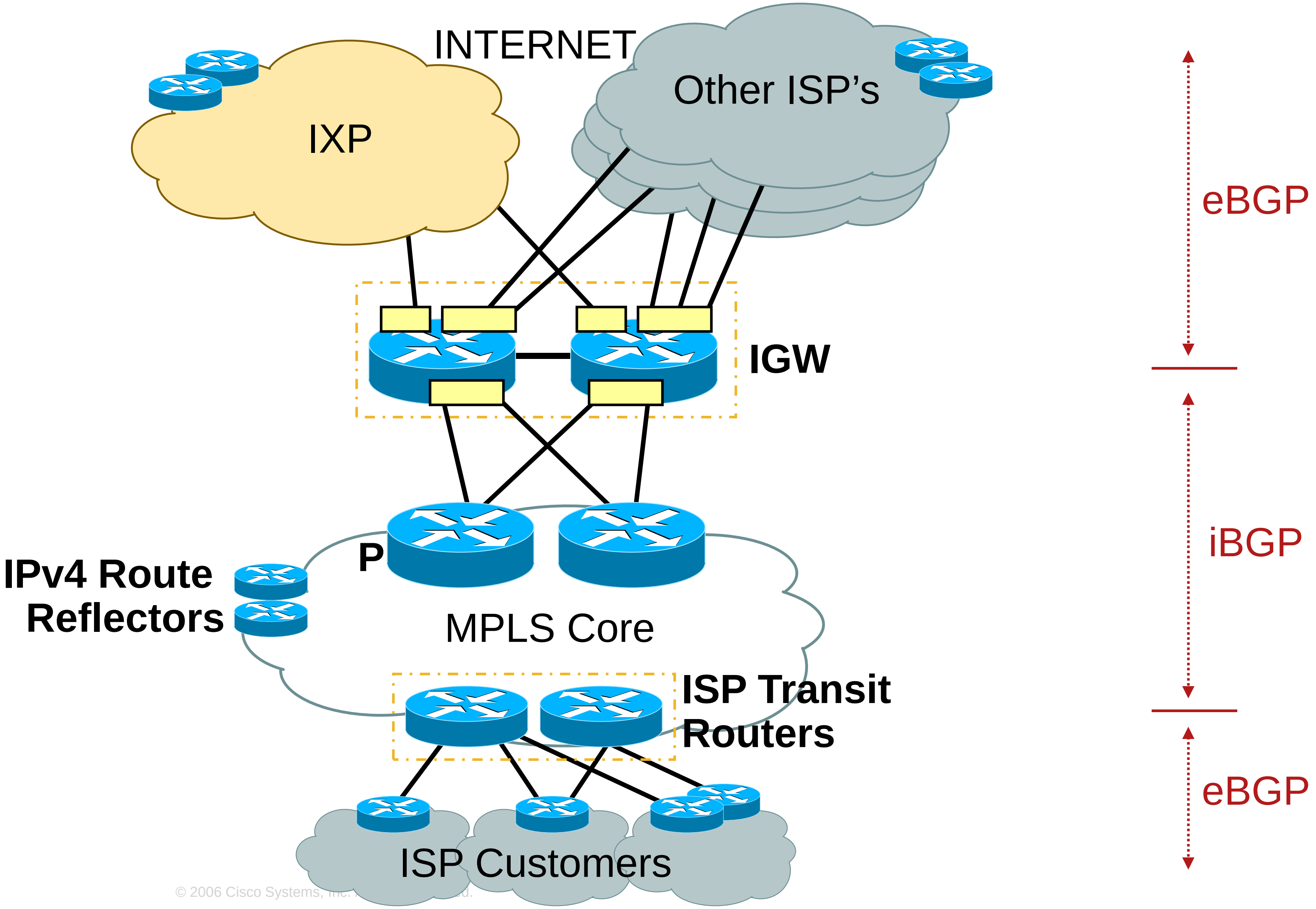
ISP design – peering layer



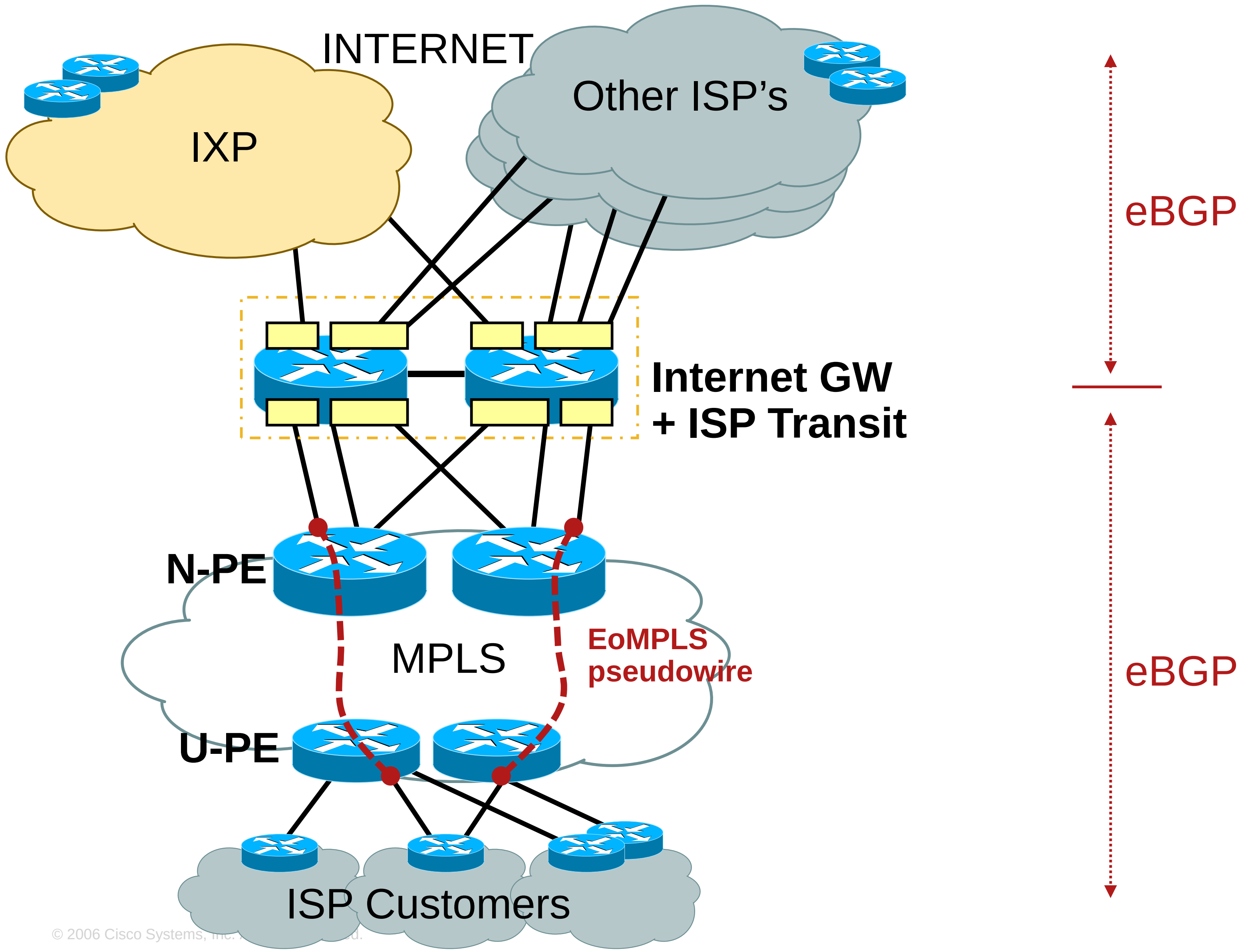
ISP design – peering layer



ISP design – peering layer

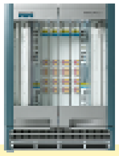


ISP design – peering layer



Internet Gateway

Cisco Internet Gateway Routers



	ASR 1000	CRS-1/4	CRS-1/8	CRS-1/16	CRS-1 MC
Throughput	20 Gbps	320 Gbps	640 Gbps	1.28 Tbps	10 Tbps
Scalability	40 Gbps	960 Gbps	1.92 Tbps	3.84 Tbps	100+ Tbps
FIB entries	2 Million	2 Million	2 Million	2 Million	2 Million
Netflow entries	2 Million	4 Million	8 Million	16 Million	100+ Million

Existing deployments (~60% marketshare)

- The most used ISP GW is Cisco 12000 (GSR)
- Many deployments are based on Cisco 7600
- Many small IGW’s are still Cisco 7200

IGW – Essential Feature set

Broad LAN and WAN interfaces support

- international links – POS STM-1/4/16/64
- national links – GE, 10GE, future full-rate 100GE

IPv4 and IPv6 Routing and Forwarding

- 2M hardware entries (IPv4 + IPv6) – no compression tricks!
- BGP, OSPF/ISIS, BFD – fast, prefix-independent convergence

IPv4 and IPv6 filters (access-lists)

- thousands of L3/L4 entries (IPv4 + IPv6) – no impact on forwarding rate!
- loose uRPF (Unicast Reverse Path Forwarding)

IPv4 and IPv6 netflow monotoring

- at least 1:1000 sampling rate, V9 export

DDoS attack protection and Control Plane protection

- in-hardware protection of router’s brain
- anti-hacking tools – management plane protection

IGW – some optional features

MPLS support

- rarely used on IGW, but sometimes yes
- MPLS Netflow is required too

Traffic Shaping with RED – per-interface or per-VLAN

- if the circuit runs over MAN or ISP subrate service
- shaping prevents unnecessary drops and improves TCP goodput

Accounting

- BGP Policy Accounting – per-AS accounting for large networks
- BGP Policy Propagation – packet marking based on BGP Communities
- MAC accounting – for peering/transit via IXP

Secure Virtualization of the router

- Logical Routers with secure resources allocation

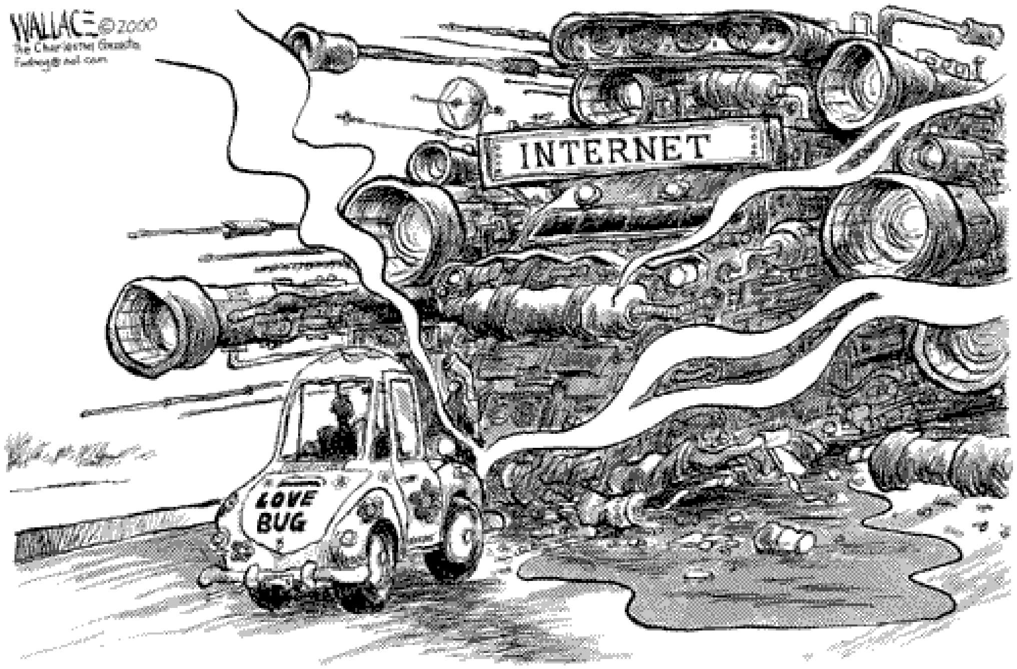
Carrier Grade NAT

- IPv4 exhaustion is close!
- large scale IPv4 NAT and IPv6 AFT with V6 Tunneling is desirable

LI (Lawful Intercept)

- if used as a ISP Transit, LI may be mandatory

ISP Security



Anti-spoofing

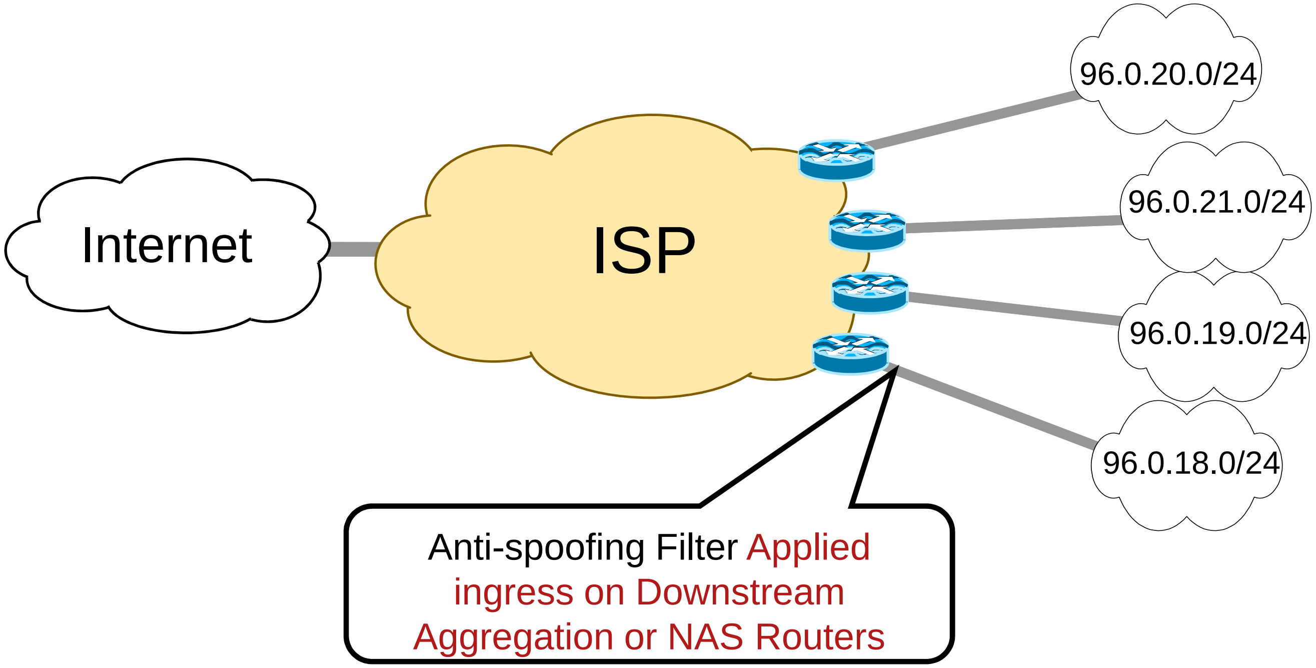
RFC2827/BCP38 Ingress Packet Filtering

Anti-spoofing filter (ingress filter on source IP)

- allow only source addresses from the customer's 96.0.X.X/24
- RFC2827 and RFC3704 (BCP 38 and 84)

Bogon filter (ingress filter on destination IP)

- Drops packets with "insane" destination IP address
- RFC1918, own block, internal IP core, NMS



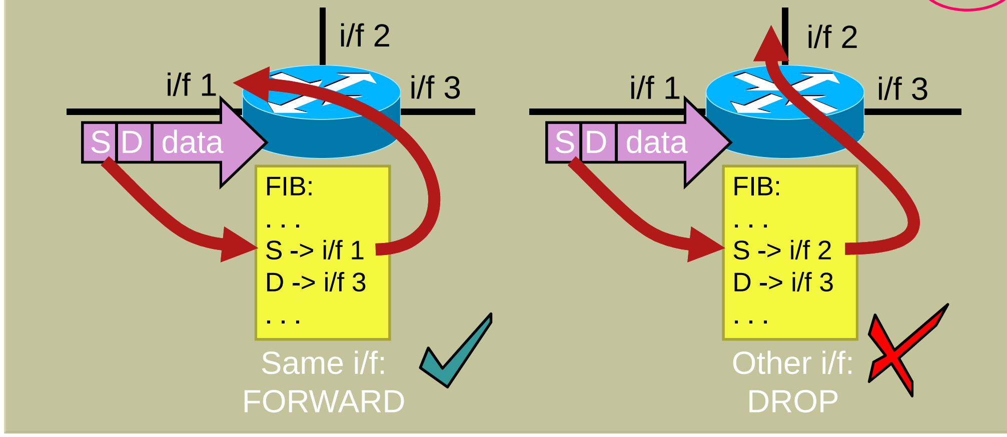
Anti-spoofing Filter Applied ingress on Downstream Aggregation or NAS Routers

ISP's Customer Allocation Block: 96.0.0.0/19

uRPF (Unicast Reverse Path Forwarding)

“Strict Mode” (v1) and “Loose Mode” (v2)

```
router(config-if)# ip verify unicast source reachable-via rx
```



FIB:
. . .
S -> i/f 1
D -> i/f 3
. . .

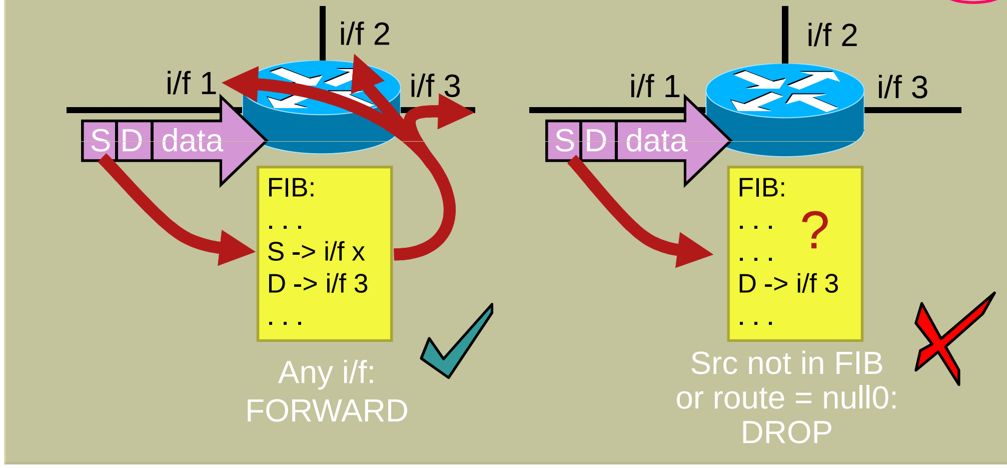
Same i/f:
FORWARD

FIB:
. . .
S -> i/f 2
D -> i/f 3
. . .

Other i/f:
DROP

“Strict Mode”
(aka “v1”)

```
router(config-if)# ip verify unicast source reachable-via any
```



FIB:
. . .
S -> i/f x
D -> i/f 3
. . .

Any i/f:
FORWARD

FIB:
. . . ?
. . .
D -> i/f 3
. . .

Src not in FIB
or route = null0:
DROP

“Loose Mode”
(aka “v2”)

Bogons

- A Bogon prefix is a route that should never appear in the Internet routing table
- Different from DSUA.

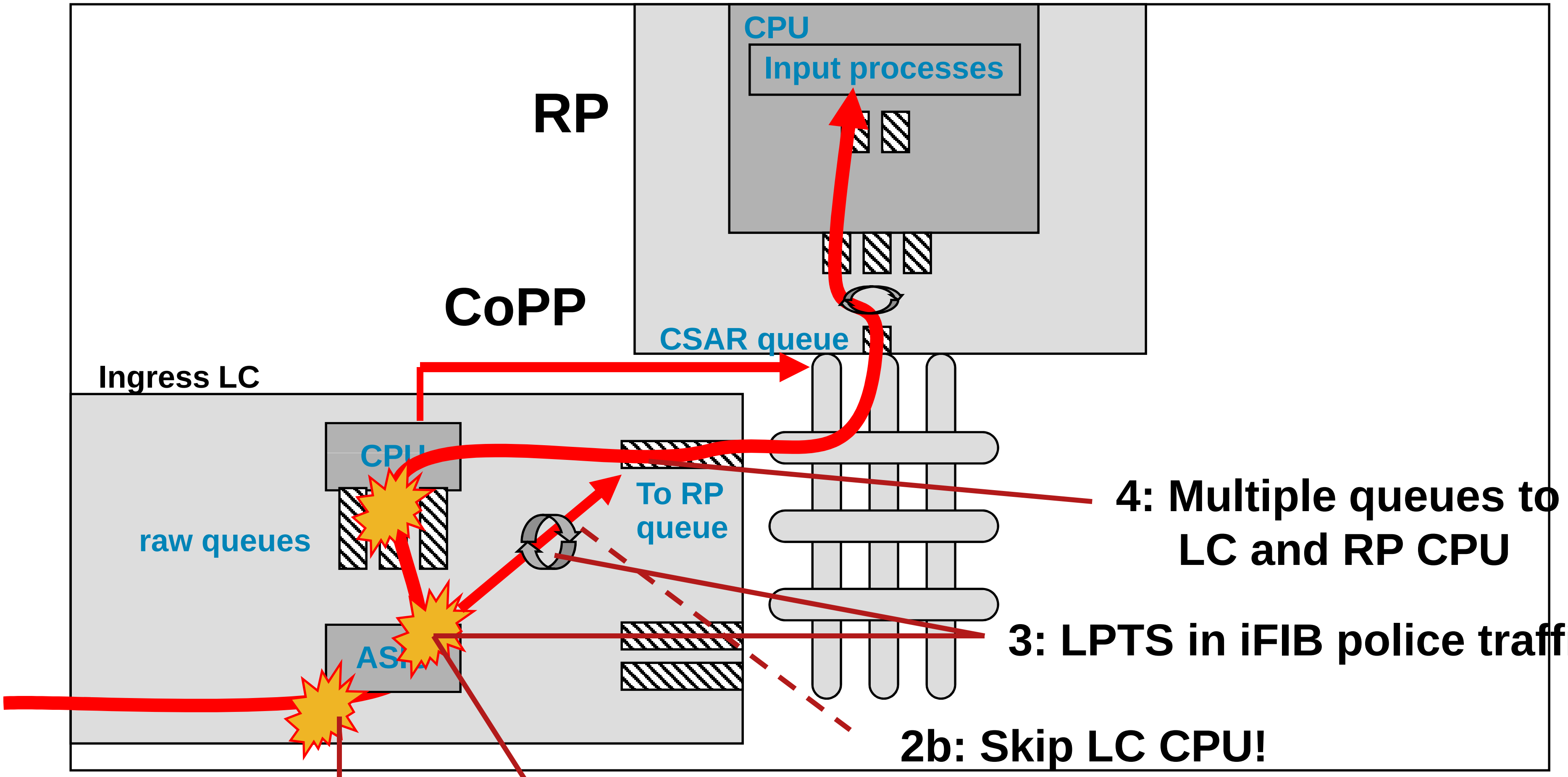
  Bogons are defined as Martians (private and reserved addresses defined by RFC 1918 and RFC 3330) and netblocks that have not been allocated to a (RIR) by IANA
- CYMRU maintains list of Bogons, works with IANA and RIR etc.
- http://www.cymru.com/Bogons/index.html
- BOGON List Keeps on Changing as IANA allocates routes.

  BE AWARE!

  The bogon prefixes are announced unaggregated by the bogon route-servers is **65333:888**; as of 14 JUL 2008 this includes **45** prefixes
- BOGON Router Server.

  Peer with CYMRU Route Server keep BOGON list upto date.

Hardware protection against DOS attacks
CRS-1 Control Plane Protection



4: Multiple queues to LC and RP CPU

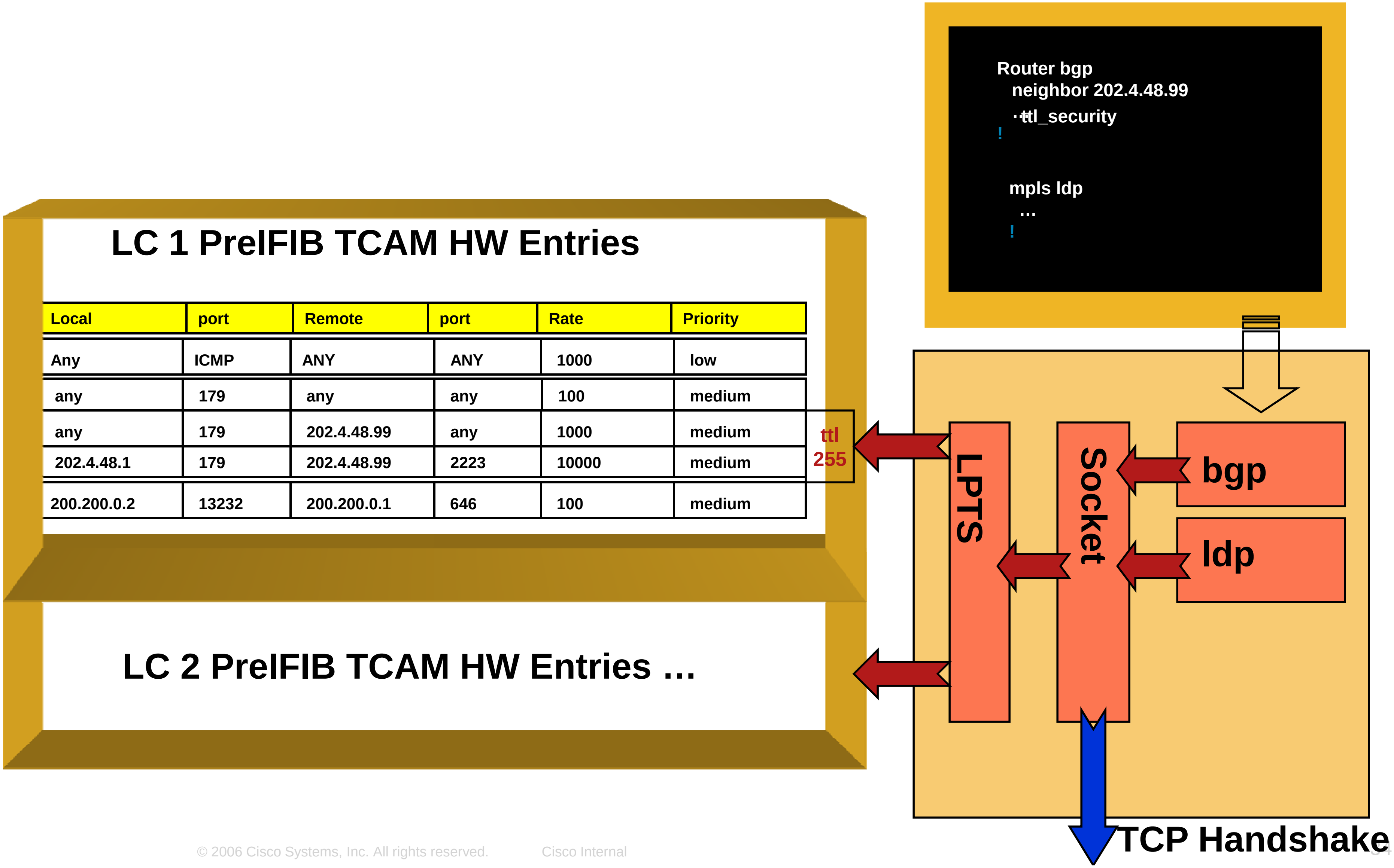
3: LPTS in iFIB police traffic

2b: Skip LC CPU!

2a: LPTS iFIB lookup (Match, BTSH/GTSM)

1: Ingress iACL, uRPF

IOS XR – Dynamic Control Plane Protection



```
Router bgp
  neighbor 202.4.48.99
   ·ttl_security
!

 mpls ldp
   ...
 !
```

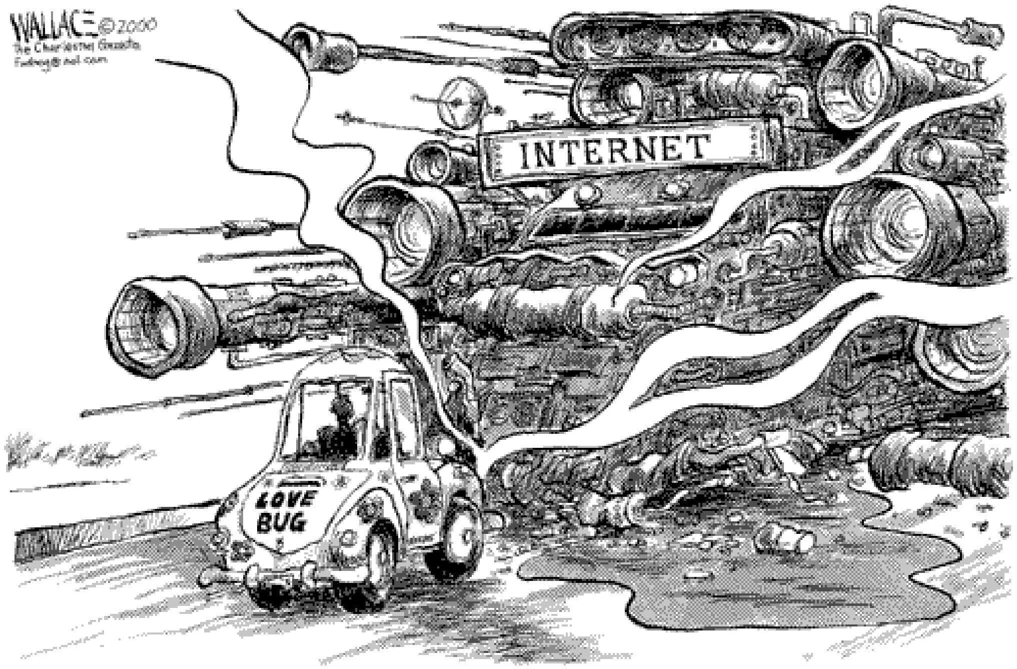
LC 1 PreIFIB TCAM HW Entries

Local	port	Remote	port	Rate	Priority
Any	ICMP	ANY	ANY	1000	low
any	179	any	any	100	medium
any	179	202.4.48.99	any	1000	medium
202.4.48.1	179	202.4.48.99	2223	10000	medium
200.200.0.2	13232	200.200.0.1	646	100	medium

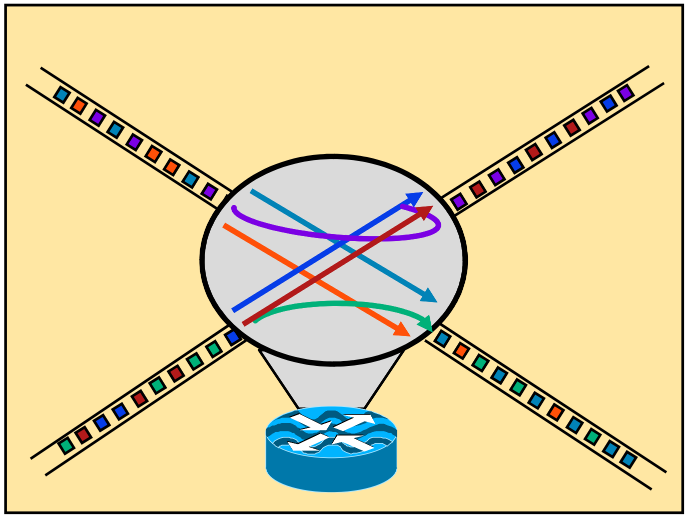
LC 2 PreIFIB TCAM HW Entries …

Detecting an attack:

Netflow



Netflow is a Security tool #1 today!



7 Keys define a flow

Source Address, Destination Address, Source Port, Destination Port, Layer 3 Protocol Type, TOS byte (DSCP), Input Logical Interface (ifIndex)

A flow is unidirectional

Turning it on (generic):

```
interface GigabitEthernet 1/1/1
  ip route-cache flow [sampled]
```

Export (optional):

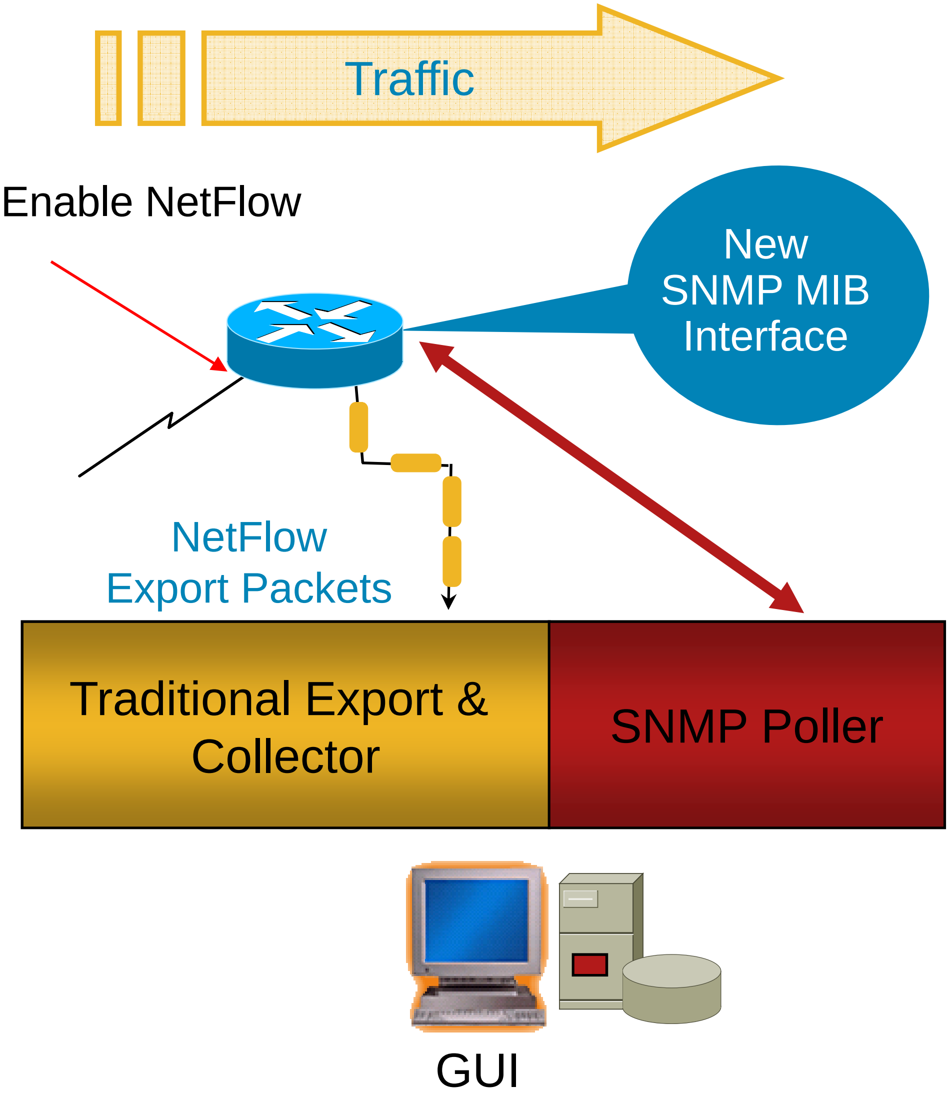
```
ip flow-export destination 172.17.246.225 9995
```

Sampled Netflow (mostly used for Security):

```
ip flow-sampling-mode packet-interval x
```

Flow Is Defined By Seven Unique Keys

- Source IP address
- Destination IP address
- Source port
- Destination port
- Layer 3 protocol type
- TOS byte (DSCP)
- Input logical interface (ifIndex)



Traffic

Enable NetFlow

New SNMP MIB Interface

NetFlow Export Packets

Traditional Export & Collector

SNMP Poller

GUI

© 2004 Cisco Systems, Inc. All rights reserved.

NetFlow Cache Example

1. Create and update flows in NetFlow cache

SrcIf	SrcIPadd	DstIf	DstIPadd	Protocol	TOS	Flgs	Pkts	Src Port	Src Msk	Src AS	Dst Port	Dst Msk	Dst AS	NextHop	Bytes/ Pkt	Active	Idle
Fa1/0	173.100.21.2	Fa0/0	10.0.227.12	11	80	10	11000	00A2	/24	5	00A2	/24	15	10.0.23.2	1528	1745	4
Fa1/0	173.100.3.2	Fa0/0	10.0.227.12	6	40	0	2491	15	/26	196	15	/24	15	10.0.23.2	740	41.5	1
Fa1/0	173.100.20.2	Fa0/0	10.0.227.12	11	80	10	10000	00A1	/24	180	00A1	/24	15	10.0.23.2	1428	1145.5	3
Fa1/0	173.100.6.2	Fa0/0	10.0.227.12	6	40	0	2210	19	/30	180	19	/24	15	10.0.23.2	1040	24.5	14

2. Expiration

- Inactive timer expired (15 sec is default)
- Active timer expired (30 min (1800 sec) is default)
- NetFlow cache is full (oldest flows are expired)
- RST or FIN TCP Flag

SrcIf	SrcIPadd	DstIf	DstIPadd	Protocol	TOS	Flgs	Pkts	Src Port	Src Msk	Src AS	Dst Port	Dst Msk	Dst AS	NextHop	Bytes/ Pkt	Active	Idle
Fa1/0	173.100.21.2	Fa0/0	10.0.227.12	11	80	10	11000	00A2	/24	5	00A2	/24	15	10.0.23.2	1528	1800	4

No

Yes

3. Aggregation

4. Export version

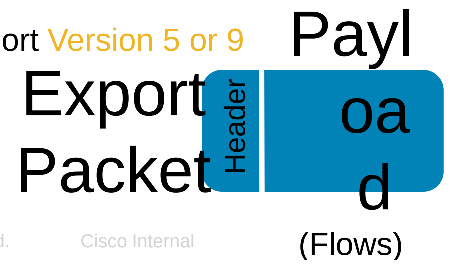
Non-Aggregated Flows—Export Version 5 or 9

e.g. Protocol-Port Aggregation Scheme Becomes

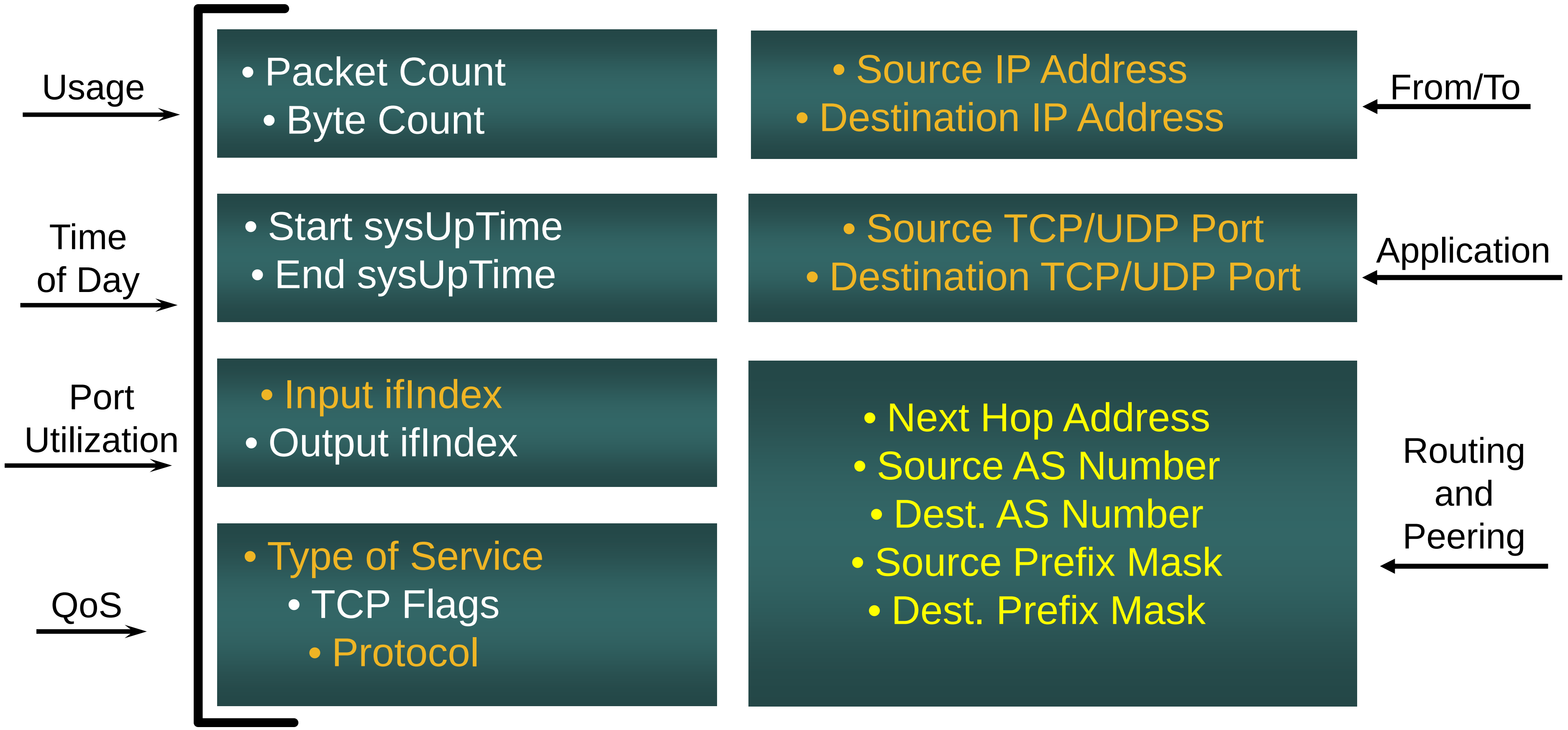
Protocol	Pkts	SrcPort	DstPort	Bytes/Pkt
11	11000	00A2	00A2	1528

Aggregated Flows—Export Version 8 or 9

5. Transport protocol



Netlow Export – V5 fixed format



Usage

- Packet Count
- Byte Count

Time of Day

- Start sysUpTime
- End sysUpTime

Port Utilization

- Input ifIndex
- Output ifIndex

QoS

- Type of Service
- TCP Flags
- Protocol

- Source IP Address
- Destination IP Address

From/To

- Source TCP/UDP Port
- Destination TCP/UDP Port

Application

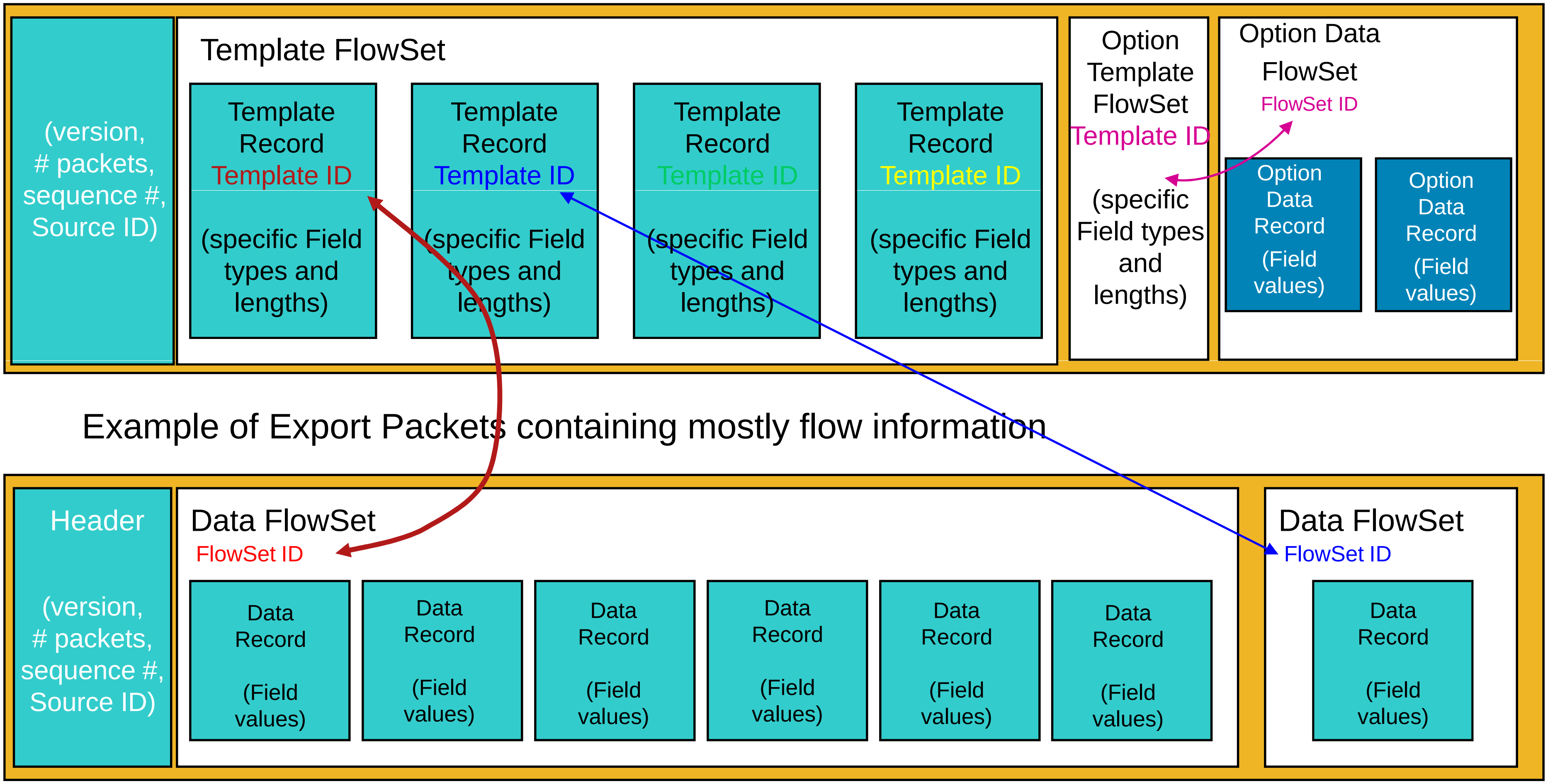
- Next Hop Address
- Source AS Number
- Dest. AS Number
- Source Prefix Mask
- Dest. Prefix Mask

Routing and Peering

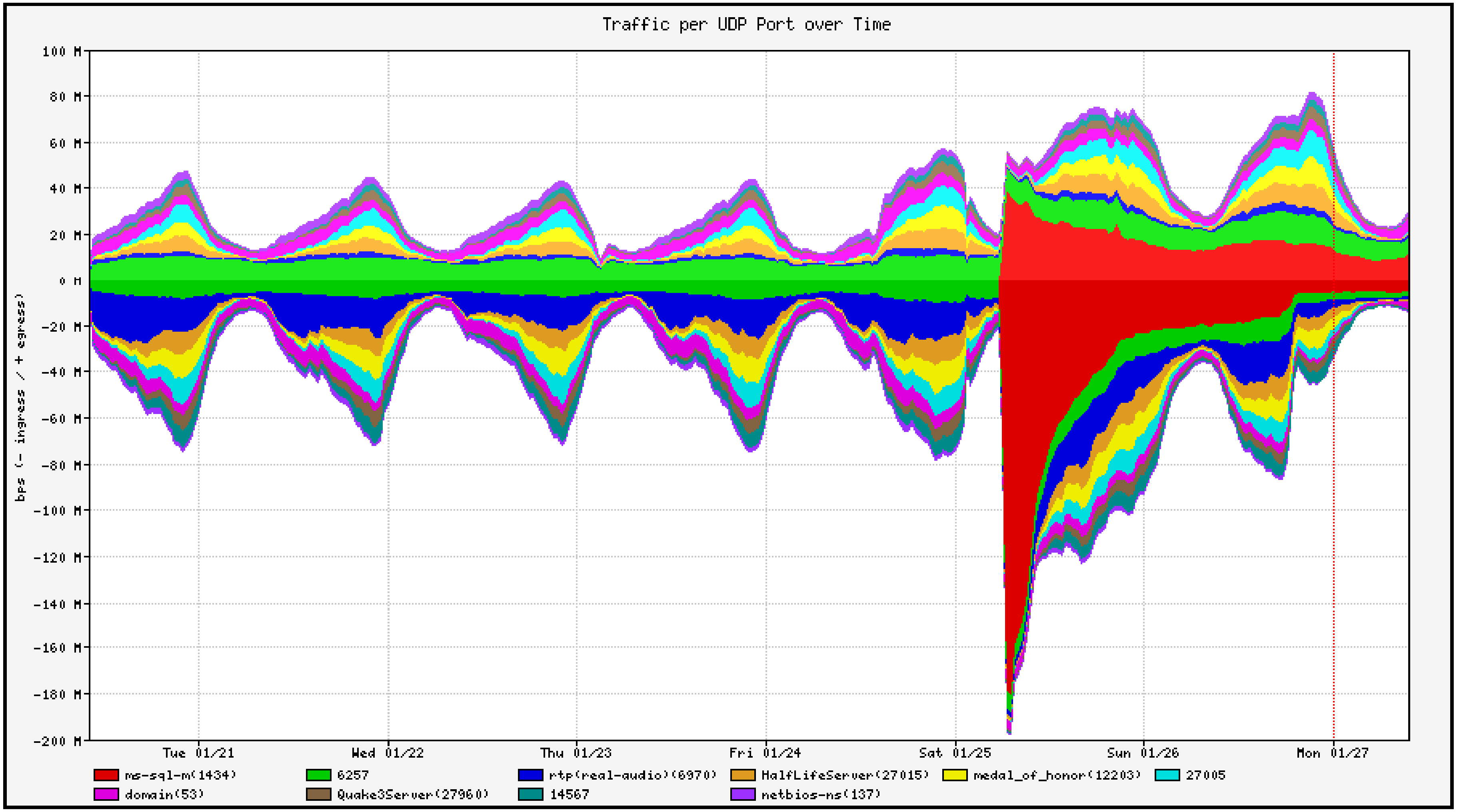
Version 5 used extensively today

NetFlow Export – V9 flexible format

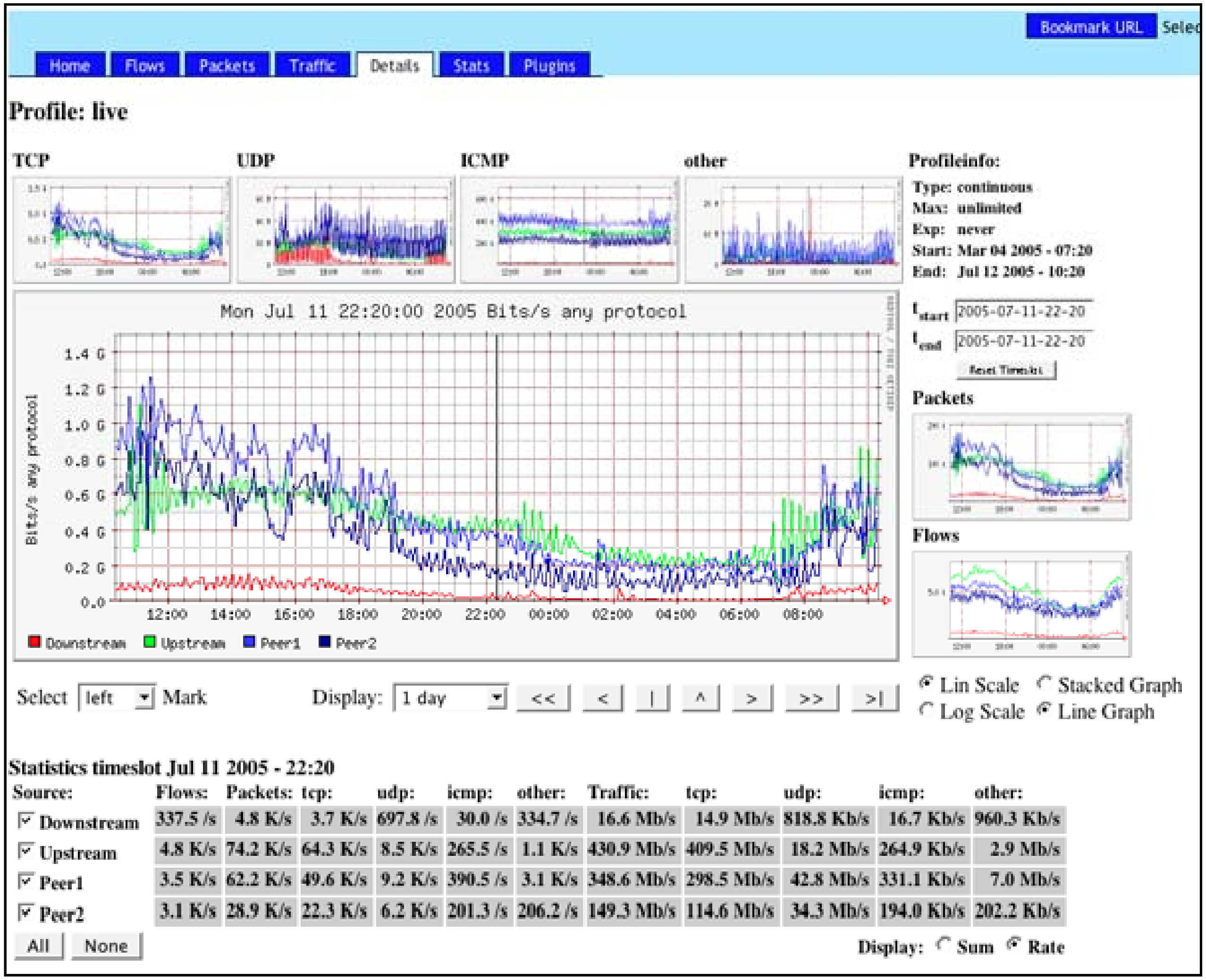
Example of Export Packet right after router boot or NetFlow configuration


(version, # packets, sequence #, Source ID)
Template FlowSet
Template Record Template ID (specific Field types and lengths)
Template Record Template ID (specific Field types and lengths)
Template Record Template ID (specific Field types and lengths)
Template Record Template ID (specific Field types and lengths)
Option Template FlowSet
Template ID
(specific Field types and lengths)
Option Data FlowSet
FlowSet ID
Option Data Record (Field values)
Option Data Record (Field values)
Example of Export Packets containing mostly flow information
Header
(version, # packets, sequence #, Source ID)
Data FlowSet
FlowSet ID
Data Record (Field values)
Data Record (Field values)
Data Record (Field values)
Data Record (Field values)
Data Record (Field values)
Data Record (Field values)
Data FlowSet
FlowSet ID
Data Record (Field values)

Example—What is an Anomaly?



NetFlow—nfdump and nfsen



Profile: live

Profileinfo:
Type: continuous
Max: unlimited
Exp: never
Start: Mar 04 2005 - 07:20
End: Jul 12 2005 - 10:20

t_{start} 2005-07-11-22-20
t_{end} 2005-07-11-22-20

Select left Mark   Display: 1 day

● Lin Scale ○ Stacked Graph
○ Log Scale ● Line Graph

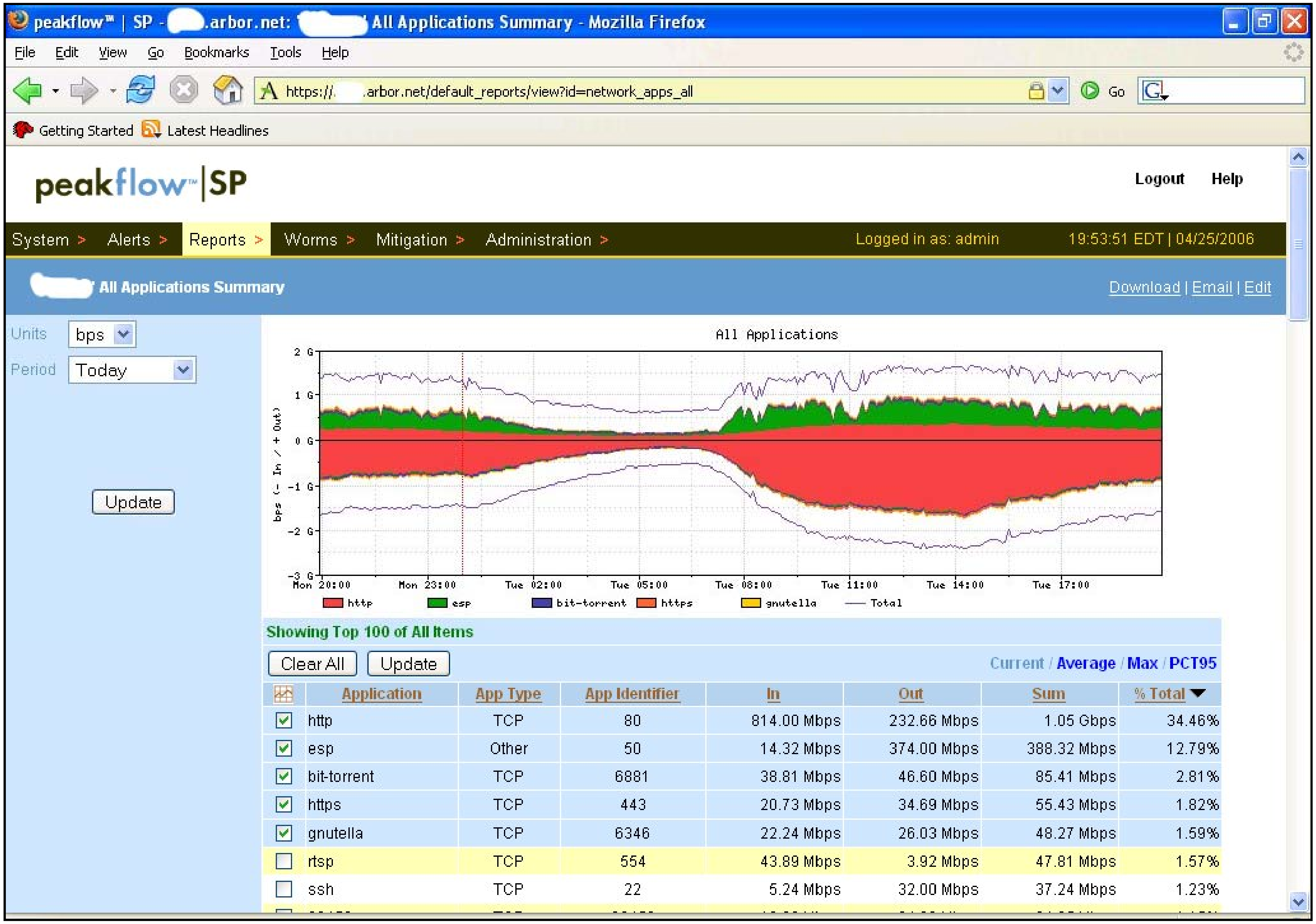
Statistics timeslot Jul 11 2005 - 22:20

Source:	Flows:	Packets:	tcp:	udp:	icmp:	other:	Traffic:	tcp:	udp:	icmp:	other:
☑ Downstream	337.5 /s	4.8 K/s	3.7 K/s	697.8 /s	30.0 /s	334.7 /s	16.6 Mb/s	14.9 Mb/s	818.8 Kb/s	16.7 Kb/s	960.3 Kb/s
☑ Upstream	4.8 K/s	74.2 K/s	64.3 K/s	8.5 K/s	265.5 /s	1.1 K/s	430.9 Mb/s	409.5 Mb/s	18.2 Mb/s	264.9 Kb/s	2.9 Mb/s
☑ Peer1	3.5 K/s	62.2 K/s	49.6 K/s	9.2 K/s	390.5 /s	3.1 K/s	348.6 Mb/s	298.5 Mb/s	42.8 Mb/s	331.1 Kb/s	7.0 Mb/s
☑ Peer2	3.1 K/s	28.9 K/s	22.3 K/s	6.2 K/s	201.3 /s	206.2 /s	149.3 Mb/s	114.6 Mb/s	34.3 Mb/s	194.0 Kb/s	202.2 Kb/s

Display: ○ Sum ● Rate

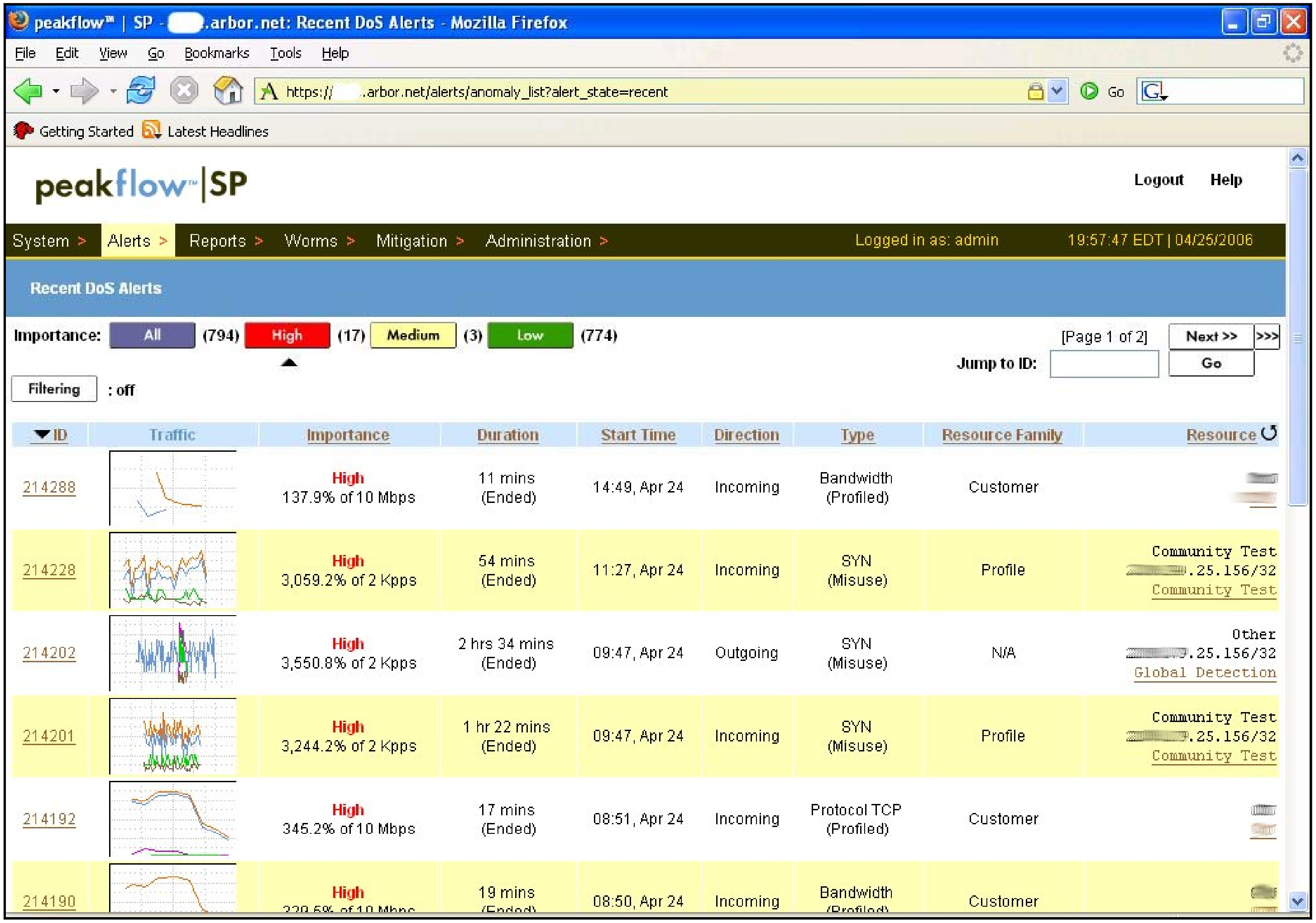
Source: http://nfsen.sourceforge.net, ev. http://software.uninett.no/stager/

Arbor Peakflow SP — Application Distribution



	Application	App Type	App Identifier	In	Out	Sum	% Total ▼
☑	http	TCP	80	814.00 Mbps	232.66 Mbps	1.05 Gbps	34.46%
☑	esp	Other	50	14.32 Mbps	374.00 Mbps	388.32 Mbps	12.79%
☑	bit-torrent	TCP	6881	38.81 Mbps	46.60 Mbps	85.41 Mbps	2.81%
☑	https	TCP	443	20.73 Mbps	34.69 Mbps	55.43 Mbps	1.82%
☑	gnutella	TCP	6346	22.24 Mbps	26.03 Mbps	48.27 Mbps	1.59%
☐	rtsp	TCP	554	43.89 Mbps	3.92 Mbps	47.81 Mbps	1.57%
☐	ssh	TCP	22	5.24 Mbps	32.00 Mbps	37.24 Mbps	1.23%

Example—Arbor Peakflow SP DoS Module


peakflow™ | SP - [redacted].arbor.net: Recent DoS Alerts - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
https://[redacted].arbor.net/alerts/anomaly_list?alert_state=recent
Go
Getting Started
Latest Headlines
peakflow™ | SP
Logout
Help
System >
Alerts >
Reports >
Worms >
Mitigation >
Administration >
Logged in as: admin
19:57:47 EDT | 04/25/2006

Recent DoS Alerts

Importance: All (794) High (17) Medium (3) Low (774)

[Page 1 of 2] Next >> >>>

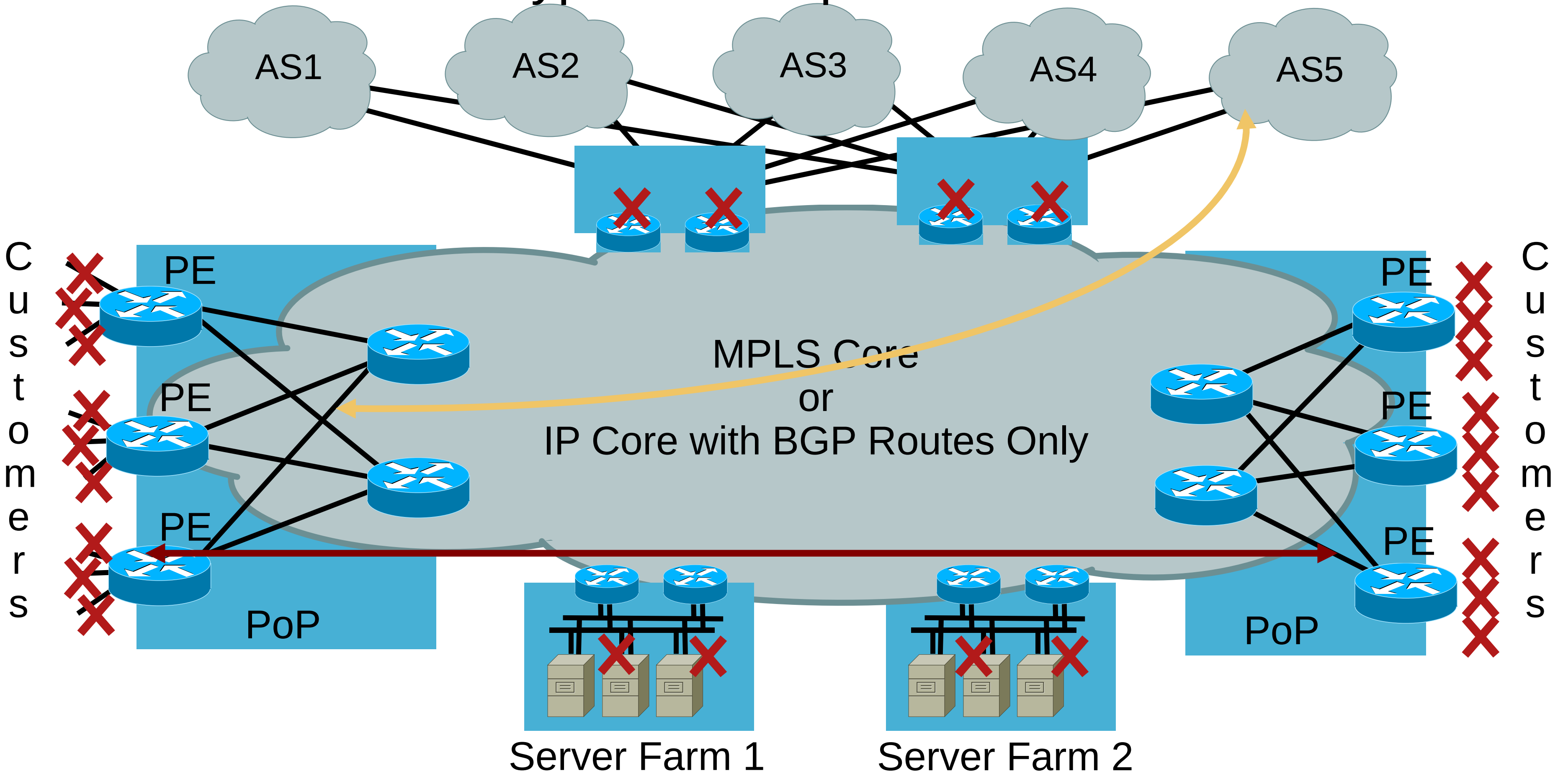
Jump to ID: Go

Filtering : off

ID	Traffic	Importance	Duration	Start Time	Direction	Type	Resource Family	Resource
214288		High 137.9% of 10 Mbps	11 mins (Ended)	14:49, Apr 24	Incoming	Bandwidth (Profiled)	Customer	[redacted]
214228		High 3,059.2% of 2 Kpps	54 mins (Ended)	11:27, Apr 24	Incoming	SYN (Misuse)	Profile	Community Test [redacted].25.156/32 Community Test
214202		High 3,550.8% of 2 Kpps	2 hrs 34 mins (Ended)	09:47, Apr 24	Outgoing	SYN (Misuse)	N/A	Other [redacted].25.156/32 Global Detection
214201		High 3,244.2% of 2 Kpps	1 hr 22 mins (Ended)	09:47, Apr 24	Incoming	SYN (Misuse)	Profile	Community Test [redacted].25.156/32 Community Test
214192		High 345.2% of 10 Mbps	17 mins (Ended)	08:51, Apr 24	Incoming	Protocol TCP (Profiled)	Customer	[redacted]
214190		High 329.5% of 10 Mbps	19 mins (Ended)	08:50, Apr 24	Incoming	Bandwidth (Profiled)	Customer	[redacted]

BGP Next Hop TOS Aggregation

Typical Example

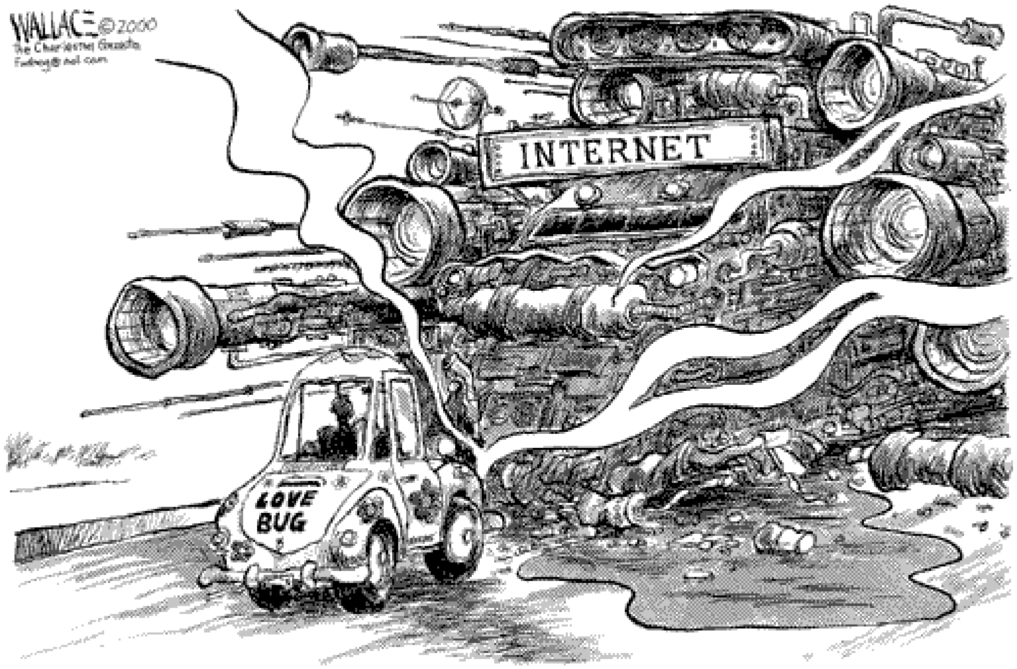


↔ Internal Traffic: “PoP to PoP”

↔ External Traffic Matrix PoP to BGP AS

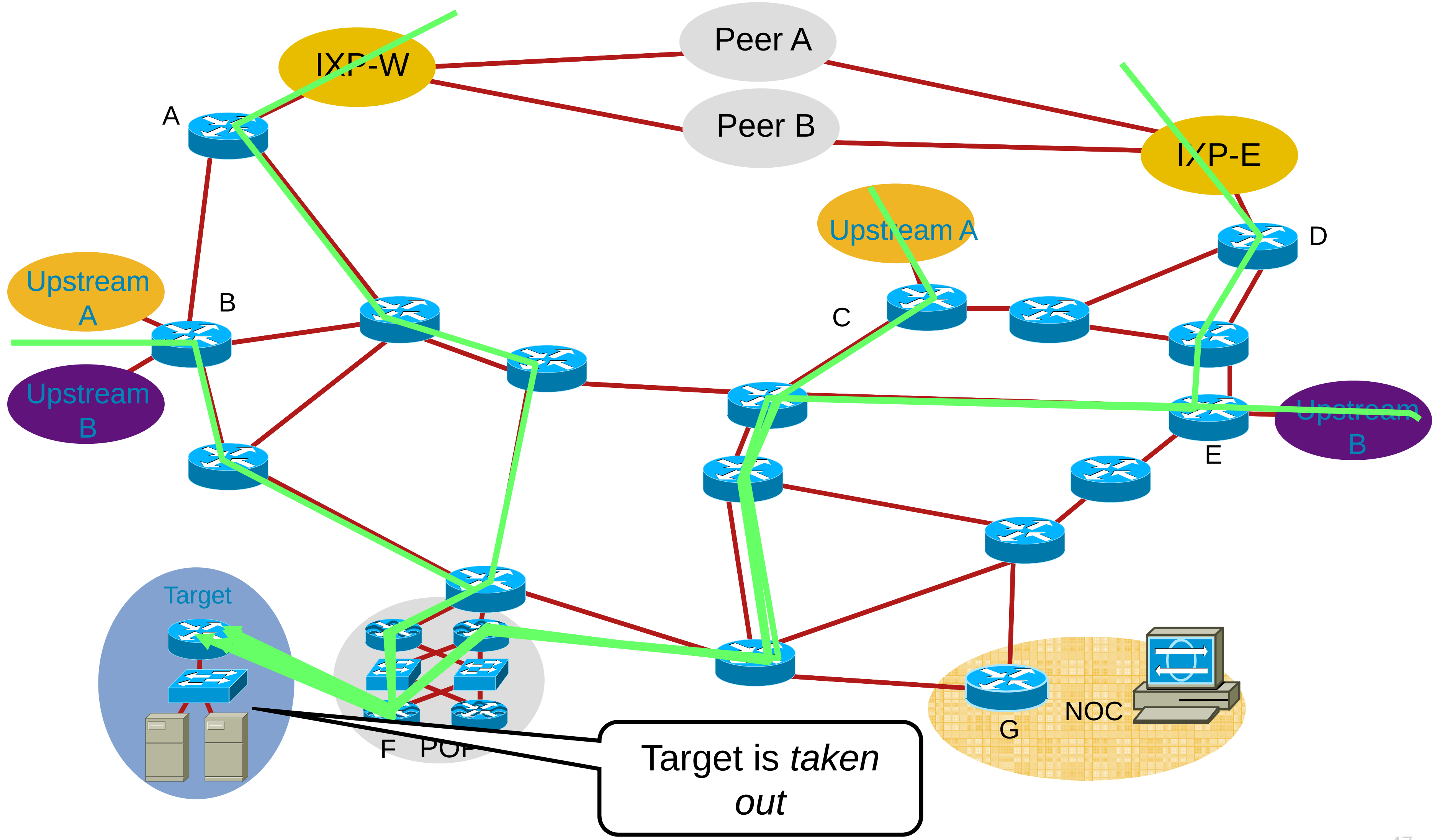
Dropping a DDoS attack:

BGP Blackholing



Customer is DOSed

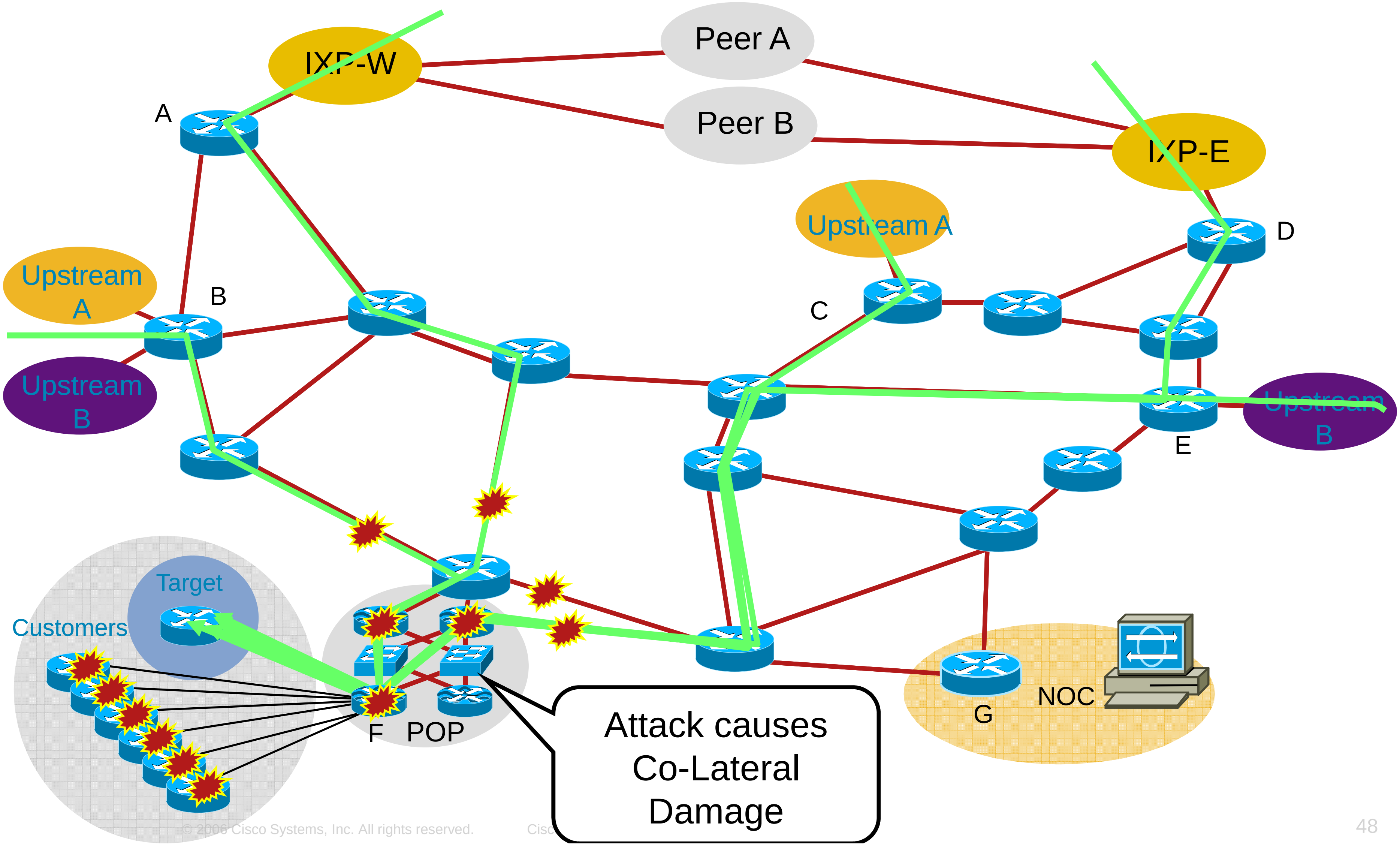
Before



Target is *taken out*

Customer is DOSed

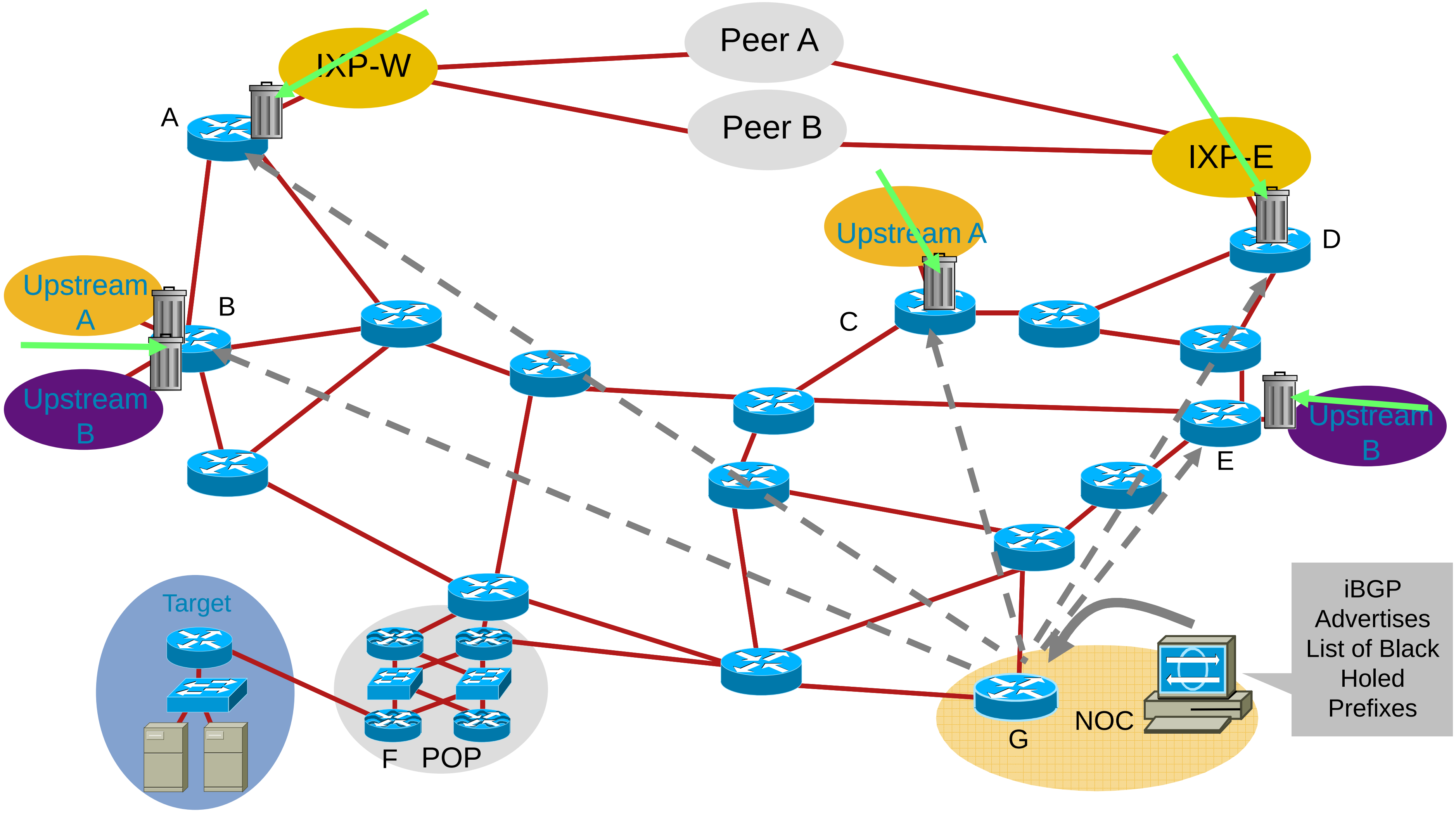
Before – Co-Lateral Damage



Attack causes Co-Lateral Damage

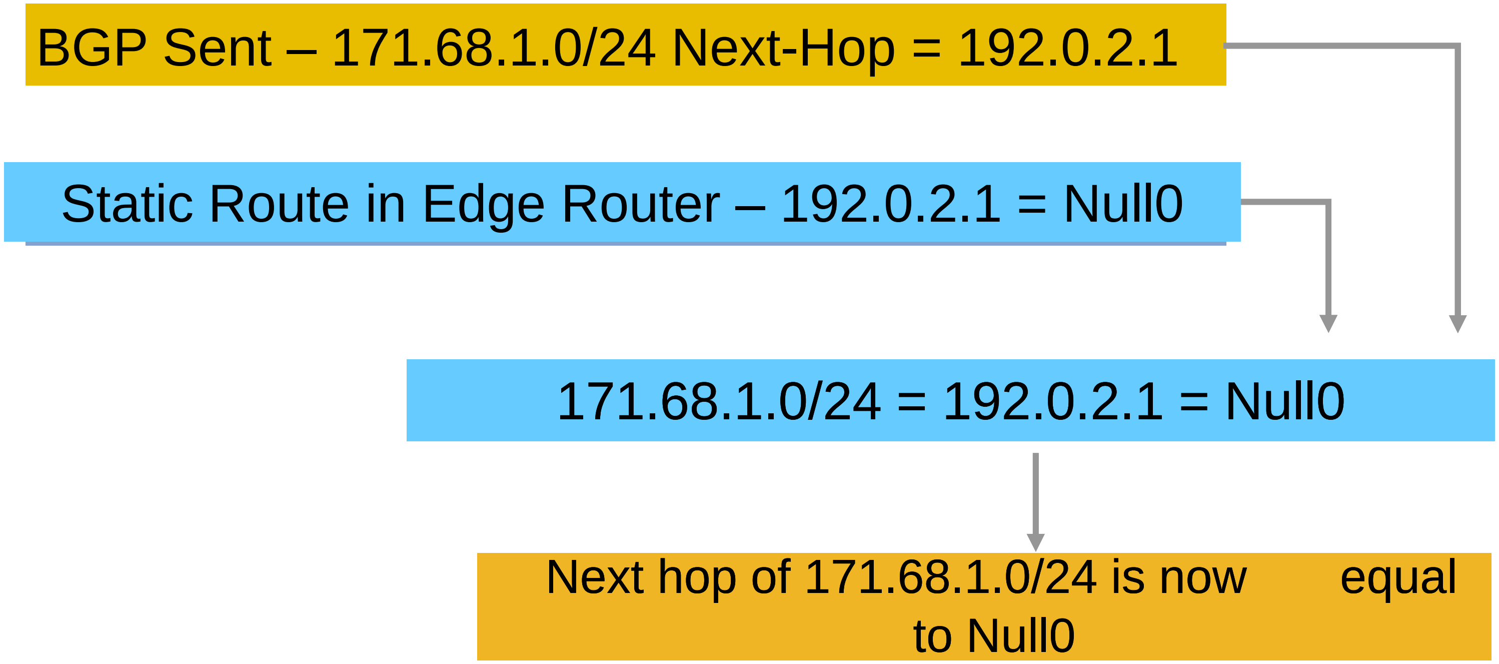
Customer is DOSed

After – Packet Drops Pushed to the Edge



iBGP Advertises List of Black Holed Prefixes

BGP Blackholing: Reacting to an Attack



BGP Sent – 171.68.1.0/24 Next-Hop = 192.0.2.1

Static Route in Edge Router – 192.0.2.1 = Null0

171.68.1.0/24 = 192.0.2.1 = Null0

Next hop of 171.68.1.0/24 is now equal
to Null0

- Remote Triggered Black Hole filtering is the foundation for a whole series of techniques to traceback and react to DDOS attacks on an ISP's network.
- Easy preparation, does not effect ISP operations or performance.
- It does adds the option to an ISP's *security toolkit.*

BGP Blackholing: IOS configuration

- place a host-route to Null on every BGP router

```
ip route 192.0.2.1 255.255.255.255 Null0
```

- prepare a injection into BGP with the blackhole next-hop

```
router bgp 10
 redistribute static route-map set-blackhole

route-map set-blackhole permit 10
 match tag 666
 set ip next-hop 192.0.2.1
 set community 10:666 no-export
 set local-preference 50
```

- simply filter it out everywhere by one command:

```
BH(config)# ip route 1.2.2.2 255.255.255.255 Null0 tag 666
```

BGP Blackholing: Filtering on source IP address

- loose uRPF (unicast reverse path forwarding)

```
ip route 192.0.2.2 255.255.255.255 Null0
int PoS 1/0/0
  ip verify unicast source reachable-via any
```

!!! packet with source IP prefix pointing to Null0 will be dropped !!!

- prepare a injection into BGP with the blackhole next-hop

```
route-map set-blackhole permit 20
 match tag 667
 set ip next-hop 192.0.2.2
 set community 10:667 no-export
 set local-preference 50
```

- simply filter it out everywhere by one command:

```
BH(config)# ip route 1.2.2.3 255.255.255.255 Null0 tag 667
```

BGP Triggered Rate Limiting

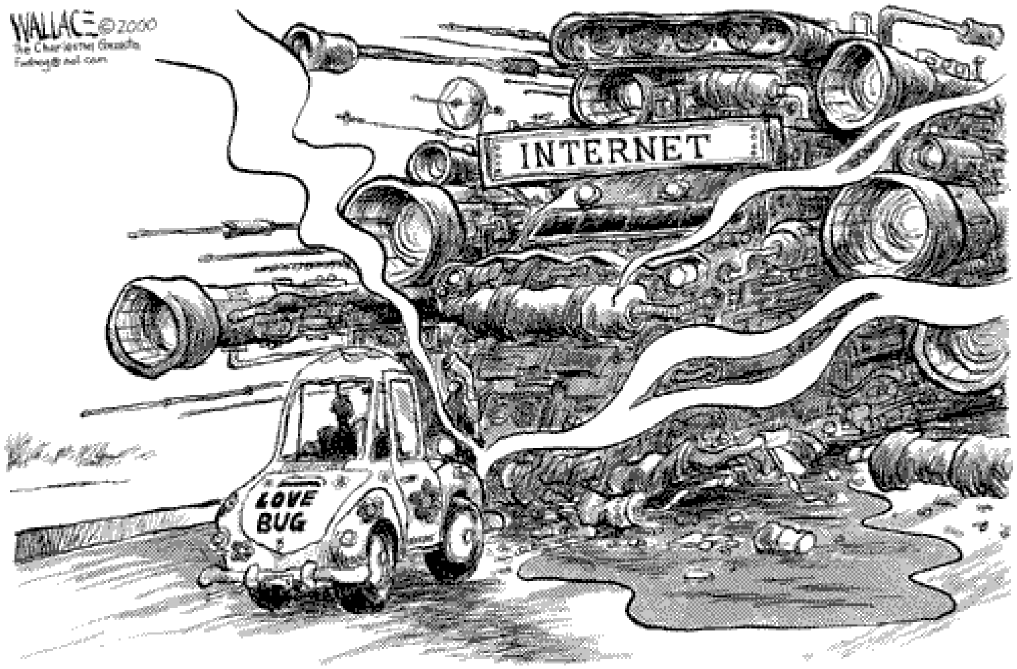
QPPB (QoS Policy Propagation via BGP)

```
router bgp 10
  table-map DOS-Activate
  neighbor 200.200.14.4 remote-as 10
  neighbor 200.200.14.4 update-source Loopback 0
  neighbor 200.200.14.4 send-community
!
ip bgp-community new-format
!
ip community-list 1 permit 10:666
!
route-map DOS-Activate permit 10
  match community 1
  set ip qos-group 66
!
route-map DOS-Activate permit 20
!
interface PoS 0/0/0
  bgp-policy source ip-qos-map
  rate-limit input qos-group 66 256000 8000 8000
      conform-action transmit
      exceed-action drop
```

- **QPPB marking is done before rate-limit or policing**
- **hardware support in Cisco 10000, 12000, CRS-1**

Dark IP space:

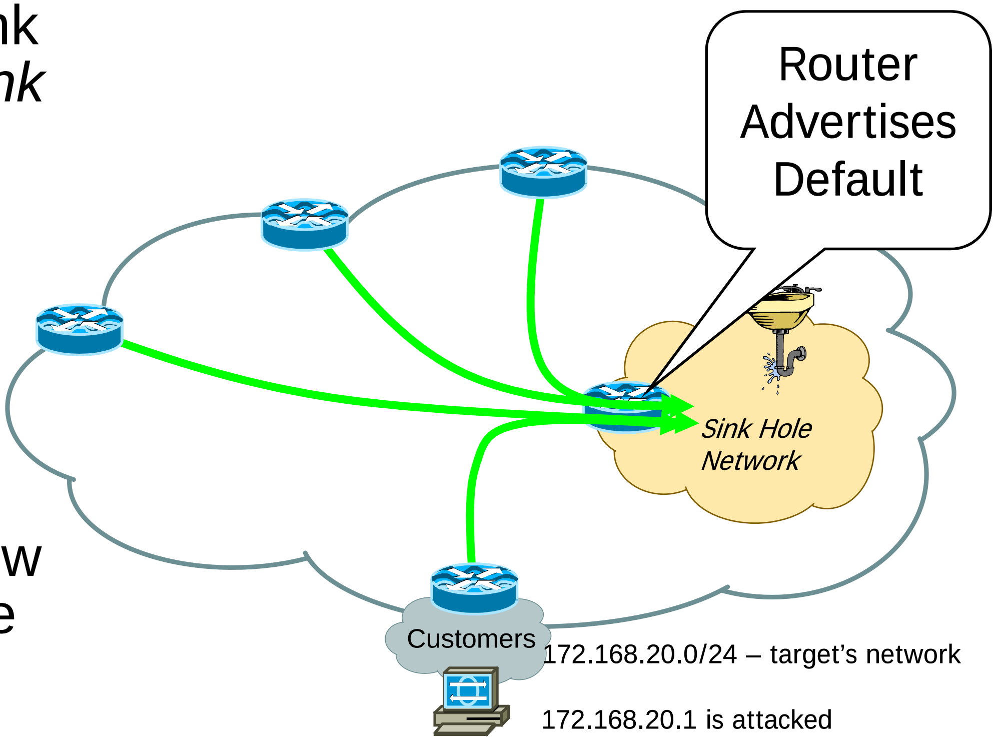
Sinkholes



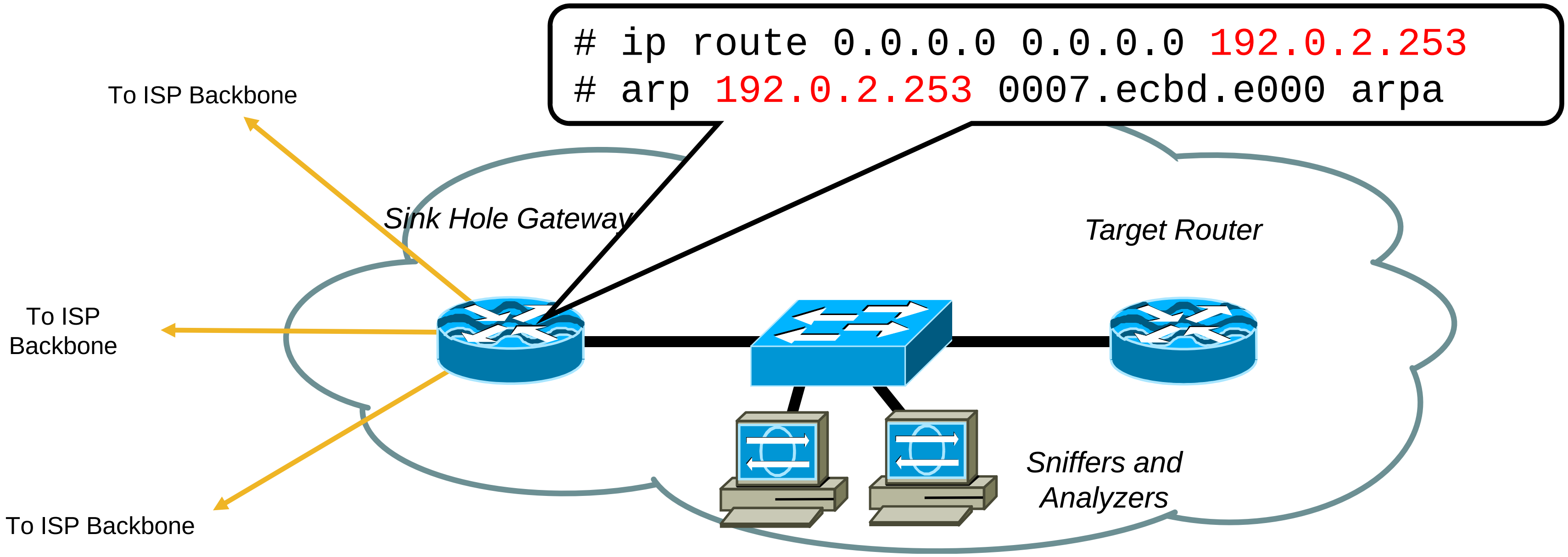
Default Route & the Internet

```
BHole(config-router)# default-information originate always
```

- Advertising Default from the Sink Hole will pull down all sort of *junk* traffic.
  - Customer Traffic when circuits flap.
  - Network Scans
  - Failed Attacks
  - Code Red/NIMDA
  - Backscatter
- Can place tracking tools (Netflow cache) and IDS in the Sink Hole network to monitor the noise.
- BCP: Default should be always a blackhole (Null0 or Static ARP) !!



Target Routers are Expendable



```
# ip route 0.0.0.0 0.0.0.0 192.0.2.253
# arp 192.0.2.253 0007.ecbd.e000 arpa
```

- Sink Hole Gateway Generates the more specific iBGP Announcement.
- Pull the DOS/DDOS attack to the sink hole and forwards the attack to the target router.
- Static ARP to the target router keeps the Sink Hole Operational – Target Router can crash from the attack and the static ARP will keep the gateway forwarding traffic to the ethernet switch.

What to Monitor in a Sinkhole?

- Scans on dark IP (allocated and announced but unassigned address space)

  Who is scoping out the network—pre-attack planning, worms…

- Scans on bogons (unallocated)

  Worms, infected machines, and Bot creation

- Backscatter from spoofed attacks

  Who is getting attacked

  don't use "`no ip icmp unreachables`"

  use "`ip icmp rate-limit unreachables`"

- Backscatter from garbage traffic (RFC-1918 leaks)

  Which customers have mis-configuration or "leaking" networks

Summary & Resources

Summary

- Transit vs. Peering
- The importance of IXP

- Anatomy of the ISP Edge
- Cisco peering platforms and features

- The importance of Netflow
- Basic ISP cecurity techniques

Cisco Networkers
25-28. januar 2010.
Barselona
Registrujte se