

Ruteo externo: BGP

Redes y Servicios Avanzados en Internet

Ruteo Externo
BGP

Nicolás Macia y Alejandro Sabolansky

Agradecimientos: Luis Marrone, Andrés Barbieri

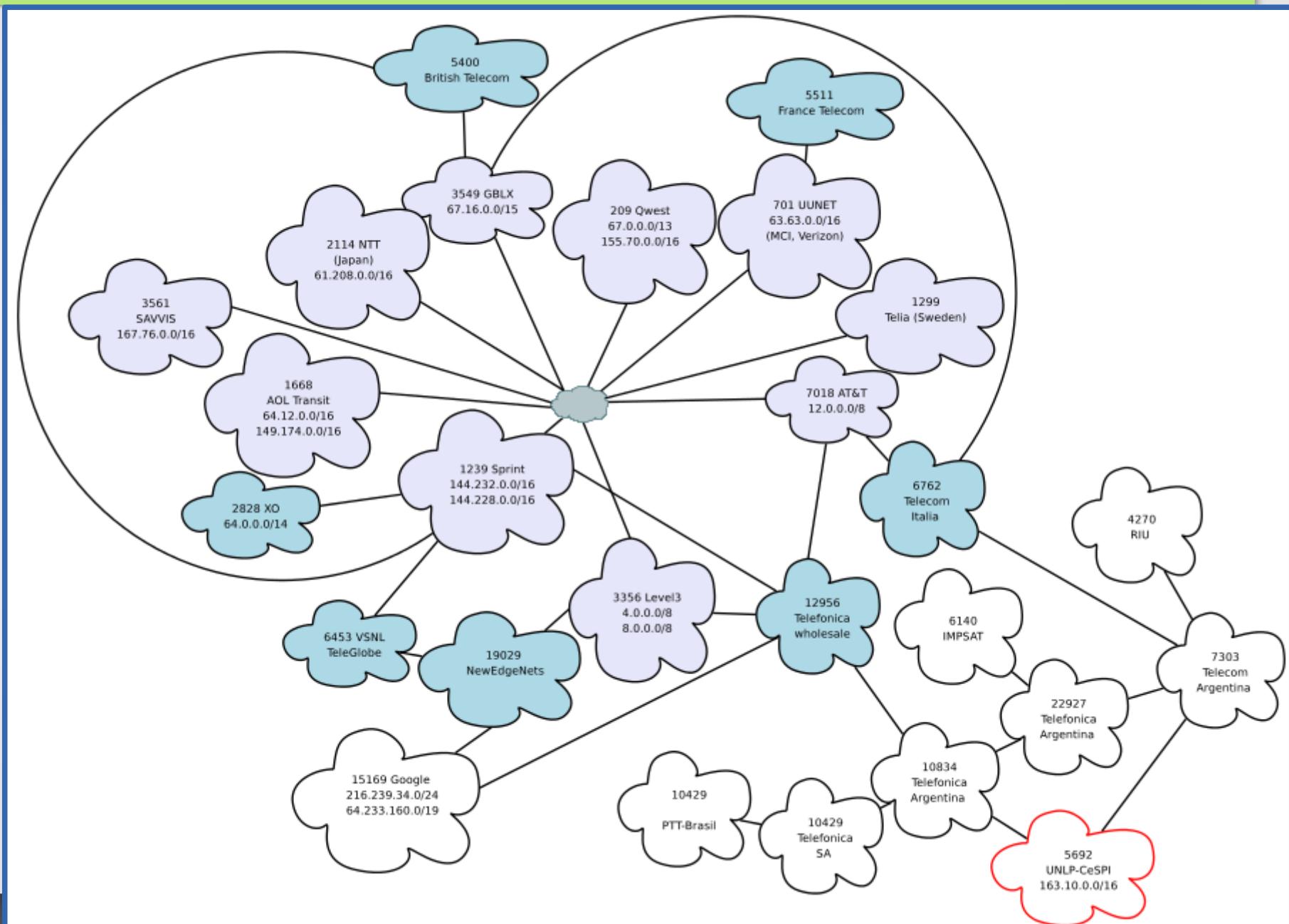
Un poco de historia

- En los 80', Internet era una red pequeña
 - Los routers (llamados por aquel entonces gateways) compartían todos con todos la información de enrutamiento.
 - Para ello se utilizaba el protocolo GGP (Gateway to Gateway Protocol – RFC 823)
 - Calculaba la mejor ruta en base a la cantidad de saltos. Similar a RIP.
 - Debido al crecimiento que tuvo la red, quedó en evidencia su poca escalabilidad.
- Para contener el crecimiento de Internet, el primer paso fue dividir la red en diferentes redes (llamadas Sistemas Autónomos).

Un poco de historia (cont)

- GGP se siguió utilizando en el backbone, pero fue rápidamente reemplazado por el protocolo EGP (Exterior Gateway Protocol)
- EGP pretendía cubrir las carencias de GGP
- EGP fue pensado para ejecutar en una red dividida en (AS) Sistemas Autónomos
- EGP comenzó a tener problemas cuando la topología de Internet comenzó a cambiar: de tener forma de árbol a forma de grafo
 - A raíz de esto el ruteo comenzó a no tomar rutas óptimas y generar loops de enrutamiento
- EGP fue remplazado por el protocolo BGP (Border Gateway Protocol)
- Actualmente se utiliza BGP-4 (RFC 4271)

Una vista cerca de la UNLP



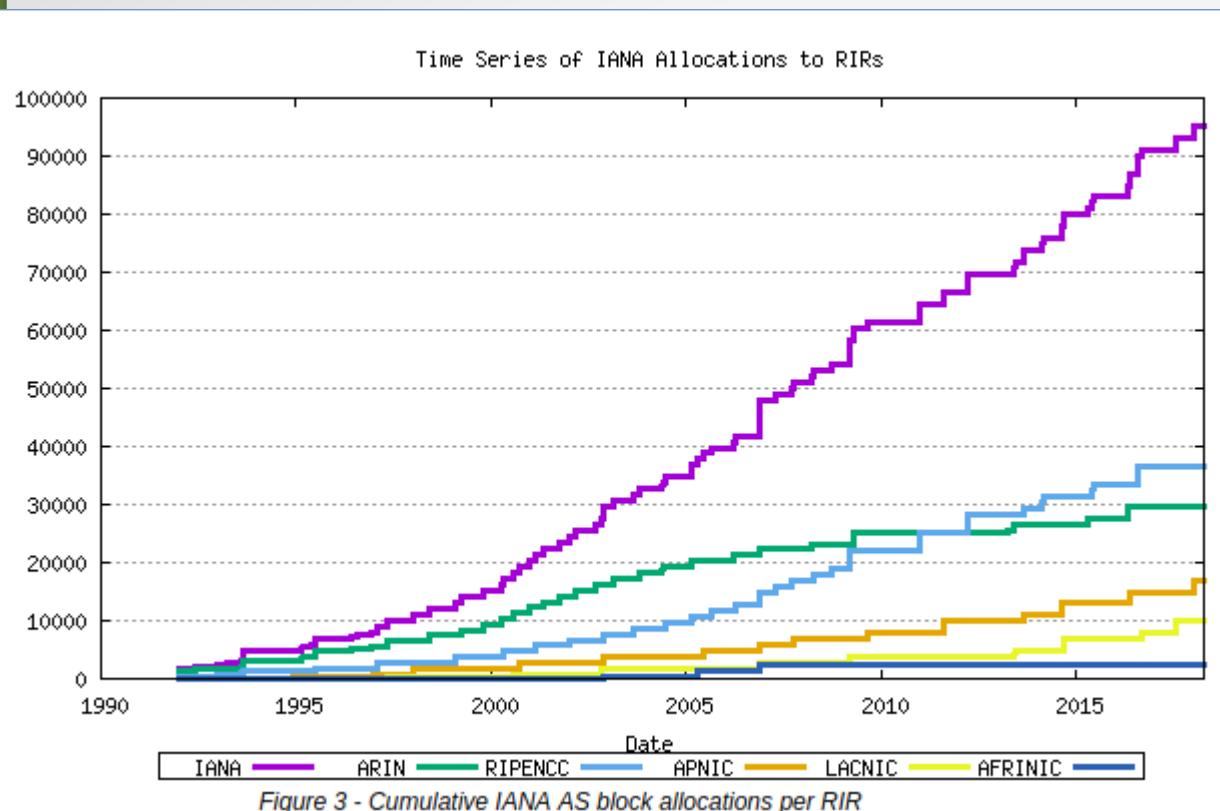
Sistema Autónomo (AS)

- “a set of routers under one or more administrations that presents a common routing policy to the internet”. RFC 1930 - BCP: 6.
- Un AS, es visto por otros AS como una unidad, que:
 - Tiene una política de ruteo coherente
 - Tiene un conjunto de redes alcanzables a través de el
- Un AS se conecta a otros AS para:
 - Publicar las redes propias de su AS
 - Intercambiar otras rutas

Números de AS (ASN)

- Identifican en forma única a un Sistema Autónomo.
- Al igual que las direcciones IP, son asignados a las entidades por los RIRs; el IANA delegó dichas funciones en los registros regionales.
- Hasta el año 2007, estaban definidos por un entero de 16 bits.
- La RFC 4893 presenta los ASN de 32 bits. LACNIC desde 2011 solo asigna de 32 bits.
- En la práctica, trabajaremos con ASN de 16 bits.

Números de AS (ASN)



http://www.caida.org/research/routing/as_growth/
<http://www.potaroo.net/tools/asn32/>

ASN de 32 bits

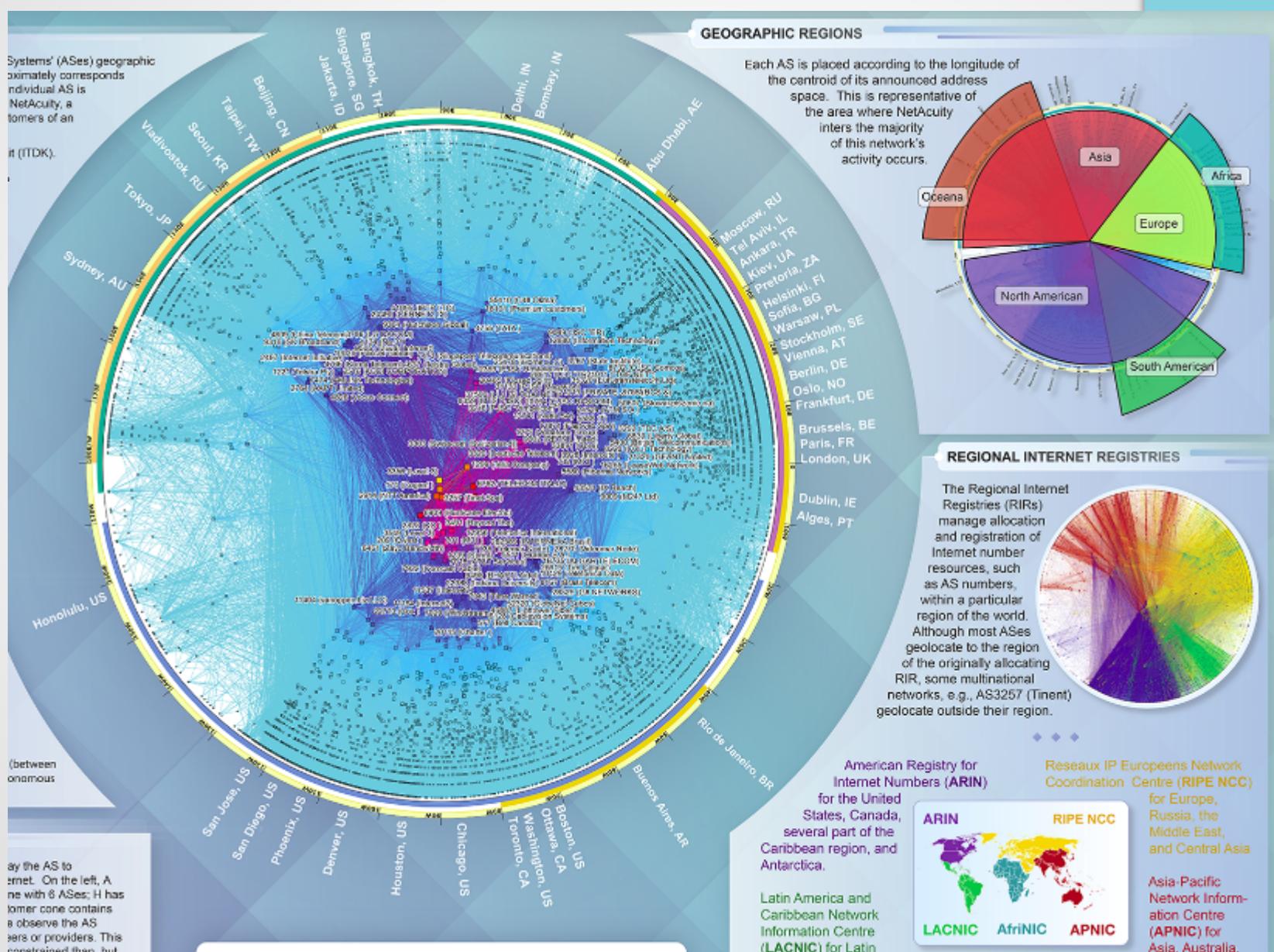
Número de AS/bloque	asignación
0.0-0.65535	Antiguos ASN de 16 bits
1.0-1.65535	reservado
2.0-2.1023	Asignado a APNIC
2.1024-2.65535	Sin asignar
3.0-3.1023	Asignado a RIPE NCC
3.1024-3.65535	Sin asignar
4.0-4.1023	Asignado a LACNIC
4.1024-4.65535	Sin asignar
5.0-5.1023	Asignado a AfriNIC
5.1024-5.65535	Sin asignar

ASN de 16 bits

Número de AS/bloque	asignación
0	Reservado
1-48127	Asignado
48128-54271	Sin asignar
54272-64511	Reservado por IANA
64512-65534	Libre para uso interno (<i>private range</i>)
65535	Reservado

CAIDA – 2017

http://www.caida.org/research/topology/as_core_network/pics/2017/ascore-2017-feb-ipv4-poster-2048x1518.png

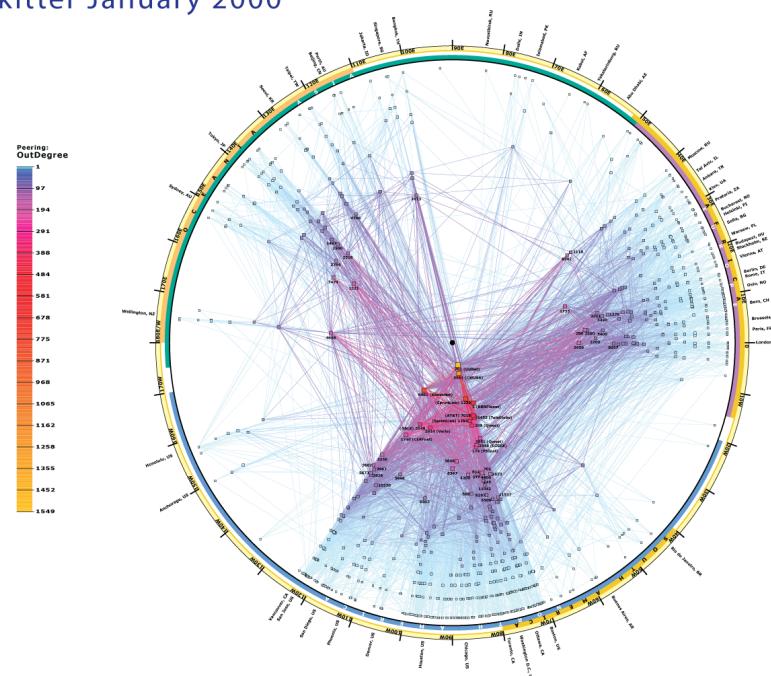


CAIDA - IPv4 - 2000

https://www.caida.org/research/topology/as_core_network/pics/ascore-2000-jan-ipv4-standalone-1600x1333.png

CAIDA's IPv4 AS Core AS-level INTERNET GRAPH

Skitter January 2000

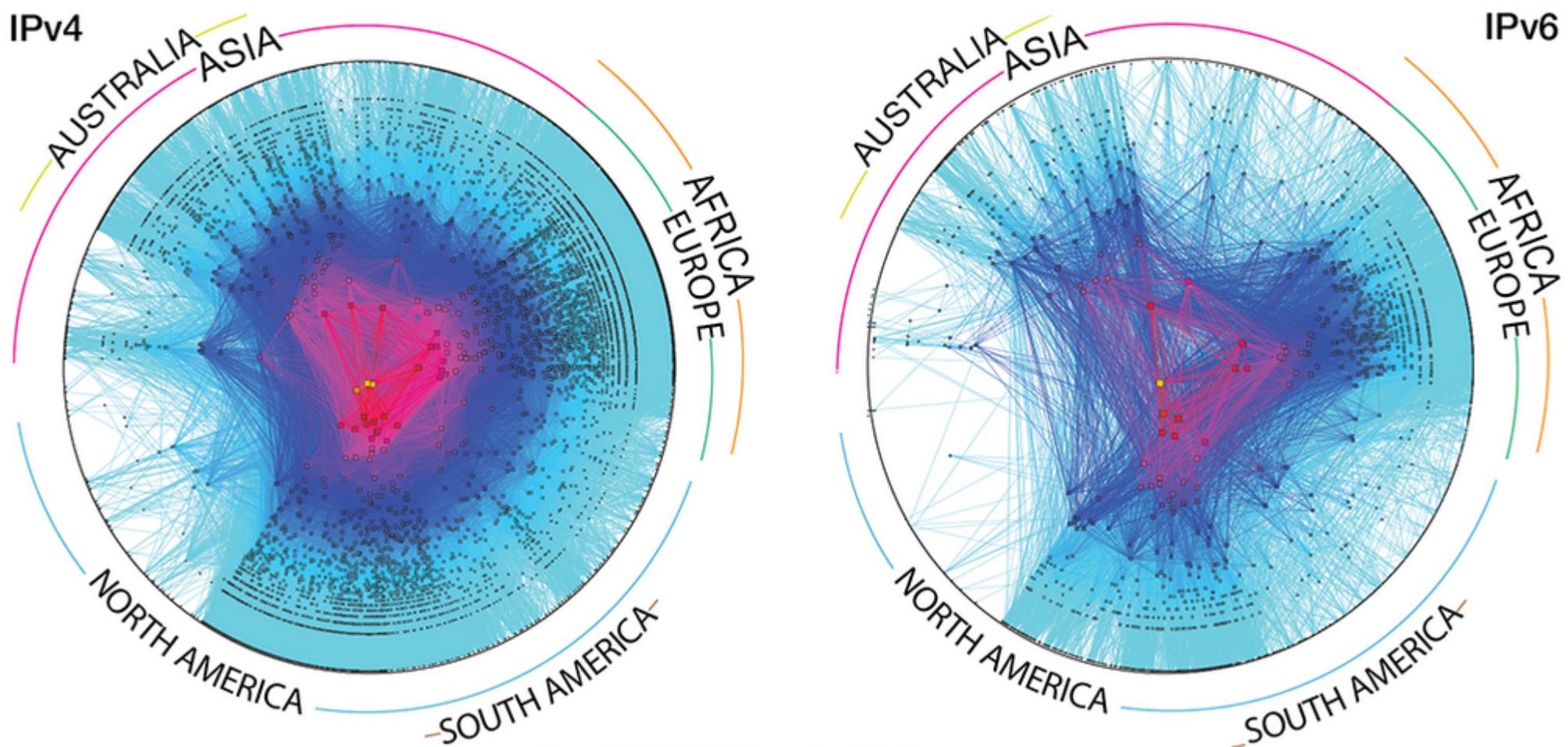


copyright © 2000 UC Regents. all rights reserved.

CAIDA – IPv4 e IPv6 - 2014

http://www.caida.org/research/topology/as_core_network/pics/2014/ascore-2014-jan-ipv4v6-standalone-1200x710.png

CAIDA's IPv4 & IPv6 AS Core
AS-level INTERNET Graph
Archipelago January 2014

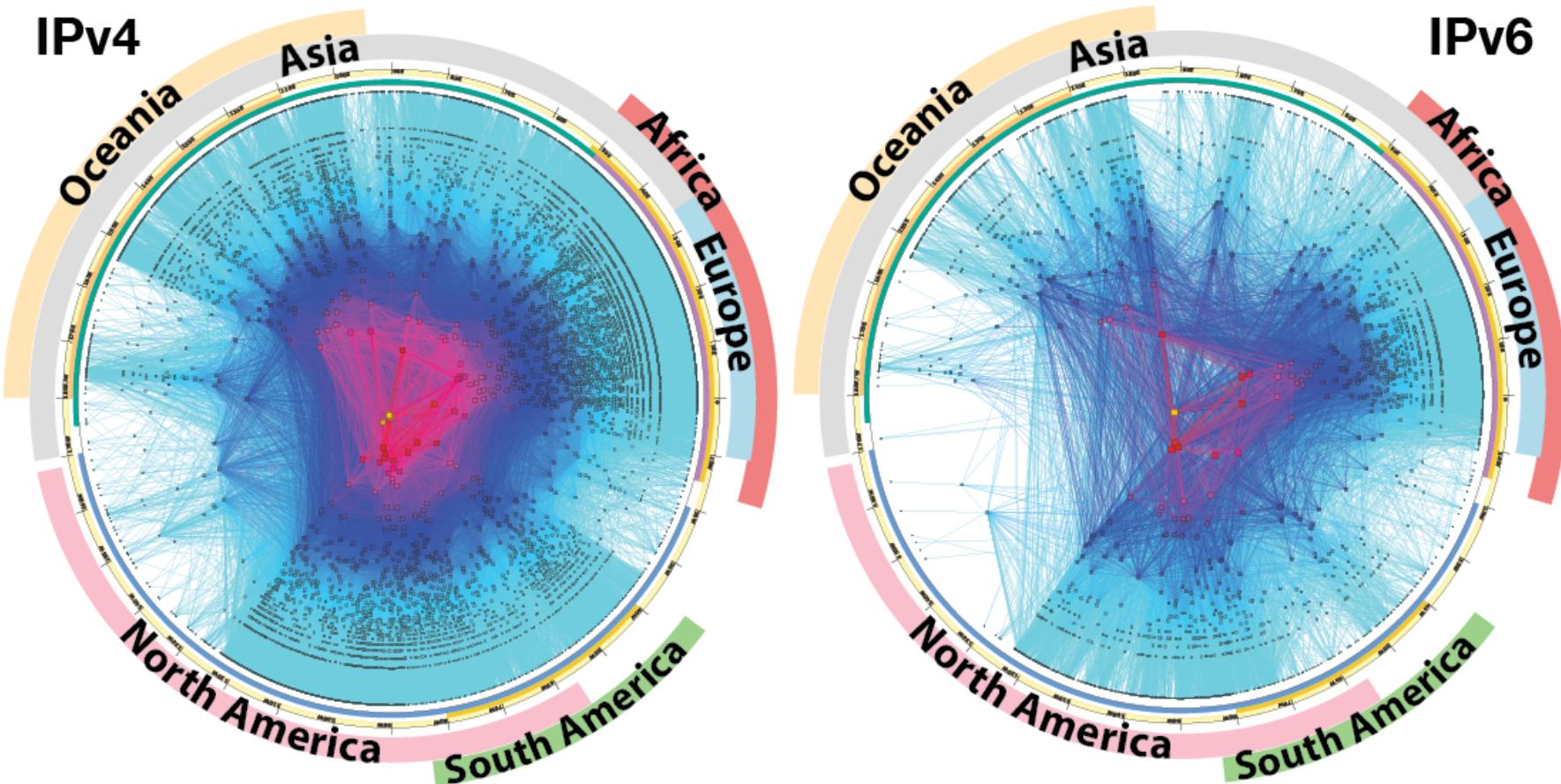


CAIDA – IPv4 e IPv6 - 2015

http://www.caida.org/research/topology/as_core_network/pics/2015/ascore-2015-jan-ipv4v6-standalone-1200x710.png

CAIDA's IPv4 vs IPv6 AS Core AS-level Internet Graph

Archipelago July 2015



Tipos de AS (Autonomous Systems)

- **Stub AS**

- Un AS conectado solamente a otro AS. Su entrada y salida depende del AS al que se conecta.
- No se necesita aprender rutas
- Podría tener un default GW.

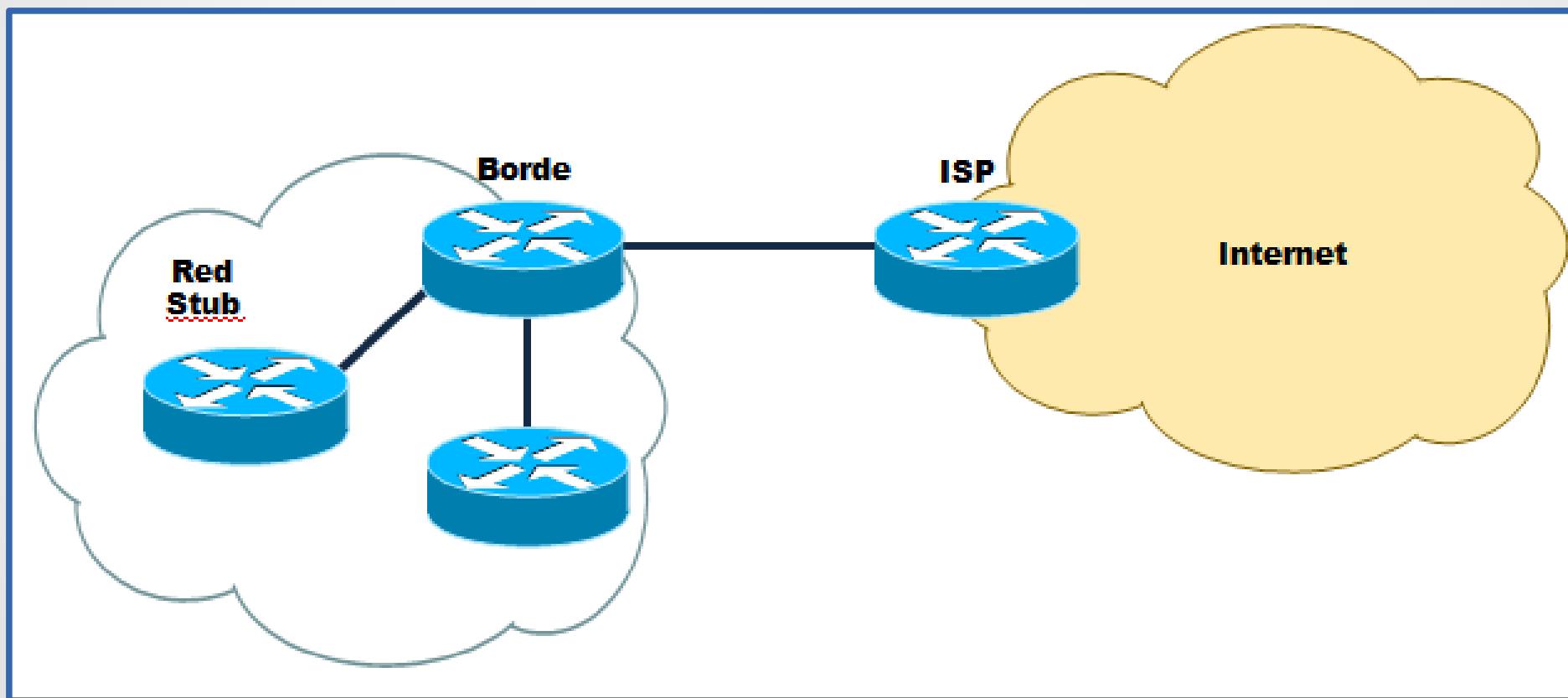
- **Multihome non Transit AS**

- AS con 2 o más conexiones
- No da servicio de ruteo a los AS que se conecta

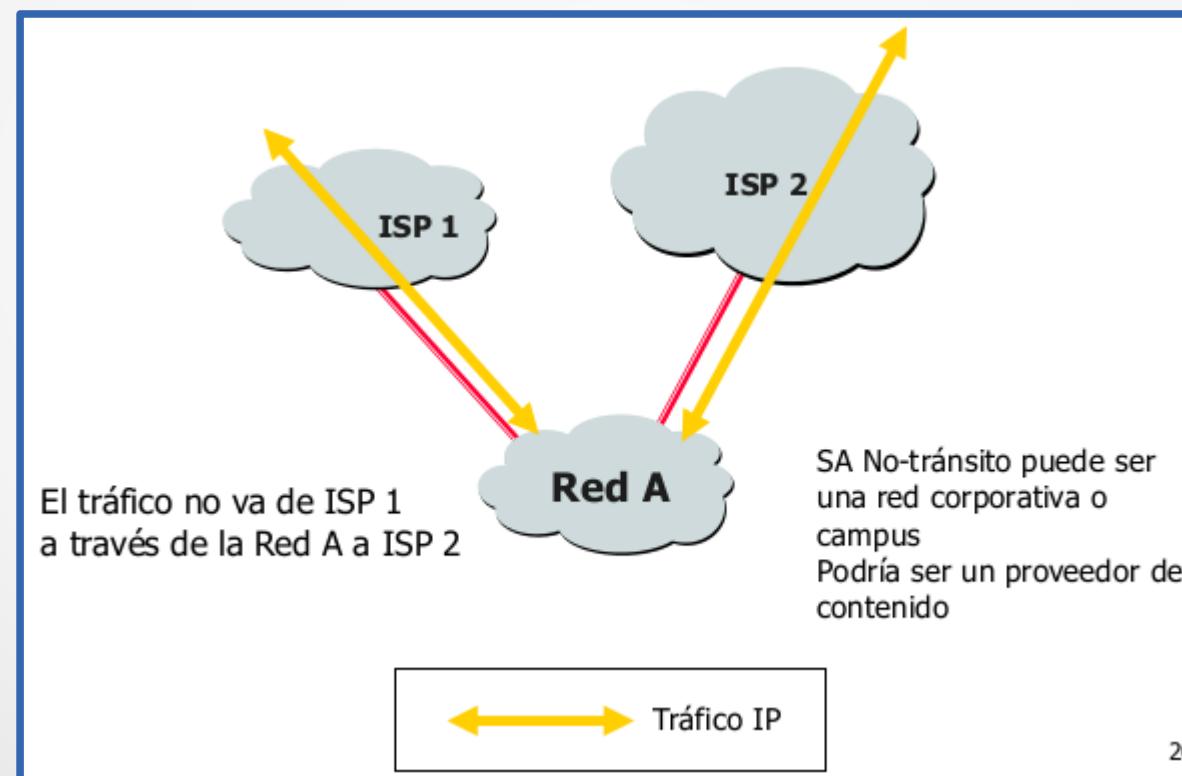
- **Multihome Transit AS**

- AS con 2 o más conexiones
- Da servicio de ruteo a los AS que conecta. Puede aplicar políticas de ruteo.

Stub AS

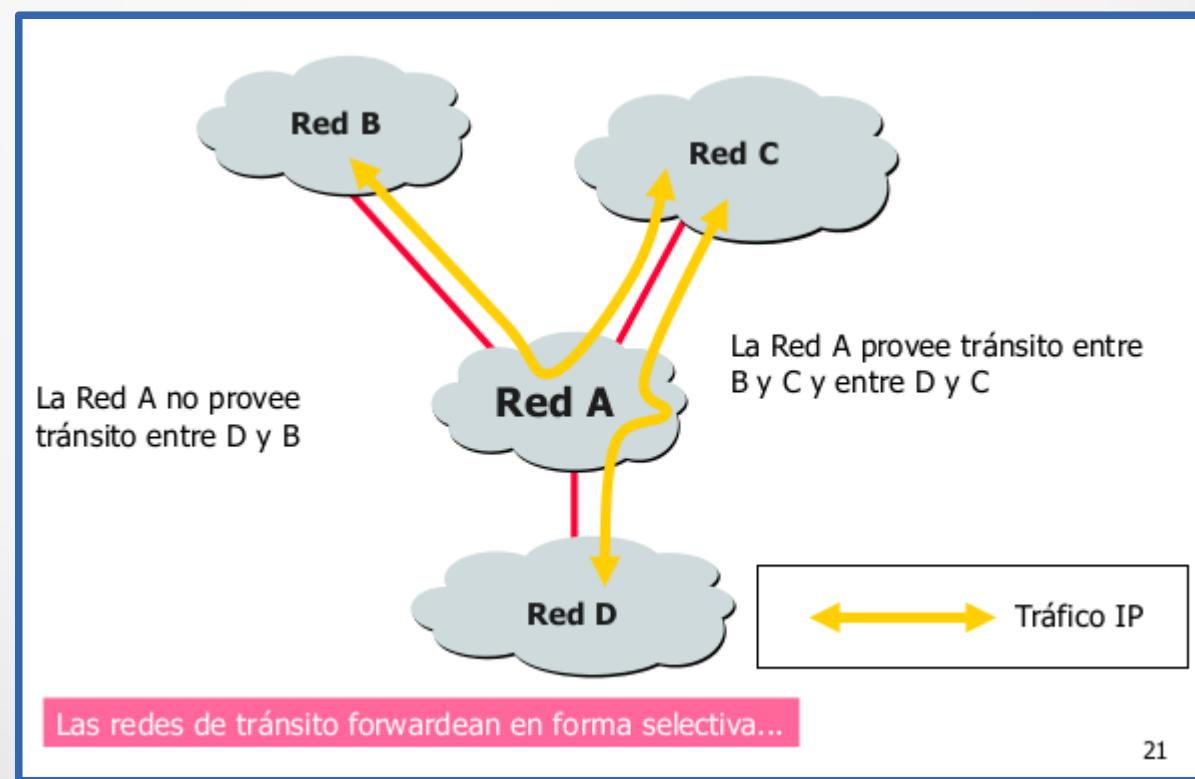


Multihome non Transit AS

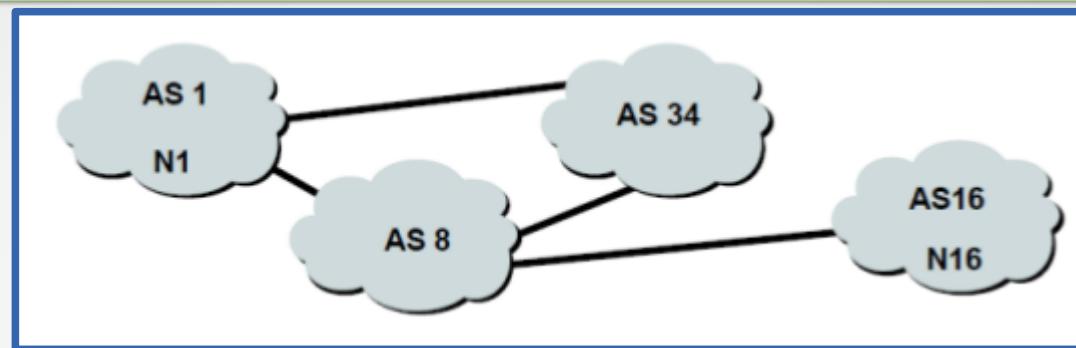


Multihome Transit AS

- Cliente: Red C
- ISP: Red A
- Otros ISPs locales: Red B y Red D



Ejemplo de conexión y funcionamiento



- En el AS1 está la red N1 y en el AS16 la red N16
- Para que desde la red N1 se pueda enviar tráfico a la red N16 se debe:
 - AS16 debe anunciar N16 a AS8
 - AS8 debe aceptar N16
 - AS8 debe anunciar N16 a AS1 y a AS34
 - AS1 debe aceptar N16
- ¿y la vuelta?
- El flujo de las rutas va en sentido contrario al flujo del tráfico.

BGP

- Protocolo desarrollado por la IETF que se utiliza para intercambiar rutas entre los AS.
- Ejecutado en **routers de borde** de un AS
- CIDR y LOOPLESS.
- Utiliza el protocolo de transporte TCP (Puerto 179)
- Keepalives periódicos para verificar conectividad TCP
- Updates incrementales
 - No hay actualizaciones periódicas
- Utiliza métrica múltiple
 - Protocolo de tipo Vector Camino. Las rutas incluyen los AS atravesados.

BGP – Distancia administrativa

Origen de la ruta	Distancia administrativa
Conectado	0
Estática	1
Ruta sumarizada EIGRP	5
BGP externo	20
EIGRP interno	90
IGRP	100
OSPF	110
IS-IS	115
RIP	120
EIGRP externo	170
BGP interno	200

Mensajes BGP

- OPEN
 - Iniciar sesión BGP. Puede requerir autenticación.
- NOTIFICATION
 - Comunicar condiciones de error
- UPDATE
 - Alta o baja de rutas. Cada ruta tiene un prefijo y atributos.
- KEEPALIVE
 - Mantenimiento de la sesión. Confirmación periódica.

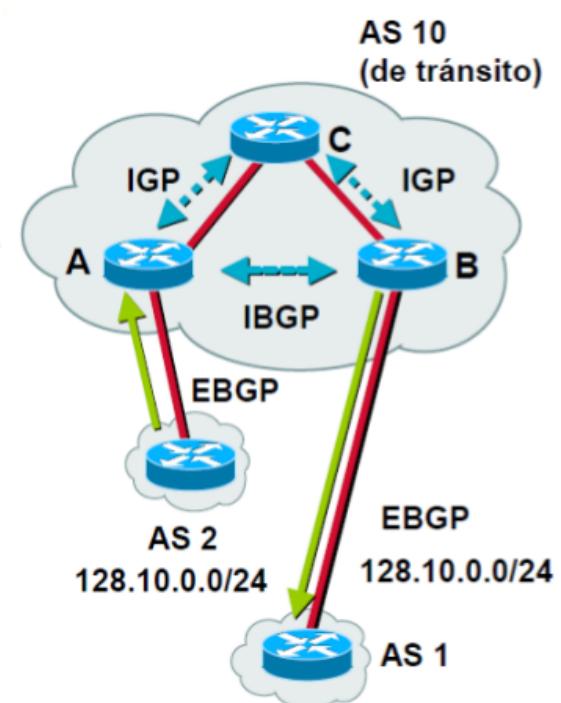
BGP: eBGP vs iBGP

- eBGP (BGP Externo):
 - Utilizado entre routers de distintos AS's para establecer peerings BGP entre los mismos.
 - Intercambia rutas entre AS's
 - Implementa políticas de ruteo
- iBGP (BGP Interno):
 - Utilizado entre los routers que hablan BGP dentro de un AS
 - Mantiene la consistencia a nivel AS (sincronización)
 - Todos los routers BGP de un AS tienen que conectarse entre si (Full Mesh)
 - Alternativamente se pueden usar route reflector o federaciones BGP.
 - Se necesita un protocolo IGP cuando los router que no están directamente conectados.
 - La información recibida por iBGP no se retransmite por iBGP
 - Se recomienda hacer el peering iBGP entre interfaces de tipo loopback

Sincronización en BGP

- iBGP se utiliza para redistribuir las rutas aprendidas por eBGP a routers iBGP.
- Si la sincronización está habilitada, indica que un router iBGP **NO** tomará como válida una ruta aprendida si no la aprendió antes a través de un protocolo IGP.
- Se puede deshabilitar si todos los routers entre los routers de borde corren BGP

- B no anuncia la red al AS 1 hasta no conocerla por IGP
- C debe conocer una ruta a la 128.10.0.0 via IGP ->se debe redistribuir la red aprendida por BGP en IGP en el router A



Atributos BGP

- Las rutas BGP tienen asociadas distintos atributos
- Algunos atributos son propagados con las rutas cuando se intercambian
- Otros atributos son preferencias locales o del AS receptor
- Los atributos que se utilizan en el proceso de determinación de la mejor ruta a un destino dado son:
 - Weight / Local Preference
 - Multi-exit discriminator / Origin
 - AS_PATH / Next hop
 - Community

Proceso de elección de la mejor ruta

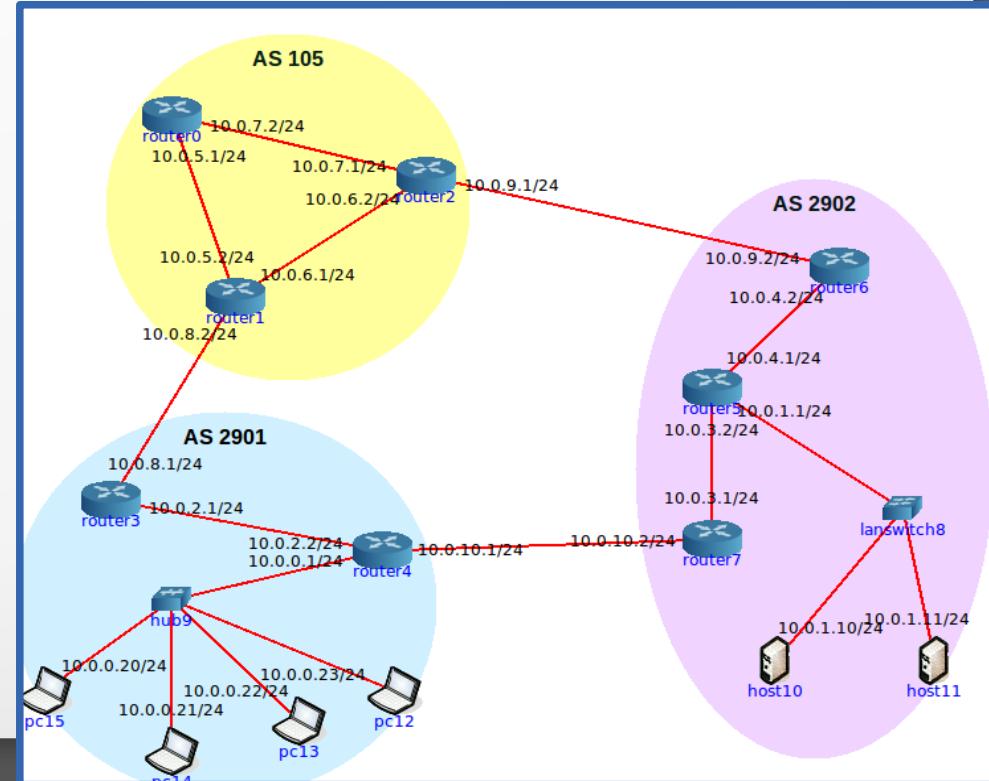
1. Si el **NextHop es inalcanzable**, descartar el update
2. Prefiere el **mayor Weight**
3. Si tienen el mismo weight prefiere el camino con **mayor Local Preference**
4. Si tienen la misma Local Preference prefiere rutas **originada por el propio router**
5. Si no fueron generadas por router propio, prefiere la que tenga el **AS_PATH más corto**
6. Si los AS_PATH son de la misma longitud, prefiere la ruta con el **origin code más bajo (IGP < EGP < INCOMPLETE)**
7. Si los origin codes son iguales, prefiere el camino con **menor MED (MULTI_EXIT_DISC)**
8. Si tienen el mismo MED, prefiere rutas **EBGP** antes que **IBGP**
9. Prefiere la ruta a través del **vecino IGP** más cercano
10. Prefiere la ruta con el valor de **BGP router ID** más bajo

Que es el NextHop en BGP

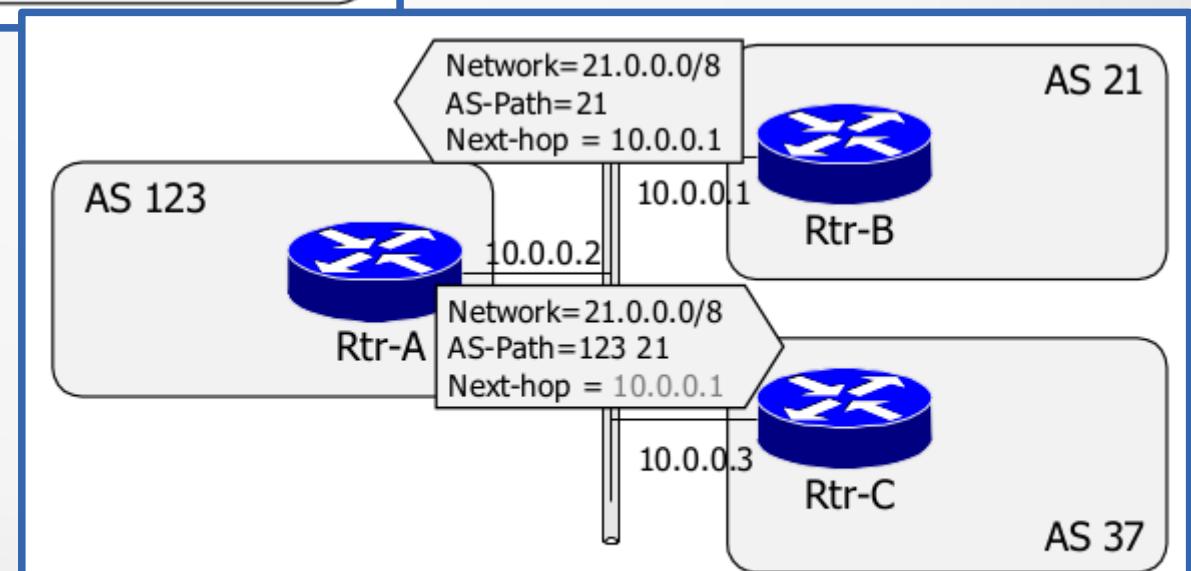
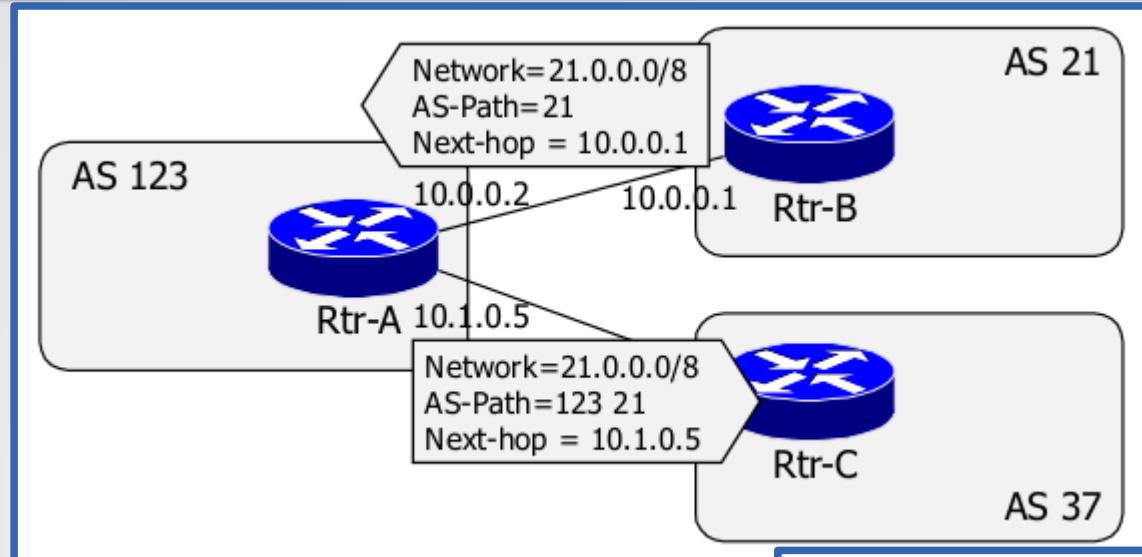
- El NextHop en una ruta es la dirección del próximo salto hacia ese destino.

```
nico@yoko:~$ route -n
Tabla de rutas IP del núcleo
Destino      Pasarela      Genmask      Indic Métric Ref    Uso Interfaz
0.0.0.0      163.10.42.254 0.0.0.0      UG     0      0        0 eth0
10.0.3.0      0.0.0.0      255.255.255.0 U       0      0        0 lxcbr0
```

- En BGP, es la IP a la que tendría que ir para que mis paquetes lleguen al AS destino.



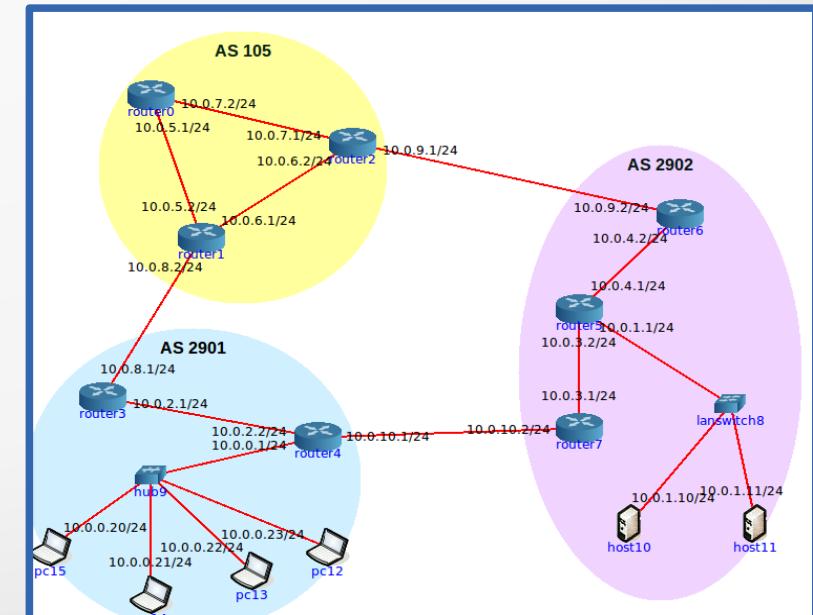
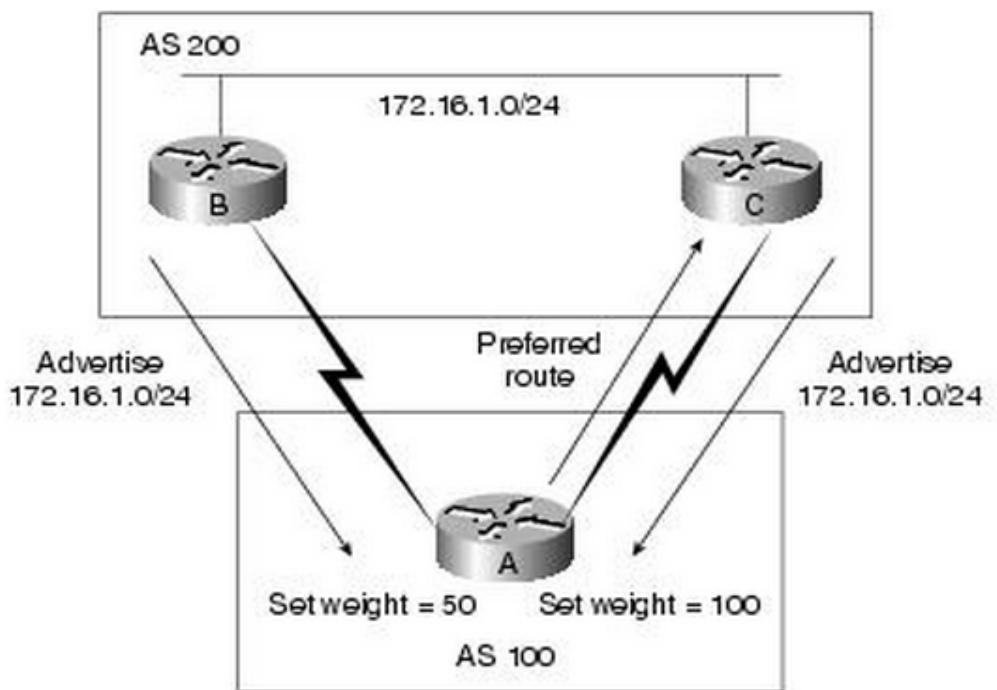
Next Hop



- Si el Router BGP receptor está en la misma red que el próximo salto, este atributo no se cambia para optimizar el ruteo.

Weight

- Es un atributo usado por Cisco en forma local en un router
- No se propaga a otros vecinos eBGP o iBGP
- La ruta con el weight más grande es la mejor

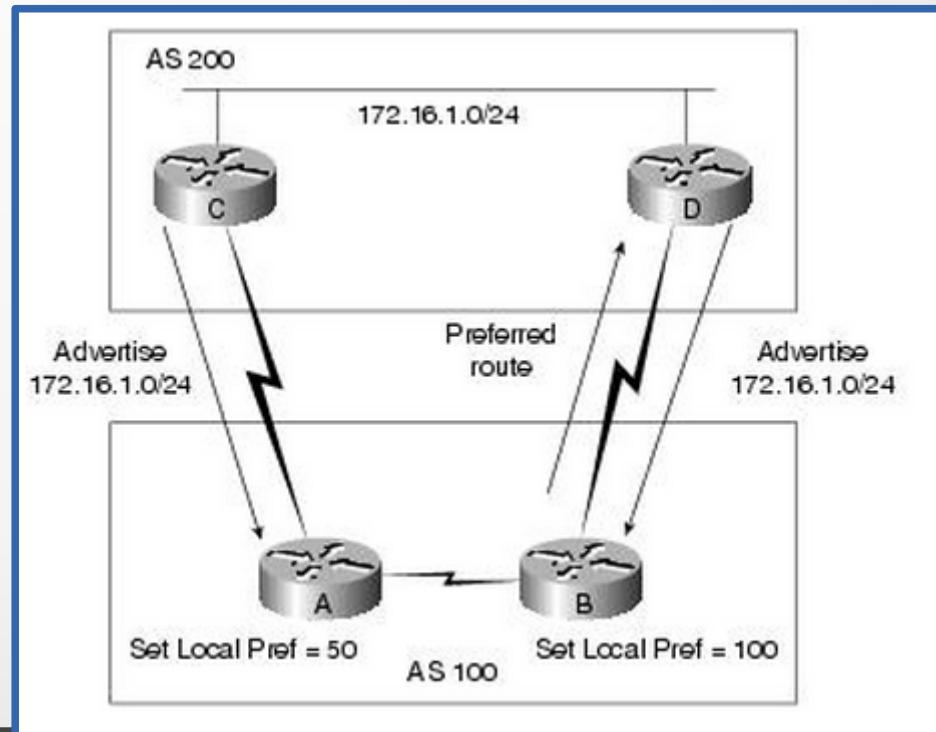


Proceso de elección de la mejor ruta

1. Si el **NextHop** es **inalcanzable**, descartar el update
2. Prefiere el **mayor Weight**
3. Si tienen el mismo weight prefiere el camino con **mayor Local Preference**
4. Si tienen la misma Local Preference prefiere rutas **originada por el propio router**
5. Si no fueron generadas por router propio, prefiere la que tenga el **AS_PATH más corto**
6. Si los AS_PATH son de la misma longitud, prefiere la ruta con el **origin code más bajo (IGP < EGP < INCOMPLETE)**
7. Si los origin codes son iguales, prefiere el camino con **menor MED (MULTI_EXIT_DISC)**
8. Si tienen el mismo MED, prefiere rutas **EBGP** antes que **IBGP**
9. Prefiere la ruta a través del **vecino IGP** más cercano
10. Prefiere la ruta con el valor de **BGP router ID** más bajo

Local Preference

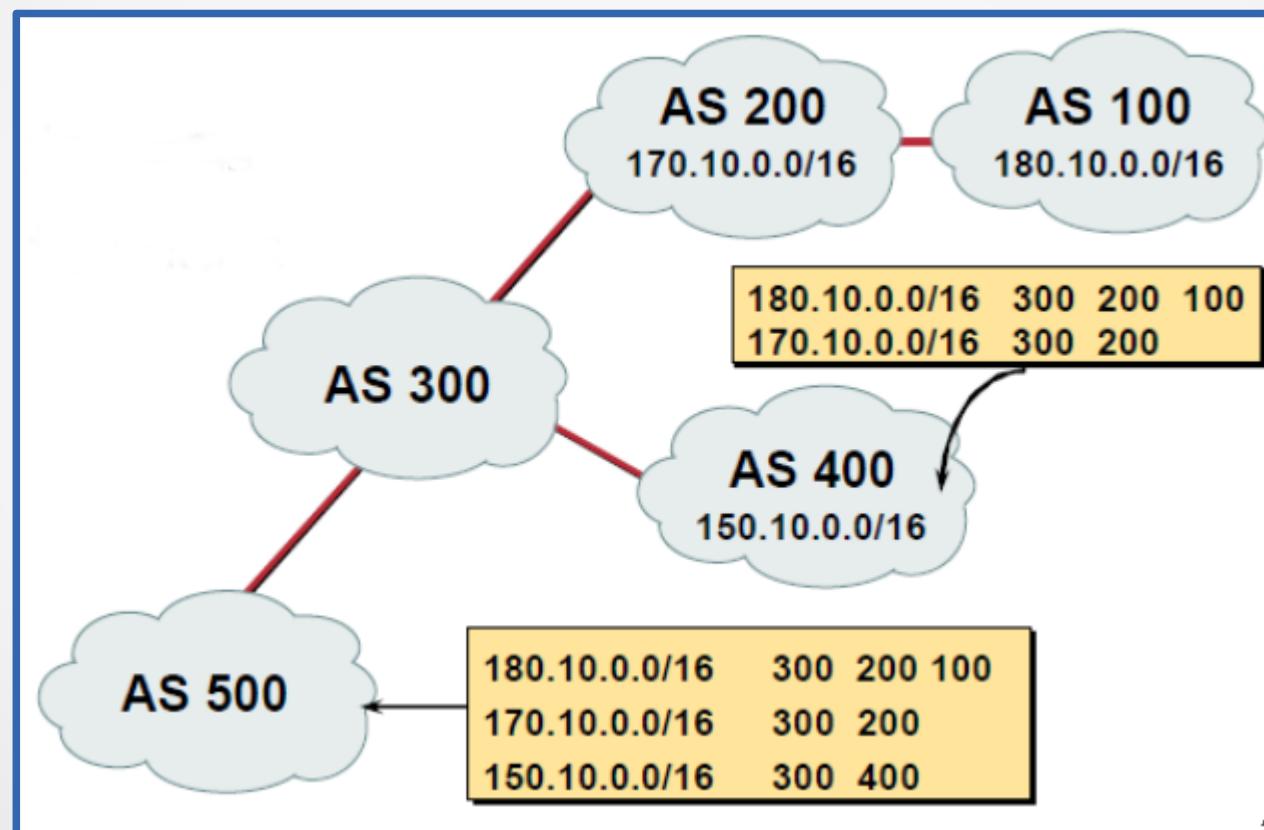
- Utilizado dentro de un AS para establecer preferencia de salida del AS a través de algún punto de salida en particular
- Se propaga a través de iBGP (dentro del AS)
- Se prefiere el mayor valor
- Se tiene una vista consistente de la ruta dentro del AS.
- Valor por defecto: 100



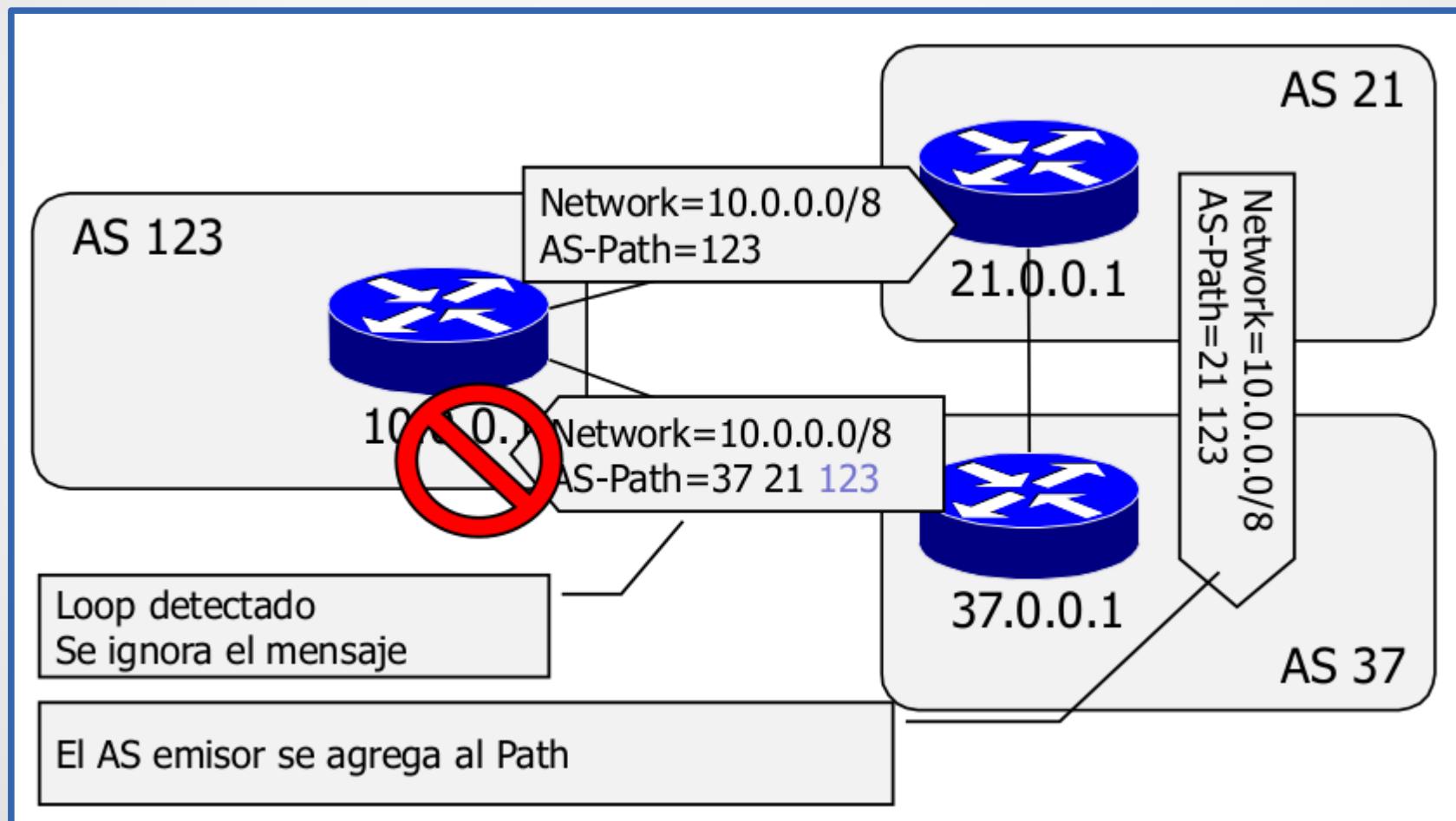
AS_PATH

- Cada ruta es anunciada junto a la lista de AS por los que fue pasando
- Al recibirse una ruta, examinando si el AS receptor está en el AS_PATH se puede determinar si hay loops o no
- Si no hay loops, se toma la ruta, y cuando se anuncia a otros AS's, se agrega el AS que la procesó al AS_PATH
- Se tiene preferencia por la ruta a un destino con el AS_PATH mas corto

Ejemplo AS_PATH



Detección de loops con AS_PATH



Proceso de elección de la mejor ruta

1. Si el **NextHop** es **inalcanzable**, descartar el update
2. Prefiere el **mayor Weight**
3. Si tienen el mismo weight prefiere el camino con **mayor Local Preference**
4. Si tienen la misma Local Preference prefiere rutas **originada por el propio router**
5. Si no fueron generadas por router propio, prefiere la que tenga el **AS_PATH más corto**
6. Si los **AS_PATH** son de la misma longitud, prefiere la ruta con el **origin code mas bajo (IGP < EGP < INCOMPLETE)**
7. Si los origin codes son iguales, prefiere el camino con **menor MED (MULTI_EXIT_DISC)**
8. Si tienen el mismo MED, prefiere rutas **EBGP** antes que **IBGP**
9. Prefiere la ruta a través del **vecino IGP** más cercano
10. Prefiere la ruta con el valor de **BGP router ID** más bajo

Origin

Indica el origen de la ruta BGP

- IGP: la ruta es una publicación BGP. Publicada con el comando “network”.
- EGP: la ruta es una publicación del protocolo EGP. Antecesor de BGP.
- Incomplete: El origen de la ruta es desconocido. Esto se debe a redistribuciones. Publicada con el comando “redistribute”

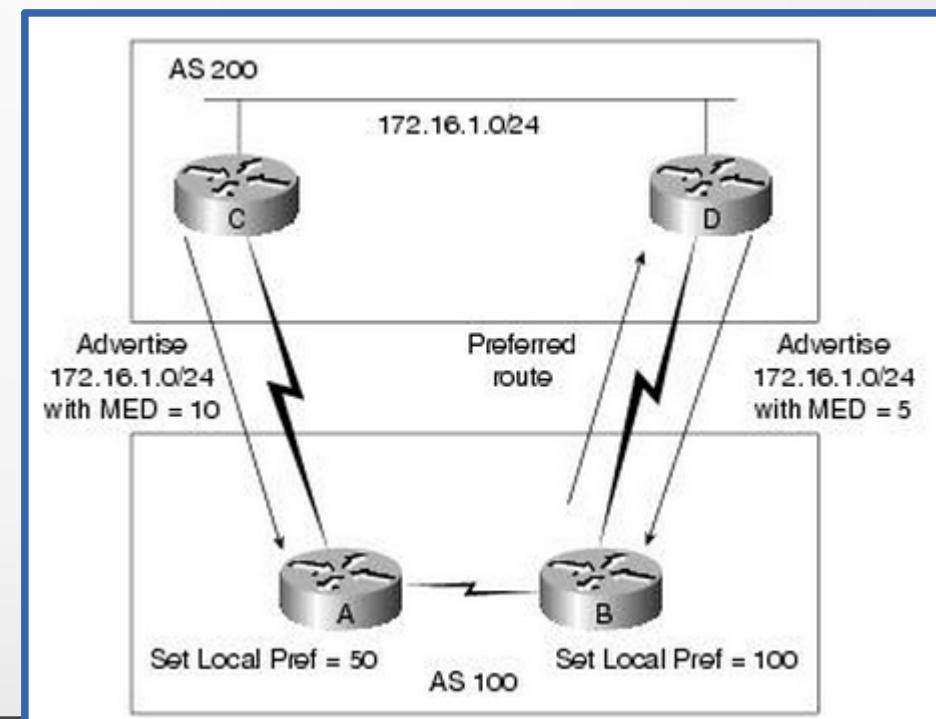
```
n10# sh ip bgp
BGP table version is 0, local router ID is 220.20.20.9
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
r RIB-failure, S Stale, R Removed
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 30.20.10.0/24	220.20.20.10			0 40 50 30	i
*> 40.30.20.0/24	220.20.20.10	0		0 40	i
*> 50.50.50.0/24	220.20.20.10		90	0 40 50	i
*> 60.60.60.0/24	220.20.20.10	0		0 40	?
*>i110.10.10.0/24	192.168.0.5	0	100	0 10	i
* i220.20.20.0/22	192.168.0.5	0	100	0	i
*>	0.0.0.0	0		32768	i

```
Total number of prefixes 6
n10#
```

Multi-exit discriminator (MED)

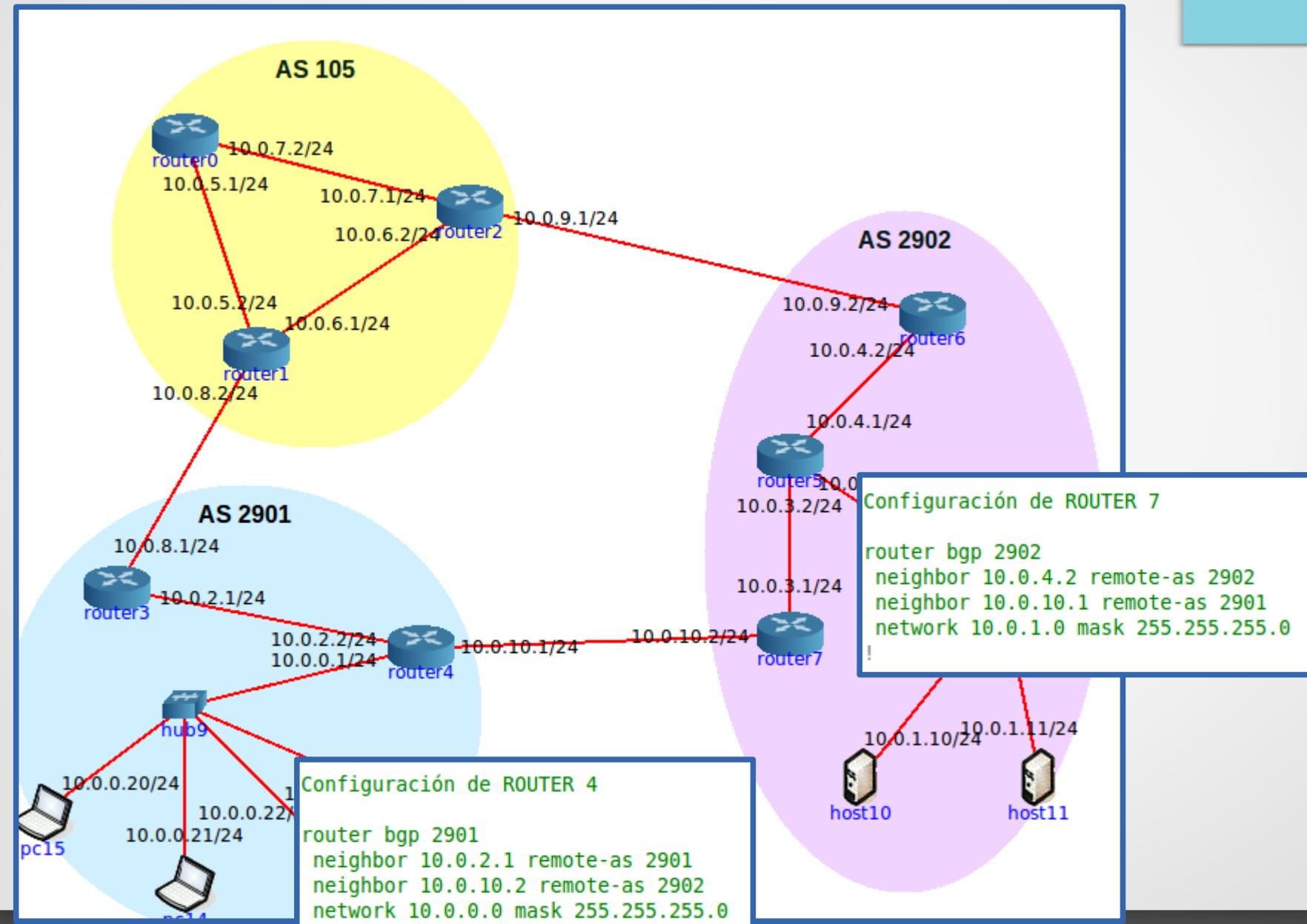
- El MED es una sugerencia a un AS externo del punto por el que prefiero que entre el tráfico a mi AS
- Si acepta la sugerencia, el valor más bajo es el preferido
- El MED se propaga dentro del AS vecino
- Por defecto es 0.



Proceso de elección de la mejor ruta

1. Si el **NextHop** es **inalcanzable**, descartar el update
2. Prefiere el **mayor Weight**
3. Si tienen el mismo weight prefiere el camino con **mayor Local Preference**
4. Si tienen la misma Local Preference prefiere rutas **originada por el propio router**
5. Si no fueron generadas por router propio, prefiere la que tenga el **AS_PATH más corto**
6. Si los **AS_PATH** son de la misma longitud, prefiere la ruta con el **origin code más bajo (IGP < EGP < INCOMPLETE)**
7. Si los origin codes son iguales, prefiere el camino con **menor MED (MULTI_EXIT_DISC)**
8. Si tienen el mismo MED, prefiere rutas **EBGP** antes que **IBGP**
9. Prefiere la ruta a través del **vecino IGP más cercano**
10. Prefiere la ruta con el valor de **BGP router ID más bajo**

Configuraciones BGP



BGP - Publicando una red

- La publicación de una red en BGP requiere que previamente dicha red se encuentre en la tabla de ruteo.
- Solamente se anuncian las redes que coincidan en forma exacta con la ruta de la tabla de ruteo.
- Se recomienda publicar la red con el comando `network` y poner una ruta estática a `null0` con peso alto.
- En Quagga no es necesario hacer esto, pero en Cisco si no se implementa de esta forma no funciona.
 - En Quagga se puede implementar como en Cisco

```
(config)#router bgp 200
(config-router)#network 140.0.0.0/16
(config-router)#bgp network import-check
(config)# ip route 140.0.0.0/16 null0
```

Cantidad de rutas BGP – Tabla global

- Rango: 1994 a 2018

