TOPIC 4: Inter-domain Routing (BGP)

Pregunta 1. Explica qué es una política de encaminamiento y cómo se implementa.

Una política de encaminamiento es la decisión de un AS de anunciar una red que éste contiene (política exterior) y de que otro AS acepte la red en su dominio.

Estas políticas se implementan mediante un protocolo llamado BGPv4. Éste funciona encapsulado en TCP, por lo tanto, las comunicaciones se establecen mediante BGP:

- OPEN: crear una conexión BGP.
- KEEPALIVE: comprueba que la conexión BGP siga activa (sirve para mantener la conexión TCP, ya que esta tiene un tiempo de vida si no se usa).
- UPDATE: envía rutas y atributos (información de encaminamiento).
- NOTIFICATION: Notificar que ha habido un error.

Pregunta 2. Explica cómo escala la tabla de encaminamiento BGP en función de la cantidad de AS's a los que está conectado un AS.

Tamaño tabla BGP <= N*m = (m/2)*Ni, i=0,1,... <= N*m, donde m=#AS, N =redes en Internet (si no se filtran las entradas).

Pregunta 3. ¿Pará qué sirve definir una dirección de loopback en un router? ¿Qué tipo de dirección es?

- Robusto a fallos a interfaces físicas.
- Tiene utilidad entre sesiones IBGP
- Permite mantener sesiones IBGP mientras haya rutas físicas (OSPF).

Se puede usar cualquier dirección IP (tanto pública como privada).

Pregunta 4. ¿Cómo resuelve BGP el problema de los bucles?

I-BGP: Se hace una malla completa (full mesh) con todos los routers BGP del dominio, de esta forma la retransmisión de paquetes I-BGP se hace en los ASBR.

E-BGP: En el AS_PATH se ponen los AS por los que el paquete E-BGP ha pasado. Si un AS detecta que en el AS_PATH vector ya aparece su id deja de transmitir el paquete. EBGP es quien actualiza el AS_PATH, si fuera IBGP entonces no sería posible detectarlo.

Pregunta 5. ¿Qué diferencia hay entre IBGP y EBGP?

E-BGP se usa para intercambiar rutas entre routers de diferentes dominios, en cambio, I-BGP se usa para routers de un mismo dominio.

Pregunta 6. ¿Qué diferencia hay entre las redes que anuncia OSPF y las que anuncia BGP (e.g. con el comando network)?

En OSPF las redes que se anuncian, tanto públicas como privadas, se quedan siempre dentro de un mismo AS, en cambio, en BGP, se anuncian redes a otros AS para que estas sean accesibles desde estos por lo tanto solo se deben anunciar redes públicas.

Pregunta 7. Explica la diferencia entre un atributo BGP conocido ("well-known") y otro opcional. Idem si el atributo es mandatorio y discrecional. Menciona algún atributo que tenga la característica de ser conocido y discrecional, otro que sea conocido y mandatorio y otro que sea opcional y transitivo.

Los atributos "well-known" són obligatorios para los AS implementarlos, mientras que los opcionales no.

Los atributos mandatorios son obligatorios usarlos cuando se transportan mensajes BGP, mientras que los discrecionales no.

- Conocido y discrecional: ATOMIC AGGREGATE
- Conocido y mandatorio: AS-PATH
- Opcional y transitivo: COMMUNITY

Pregunta 8. ¿Qué significa que en una tabla BGP aparezca el atributo ORIGEN como incompleto? ¿Qué acción ha ejecutado el administrador del sistema para que aparezca como incompleto? ¿Qué efectos tiene dicha acción?

Quiere decir que las redes que hemos recibido, se han redistribuido mediante un protocolo interno del AS, como por ejemplo OSPF, RIP o IS-IS.

Al programar el ASBR no se ha puesto el passive interface.

Por lo tanto, estas anunciando todas tus redes internas a Internet, tanto públicas como privadas.

Pregunta 9. ¿Qué relación hay entre los atributos ATOMIC AGGREGATE y AGGREGATOR?

AGGREGATOR sirve para agregar direcciones IP, y ATOMIC AGGREGATE sirve para informar que en esa ruta hay direcciones agregadas. Aún que, AGGREGATOR, ya incluye una opción, que se llama AS-SET, para indicar que AS's se han agregado. Al ser opcional el AS-SET, si no se incluye es mandatorio activar el ATOMIC AGGREGATE.

Pregunta 10. Qué diferencia hay entre una política BGP inbound y una outbound. ¿Qué atributo BGP te permite generar una política outbound?

Una política BGP inbound sirve para decidir por dónde quieres que entre el tráfico, y una política BGP outbound sirve para decidir por dónde quieres que salga el tráfico.

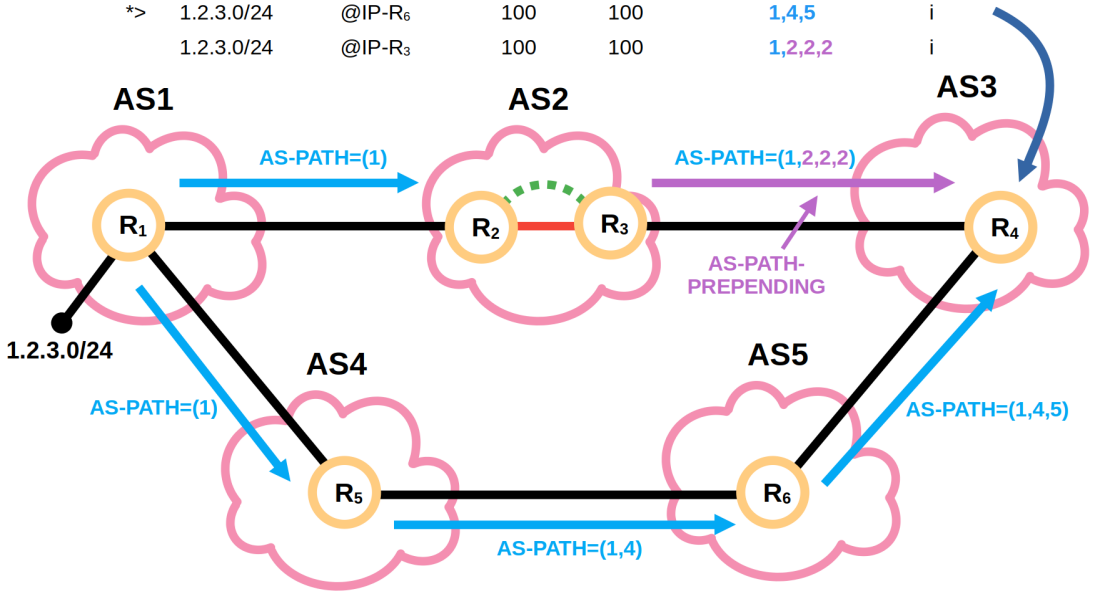
Un atributo outbound es el Local-Preference.

Pregunta 11. a) ¿Qué es una política de "AS-path-prepending"?. Explica mediante un ejemplo sencillo como un ISP puede usar esta política. **b)** ¿Qué atributo BGP permite definir a un ISP una política de tráfico de tipo "outbound"? Explica mediante un ejemplo sencillo como un ISP puede usar esta política.

a) El "AS-path-prepending" es cuando inflas el AS-PATH vector de un mensaje BGP con tu propio número de AS para que así, un router de otro AS elija otro AS para enviar mensajes. Hay que tener en cuenta que no afecta al tiempo de transmisión real de paquetes, simplemente se hace para que otros AS al ver el número de saltos del AS-PATH escojan otro AS que tenga menos saltos en el AS-PATH. Por ejemplo:

Tabla BGP R_4

	Net/Mask	Next-Hop	MED	Local-Pref	AS-PATH	Origin
*>	1.2.3.0/24	@IP-R_6	100	100	1,4,5	i
	1.2.3.0/24	@IP-R_3	100	100	1,2,2,2	i

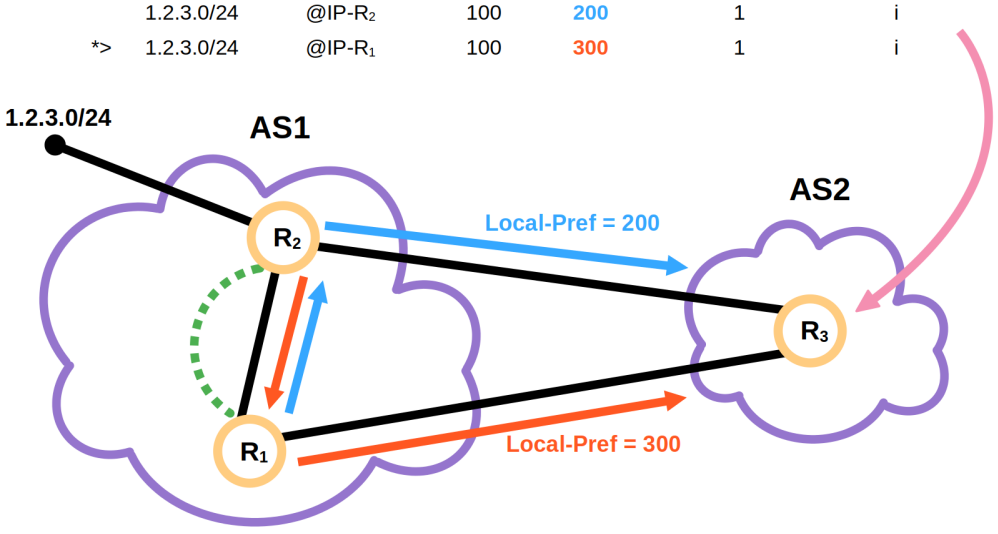


Podemos ver en la figura que el AS2 infla con su propio número de AS el AS-PATH vector, para que así, el AS3 elija el router R5 para enviar mensajes a la red 1.2.3.0/24 aún que en realidad haya más saltos reales en esa ruta.

b) El Local-Preference es un atributo BGP con política outbound que sirve para decidir por dónde quieres que entre el tráfico. Por ejemplo:

Tabla BGP R_3

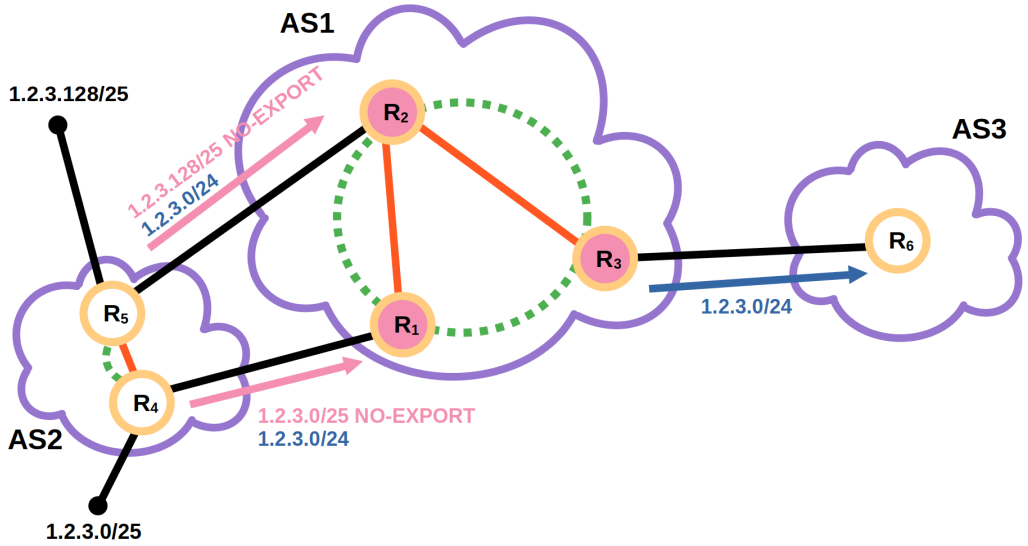
	Net/Mask	Next-Hop	MED	Local-Pref	AS-PATH	Origin
	1.2.3.0/24	@IP-R_2	100	200	1	i
*>	1.2.3.0/24	@IP-R_1	100	300	1	i



Si programamos los routers R2 y R1 de tal forma que ponemos un Local-Preference más alto (= 300) en la ruta más larga (R2,R1,R3); los mensajes, si no hay un fallo en la red, siempre pasarán por esta ruta, ya que el router R3 al enviar un mensaje a la red 1.2.3.0/24 al ver la tabla BGP verá que el Local-Preference-R1 = 300 > Local-Preference-R2 = 200.

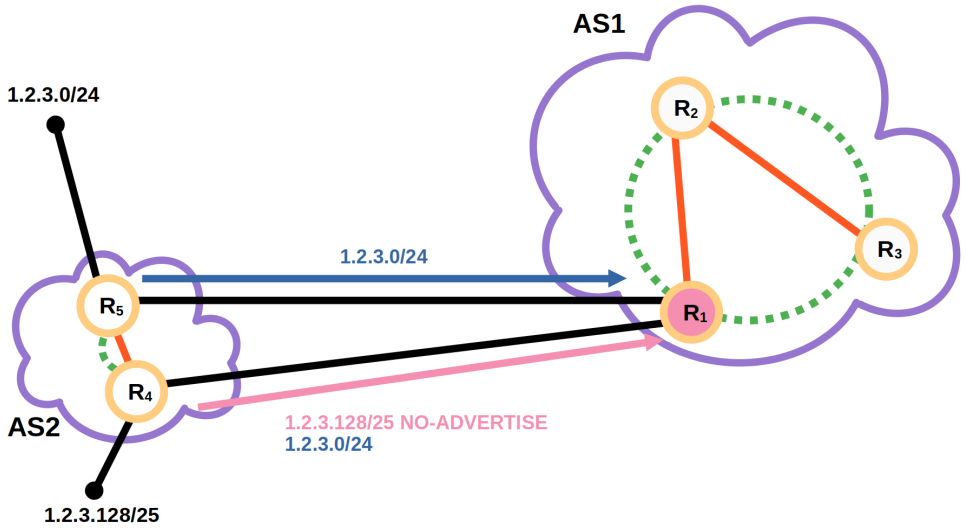
Pregunta 12. Explica la diferencia entre una comunidad "NO-EXPORT" y una comunidad "NO-ADVERTISE". Pon un ejemplo de uso de cada una de ellas.

La comunidad "NO-EXPORT" comparte el atributo con todos los routers del AS y no sale de este. Ejemplo:



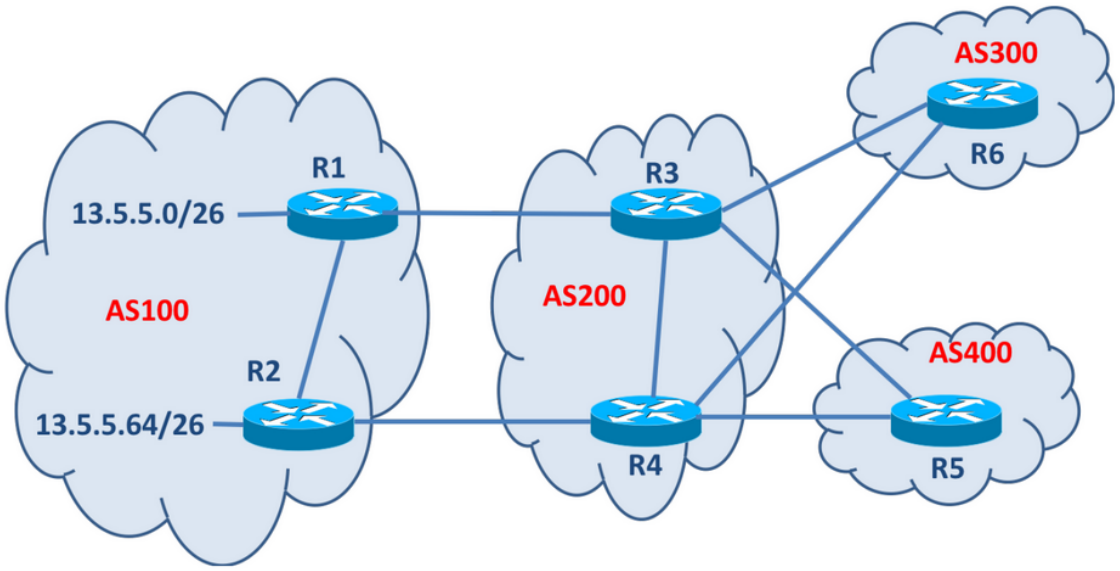
Cuando exportamos las redes 1.2.3.0/25 y 1.2.3.128/25 desde el AS2 hacia el AS1, queremos que sean visibles sólo para el AS1 y para el AS3 solamente sea visible la dirección 1.2.3.0/24. Si ponemos una comunidad "NO-EXPORT" sólamente los routers de AS1 sabrán diferenciar estas dos rutas, ya que si R3 quiere enviar un mensaje a 1.2.3.0/25 lo hará por R1 y si quiere enviar un mensaje a 1.2.3.128/25 lo hará por R2.

En cambio, el "NO-ADVERTISE" comparte el atributo sólo con el ASBR del AS y no se propaga por el AS. Ejemplo:



Si tenemos la red en 1.2.3.0/24 en AS2 y R3 quiere enviar un mensaje a 1.2.3.128/25, lo enviará por R1, y este al ver su tabla de encaminamiento verá que hay un atajo por R4 y lo enviará por allí.

Pregunta 13. El AS100 dispone de la red 13.5.5.0/24 que ha divido en 4 subredes /26. El AS100 quiere qué el tráfico dirigido a la subred 13.5.5.0/26 desde AS200 entre por R1 y el tráfico dirigido a la subred 13.5.5.64/26 desde AS200 entre por R2. Los AS300 y AS400 no tienen que aprender la división en subredes /26 que ha creado AS100, pero sí han de ser capaces de llegar a ellas. Indica qué routers reciben qué redes en cada uno de estos casos:



a) AS100 envía por R1 la red 13.5.5.0/26 con la comunidad no-export y la red 13.5.5.0/24 sin ningún tipo de comunidad. AS100 envía por R2 la red 13.5.5.64/26 con la comunidad no-export y la red 13.5.5.0/24 sin ningún tipo de comunidad.

Las /26 las reciben sólo los routers R3 y R4, y las /24 las reciben todos los routers.

b) AS100 envía por R1 la red 13.5.5.0/26 con una comunidad 100:30 que pone un local-pref=200 en quien lo reciba y la red 13.5.5.0/24 sin ningún tipo de comunidad. AS100 envía por R2 la red 13.5.5.64/26 con una comunidad 100:40 que pone un local-pref=200 en quien lo reciba y la red 13.5.5.0/24 sin ningún tipo de comunidad.

Las redes /26 las reciben todos los routers de AS200. Y las /24 las reciben todos los routers porque se sumariza.

c) AS100 envía por R1 la red 13.5.5.0/26 con la comunidad no-export. AS100 envía por R2 la red 13.5.5.64/26 con la comunidad no-export.

Las reciben R3 y R4.

d) AS100 envía por R1 la red 13.5.5.0/26 con una comunidad 100:30 que pone un local-pref=200 en quien lo reciba. AS100 envía por R2 la red 13.5.5.0/26 con una comunidad 100:40 que pone un local-pref=200 en quien lo reciba.

La red /26 enviada por R1 sólo la recibe R3, y la red /26 enviada por R2 sólo la recibe R4.

Pregunta 14. ¿Qué diferencia hay entre asignar un "route-map" con el comando neighbor en modo "in" o en modo "out"? Explica qué efectos tienen ambas acciones sobre las tablas BGP del router emisor del UPDATE BGP y sobre el router receptor del UPDATE BGP. Indica un atributo que se use en modo "in" y otro en modo "out". Explica qué relación y que diferencia hay entre la tabla de encaminamiento y la tabla BGP.

Si asignamos con el "route-map" en modo in, se usará cuando entren paquetes en el router, en cambio si usamos el modo out, se usará cuando salgan paquetes del router.

MODO IN		MODO OUT	
EMISOR UPDATE	RECEPTOR UPDATE	EMISOR UPDATE	RECEPTOR UPDATE
Su tabla BGP es modificada	-	-	Su tabla BGP es modificada

La tabla BGP contiene todas las posibles rutas para llegar a una red desde ese router, en cambio, en la tabla de encaminamiento sólo aparece la entrada elegida por BGP para llegar a esa ruta.

Pregunta 15. Justifica por qué los routers BGP tienen que estar i-BGP totalmente mallados y explica la diferencia entre el funcionamiento i-BGP y e-BGP respecto al anuncio de rutas.

Diferencias al anuncio de redes:

- Las rutas aprendidas por E-BGP se pueden anunciar por I-BGP y por E-BGP
- Las rutas aprendidas por I-BGP solo se pueden anunciar por E-BGP

Los mensajes aprendidos por I-BGP sólo se pueden anunciar mediante E-BGP, esto se hace para evitar que los bucles internos no paren de anunciar las mismas rutas debido a los bucles internos. Si usamos la política anterior solucionamos el problema de los bucles, peró aparece un nuevo problema: sólo los routers a un salto del ASBR podrían aprender las rutas. Para solucionar este problema, tenemos que hacer una malla completa con sesiones I-BGP de todos los routers BGP.

Pregunta 16. Explica que es multi-homing y explica cómo se puede implementar una línea de back-up con un ISP.

Multi-homing es cuando un cliente mantiene más de una conexión con un mismo proveedor. Se puede usar como backup usando el Local-Preference poniendo que todo el tráfico tenga preferencia por una conexión y si esta falla, que use la otra (ejemplo en el 11.b).

Pregunta 17. Explica que significa que el encaminamiento externo e interno estén sincronizados.

Un AS generalmente tiene un protocolo de encaminamiento interno activo que gestiona las tablas de encaminamiento de todos los routers. Para configurar BGP, se tiene que haber configurado previamente este protocolo interno (normalmente OSPF o IS-IS) para que de esta manera, los dos estén sincronizados (interno y externo).

Pregunta 18. Asume que tienes un ISP con 100 routers BGP. Indica cuantas sesiones I-BGP necesita para funcionar correctamente. Indica que técnicas hay para reducir el número de sesiones I-BGP y explica brevemente el funcionamiento de una de ellas. Pon un ejemplo de las técnicas que has explicado enseñando la reducción de sesiones BGP a los 100 routers.

#sesiones = N * (N-1) / 2 = 100 routers * 99 routers / 2= **4950 sesiones IBGP bidireccionales**

Tenemos dos técnicas para reducir este número:

1) **Reflectores de rutas:** Seleccionamos un número de routers que hagan de reflectores, seguidamente los routers que no han sido seleccionados, hacemos una sesión IBGP con uno de los routers reflectores. Entonces cada reflector tiene una topología de estrella con sus clientes y una malla completa entre reflectores. Los RR transmiten la información a todos sus clientes.

   Ejemplo: #reflectores = 10 RR -> #clientesRR = 9 routers/RR

   #sesiones = SUM(#clientesRRi)+#reflectores*(#reflectores - 1)/2 = 9 clients * 10 + 10 RR * 9 RR/2

   = **135 sesiones IBGP bidireccionales**

2) **Confederaciones:** Dividimos el AS en múltiples AS más pequeños (identificados con números de AS privados), como cada sub-AS tiene una cantidad menor de routers, si hacemos una malla completa en cada confederación, usaremos menos sesiones y solo falta conectar los AS con una topología de trayecto mediante E-IBGP.

   Ejemplo: #confederaciones = 4 conf -> #routers_conf = 100 / 4 = 25 routers/conf

   #sesiones = #routers_conf * (#routers_conf - 1)/2 * #confederaciones = (25 * 24)/2 * 4 =

   = **1200 sesiones IBGP bidireccionales**

Pregunta 19. Asume que tienes un ISP con 1000 routers BGP.

a) Indica cuántas sesiones I-BGP necesita para funcionar correctamente.

#sesiones = N * (N-1) / 2 = 1000 routers * 999 routers / 2= **499500 sesiones IBGP bidireccionales**

b) Definimos 10 reflectores de rutas con 99 clientes por cada reflector. ¿Calcula el número de sesiones BGP que se necesitan?

#sesiones = SUM(#clientesRR) + #RR * (#RR - 1) / 2 = 10 * 99 clientes + 10 RR * 9 RR / 2 =

= **1035 sesiones IBGP bidireccionales**

c) Definimos 10 confederaciones con 100 routers por confederación, ¿Calcula el número de sesiones BGP que se necesitan?

#sesiones = 10 conf * (100 routers * 99 routers /2) + 9 conexiones-EIBGP= **495509 conexiones BGP bidir**

d) Definimos 5 confederaciones con 200 routers por confederación y dentro de cada confederación, definimos 5 reflectores de rutas con 39 clientes, ¿Calcula el número de sesiones BGP que se necesitan?

#sesiones = 5 conf * (5 RR * (39 clientes + 4 RR)) + 4 conexiones-EIBGP =

= **1079 sesiones BGP unidireccionales**

Pregunta 20. Explica el funcionamiento de los reflectores de rutas en BGP.

En un dominio se eligen un número X de RR (reflectores de rutas) y se hace una malla completa con estos mediante I-BGP. Los routers que no han sido elegidos como RR, pasan a ser clientes de sólo un RR mediante una conexión I-BGP teniendo en cada cluster una topología de estrella (con el RR en el centro). El criterio que se usa el RR para compartir los mensajes BGP es la siguiente:

1) Si el mensaje no proviene de un cliente (otro RR), lo refleja a todos sus clientes.
2) Si el mensaje proviene de un cliente, lo refleja a todos sus vecinos y clientes.
3) Si el mensaje proviene de un vecino mediante E-BGP (otro AS), lo refleja a todos los clientes y vecinos.

Pregunta 21. Explica el funcionamiento de las confederaciones en BGP.

Las confederaciones ayudan a reducir el número de conexiones I-BGP en un AS. Lo hacen separando el AS (con número público [0, 64512)), en múltiples AS más pequeños (con números privados [64512,65535]), así solo tienes que hacer una malla completa con los routers de cada confederación y no con todos los routers del AS. Este subconjunto de AS, de cara al exterior del AS público, comparten el número de AS público, pero dentro de este, cada uno tiene un número privado asignado. Cada confederación tiene una malla completa con todos sus routers mediante I-BGP, y se conecta con las otras confederaciones del AS mediante EI-BGP (no hace falta hacer una malla completa con EI-BGP).

Pregunta 22. Explica los conceptos de escalabilidad, sincronización y convergencia en BGP y como se solucionan cada uno de ellos.

La **sincronización** nos dice que todos los routers de nuestra red, tienen que tener una tabla BGP en común. Para solucionarlo:

- Haciendo una malla completa con sesiones I-BGP, ya que todos los routers entre dos routers BGP también son BGP.
- Antes de configurar BGP, tenemos que configurar con OSPF y anunciar todas las redes entre los routers del dominio.

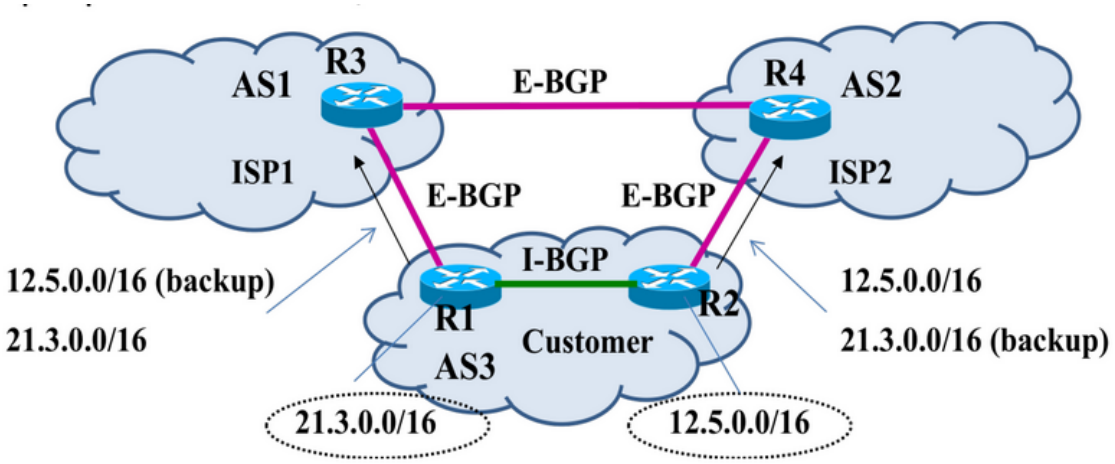
La **escalabilidad** nos dice que cuantos más routers tengamos en una red, más sesiones IBGP tendremos que hacer (#routers * (#routers -1)) debido a la malla completa. Para solucionarlo:

- Usar reflectores de rutas
- Usar confederaciones

La **convergencia** nos dice que como muchas redes van cambiando envían actualizaciones a otros routers BGP para que modifiquen su tabla BGP. Hay veces que las CPUs de los routers no tienen tanta potencia de computación y las redes fallan debido a que no paran de computar nuevas rutas ("meltdown"). Para solucionarlo:

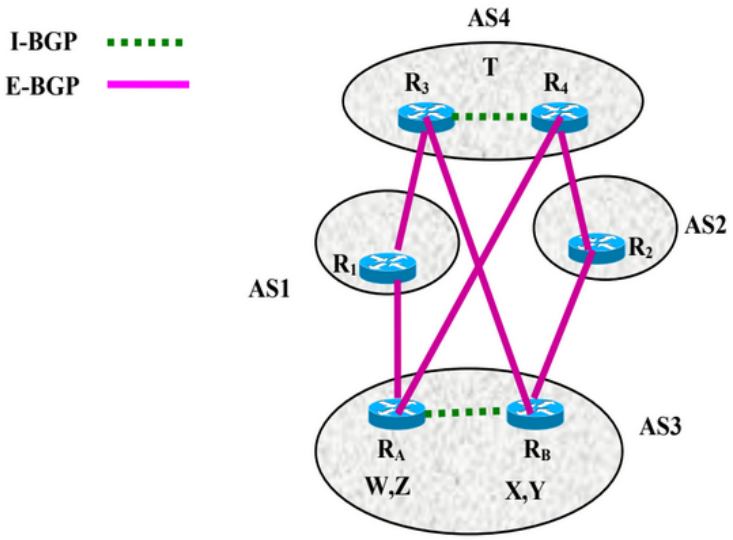
- Minimizar el número de mensajes update en la red usando "exponential back-off"

Pregunta 23. Explica cómo puede el AS3 forzar que la línea R3-R1 es backup para la red 12.5.0.0/16 y principal para la 21.3.0.0/16 y la línea R4-R2 es backup para la red 21.3.0.0/16 y principal para la 12.5.0.0/16.

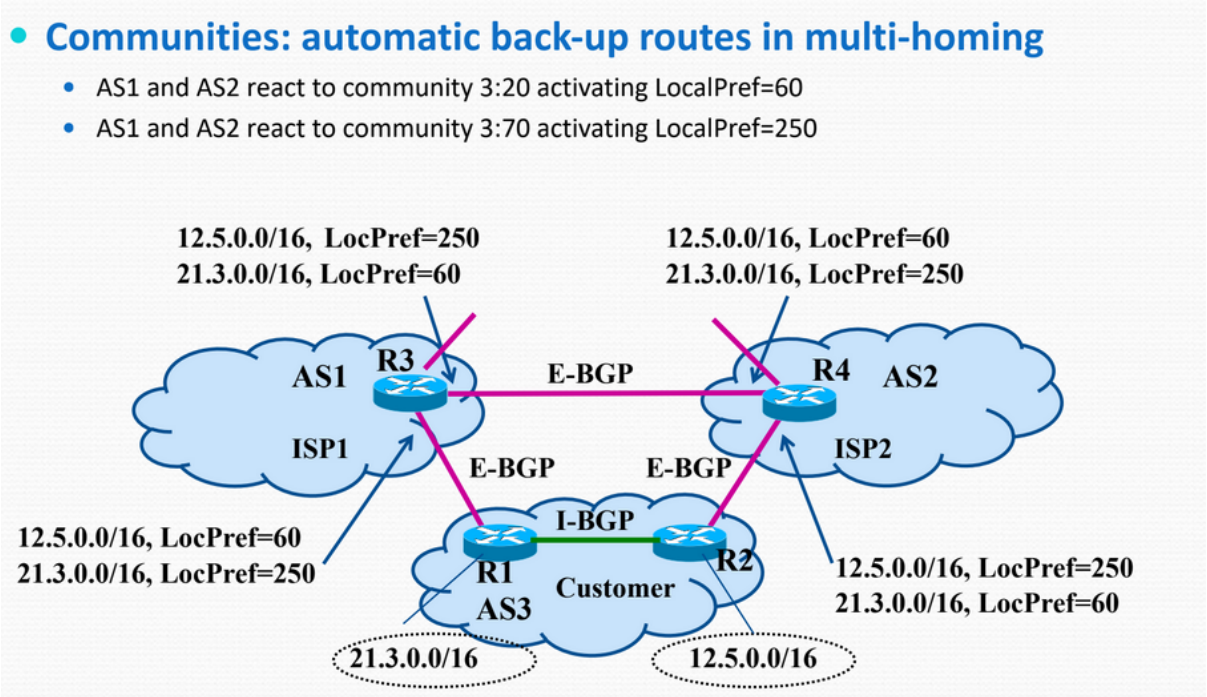


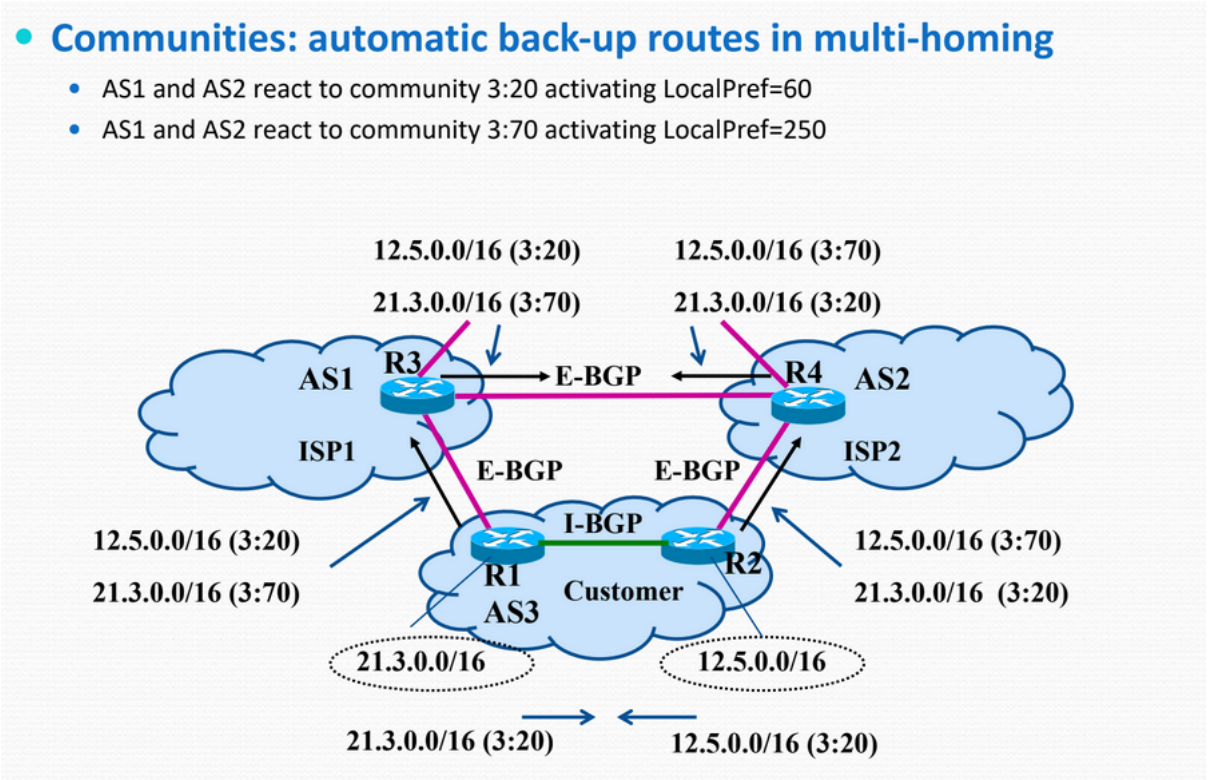
Se puede configurar mediante comunidades. El AS3 anuncia la red 21.3.0.0/16 con el atributo de comunidad 3:30 (que por ejemplo pondrá un Local-Preference a 300 de esa red) al AS1 y anuncia la red 12.5.0.0/16 con un atributo de comunidad 3:20 (que por ejemplo pondrá un Local-Preference a 200 de esa red). Si el AS1 usa "route-maps" para filtrar el tráfico, puede poner la política que si recibe la comunidad 3:30 poner un local-preference de 300 hacia esa red, y si recibe la comunidad 3:20 pondrá un local-preference de 200 en esa red. El AS3 puede hacer la misma configuración para el AS2, pero intercambiando las comunidades de las redes anunciadas (poner a 21.3.0.0/16 la comunidad 3:20 y a la 12.5.0.0/16 la comunidad 3:30). Si AS1 anuncia las redes de AS3 hacia el AS2, lo hará con las comunidades que le ha anunciado AS3 para que así, si el AS2 recibe la red 21.3.0.0/16 la dirija hacia el AS1, y viceversa con el AS2.

Pregunta 24. Asume que tienes la siguiente red. Explica cómo AS3 puede definir una comunidad para que el tráfico que va desde AS4 hacia las redes X,Y en AS3, vaya preferentemente vía R3-RB en vez de usar otras rutas y que el tráfico que va desde AS4 hacia las redes W,Z en AS3, vaya preferentemente vía R4-RA.

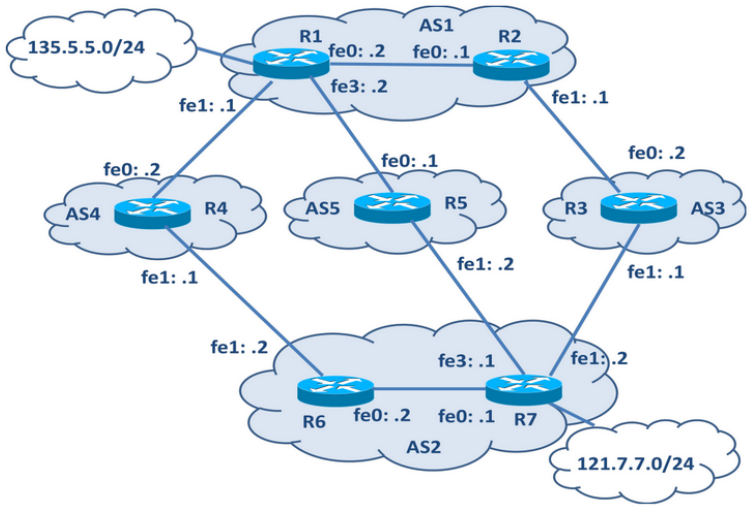


Se pueden definir dos comunidades, por ejemplo: 3:30 y 3:50. La comunidad 3:30 pone un local-preference de 80 en la ruta que ha anunciado esa red, y el 3:50 pone un local-preference de 120 en la ruta que ha anunciado esa red. Si queremos que el tráfico de las redes X,Y venga por la vía R3-RB, anunciamos desde RB hacia R3 las rutas X,Y con la comunidad 3:50, y las redes W,Z con las comunidades 3:30. Hacemos lo mismo con la vía R4-RA intercambiando las comunidades anunciadas.





Pregunta 25. Asume que tienes la arquitectura de la figura. Explica ya a la red 121.7.7.0/24 conectada al router R7 preferentemente vía R1-R5 como primer opción, vía R1-R2 como segunda opción y finalmente vía R1-R4 como tercera opción



Pregunta 26. Asume que tienes un ISP con 100 routers BGP. Para que funcionen correctamente necesitas una red totalmente mallada i-BGP. Obtén el número total de sesiones iBGP necesarias para que funcione correctamente el AS. Definimos ahora una configuración con 5 confederaciones: en las 3 primeras confederaciones se configuran reflectores de routers (4 Reflectores con 4 clientes

cada uno), mientras que en las 2 últimas confederaciones no hay reflectores. Obtén el número total de sesiones i-BGP de cada confederación y el total de la nueva configuración en el AS.

#sesiones1 = #routers * (#routers-1) = 100 * 99 = 9900 sesiones I-BGP unidireccionales

En las 2 confederaciones si routers hay 10 routers en cada una, ya que:

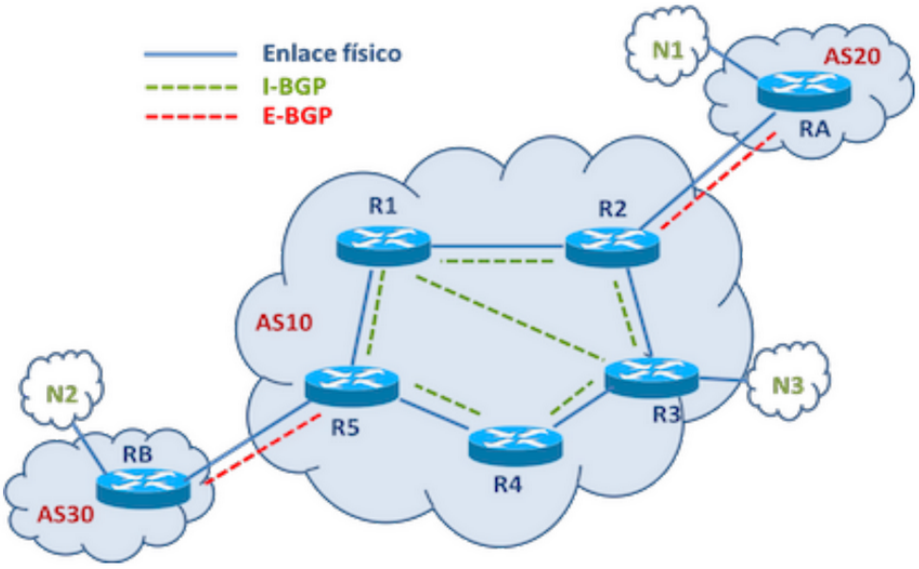
N = (100 routers - (4 clientes * 4 RR + 4 RR)* 3 conf)/2 conf = 10 routers

#sesionesIBGP = 3 conf * (#clientes * #RR + #RR * (#RR-1)) + 2 conf * (N * (N-1)) =

= 3 * (4 * 4 + 4 * 3) + 2 * (10 * 9) = **264 sesiones I-BGP unidireccionales**

#sesionesBGP = #sesionesIBPG + #conf - 1 = 264 IBGP + 4 EBGP = **268 sesiones BGP**

Pregunta 27. En la figura siguiente tienes una red con enlaces I-BGP y E-BGP, busca una solución con reflectores de rutas, dadas las conexiones existentes para que las redes N1, N2 y N3 sean alcanzables desde todos los routers.



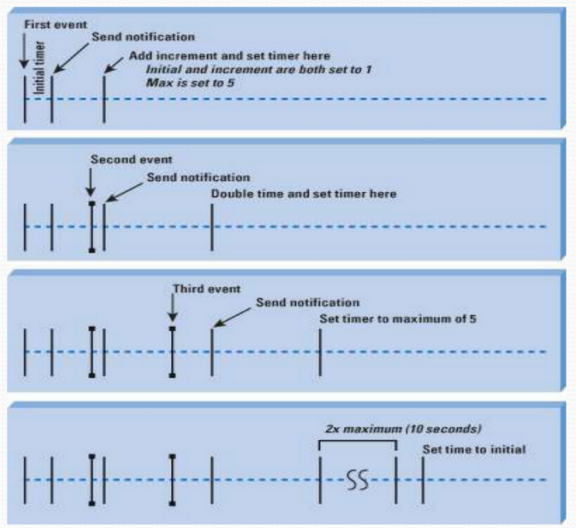
R1 y R3 serían los reflector routers, y R1 por ejemplo tendría como clientes a R5; y R3 tendría como clientes a R4 y R2.

Pregunta 28. Explica que es el "flapping". ¿Para qué sirven las técnicas de "slow-down" en BGP? Explica el "exponential back-off".

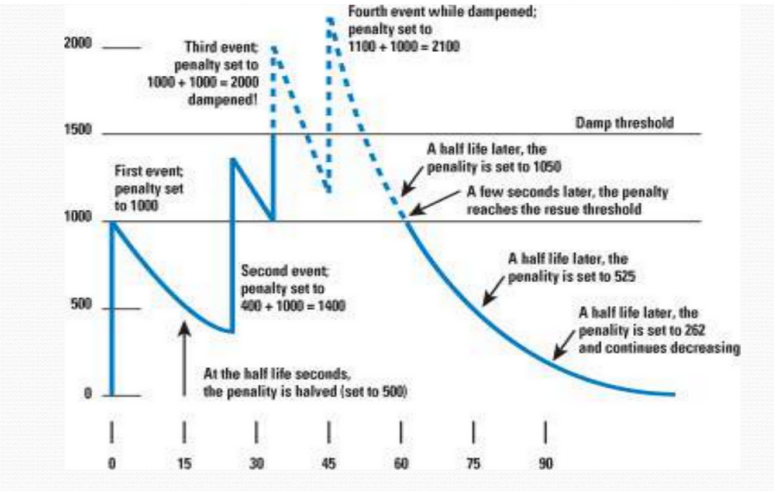
El "flapping" se produce cuando una interfaz de un router cambia de estado constantemente (de up a down), generando muchos mensajes update y consecuentemente una bajando la convergencia de la red, bucles y fallos.

Los "slow-down" sirven para reducir la frecuencia de estos paquetes update a otros routers BGP debido al "flapping" de algunas interfaces de los routers.

El “exponential back-off” es un método para reducir la continuidad de mensajes update hacia la red. Lo que hace es poner un temporizador que va creciendo exponencialmente y se reinicia una vez que ha llegado al máximo. Si se genera un evento antes que haya pasado el tiempo de notificación, este se esperará hasta que el tiempo haya transcurrido.



Pregunta 29. Explica qué es el dampening en BGP y para qué sirve. Explica el mecanismo de dampening y la relación entre el temporizadores half-time, max-supress-limit y el valor máximo del supress-limit (dampening threshold).



El dampening es lo que usa BGP para reducir los mensajes de update en la red debido al “flapping”. Funciona de la siguiente manera:

Inicializa un contador que se va incrementado cada vez que ocurre un evento, este se incrementa según el “**penalty**” que tengamos (por definición 1000). Mientras el contador no supere el umbral “**Damp threshold**” este va enviando los update que se van generando, pero si este lo supera deja enviar updates hasta que sea inferior al **umbral de reuso**. El contador va disminuyendo exponencialmente (se divide entre 2) cada vez que pasen “**half-time**” segundos.

max-penalty = reuse-limit * $2^{(\text{max-suppress-limit/half-life})}$

Pregunta 30. ¿Qué es y qué implicaciones tiene el max-penalty en dampening? Si tienes un penalty = 1000, un reuse-limit = 2000, un half-life = 15 minutos, y un max-suppress-limit = 60 minutos, ¿Cuál es el valor máximo del suppress-limit (dampening theshold) que puedes configurar?

El max-penalty nos asegura que al configurar el dampening, no pongamos un suppress-limit tan alto que este nunca llegue a activarse.

max-penalty = reuse-limit * $2^{(\text{max-suppress-limit/half-life})}$

max-penalty = 2000 * $2^{(60\ \text{min}/\ 15\ \text{min})}$ = **32000**