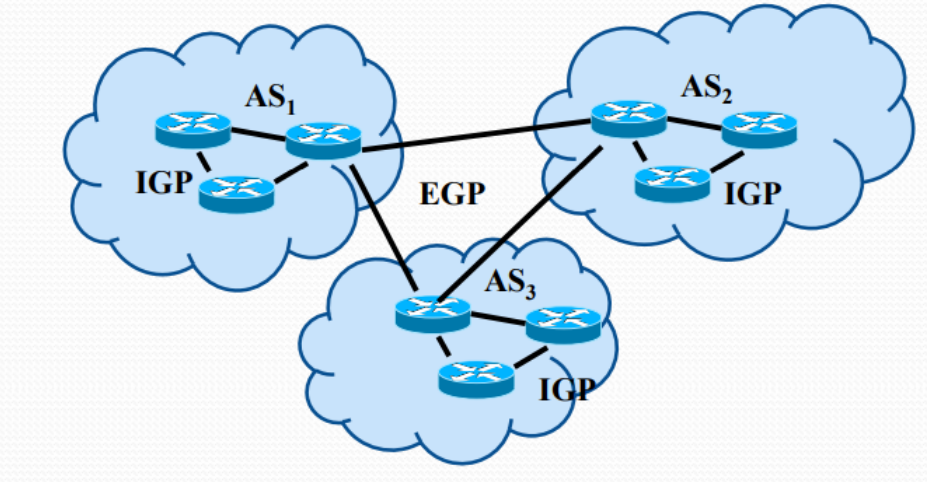
TEMA 4 – Inter-domain routing

Objetivos

1. Introducir conceptos básicos de inter-domain routing
2. Entender atributos BGP
3. Entender relaciones Peer-to-peer entre ISPs
4. Aprender técnicas multi-homing

Conceptos importantes

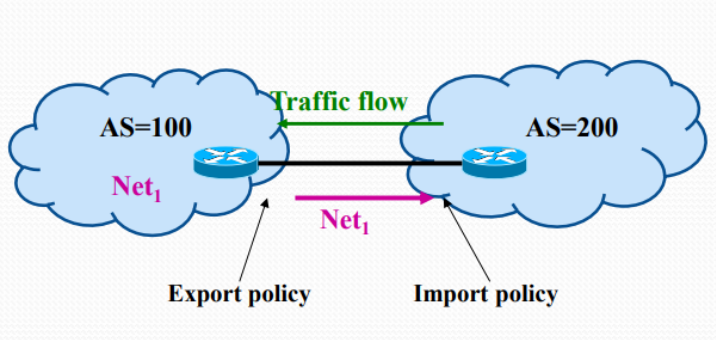
- **Sistemas autónomos** (AS) → Conjunto de routers con la misma política de enrutamiento en un único dominio administrativo. Se identifican con con un número (ASNs) de 16 bits (65535 AS's). AS's >= 65512 → Privados



- **BGPv4** → Es un protocolo de enrutamiento basado en políticas. **NO** usa **routing metrics** (con son hops, ancho de banda, delay…). Utiliza **routing atributes** que permiten definir políticas de enrutamiento. Se encapsula en paquetes TCP (lo que significa que entre dos routers BGP, debe hacer conexión TCP en ambas direcciones)
- **ISP** → Un ISP es una entidad administrativa que puede tener uno o más ASNs asignados dependiendo su arquitectura y situación geográfica. En general, los ASNs son asignados a un ISP o a una red corporativa. → No todos los ASs son ISP, pero ISPs tienen ASs asignados.
- **Tipos de operaciones de ASs** → Hay de dos tipos:
    - **Single-homed** → AS's que llegan a routers de otros As's usando un solo punto de conexión.
    - **Multi-homed** → As's que llegan a routers de otros As's usando más de un punto de conexión, pero, pueden no transitar routers de otros AS (**Multi-homed non-transit**) o si pueden (**Multi-homed transit**)

BGPv4 (protocolo de enrutamiento)

- Anuncia rutas y subredes usando políticas de enrutamiento a otros AS's.
- **Política de enrutamiento** → Suponiendo que una subred pertenece a un AS, significa la decisión de un AS de anunciar esa ruta a otro AS ("**Export Policy**") y la decisión del otro AS de aceptar la ruta que le envían. ("**Import Policy**")



- **Paquetes BGP**:

  Los routers BGP envían paquetes encapsulados en segmentos TCP. Los diferentes tipos son:

  - **OPEN** → Establecer conexiones BGP
  - **KEEP ALIVE** → Testea si la conexión TCP sigue funcionando
  - **UPDATE** → Cuando hay una modificación de una ruta o se ha encontrado una mejor
  - **NOTIFICATION** → En caso de error, envía una notificación.

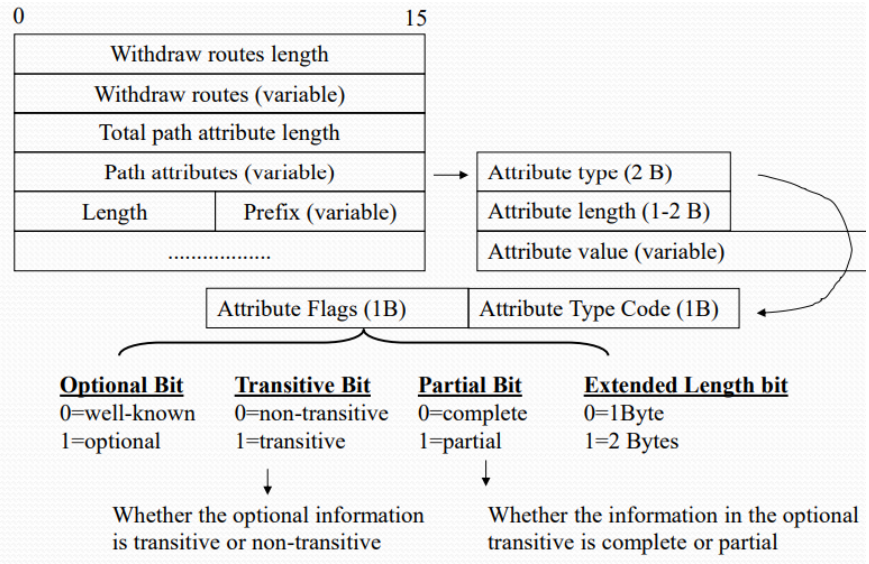
- **Atributos BGP:**

  Los mensajes de UPDATE son los encargados de llevar los atributos que indican las políticas de enrutamiento que BGP debe tener. Puede ser de diferentes tipos:

  - **Well-known** → Atributos deben ser soportados por todas las implementaciones de BGP.
  - **Optional** → Atributos no necesariamente deben ser soportados.
  - **Mandatory** → Atributos siempre son enviados en mensajes UPDATE.
  - **Dicretional** → No se envían con mensajes UPDATE.
  - **Transitive** / **Non-transitive** → Routers transitan o no otros routers.
  - **Complete** → Se utilizan si todos los routers que transitan un atributo implementan atributos opcionales. (…)

  Las combinaciones entre ellos son posibles, pero solo las siguientes; Well-known y Mandatory / Well-known y discretional / Optional y transitive / Optional y Non-transitive.

- **Estructura Update messages** (no lo pregunta)



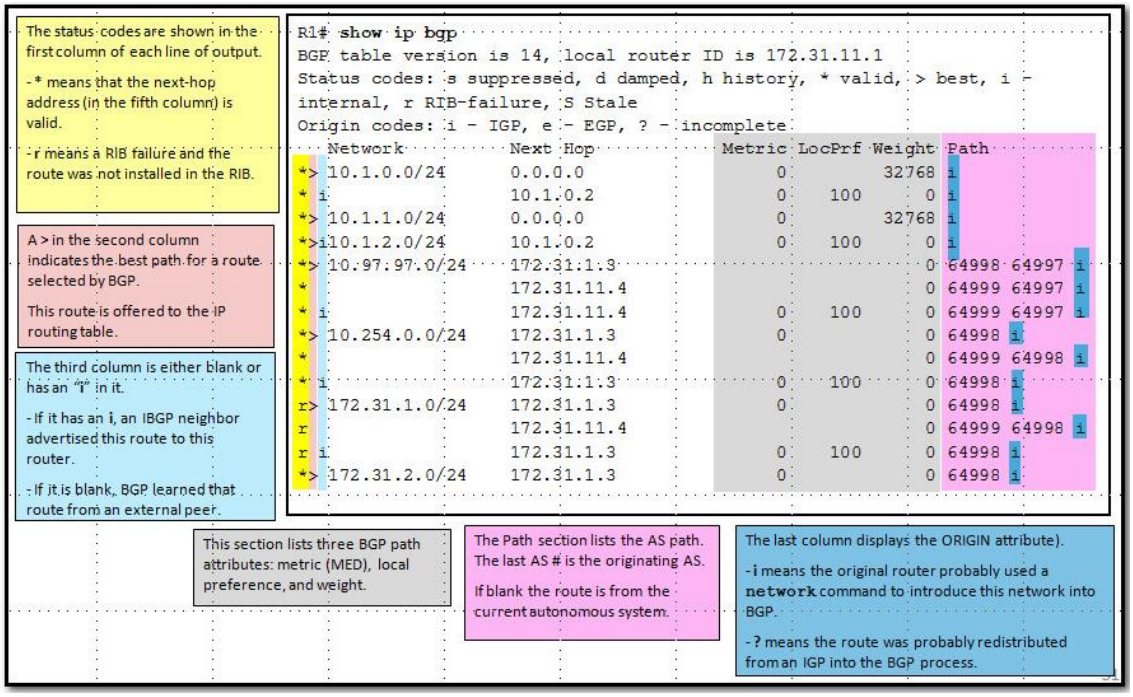
- **Tabla de enrutamiento BGP**

  Incluye: **subnet y mascara**, **next-hop**, **métrica**, **Local_pref**, **AS-path-vector** y **origin**.

  Un AS puede estar conectado con N AS's, por lo tanto, recibirá un mensaje UPDATE con una posible ruta para cada conexión BGP. → Lo que significa N entradas por router.

  Entonces, se usa el **decision process** que escoge la mejor entrada del router en función de los atributos. Esta decisión depende de: (se ve más adelante)

  - Implementación del fabricante
  - Mantiene una base de datos por sesión activa de BGP
  - El símbolo **">"** que indica la mejor ruta hacia un router
  - Un router solo anuncia su mejor ruta con mensajes UPDATE BGP

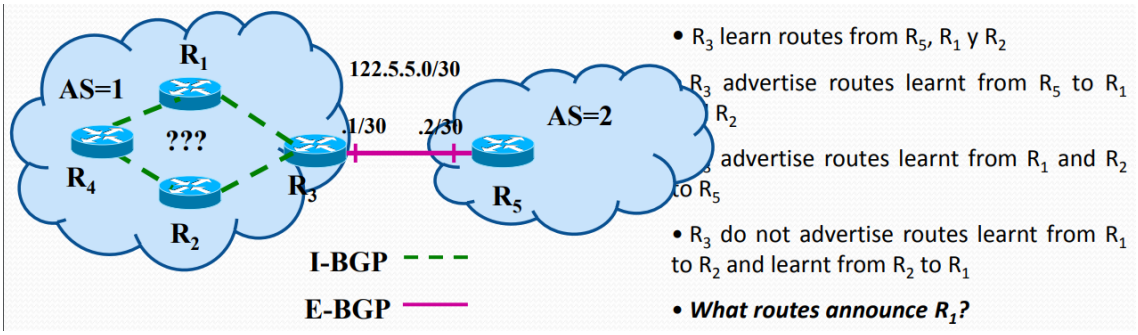


- Los routers BGP intercambian rutas. Cada router tiene una lista de atributos que permiten a otros routers BGP fijar una política con respecto a ese router.
- **Sesiones BGP** → Dos routers abren una sesión TCP con el puerto 179. Se le llaman **vecinos o peers**.
- Hay que diferenciar…
    - Routers BGP que pertecen a la misma AS → **Internal BGP (I-BGP)**
    - Routers BGP que pertenecen a diferentes AS's → **External BGP (E-BGP)**
    - ¡SOLO HAY **UN** PROTOCOLO BGP! Solo que trabajan diferente.
- **AS-PATH-VECTOR** (de las tablas BGP)

  Representan la ruta que se sigue desde el AS origen. → Es decir, da todos los AS's que atraviesa. BGP siempre prefiere la ruta más corta (en saltos de AS's)

  Cada AS, añade su ASN al path → EJ. (3,2,1) donde 1 es el AS destino.
- **Internal BGP** (I-BGP)

  Se usa para **coordinar la política de enrutamiento dentro de un AS**, además es necesario para permitir tránsito a rutas externas a través del AS.
    - Subredes aprendidas vía E-BGP puede ser anunciadas vía E-BGP y I-BGP.
    - Subredes aprendidas vía I-BGP solo pueden ser anunciadas vía E-BGP.
    - Router I-BGP NO ANUNCIAN rutas vía i-BGP.



- ¿Por qué los routers I-BGP no anuncian rutas aprendidas vía I-BGP? Ya tienen el atributo AS-path-vector, que sirve para detectar loops. → Si hay un loop, no anuncian la ruta.
- **Loopback dummy interface**

  En caso de que haya una sesión I-BGP entre dos routers y el enlace falle, la **sesión I-BGP se pierde.** → Truco: usar una loopback interface con direcciones IP públicas o privadas diferentes de 127.0.0.0/8.

  + En el momento de configurar el protocolo interno de un AS, hay que tener en cuenta que debe configurarse antes OSPF (en caso de que se escoja este protocolo) que BGP para que queden bien sincronizados. Y también es importante, que las interfaces conectadas a otros AS's no anuncien las redes internas por OSPF. → Para ello, se usa el comando **"passive interface interfaceName".** (usado en la práctica)

- **Next-hop**

  Para una sesión BGP, es la @IP del router BGP el que anuncia la ruta. Las sesiones I-BGP NO pueden cambiar el next-hop con una E-BGP UPDATE message.

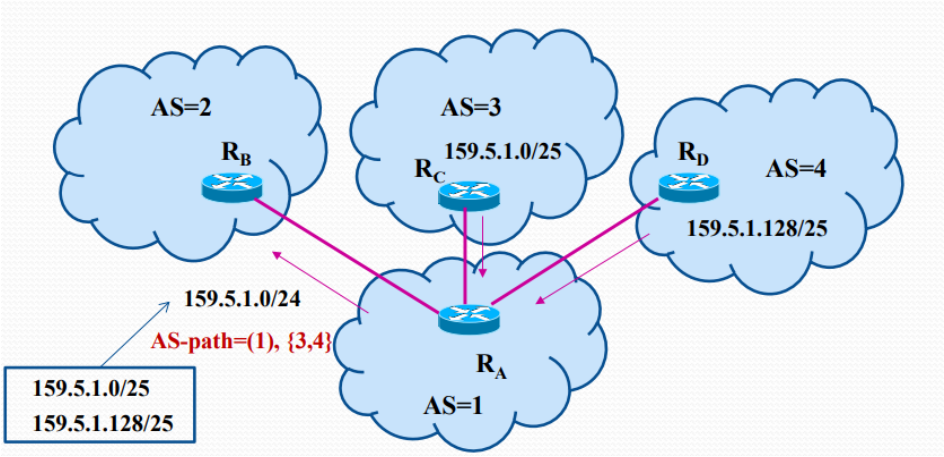
- **Origin** → Indica quien originó la ruta. Pueden ser 3 tipos:

  - **IGP.** La ruta fue originada por un mecanismo interno y se indica con el carácter "i" en la tabla de enrutamiento BGP. En CISCO, se activa usando el comando "network".
  - **EGP.** La ruta fue originada por el protocolo EGP desde un AS externo y se indica con "e" en la tabla de enrutamiento. No se usa actualmente, está obsoleto.
  - **Incomplete.** No se sabe el origen. Y se indica con el carácter "?" en la tabla de enrutamiento.

- **AGGREGATOR** (Solo tipo Optional y transitivo)

  Es un atributo. Un mensaje UPDATE BGPv4 envía una subred/máscara que puede agregarse. El router BGP que agrega, puede indicar en el AS-PATH-VECTOR la partición de la subred agregada. (opción AS-SET).

  No tiene influencia en la selección del path.



- **ATOMIC AGGREGATE** (Solo tipo well-known y discretional)

  El propósito de este atributo es alertar a los BGP speakers a lo largo de la ruta que se ha perdido cierta información debido al proceso de agregación de rutas y que la ruta agregada podría no ser la mejor ruta al destino.

  Si al agregar, AS-SET no se ha activado, entonces el AS-PATH-VECTOR puede perder información de los paths originales antes de la agregación. → Es obligatorio entonces, que el Atomic Aggregate esté activo.

- **LOCAL-PREFERENCE** (Solo tipo well-known y discretional)

  Atributo que indica el enlace de salida preferido. Los valores más altos de Loc-Prf tienen mayor preferencia. (valor por default = 100)

- **Route Maps**

  Los route maps se usan en BGP para controlar y modificar la información de la tabla de encaminamiento BGP, para definir las condiciones por las cuales una ruta es distribuida entre dos routers y para modificar los atributos incluidos en los mensajes BGP.

  Se usa el siguiente comando:

  route-map map-tag [**permit | deny**] [seq-number]
  match: comando que especifica el criterio que debe ser comprobado
  set: comando que indica la acción a ejecutar si el match aplica

  Que es parecido a un if-else. (Hay un ejemplo de uso de route maps en el **dossier de lab**)

- **MED (Multi-Exit-Discriminator) (Solo si son optional y non-transitive)**

  ¡¡¡También llamado **métrica**!!! (ver tabla de enrutamiento BGP), indica a los vecinos cuál es el enlace de entrada preferido. → Siempre el **menor valor**.

- **Comunidades**

  Ofrece la posibilidad de asociar un identificador con una ruta. Permite que esta ruta reciba la misma política que todos los AS's asociados a esa política.

  Codificado con 32 bits (4 bytes) como:

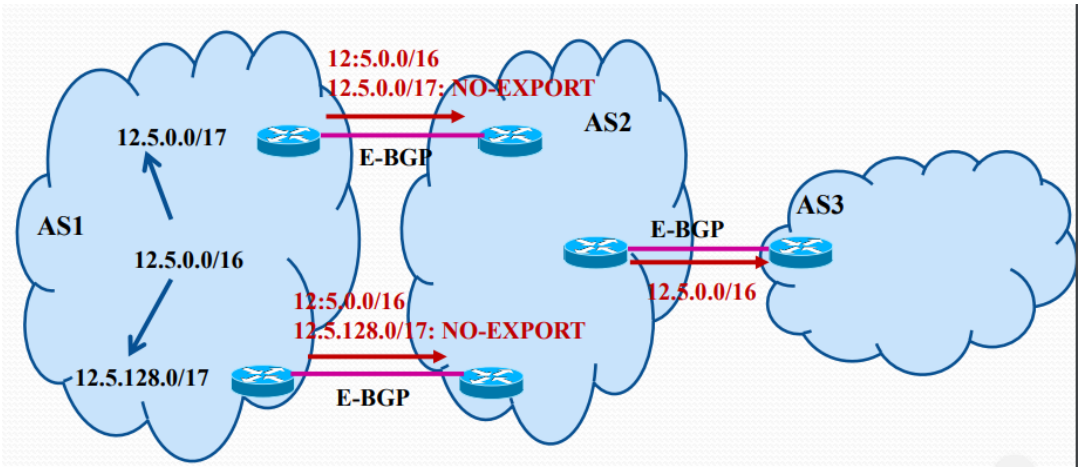
  AS: value (decimal)

  2 bytes (AS) → AS que crea la comunidad

  2 bytes (value) → Definido por el AS

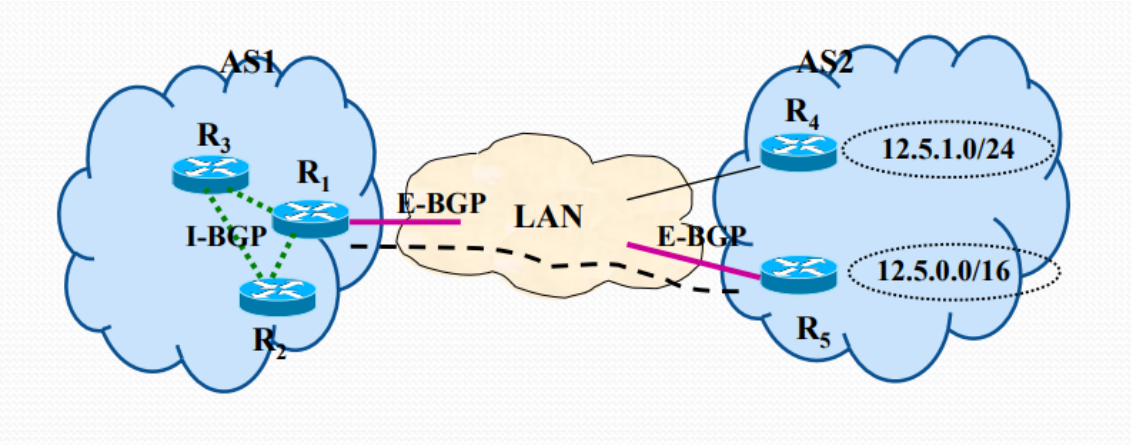
  Las comunidades estándar son:

  - **NO_EXPORT (0xFFFFFF01)** → Todas las rutas recibidas con este atributo NO DEBEN publicarse fuera del AS.



  AS1 exporta a AS2 la /17 con NO-EXPORT community y /16 sin community (porque no puede publicar las /17).

- **NO_ADVERTISE (0xFFFFFF02)** → Todas las rutas recibidas con este atributo NO DEBEN publicarse a otros vecino BGP (dentro del mismo AS)



Tanto R3 como R2, no entienden la política en AS1, solo R1 la entiende.

- **NO_EXPORT_SUBCONFED (0xFFFFFF03)** → Todas las rutas recibidas con este atributo NO DEBEN publicarse en routers BGP externos. (¡otra confederación, visto más adelante!)

- **Decision Process** → ¿Cómo se escoge la mejor ruta? (Best Path Selection)

  Depende de la implementación. Por ejemplo, en los routers CISCO:

  1. Highest WEIGHT
  2. Highest LOCAL_PRF
  3. Lowest AS-PATH
  4. Lowest origin type
  5. Lowest MED (métrica)
  6. Prefer eBGP que iBGP
  7. When both paths are external, prefer the path that was received first
  8. Lowest ROUTER-ID
  9. Lowest interface @IP

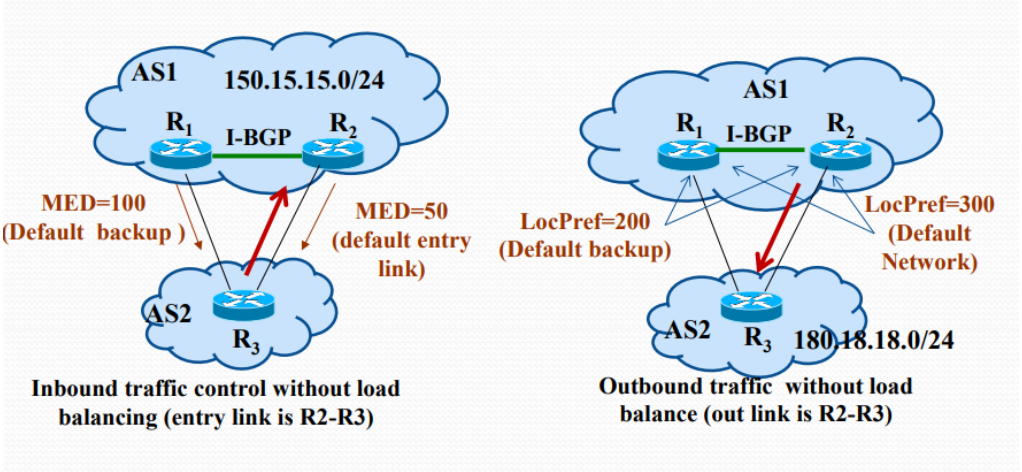
- **Multi-homing**

  (visto más arriba) **Single-homed** o **Multi-homed** entre cliente y ISPs

  Multi-homing incrementa la confiablidad de acceso, ya que, si un enlace falla, el cliente tiene un back-up (debido a que tiene una conexión con uno o más ISPs)

  Balanceo de carga (load balancing) → Equilibrio de tráfico entre los enlaces permitiendo el control del tráfico entrante (**inbound**) y el control del tráfico saliente (**outbound**).

Ejemplo1: (**sin balanceo de carga**)

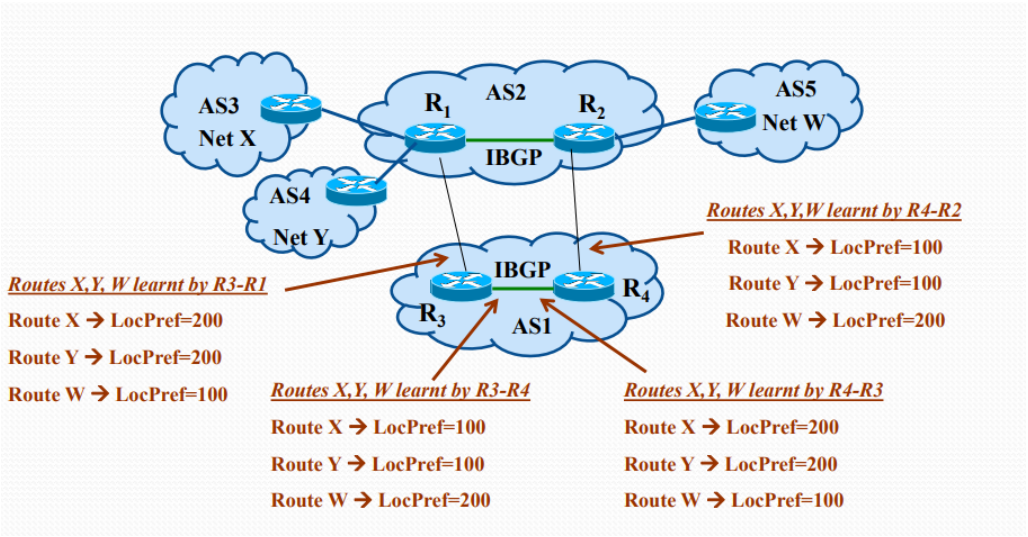


Primer caso: AS1 usa BGP y **exporta** rutas con diferentes atributos MED para forzar AS2 a usar el link de entrada R2-R3. → **Lowest MED**

Segundo caso: AS1 usa BGP y **importa** rutas teniendo en cuenta el atributo Local-Pref para seleccionar el link de salida R2-R3. → **Highest Local-pref**

Acordarse ISP = AS.

Ejemplo2: (**con balanceo de carga visto con outbound traffic**)



El tráfico hacia los clientes AS3/AS4 sale usando R3 y el tráfico hacia AS5 sale usando R4 → **Lowest MED**

- **Back-up automático en multi-homing**

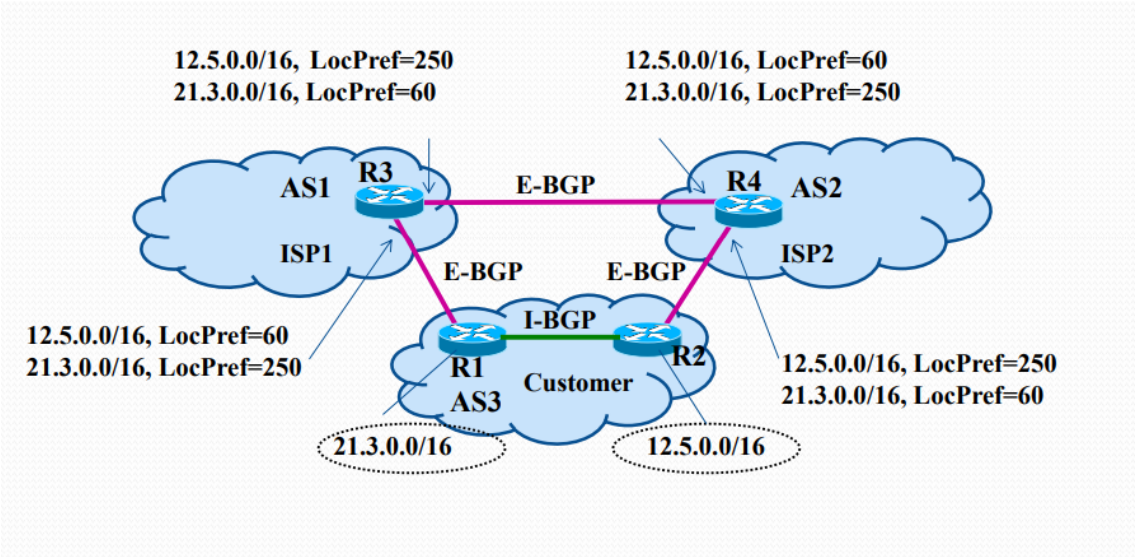
  En caso de fallo, el cliente en multi-homing siempre tiene una línea de back-up.

  Ejemplo práctico:

  AS3 (cliente) quiere muti-homing con ISP1 y ISP2. El tráfico hacia 12.5.0.0/16 entra por ISP2 (back-up = 21.3.0.0/16) y el tráfico hacia 21.3.0.0/16 entra por ISP1 (back-up = 12.5.0.0/16). Ambas conexiones actúan como backup respecto a la otra red.

¿Cómo conseguimos lo anterior? (valores de localPref y value pueden ser otros)

- AS1 y AS2 reaccionan a la **comunidad 3:20 activando LocalPref = 60**
- AS1 y AS2 reaccionan a la **comunidad 3:70 activando LocalPref = 250**



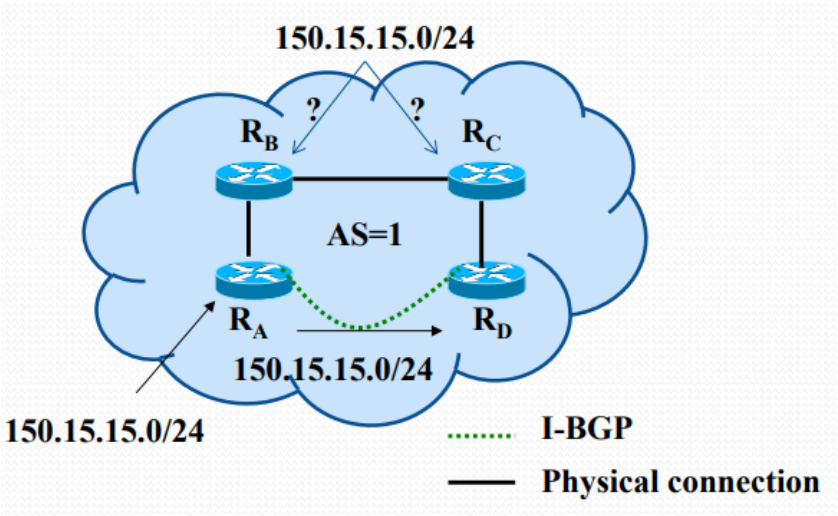
La ruta con LocPref menor será la que quede como back-up, mientras que con mayor será la buena.

(mirarse **configuración communities** en el **dossier de lab y power**)

- **Sincronización BGP**

  Antes de poder propagar una ruta interna a otro AS es necesario que el protocolo de enrutamiento IGP esté sincronizado con BGP (que es EGP). La sincronización se asegura que, si el tráfico se envía al AS, el protocolo IGP conocerá su destino.

  Las rutas recibidas por Ra se envían a través de iBGP a Rd. Sin embargo, los routers Rb y Rc no conocen la ruta 150.15.15.0/24. Entonces, se dice que Rb y Rc **no están sincronizados**. → Solución: redistribuir BGP con protocolos IGP o crear una red IBGP con malla completa.



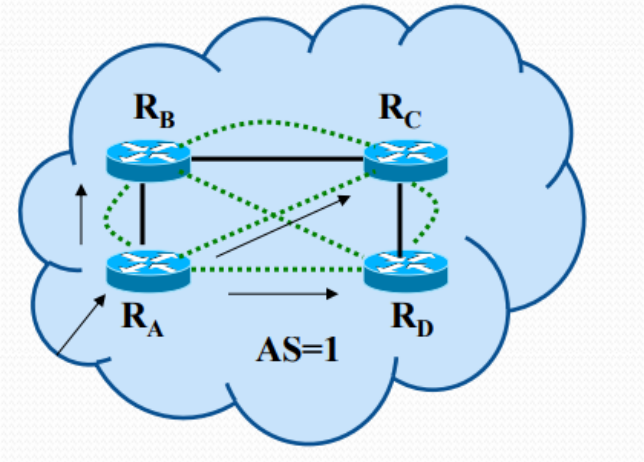
- **Escalabilidad BGP**

  Regla de Split-horizon: (horizonte dividido)

  Una ruta aprendida por I-BGP no seran propagadas nunca a otros routers I-BGP. Al igual que otros protocolos de routing es necesario para evitar bucles de red.

  Como resultado, es necesario una **red I-BGP con malla completa**. Cumple la siguiente formula, dado N routers:

  N*(N-1)/2 conexiones/sesiones I-BGP.



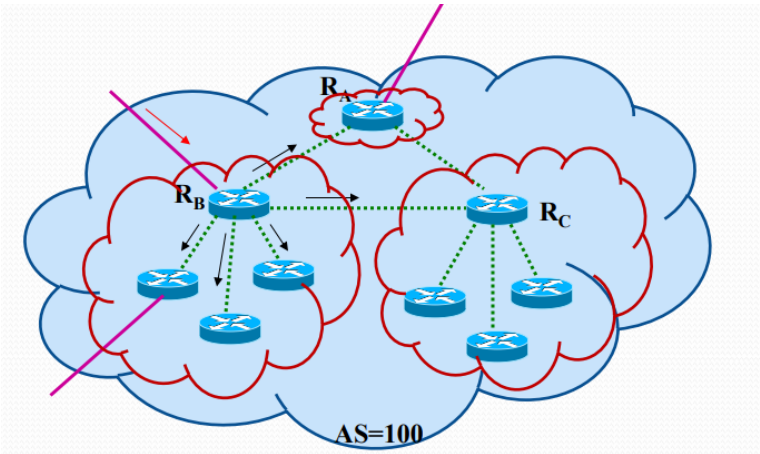
  Soluciones: (para reducir el número de sesiones I-BGP que utiliza una red de malla completa)

  Los route reflectors o confederaciones.

- **Route Reflectors**

  Se modifica la regla Split-Horizon para que el route reflector pueda propagar las rutas aprendidas por conexiones I-BGP a sus vecinos a través del mismo AS.

  Son necesarios los **cluster**. Usados para definir una red, son la unión de un route reflector y sus clientes. El route reflector actúa como **cabeza de cluster**. Deben formar una red de malla entre ellos, pero los clientes no necesitan formar una malla.



  3*(3-1)/2 = 3 conexiones i-bgp (las que hay entre clusters)

+ Cuando una ruta Reflector reenvía actualizaciones se activa el **atributo Originator-ID**. Si el Route Reflector vuelve a recibir una actualización con su Originator-ID, la descartará, así evitar bucles. Lo mismo pasa con el **atributo Cluster-ID**, cuando existen múltiples Route Reflector.

En concreto el router RR sigue estas reglas al recibir un mensaje BGP:

- Si el mensaje BGP proviene de un vecino no cliente (por ejemplo otro RR), entonces el RR la refleja a todos sus clientes dentro de su cluster.
- Si el mensaje BGP proviene de un cliente, el RR la refleja a todos los vecinos clientes y no clientes.
- Si el mensaje BGP se aprende de un vecino eBGP, éste se envía a todos los vecinos clientes y no clientes.

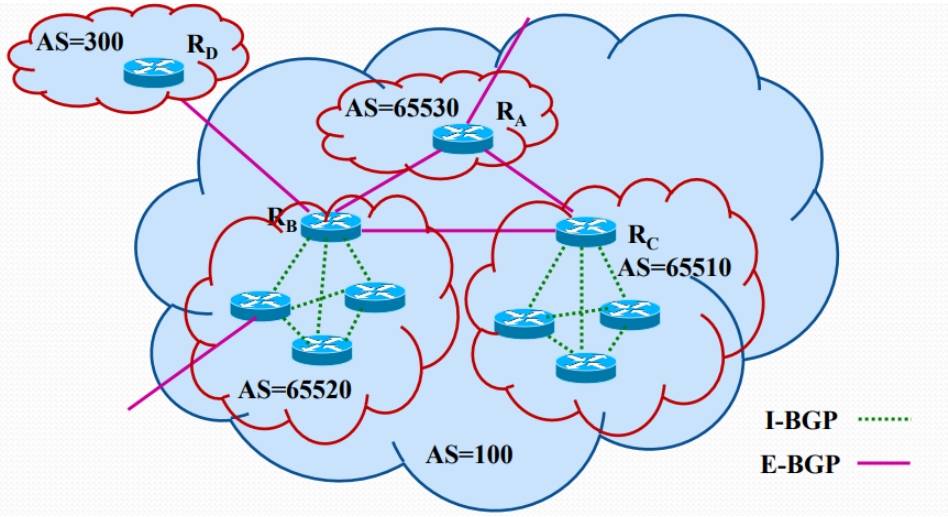
(mirar configuración para la práctica)

- **Confederaciones BGP**

Una confederación es otra de las soluciones posibles para reducir el número de sesiones I-bgp en un AS.

Se basa en crear mini-ASs usando ASNs privados dentro de un AS. Cada uno de estos mini-ASs están formados por una y necesita sesiones e-bgp con otros mini-AS.

Pero desde el punto de vista "externo", es un solo AS, solo que dentro hay varios mini-ASs.



(mirar configuración de confederaciones para la práctica)

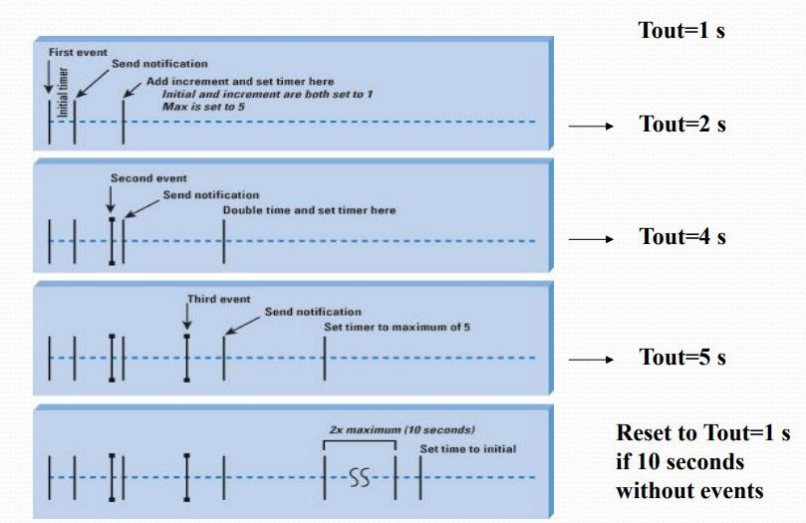
- **Convergencia BGP** → Es el momento en que TODOS los routers de una red alcanzan a un conocimiento común entre ellos y la red es operativa para enviar paquetes.

  Tenemos el problema del **flapping.** Un estado cambia constantemente de un estado a otro (es decir, up o down), lo **que provoca una actualización de los mensajes** bastante a menudo → una baja convergencia de la red, bucles, fallos de red y colapso.

  Solución (para reducir flapping):

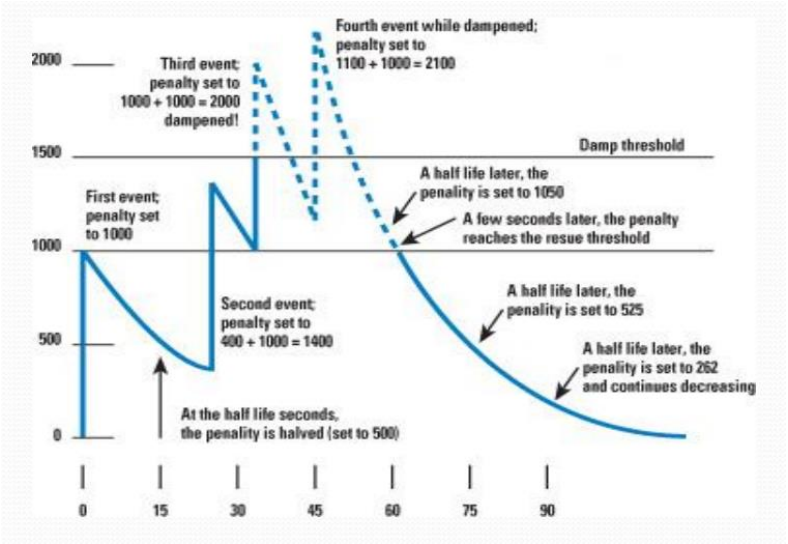
  Slow-down techniques → Reducir la frecuencia con la que los mensajes UPDATE se envían a otros routers BGP. Cuantos más cambios, menos frecuencia. Hay dos tipos:

  - **Exponential back-off** → Consiste en ralentizar la notificación de mensajes



  - **Dampening** → Consiste en no reportar un evento si este ocurre con frecuencia.

    Cada vez que ocurre un evento, un contador se incrementa por un valor de penalización. Después de un rato sin ocurrir ningún evento, el contador se va decrementando. Si el contador llega a **"damp threshold"** → Entra en **estado "DAMPENED".** (state down) Si el contador alcanza el **"reuse threshold"** pasa a estado up.



  (Mirar cómo usarlo con route-maps y ejercicios BGP)