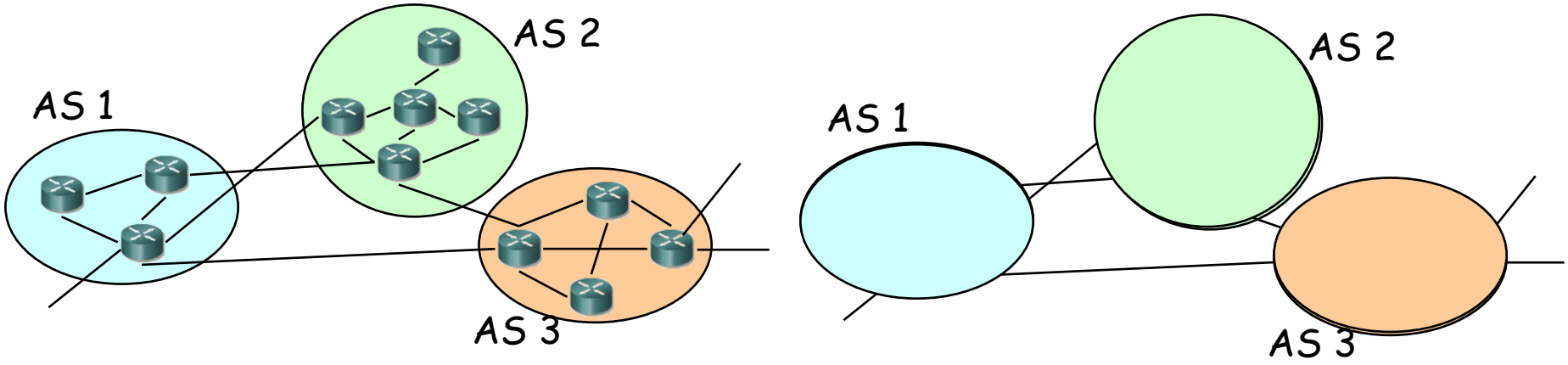
BGP-4

Enrutamiento jerárquico

- ¿Un solo grafo para toda la Internet?
  - Problemas de escala
  - Problemas de coordinación (¿métrica?)
- Enrutamiento jerárquico
  - IGP: Interior Gateway Protocol
  - EGP: Exterior Gateway Protocol
  - Interior/exterior respecto a "sistemas autónomos" (*Autonomous Systems*)
  - *"An AS is a connected group of one or more IP prefixes run by one or more network operators which has a SINGLE and CLEARLY DEFINED routing policy" (BCP 6)*

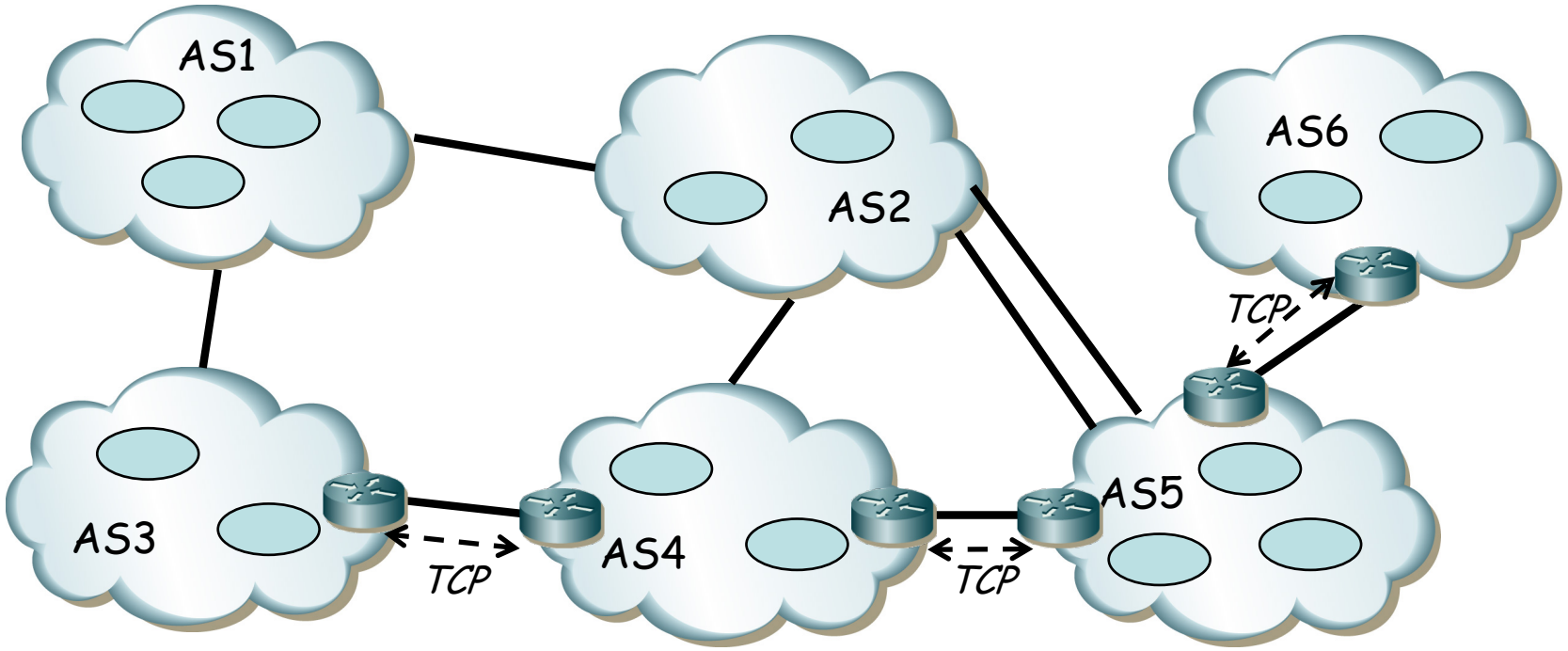


upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Redes de Nueva Generación
Área de Ingeniería Telemática

BGP: Introducción

BGP

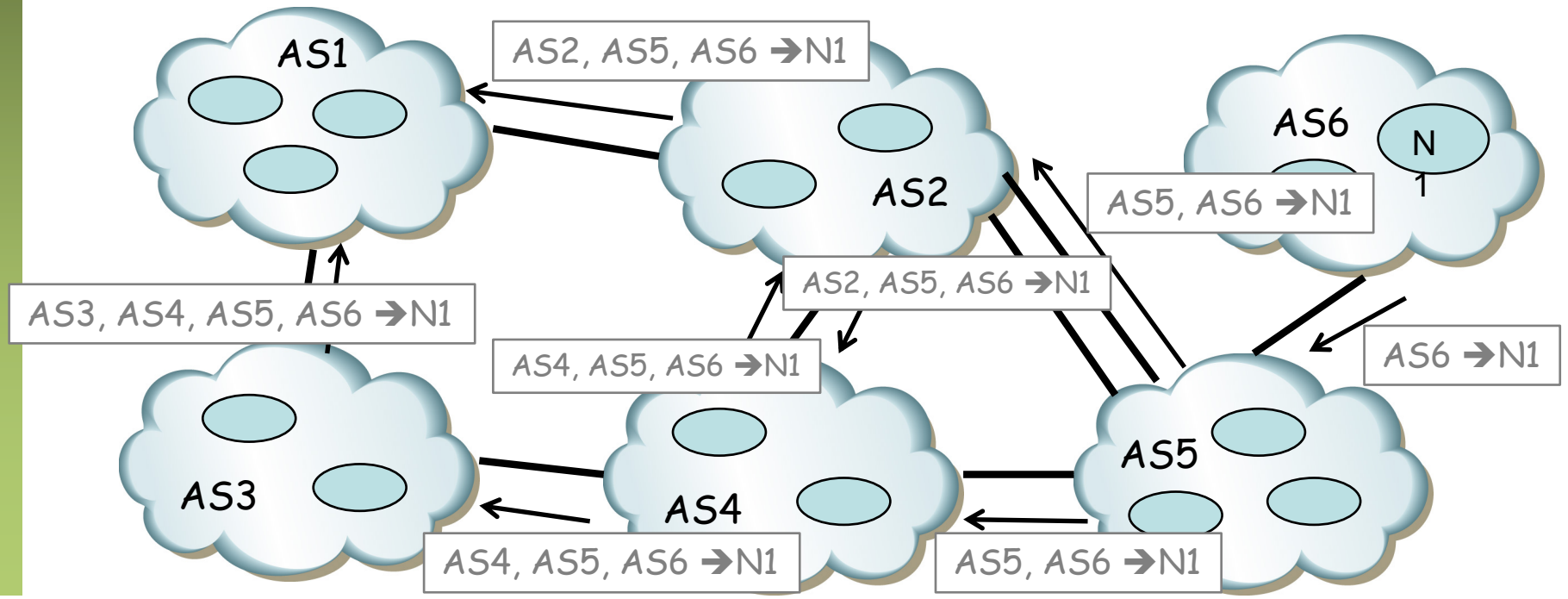
- *Border Gateway Protocol*
- BGP-4, RFC 4271
- BGP-4 primera versión classless
- Protocolo Interdomain estándar *de facto*
- Comunicación fiable mediante conexión TCP entre routers adyacentes
- Puerto 179



BGP

Path Vector

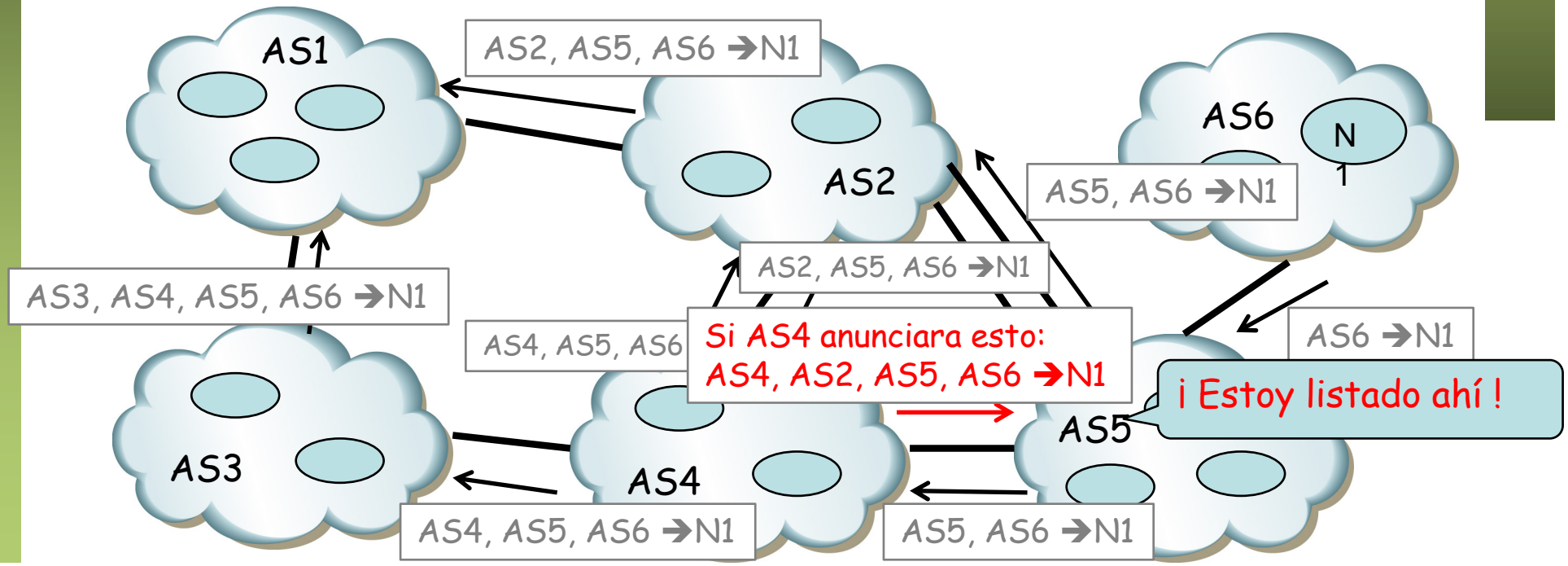
- Calcula caminos a prefijos
- Como DV recibe de vecinos, calcula sus rutas y envía a vecinos
- En vez de métrica anuncia la lista de AS en cada camino (. . .)
- Por defecto elige el camino que pasa por menor número de ASs



BGP

Path Vector

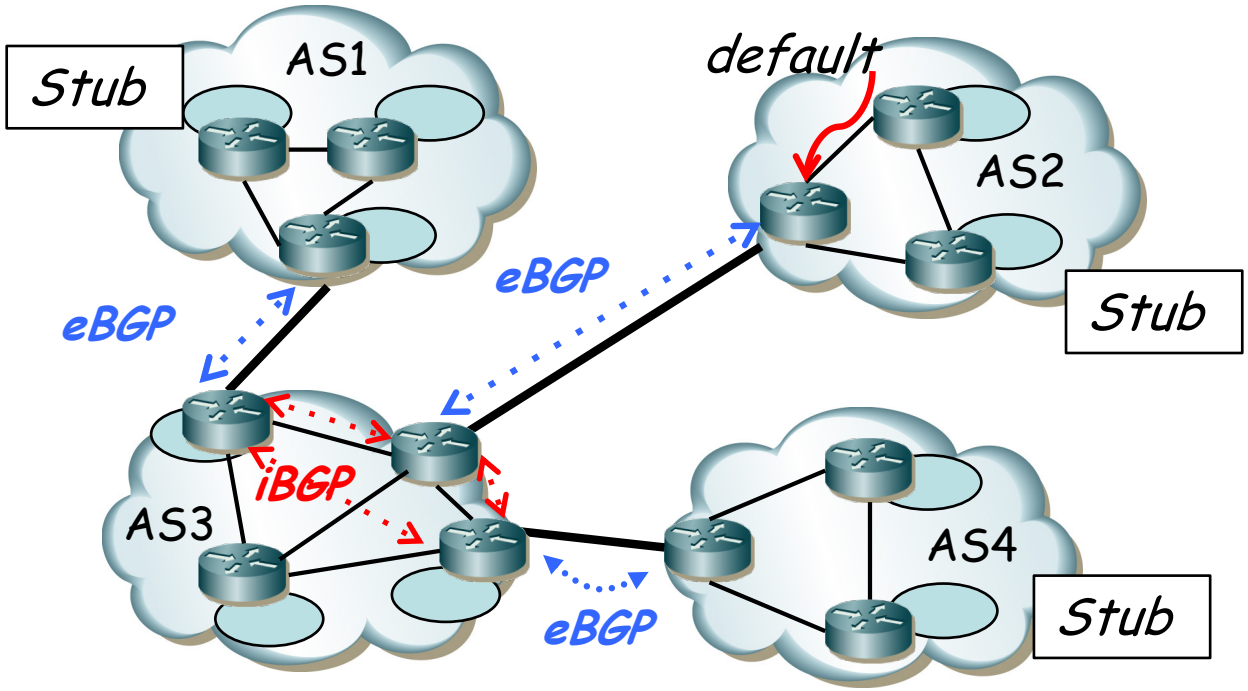
- Anunciar el camino permite evitar los ciclos
- El menor número de ASs no quiere decir que sea el menor número de saltos por routers



eBGP vs iBGP

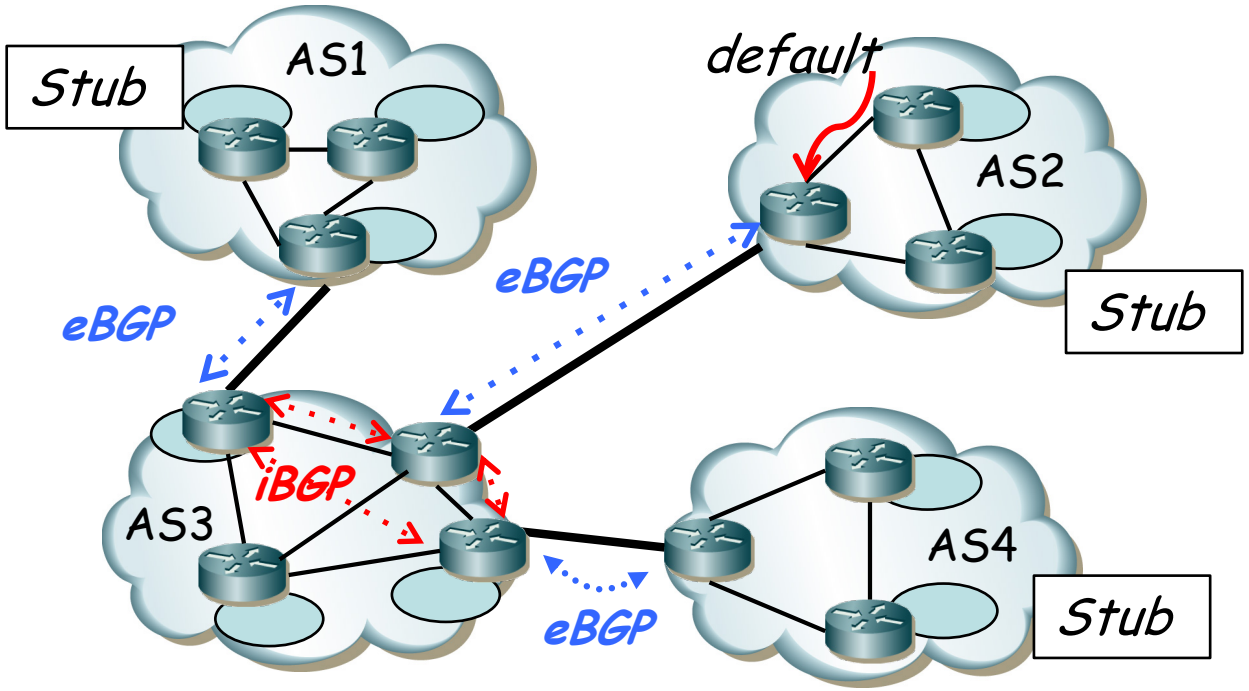
Peering en BGP

- Los *peers* de un proceso BGP pueden estar:
  - En otro AS: *external peer* ⇒ ***eBGP***
  - En el mismo AS: *internal peer* ⇒ ***iBGP***
- (…)



Peering en BGP

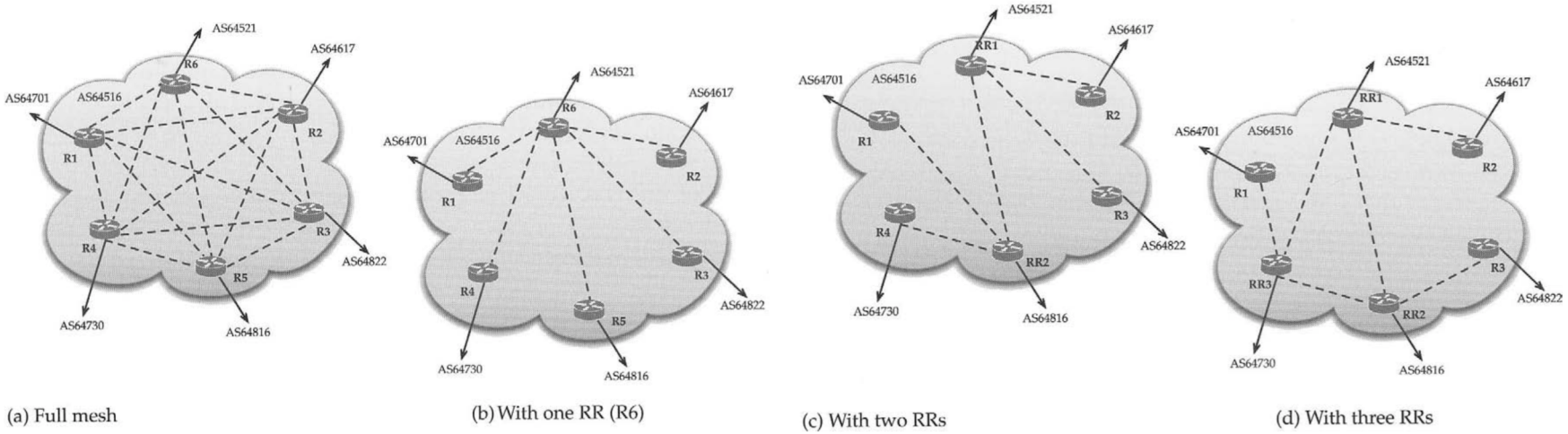
- En el mismo AS el *peering* iBGP forma una malla porque...
- No se pasan por iBGP prefijos aprendidos por iBGP
- Reconoce si es del mismo AS porque en el OPEN anuncia el ASN
- No interesa difundir todas las rutas al IGP (escalabilidad)
- iBGP permite que otros ASBRs aprendan los prefijos a anunciar
- El ASN se añade a la ruta al hacer anuncio a otro *eBGP*



ASBR = Autonomous System Border Router

Route Reflectors

- Problema de escalabilidad en iBGP debido al full-mesh
- RFC 4456 “BGP Route Reflection: An Alternative to Full Mech Internal BGP (IBGP)”
- En lugar de *full-mesh* conectan todos con el RR del *cluster*
- El RR sí reenvía rutas aprendidas por iBGP
- Un RR puede ser un cliente para otro RR

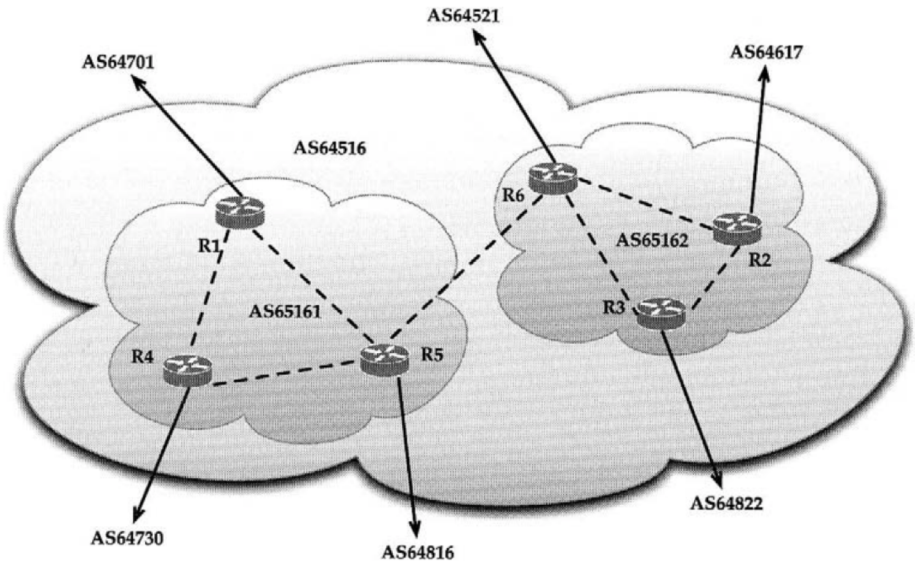


(a) Full mesh (b) With one RR (R6) (c) With two RRs (d) With three RRs

D. Medhi, K. Ramasamy, “Network Routing”, Ed. Elsevier

Confederations

- Otra solución al problema de escalabilidad de iBGP
- Internamente el AS se divide en sub-ASs, por ejempo con ASNs privados
- Externamente se anuncia como un solo AS (el identificador de la Confederación)
- Internamente hay *full-mesh* en cada sub-AS pero no globalmente al AS
- La estructura interna no es visible externamente



D. Medhi, K. Ramasamy, "Network Routing", Ed. Elsevier

Atributos en BGP

Path Attributes

- Son características de una ruta BGP, incluidos en el anuncio de la misma

Tipos según se soporten:

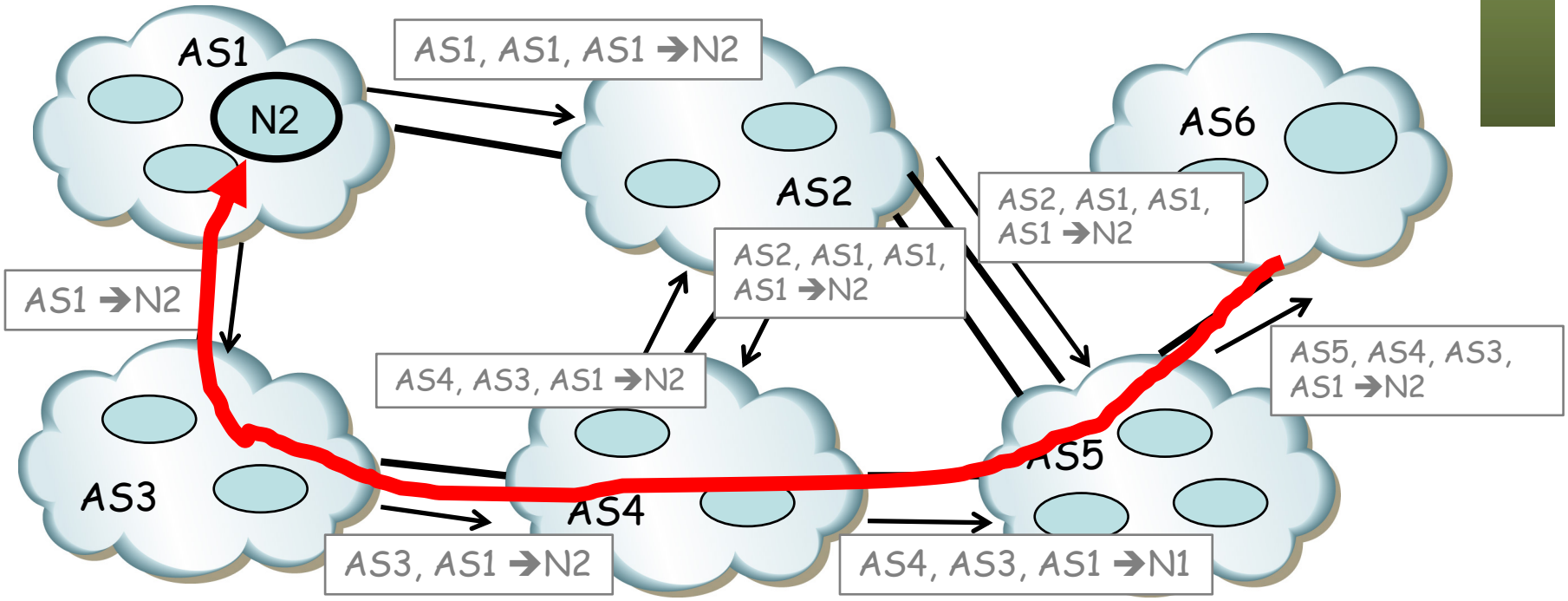
- *Well-known: mandatory* (en update) o *discretionary*
- *Optional: transitive* o *nontransitive*

"well-known" : Debe soportarlo
"Optional" : No está obligado a soportarlo
"mandatory" : Debe aparecer en los mensajes
"discretionary" : Puede no aparecer en los mensajes
"Transitive" : Debe reenviarlo
"Nontransitive" : No debe reenviarlo

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 (. . .)



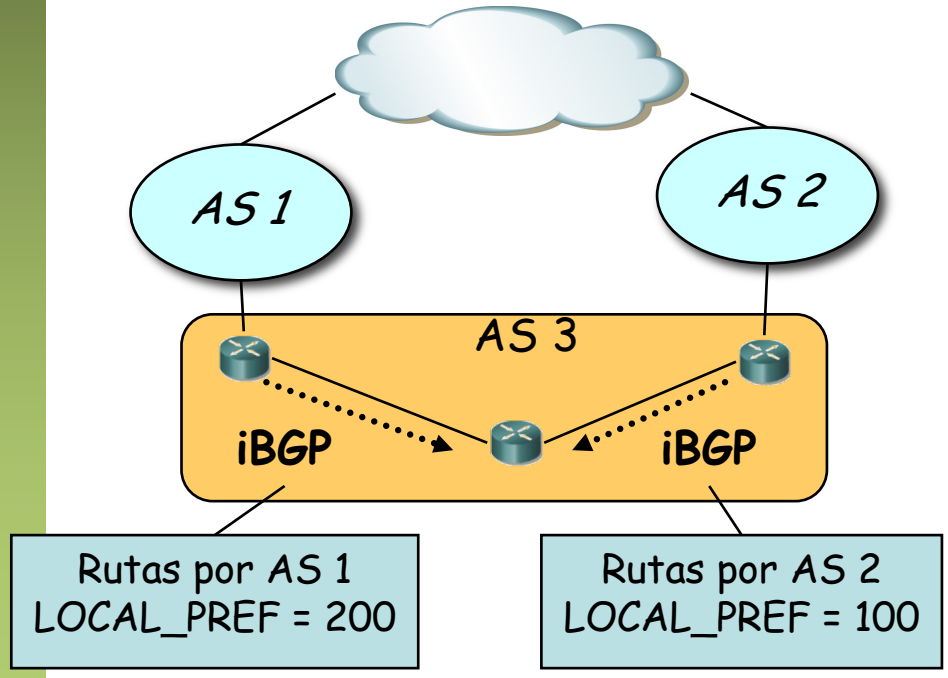
Path Attributes

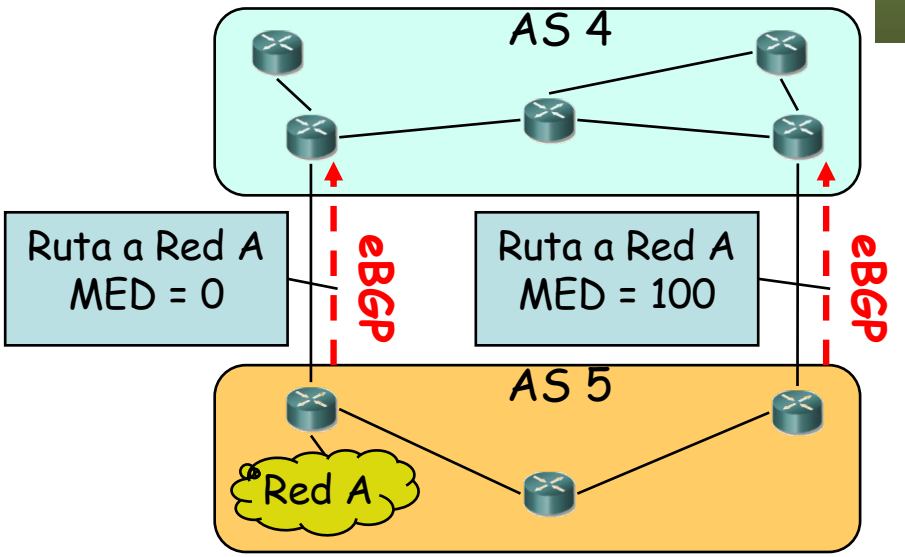
LOCAL_PREF (well-known discretionary, nontransitive)

- 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

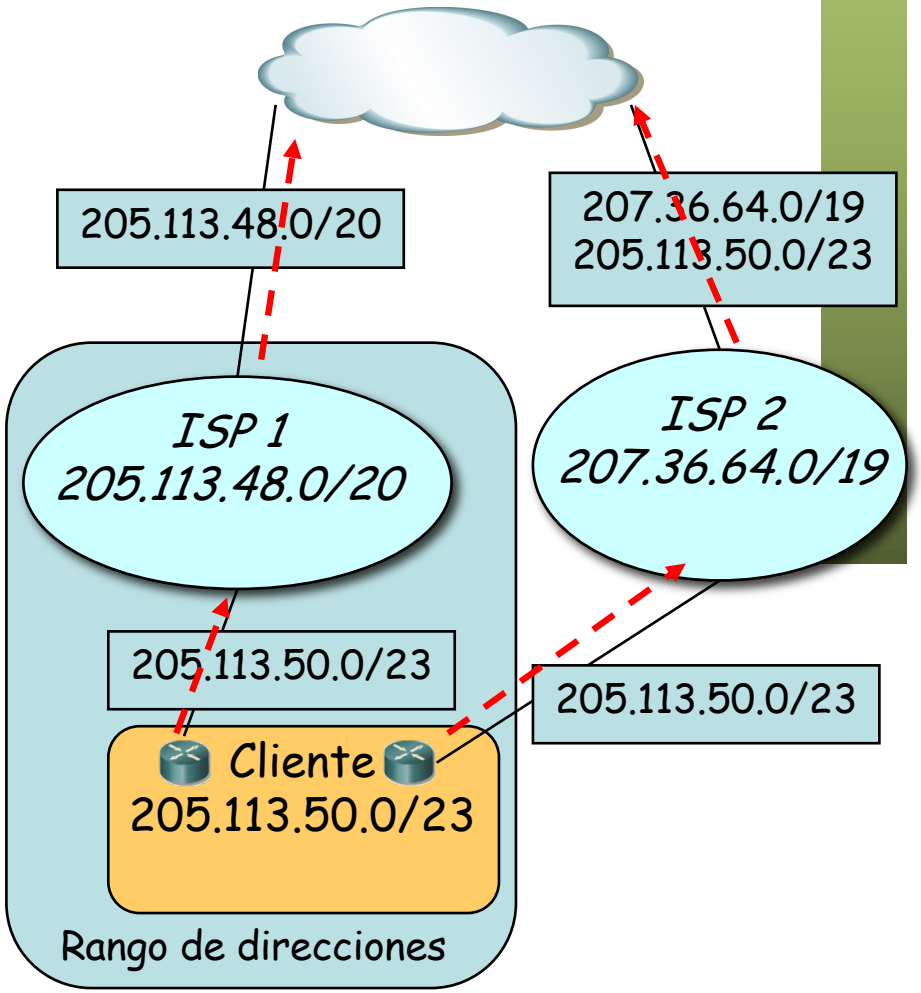
1. Ruta con el mayor **LOCAL_PREF**
2. Si iguales, la ruta de **AS_PATH** más corto
3. Si iguales, la ruta de origen menor (**ORIGIN** IGP < EGP < Incomplete)
4. Si iguales y van al mismo AS, la de menor **MED**
5. Si igual, la de menor **métrica** del IGP hasta el NEXT_HOP
6. Si iguales y van al mismo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

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Redes de Nueva Generación
Área de Ingeniería Telemática

BGP e Internet

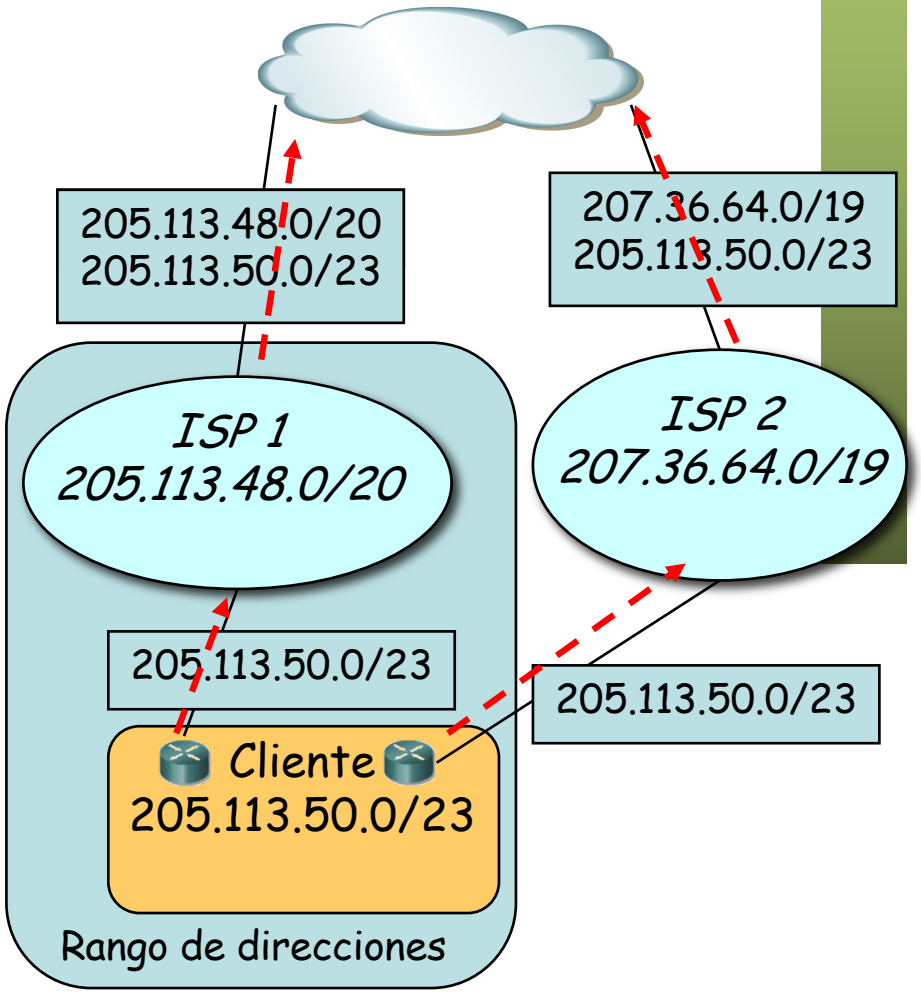
Multihoming

- 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 (…)



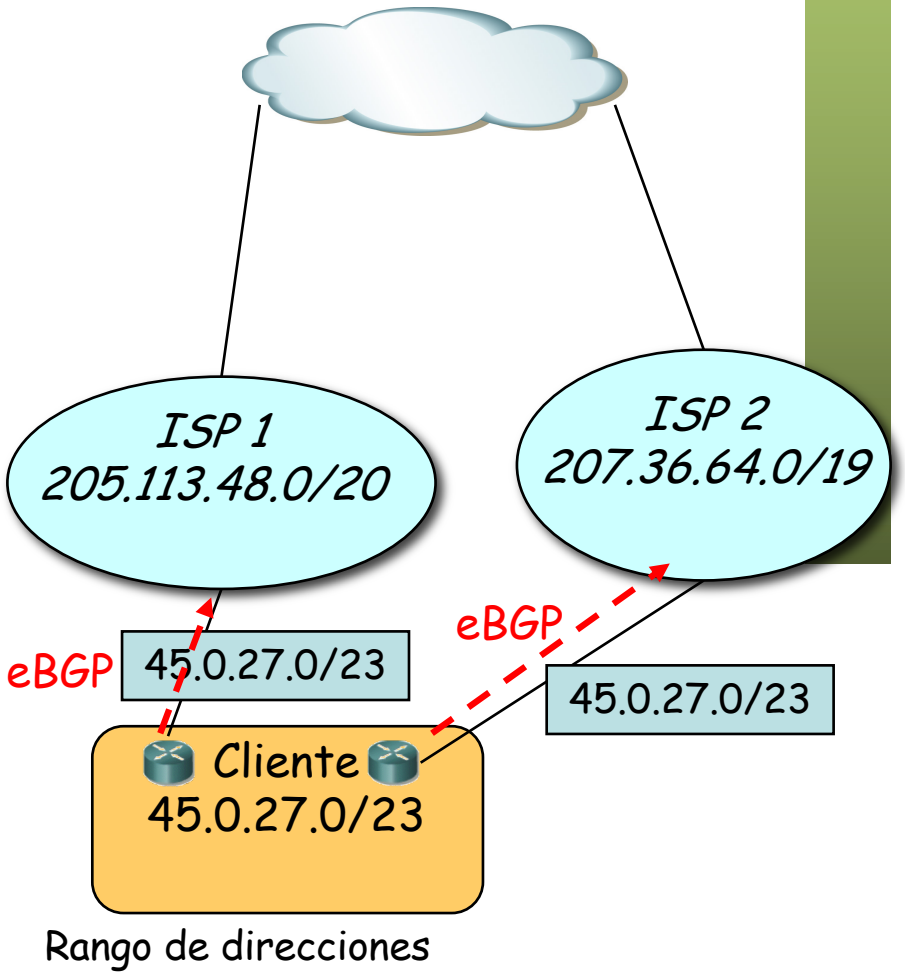
Multihoming

- 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
- Las dos de igual long. prefijo; anunciar 2x /24 permitiría forzar un camino
- (…)



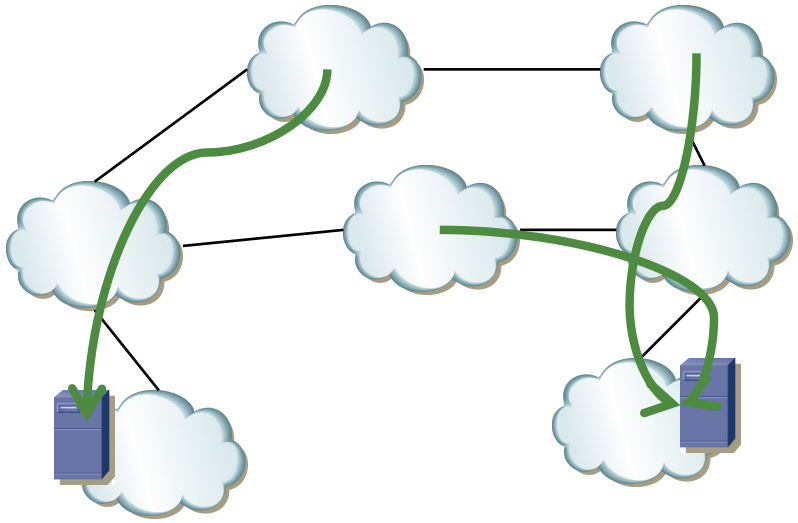
Multihoming

- 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
- Más habitual tener un espacio de direcciones propio
- Ser un AS y correr BGP



Anycast

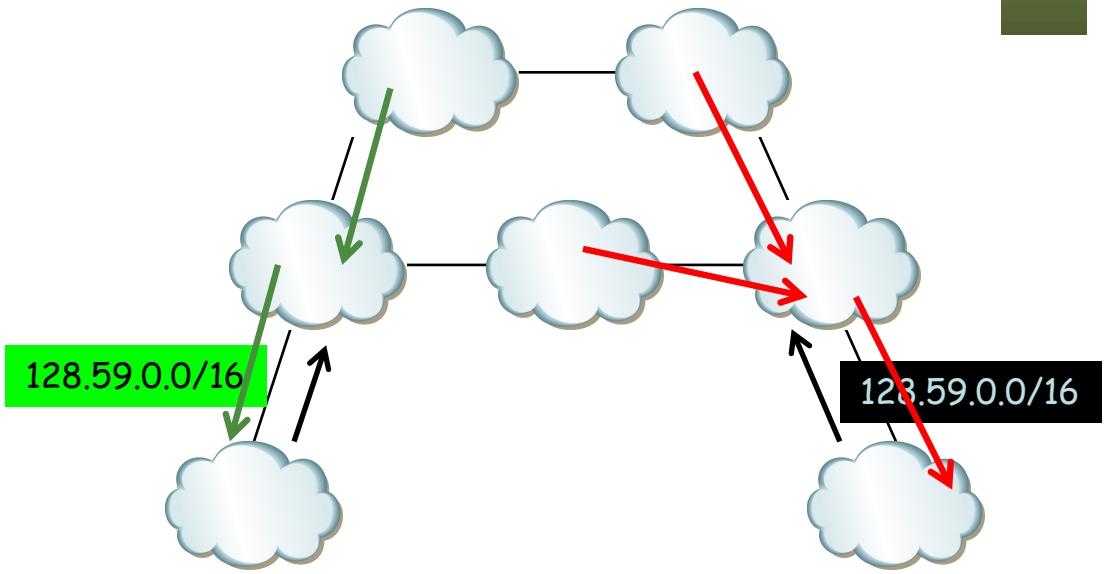
- Servidores con misma dirección IP (contenido replicado o no)
- Todos en la misma red física o en diferentes
- Anuncios por ejemplo por diferentes proveedores
- Clientes acceden a servidor según proximidad
- Permite distribución de contenidos
- También se puede hacer en el IGP
- Ejemplo: F-root name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



Precauciones

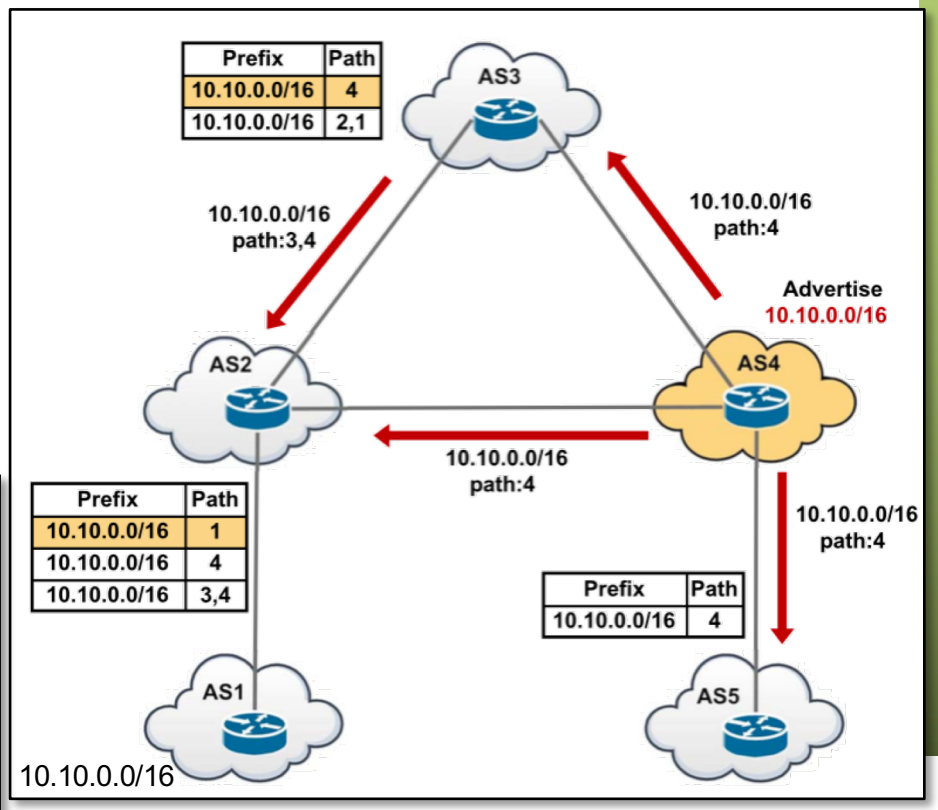
Black holes

- 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

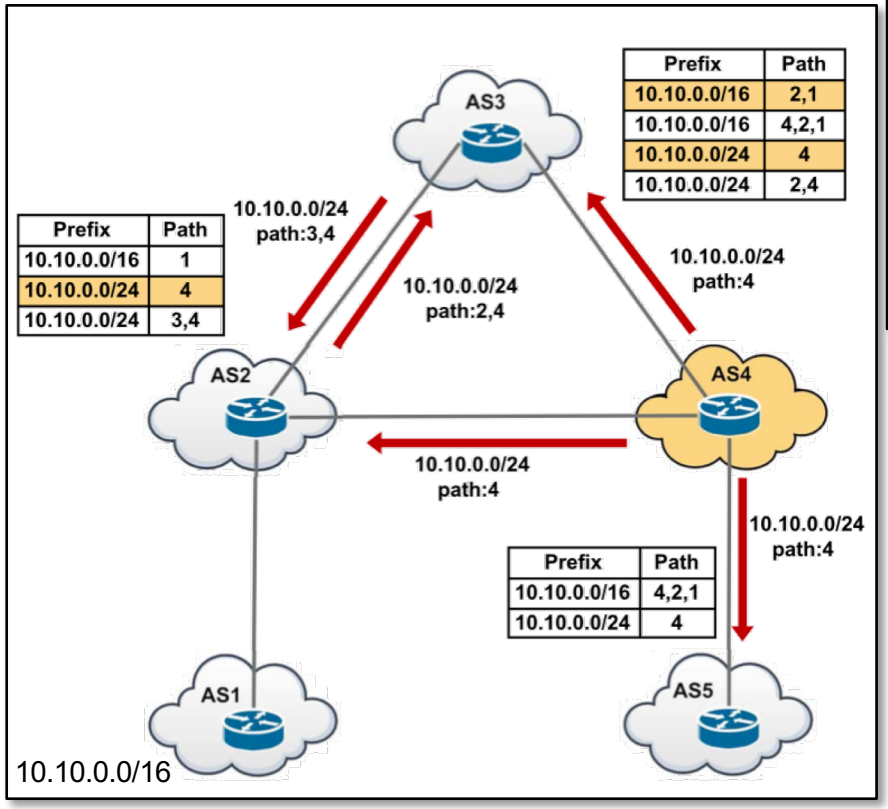


Ejemplos de problemas

- Prefix hijacking



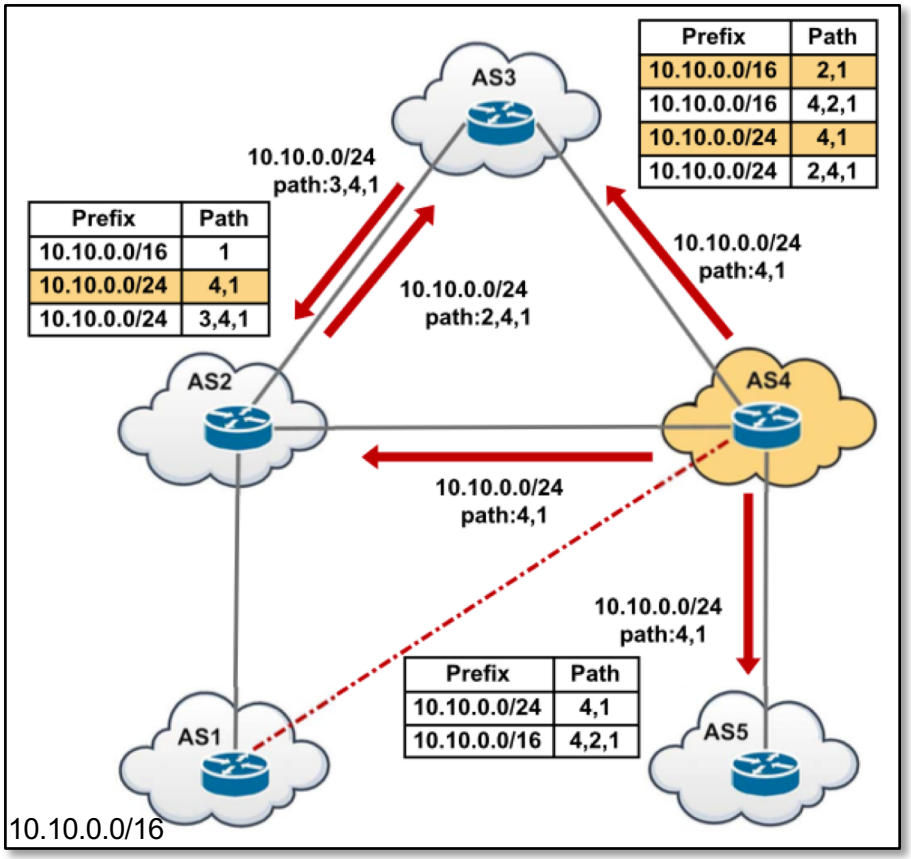
- Sub-prefix hijacking

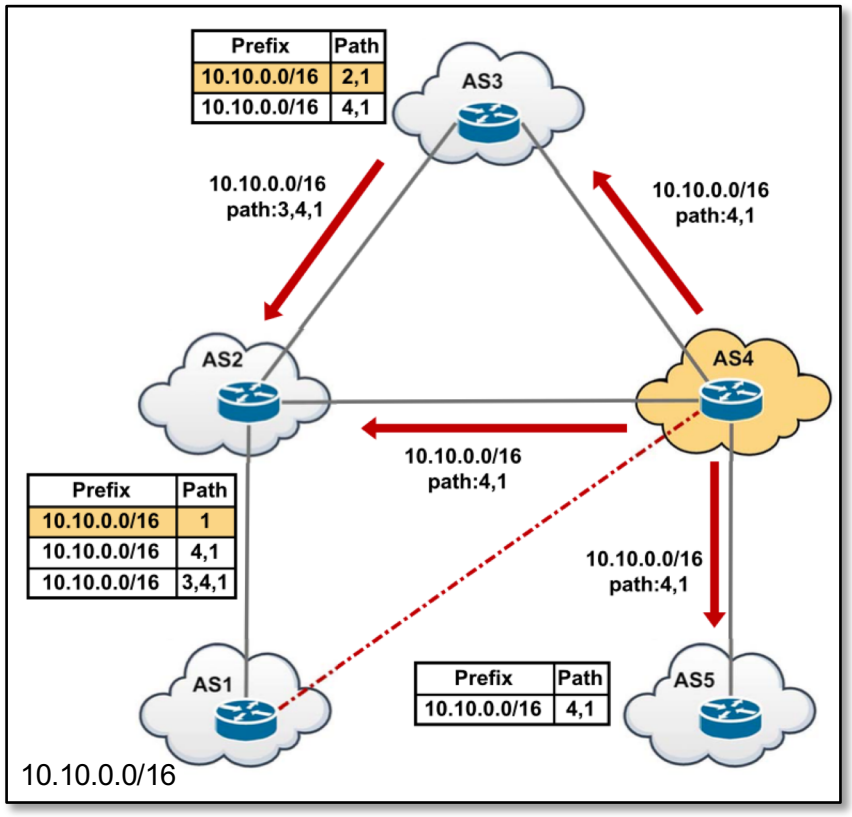


B. Al-Musawi, P. Branch y G. Armitage, “BGP Anomaly Detection Techniques: A Survey”, IEEE Communications Surveys & Tutorials, vol. 19 n.1

Ejemplos de problemas

- Prefix and its AS hijacking
- Sub-prefix and its AS hijacking

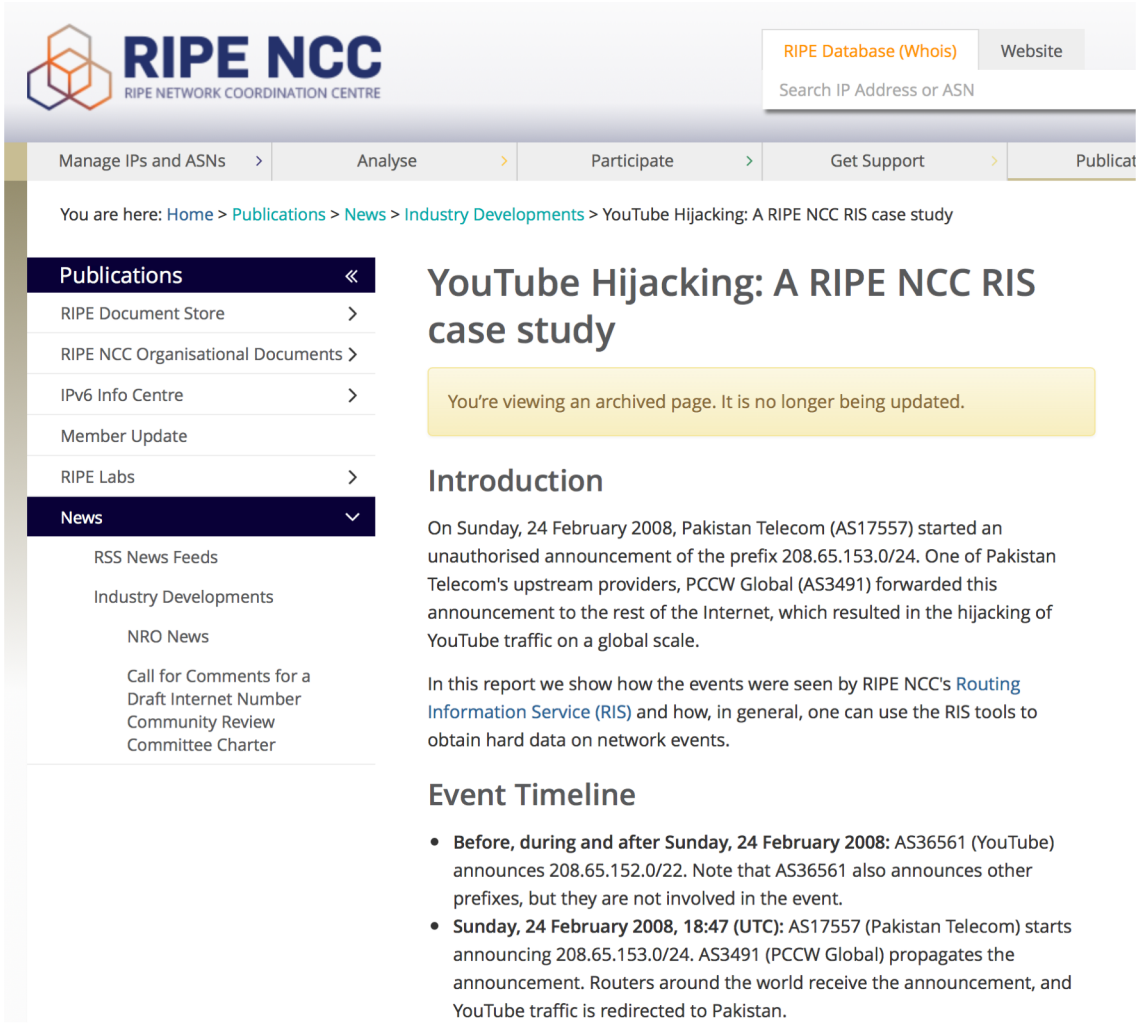




B. Al-Musawi, P. Branch y G. Armitage, "BGP Anomaly Detection Techniques: A Survey", IEEE Communications Surveys & Tutorials, vol. 19 n.1

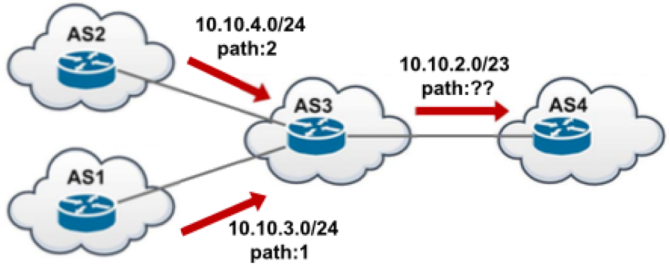
Ejemplo de secuestro

- YouTube Hijacking: A RIPE NCC RIS case study
  - https://www.ripe.net/publications/news/industry-developments/youtube-hijacking-a-ripe-ncc-ris-case-study
  - https://youtu.be/IzLPKuAOe50



Otros temas

- Agregación de rutas
  - Combinar prefijos de dos o más ASs y anunciar el combinado
  - Gracias a CIDR
  - Atributo para marcarlo
  - Se pierde detalle (en el AS_PATH)



- *Route Flap Dampening*
  - Para evitar rápidas oscilaciones en una ruta
  - Aumenta el tiempo de convergencia
  - Algunos RIRs han dejado de recomendarlo
- RouteViews
  - http://www.routeviews.org
  - Equipos haciendo peering con routers, solo para obtener los anuncios
  - (...)

Ejemplo RouteViews

- http://www.routeviews.org

```
$ telnet route-views.routeviews.org
Trying 128.223.51.103...
Connected to route-views.routeviews.org.
Escape character is '^]'.
******************************************************************
                    Oregon Exchange BGP Route Viewer
          route-views.oregon-ix.net / route-views.routeviews.org
BLA BLA BLA…

route-views>show ip bgp summary
BGP router identifier 128.223.51.103, local AS number 6447
BGP table version is 49790349, main routing table version 49790349
729435 network entries using 180899880 bytes of memory
28374069 path entries using 3404888280 bytes of memory
4384482/127370 BGP path/bestpath attribute entries using 1087351536 bytes of memory
4042805 BGP AS-PATH entries using 200965126 bytes of memory
3 BGP ATTR_SET entries using 120 bytes of memory
156666 BGP community entries using 19192598 bytes of memory
1165 BGP extended community entries using 53954 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 4893351374 total bytes of memory
BGP activity 2242697/1459312 prefixes, 269430605/238965595 paths, scan interval 60 secs
…
```

Ejemplo RouteViews

```
route-views>show ip route
show ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       bla bla bla…

Gateway of last resort is 128.223.51.1 to network 0.0.0.0

S*    0.0.0.0/0 [1/0] via 128.223.51.1
      1.0.0.0/8 is variably subnetted, 2652 subnets, 20 masks
B        1.0.0.0/24 [20/0] via 64.71.137.241, 14:06:54
B        1.0.4.0/22 [20/0] via 114.31.199.1, 2w6d
B        1.0.4.0/24 [20/0] via 114.31.199.1, 2w6d
B        1.0.5.0/24 [20/0] via 114.31.199.1, 2w6d
B        1.0.6.0/24 [20/0] via 114.31.199.1, 2w6d
B        1.0.7.0/24 [20/0] via 114.31.199.1, 2w6d
B        1.0.16.0/24 [20/0] via 202.232.0.2, 2d03h
B        1.0.64.0/18 [20/0] via 202.232.0.2, 1w5d
B        1.0.128.0/17 [20/0] via 64.71.137.241, 2d01h
B        1.0.128.0/18 [20/0] via 64.71.137.241, 2d01h
B        1.0.128.0/19 [20/0] via 64.71.137.241, 2d01h
B        1.0.128.0/24 [20/0] via 114.31.199.1, 1w5d
B        1.0.129.0/24 [20/0] via 64.71.137.241, 6d00h
B        1.0.131.0/24 [20/0] via 114.31.199.1, 1w5d
B        1.0.132.0/22 [20/0] via 114.31.199.1, 2d03h
B        1.0.136.0/24 [20/0] via 64.71.137.241, 2d03h
B        1.0.138.0/24 [20/0] via 64.71.137.241, 5d07h
B        1.0.139.0/24 [20/0] via 114.31.199.1, 1w5d
Bla bla bla…
```

Ejemplo RouteViews

```
route-views>show ip bgp 130.206.162.158 bestpath
BGP routing table entry for 130.206.0.0/16, version 36631792
Paths: (43 available, best #13, table default)
  Not advertised to any peer
  Refresh Epoch 1
  6939 766
    64.71.137.241 from 64.71.137.241 (216.218.252.164)
      Origin IGP, localpref 100, valid, external, best
      rx pathid: 0, tx pathid: 0x0
```

Ejemplo RouteViews

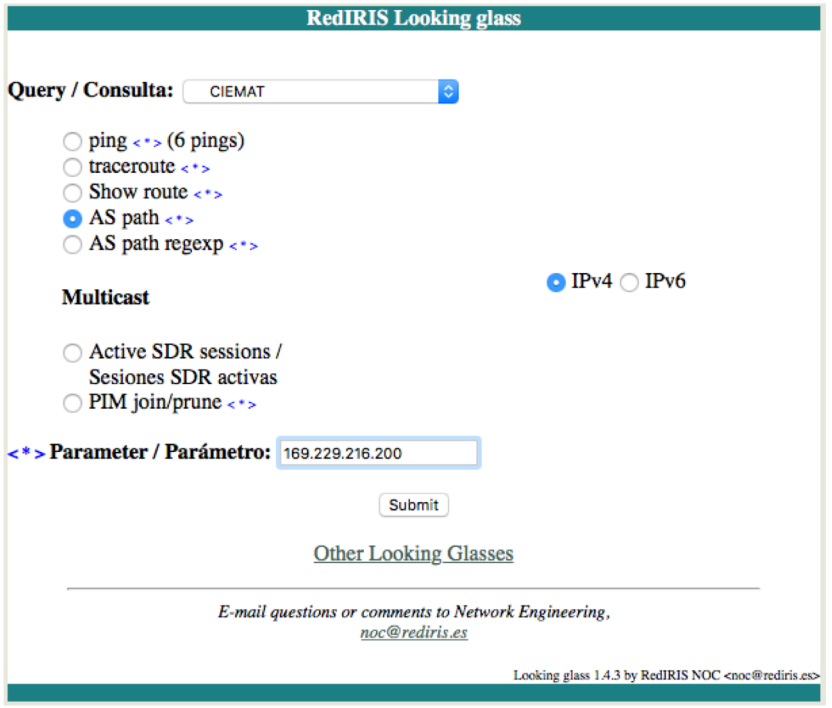
```
route-views>show ip bgp 130.206.162.158 bestpath
BGP routing table entry for 130.206.0.0/16, version 36631792
Paths: (43 available, best #13, table default)
  Not advertised to any peer
  Refresh Epoch 1
  49788 12552 2603 21320 766
    91.218.184.60 from 91.218.184.60 (91.218.184.60)
      Origin IGP, metric 0, localpref 100, valid, external
      Community: 12552:12000 12552:12100 12552:12101 12552:22000
      rx pathid: 0, tx pathid: 0
  Refresh Epoch 1
  2914 174 766 766 766 766 766
    129.250.1.66 from 129.250.1.66 (129.250.0.12)
      Origin IGP, metric 15, localpref 100, valid, external
      Community: 2914:420 2914:1008 2914:2000 2914:3000 65504:174
      rx pathid: 0, tx pathid: 0
  Refresh Epoch 1
  200130 2914 766
    95.85.0.2 from 95.85.0.2 (95.85.0.2)
      Origin IGP, localpref 100, valid, external
      Community: 2914:420 2914:1204 2914:2205 2914:3200 14061:2100 14061:2103 14061:4000
14061:4001 65504:766
      rx pathid: 0, tx pathid: 0
  Refresh Epoch 1
  202018 2914 766
    5.101.110.2 from 5.101.110.2 (5.101.110.2)
      Origin IGP, localpref 100, valid, external
      bla bla bla…
Bla bla bla… (hasta las 43 entradas)
```

Otros

- Internet Routing Registry
  - http://www.irr.net/docs/overview.html
- RIPE Routing Information Service (RIS)
  - https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris
- BGPmon
  - https://bgpmon.net
  - Comprado por OpenDNS

Otros

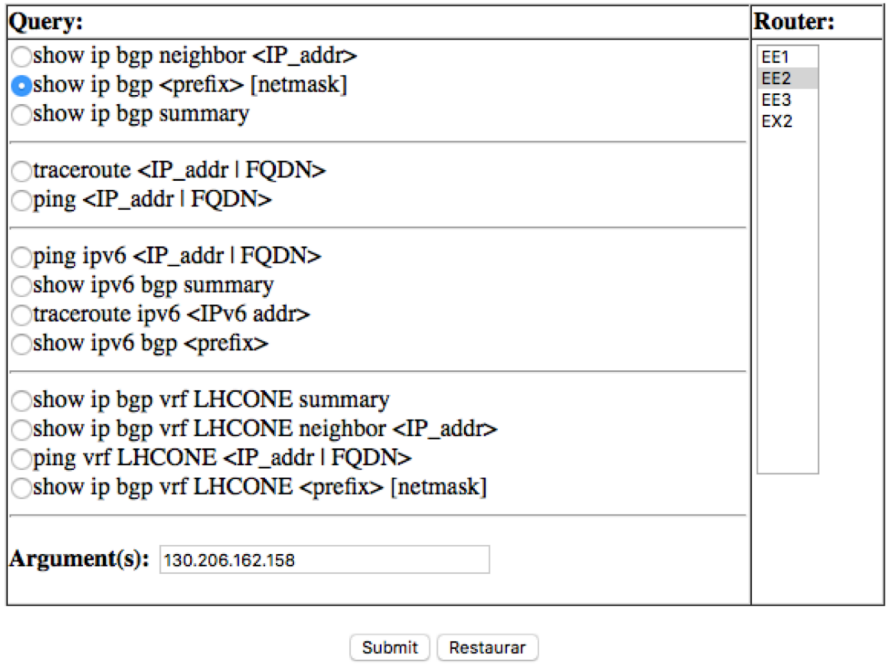
- Looking Glass
  - Por ejemplo: http://www.rediris.es/red/lg/




```
Espere, por favor...
Please wait...
169.229.0.0/16     *[BGP/170] 2d 18:26:41, MED 116, localpref 161, from 130.206.206.250
                      AS path: 20965 11537 2153 25 I, validation-state: unverified
                    [BGP/170] 2d 18:26:41, MED 142, localpref 150
                      AS path: 20965 11537 2153 25 I, validation-state: unverified
                    [BGP/170] 6d 01:47:08, MED 23040, localpref 110
                      AS path: 174 3356 3356 3356 2152 2152 2152 25 I, validation-state: unverified
                    [BGP/170] 6d 01:47:08, MED 23040, localpref 110
                      AS path: 174 3356 3356 3356 2152 2152 2152 25 I, validation-state: unverified
{master}
```

Otros

- Looking Glass
  - Por ejemplo: http://lg.cern.ch




```
Query: show ip bgp
Argument(s): 130.206.162.158

Number of BGP Routes matching display condition : 7
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
*>i 130.206.0.0/16     192.65.184.1    20         65000      100    20965 766 i
*   130.206.0.0/16     192.65.184.69   20         65000      100    559 20965 766 i
*   130.206.0.0/16     192.65.184.173  25         65000      100    559 20965 766 i
*   130.206.0.0/16     192.91.246.109  30         65000      100    10764 10764 11537 20965 766 i
*   130.206.0.0/16     192.91.246.125  30         65000      100    11537 11537 20965 766 i
*   130.206.0.0/16     62.179.22.53    20         64000      100    6830 6830 2603 21320 766 i
*   130.206.0.0/16     195.141.200.17  10         64000      100    6730 6730 2603 21320 766 i
       Last update to IP routing table: 6d19h11m40s, 1 path(s) installed:
       Route is advertised to 3 peers:
        192.65.184.3(513)         192.65.184.4(513)           192.65.184.24(513)
```