

Maestría y Especialización en Tecnologías de la Información y la Comunicación

R4. Tópicos de IPv6 y Enrutamiento Avanzado

Profesor:

Dr. Juan Carlos Fabero Jiménez (UCM)

jcfabero@ucm.es



Contenidos

- Tema 1: IP de nueva generación: IPv6
- Tema 2: Encaminamiento interno: OSPF
- Tema 3: Encaminamiento externo: BGPv4
- Tema 4: Encaminamiento troncal: MPLS
- Tema 5: Redes definidas por software: SDN
- Tema 6: Encaminamiento multicast

BGP

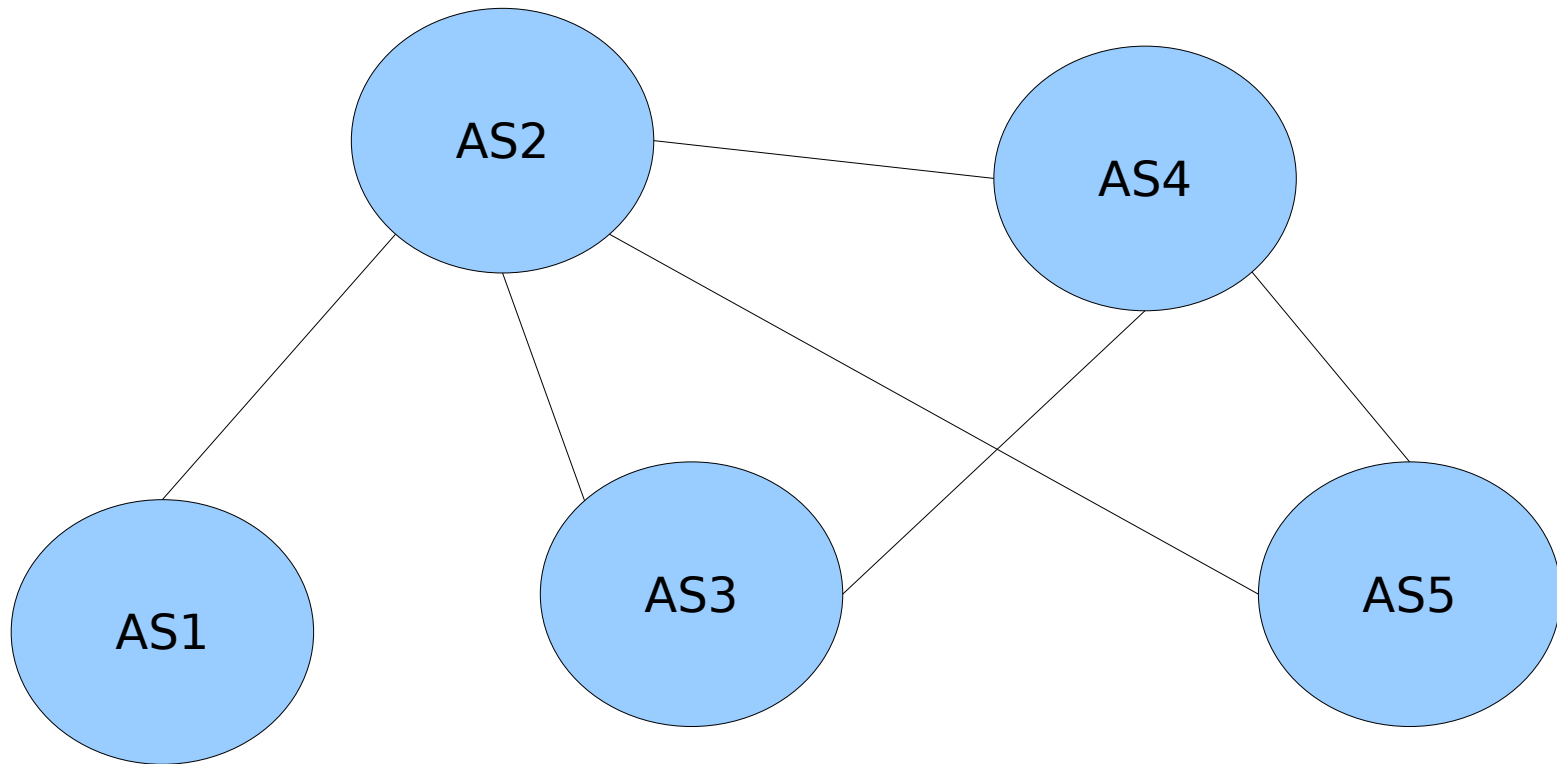


Introducción

Infraestructura de Internet

■ ¿Cómo se construye Internet?

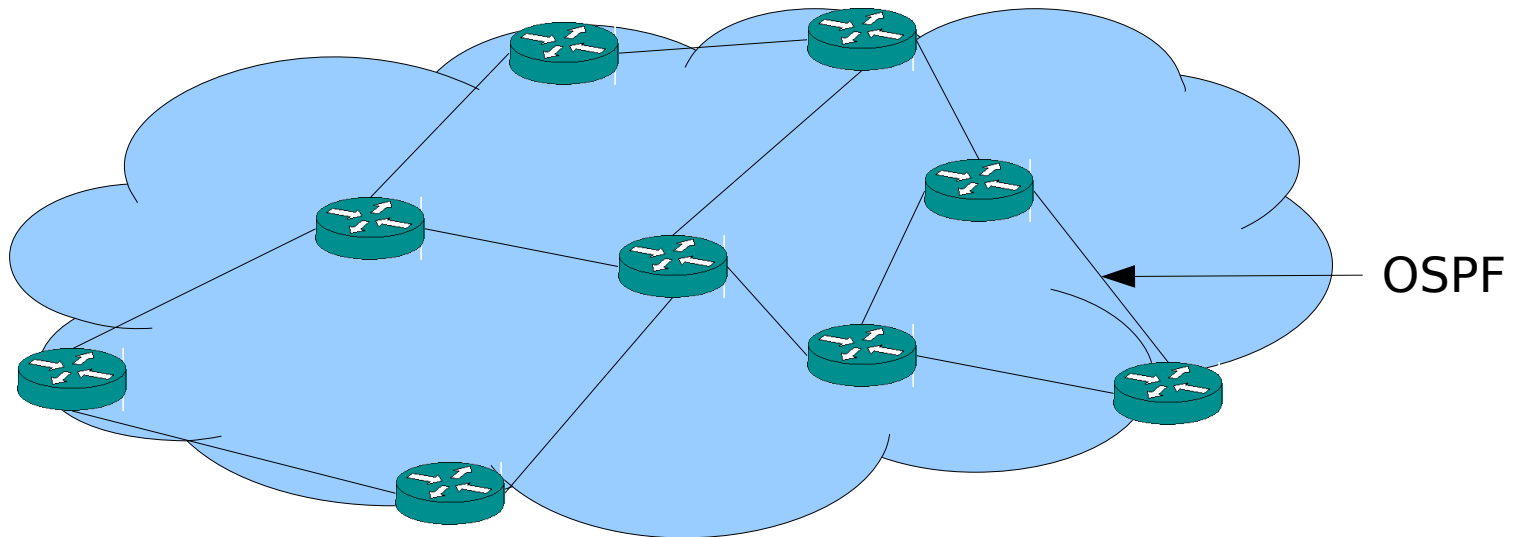
- Mediante la interconexión de Sistemas Autónomos (AS)
- Los protocolos de encaminamiento permiten intercambiar información de encaminamiento resumida.



Sistemas Autónomos

■ Definición 1:

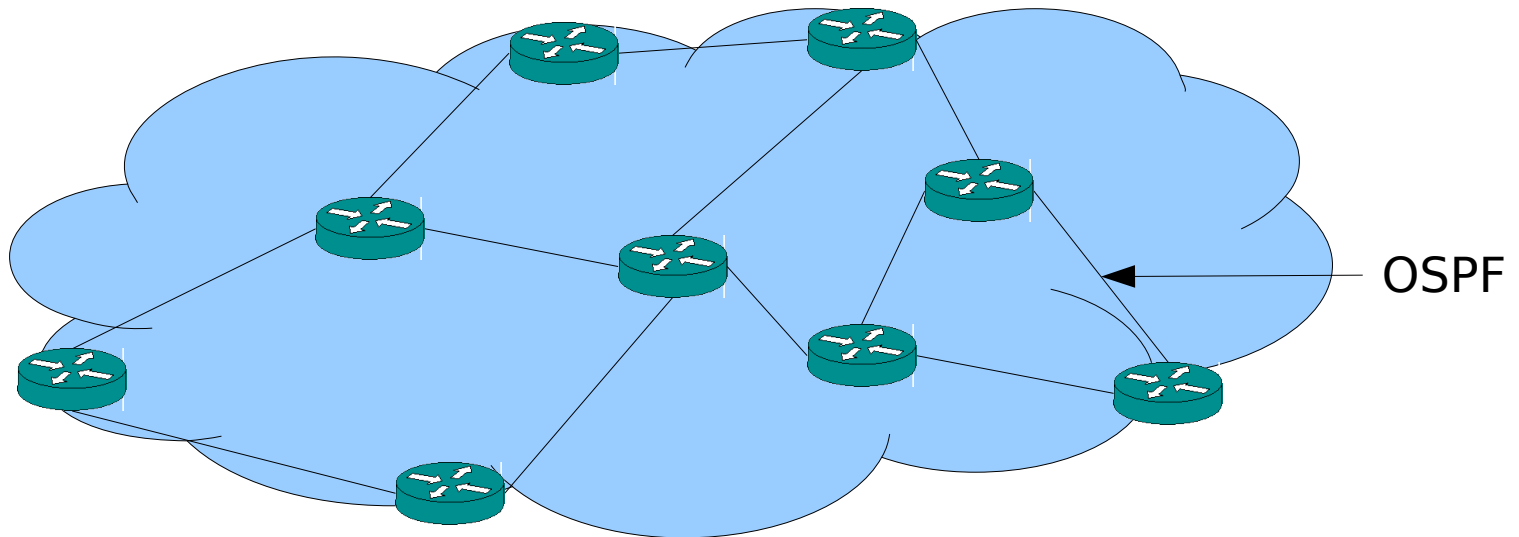
- Conjunto conexo de redes IP y encaminadores bajo el control de una o varias organizaciones y con política de encaminamiento común (RFC1930)
 - Utiliza un protocolo de pasarela interior (IGP, Internal Gateway Protocol). Generalmente OSPF o RIP.
 - Se comunica con otros AS mediante un protocolo de pasarela de frontera (BGP, Border Gateway Protocol)



Sistemas Autónomos

■ Definición 1:

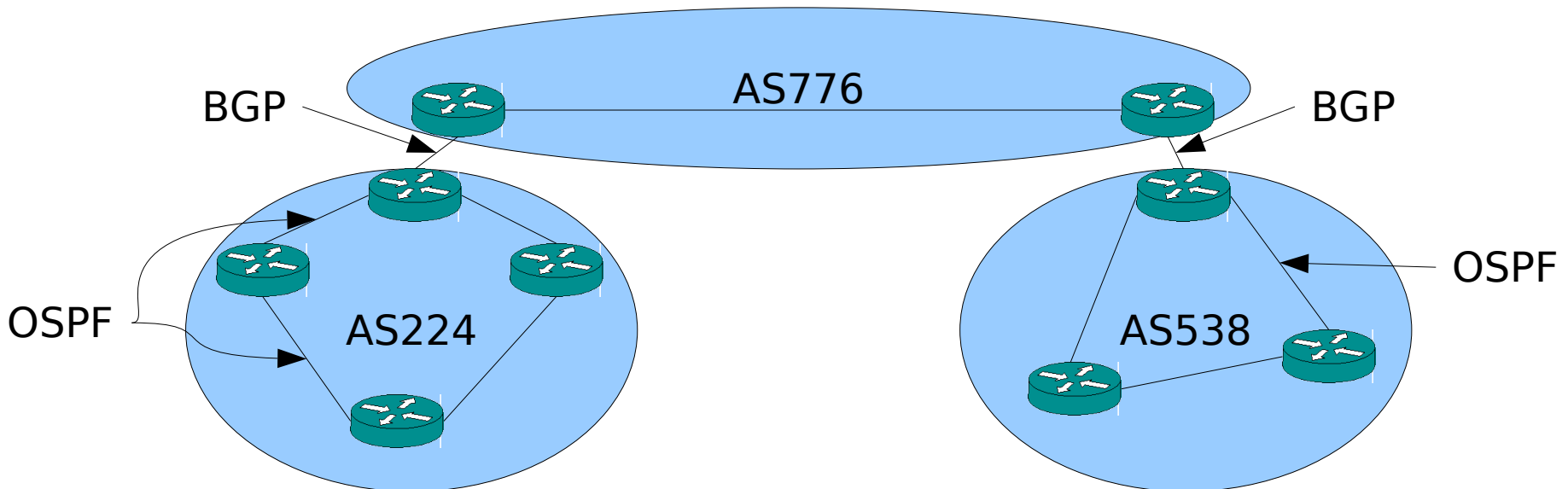
- Conjunto conexo de redes IP y encaminadores bajo el control de una o varias organizaciones y con política de encaminamiento común (RFC1930)
 - Utiliza un protocolo de pasarela interior (IGP, Internal Gateway Protocol). Generalmente OSPF o RIP.
 - Se comunica con otros AS mediante un protocolo de pasarela de frontera (BGP, Border Gateway Protocol)



Sistemas Autónomos

■ Definición 1:

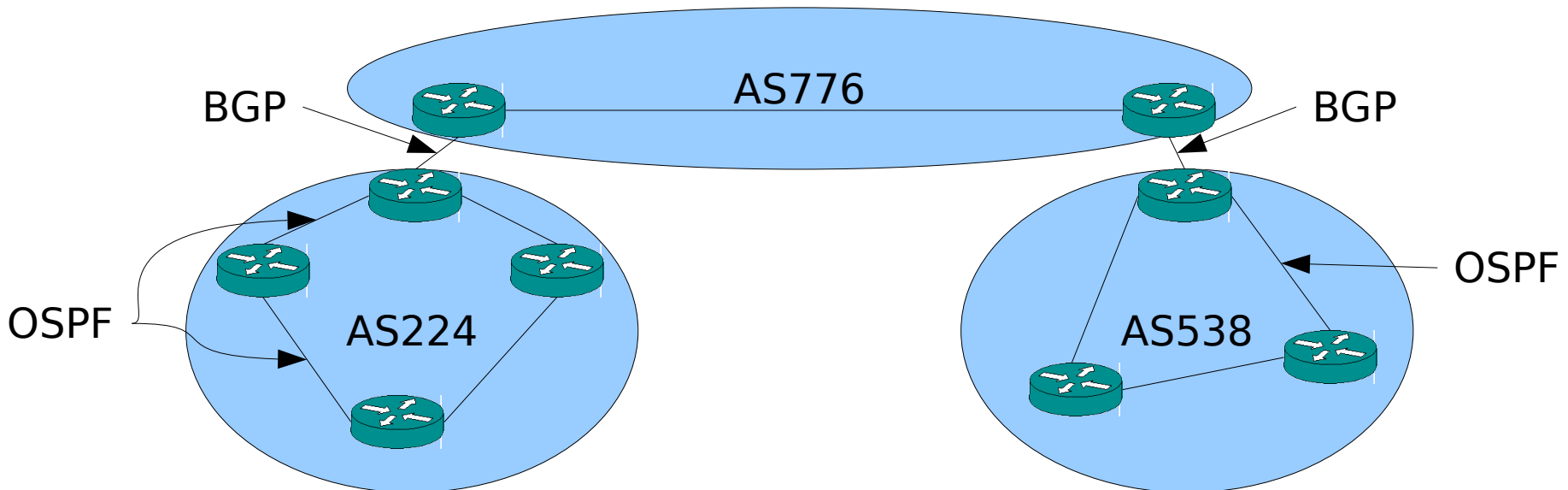
- Conjunto conexo de redes IP y encaminadores bajo el control de una o varias organizaciones y con política de encaminamiento común [RFC1930]
 - Utiliza un protocolo de pasarela interior (IGP, Internal Gateway Protocol). Generalmente OSPF o RIP.
 - Se comunica con otros AS mediante un protocolo de pasarela de frontera (BGP, Border Gateway Protocol)



Sistemas Autónomos

■ Definición 2:

- Entidad que conecta una o más redes a Internet, funciona con sus propias políticas, precios, servicios y clientes, toma decisiones independientes sobre el intercambio de tráfico, y es reconocido globalmente mediante un identificador único (ASN) [Weidong & Bingxin]



Sistemas Autónomos

■ Identificación

- Cada AS está identificado por un número único denominado ASN (Autonomous System Number)
- Los ASN están delegados por IANA (Internet Assigned Number Authority) a los RIR (Regional Internet Registries) por bloques.
- Cada RIR asigna un ASN a cada organización.
- Hasta 2007 los ASN eran un número de 16 bits. Ahora se ofrecen de 32 bits, aunque no todos los sistemas son compatibles con la nueva numeración.
- Los ASN del 65512 al 65534 están reservados para uso privado y no pueden anunciarse en Internet.

RIR (Regional Internet Registries)

- **Existen 5 Registros Regionales de Internet:**
 - **AfriNIC:** Africa (8 de abril de 2005)
 - **APNIC:** Asia y Pacífico
 - **ARIN:** Norte América
 - **LACNIC:** Latino América y Caribe
 - **RIPE NCC:** Europa, Oriente Medio y Asia Central



Sistemas Autónomos

■ Definiciones

- Vecinos (Neighbours)
 - AS que intercambian información de encaminamiento.
 - Encaminadores que intercambian información de encaminamiento.
- Anuncio (Announce)
 - Envío de información de encaminamiento a un vecino.
- Aceptación (Accept)
 - Uso de la información recibida desde un vecino.
- Originar (Originate)
 - Insertar información de encaminamiento en un anuncio (generalmente de manera explícita o como resultado de un IGP).
- Parejas (Peers)
 - Encaminadores en AS vecinos o dentro de un AS que intercambian información de encaminamiento y políticas.

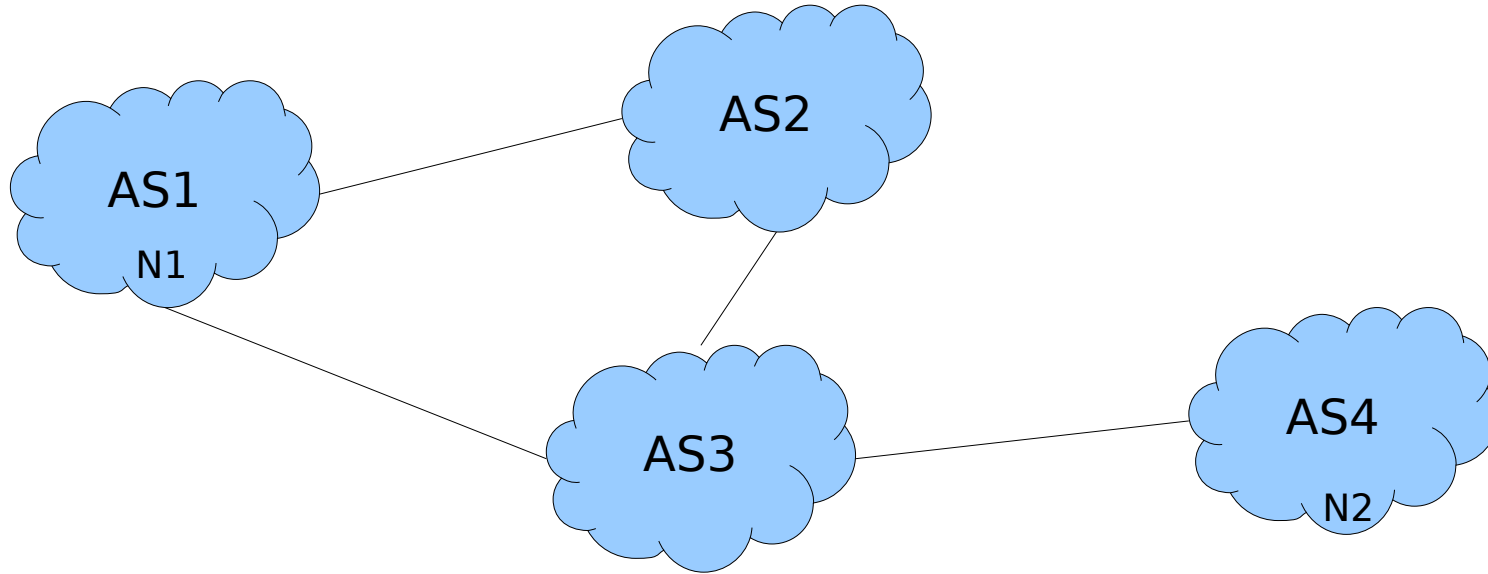
Sistemas Autónomos



■ Rutas y flujo de datos entre Sistemas Autónomos

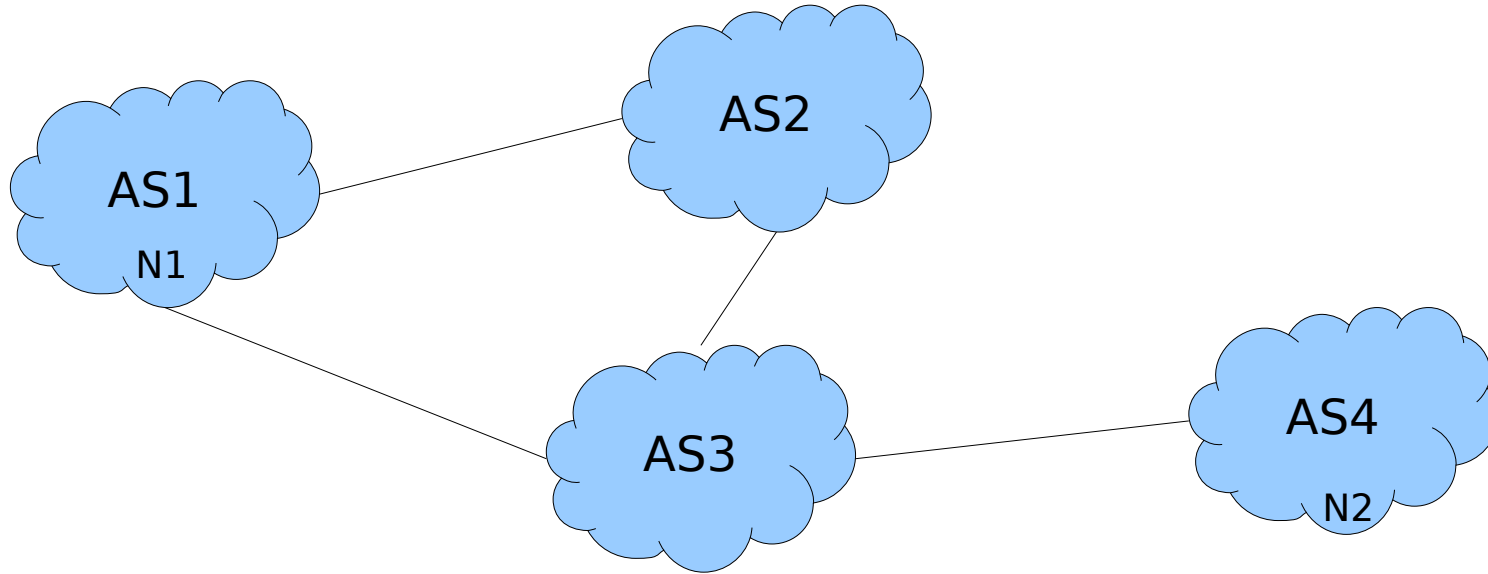
- AS1299 debe anunciar rutas a AS766
- AS766 debe aceptar las rutas de AS1299
- AS766 debe anunciar rutas a AS1299
- AS1299 debe aceptar las rutas de AS766
- **El flujo de datos es en sentido contrario a los anuncios de rutas**

Sistemas Autónomos



- **Rutas y flujo de datos entre varios Sistemas Autónomos**
 - **Para que N1 pueda enviar datos a N2**
 - **AS4 debe originar y anunciar N2 a AS3**
 - **AS3 debe aceptar el anuncio de AS4**
 - **AS3 debe anunciar N2 a AS2 y/o AS1**
 - **AS2 debe aceptar el anuncio de AS3**
 - **AS2 debe anunciar N2 a AS1**
 - **AS1 debe aceptar el anuncio de AS2 y/o AS3**

Sistemas Autónomos



Cuando existen varios caminos posibles,
la decisión de encaminamiento no es trivial

Sistemas Autónomos

■ Rutas entre Sistemas Autónomos

- Se puede hacer una traza de los AS atravesados en una conexión en:
 - <http://www.rediris.es/red/lg>
 - http://logbud.com/visual_trace (desactivado)
- Ejemplo:
 - Desde UCM a UNA:
 - 766 (REDIRIS)
 - 1299 (TELIANET)
 - 6762 (SEABONE)
 - 7303 (TELECOM AR)
 - 27768 (COPACO PY)
 - 27733 (CNC PY)

Tipos de Sistemas Autónomos

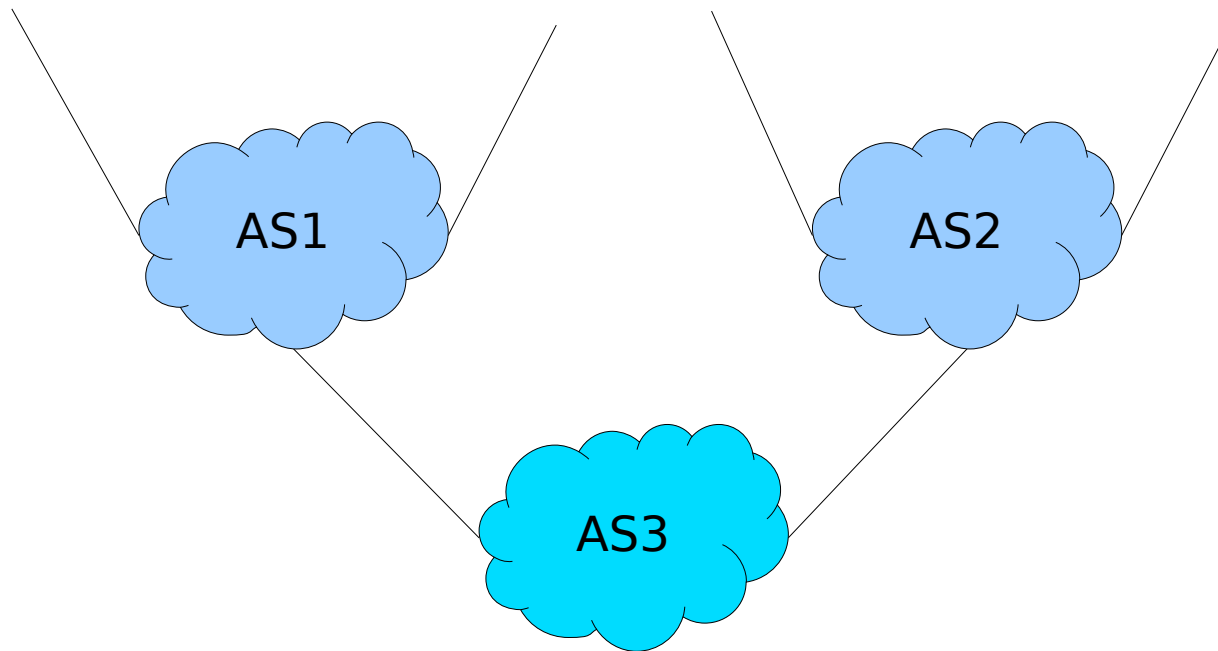
■ Los Sistemas Autónomos pueden ser:

- Multihomed: está conectado a más de un proveedor.
 - Enlaces redundantes.
- Stub: es un nodo final (no es proveedor para ningún otro AS).
 - Sistemas autónomos clientes.
- Transit: es un proveedor para otros AS.
 - Permite tráfico que no ha sido originado ni va dirigido al propio sistema autónomo.
 - Establece políticas que restringen el tráfico de tránsito.

Tipos de Sistemas Autónomos

■ Multihomed

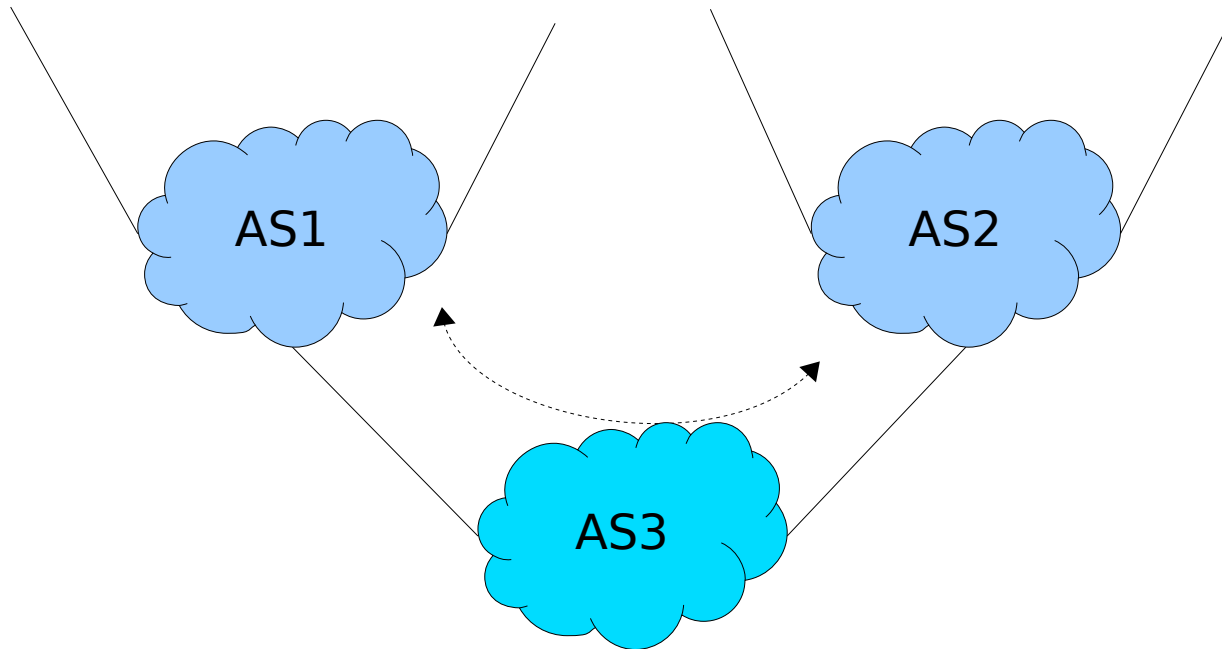
- Está conectado a más de un AS, de manera que puede permanecer conectado a Internet en el caso de fallo de una de las conexiones.
- No permite tráfico de tránsito.



Tipos de Sistemas Autónomos

■ Transit

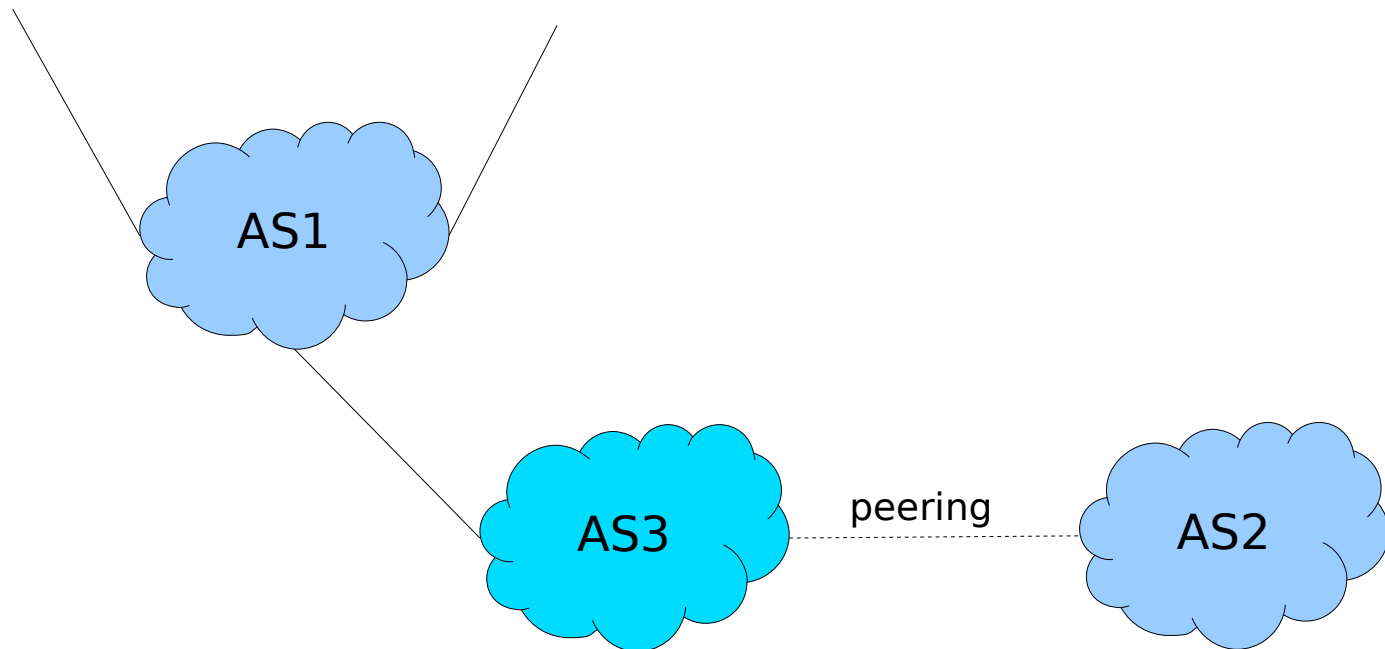
- Proporciona conexión entre distintos AS.
- Un ISP (Internet Service Provider) es siempre un AS de tránsito.
- Las troncales de Internet son AS de tránsito.



Tipos de Sistemas Autónomos

■ Stub

- Sólo está conectado a otro AS.
- Puede haber establecido una relación con otro AS que no se refleja en los servidores públicos de rutas (“looking glass servers”)
 - Por ejemplo: en entornos financieros y del sector transportes.

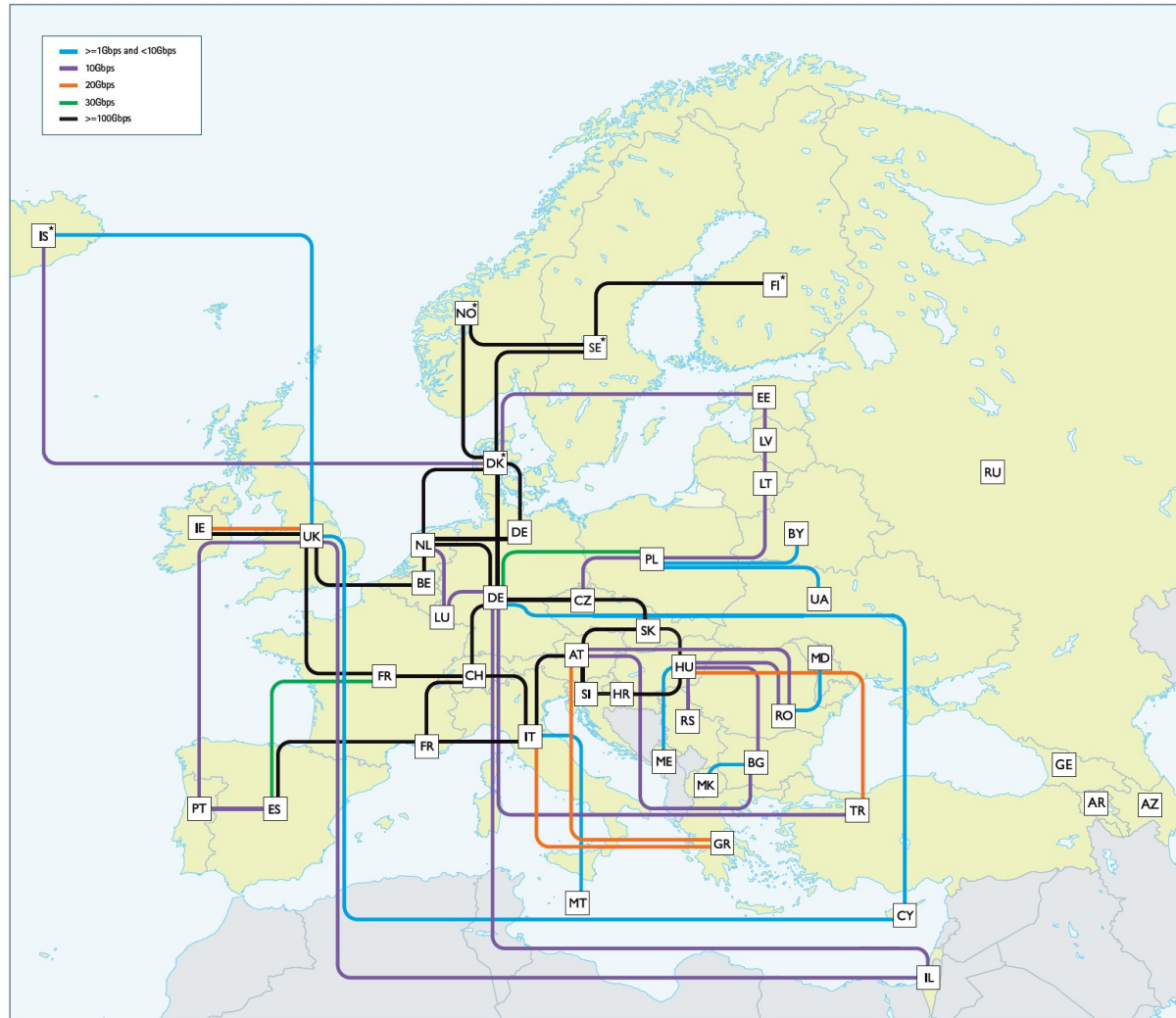


BGP

Troncales y la estructura de Internet

Troncales

■ Europa: GEANT2 (<http://www.geant.net>)



GEANT connectivity as at January 2014. GEANT is operated by DANTE on behalf of Europe's NRENs.

Troncales

■ Iberoamérica: CLARA / Alice (<http://www.redclara.net>)



Troncales

■ Global (<http://www.geant.net>)



Troncales

■ España: RedIris (<http://www.rediris.es>)



Whois

■ Definición y uso:

- La base de datos *whois* mantiene información sobre las redes (destinos) que existen en Internet, sus propietarios, etc.
- Está mantenida por los diversos RIR y se puede consultar mediante la orden `whois`.

■ Confusión:

- La base de datos de Internic (o sus delegados) que mantiene información sobre nombres de dominio también es accesible mediante `whois`.

■ Consultas:

- Se puede consultar sobre:
 - Una dirección IP
 - Un ASN
 - Una persona
 - Etc.

Whois

■ Ejemplos:

- Consulta sobre una dirección IP.

\$ whois 200.10.228.135

% Copyright LACNIC lacnic.net

% 2010-06-01 06:48:40 (BRT -03:00)

```
inetnum: 200.10.228/22
status: assigned
owner: Centro Nacional de Computacion
ownerid: PY-CNCO-LACNIC
responsible: Gustavo Amarilla
address: Campus Universitario - San Lorenzo, 1,
address: 1439 - San Lorenzo - ce
country: PY
phone: +595 21 585550 []
owner-c: GUA4
tech-c: GUA4
abuse-c: GUA4
inetrev: 200.10.228/22
nserver: NS.CNC.UNA.PY
nsstat: 20100530 AA
nslastaa: 20100530
nserver: SCE.CNC.UNA.PY
nsstat: 20100530 AA
created: 19940610
changed: 19960419
```

Whois

■ Ejemplos:

■ Consulta sobre un ASN

\$ **whois AS27733**

aut-num: AS27733
owner: Centro Nacional de Computacion
ownerid: PY-CNCO-LACNIC
responsible: Gustavo Amarilla
address: Campus Universitario - San Lorenzo, 1,
address: 1439 - San Lorenzo - ce
country: PY
phone: +595 21 585550 []
owner-c: GUA4
routing-c: GUA4
abuse-c: GUA4
created: 20040823
changed: 20050810

nic-hdl: GUA4
person: gustavo amarilla
e-mail: gamarila@CNC.UNA.PY
address: san lorenzo, 1432,
address: 1432 - san lorenzo - dc
country: PY
phone: +000 00 585550 []
created: 20031224
changed: 20031224

Whois

■ Ejemplos:

■ Consulta sobre un identificador

\$ **whois PY-CNCO-LACNIC**

owner: Centro Nacional de Computacion
ownerid: PY-CNCO-LACNIC
responsable: Gustavo Amarilla
address: Campus Universitario - San Lorenzo, 1,
address: 1439 - San Lorenzo - ce
country: PY
phone: +595 21 585550 []
owner-c: GUA4
created: 20040517
changed: 20050810

nic-hdl: GUA4
person: gustavo amarilla
e-mail: gamarila@CNC.UNA.PY
address: san lorenzo, 1432,
address: 1432 - san lorenzo - dc
country: PY
phone: +000 00 585550 []
created: 20031224
changed: 20031224

aut-num: 27733
inetnum: 200.10.228/22
inetnum: 201.217.14/24
inetnum: 2001:1320::/32

Whois

■ Ejemplos:

- Consulta sobre una persona.

\$ **whois MHU3-RIPE**

% This is the RIPE Whois query server #3.

```
person:      Manuel Hernandez Urrea
address:     Centro de Proceso de Datos de la UCM
address:     Avda. Complutense, s/n
address:     Ciudad Universitaria
address:     E-28040 Madrid
address:     SPAIN
phone:       +34 913944686
fax-no:      +34 913944773
nic-hdl:     MHU3-RIPE
abuse-mailbox: abuse@rediris.es
abuse-mailbox: abuse@rediris.es
mnt-by:      REDIRIS-NMC
source:      RIPE # Filtered
```

Whois

■ Interconexión de AS

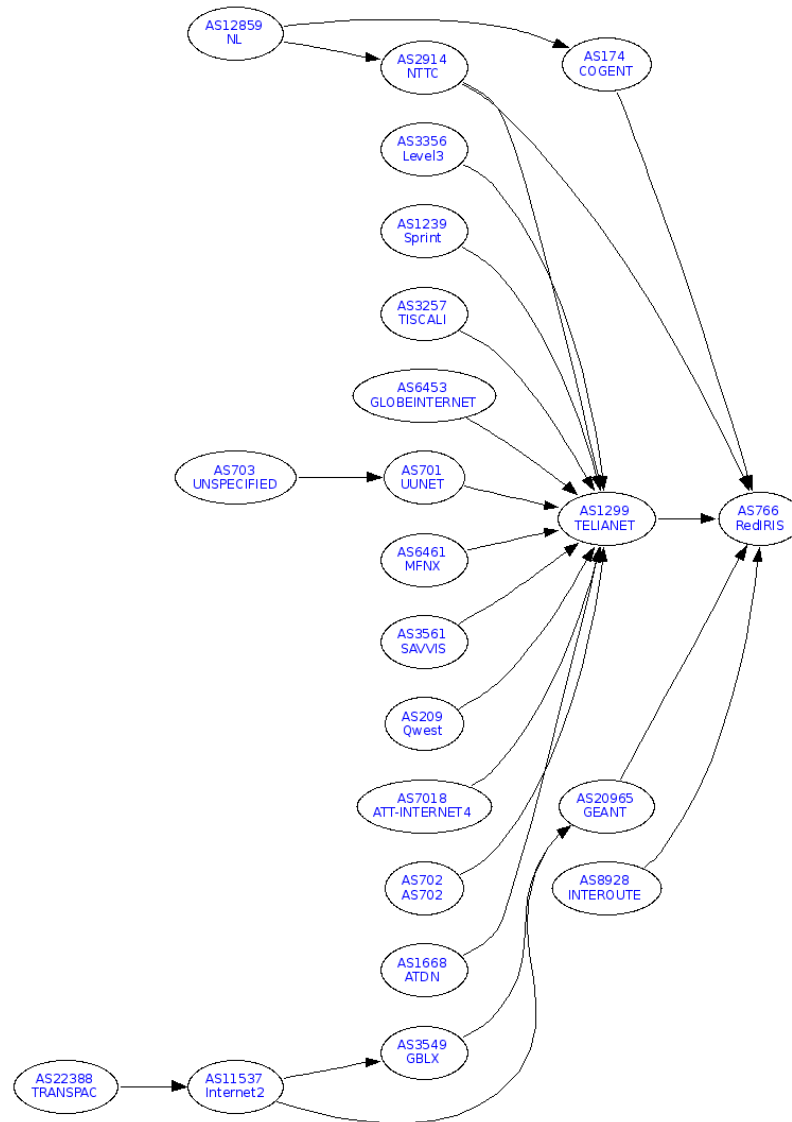
- Una consulta whois puede darnos mucha información sobre el sistema autónomo. Por ejemplo:

\$ whois as766

```
import:      from AS288 action pref=100; accept { 192.171.2.0/24,
131.176.160.0/24, 131.176.161.0/24, 131.176.162.0/24, 131.176.163.0/24,
131.176.164.0/24, 131.176.165.0/24, 131.176.166.0/24, 131.176.167.0/24,
131.176.168.0/24, 131.176.169.0/24, 131.176.170.0/24, 131.176.171.0/24 }
import:      from AS12457 action pref=100; accept AS-TELIBERIA
import:      from AS12715 action pref=100; accept AS-JAZZTRANS
import:      from AS13041 action pref=100; accept AS-ANELLA
import:      from AS13237 action pref=100; accept AS-LNCESPANIX
import:      from AS15488 action pref=100; accept AS15488
import:      from AS15630 action pref=100; accept AS15630
import:      from AS15915 action pref=100; accept AS-IBERCOM
export:      to AS288 announce ANY
export:      to AS1299 announce AS-REDIRIS {192.243.16.0/22,
192.171.2.0/24, 131.176.160.0/24, 131.176.161.0/24, 131.176.162.0/24,
131.176.163.0/24, 131.176.164.0/24, 131.176.165.0/24, 131.176.166.0/24,
131.176.167.0/24, 131.176.168.0/24, 131.176.169.0/24, 131.176.170.0/24,
131.176.171.0/24 }
```

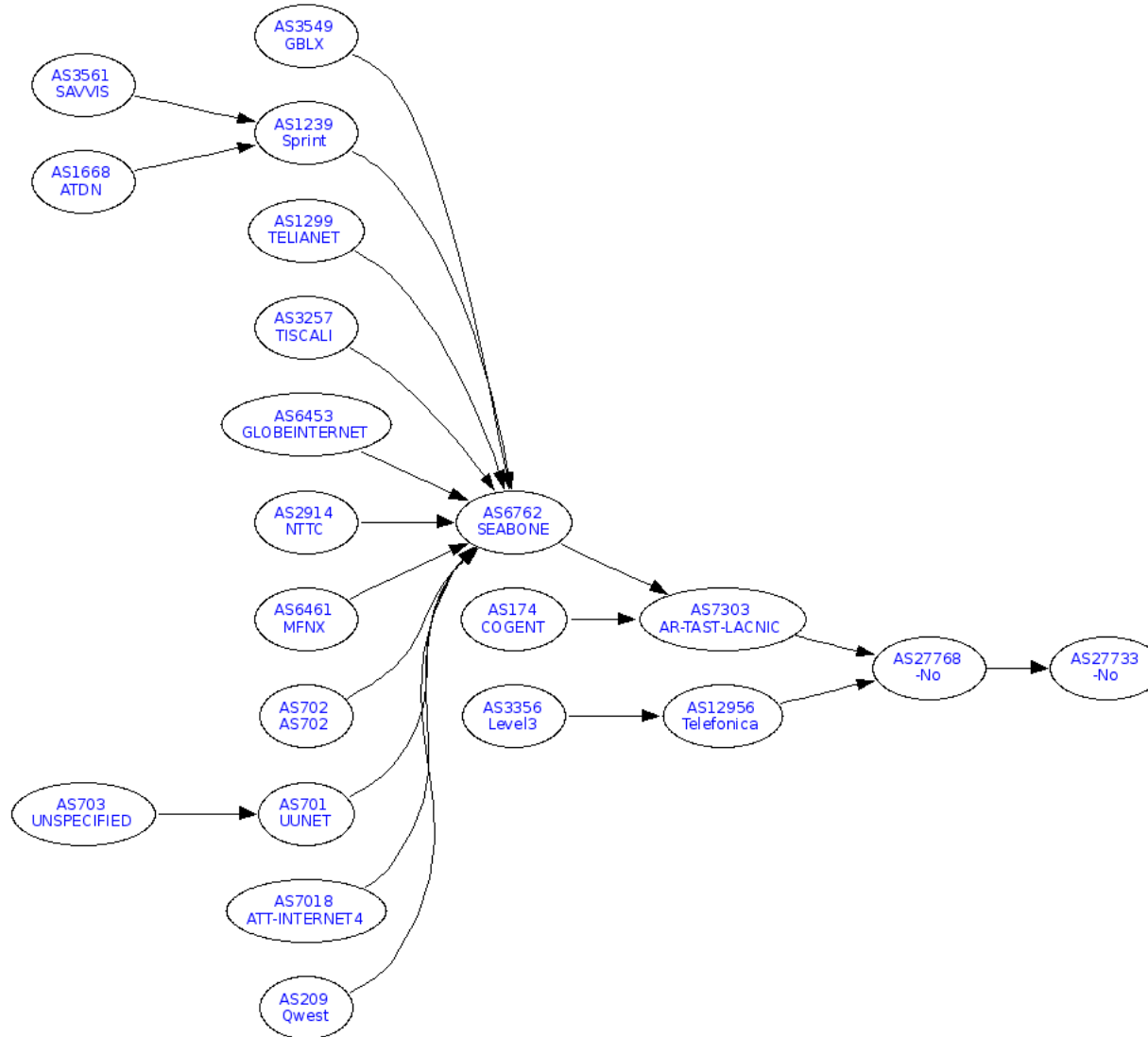
Whois

■ AS766



Whois

■ AS27733



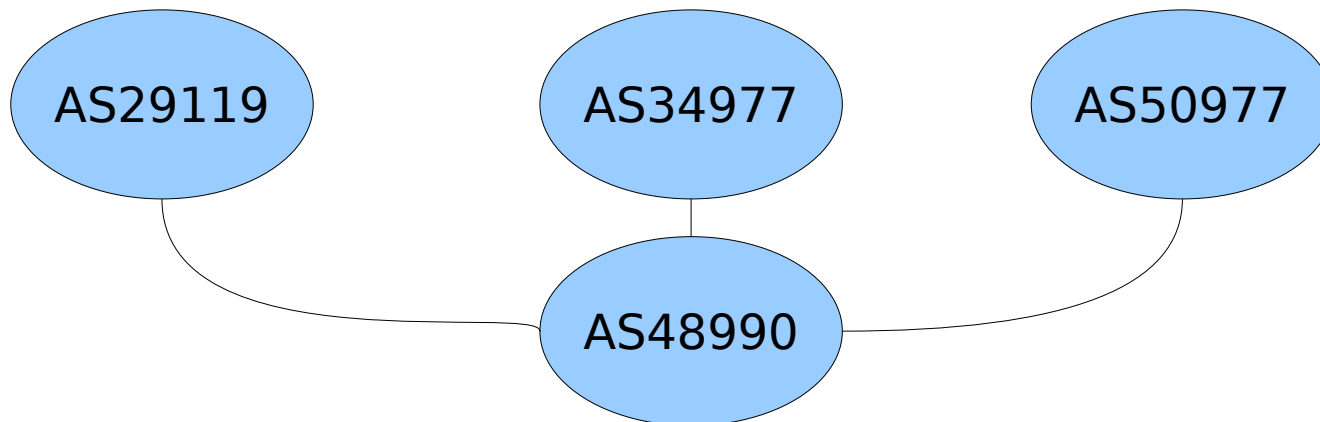
Whois

■ AS48990

import: from AS29119 action pref=100; accept ANY
import: from AS34977 action pref=100; accept ANY
import: from AS50977 accept ANY
export: to AS29119 announce AS48990
export: to AS34977 announce AS48990
export: to AS50977 announce AS48990

■ AS29119

import: from AS48990 action pref=300; accept AS48990
export: to AS48990 announce ANY



BGP

El encaminamiento inter-AS

Encaminamiento entre AS

■ El problema de las clases

- IPv4 sólo permite un encaminamiento jerárquico en 2 niveles: netid y hostid.
- Hay que mantener una entrada en la tabla de encaminamiento para cada destino (red) posible en Internet.
- 128 redes de clase A
- 60000 (aprox.) redes de clase B
- $>2 \cdot 10^6$ redes de clase C

¡Hay que mantener una tabla con más de 2 millones de entradas!

Encaminamiento entre AS

- **La solución: CIDR (Classless Interdomain Routing)**
 - Las redes de Clase C que no estaban asignadas se han agrupado en bloques (superredes) con máscaras de redes menores de /24.
 - Los bloques se han delegado a los cinco RIR.
 - Una sola entrada en la tabla de encaminamiento agrupa miles de redes.
 - Ejemplo:
 - Asignados a LACNIC
 - 200/8
 - 201/8
 - Asignados a RIPE
 - 193/8
 - 194/8
 - 195/8
 - Cada prefijo /8 agrupa 65535 redes de clase C.

Encaminamiento entre AS

■ Políticas de encaminamiento

- Para encaminar entre sistemas autónomos hay que considerar múltiples factores.
 - Las organizaciones que gestionan cada AS deben establecer acuerdos particulares.
 - Factores económicos.
 - Conectividad física.
 - Consideraciones geopolíticas.
- La política de encaminamiento por siguiente salto puede no ser adecuada.
 - Es posible que interese evitar atravesar determinados sistemas autónomos.
- Cuando se recibe una ruta hacia un destino, han de conocerse los AS que se atraviesan para alcanzarla.

BGP

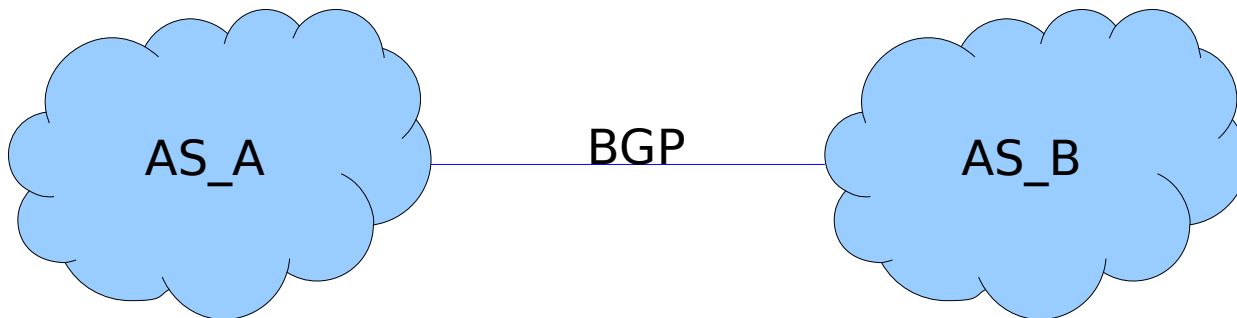


El protocolo BGP

BGP (Border Gateway Protocol)

■ BGP

- Es el EGP (Exterior Gateway Protocol) más utilizado.
- Descrito en los RFC 4271 (versión 4) y 4760 (versión 4+)
- Permite intercambiar información de encaminamiento entre sistemas autónomos.



■ Tipos de mensajes:

- Establecimiento (**OPEN**): establece relación con otro encaminador frontera de AS.
- Actualización (**UPDATE**): transmite información de encaminamiento.
- Encaminador activo (**KEEPALIVE**): confirma la recepción de un mensaje de establecimiento y también periódicamente la relación con otro encaminador frontera de AS.
- Notificación (**NOTIFICATION**): informa de la detección de una condición de error y termina la conexión.
- Los mensajes se intercambian mediante TCP (puerto 179)

■ Mensaje OPEN

- Establece la sesión BGP.
- Contiene:
 - El ASN del encaminador originante y su identificador BGP.
 - Una propuesta de valor para el *hold timer*.
 - Parámetros opcionales:
 - Autenticación del vecino.

■ Mensaje KEEPALIVE

- Confirma un mensaje OPEN.
- Mantiene la sesión abierta.
 - Si no se recibe un KEEPALIVE durante el tiempo de *hold timer*, se supone que ha ocurrido un error y se cierra la sesión.
 - RFC 4271 recomienda valores de 90s para el *hold timer* y 30s para el intervalo de keepalive.

■ Mensaje NOTIFICATION

- Se envía cuando se ha detectado una condición de error.
 - Por ejemplo, no se han recibido *keepalives* durante el periodo *hold timer*.
- Una vez enviado, se cierra la sesión BGP. Todas las rutas asociadas con ese vecino se marcan como no válidas.
- El mensaje incluye un código de error y un subcódigo.
 - Error de cabecera de mensaje.
 - Error de mensaje OPEN.
 - Error de mensaje UPDATE.
 - *Hold timer* superado.
 - Error en la máquina de estados finitos.
 - Cierre administrativo.

■ Mensaje UPDATE

- Transporta información de encaminamiento (redes alcanzables).
- Tras el mensaje OPEN, se envía la tabla completa al vecino.
- Luego sólo se envían las actualizaciones.
 - Nueva ruta.
 - Rutas eliminadas.

Common Header (Type = 2)	19
Longitud de rutas inalcanzables	2
Rutas inalcanzables	variable
Longitud total de atributos de ruta	2
Atributos de ruta	variable
Información de alcanzabilidad de red	variable

■ Anuncios

- Un mensaje UPDATE **anuncia** información de encaminamiento hacia una ruta o rutas concretas, o enumera un conjunto de rutas que deben retirarse.
- Un **anuncio** consta de un **prefijo** (notación CIDR) y uno o más **atributos**.

■ Selección de ruta

- BGP selecciona la ruta en función de los dominios (o confederaciones) que deben atravesarse para alcanzar el destino.
- La mejor ruta es la que atraviesa el menor número de dominios, pero...
- ¡Hay otros factores que tener en cuenta!
 - Costes económicos.
 - Acuerdos comerciales.
 - Política internacional...

■ Atributos de ruta

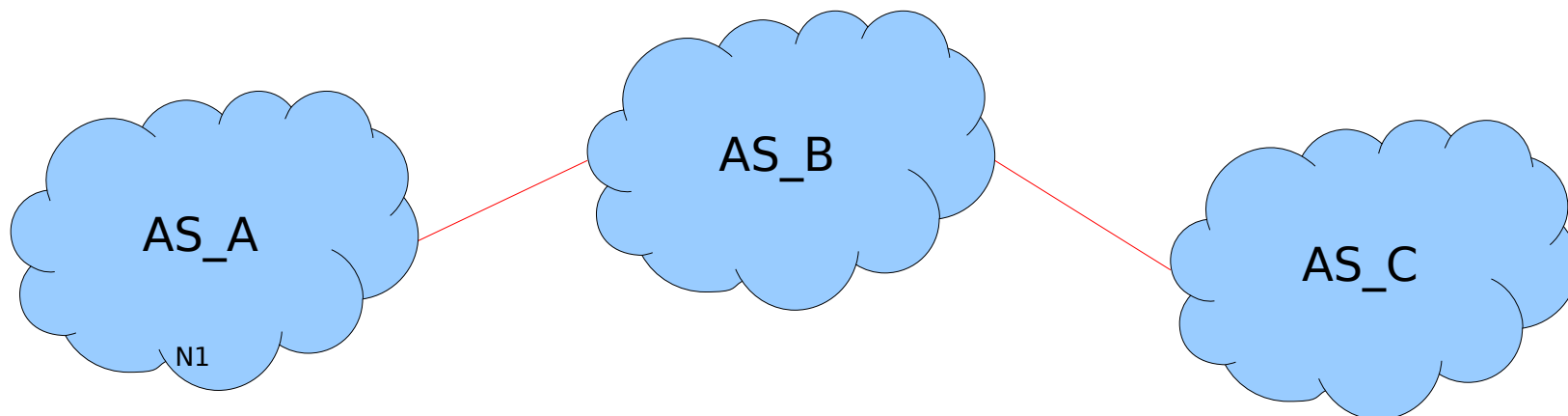
- Se utilizan para dirigir el proceso de selección de ruta, según determinadas políticas.
- Diversos tipos de atributos:
 - Bien conocidos y obligatorios.
 - El atributo debe ser reconocido por todas las implementaciones BGP y estar presente en cada mensaje UPDATE.
 - Bien conocidos y no obligatorios.
 - El atributo debe ser reconocido por todas las implementaciones BGP pero su presencia en los mensajes UPDATE es opcional.
 - Opcionales transitivos.
 - No se requiere que sean reconocidos por todas las implementaciones BGP, pero se debe reenviar a otros vecinos.
 - Opcionales no transitivos.
 - No se requiere que sean reconocidos por todas las implementaciones BGP, y pueden ignorarse y no reenviarse a otros vecinos.

■ Atributos más comunes

- BGP define 7 atributos:
 - ORIGIN
 - AS_PATH
 - NEXT_HOP
 - LOCAL_PREF
 - MULTI_EXIT_DISC
 - ATOMIC_AGGREGATE
 - AGGREGATOR

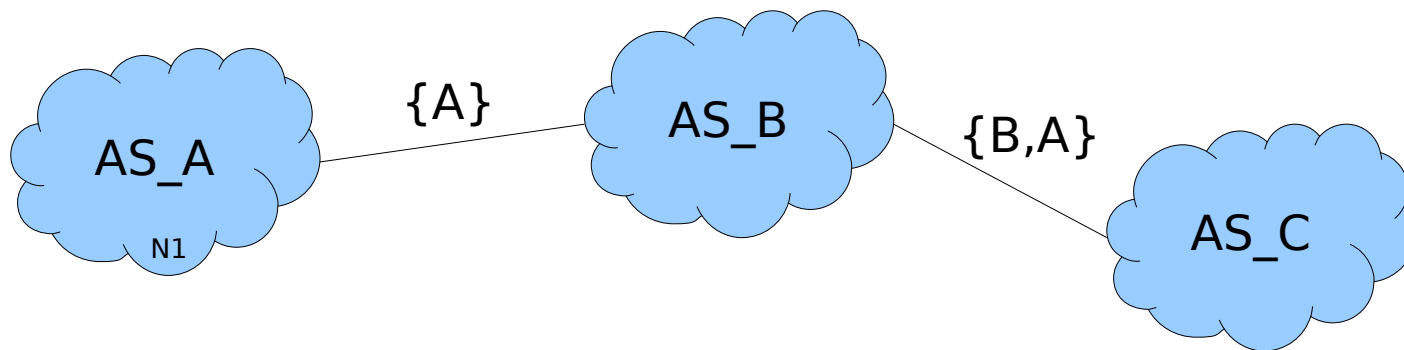
■ ORIGIN

- Es un atributo bien conocido y obligatorio que identifica el origen de una ruta.
- El origen puede ser:
 - IGP: el origen de la ruta está en el propio AS.
 - EGP: la ruta es externa al AS (obsoleto)
 - Incomplete: la ruta se ha inyectado a BGP por algún otro método.
- Ejemplo:
 - AS_A anuncia la red N1 como interna (IGP).
 - AS_B propaga el anuncio de la red N1.



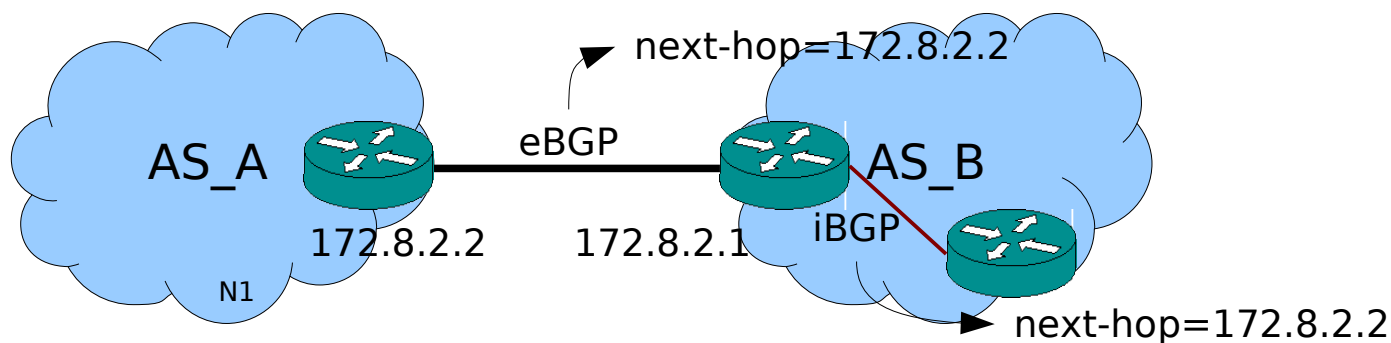
■ AS_PATH

- Atributo bien conocido y obligatorio.
- Incluye información sobre los AS que deben atravesarse para alcanzar el destino.
- Un AS que origina una ruta incluye su propio ASN.
- Cuando un encaminador anuncia a otro una ruta que no tiene origen en él mismo, añade su propio ASN.
- Si un encaminador recibe un anuncio en el que figura su propio ASN en el AS_PATH, rechaza la ruta (detección de bucles).



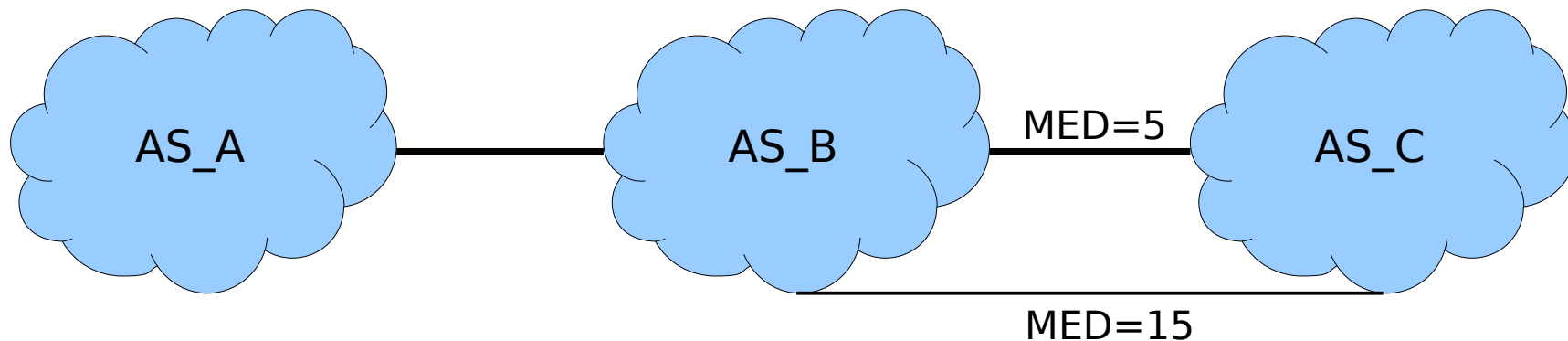
■ NEXT_HOP

- Atributo bien conocido y obligatorio.
- Establece la dirección IP del encaminador frontera que debe utilizarse como siguiente salto en la ruta.
- Cuando se recibe a través de eBGP (BGP exterior), generalmente contiene la dirección del encaminador en el AS remoto
- Se reenvía sin cambios en iBGP (BGP interior).
 - Permite que sea el IGP el que decida la mejor ruta dentro del AS.
 - El encaminador iBGP debe tener una ruta hacia la dirección del *next hop*.



■ MULTI_EXIT_DISC(riminator)

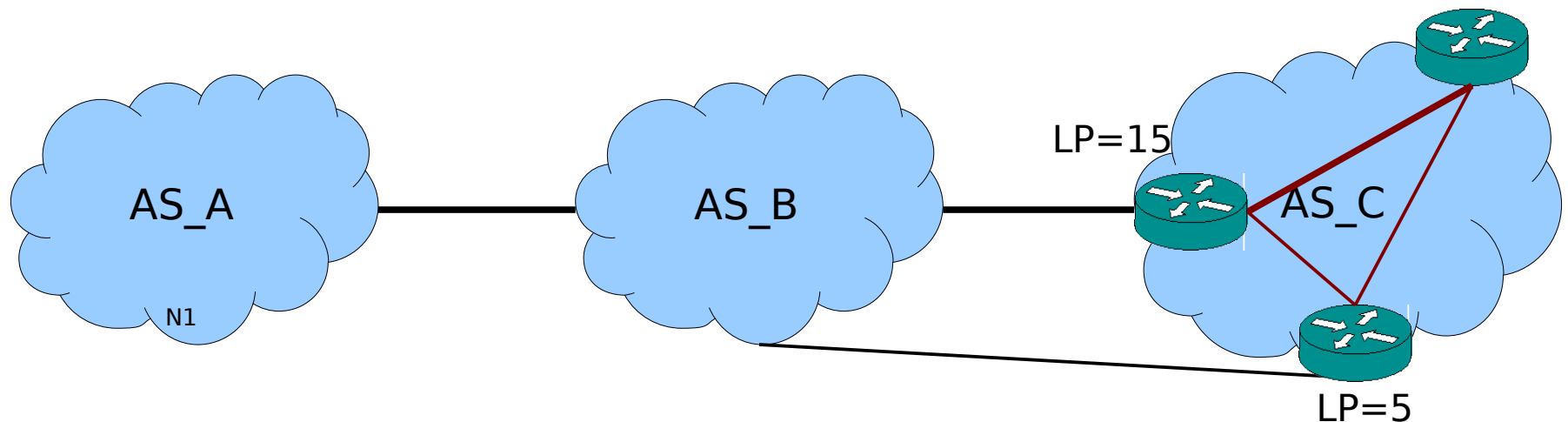
- Atributo opcional no transitivo.
- Cuando existe más de una conexión entre dos AS, el originante puede sugerir al otro extremo una ruta preferida.
- Parámetro menor, ruta preferible.
- Por ejemplo, AS_B **sugiere** a AS_C que la ruta con MED=5 es preferible a la que tiene MED=15 cuando encamine a través de AS_B.
 - Sin embargo, AS_C puede estar usando otra métrica.
 - AS_C puede propagar el valor de MED dentro de su AS (mediante iBGP), pero nunca anunciarlo a otro AS (no transitivo)



BGP

■ LOCAL_PREF

- Atributo bien conocido no obligatorio.
- Indica la preferencia local hacia rutas externas.
- Sólo válido entre encaminadores BGP dentro del mismo AS (se propaga sólo en iBGP, no en eBGP).
- Se prefiere el valor más alto.
 - Por ejemplo, se prefiere la ruta con LP=15 frente a la LP=5.



■ **ATOMIC_AGGREGATE**

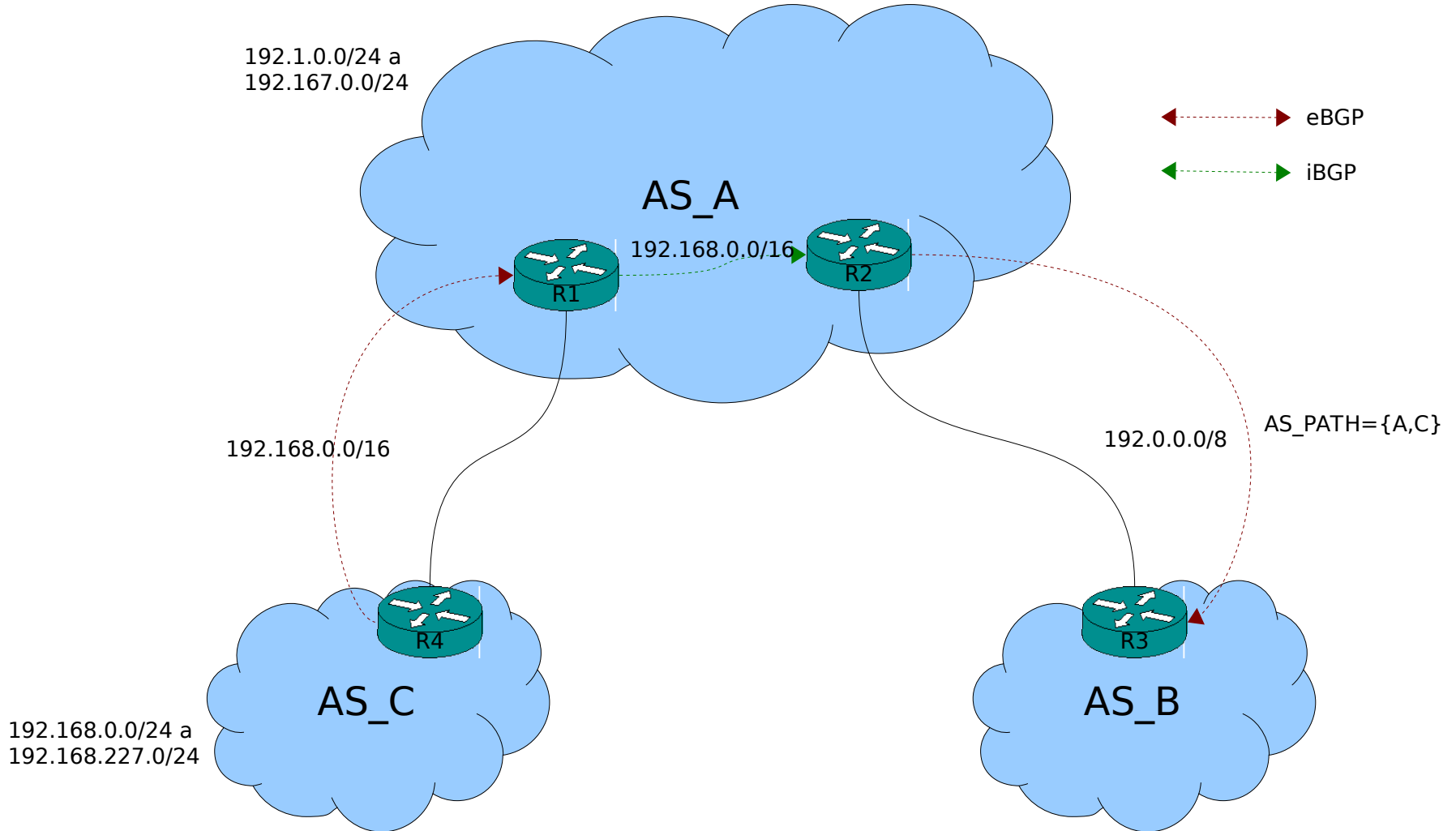
- Atributo bien conocido no obligatorio.
- Cuando un encaminador recibe varios anuncios para redes solapadas a través de distintas rutas, puede optar por anunciar sólo la red que engloba a las demás (la menos específica, el prefijo más corto).
- Este atributo indica que se ha producido un agregado de direcciones.
- Una ruta que incluya este atributo puede, en realidad, atravesar otros AS no incluidos en el AS_PATH.

■ **AGGREGATOR**

- Atributo opcional transitivo.
- Indica el último ASN que formó el agregado de rutas, y la dirección IP del encaminador que lo hizo.

BGP

■ Agregado de rutas



■ Proceso de decisión

- Cuando se han recibido varios anuncios de ruta para un mismo destino, es preciso decidir cuál de ellos se acepta y se usa para encaminar el tráfico.
 1. Se elige aquel que tenga mejor (mayor) Local Preference.
 2. En caso de igualdad, elegir el que se haya originado localmente (en el propio encaminador).
 3. En caso de igualdad, aquel que tenga el AS_PATH más corto.
 4. En caso de igualdad, el que tenga ORIGIN menor (IGP < EGP < Incomplete)
 5. En caso de igualdad, el que tenga mejor (menor) MED.
 6. En caso de igualdad, se prefiere eBGP frente a iBGP.
 7. En caso de igualdad, la ruta que tenga un coste IGP al siguiente salto menor.
 8. En caso de igualdad, el vecino BGP con identificador menor.
 9. En caso de igualdad, el vecino BGP con dirección IP menor.

BGP



Problemas de BGP

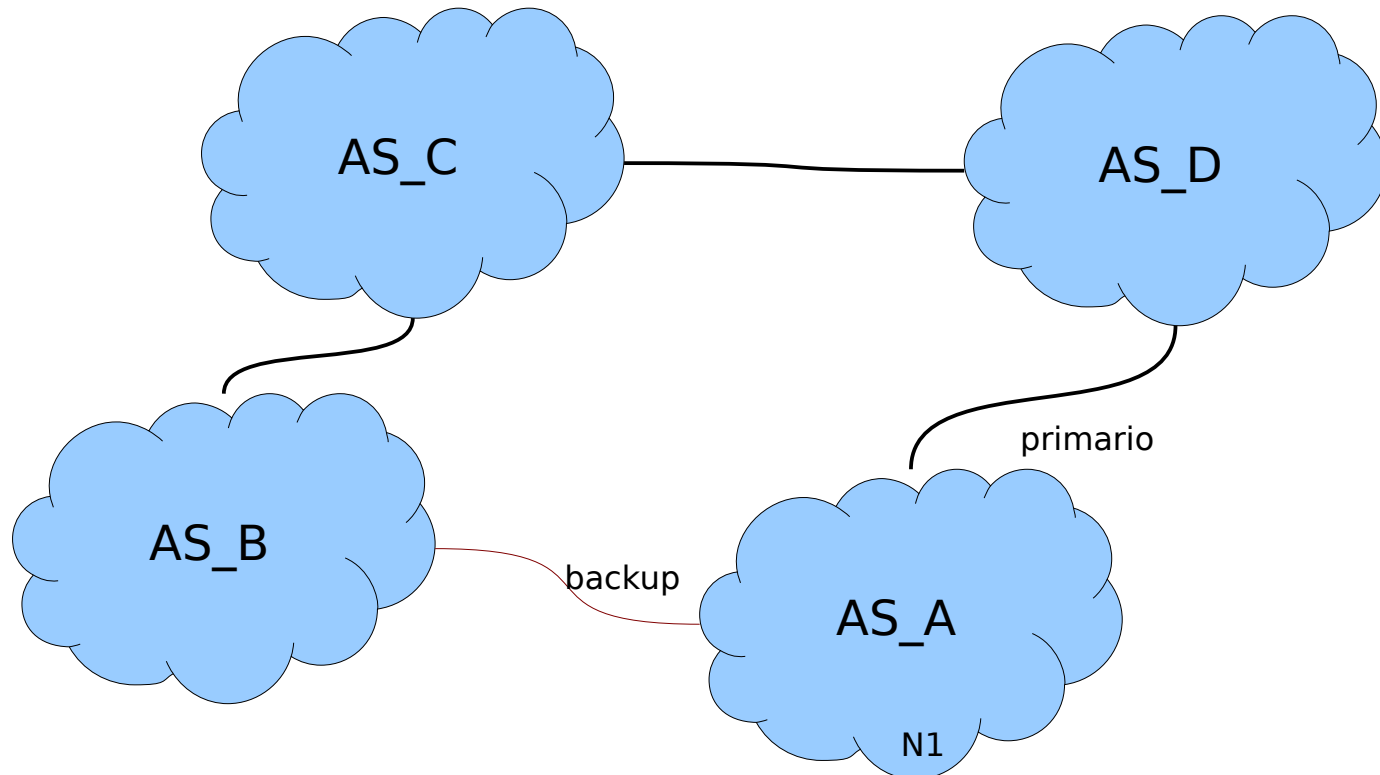
Problemas de BGP

- **BGP es un protocolo de encaminamiento global...**
- **¡Pero las decisiones se toman localmente!**
 - Esto produce:
 - Incongruencias.
 - Comportamientos no deterministas.
 - Oscilaciones (no convergencia).
 - No recuperación ante fallos.
 - Colapso de Internet... fin de la civilización...

Problemas de BGP

■ Wedgies (RFC4264)

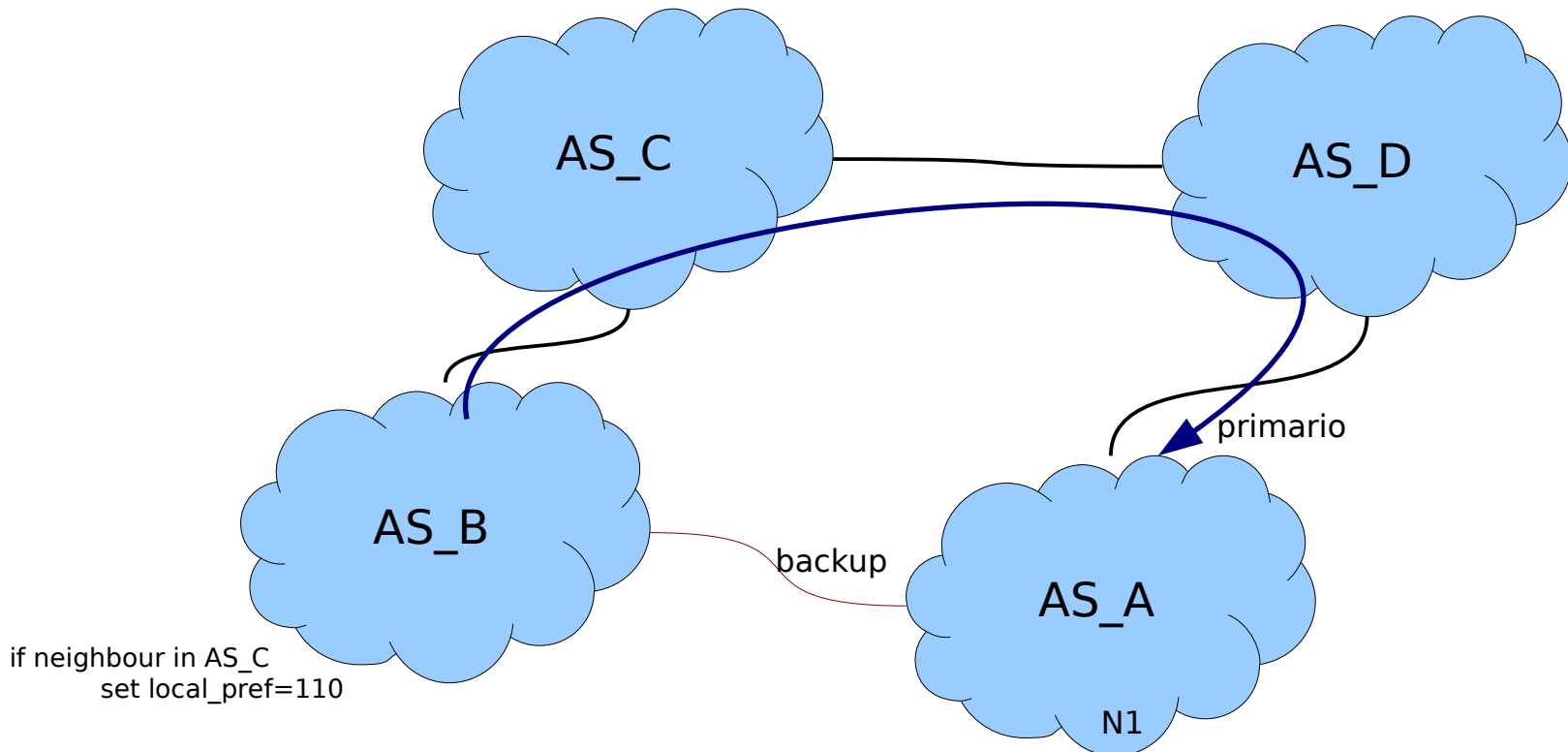
- Son comportamientos no deterministas de BGP.
 - Ante la misma situación de la red, las rutas elegidas no son siempre las mismas.
 - Pueden aparecer estados que no son los esperados.
 - El tráfico sigue rutas no intencionadas.



Problemas de BGP

■ Wedgies (cont.)

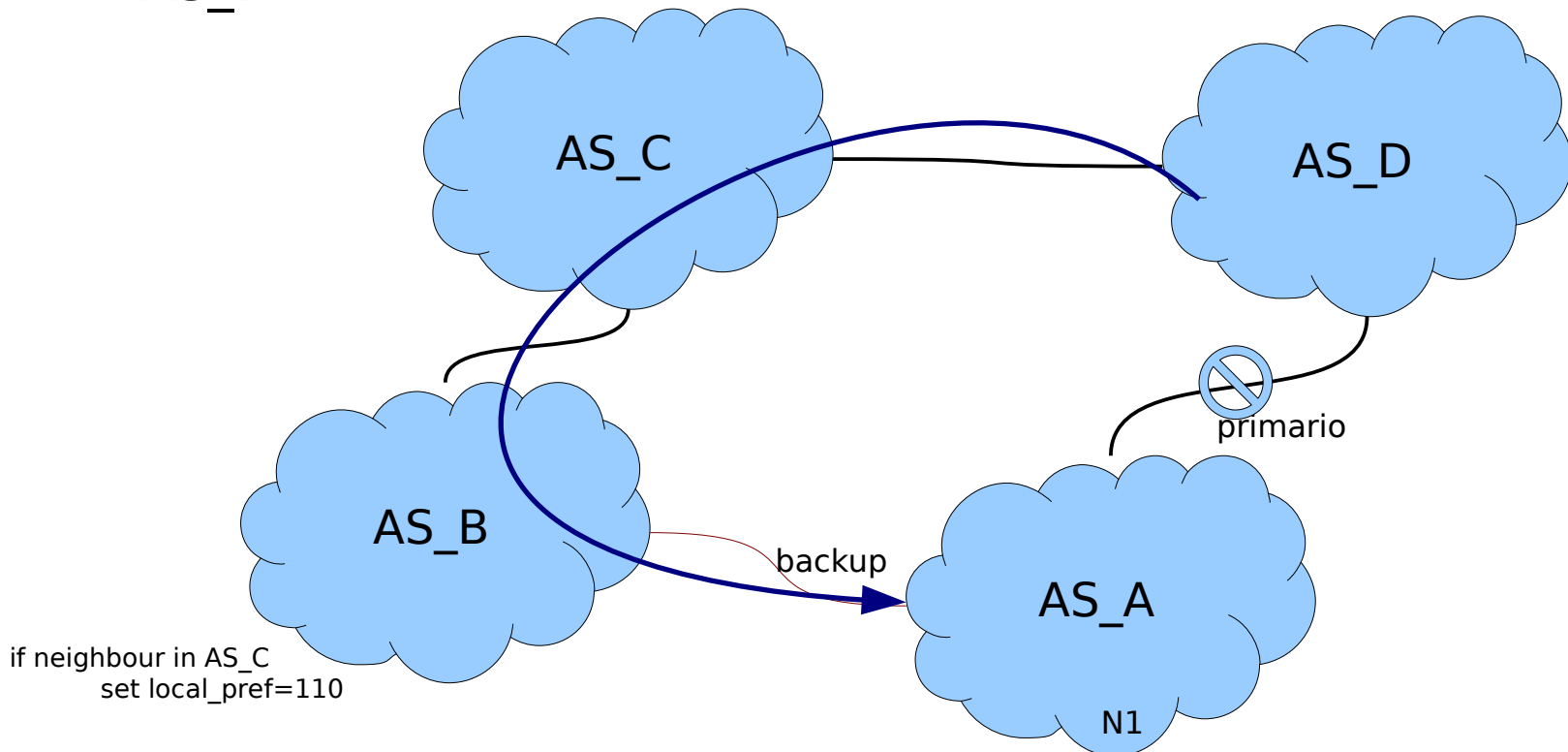
- El Sistema Autónomo A tiene dos enlaces, con AS_D y AS_B.
- El enlace con AS_B es un enlace de *backup*.
- El tráfico hacia la red N1 debe seguir el enlace primario.



Problemas de BGP

■ Wedgies (cont.)

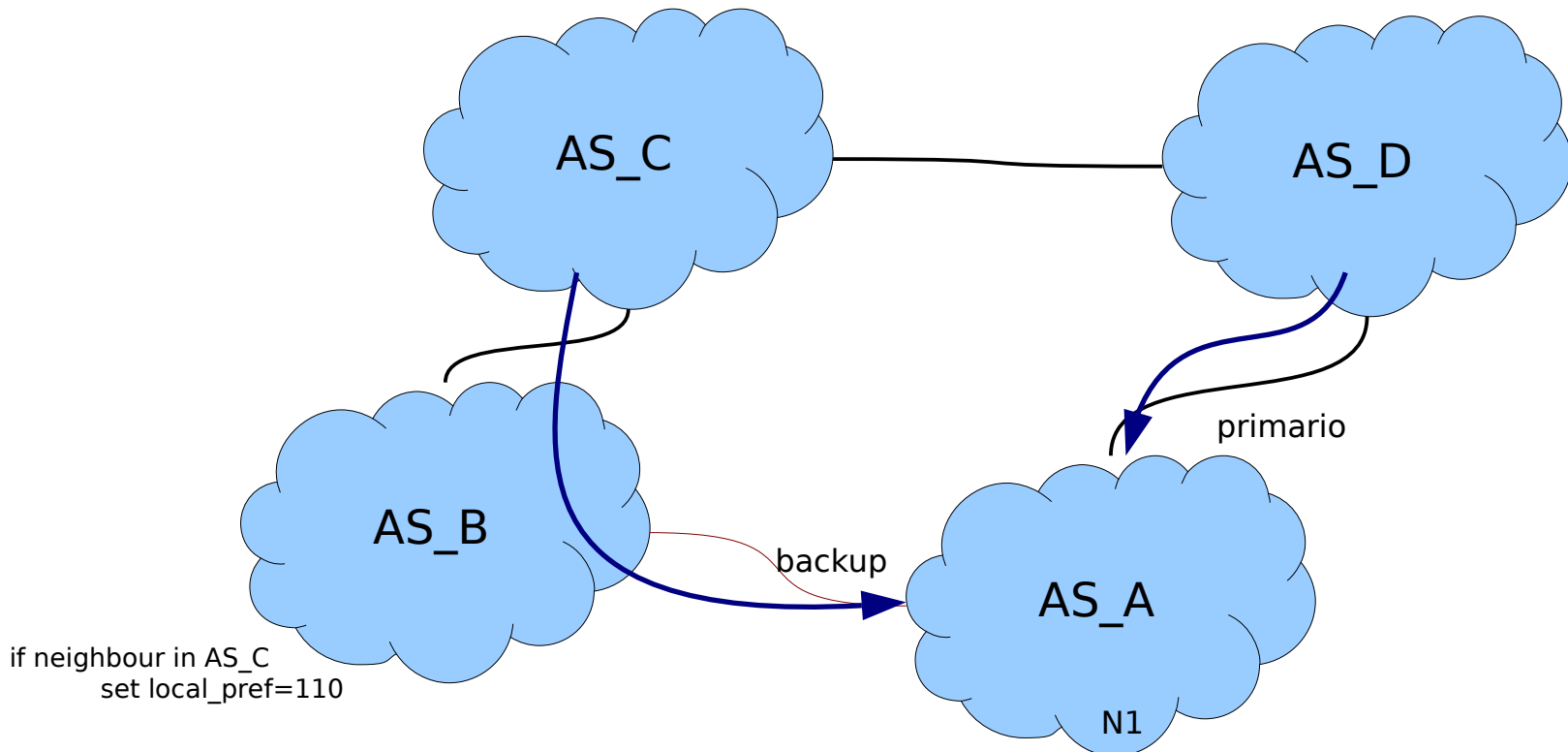
- Si falla el enlace primario, AS_D marcará la ruta a N1 como inaccesible y anunciará el cambio a AS_C.
- AS_C anuncia el cambio a AS_B
- AS_B elige el enlace de *backup* y anuncia N1 a AS_C, y éste a AS_D.



Problemas de BGP

■ Wedgies (cont.)

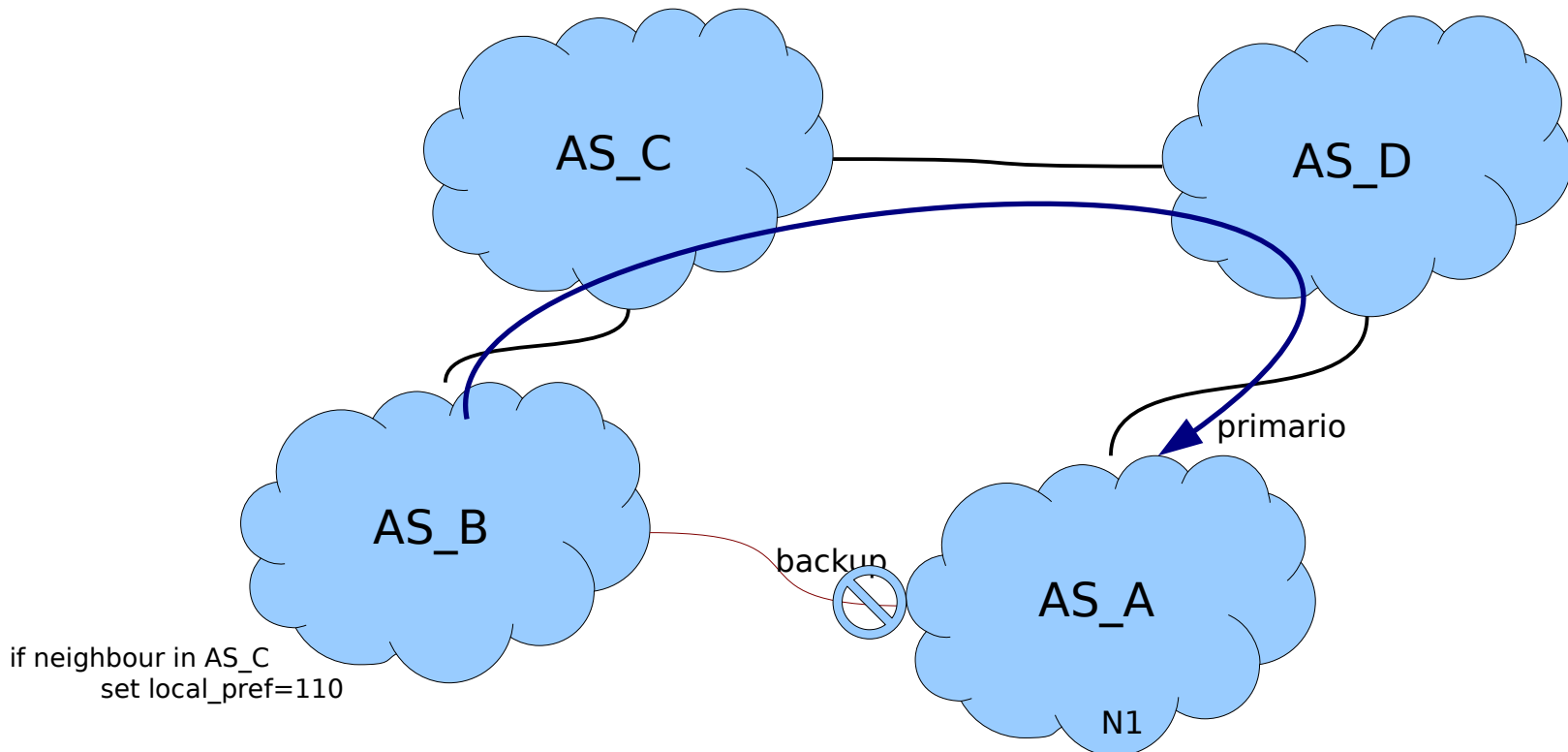
- Cuando se reestablece el enlace primario, AS_D anunciará la ruta a AS_C.
- Pero AS_C seguirá encaminando a través de AS_B.



Problemas de BGP

■ Wedgies (cont.)

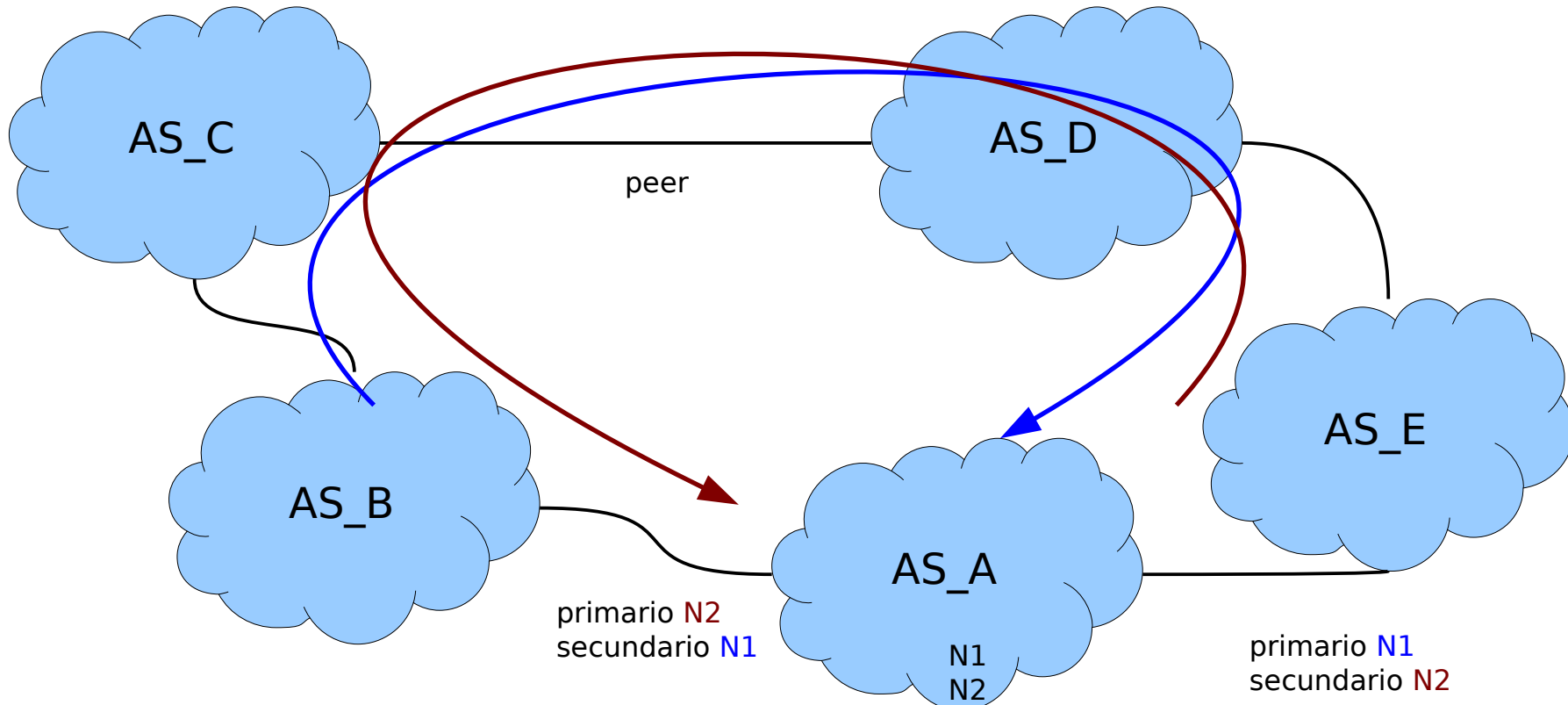
- Para volver a la situación inicial, AS_A cierra la sesión BGP con AS_B.



Problemas de BGP

■ Wedgies 2

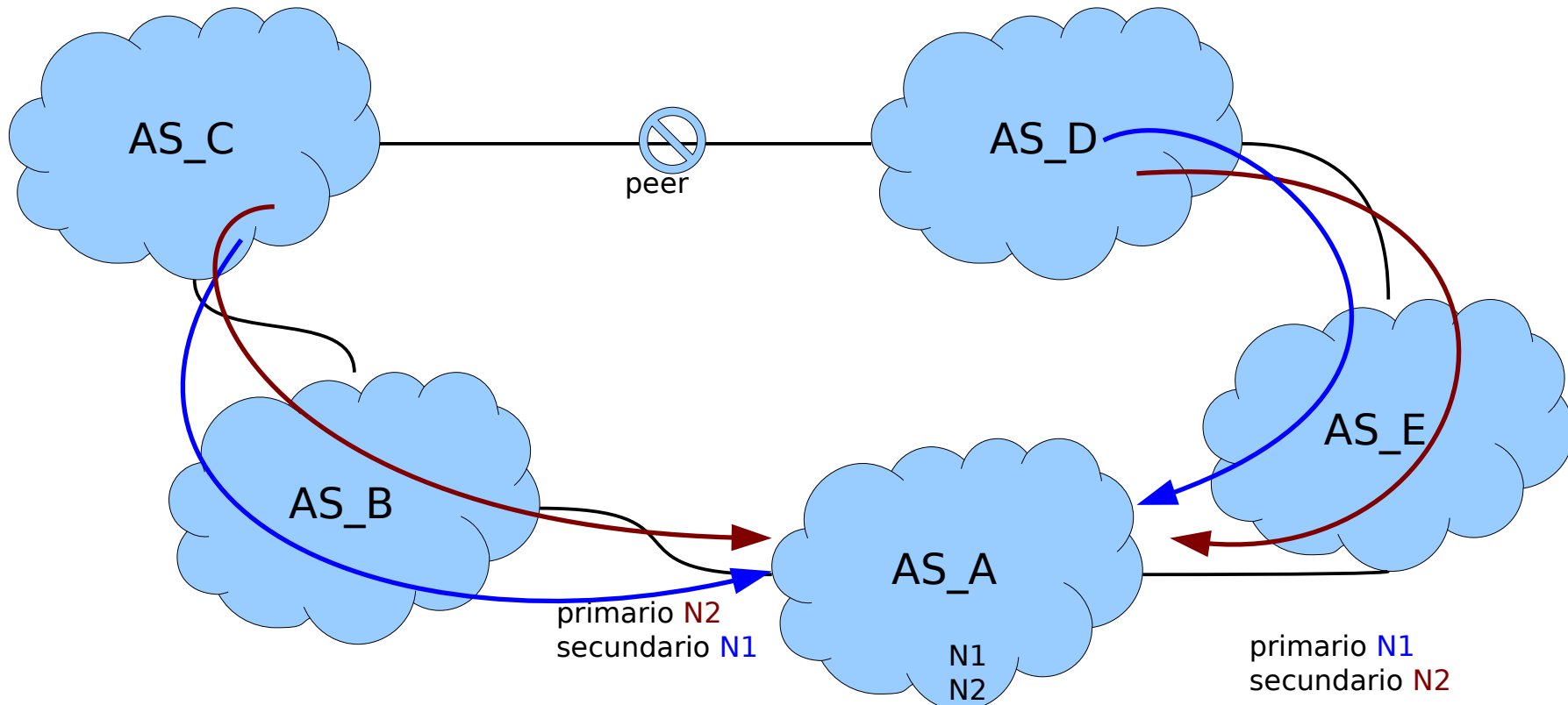
- Una situación más compleja es la siguiente.
- AS_A balancea el tráfico entre sus dos enlaces:
 - Hacia **N1** a través de AS_E.
 - Hacia **N2** a través de AS_B.



Problemas de BGP

■ Wedgies 2

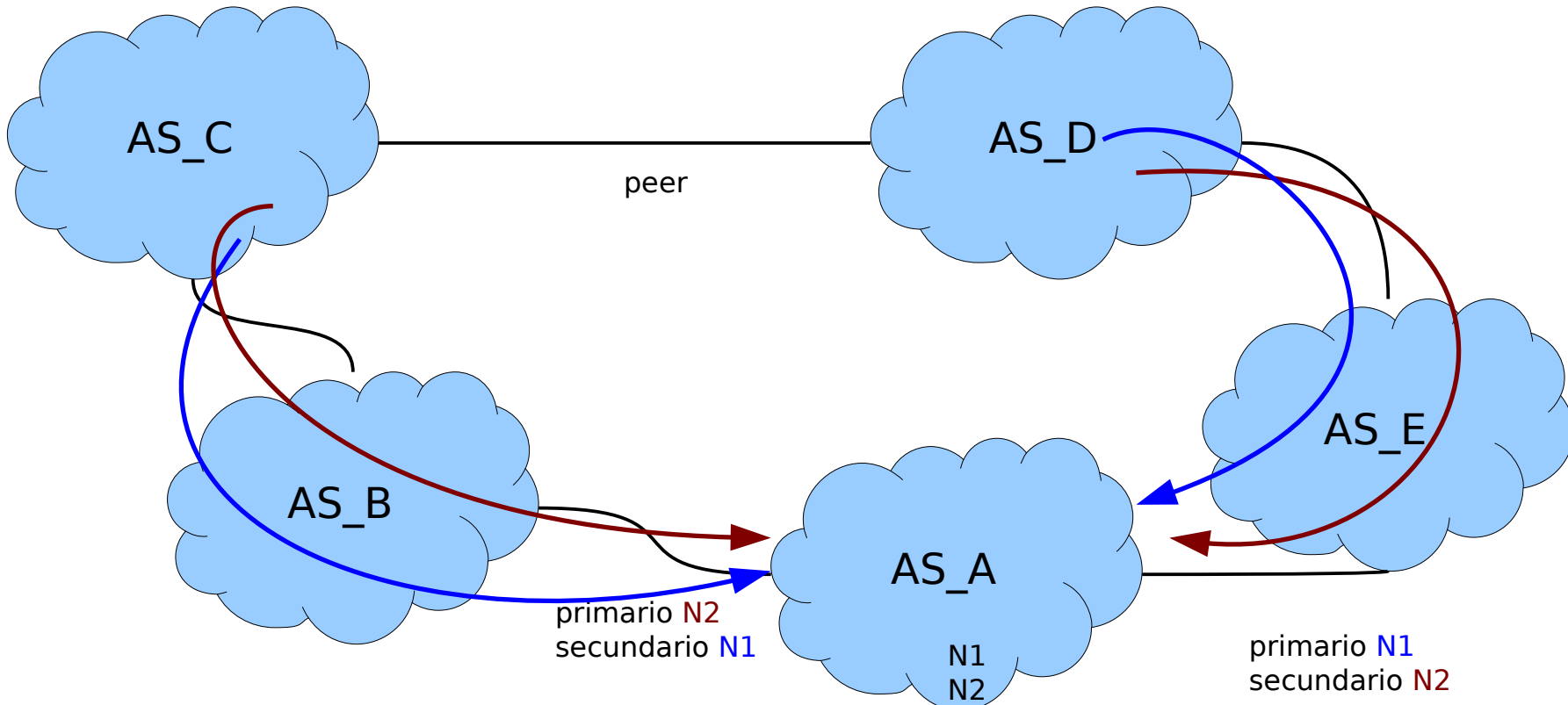
- Si falla el enlace entre AS_C y AS_D:
 - AS_C encamina por AS_B.
 - AS_D encamina por AS_E



Problemas de BGP

■ Wedgies 2

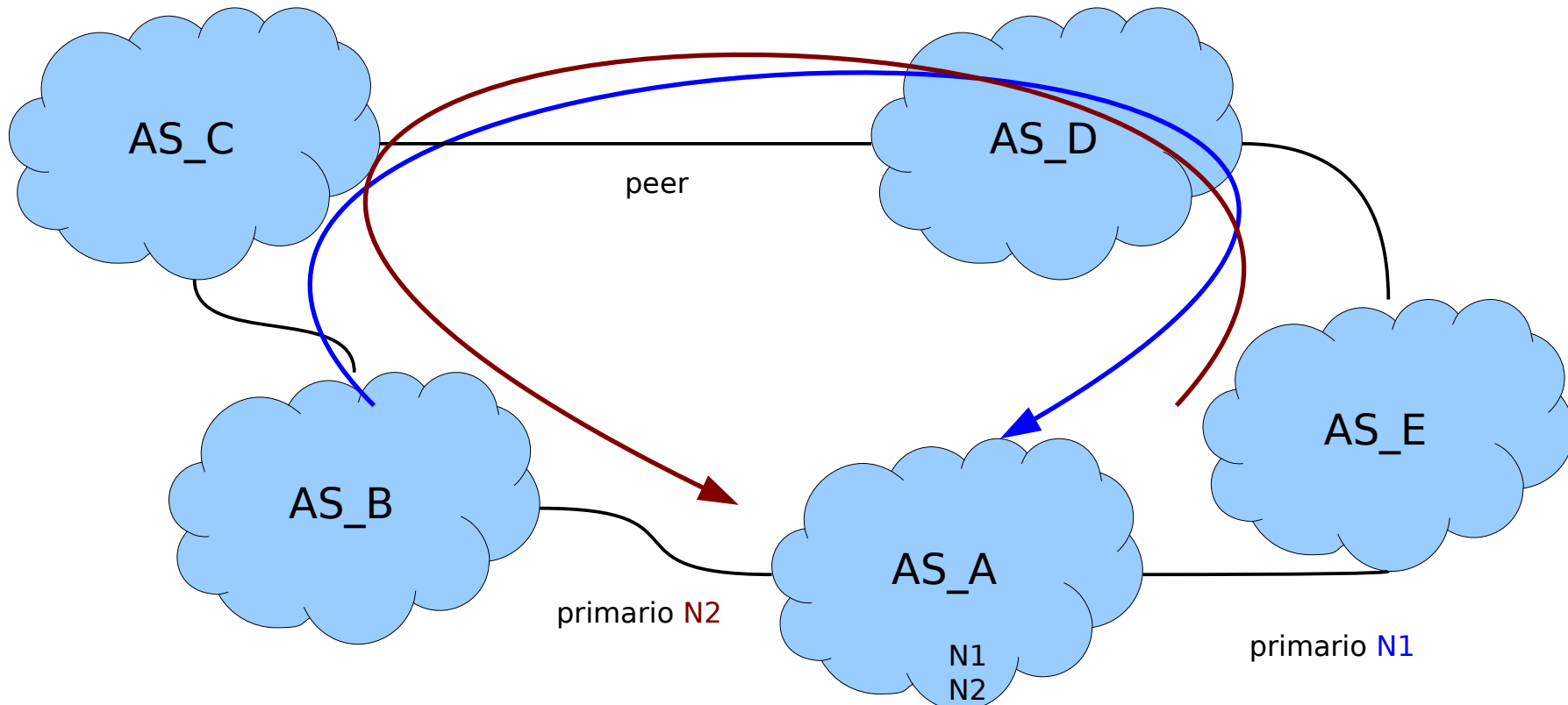
- Cuando se reestablece el enlace entre AS_C y AS_D:
 - Se prefieren las rutas a través de enlaces proveedor-cliente frente a las rutas entre pares (*peer*).
 - AS_A sólo puede dejar de anunciar las rutas secundarias.



Problemas de BGP

■ Wedgies 2

- AS_B elimina la ruta hacia N1.
- AS_C elimina la ruta hacia N1 a través de AS_B.
- AS_C aprende la ruta hacia N1 a través de AS_D.
- AS_B aprende la ruta hacia N1 a través de AS_C.



Problemas de BGP

■ Errores de configuración

- ¿Intencionados o accidentales?
 - 27 de enero de 2011, 22:28UTC
 - El gobierno egipcio ordena la desconexión de Internet

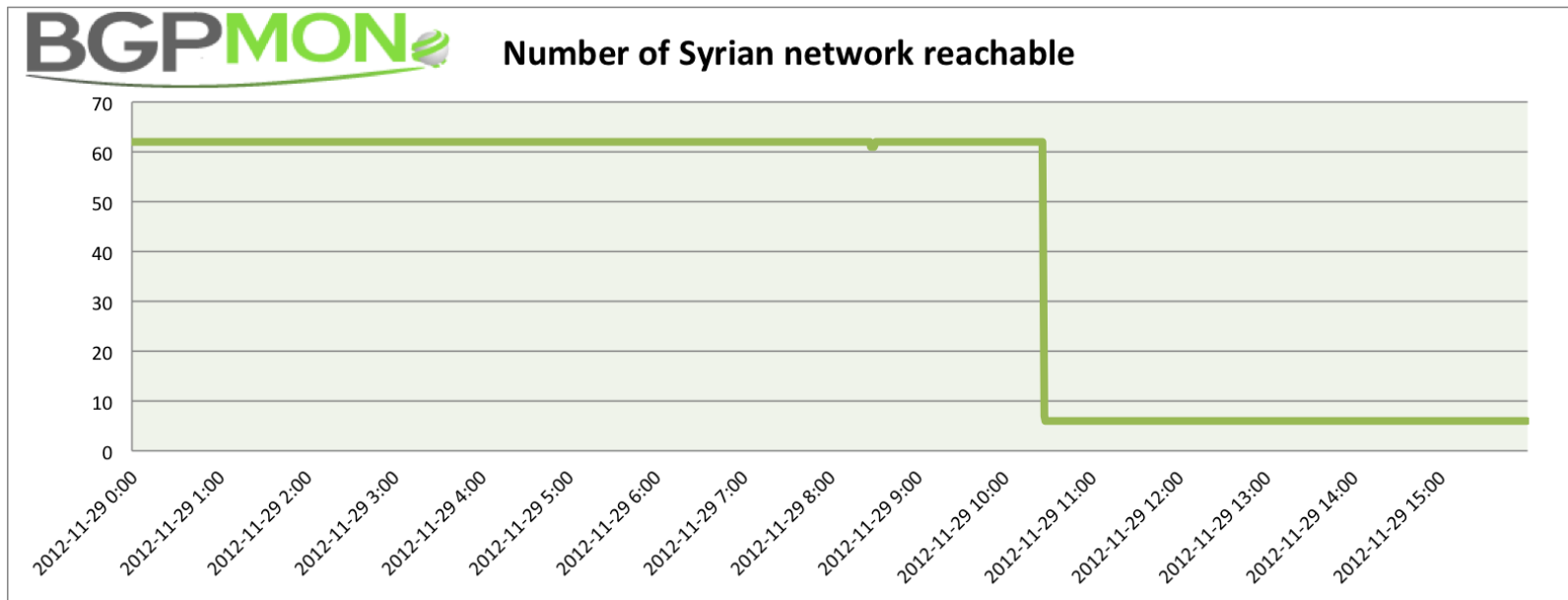
All mobile operators in Egypt have been instructed to suspend services in selected areas. Under Egyptian legislation the authorities have the right to issue such an order and we are obliged to comply with it. The Egyptian authorities will be clarifying the situation in due course .

<http://www.vodafone.com/content/index/press.html>

Problemas de BGP

■ Errores de configuración

- ¿Intencionados o accidentales?
 - 29 de noviembre de 2012, 10:27UTC
 - Siria desaparece de Internet



- 1 de diciembre de 2012, 14:32UTC
 - Se recupera la conectividad

Fuente: www.bgpmon.net

Problemas de BGP

■ Errores de configuración

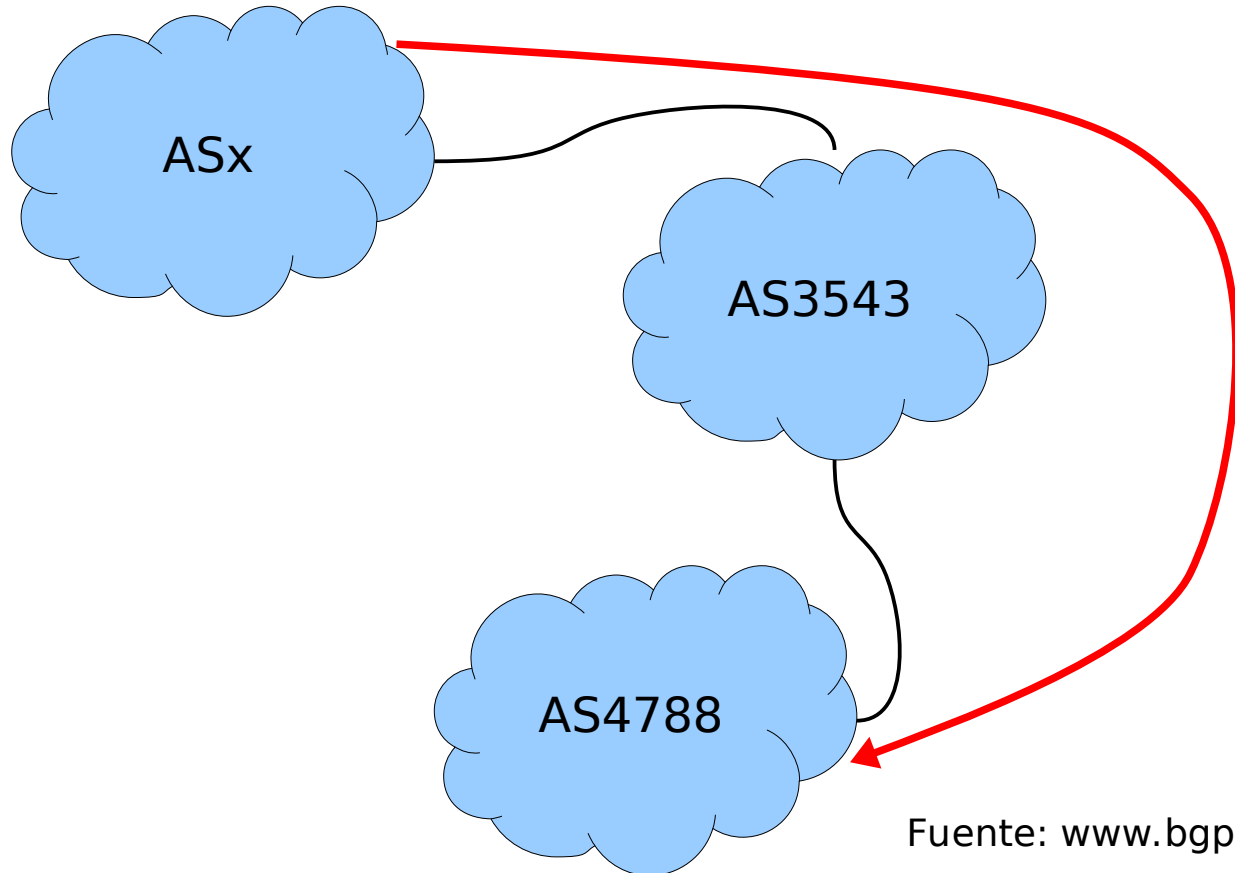
- ¿Intencionados o accidentales?
 - 12 de marzo de 2015, 08:58UTC
 - Los servicios de Google son inaccesibles desde Europa
 - Google AS15169 origina los anuncios y los envía a AS17488 Hathway.
 - AS17488 Hathway los filtra a AS9498 Airtel.
 - Airtel tiene fuerte presencia en Europa.
 - Muchos AS europeos empiezan a redirigir el tráfico hacia Airtel y Hathway.

Fuente: www.bgpmon.net

Problemas de BGP

■ Errores de configuración

- ¿Intencionados o accidentales?
 - 12 de junio de 2015, 08:43UTC
 - AS4788 Telekom Malaysia anuncia unos 179K prefijos a AS3543 Level3



Fuente: www.bgpmn.net

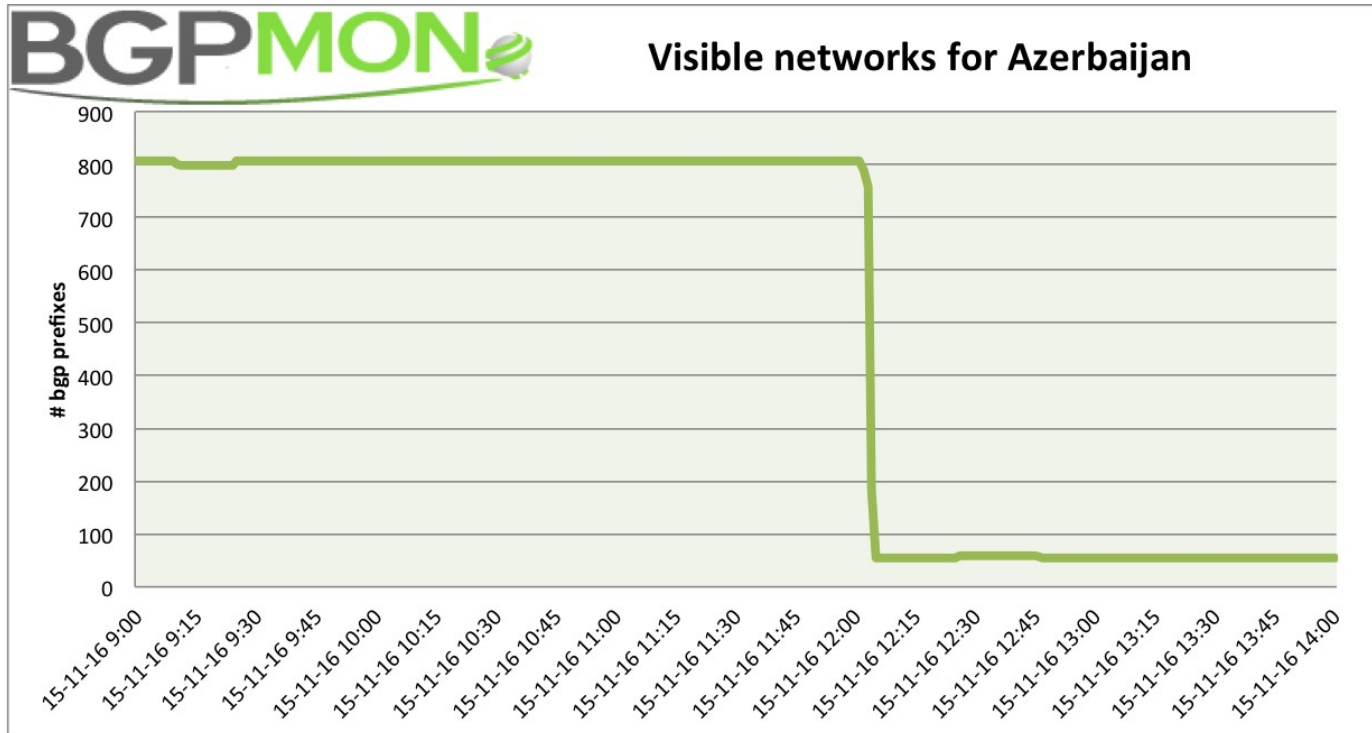
Problemas de BGP

■ Errores de configuración

■ ¿Intencionados o accidentales?

■ 16 de noviembre de 2015, 12:04UTC

- El principal proveedor en Azerbaijan es AS29049, Delta Telecom Ltd.
 - Por un incendio, el 94% de las redes nacionales quedan inaccesibles.



Fuente: www.bgpmon.net

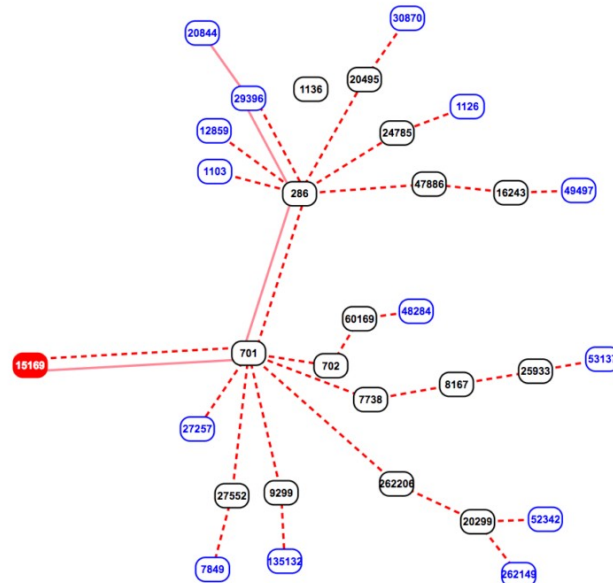
Problemas de BGP

■ Errores de configuración

■ ¿Intencionados o accidentales?

■ 25 de agosto de 2017, 03:22UTC

- OCN (AS4713) es uno de los principales proveedores de Japón y mantiene *peering* con Google.
- Google (AS15169) filtra sus tablas desagregadas a Verizon (AS701) y se convierte en proveedor de tránsito.
- El tráfico hacia OCN empieza a fluir a través de Verizon y Google, saturando sus enlaces o siendo filtrado por alguna ACL.



Fuente: www.bgpmmon.net

BGP

Configuración en Quagga

BGP

■ Configuración

■ Sección *router bgp*

The diagram shows a list of BGP configuration commands. Each command is followed by a red oval highlighting a specific value. A line with an arrow points from a descriptive label to the highlighted value.

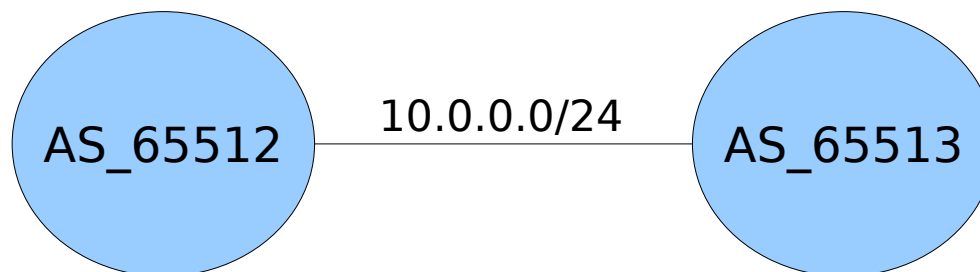
- `router bgp 65512`: The value `65512` is circled in red. An arrow points from the label "Mi ASN" to it.
- `bgp router-id 10.0.0.1`: The value `10.0.0.1` is circled in red. An arrow points from the label "Identificador BGP" to it.
- `network 10.0.0.0/8`: The value `10.0.0.0/8` is circled in red. An arrow points from the label "Anuncio BGP explícito" to it.
- `neighbor 10.0.0.2 remote-as 65513`: The entire line is circled in red. An arrow points from the label "Vecino" to it.
- `aggregate-address 192.168.0.0/16`: The value `192.168.0.0/16` is circled in red. An arrow points from the label "Resumen CIDR" to it.
- `redistribute ospf`: The value `ospf` is circled in red. An arrow points from the label "Inyecta otros protocolos" to it.

BGP

■ Ejemplo (sólo IPv4)

- Los AS 65512 y 65513 se conectan entre sí mediante encaminadores frontera
 - Prefijos anunciados:
 - AS_65512: 10.12.0.0/16, 172.16.12.0/24, 192.168.0.0/23
 - AS_65513: 10.13.0.0/16
- Configuración del encaminador del AS_65512:

```
router bgp 65512
  neighbor 10.0.0.3 remote-as 65513
  network 10.12.0.0/16
  network 172.16.12.0/24
  network 192.168.0.0/23
```



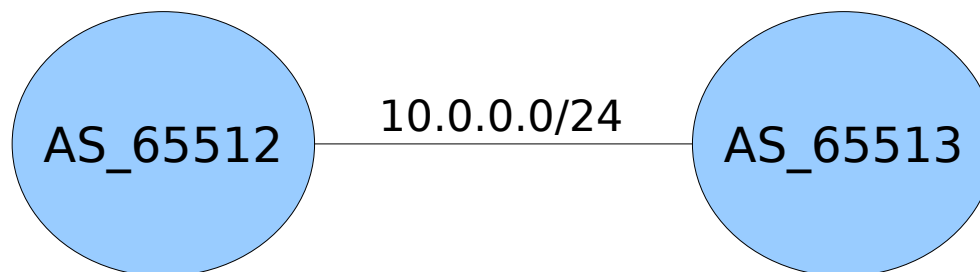
BGP

■ Ejemplo (con IPv4 + IPv6)

- Los AS 65512 y 65513 se conectan entre sí mediante encaminadores frontera, por IPv4 e IPv6.
 - Prefijos anunciados:
 - AS_65512: 147.80.0.0/20, 2001:db8:53f2::/48

```
router bgp 65512
  neighbor 10.0.0.3 remote-as 65513
  network 147.80.0.0/20

  address-family ipv6
    neighbor 10.0.0.3 activate
    network 2001:db8:53f2::/48
  exit-address-family
```



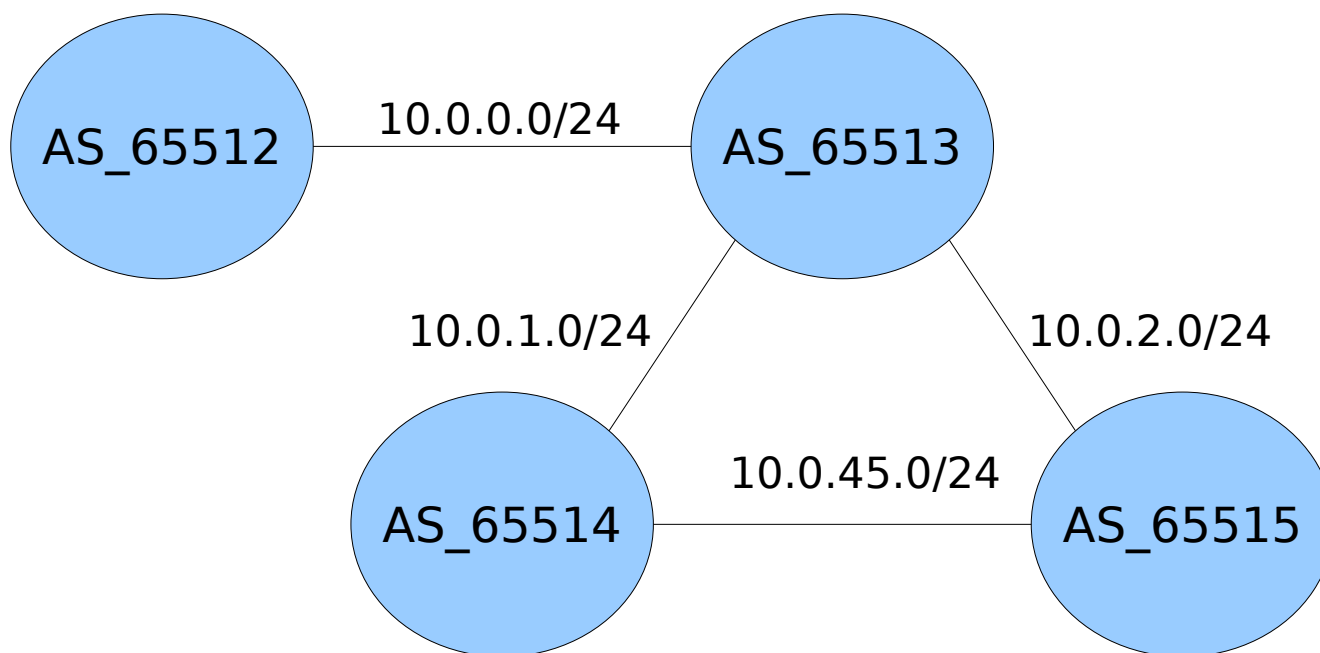
BGP



Práctica

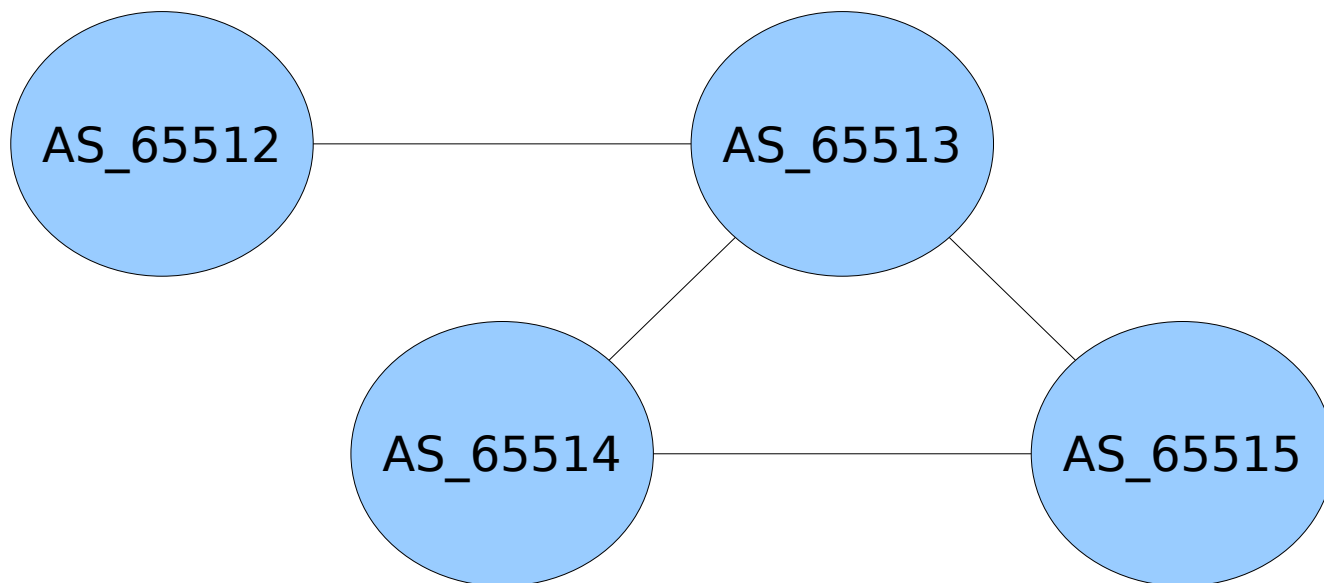
■ Práctica

- Haciendo uso de las máquinas virtuales UML, crear la siguiente configuración:
 - Prefijos anunciados:
 - AS_65512: 10.12.0.0/16, 172.16.12.0/24, 192.168.0.0/23
 - AS_65513: 10.13.0.0/16
 - AS_65514: 10.14.0.0/16, 172.16.14.0/24
 - AS_65515: 10.15.0.0/16, 172.16.15.0/24, 192.168.128.0/23



■ Práctica

- Modificando la práctica anterior, se debe incorporar el soporte para IPv6.
 - Prefijos anunciados (además de los anunciados por IPv4):
 - AS_65512: 2001:db8:12::/52, 2001:db8:12:1000::/52, 2001:db8:12:2000::/52
 - AS_65513: 2001:db8:13:1::/64, 2001:db8:13:2::/64
 - AS_65514: 2001:db8:14::/64, 2001:db8:14:2::/64, 2001:db8:14:3::/64
 - AS_65515: 2001:db8:15::/64, 2001:db8:15:1::/64, 2001:db8:15:2::/64



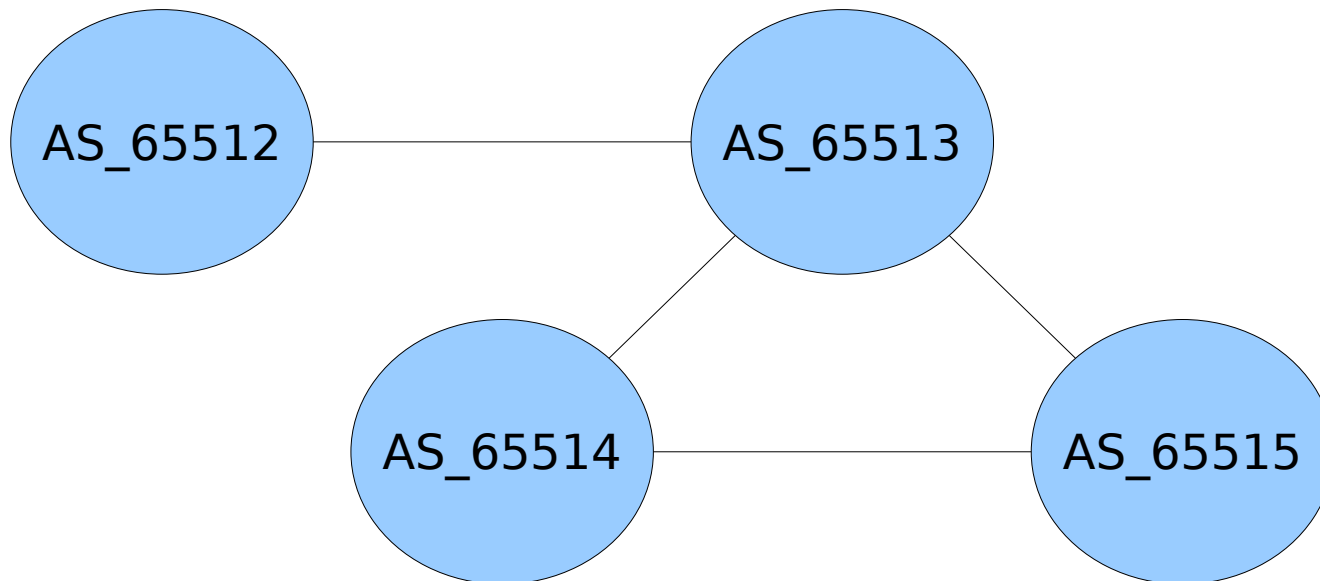
BGP

■ Práctica

- Comprobar las rutas y datos de BGP:

vttysh> show ip[v6] bgp

vttysh> show ip bgp neighbor

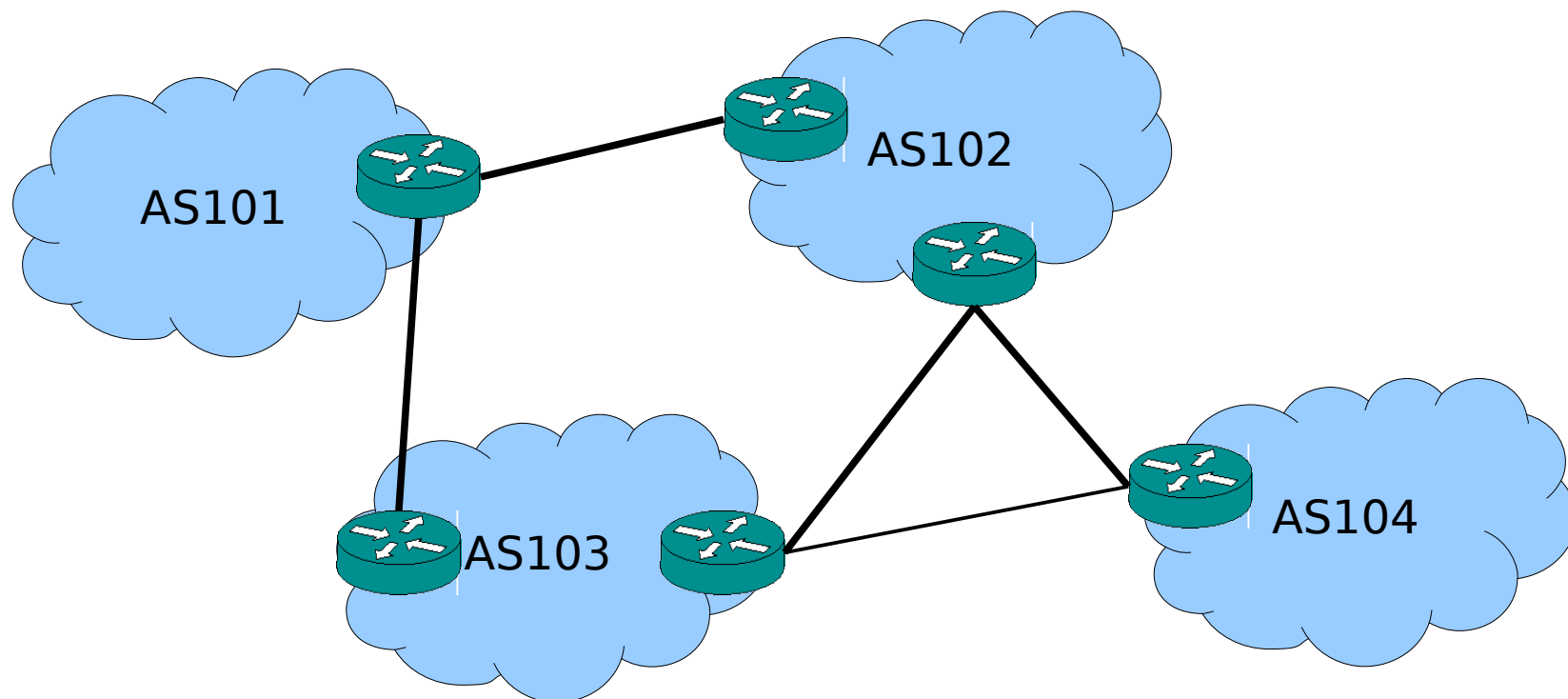


BGP

Encaminamiento basado en políticas

■ Filtrado de rutas

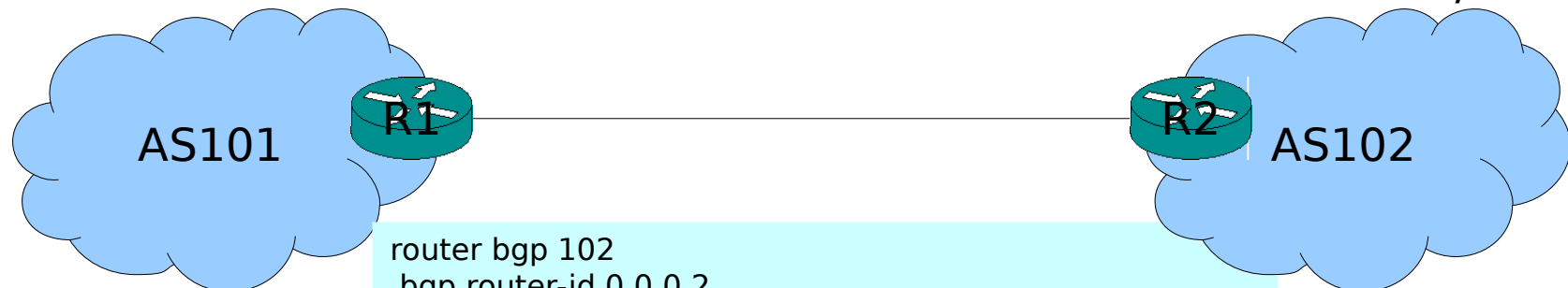
- Se puede establecer un encaminamiento basado en políticas
 - **prefix-list** (filtra prefijos)
 - **filter-list** (filtra ASN)
 - **route-map** (filtra y modifica atributos)
 - **distribute-list** (filtra anuncios de prefijos)



BGP

■ Filtrado de rutas

- prefix-list: filtrado de prefijos (in|out)



3ffe:2800:e800::/40
3ffe:202::/33
2001:db8:1::/64

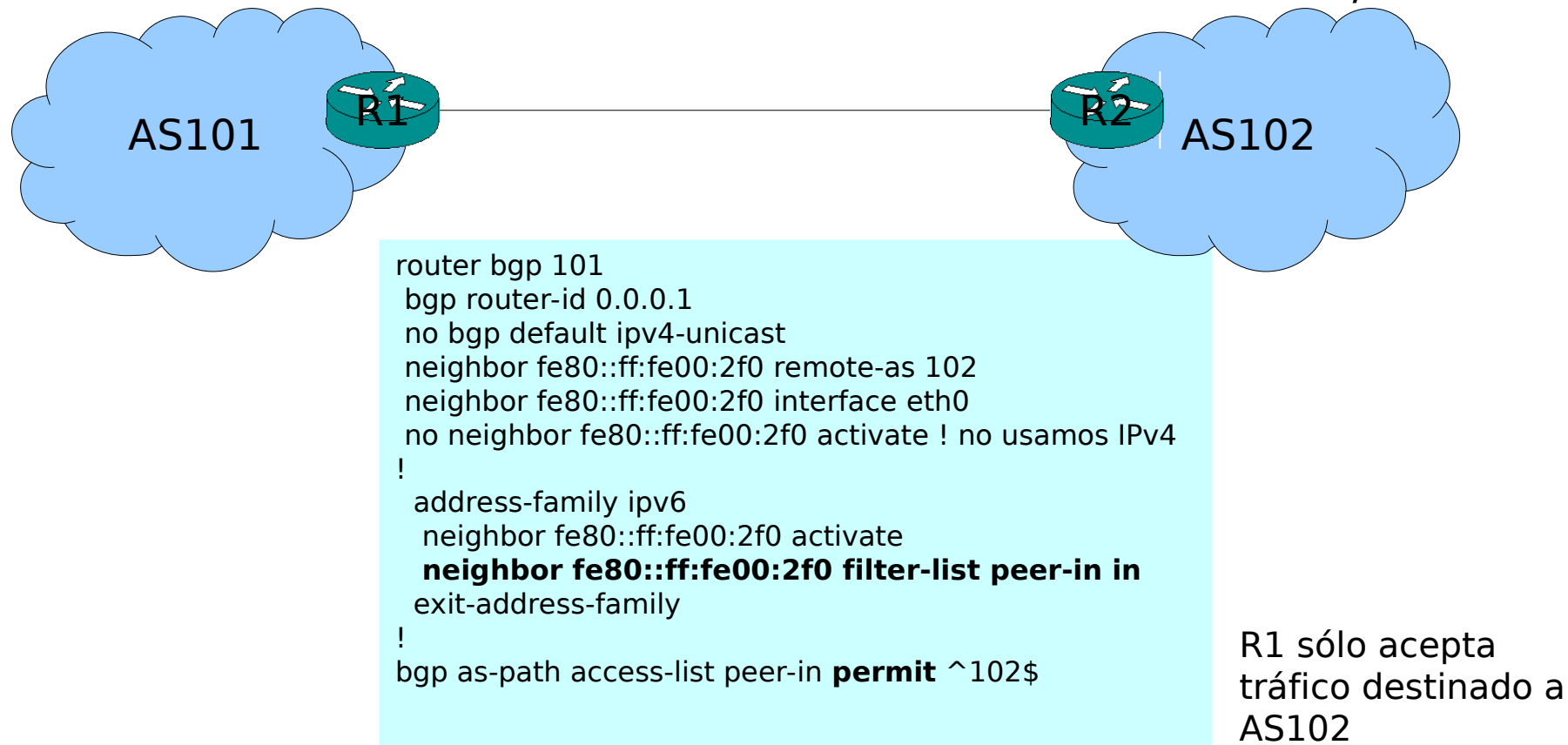
```
router bgp 102
  bgp router-id 0.0.0.2
  no bgp default ipv4-unicast
  neighbor fe80::ff:fe00:1f0 remote-as 101
  neighbor fe80::ff:fe00:1f0 interface eth0
  no neighbor fe80::ff:fe00:1f0 activate ! no usamos IPv4
  !
  address-family ipv6
    network 3ffe:202::/33
    network 3ffe:2800:e800::/40
    network 2001:db8:1::/64
    neighbor fe80::ff:fe00:1f0 activate
    neighbor fe80::ff:fe00:1f0 prefix-list peer-out out
  exit-address-family
  !
  ipv6 prefix-list peer-out deny 3ffe:2800::/32 le 64
  ipv6 prefix-list peer-out permit 3ffe::/16 ge 20 le 64
  ipv6 prefix-list peer-out permit 2001::/16 le 64
```

R2 no acepta tráfico
hacia el prefijo
3ffe:2800::/32
proveniente de R1

BGP

■ Filtrado de rutas

- filter-list: filtrado por ASN (in|out)



■ Filtrado de rutas

- as-path access-list: expresiones regulares
 - Se pueden emplear expresiones regulares en el AS path
 - . Cualquier carácter.
 - * 0 o más apariciones del patrón.
 - + 1 o más apariciones del patrón.
 - ? 0 ó 1 apariciones del patrón.
 - ^ Comienzo de línea.
 - \$ Final de línea.
 - _ Concuerda con el espacio y la coma, así como con el delimitador de conjunto de ASN { y }, y con el delimitador de confederación (y). También concuerda con el principio y final de línea.
 - **show ip bgp regexp _1180_** muestra todas las rutas BGP que contienen el ASN 1180.

■ Filtrado de rutas

- as-path access-list: expresiones regulares

- Ejemplos:

- . * cualquier cosa

- . + al menos un carácter

- ^ \$ rutas locales a este AS

- _ **100** \$ con origen en AS100

- ^ **100** _ recibidas desde AS100

- _ **100** _ vía AS100

- _ **200_100** _ vía AS100 y luego AS200

- _ (**100** _) + varios AS100 consecutivos (para detectar *AS_PATH prepends*)

■ Filtrado de rutas

- as-path access-list: sintaxis

bgp as-path access-list *<nombre>* {permit|deny} *<regex>*

- Ejemplos:

- bgp as-path access-list **n1** permit _100\$

- bgp as-path access-list **n2** permit _102_

- bgp as-path access-list **n3** deny _105_

bgp as-path access-list **n3** permit .*

- Explicación:

- **n1** sólo permite rutas originadas en el AS100.

- **n2** sólo permite rutas que pasen por AS102.

- **n3** prohíbe todas las rutas que pasen (o provengan) de AS105.

■ Filtrado de rutas

■ route-map

- Tiene una sección **match**
- Puede tener una sección **set**
- La sección **set** se ejecuta si se cumple toda la sección **match**
- Sintaxis:

```
route-map <nombre> {permit|deny} <seq>  
    [no] match ...  
    [no] set ...
```

▪ Ejemplo:

```
router bgp 100  
    neighbour a.b.c.d route-map ejemplo in  
route-map ejemplo permit 10  
    match as-path origen102  
    set local-preference 80  
route-map ejemplo permit 20  
    match ipv6 address prefix-list remoto  
    match origin incomplete  
    set metric +100
```

■ Filtrado de rutas

■ route-map: sintaxis

neighbour <vecino> route-map <nombre> {in|out}

route-map <nombre> {permit|deny} <seq>

[description <texto>]

[no] match {

as-path <as_path>|

community {<1-99>|<100-500>|<nombre>}|

ip {address|next-hop|route-source}{<access_list>|prefix-list
<prefix_list>}|

ipv6 {address|next-hop} {<access_list>|prefix-list <prefix_list>}|

metric <valor>|

origin {egp|igp|incomplete}|

pathlimit as <asn>|

peer {<A.B.C.D>|<X:X::X:X>|local}

}

■ Filtrado de rutas

■ route-map: sintaxis

```
route-map <nombre> {permit|deny} <seq>
```

```
[no] match ...
```

```
[no] set {
```

```
as-path exclude .<1-4294967295>|
```

```
set as-path prepend .<1-4294967295>|
```

```
community {<AA:NN>|none}|
```

```
ip next-hop {<A.B.C.D>|peer-address}|
```

```
ipv6 next-hop {global|local} <X:X::X:X>|
```

```
local-preference <0-4294967295>|
```

```
metric [{+|-}]<0-4294967295>|
```

```
originator-id <A.B.C.D>|
```

```
pathlimit ttl <1-255>|
```

```
vpn4 next-hop <A.B.C.D>|
```

```
weight <0-4294967295>|
```

```
}
```

■ Filtrado de rutas

- route-map
 - Tiene una sección **match**
 - Puede tener una sección **set**

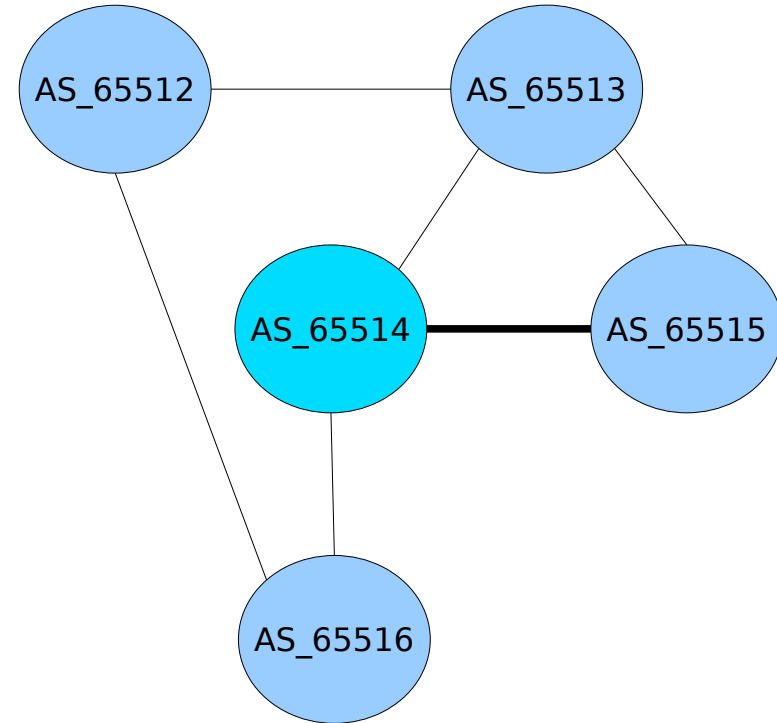
```
router bgp 65513
  bgp router-id 0.0.0.2
  network 180.3.3.0/24
  network 172.16.0.0/16
  neighbor 10.0.0.1 remote-as 65512
  neighbor 10.0.0.1 route-map internal-rtm out
  neighbor 10.0.0.1 route-map 65512-only in
!
access-list internal permit 172.16.0.0/16
access-list internal permit 180.3.3.0/24
access-list internal deny any ! implícito
!
route-map internal-rtm permit 10
  match ip address internal
  set as-path prepend 65513 65513 65513
  set metric 100
```

■ Filtrado de rutas

■ route-map

■ Ejemplo:

- El AS65514 desea usar preferentemente su enlace con AS65515, y mantener los otros dos como respaldo (*backup*)
- Implementar esa política mediante el uso de route-map.



BGP

■ Filtrado de rutas

■ route-map

■ Ejemplo:

```
router bgp 65514
  bgp router-id 10.0.44.4
  neighbor 10.0.23.2 remote-as 65513
  neighbor 10.0.34.4 remote-as 65515
  neighbor 10.0.34.4 route-map as65515 in
  neighbor 10.0.35.5 remote-as 65516
  network 172.16.14.0/24
```

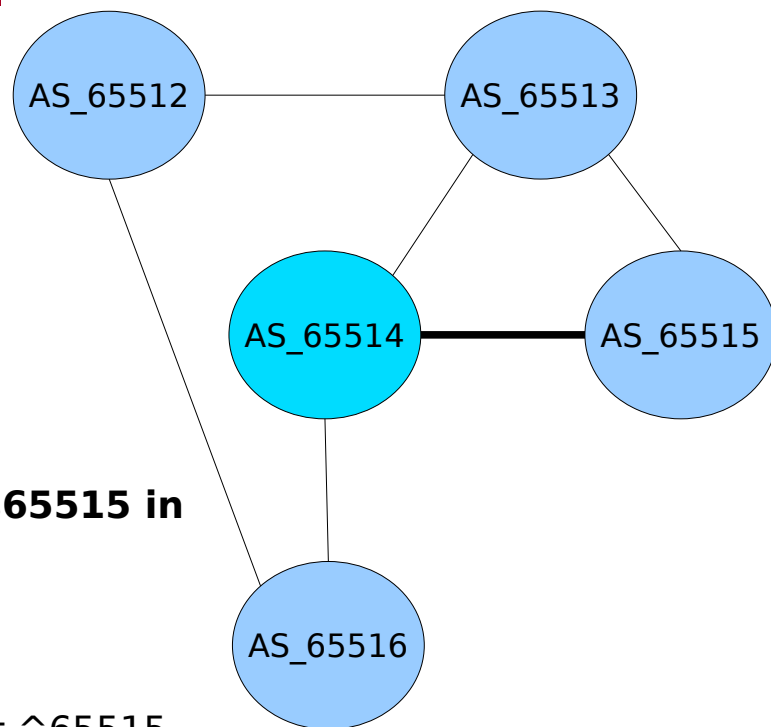
```
bgp as-path access-list preferente permit ^65515_
bgp as-path access-list cualquiera permit .*
```

!

```
route-map as65515 permit 10
  match as-path preferente
  set local-preference 200
```

!

```
route-map as65515 permit 20
  match as-path cualquiera
```



BGP

■ Filtrado de rutas

■ route-map

■ Ejemplo (más simple):

```
router bgp 65514
```

```
  bgp router-id 10.0.44.4
```

```
  neighbor 10.0.23.2 remote-as 65513
```

```
  neighbor 10.0.34.4 remote-as 65515
```

```
  neighbor 10.0.35.5 remote-as 65516
```

```
  neighbor 10.0.34.4 route-map primaria in
```

```
  neighbor 10.0.23.2 route-map backup out
```

```
  neighbor 10.0.35.5 route-map backup out
```

```
  network 172.16.14.0/24
```

```
!
```

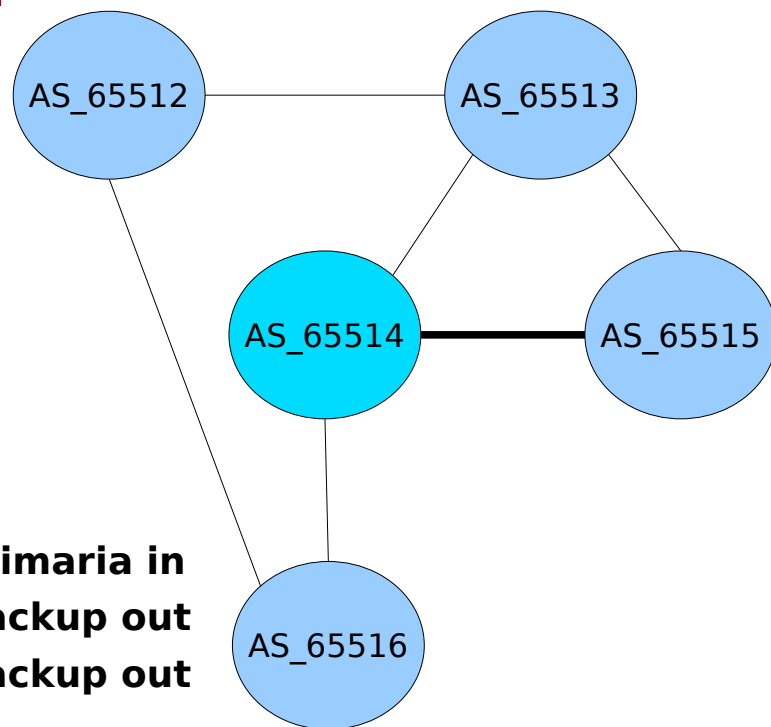
```
route-map primaria permit 10
```

```
  set local-preference 200
```

```
!
```

```
route-map backup permit 10
```

```
  set as-path prepend 65514 65514 65514 65514 65514 65514
```



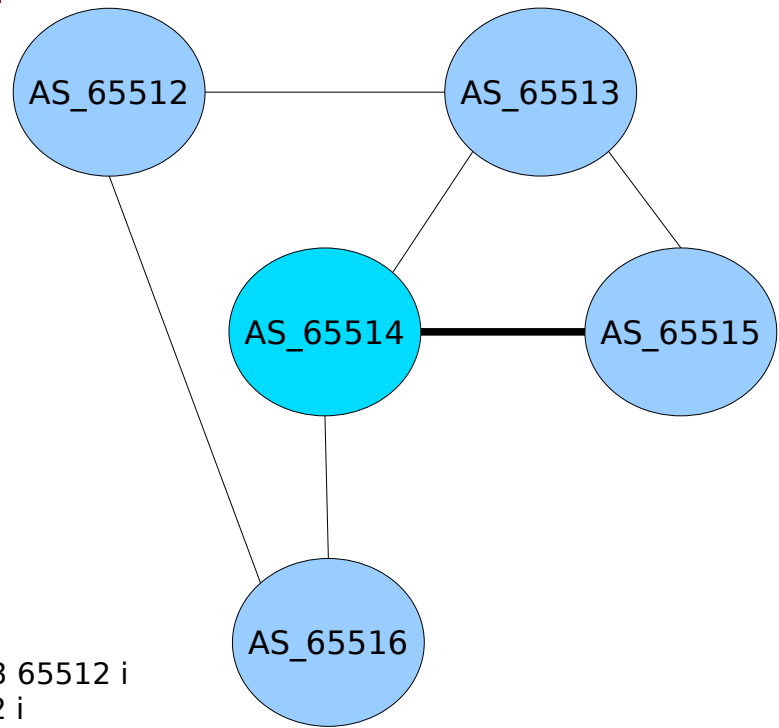
Filtrado de rutas

- route-map
- Ejemplo:

show ip bgp
BGP table version is 0, local router ID is 10.0.35.3

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 172.16.12.0/24	10.0.34.4	200	0	0	65515 65513 65512 i
*	10.0.35.5			0	65516 65512 i
*	10.0.23.2			0	65513 65512 i
*> 172.16.13.0/24	10.0.34.4	200	0		65515 65513 i
*	10.0.35.5			0	65516 65512 65513 i
*	10.0.23.2	0	0	0	65513 i
*> 172.16.14.0/24	0.0.0.0	0		32768	i
*> 172.16.15.0/24	10.0.34.4	0	200	0	65515 i
*	10.0.23.2			0	65513 65515 i
* 172.16.16.0/24	10.0.23.2			0	65513 65512 65516 i
*>	10.0.35.5	0		0	65516 i

Total number of prefixes 5



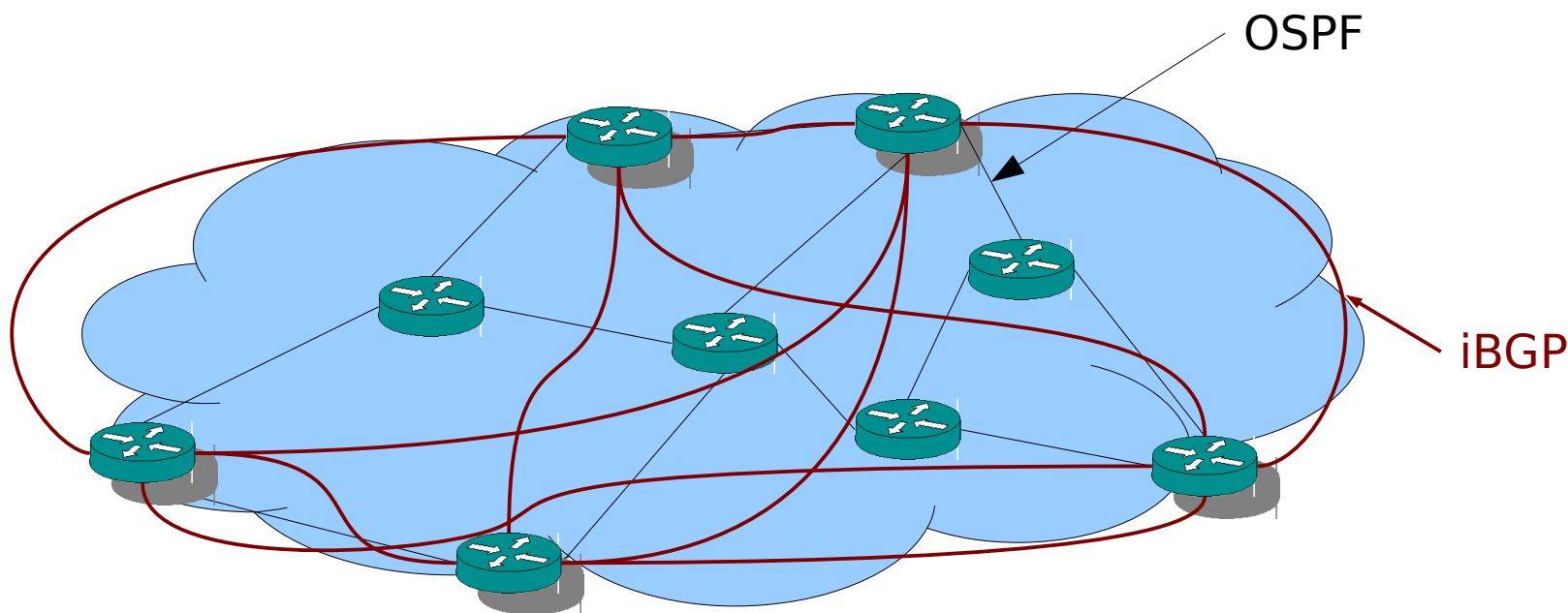
← ¡No es el esperado!

BGP



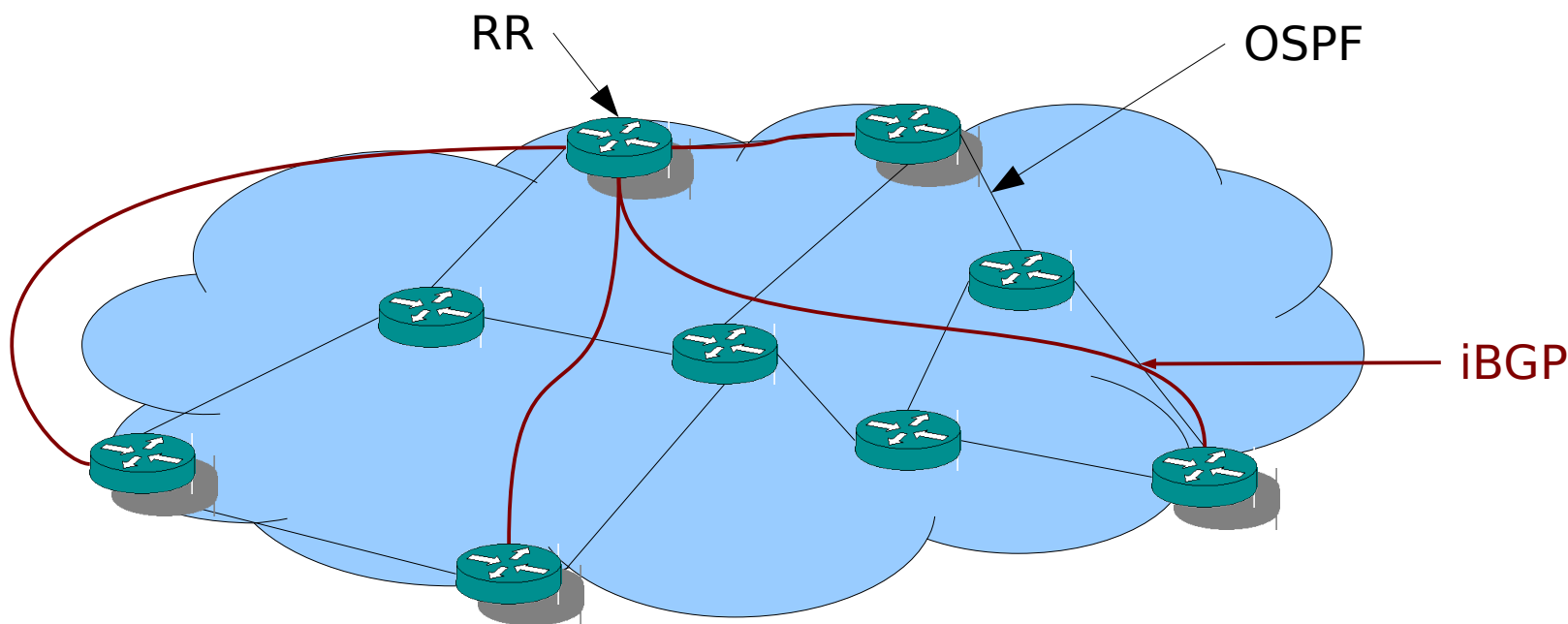
■ Internal BGP

- Se emplea para comunicar información BGP entre encaminadores de un mismo Sistema Autónomo.
 - Podría emplearse cualquier IGP, pero se pierde información de atributos.
- Se exige, en principio, que todos los encaminadores frontera de AS estén comunicados entre sí (full-mesh).



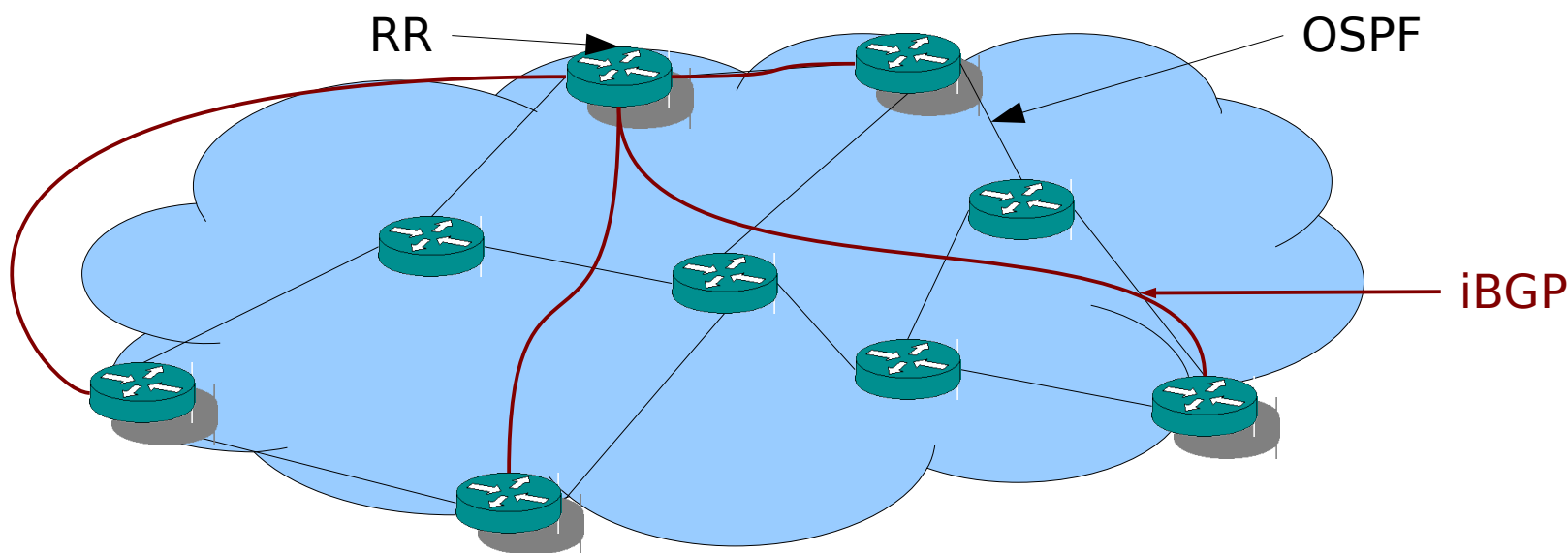
■ Internal BGP

- El número de conexiones entre encaminadores frontera puede crecer de manera cuadrática.
 - Se utilizan mecanismos para reducir el número de conexiones activas:
 - Reflectores de rutas.
 - Confederaciones.



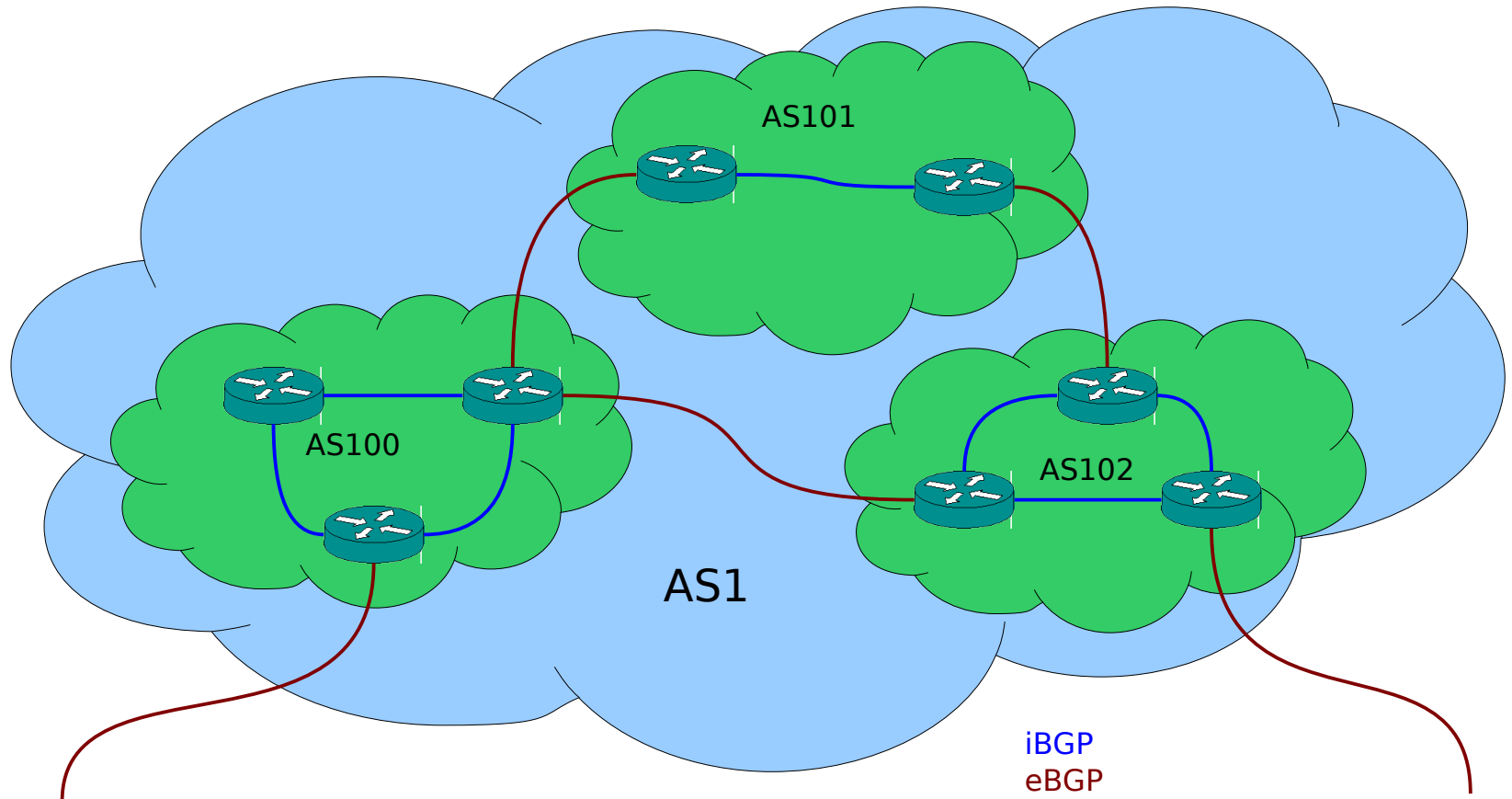
■ Reflectores de rutas

- Cuando un encaminador BGP recibe una actualización mediante iBGP, sólo la reenvía a través de eBGP.
- RR está configurado como Router Reflector.
- El resto de encaminadores BGP dentro del AS están configurados como clientes del RR.
 - Cuando un encaminador recibe una actualización a través de eBGP, la reenvía al RR, y éste al resto de clientes.



■ Confederaciones

- Un AS se puede dividir en un conjunto de AS.
- De cara al exterior, se comporta como un solo AS.

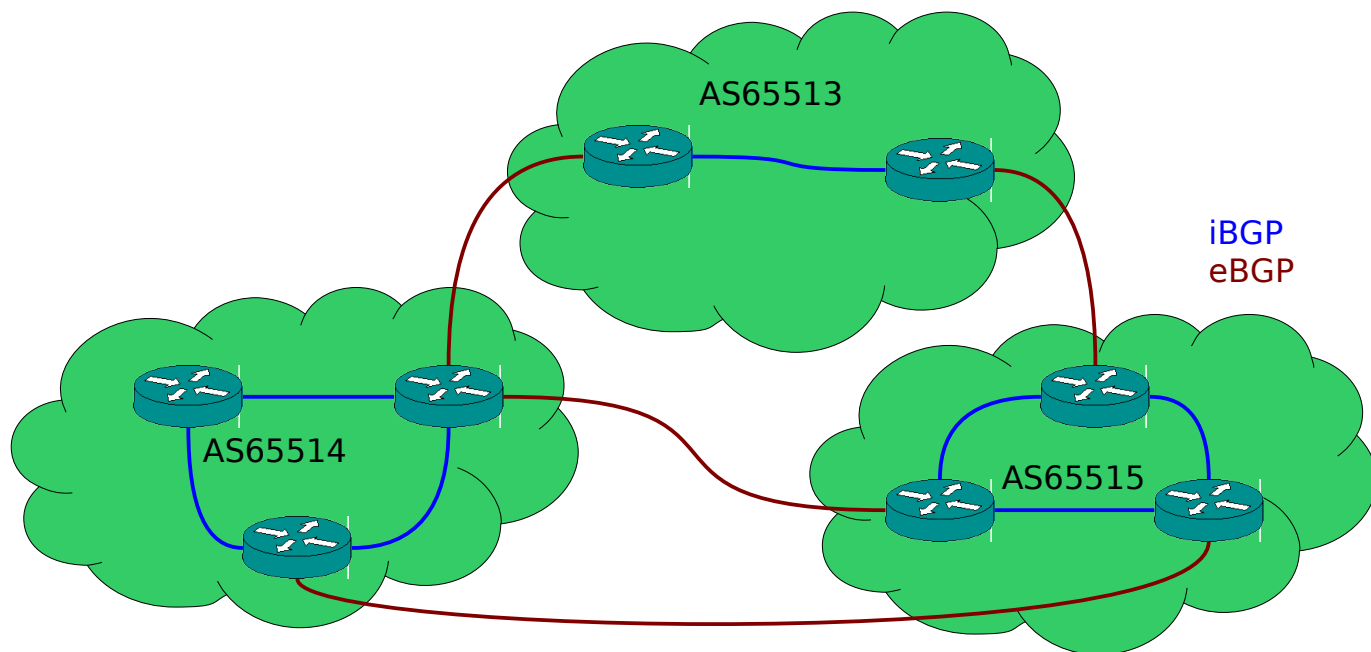


■ Confederaciones: RedIris (AS766) (<http://www.rediris.es>)



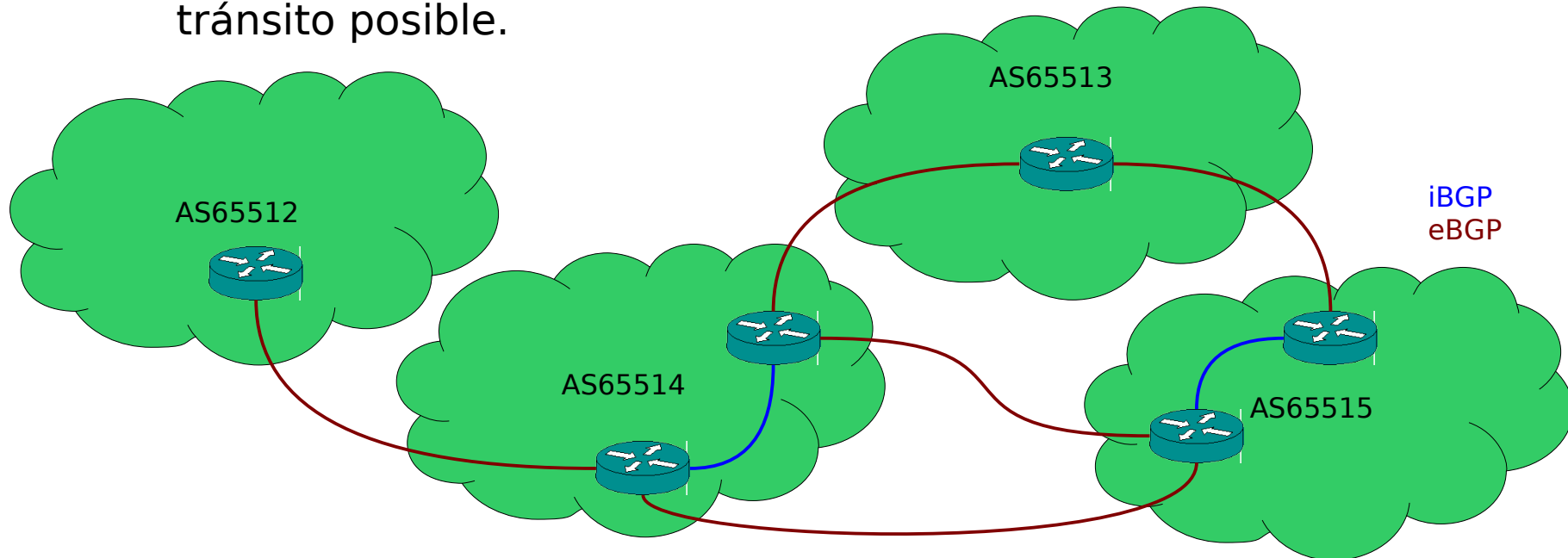
■ Práctica

- Configurar BGP (iBGP y eBGP)
 - Prefijos anunciados:
 - AS65513: 172.16.13.0/24, 2001:db8:13::/48
 - AS65514: 172.16.14.0/24, 2001:db8:14::/48
 - AS65515: 172.16.15.0/24, 2001:db8:15::/48
 - De los enlaces entre AS65514 y AS65515, uno debe ser el principal y otro el de respaldo.



■ Práctica

- Configurar BGP (iBGP y eBGP)
 - Prefijos anunciados:
 - AS65512: 172.16.12.0/24, 2001:db8:13::/48
 - AS65513: 172.16.13.0/24, 2001:db8:13::/48
 - AS65514: 172.16.14.0/24, 2001:db8:14::/48
 - AS65515: 172.16.15.0/24, 2001:db8:15::/48
 - AS65514 debe intentar que le atraviese el menor tráfico de tránsito posible.



BGP



Práctica

BGP

■ Práctica: BGP+OSPF

- Configurar como principales los enlaces marcados.

