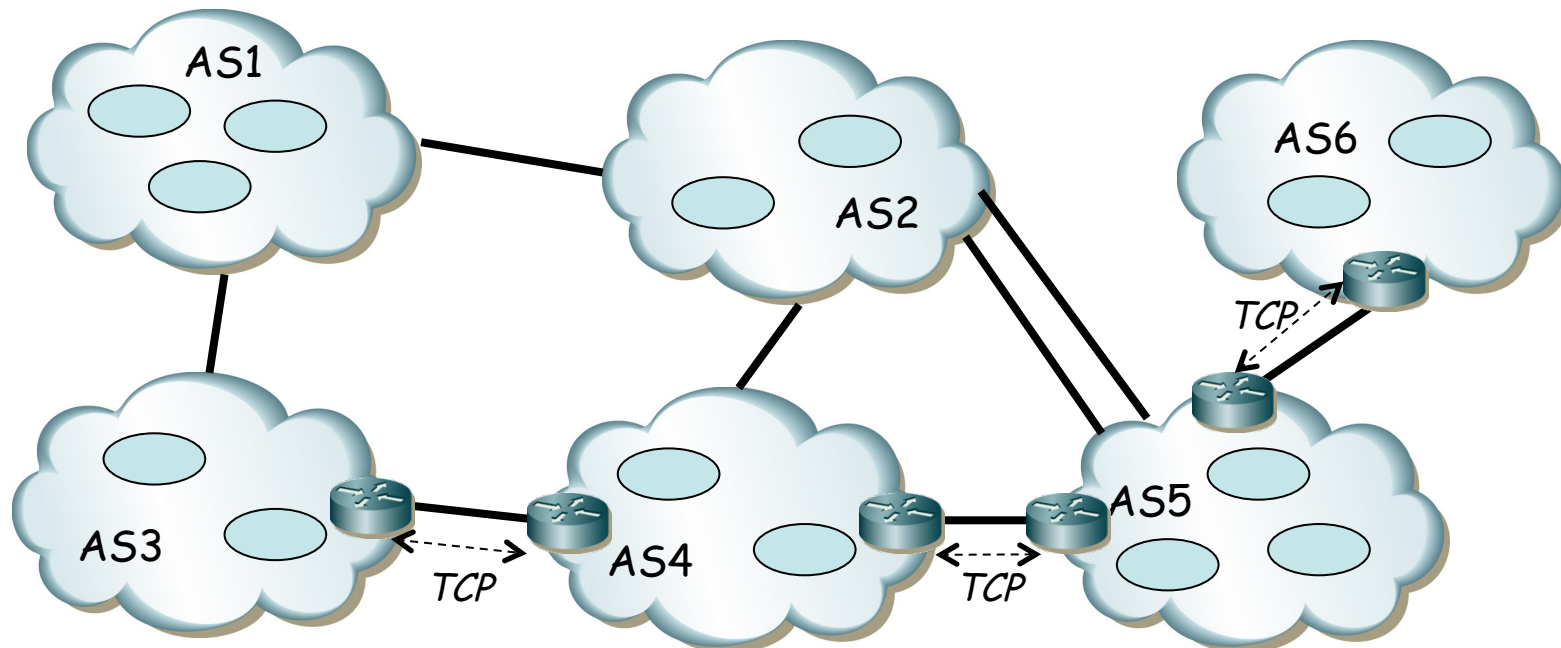


BGP-4

BGP: Introducción

BGP

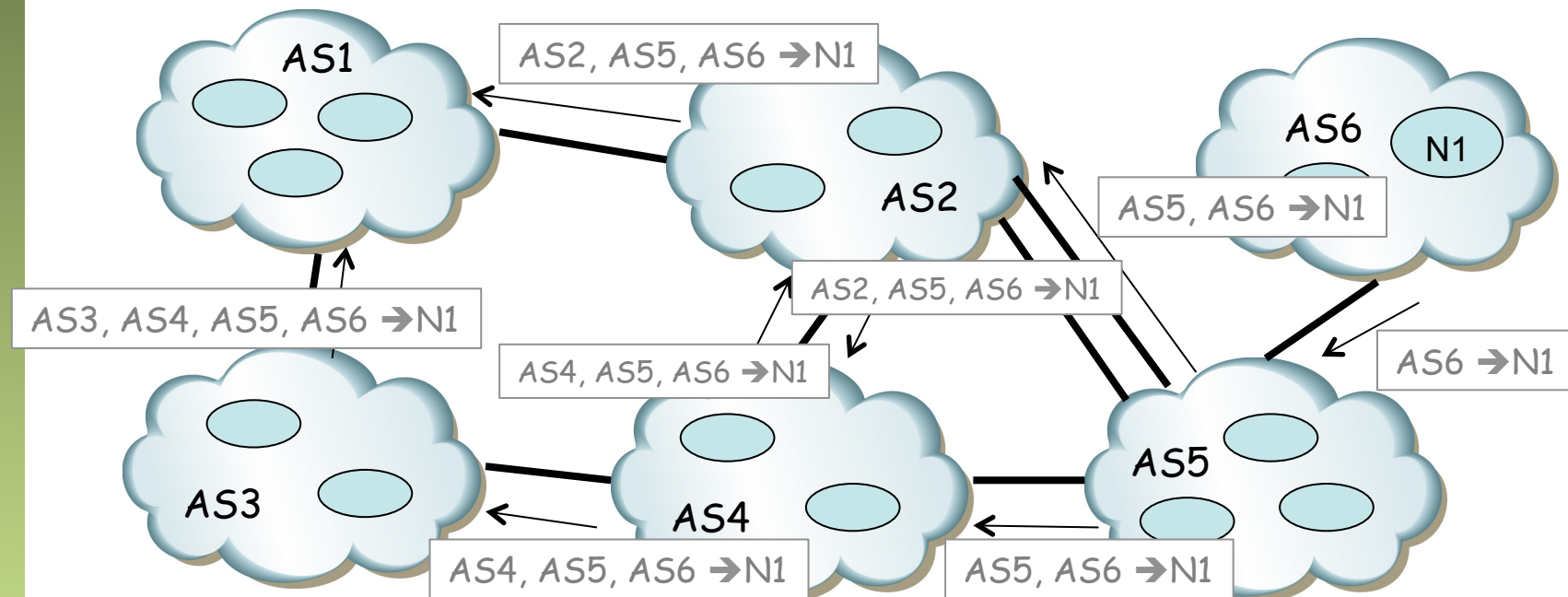
- *Border Gateway Protocol*
- BGP-4, RFC 4271
- BGP-4 primera versión classless
- Protocolo Interdomain estándar *de facto*
- Comunicación fiable mediante conexión TCP entre routers adyacentes
- Puerto 179



BGP

Path Vector

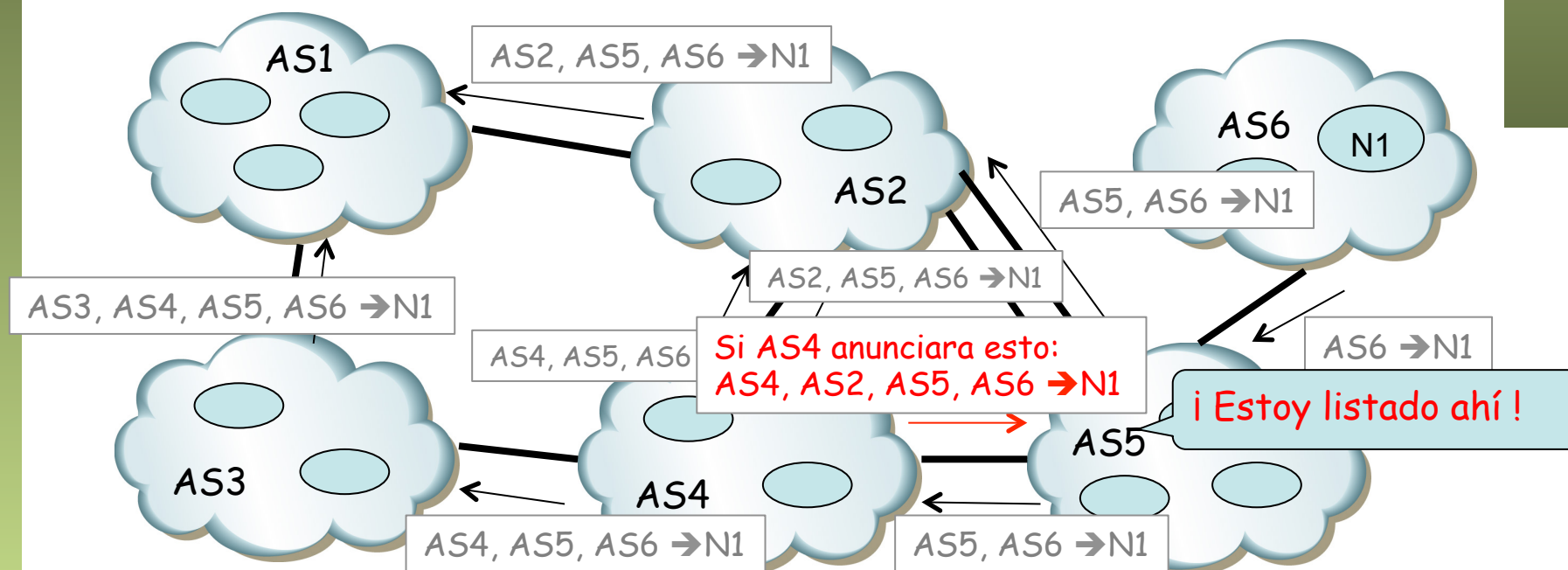
- Calcula caminos a prefijos
- Como DV recibe de vecinos, calcula sus rutas y envía a vecinos
- En vez de métrica anuncia la lista de AS en cada camino (. . .)
- Por defecto elige el camino que pasa por menor número de ASs



BGP

Path Vector

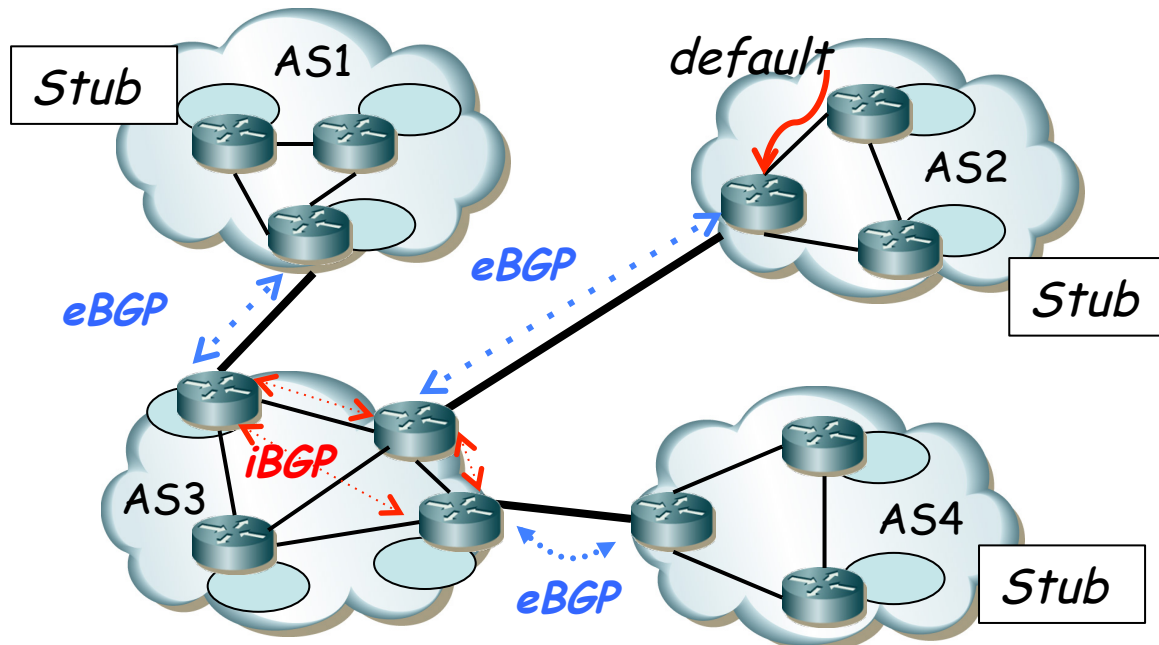
- Anunciar el camino permite evitar los ciclos
- El menor número de ASs no quiere decir que sea el menor número de saltos por routers



eBGP vs iBGP

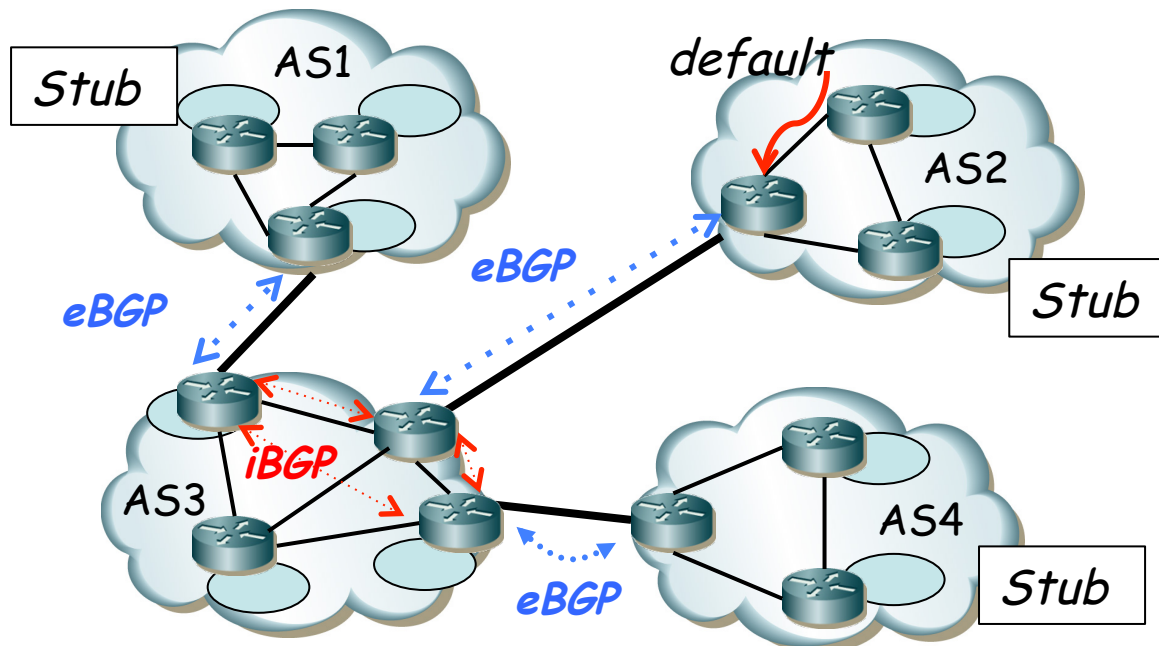
Peering en BGP

- Los *peers* de un proceso BGP pueden estar:
 - En otro AS: *external peer* $\Rightarrow$ **eBGP**
 - En el mismo AS: *internal peer* $\Rightarrow$ **iBGP**
- (...)



Peering en BGP

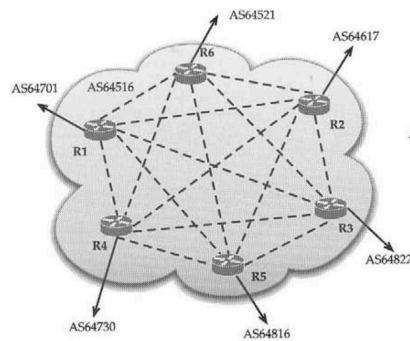
- En el mismo AS el *peering* iBGP forma una malla porque...
- No se pasan por iBGP prefijos aprendidos por iBGP
- Reconoce si es del mismo AS porque en el OPEN anuncia el ASN
- No interesa difundir todas las rutas al IGP (escalabilidad)
- iBGP permite que otros ASBRs aprendan los prefijos a anunciar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



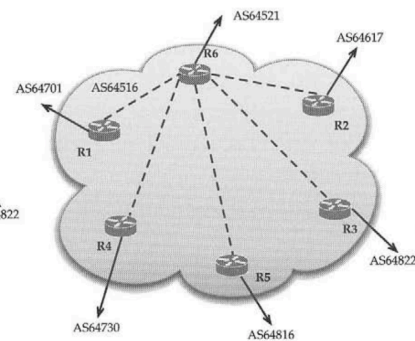
ASBR = Autonomous
System Border Router

Route Reflectors

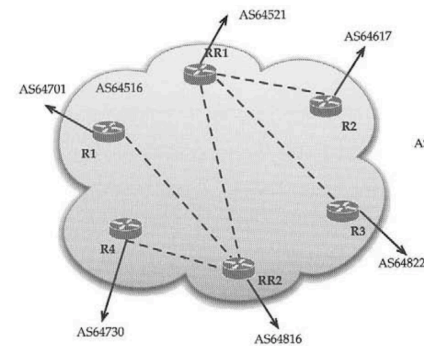
- Problema de escalabilidad en iBGP debido al full-mesh
- RFC 4456 “BGP Route Reflection: An Alternative to Full Mech Internal BGP (IBGP)”
- En lugar de *full-mesh* conectan todos con el RR del *cluster*
- El RR sí reenvía rutas aprendidas por iBGP
- Un RR puede ser un cliente para otro RR



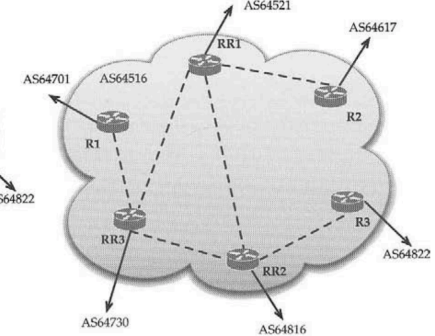
(a) Full mesh



(b) With one RR (R6)



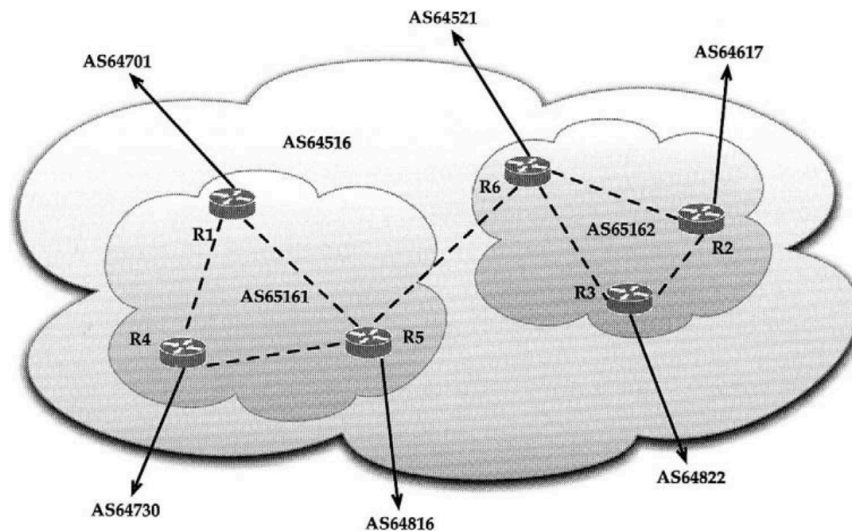
(c) With two RRs



(d) With three RRs

Confederations

- Otra solución al problema de escalabilidad de iBGP
- Internamente el AS se divide en sub-ASs, por ejemplo 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



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*

ORIGIN (well-known mandatory)

- IGP, EGP o Incompleto (rutas estáticas)

"*well-known*" : Debe soportarlo

"*Optional*" : No está obligado a soportarlo

"*mandatory*" : Debe aparecer en los mensajes

"*discretionary*" : Puede no aparecer en los mensajes

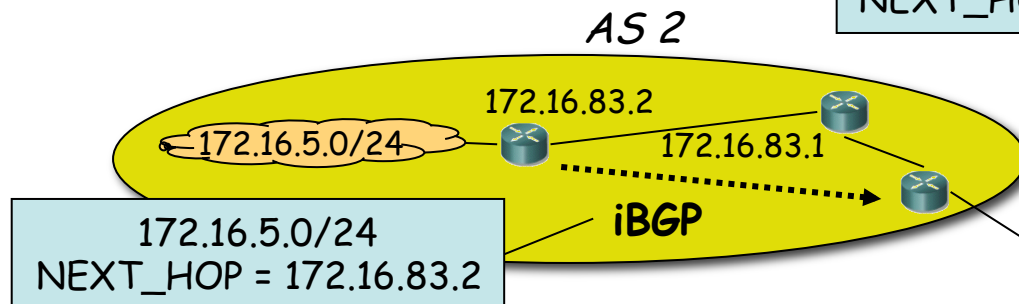
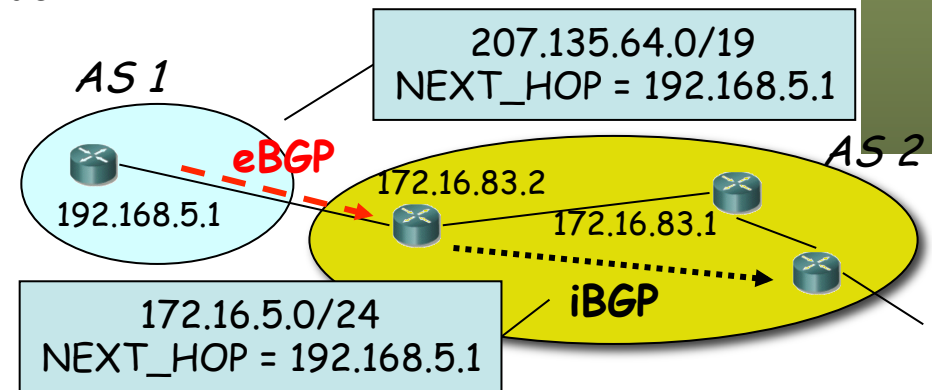
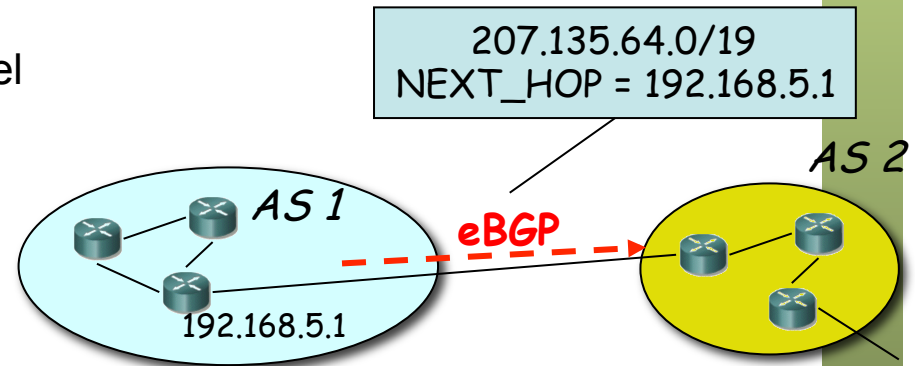
"*Transitive*" : Debe reenviarlo

"*Nontransitive*" : No debe reenviarlo

Path Attributes

NEXT_HOP (well-known mandatory)

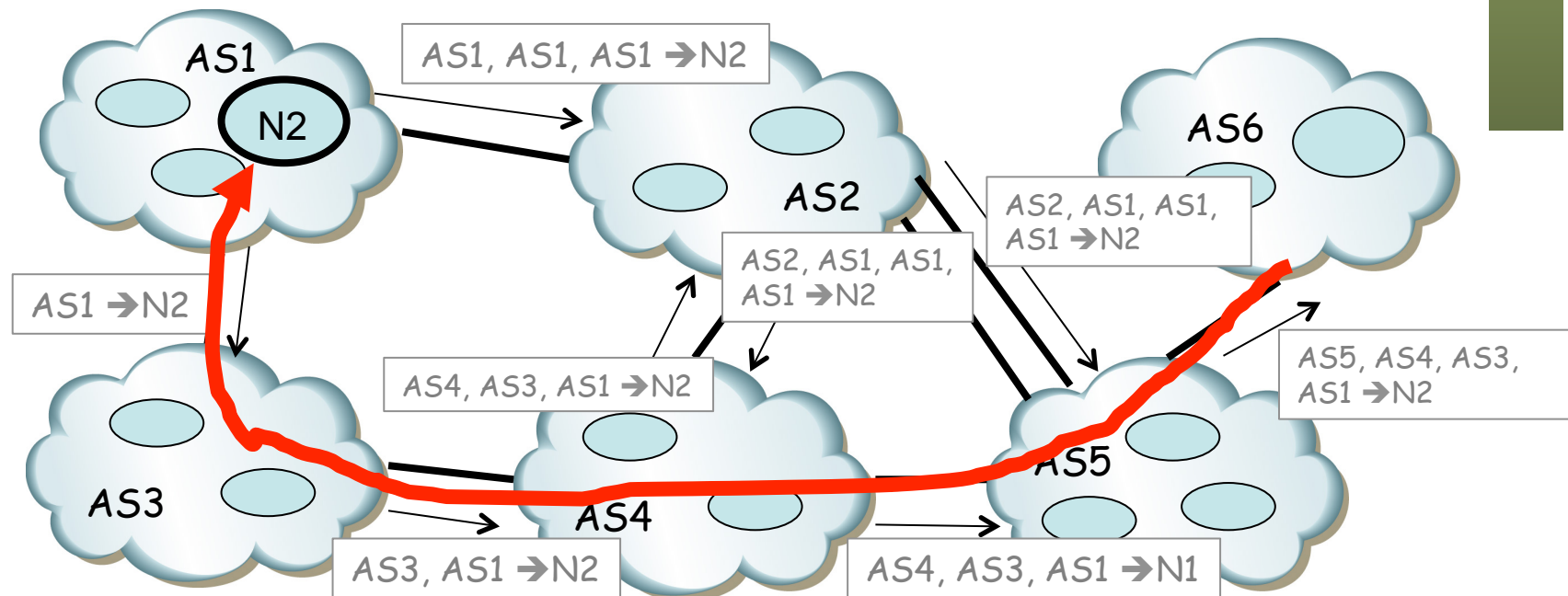
- Si son *External Peers* es la IP del interfaz del router anunciante
- Si son *Internal Peers* y
 - Destino fuera del AS: IP del *peer* externo
 - Destino en el mismo AS: IP del anunciante (*recursive lookups*)



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*: añadir el ASN *más veces* para desalentar usar este camino (. . .)



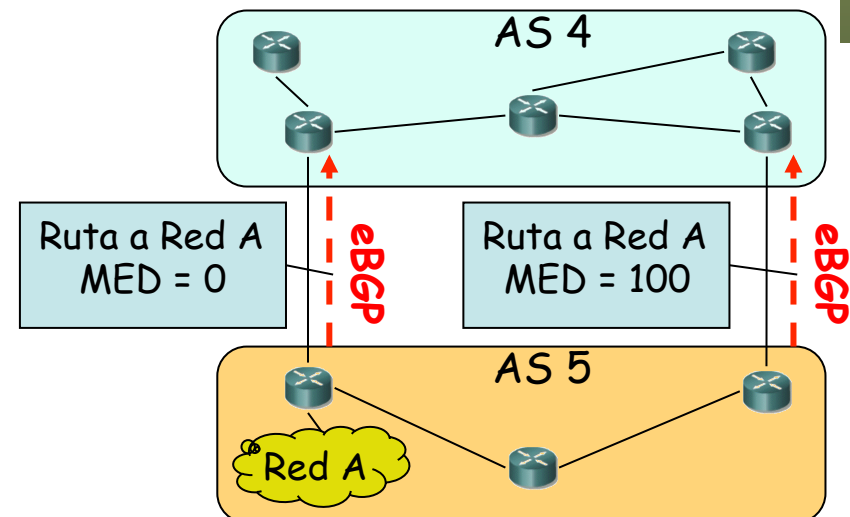
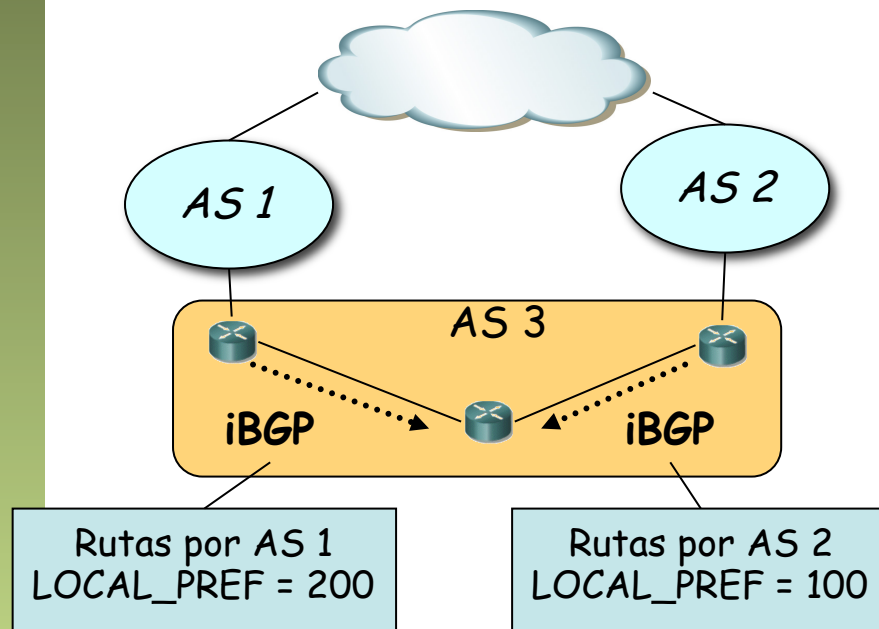
Path Attributes

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)



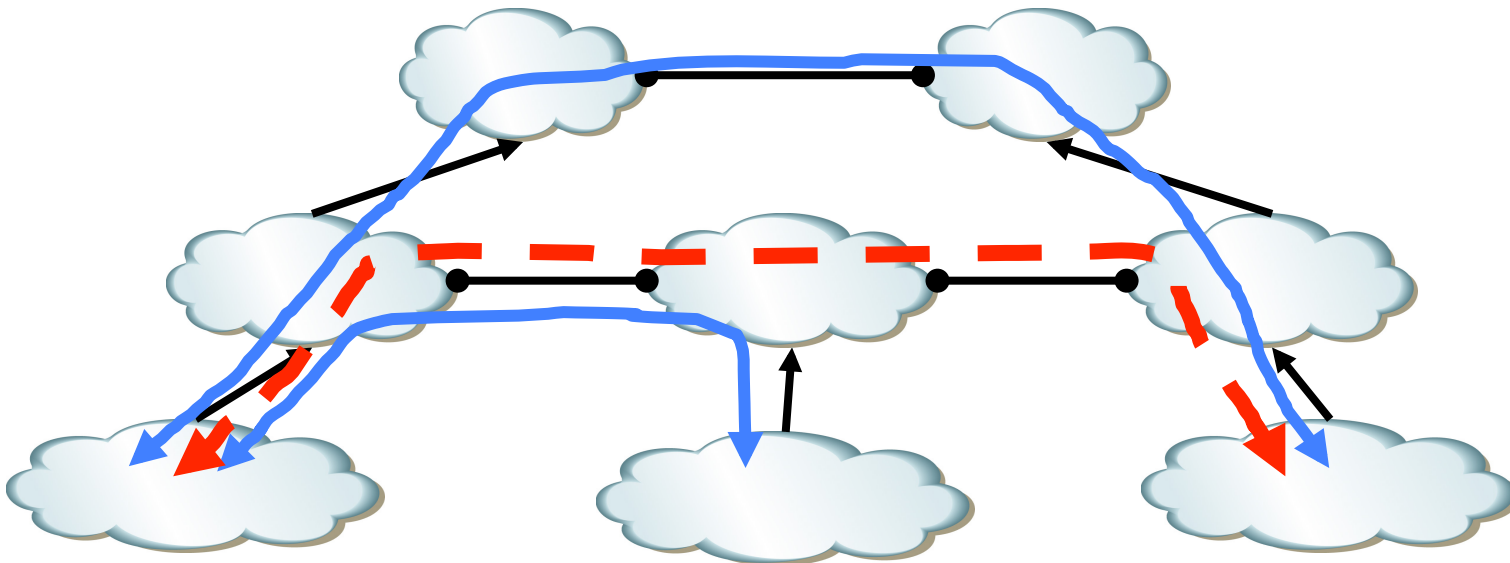
Un criterio de selección

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

BGP e Internet

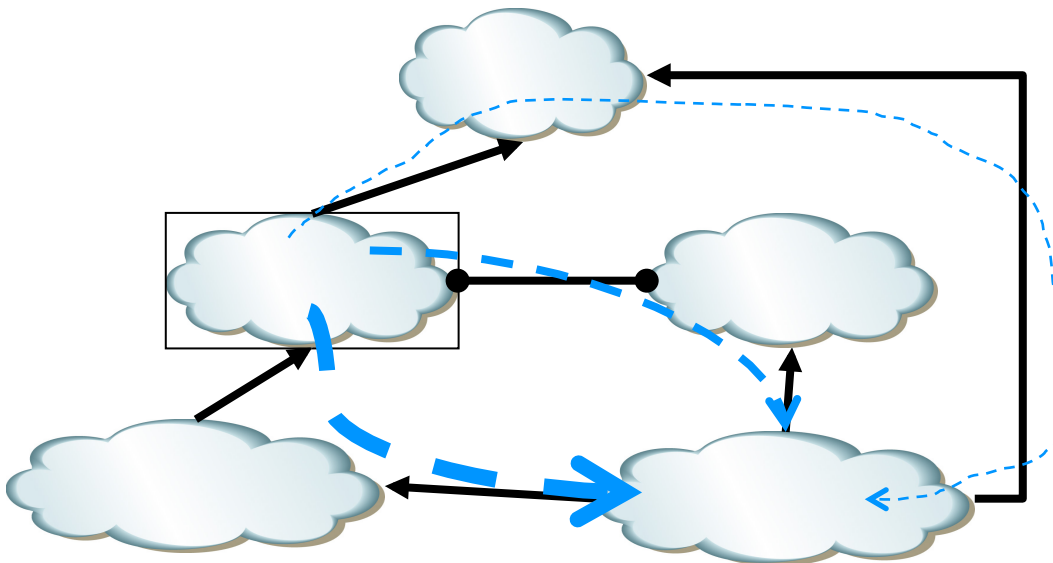
Jerarquía y economía

- 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:

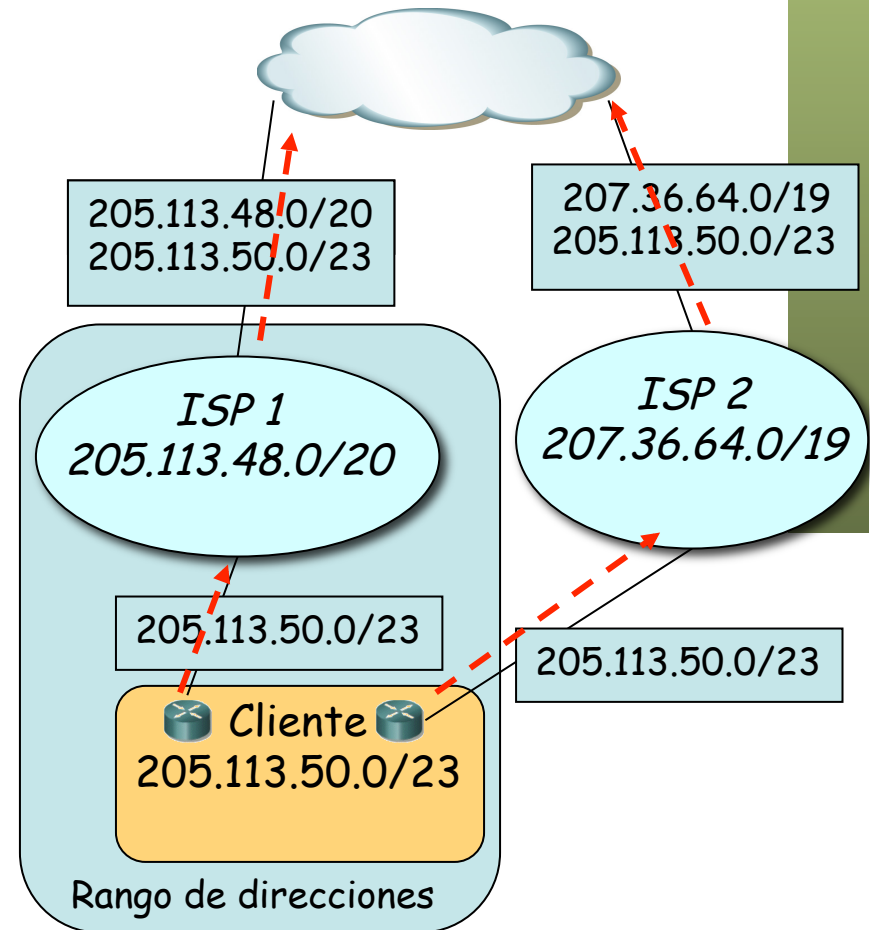
- No anunciar un destino a un vecino
- No usar caminos que pasen por cierto AS
- Ignorar el MED y usar *shortest-paths* (*hot potato routing*)
- Añadir varias veces su ASN
- Etc.

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

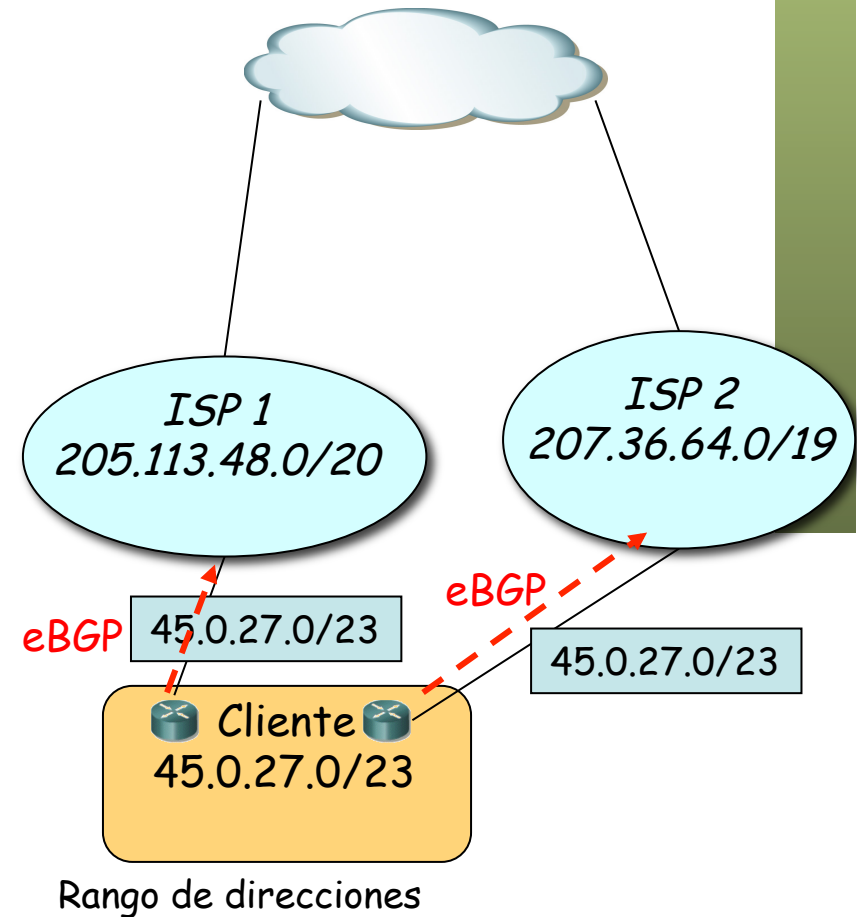
Multihoming

- 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
- (...)



Multihoming

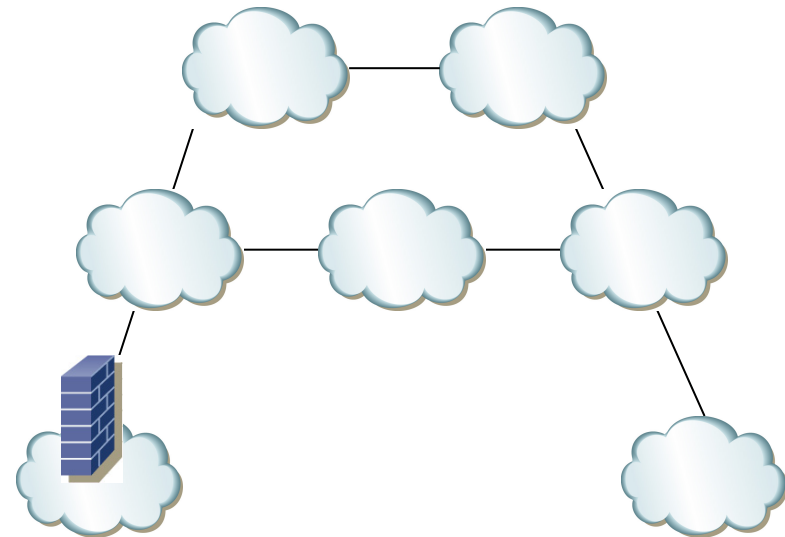
- Para ofrecer redundancia
- El rango de direcciones pertenece al ISP 1
- Habrá que anunciarlo también al ISP 2
- Ahora la ruta por ISP 2 es más específica
- *Address leaking*: ISP 1 debería anunciar también la ruta específica
- Más habitual tener un espacio de direcciones propio
- Ser un AS y correr BGP



Precauciones

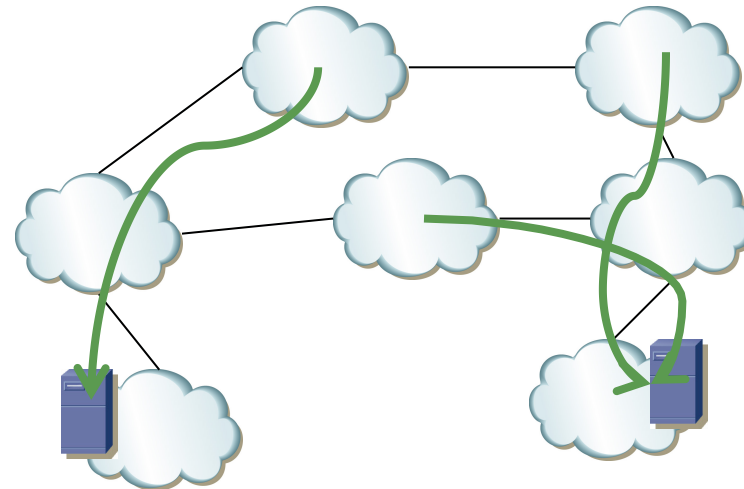
Martians

- 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



Anycast

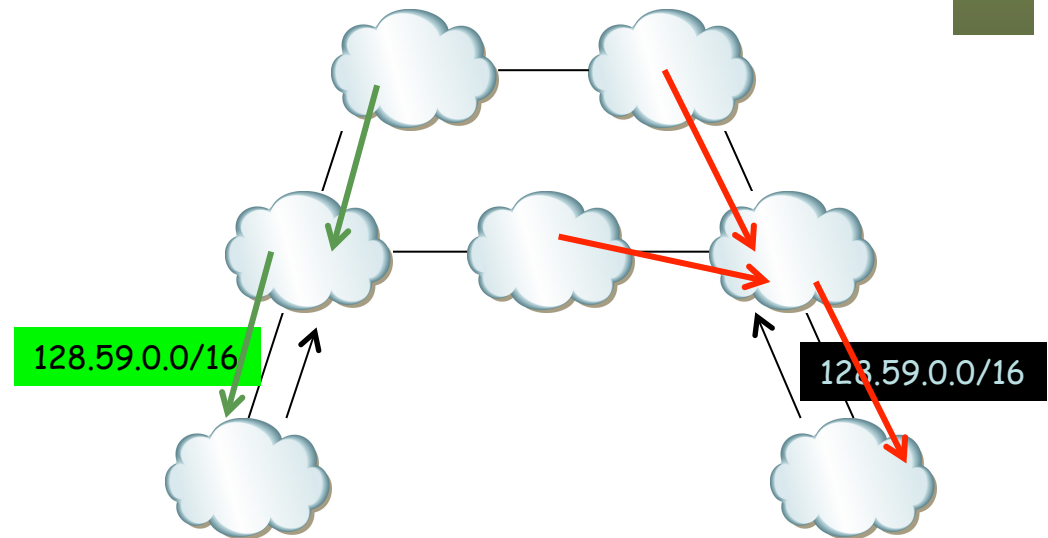
- Servidores con misma dirección IP (contenido replicado o no)
- Todos en la misma red física o en diferentes
- Anuncios por ejemplo por diferentes proveedores
- Clientes acceden a servidor según proximidad
- Permite distribución de contenidos
- También se puede hacer en el IGP
- Ejemplo: F-root name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



Precauciones

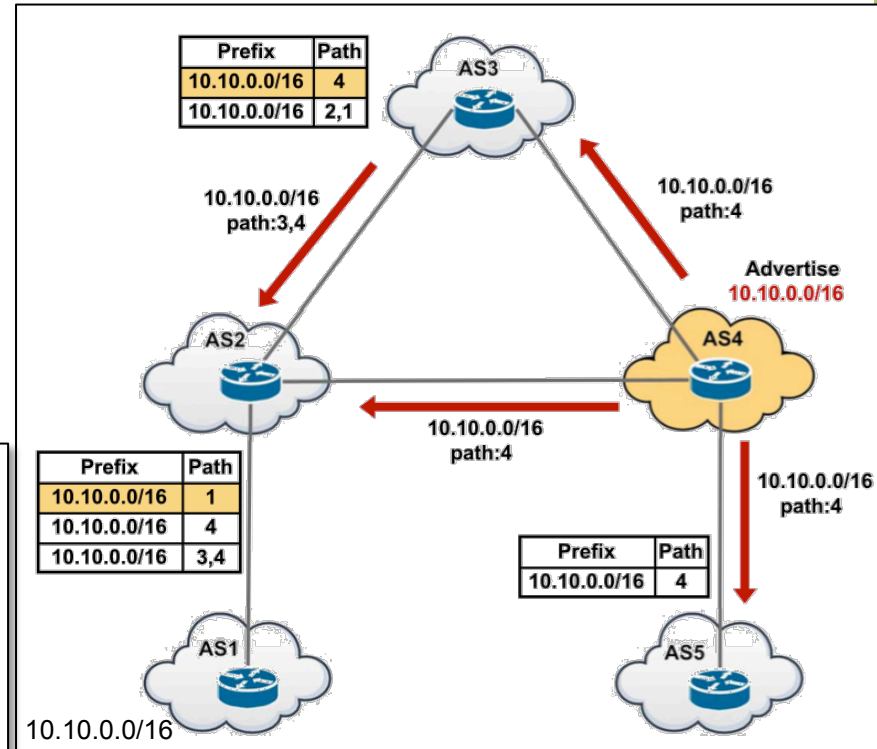
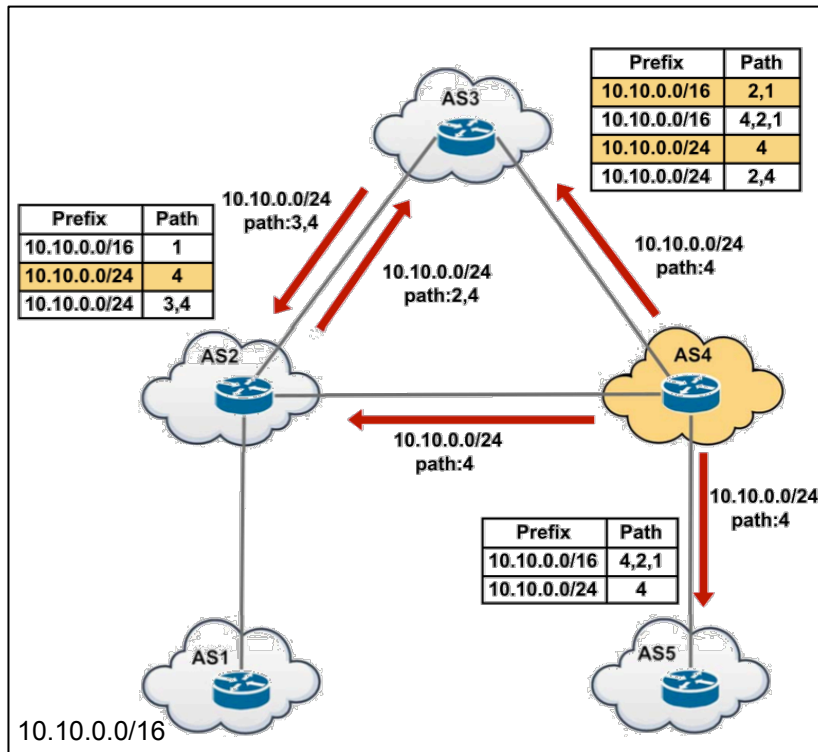
Black holes

- 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



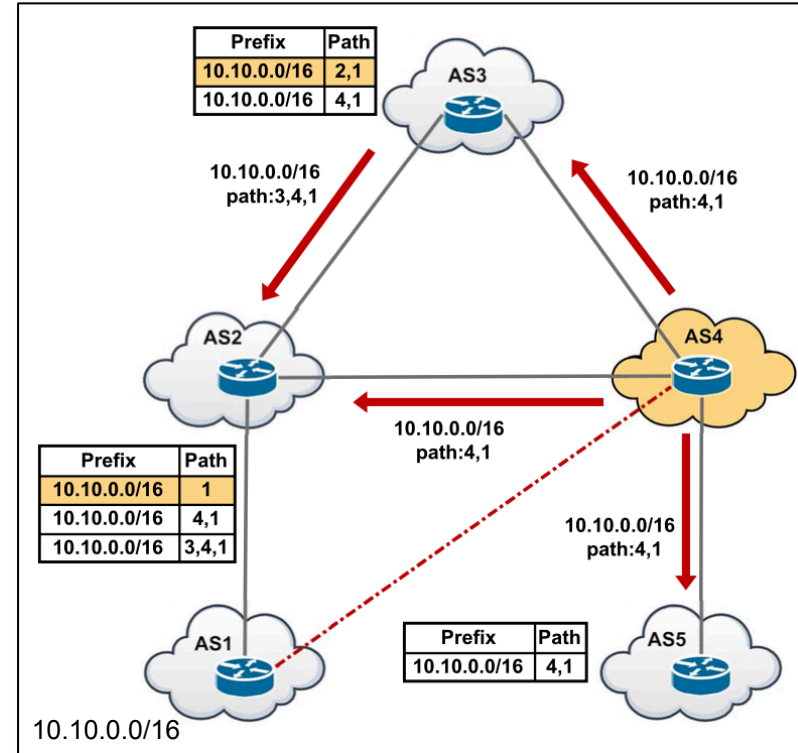
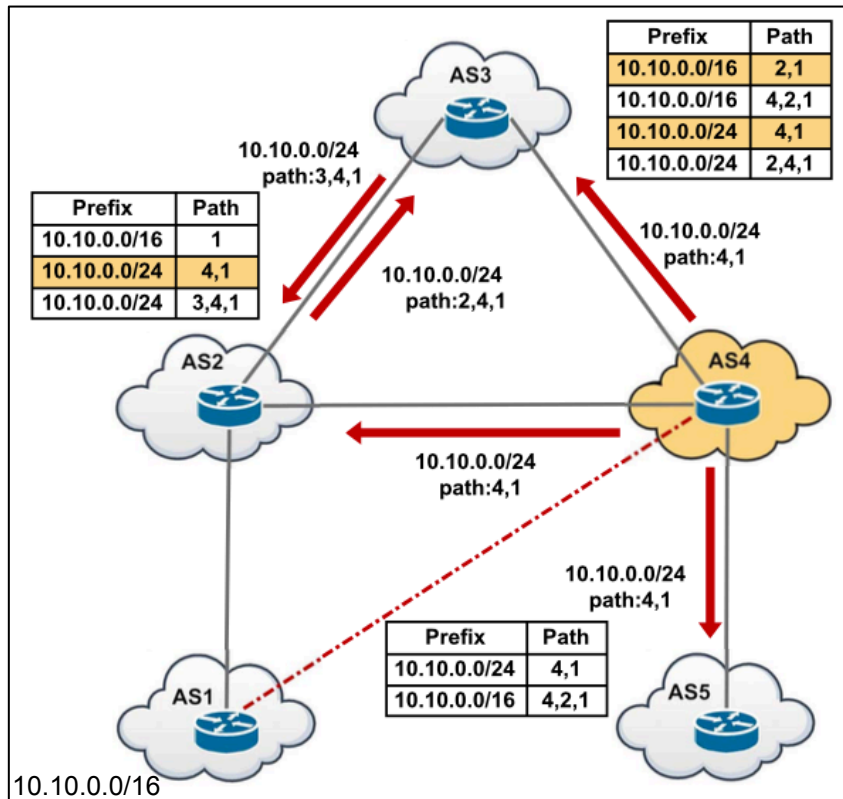
Ejemplos de problemas

- Prefix hijacking
- Sub-prefix hijacking



Ejemplos de problemas

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



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
 - (...)

