

# Routing: BGP-4

Area de Ingeniería Telemática  
<http://www.tlm.unavarra.es>

Grado en Ingeniería en Tecnologías de  
Telecomunicación, 3º

# Temas de teoría

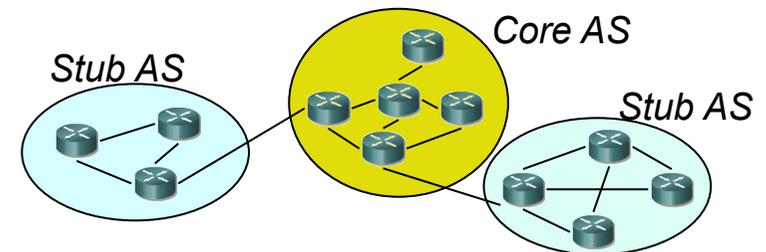
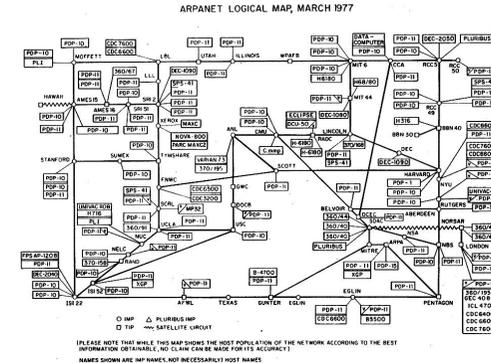
1. Introducción
2. QoS
3. Encaminamiento dinámico en redes IP
4. Tecnologías móviles
5. Otros temas

# Objetivos

- Comprender el funcionamiento básico de BGP-4
- Conocer la relación entre BGP-4 y la arquitectura de Internet

# Un poco más de historia

- En el comienzo de los 80 ARPANET empleaba GGP
  - *Gateway-to-Gateway Protocol*
  - Distance Vector
  - Todo router conocía ruta a todas las redes
  - Distancia en saltos
  - No escala
- RFC 827 (1982) propuso:
  - Migración a un sistema de redes autónomas interconectadas: *Autonomous Systems*
  - Añade un nuevo nivel en la jerarquía
  - Red de redes  $\Rightarrow$  Red de sistemas autónomos
  - Los routers *externos* compartirían información de enrutamiento mediante un protocolo llamado EGP
- GGP se convierte en el primer IGP, para el AS Core



# EGP (RFC 904)

- Distance Vector
- Da alcanzabilidad de redes
- No tiene un algoritmo para buscar el mejor camino
- No envía suficiente información como para evitar ciclos
- Necesita una **topología sin ciclos**
- Intercambia mensajes entre *vecinos o peers*
  - *vecinos interiores (interior neighbors)*: en el mismo AS
  - Si no, son *exterior neighbors*
  - Configuración manual de quiénes son los vecinos

## Core AS

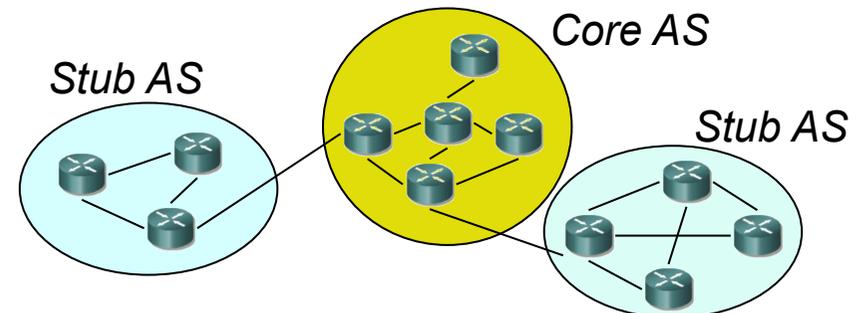
- Solo uno, a él se conectan los demás
- Sus routers pueden enviar información sobre redes diferentes de su AS

## Stub AS

- Solo envían información sobre este AS

## Problemas:

- Necesidad de topología más conectada
- Detectar bucles
- Reducir tiempos de convergencia
- Especificar *routing policies*



# El nuevo protocolo

## ¿Distance vector?

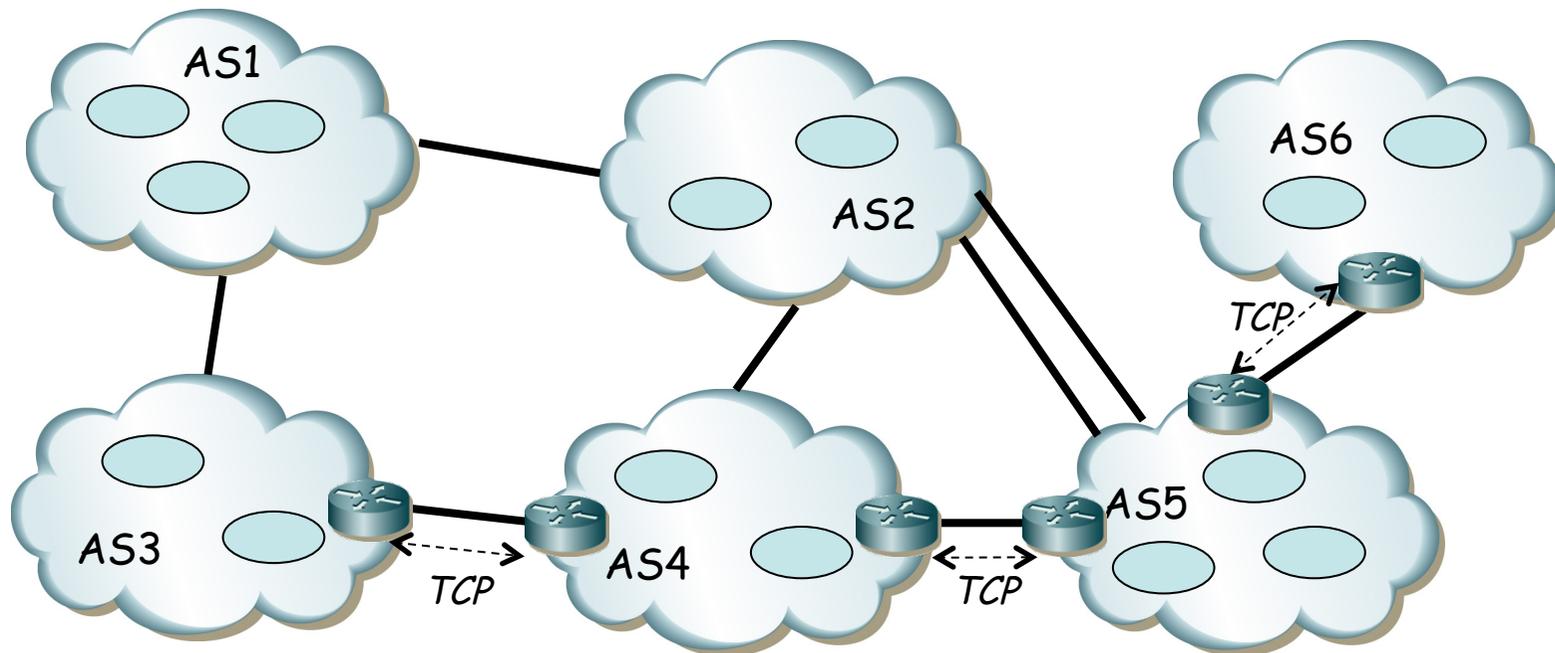
- En ocasiones no se quiere ir por la ruta más corta (políticas)
- Inestable

## ¿Link State?

- No escala bien (aunque se puede lograr, como PNNI)
- Base de datos de LSPs grande
- Coste de calcular las rutas mediante Dijkstra

# BGP

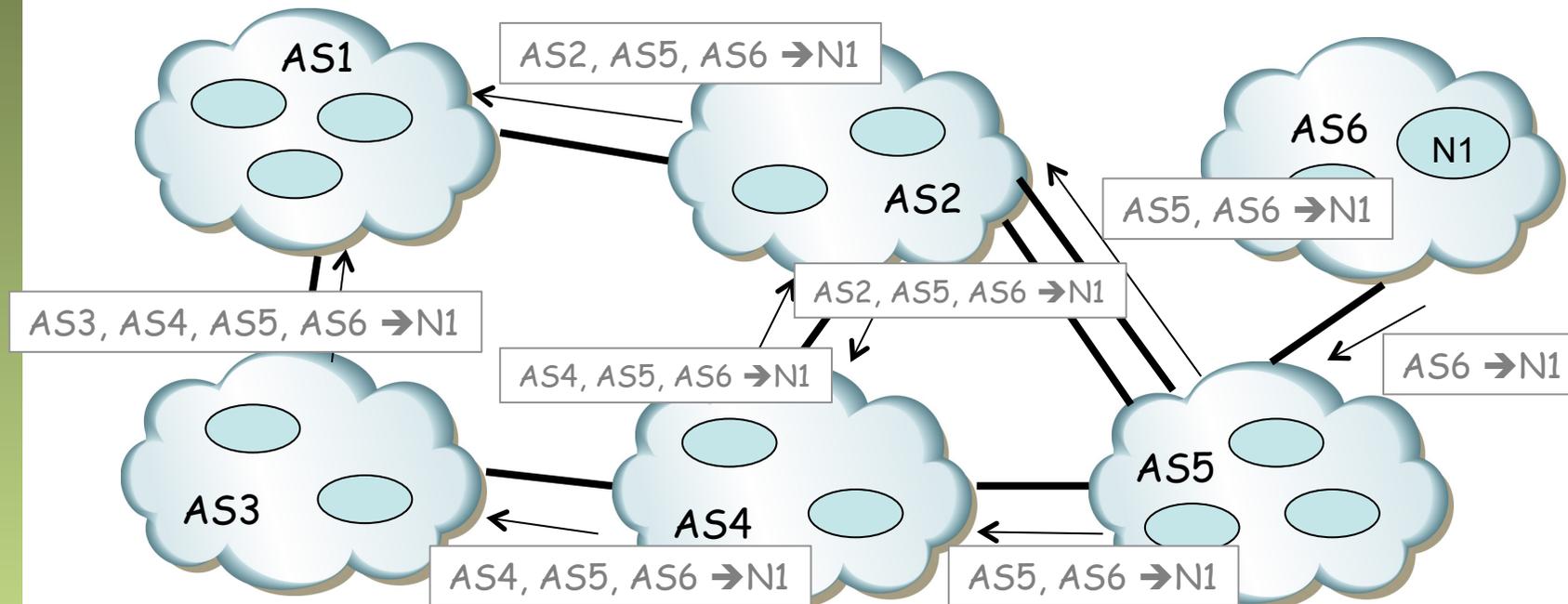
- *Border Gateway Protocol*
- BGP-4, RFC 4271
- BGP-4 primera versión classless
- Protocolo Interdomain estándar *de facto*
- Comunicación fiable mediante conexión TCP entre routers adyacentes
- Puerto 179



# BGP

## Path Vector

- Calcula caminos a prefijos
- Como DV recibe de vecinos, calcula sus rutas y envía a vecinos
- En vez de métrica anuncia la lista de AS en cada camino (. . .)
- Anunciar el camino permite evitar los ciclos
- Por defecto elige el camino que pasa por menor número de ASs
- Eso no quiere decir que sea el menor número de saltos por routers



# Mensajes

- Primero se establece la conexión TCP entre los dos *BGP speakers*
- Cuatro mensajes obligatorios

## OPEN

- Tras establecerse la conexión
- Router especifica parámetros de operación: versión, identificador, AS number, *hold time*, *capabilities*, etc.
- Suele ir seguido de un intercambio de todas las rutas

## KEEPALIVE

- Para comprobar periódicamente el *peering*
- Se da por rota la sesión si pasa el *hold time* sin recibirlo

## NOTIFICATION

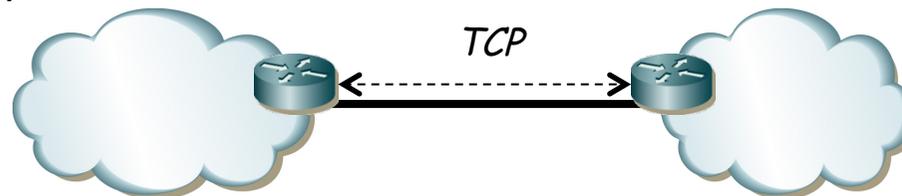
- Cuando se detecta un error
- Termina la conexión

## UPDATE

- Anuncia información de enrutamiento (nuevas rutas o eliminar otras – *withdraw* –)
- Anuncia un solo camino por mensaje
- Anuncia cuando ha calculado una nueva mejor ruta al destino
- Si deja de poder alcanzarlo anuncia eso también
- Prefijo / Longitud
- **Atributos del camino:** permiten a BGP elegir el mejor

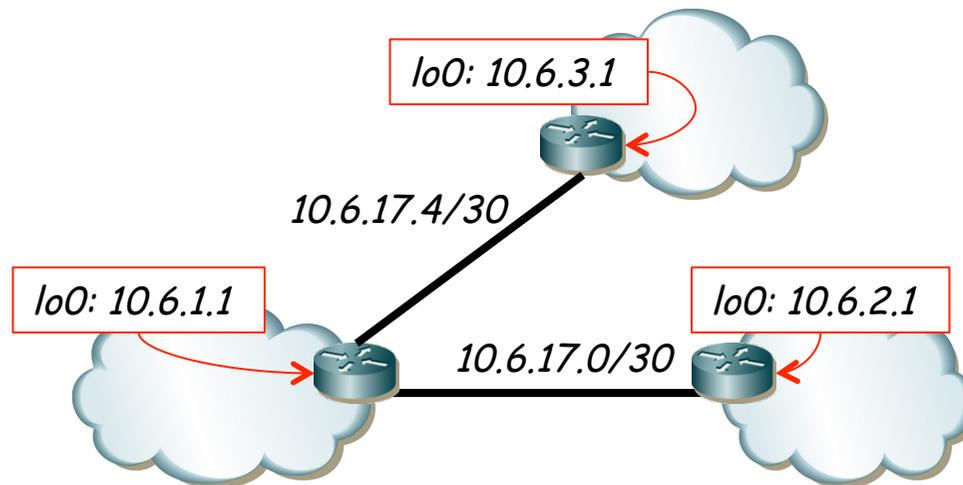
## ROUTE-REFRESH (opcional)

- Pare pedir que vuelva a anunciar los prefijos que conoce



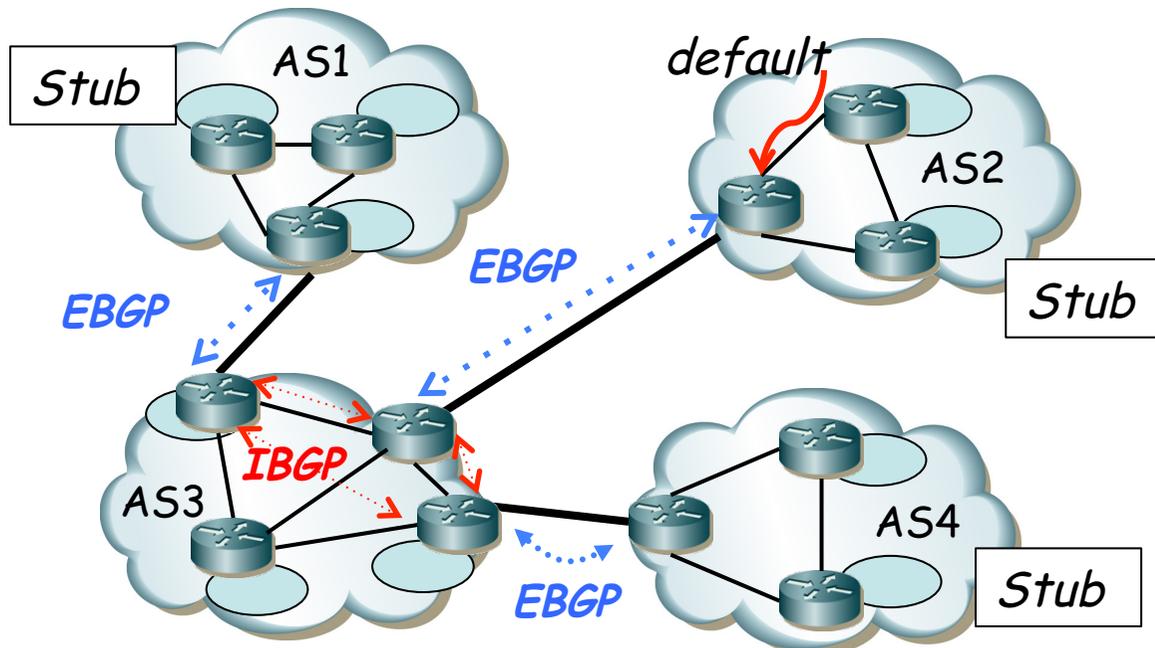
# Direccionamiento

- En enlace entre dos ASBRs empleará un direccionamiento
- Frecuentemente es parte de la asignación de uno de los dos ASs
- Un router con más de un *peer* tendría una dirección diferente para cada uno, lo cual complica la gestión
- Para tener una única identificación se emplea la de un interfaz de *loopback*
- Se envían los paquetes con TTL=2 si se usa *loopback* y con TTL=1 si no se usa



# Peering en BGP

- Los *peers* de un proceso BGP pueden estar:
  - En otro AS: *external peer* ⇒ **EBGP**
  - En el mismo AS: *internal peer* ⇒ **IBGP**
- En el mismo AS el *peering* forma una malla
- Reconoce si es del mismo porque en el OPEN anuncia el ASN
- No interesa difundir todas las rutas al IGP (escalabilidad)
- IBGP permite que otros ASBRs aprendan los prefijos a anunciar
- EL ASN se añade a al ruta al hacer anuncio a otro **EBGP**



# Path Attributes

- Son características de una ruta BGP

## Tipos según se soporten:

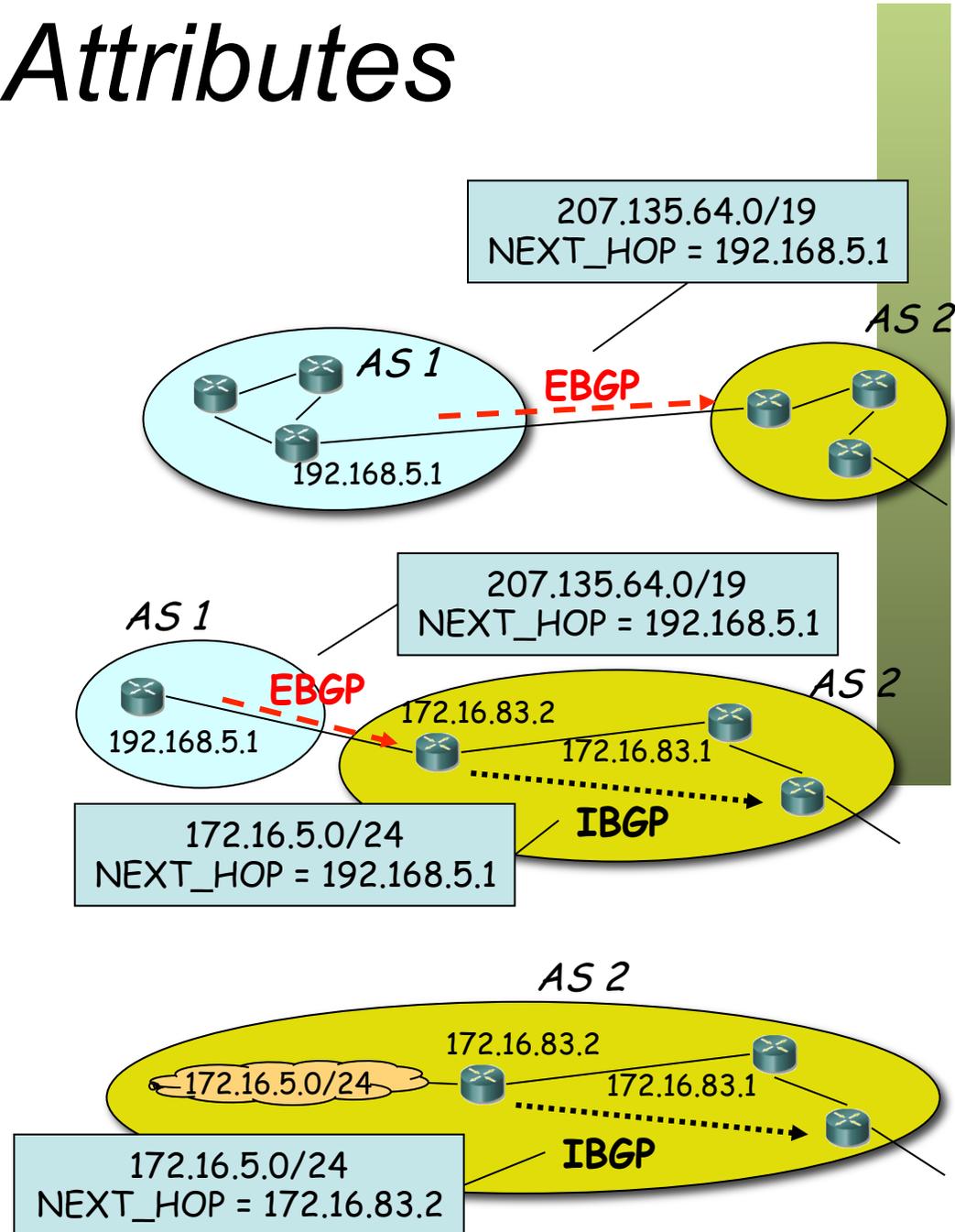
- *Well-known: mandatory* (en update) o *discretionary*
- *Optional: transitive* o *nontransitive*

## ORIGIN (well-known mandatory)

- IGP, EGP o Incompleto (rutas estáticas)

## NEXT\_HOP (well-known mandatory)

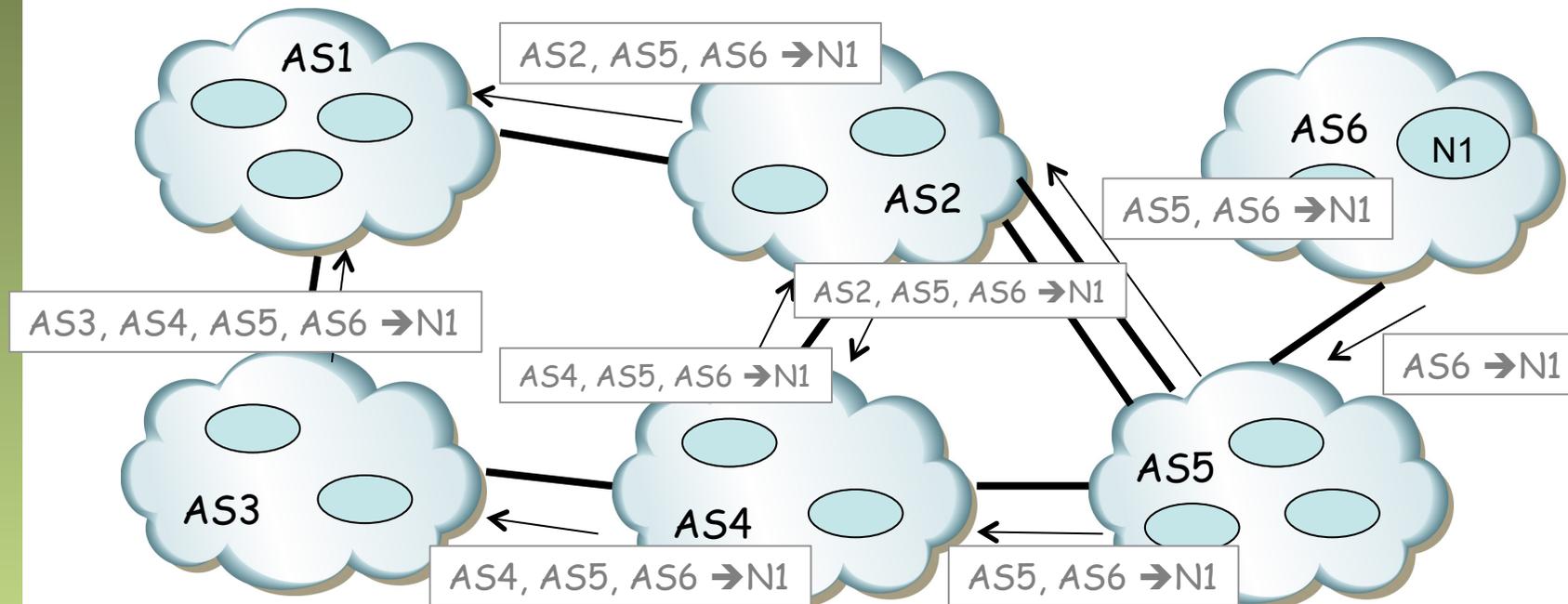
- Si son *External Peers* es la IP del interfaz del router anunciante
- Si son *Internal Peers* y
  - Destino fuera del AS: IP del peer externo (...)
  - Destino en el mismo AS: IP del anunciante (*recursive lookups*) (...)



# Path Attributes

## AS\_PATH (well-known mandatory)

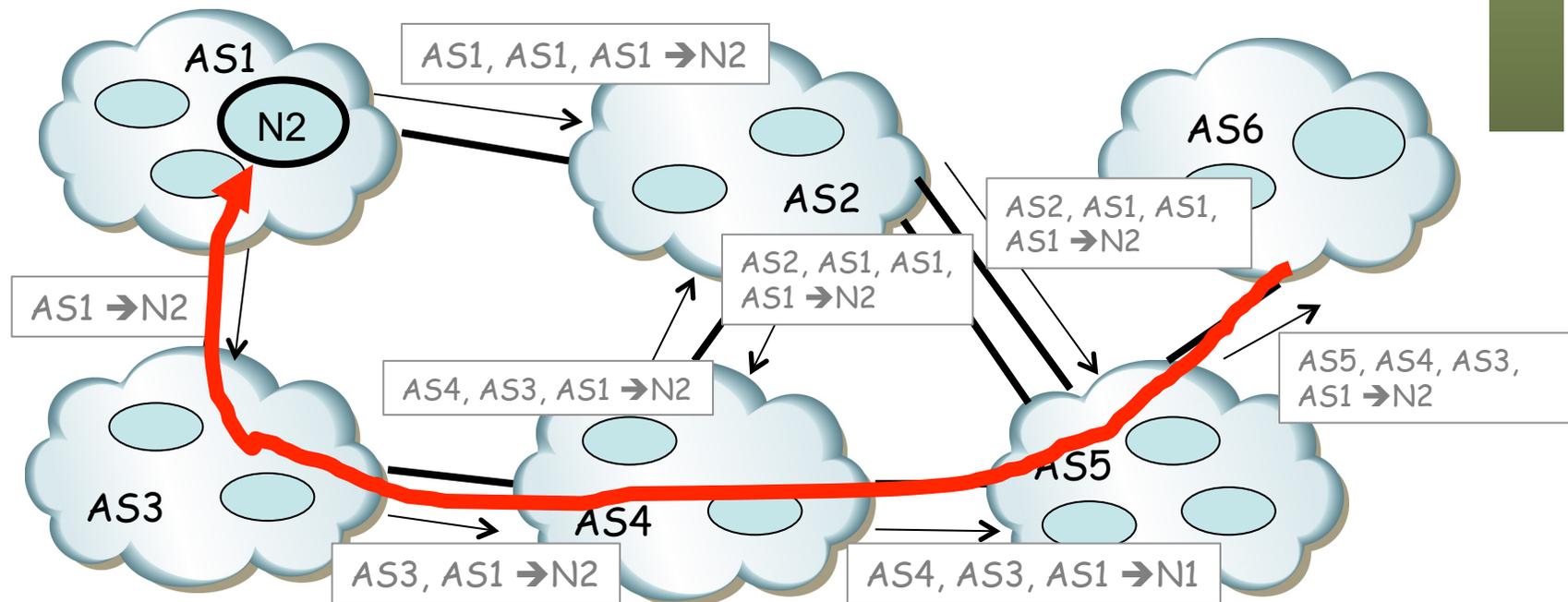
- Secuencia de ASs hasta el destino
- Al mandar un *update* por EBGP se añade el ASN a la secuencia
- Si se manda por IBGP no se añade el ASN
- *AS path prepending*: (...)



# Path Attributes

## AS\_PATH (well-known mandatory)

- Secuencia de ASs hasta el destino
- Al mandar un *update* por EBGP se añade el ASN a la secuencia
- Si se manda por IBGP no se añade el ASN
- *AS path prepending*: añadir el ASN *más veces* para desalentar usar este camino (. . .)



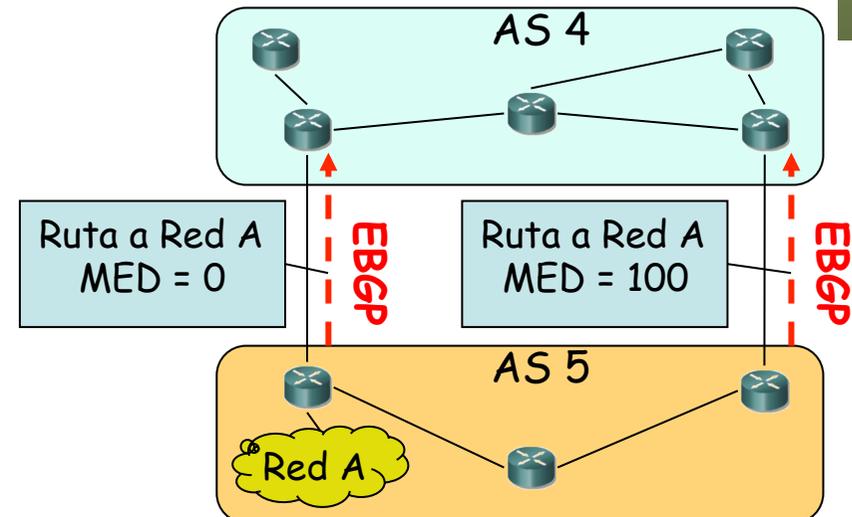
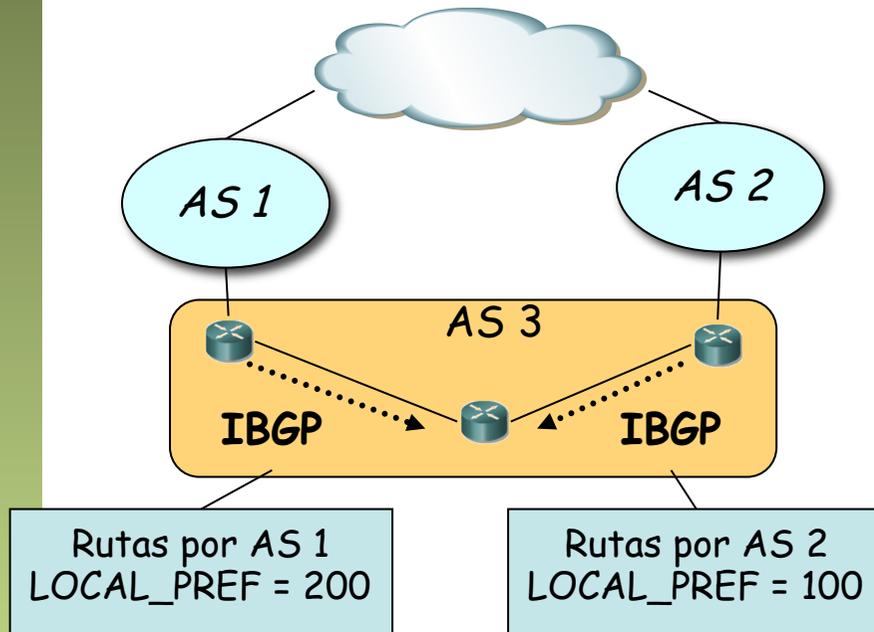
# Path Attributes

## LOCAL\_PREF (well-known discretionary)

- Solo en IBGP
- Comunica el grado de preferencia por una ruta
- La ruta de mayor valor es seleccionada

## MED (optional nontransitive)

- Multi-Exit-Discriminator
- Cuando hay múltiples links a un AS
- Anuncia el *ingress point* preferido
- Es una métrica y se selecciona el de menor MED
- No se propaga a más ASs (debe borrarlo al pasar la ruta a otro AS)

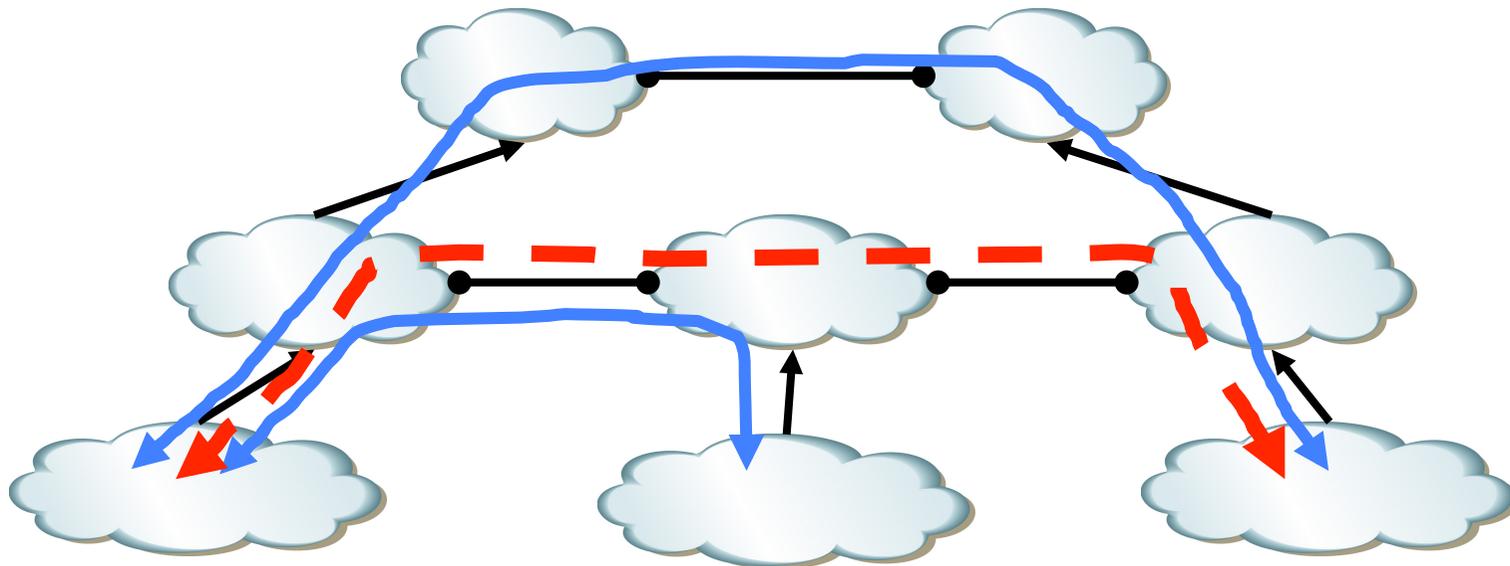


# Un criterio de selección

1. Ruta con el mayor **LOCAL\_PREF**
2. Si iguales, la ruta de **AS\_PATH** más corto
3. Si iguales, la ruta de origen menor (**ORIGIN** IGP < EGP < Incomplete)
4. Si iguales y van al mismo AS, la de menor **MED**
5. Si igual, la de menor **métrica** del IGP hasta el NEXT\_HOP
6. Si iguales y van al mismo AS, se puede instalar todas las rutas o escoger la de menor identificador de router

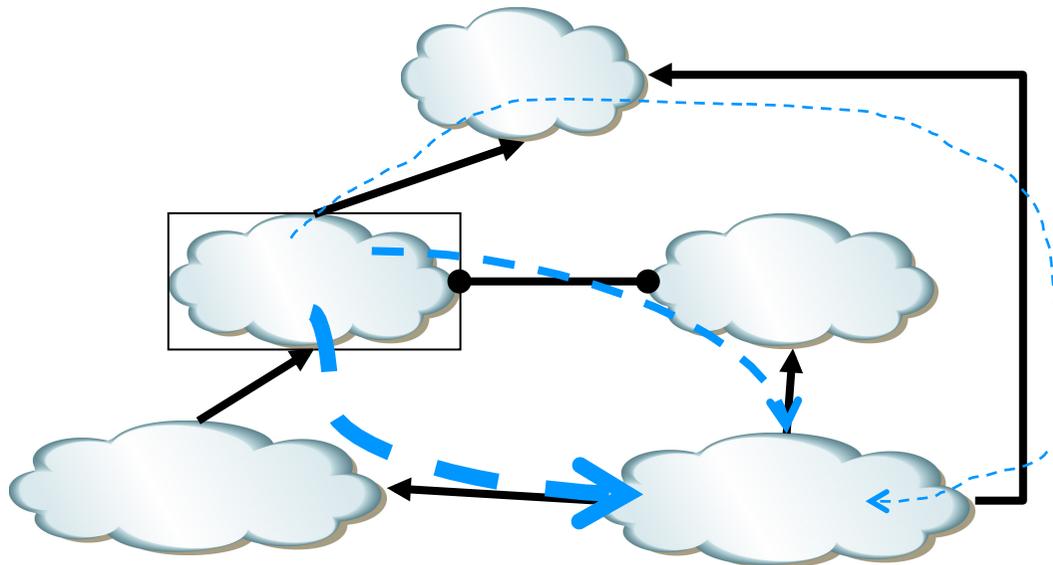
# Jerarquía y economía

- En la Internet tenemos enlaces
  - Cliente-Proveedor (de pago)
  - Entre iguales (normalmente no se pagan)
- Por un enlace entre pares no se hace tránsito (...)
- Preferencia habitual:
  - (...)



# Jerarquía y economía

- En la Internet tenemos enlaces
  - Cliente-Proveedor (de pago)
  - Entre iguales (normalmente no se pagan)
- Por un enlace entre pares no se hace tránsito (...)
- Preferencia habitual:
  1. Por cliente
  2. Por *peer*
  3. Por proveedor



# Políticas

- Anunciar una ruta implica que se está dispuesto a encaminar el tráfico a ese destino

## Los administradores pueden implementar diferentes políticas:

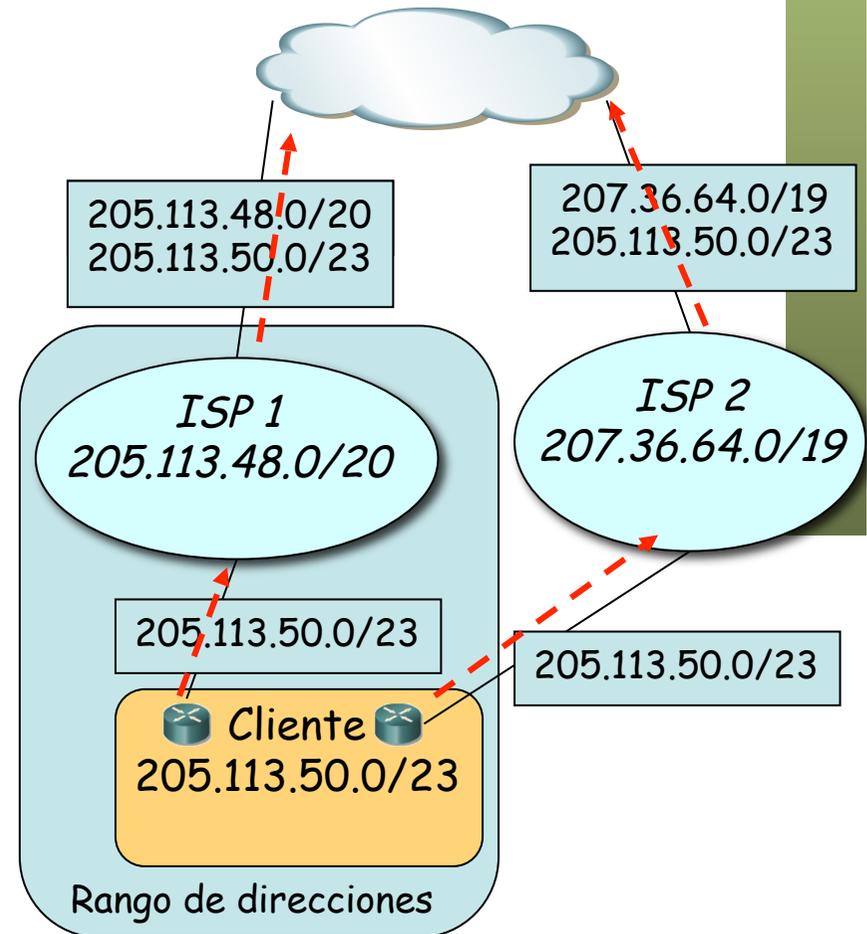
- No anunciar un destino a un vecino
- No usar caminos que pasen por cierto AS
- Ignorar el MED y usar *shortest-paths (hot potato routing)*
- Añadir varias veces su ASN
- Etc.

## Problemas

- Hay políticas que no convergen
- Hay políticas que pueden converger dependiendo del orden de los mensajes
- Hay políticas que convergen pero dejan de hacerlo si un enlace se cae
- Dadas las políticas y la topología, decidir si convergerá es NP-completo

# Multihoming

- Para ofrecer redundancia
- El rango de direcciones pertenece al ISP 1
- Habrá que anunciarlo también al ISP 2
- Ahora la ruta por ISP 2 es más específica
- *Address leaking*: ISP 1 debería anunciar también la ruta específica (...)



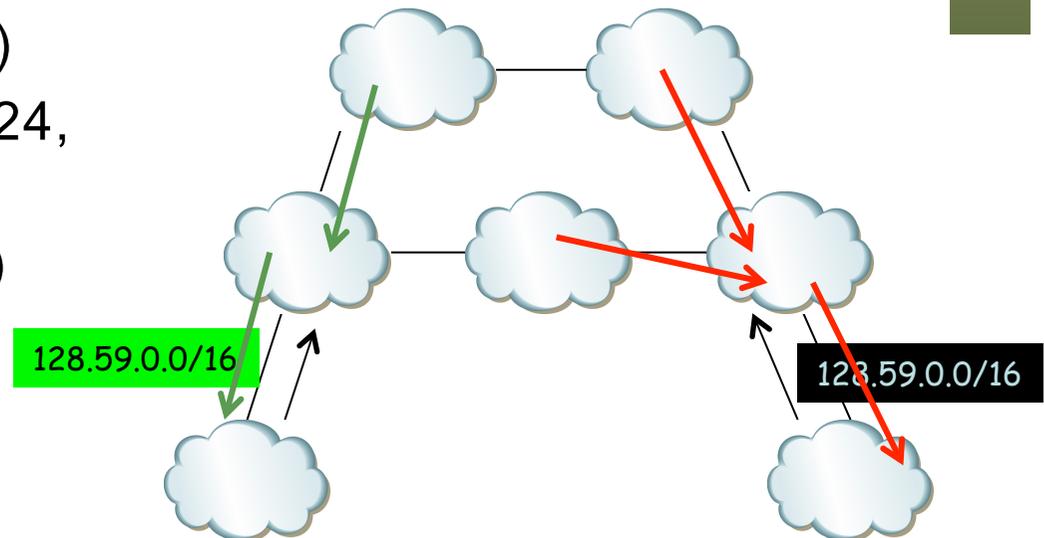
# Precauciones

## ***Martians***

- Algunos prefijos no se deben anunciar ni enrutar paquetes de ellos
- Ruta por defecto (0.0.0.0/0)
- Direccionamiento privado
  - 10.0.0.0/8
  - 172.16.0.0/12
  - 192.168.0.0/16
- *Link-local* (169.254.0.0/16)
- TEST\_NET (192.0.2.0/24, etc.)
- Clases D y E (224.0.0.0/3)
- Reservados para IANA

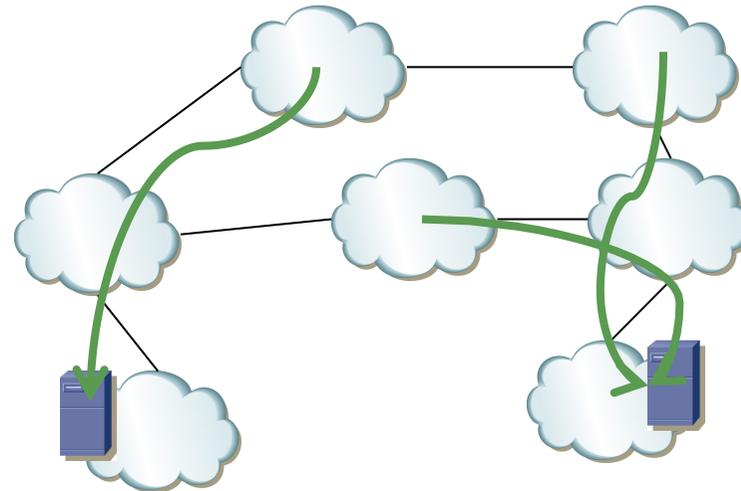
## ***Black holes***

- Si un AS anuncia un prefijo al que no está conectado
- El real puede dejar de ser accesible desde ciertas redes
- O puede hacer pasar tráfico por él



# Anycast

- Servidores con misma dirección IP (contenido replicado o no)
- Todos en la misma red física o en diferentes
- Anuncios por ejemplo por diferentes proveedores
- Clientes acceden a servidor según proximidad
- Permite distribución de contenidos
- También se puede hacer en el IGP
- Ejemplo: F-root name server



# Otras características

- Agregación de rutas
  - Gracias a CIDR
  - Combinar prefijos de dos o más ASs y anunciar el combinado
- *Route Reflectors*
  - Mejorar escalabilidad de IBGP (que crea un *full-mesh*)
  - Un router reflector actúa como un concentrador
- *Confederations*
  - Mejora escalabilidad de IBGP
  - Dividir AS en varios de forma que entre ellos sea EBGP
  - La confederación tiene un ASN y cada sub-AS puede tenerlo o usar uno privado
- *Route Flap Dampening*
  - Para evitar rápidas oscilaciones en una ruta
  - Aumenta el tiempo de convergencia

# Ejemplos

- Ver <http://www.routeviews.org>
- Probad un *Looking Glass*, por ejemplo: <http://www.rediris.es/red/lg/lg.pl>
- Ejemplo: Desde CIEMAT, AS PATH a 169.229.216.200 :

**Espere, por favor...**

**Please wait...**

```
169.229.0.0/16      *[BGP/170] 03:30:11, MED 390, localpref 160, from 130.206.206.250
                   AS path: 20965 11537 2153 25 I
                   [BGP/170] 03:30:12, MED 390, localpref 150
                   AS path: 20965 11537 2153 25 I
                   [BGP/170] 1w2d 10:32:33, MED 28040, localpref 110
                   AS path: 174 3356 3356 3356 2152 2152 2152 25 I
                   [BGP/170] 1w2d 10:32:34, MED 28040, localpref 110
                   AS path: 174 3356 3356 3356 2152 2152 2152 25 I
```

```
{master}
```

# Ejemplos

- Otro *Looking Glass*: <http://lg.cern.ch>
- Ejemplo: Desde EE1-EXT, show ip bgp summary:

## BGP4 Summary

```
Router ID: 192.65.184.1   Local AS Number: 513
Confederation Identifier: not configured
Confederation Peers:
Cluster ID: 513
Maximum Number of IP ECMP Paths Supported for Load Sharing: 4
Number of Neighbors Configured: 14, UP: 14
Number of Routes Installed: 1277047, Uses 109826042 bytes
Number of Routes Advertising to All Neighbors: 1475685 (909601 entries), Uses 43660848 bytes
Number of Attribute Entries Installed: 433212, Uses 38989080 bytes
```

Neighbor Address	AS#	State	Time	Rt:Accepted	Filtered	Sent	ToSend
192.65.184.2	513	ESTAB	6d11h20m	345414	0	115703	0
192.65.184.4	513	ESTAB	26d 9h 9m	0	0	454748	0
192.65.184.6	513	ESTAB	37d 2h59m	4397	0	450358	0
192.65.184.154	1297	ESTAB	13d 3h 8m	165	0	454739	0
192.65.184.210	559	ESTAB	37d 2h 7m	12204	0	23	0
192.65.184.221	559	ESTAB	13d 3h 9m	12204	0	23	0
194.12.145.29	513	ESTAB	37d 2h15m	9	0	1	0
194.12.145.33	513	ESTAB	37d 2h15m	9	0	1	0
194.12.145.37	513	ESTAB	6d11h19m	9	0	1	0
194.12.145.41	513	ESTAB	6d11h19m	9	0	1	0
198.32.11.85	11537	ESTAB	13d 3h 8m	13097	0	23	0
198.124.216.213	293	ESTAB	7d 2h 2m	152	0	23	0
212.203.96.129	8220	ESTAB	37d 2h 4m	446732	0	23	0
213.242.73.21	3356	ESTAB	37d 2h 1m	442636	0	23	0

# Ejemplos

- Otro *Looking Glass*: <http://lg.cern.ch>
- Ejemplo: Desde EE1-EXT, show ip bgp 130.206.162.158:

```

Number of BGP Routes matching display condition : 6
Status codes: s suppressed, d damped, h history, * valid, > best, i internal
Origin codes: i - IGP, e - EGP, ? - incomplete
  Network          Next Hop          MED    LocPrf    Weight Path
*> 130.206.0.0/16  192.65.184.210   20     65000     100    559 20965 766 i
*i 130.206.0.0/16  192.65.184.2     20     65000     100    559 20965 766 i
*  130.206.0.0/16  192.65.184.221  20     65000     100    559 20965 766 i
*  130.206.0.0/16  198.32.11.85    30     65000     100    11537 11537 20965 766 i
*  130.206.0.0/16  212.203.96.129  30     64000     100    8220 8220 8220 766 i
*  130.206.0.0/16  213.242.73.21   30     64000     100    3356 3356 3356 766 766 766 766 766 i

  Last update to IP routing table: 4h49m54s, 2 path(s) installed:
  Route is advertised to 4 peers:
  192.65.184.154(1297)    192.65.184.2(513)    192.65.184.4(513)    192.65.184.6(513)
  
```

# Resumen

- Es el EGP empleado en Internet
- Path Vector
- I-BGP y E-BGP
- La influencia de las políticas es crítica