

Xarxes de computadors II

Tema 4 - Encaminamiento inter-dominio

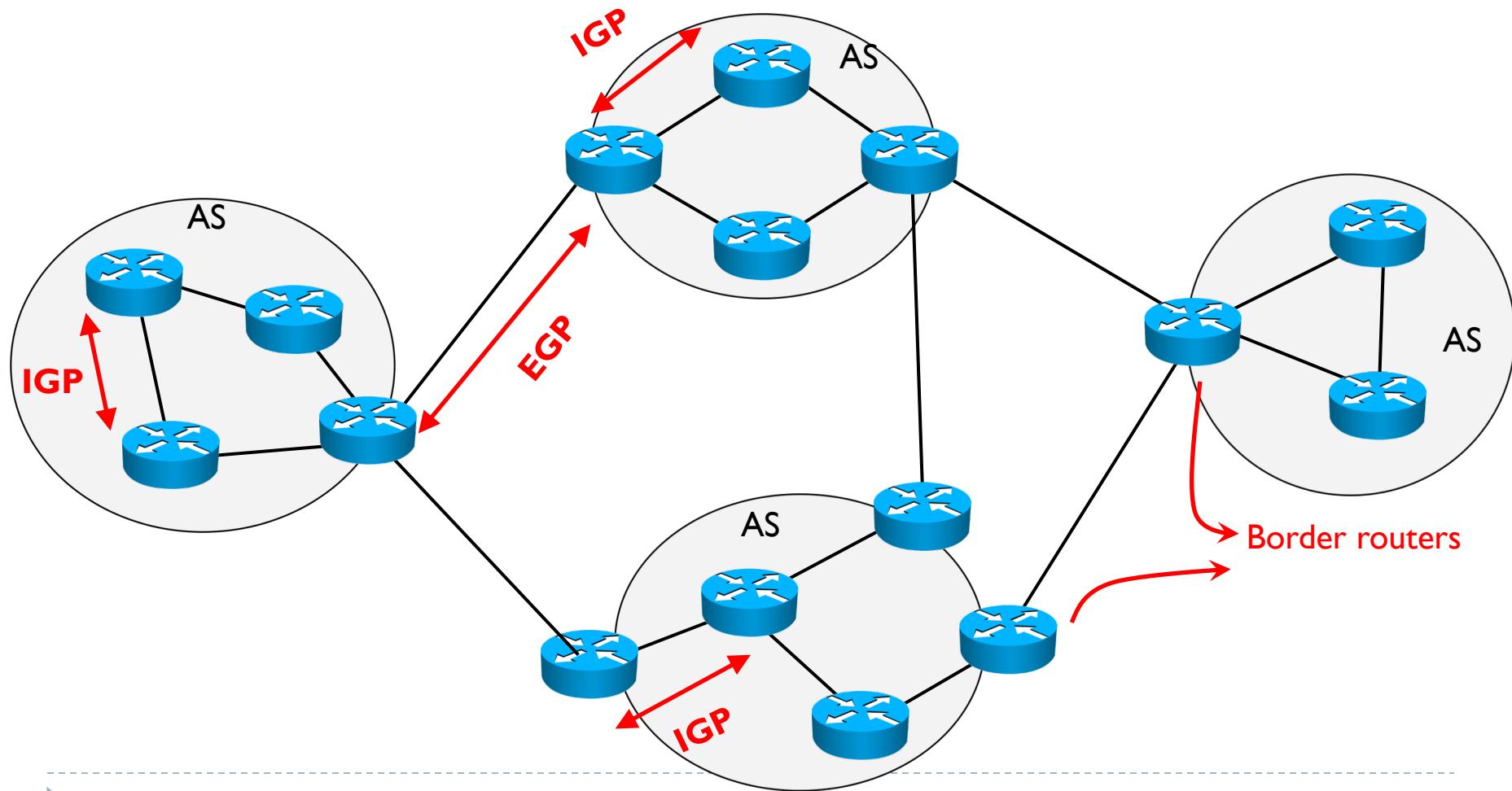
Temas

- ▶ **Tema 1. Introducción**
- ▶ **Tema 2. Arquitectura y direccionamiento en Internet**
- ▶ **Tema 3. Encaminamiento intra-dominio**
- ▶ **Tema 4. Encaminamiento inter-dominio**
- ▶ **Tema 5. Temas de investigación**
- ▶ **Tema 6. Conceptos avanzados**

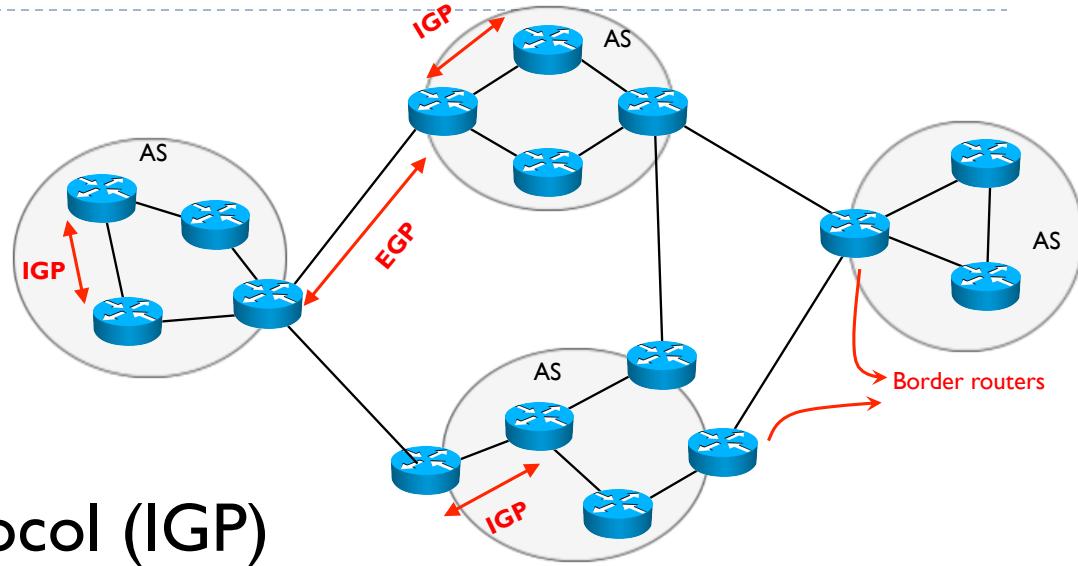


4. Intra-dominio vs. Inter-dominio

- Internet está formada por diferentes AS interconectados



4. Intra-dominio vs. Inter-dominio



- ▶ **Interior Gateway Protocol (IGP)**
 - ▶ RIP - RFC 2453 (versión 2) RFC 2080 para IPv6
 - ▶ OSPF - RFC 2328 (versión 2) RFC 5340 para IPv6
 - ▶ IS-IS - RFC 1142 RFC 5308 para IPv6
 - ▶ **Exterior Gateway Protocol (EGP)**
 - ▶ EGP - RFC 904
 - ▶ BGP - RFC 1771 (versión 4) RFC 2545 para IPv6

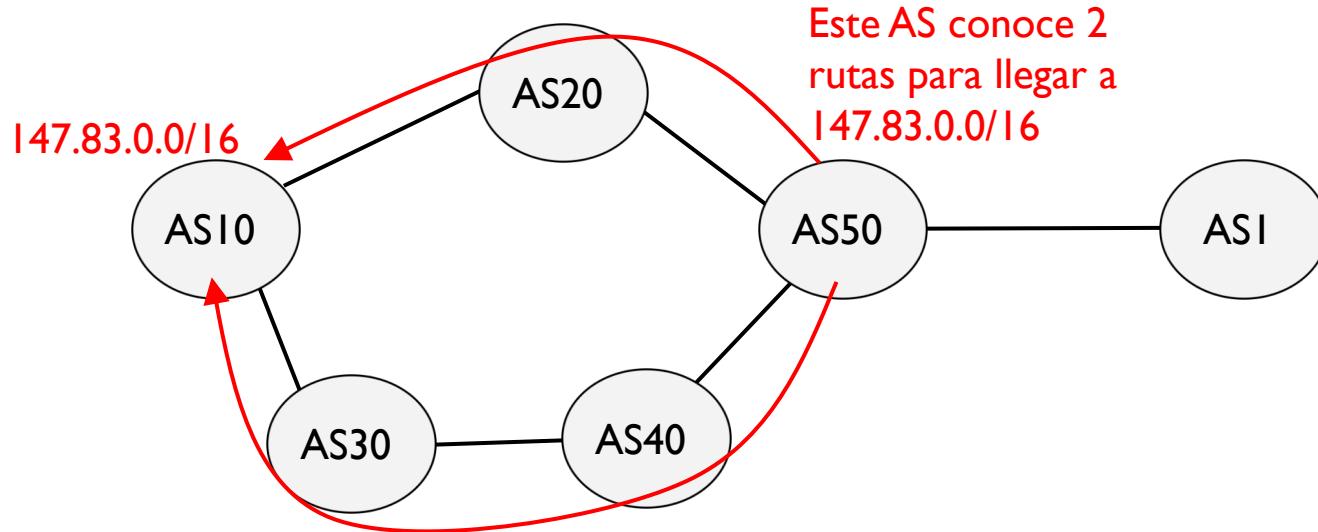
4. Border Gateway Protocol (BGP)

- RFC 1771 → 4271 → 6286
-
- ▶ Protocolo de encaminamiento dinámico entre AS usado en Internet
 - ▶ Basado en vector-distancia (realmente se dice que es path-vector)
 - ▶ El camino entre origen y destino se construye como composición de próximo salto (como RIP)
 - ▶ Eso implica que el conocimiento de un router depende de la decisión de otros
 - ▶ Encaminamiento basado en políticas según unos atributos
 - ▶ Estas políticas influyen en la selección de la ruta
 - ▶ Permite no enviar información considerada confidencial
 - ▶ p.e., topología del AS, número de routers, velocidad de transmisión
 - ▶ Solo se intercambian aquellos prefijos que se quiere que los AS vecinos sepan
 - ▶ En BGP se usa el término genérico prefijos: ya que un prefijo puede no ser una red real



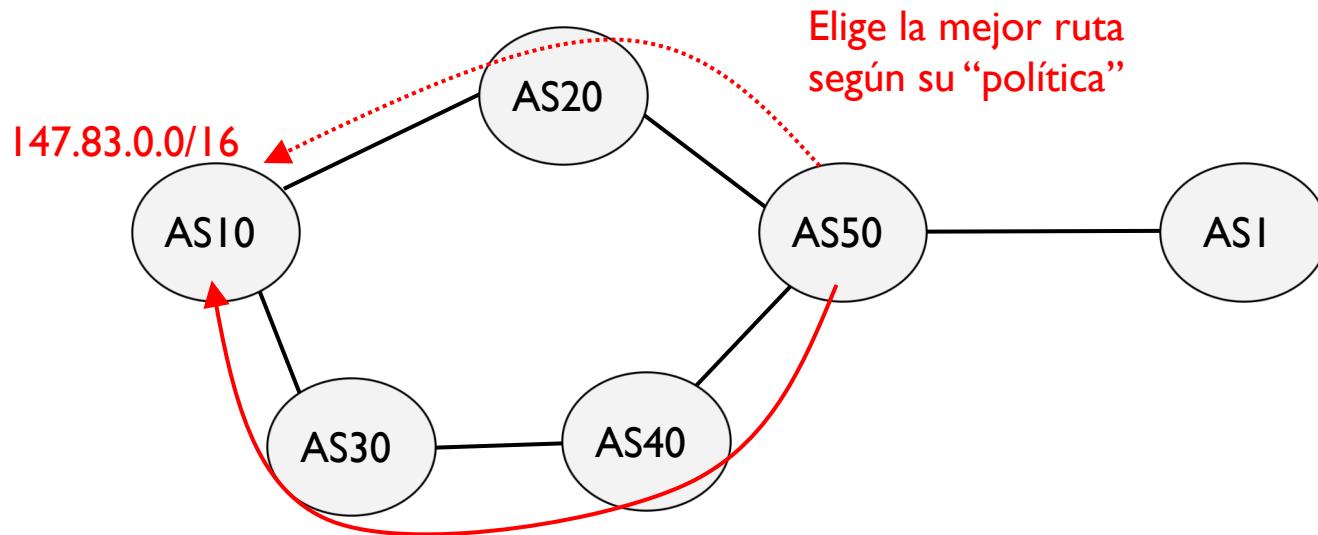
4. Border Gateway Protocol (BGP)

- ▶ Basado en vector-distancia (realmente se dice que es path-vector)
 - ▶ El camino entre origen y destino se construye como composición de próximo salto (como RIP)
 - ▶ Eso implica que el conocimiento de un router depende de la decisión de otros



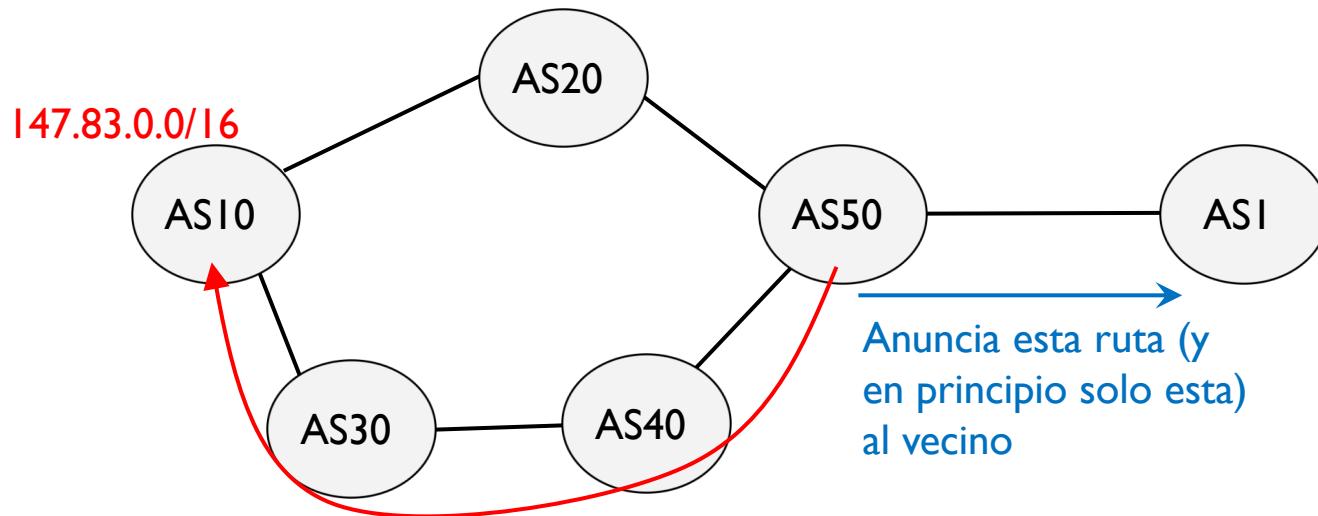
4. Border Gateway Protocol (BGP)

- ▶ Basado en vector-distancia (realmente se dice que es path-vector)
 - ▶ El camino entre origen y destino se construye como composición de próximo salto (como RIP)
 - ▶ Eso implica que el conocimiento de un router depende de la decisión de otros



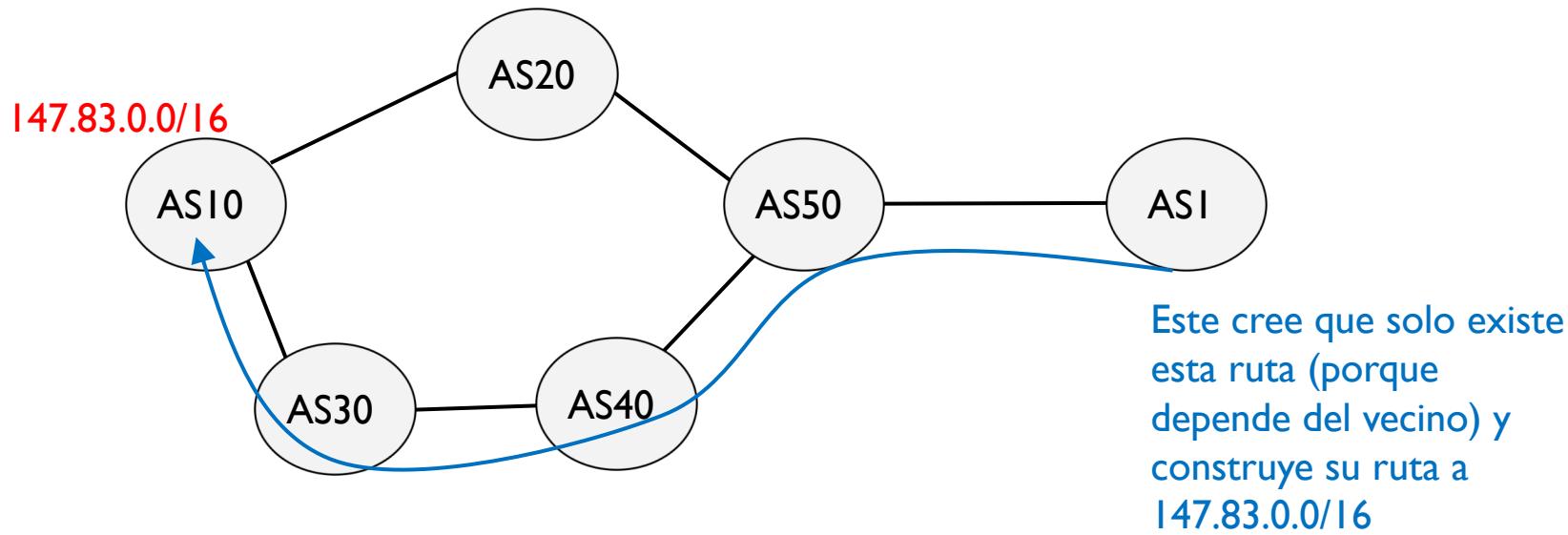
4. Border Gateway Protocol (BGP)

- ▶ Basado en vector-distancia (realmente se dice que es path-vector)
 - ▶ El camino entre origen y destino se construye como composición de próximo salto (como RIP)
 - ▶ Eso implica que el conocimiento de un router depende de la decisión de otros

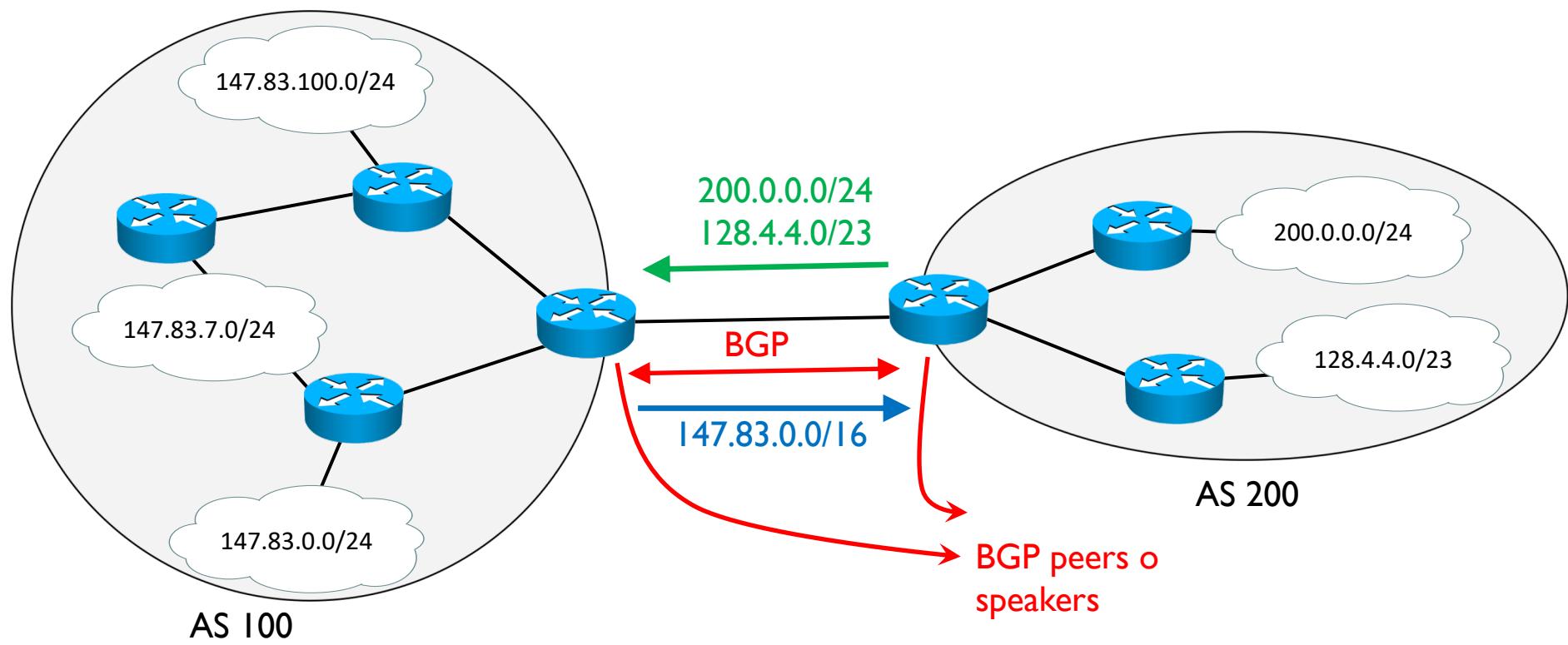


4. Border Gateway Protocol (BGP)

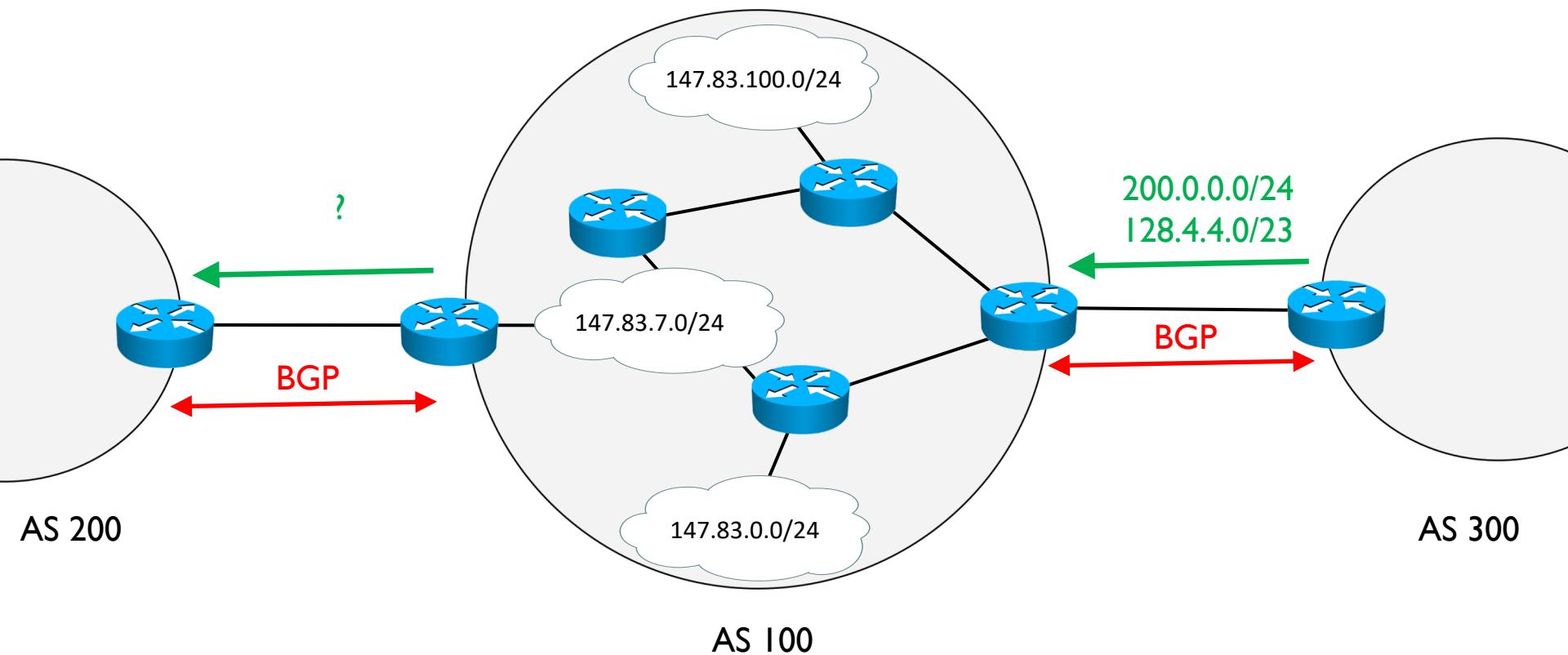
- ▶ Basado en vector-distancia (realmente se dice que es path-vector)
 - ▶ El camino entre origen y destino se construye como composición de próximo salto (como RIP)
 - ▶ Eso implica que el conocimiento de un router depende de la decisión de otros



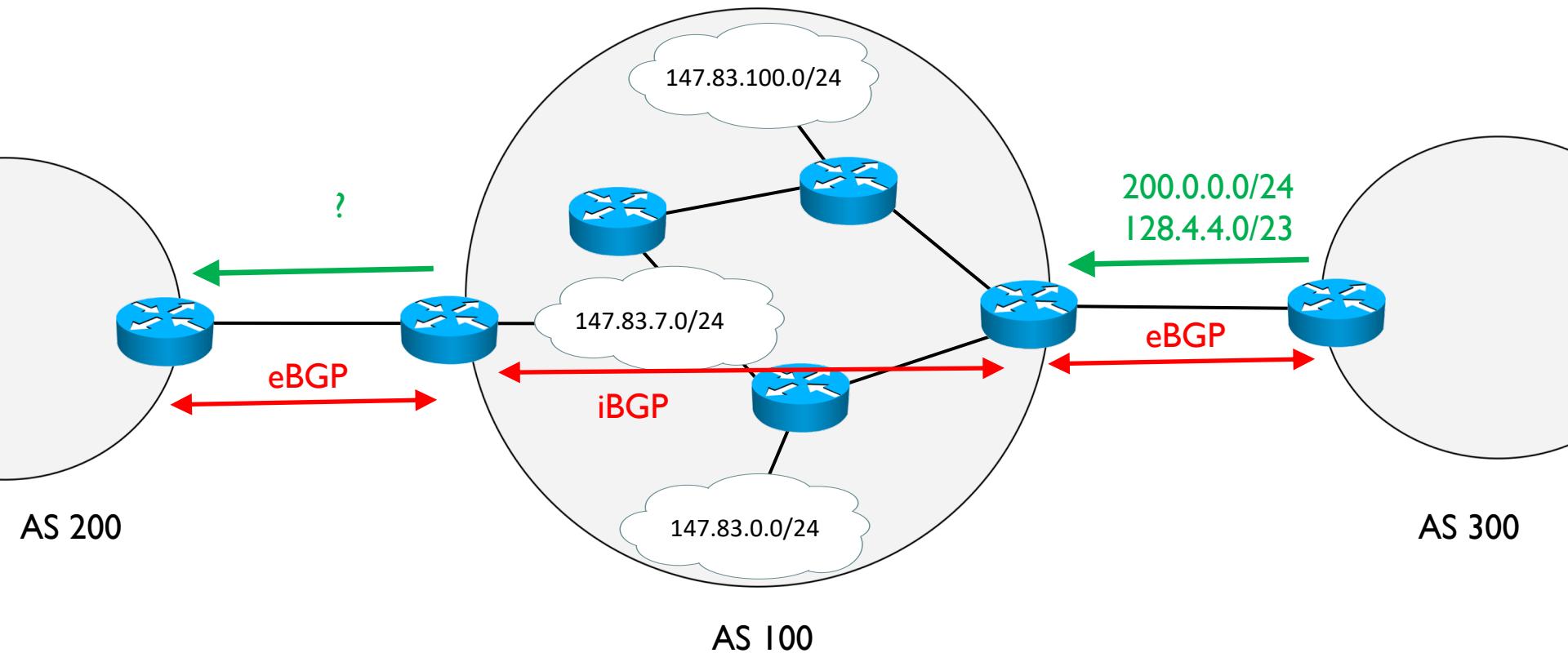
4. Border Gateway Protocol (BGP)



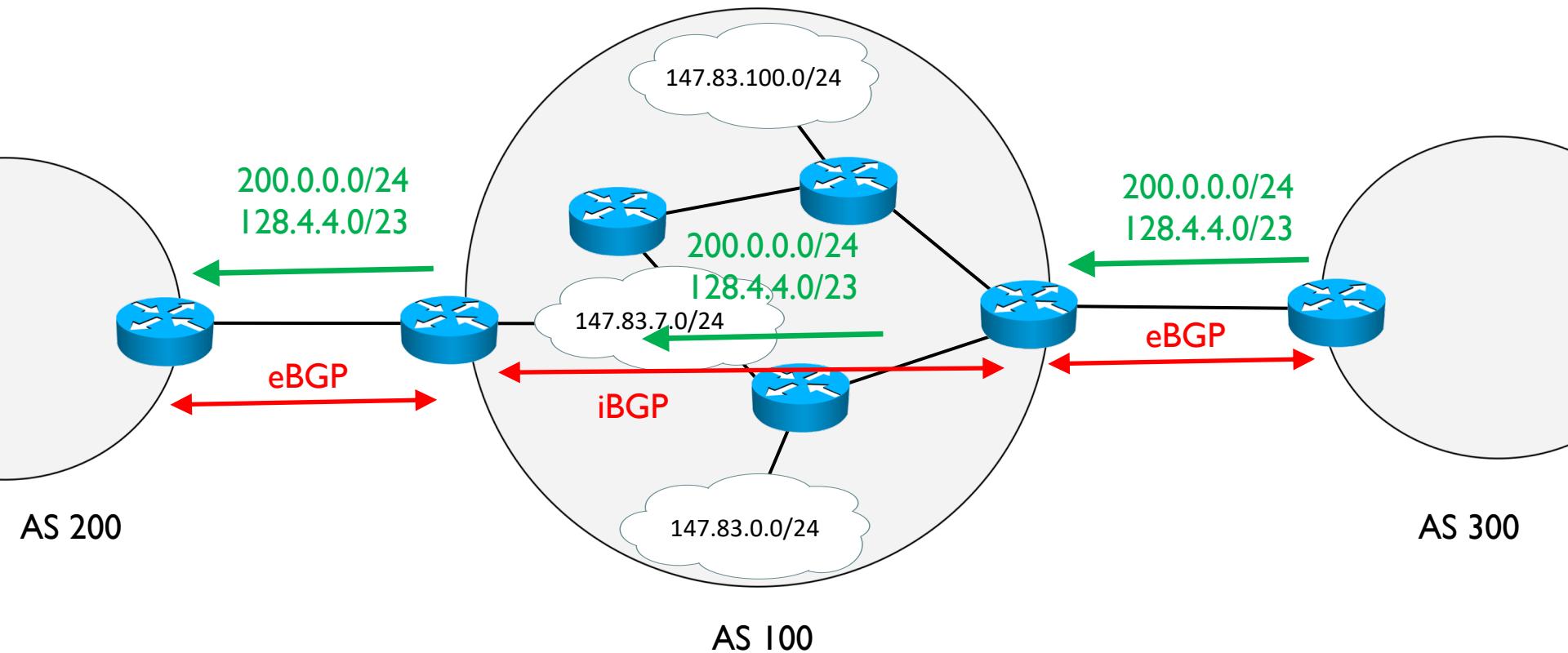
4. external BGP vs. internal BGP



4. external BGP vs. internal BGP

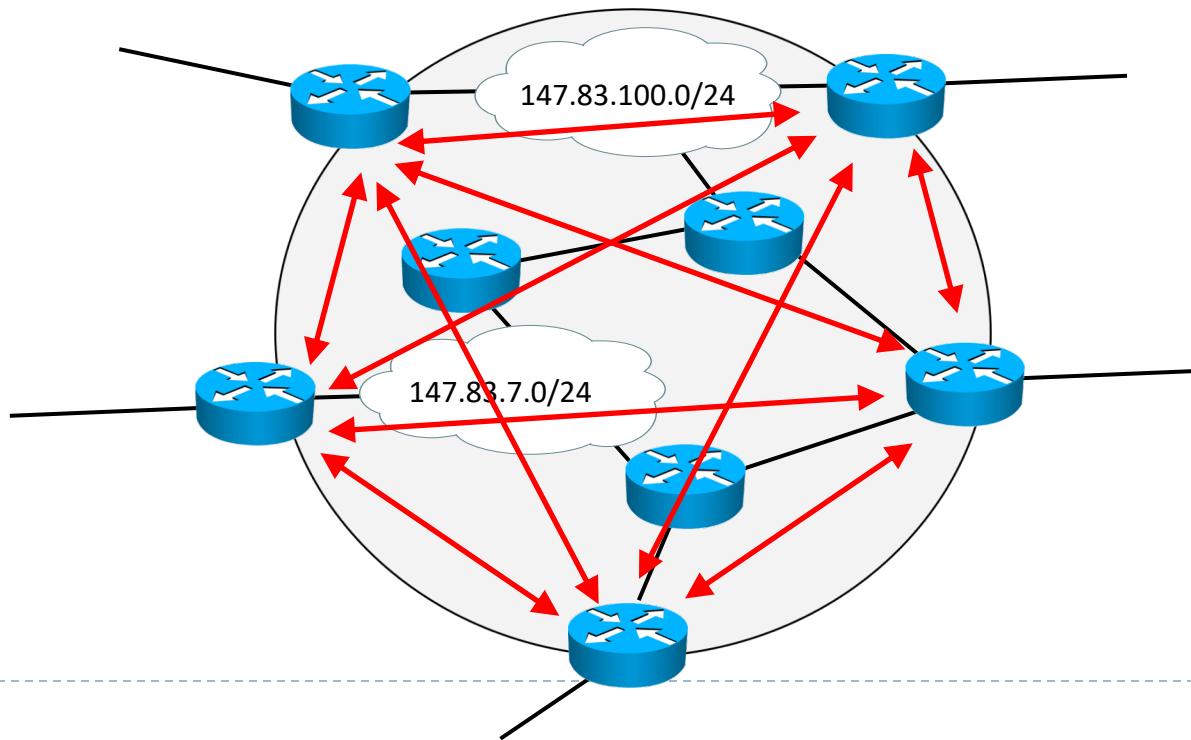


4. external BGP vs. internal BGP



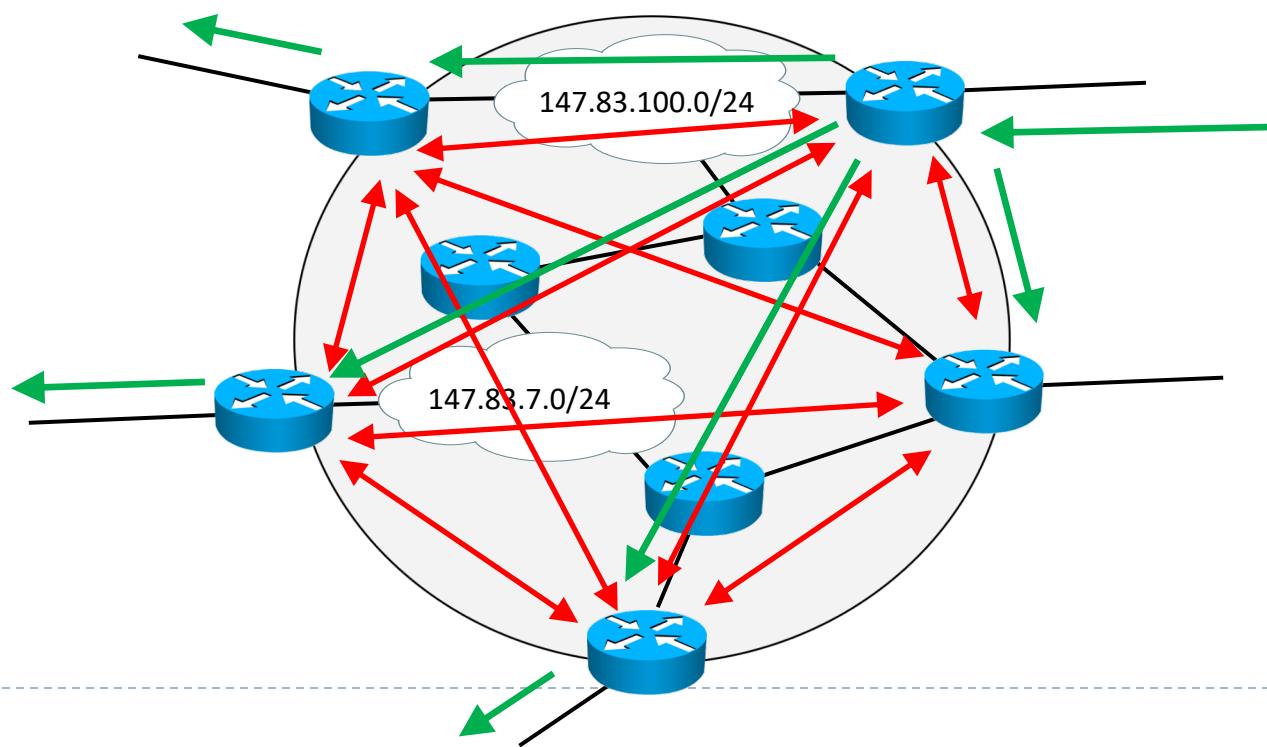
4. BGP

- ▶ Se necesita que todos los BGP speakers de un AS establezcan una sesión iBGP entre ellos
 - ▶ Se crea una full-mesh (malla completa) de iBGP (entre routers que tienen por lo menos una sesión BGP con otro AS)
 - ▶ De esta manera se evitan bucles de mensajes BGP



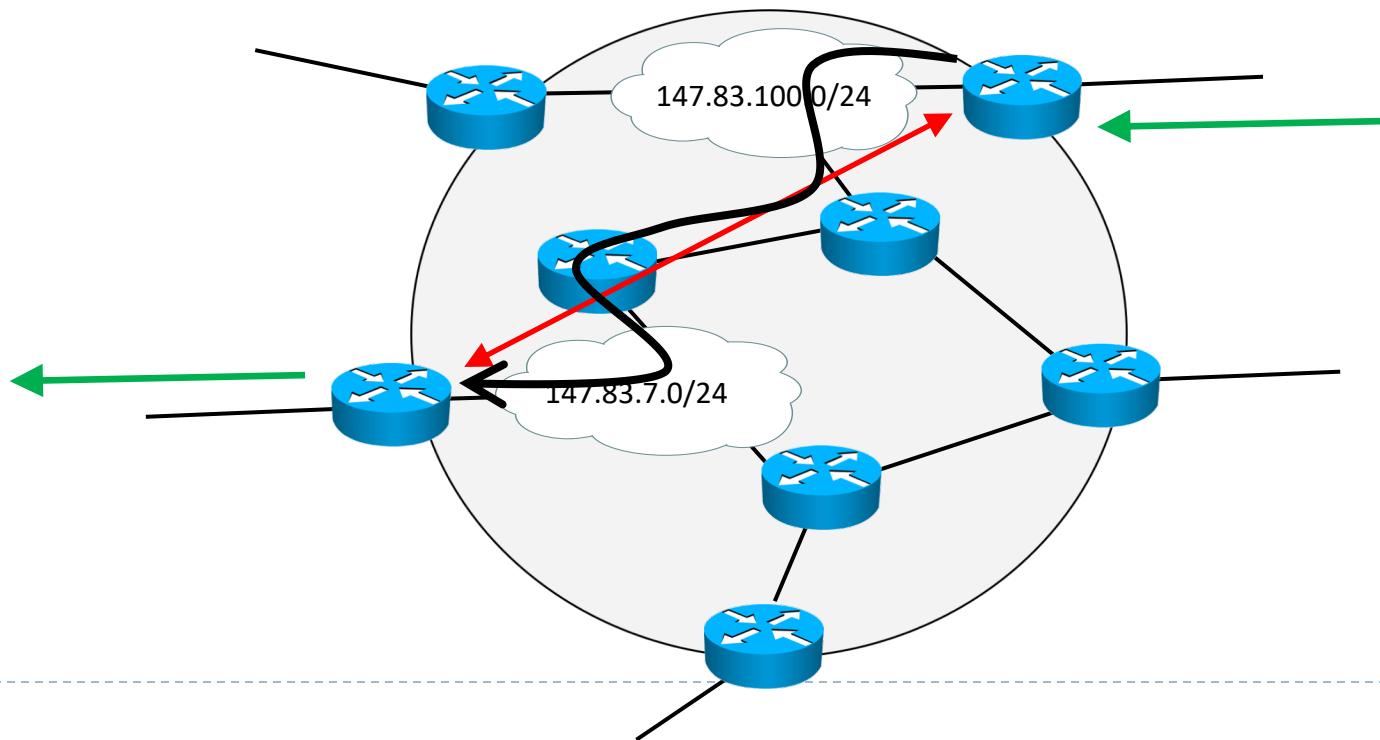
4. BGP

- ▶ Los iBGP speakers usan los mismos mensajes que los eBGP
- ▶ Los iBGP speakers anuncian solo los prefijos que aprenden de los eBGP pero no pueden re-enviar los recibidos de otro iBGP



4. BGP

- ▶ Los mensajes iBGP se envían como si fueran paquetes normales encaminados según las tablas de encaminamiento de los routers
- ▶ El origen y destino de estos mensajes son los iBGP speakers
- ▶ No confundir iBGP con el protocolo de encaminamiento interno IGP



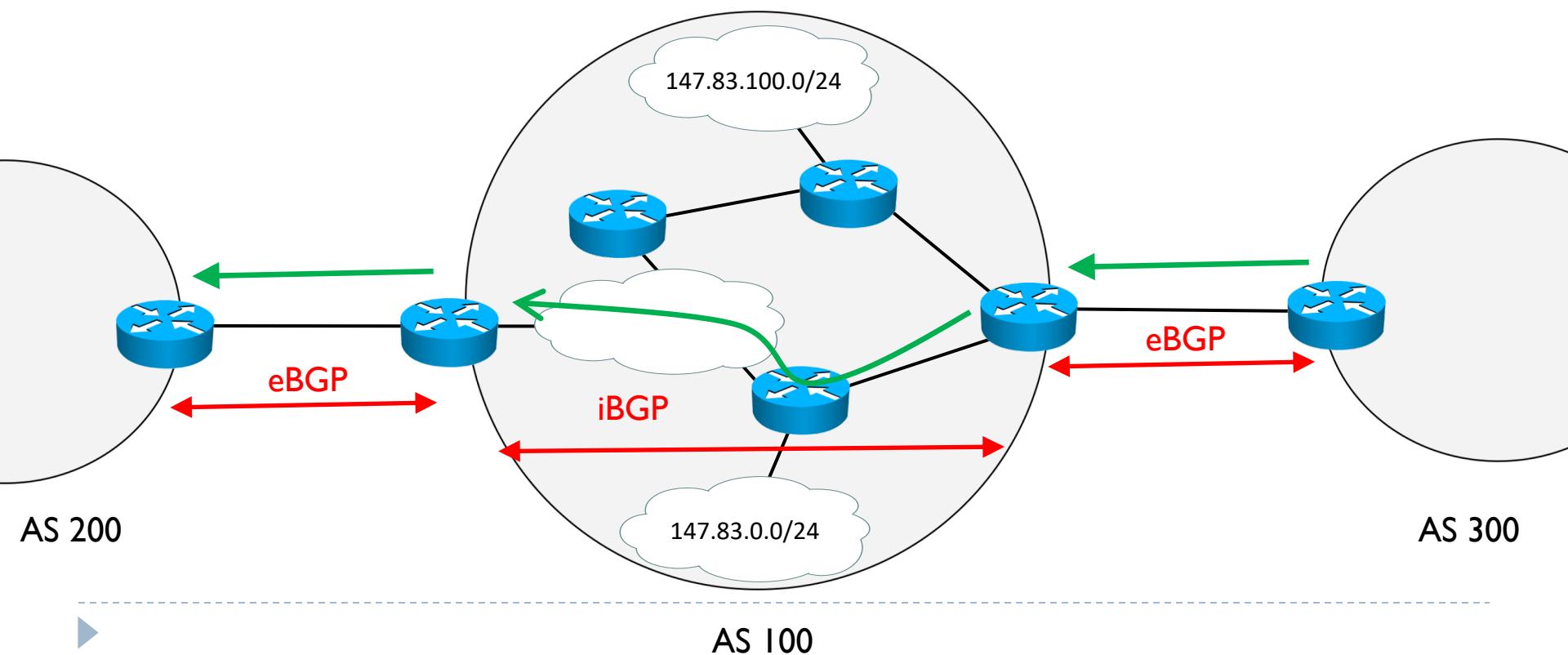
4. Establecimiento sesión BGP

- ▶ A la hora de establecer una sesión BGP se puede elegir una interfaz real o una virtual. Se suele usar
 - ▶ una interfaz real para eBGP
 - ▶ una interfaz virtual para iBGP



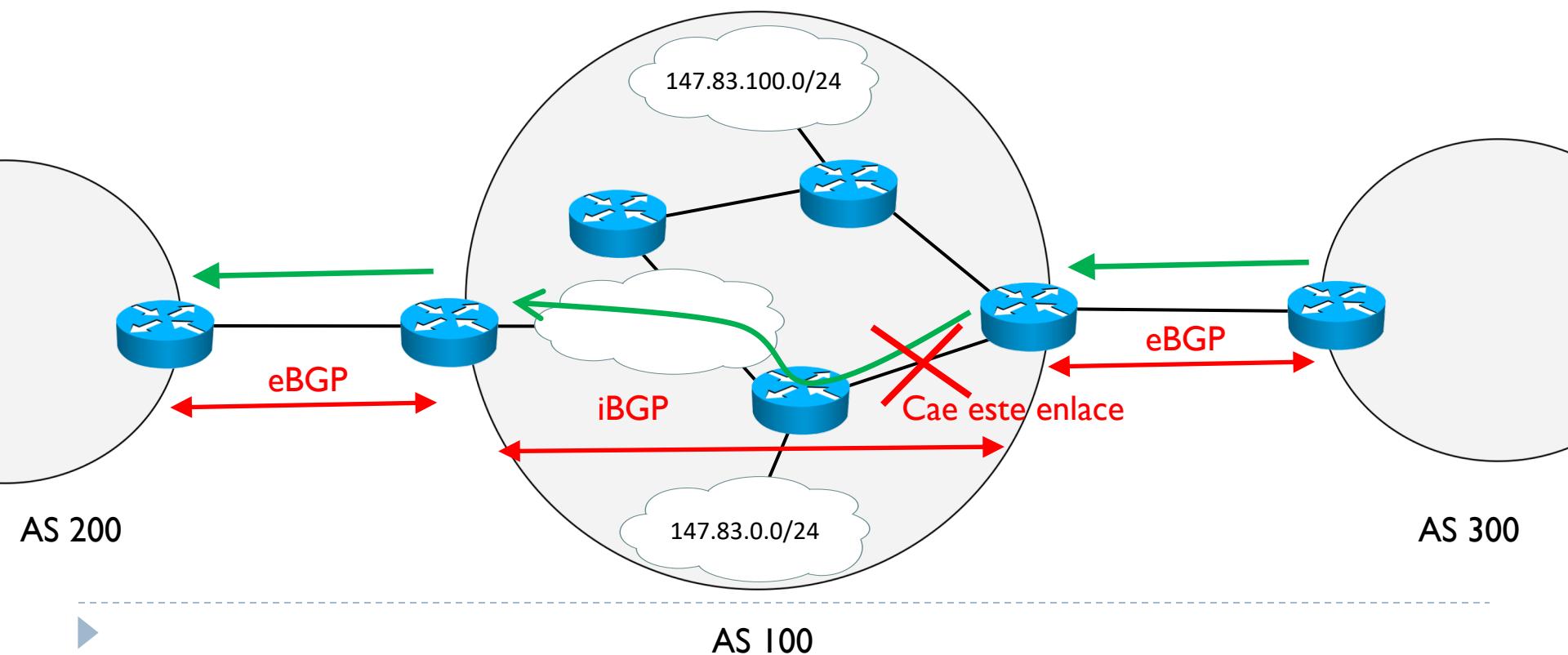
4. Establecimiento sesión BGP

- ▶ A la hora de establecer una sesión BGP se puede elegir una interfaz real o una virtual. Se suele usar
 - ▶ una interfaz real para eBGP
 - ▶ una interfaz virtual para iBGP



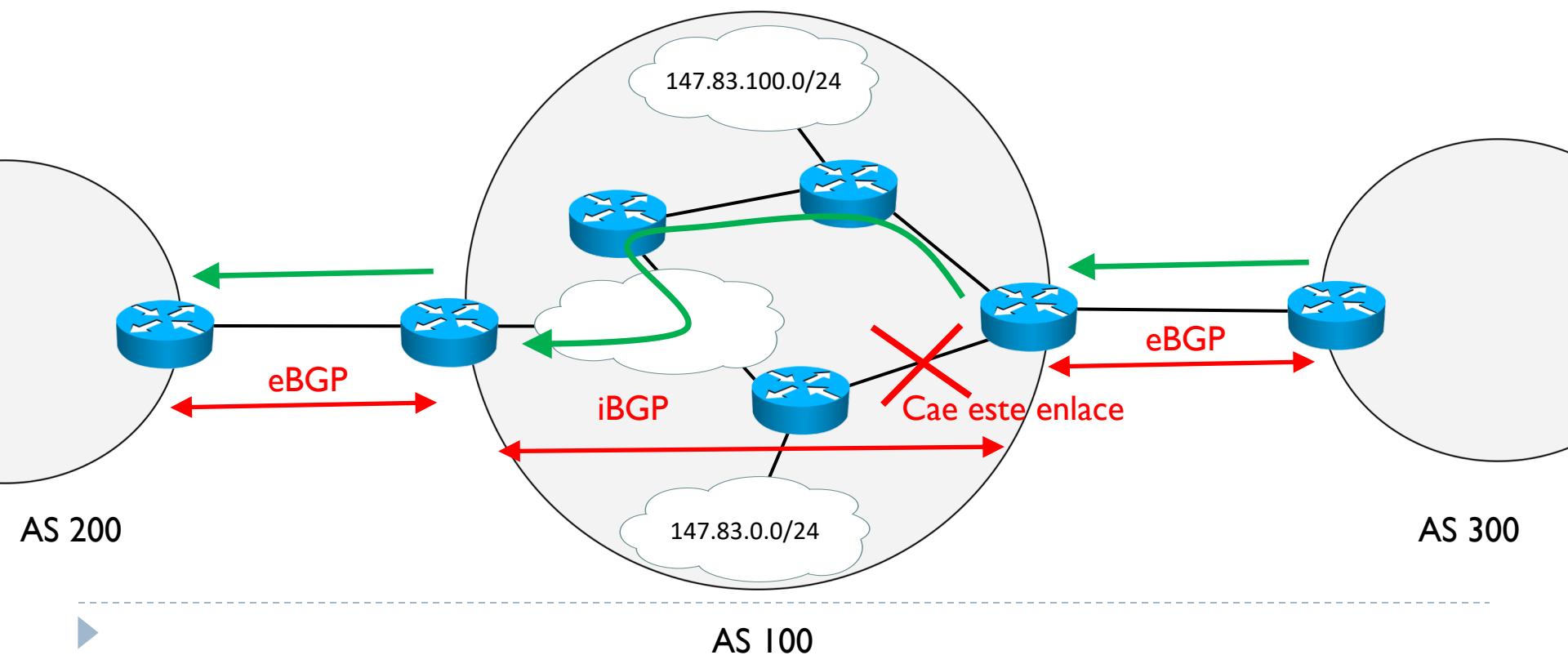
4. Establecimiento sesión BGP

- ▶ Si se usara una interfaz real para iBGP y cae un enlace que se usa para encaminar los mensajes iBGP
- ▶ → Cae la sesión iBGP
- ▶ Hay que volver a configurarla manualmente usando otra interfaz



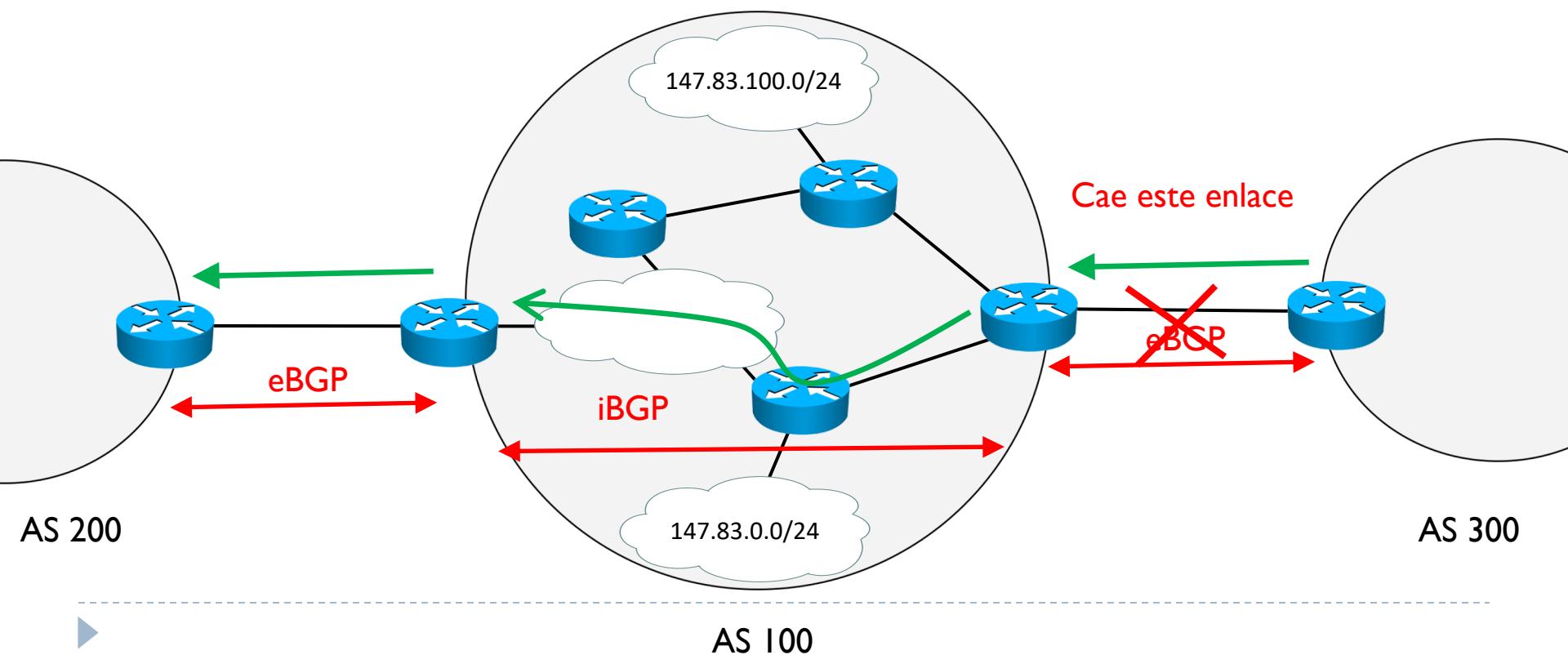
4. Establecimiento sesión BGP

- ▶ Si en cambio se usara una interfaz virtual para iBGP y cae un enlace que se usa para encaminar los mensajes iBGP
- ▶ → el protocolo IGP (interno) encontraría otra ruta para estos mensajes
- ▶ La sesión iBGP sigue activa



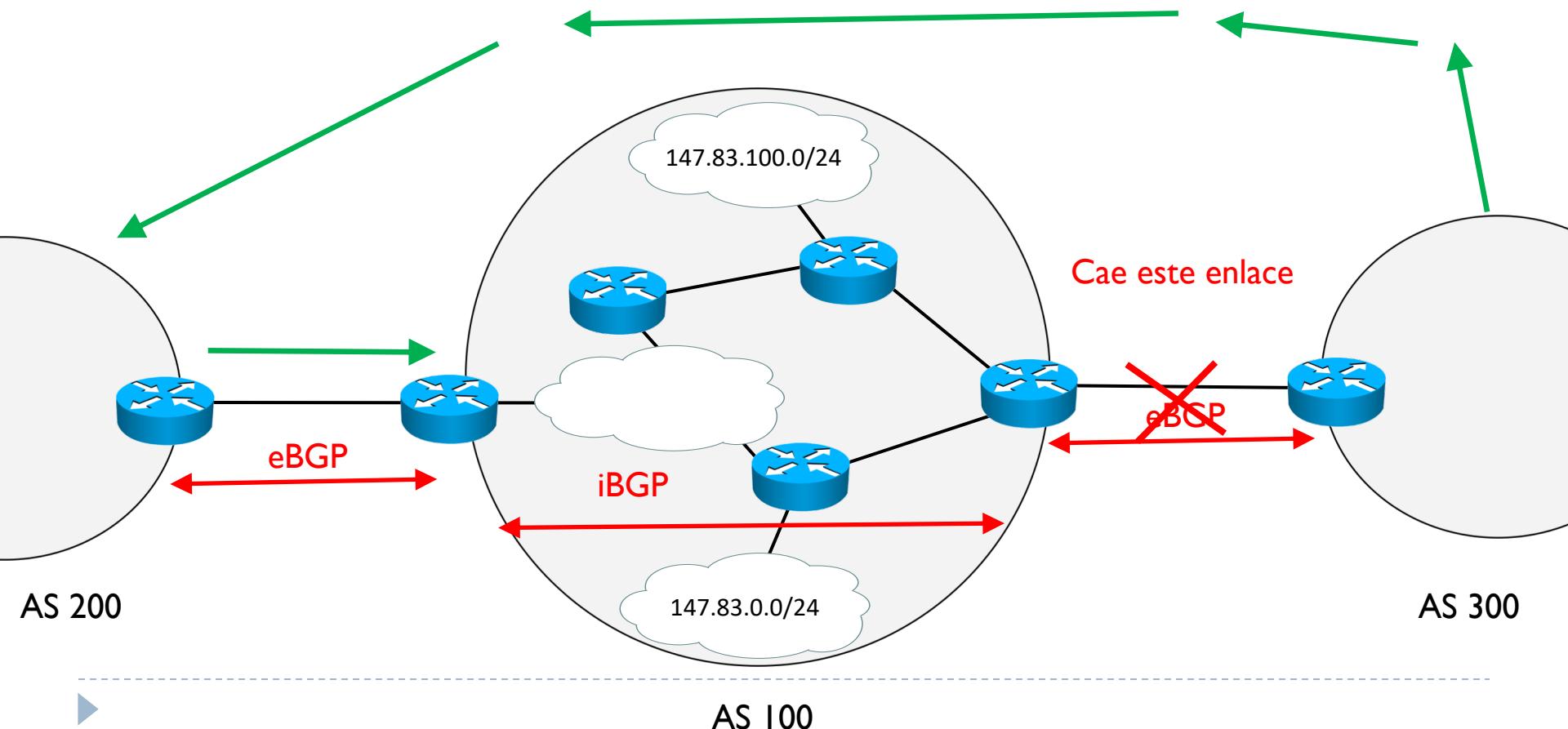
4. Establecimiento sesión BGP

- ▶ Se usa una interfaz real en el caso eBGP porque se quiere que sea BGP que se ocupe de encontrar una ruta alternativa
- ▶ Y BGP busca otra ruta solo si se entera del fallo



4. Establecimiento sesión BGP

- ▶ Se usa una interfaz real en el caso eBGP porque se quiere que sea BGP que se ocupe de encontrar una ruta alternativa
- ▶ Y BGP busca otra ruta solo si se entera del fallo



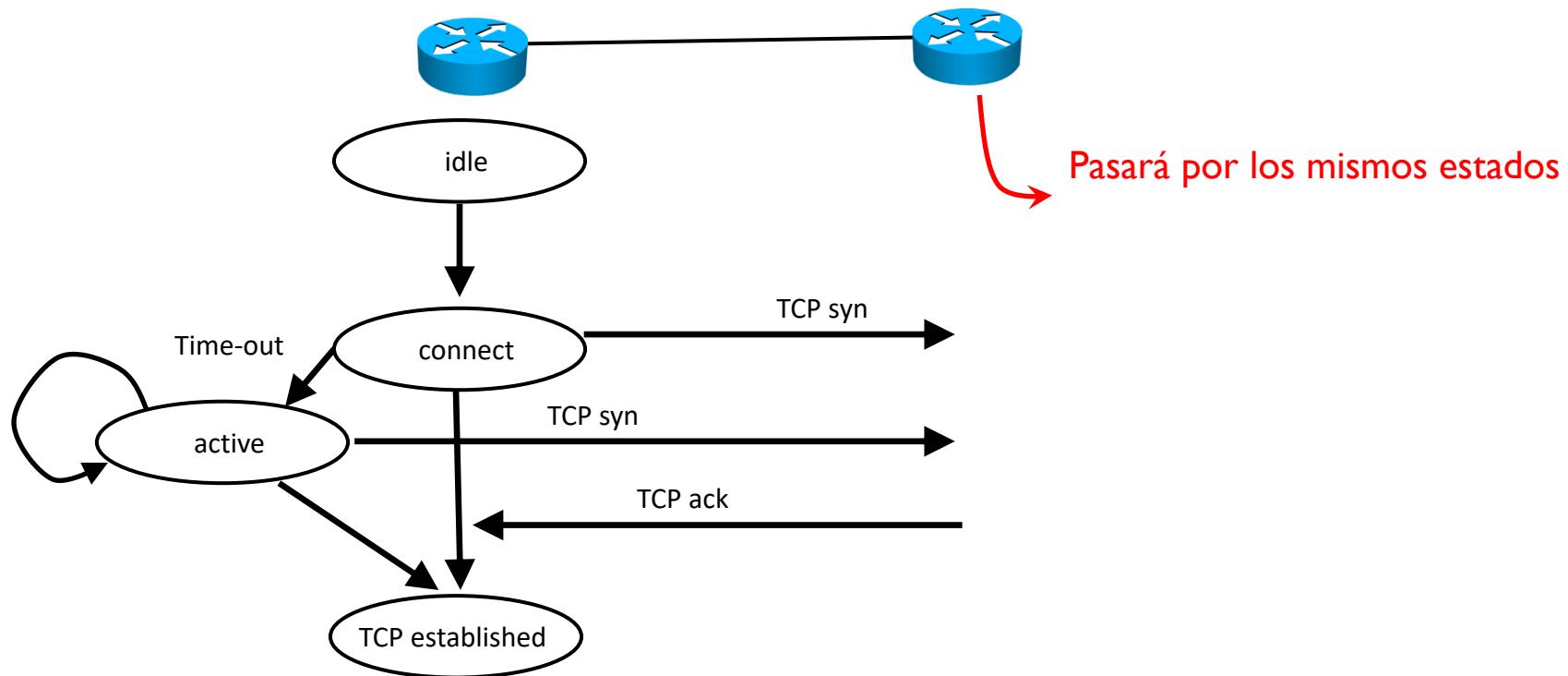
4. Establecimiento sesión BGP

- ▶ La información que se intercambian los BGP speakers debe ser fiable
 - ▶ Se abren una conexión TCP entre ellos
 - ▶ Los dos extremos del TCP son los dos routers
 - ▶ Se usa el puerto 179 (BGP)



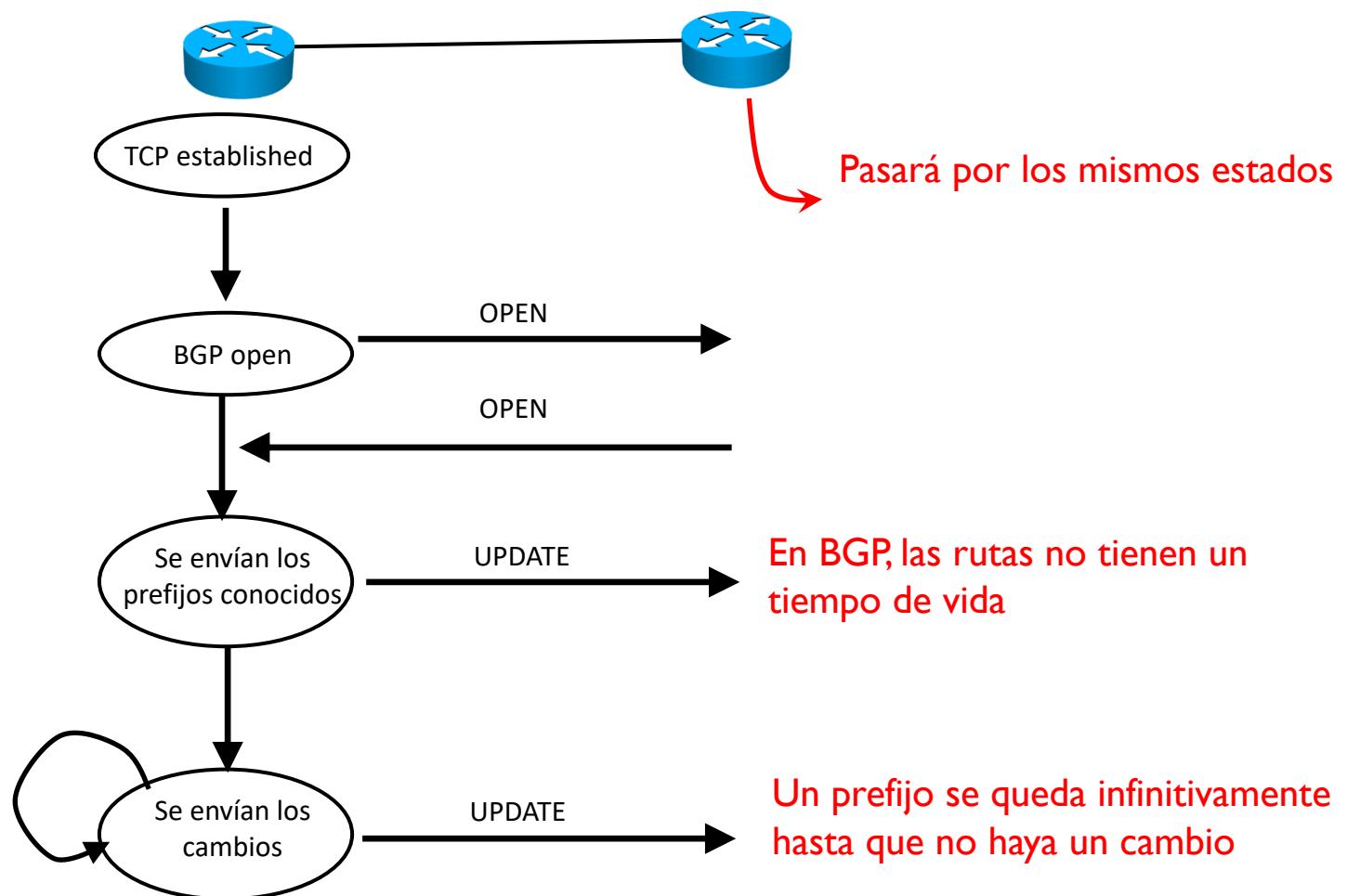
4. Establecimiento sesión BGP

▶ Estados



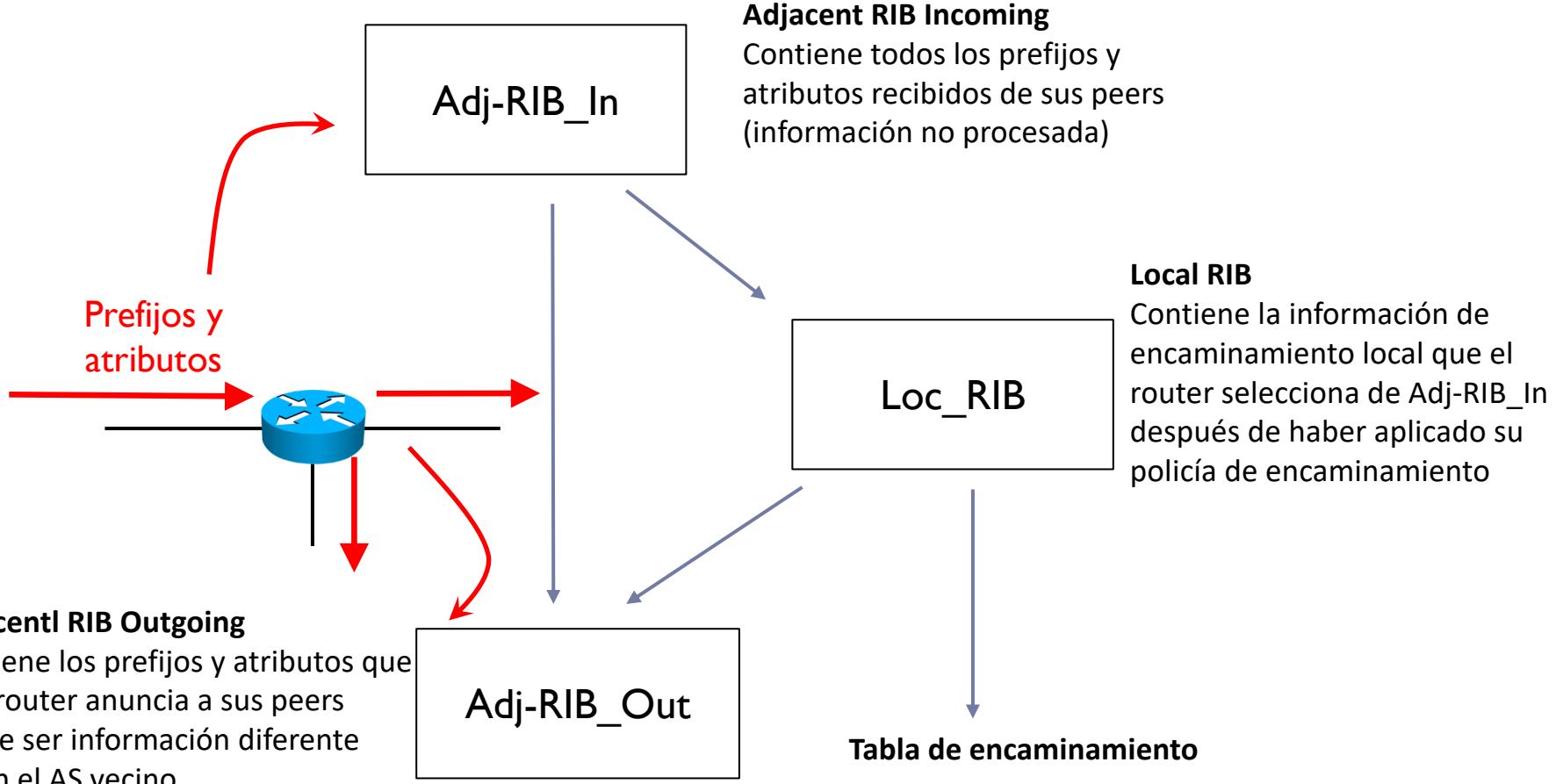
4. Establecimiento sesión BGP

▶ Estados



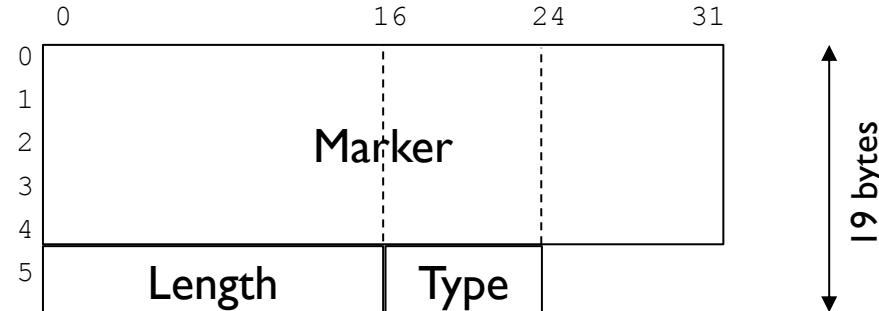
4. Bases de datos en BGP

- ▶ Un router BGP mantiene 3 bases de datos



4. Mensajes BGP

▶ Cabecera común



▶ Marker

- ▶ Seguridad
- ▶ Si primer mensaje o no hay seguridad, son todos 1
- ▶ En otros casos, se aplica la seguridad negociada entre los dos peers

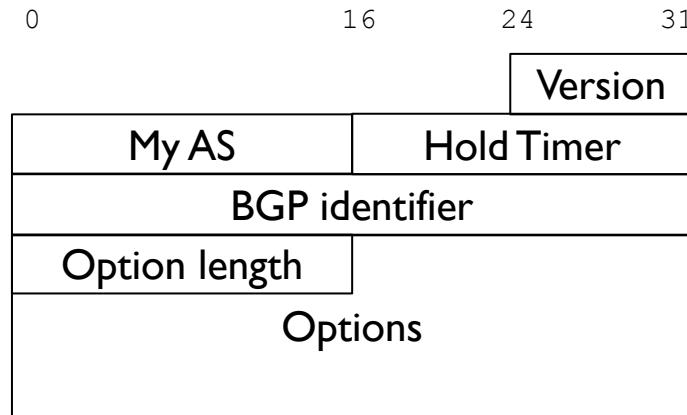
▶ Length

- ▶ Longitud de todo el mensaje BGP (cabecera + payload)

▶ Type

- ▶ 1. Open
- ▶ 2. Update
- ▶ 3. Notification
- ▶ 4. Keepalive

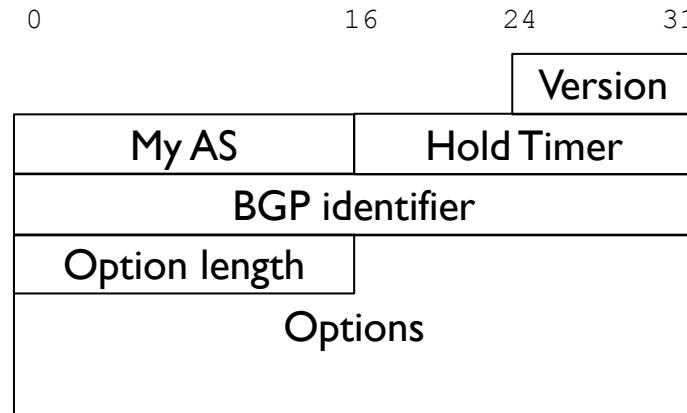
4. Mensaje OPEN



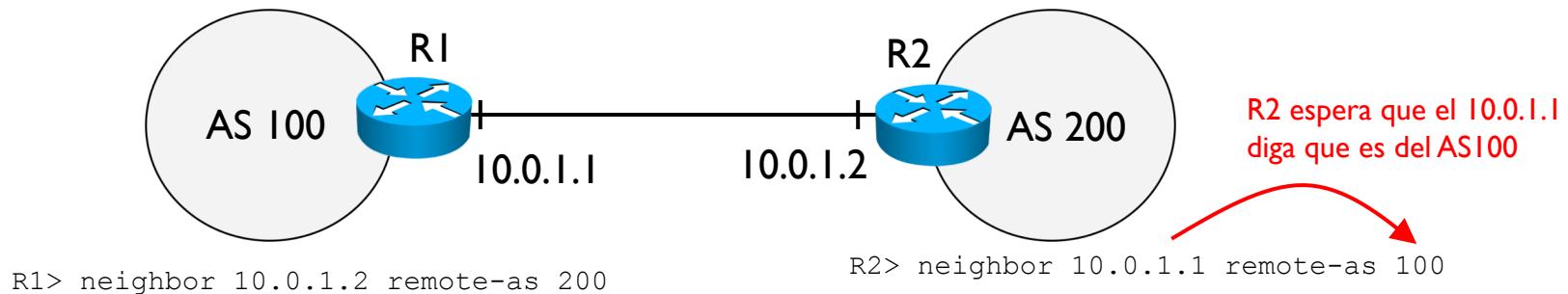
- ▶ Después de el establecimiento de la sesión TCP, los routers se envían un OPEN
- ▶ Los objetivos son
 - ▶ Identificarse
 - ▶ Negociar los parámetros del BGP



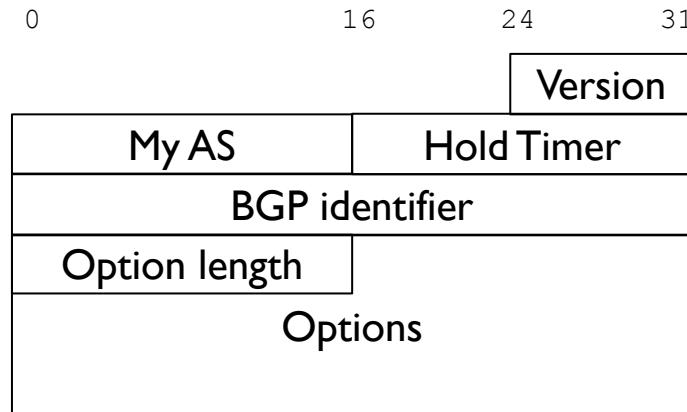
4. Mensaje OPEN



- ▶ Version: 4 (solo se usa esta versión actualmente)
- ▶ My AS:
 - ▶ El router envía el número de su AS
 - ▶ El receptor compara este número con el que se espera recibir de este vecino

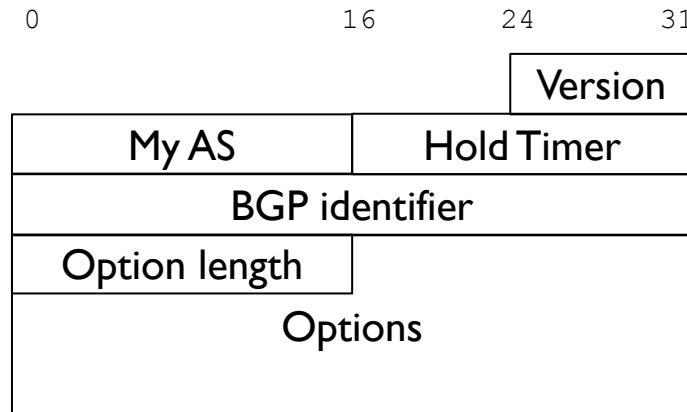


4. Mensaje OPEN



- ▶ **Hold Timer**
 - ▶ Especifica los segundos que el router quiere usar como Hold Timer
 - ▶ Es el tiempo máximo que un router puede esperar sin recibir mensajes BGP del otro peer (por defecto se usan 60 segundos)
 - ▶ Este Hold Timer se resetea cada vez que el router recibe un UPDATE o un KEEPALIVE
 - ▶ Generalmente se esperan n Hold Timer (por defecto 3) antes de considerar que ha pasado algo a la sesión BGP
 - ▶ Una vez pasados n Hold Timers, se cierra la sesión BGP que pasa a idle
 - ▶ Si los dos routers se envían 2 valores distintos, generalmente se usa el menor de los dos
 - ▶ Se puede configurar a 0, que significa que no se envían los mensajes KEEPALIVE

4. Mensaje OPEN



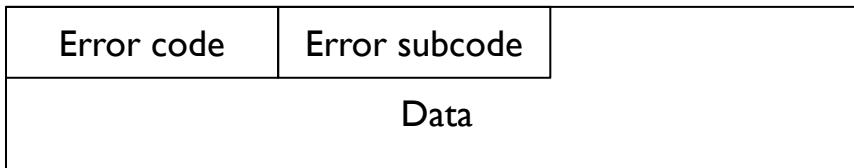
- ▶ **BGP identifier**
 - ▶ Es el RID del router
- ▶ **Option length**
 - ▶ Longitud del campo options
- ▶ **Options**
 - ▶ Contiene toda la información que un router quiere anunciar al otro sobre sus “capacidades” RFC 5492
 - ▶ Por ejemplo, métodos de autentificación y seguridad, mejoras como Comunidades, etc.

4. Mensaje KEEPALIVE

- ▶ Se usa para verificar la conectividad entre dos peers
- ▶ Se envía cada vez que expira el Hold Timer establecido durante la etapa OPEN
 - ▶ Recordar que el Hold Timer se reinicia también si se envía un UPDATE
- ▶ El mensaje solo consiste de la cabecera común de 19 bytes



4. Mensaje NOTIFICATION



- ▶ Si hay un error durante la sesión BGP, se usa este tipo de mensaje
- Al enviar este mensaje, se cierra la sesión BGP
- ▶ Error code proporciona información general
- ▶ Ejemplo de errores
 - ▶ Error code: error en la cabecera común BGP subcode: longitud incorrecta
 - ▶ Error code: error durante el proceso OPEN subcode: versión 3 vs versión 4
 subcode: AS no aceptable
 - ▶ Error code: error al procesar un UPDATE subcode: atributo no valido
 - ▶ Error code: Hold Timer expirado y no se han recibido mensajes
- ▶ Error code proporciona información general, error subcode proporciona detalles más específicos y el campo dato completa la información



4. Mensaje UPDATE

Withdraw routes length	2 bytes
Withdraw routes	variable
Total path attribute length	2 bytes
Path attribute	variable
Network Layer Reachability Information (NLRI)	variable

- ▶ Se usa para enviar información de encaminamiento entre peers
- ▶ Withdraw routes length
 - ▶ Contiene la longitud del siguiente campo que es variable y puede ser 0 (no se eliminan rutas)
- ▶ Withdraw routes
 - ▶ Un peer puede notificar al otro que algunos prefijos previamente notificado ya no son validos y hay que borrarlos



4. Mensaje UPDATE

Withdraw routes length	2 bytes
Withdraw routes	variable
Total path attribute length	2 bytes
Path attribute	variable
Network Layer Reachability Information (NLRI)	variable

▶ NLRI

- ▶ Contiene la lista de prefijos anunciados por este peer
- ▶ El número de prefijos está limitado por el tamaño máximo de un mensajes BGP (4096 bytes)
- ▶ Formato:
 - ▶ longitud prefijo – net-id del prefijo
 - ▶ Por ejemplo para informar del prefijo 147.83.0.0/16 → se envía 16 - 147.83
- ▶ La longitud de este campo no está especificada directamente pero se puede deducir por los otros campos

Longitud NLRI = Length – 19 bytes – 2 bytes – Longitud de withdraw routes – 2 bytes – Longitud de path attribute

Campo de la cabecera común	Longitud cabecera común	Longitud withdraw routes length	Valor de Withdraw routes length	Longitud withdraw routes length	Valor de Total path attribute Length
---	--	--	--	--	---

4. Mensaje UPDATE

Withdraw routes length	2 bytes
Withdraw routes	variable
Total path attribute length	2 bytes
Path attribute	variable
Network Layer Reachability Information (NLRI)	variable

- ▶ **Total path attribute length**
 - ▶ Contiene la longitud del siguiente campo que es variable (no puede ser 0 ya que hay atributos obligatorios)
- ▶ **Path attribute**
 - ▶ La gran ventaja de BGP es que funciona a través de políticas que se determinan usando atributos
 - ▶ Una política indica las preferencias a la hora de seleccionar una ruta (veremos más adelante)
 - ▶ Estos atributos son comunes a todos los prefijos que se ponen en NLRI
 - ▶ Si hay prefijos que no tienen los mismos atributos, hay que separarlos en UPDATE diferentes
 - ▶ Se pueden definir atributos estandar o definir de nuevos



4. Mensaje UPDATE

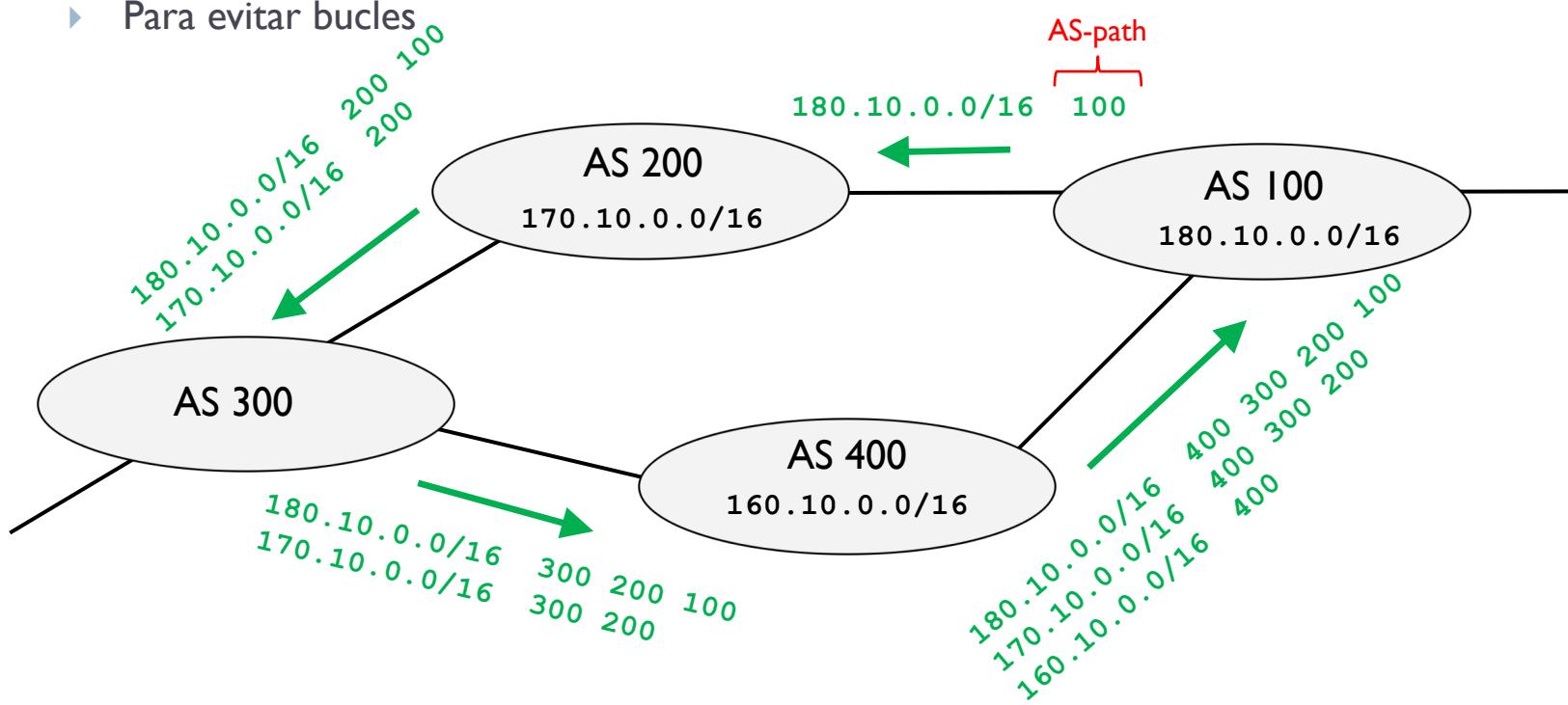
Withdraw routes length	2 bytes
Withdraw routes	variable
Total path attribute length	2 bytes
Path attribute	variable
Network Layer Reachability Information (NLRI)	variable

- ▶ Path attribute
- ▶ Estandar
 - ▶ ORIGEN
 - ▶ AS-PATH
 - ▶ NEXT HOP
 - ▶ MULTI-EXIT DISCRIMINATOR
 - ▶ LOCAL-PREFERENCE
 - ▶ AGGREGATOR
- ▶ Propietarios
 - ▶ Weight (CISCO)



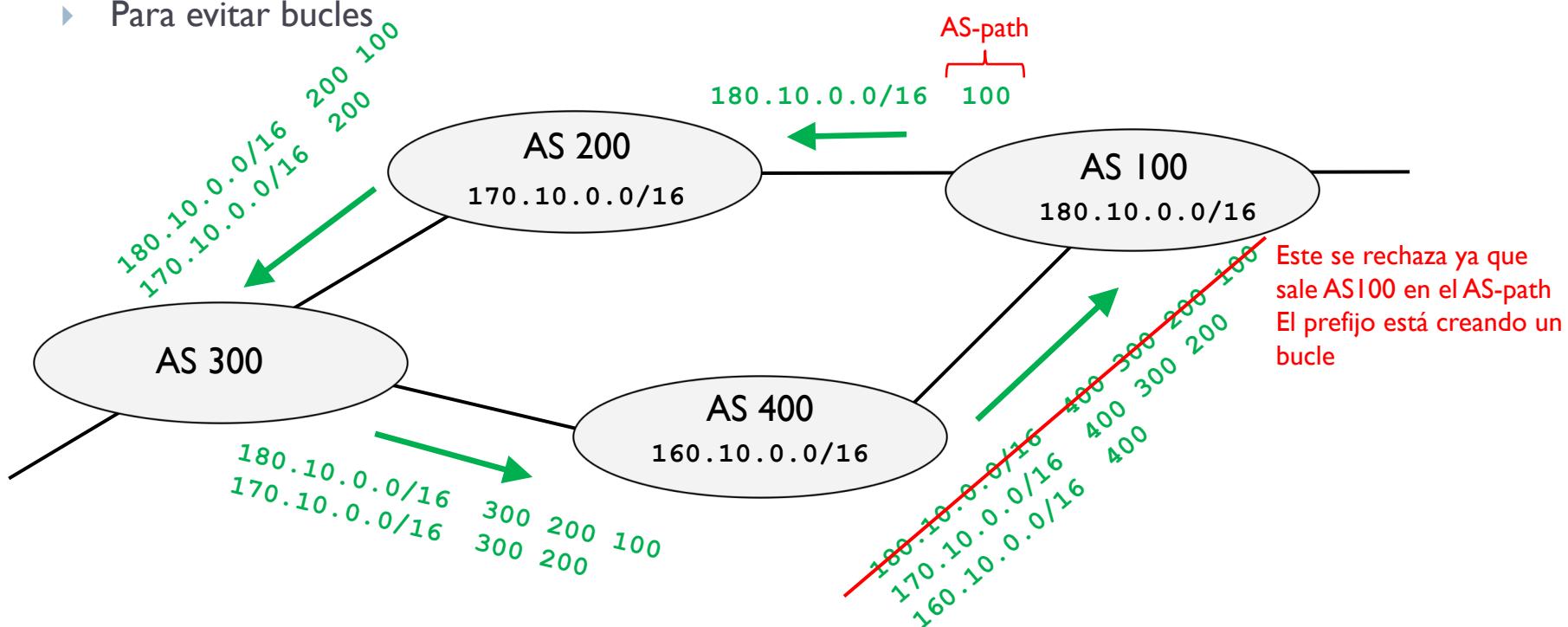
4. Atributos: AS-PATH

- ▶ Obligatorio
- ▶ Secuencia de ASN por donde ha pasado un prefijo
- ▶ Objetivos
 - ▶ Para evitar bucles



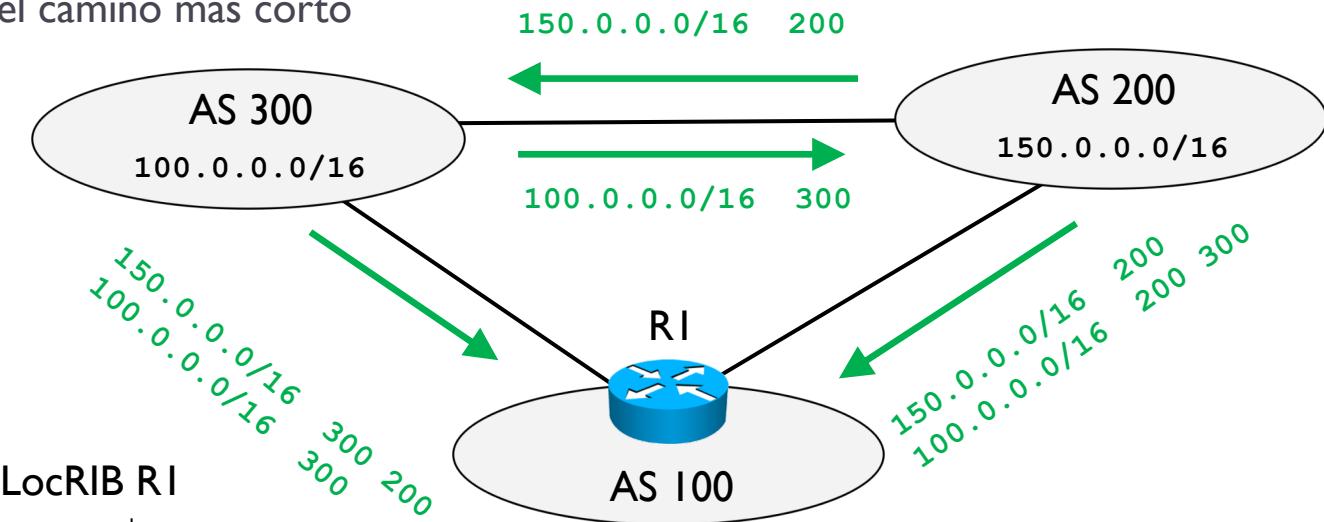
4. Atributos: AS-PATH

- ▶ Obligatorio
- ▶ Secuencia de ASN por donde ha pasado un prefijo
- ▶ Objetivos
 - ▶ Para evitar bucles



4. Atributos: AS-PATH

- ▶ Obligatorio
- ▶ Secuencia de ASN por donde ha pasado un prefijo
- ▶ Objetivos
 - ▶ Para evitar bucles
 - ▶ Para determinar el camino más corto

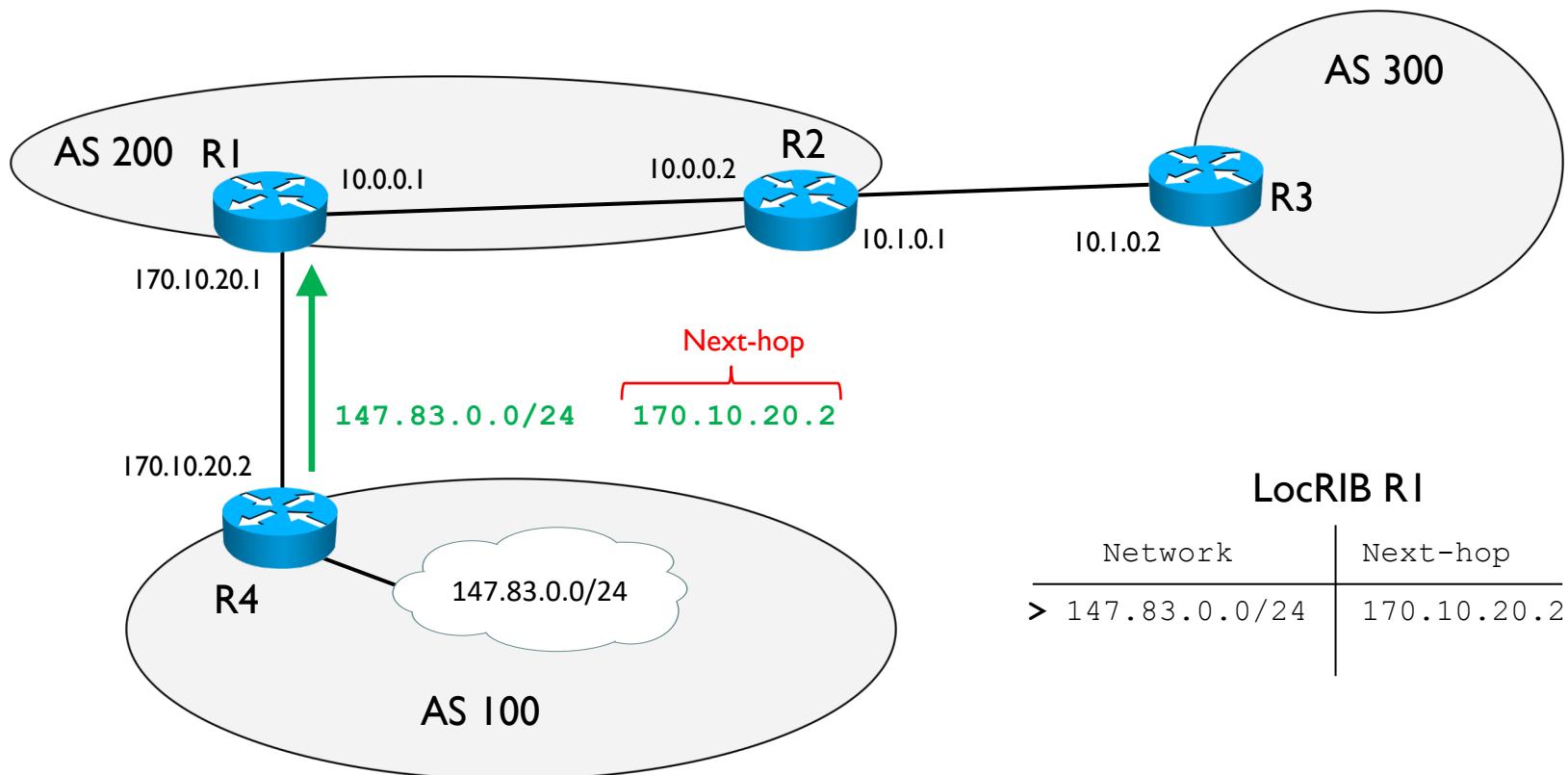


Network	AS-path
> 100.0.0.0/16	200
	200 300
> 150.0.0.0/16	300 200
>	300

BGP mantiene múltiples entradas
Para la tabla de encaminamiento, elige la “mejor” ruta y la marca con >
Si no hay otros atributos, elige la ruta con el AS-path más corto

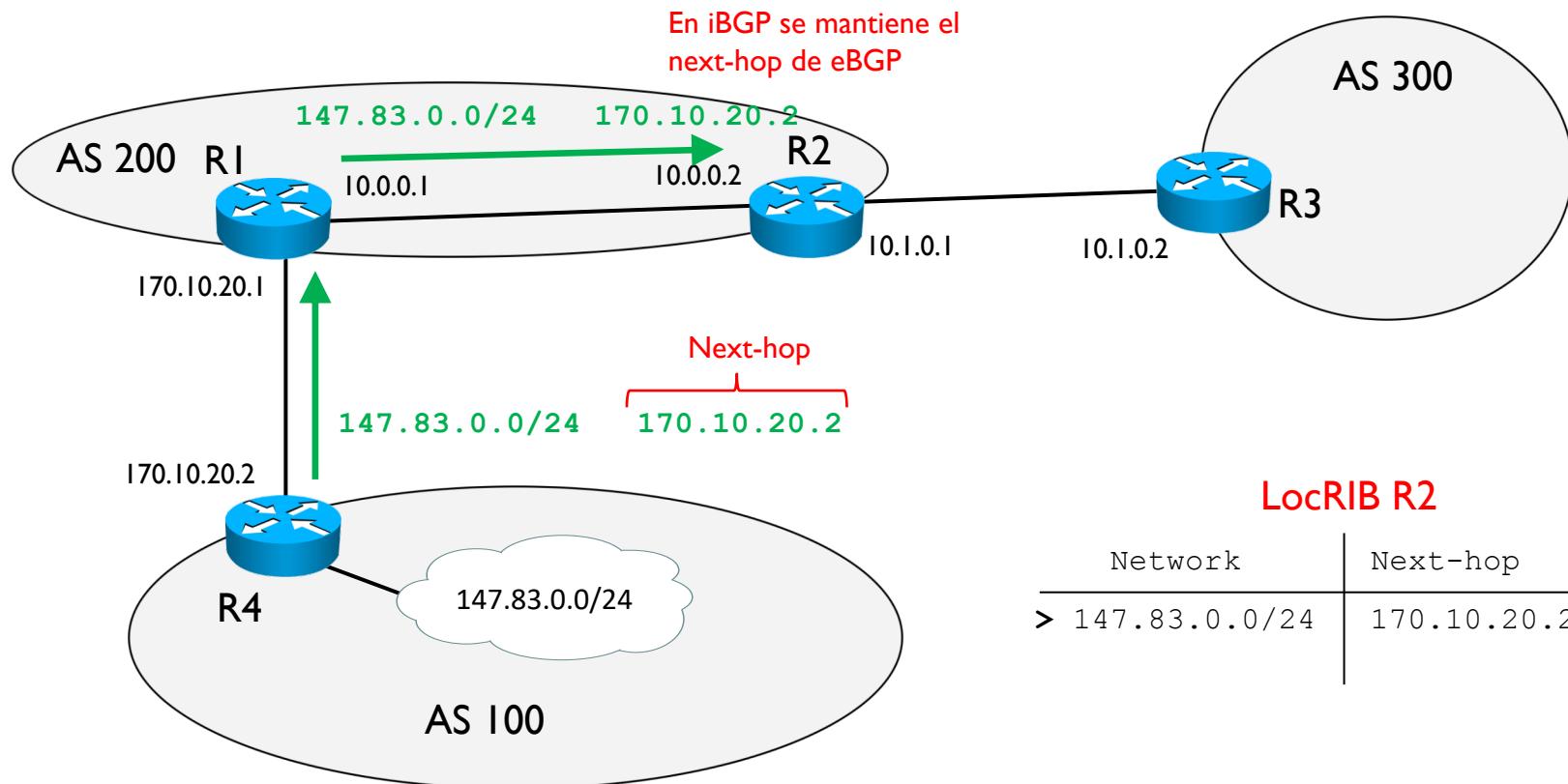
4. Atributos: next-hop

- ▶ Obligatorio
- ▶ Indica la @IP del router que hace de gateway entre AS



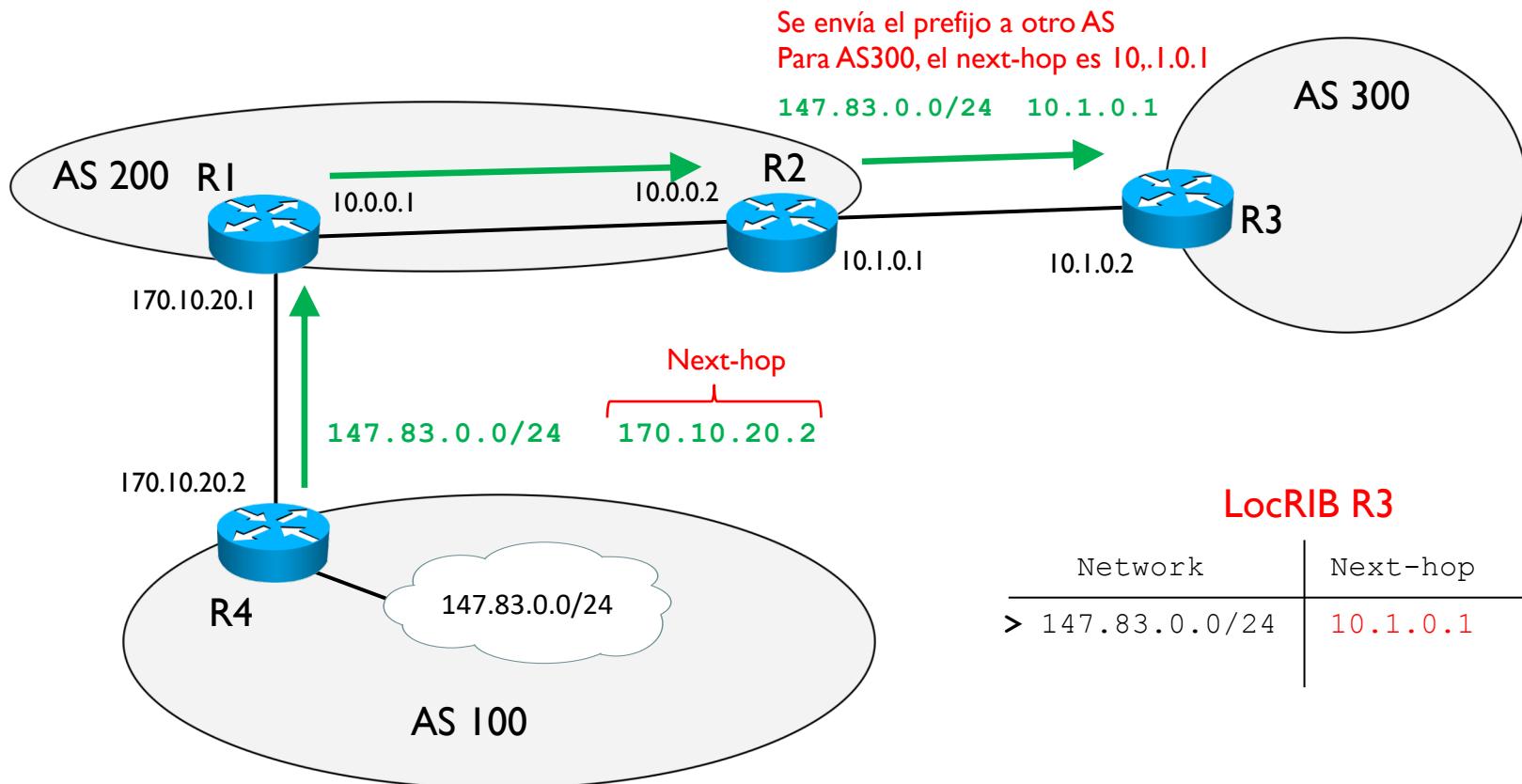
4. Atributos: next-hop

- ▶ Obligatorio
- ▶ Indica la @IP del router que hace de gateway entre AS



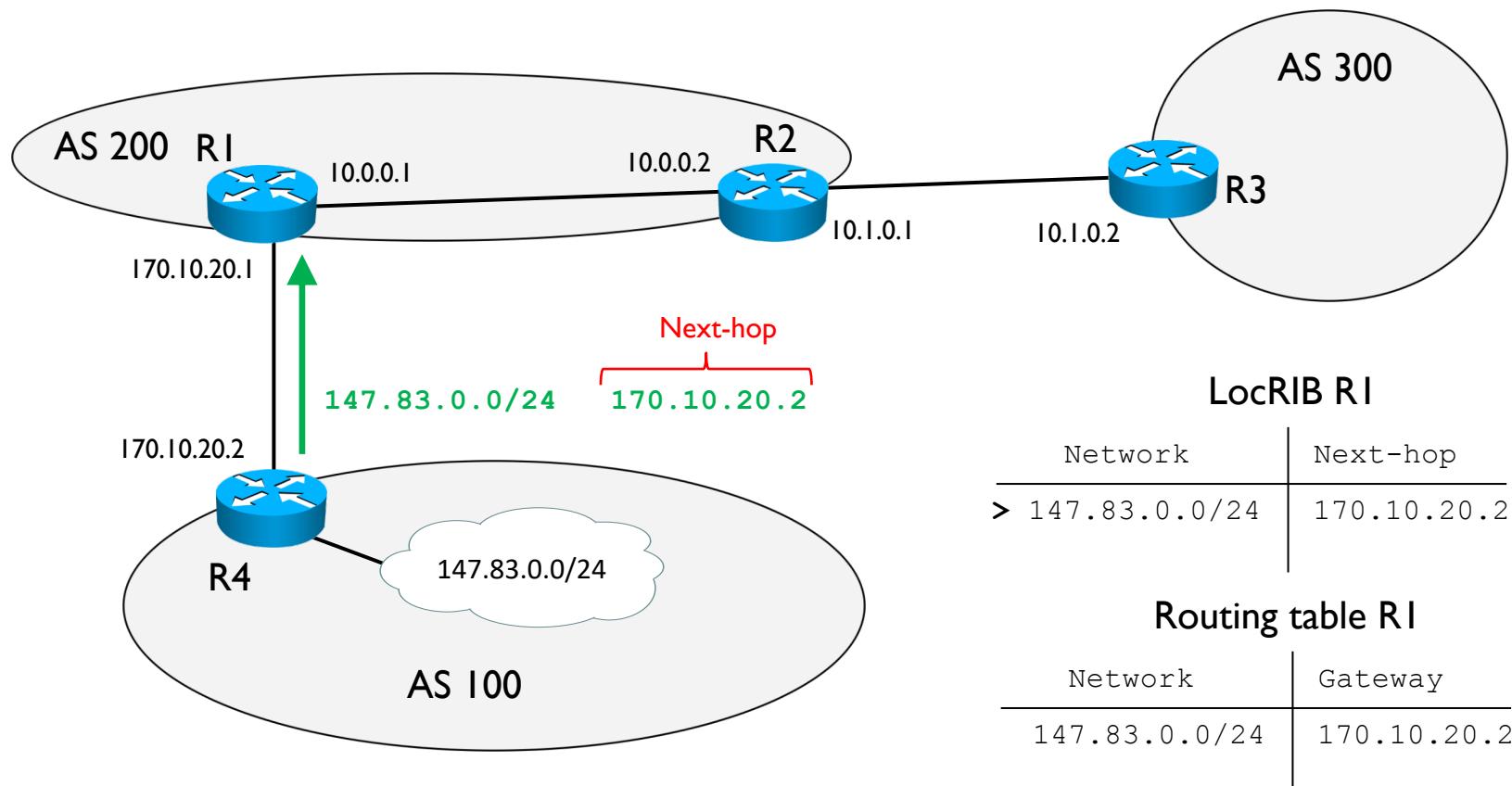
4. Atributos: next-hop

- ▶ Obligatorio
- ▶ Indica la @IP del router que hace de gateway entre AS



4. Atributos: next-hop

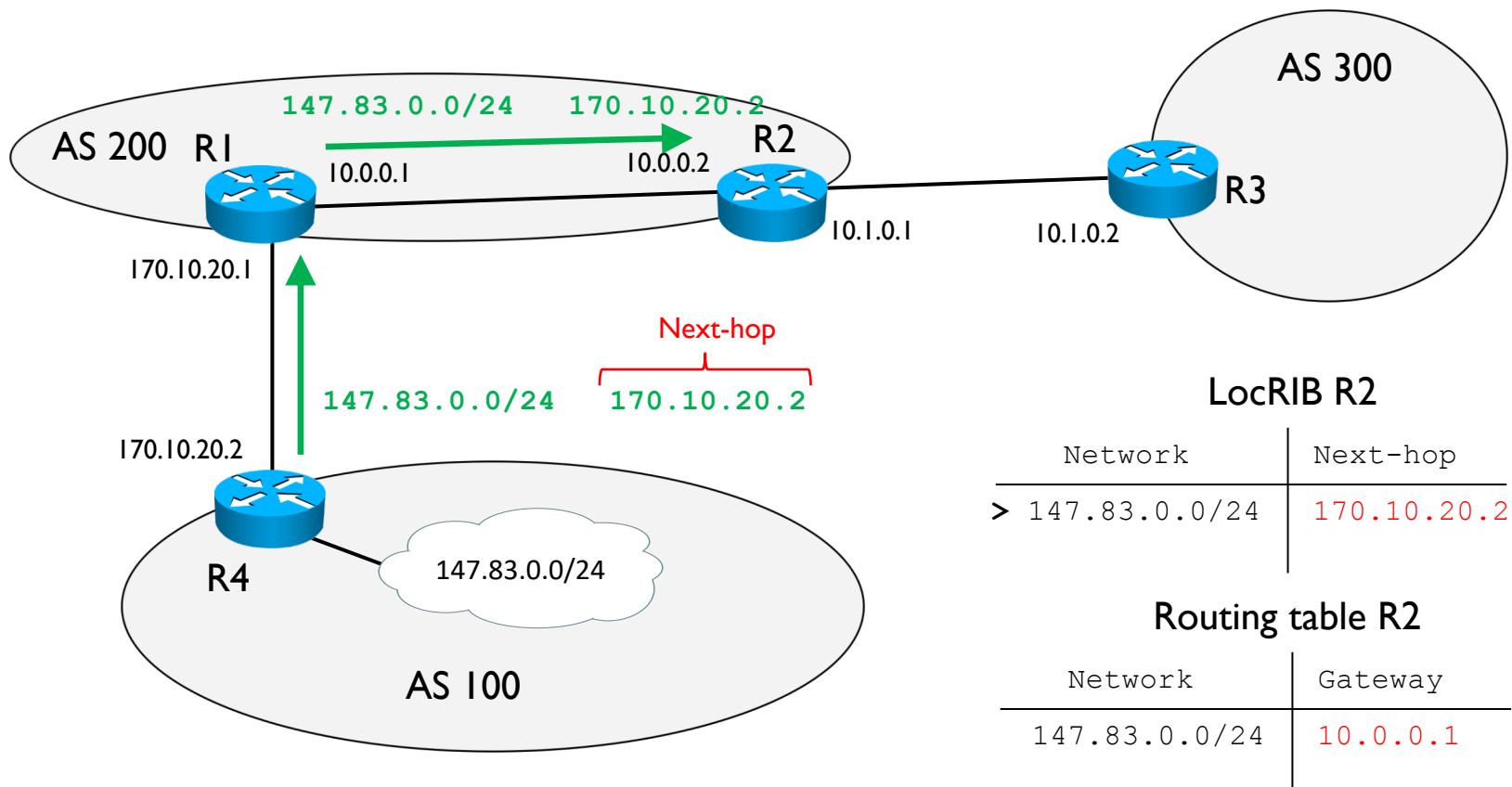
- ▶ No confundir gateway de una tabla de encaminamiento con next-hop
- ▶ Gateway es a nivel de routers, next-hop a nivel de AS



▶ Para RI, gateway y next-hop coinciden

4. Atributos: next-hop

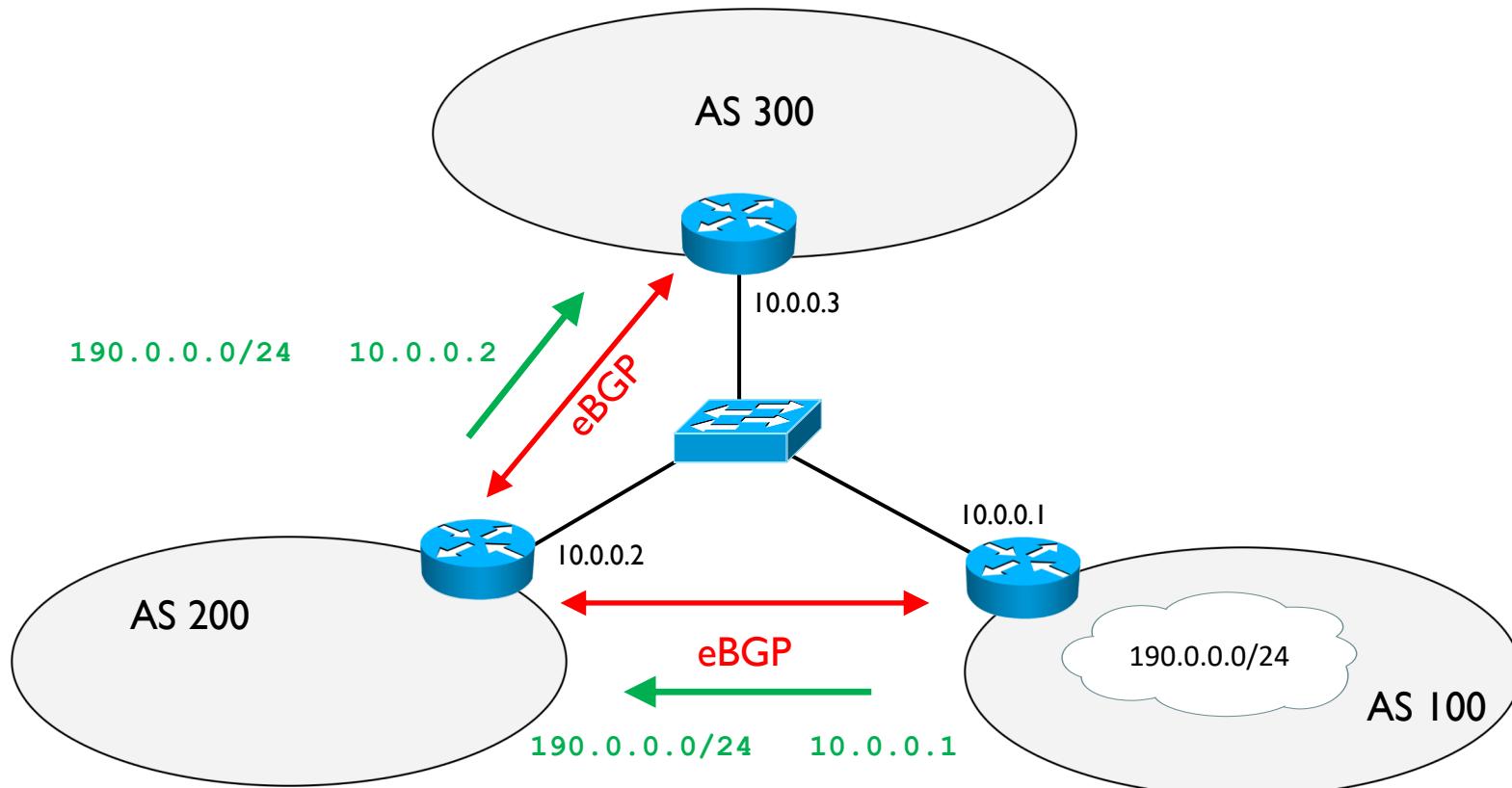
- ▶ No confundir gateway de una tabla de encaminamiento con next-hop
- ▶ Gateway es a nivel de routers, next-hop a nivel de AS



Para R2, gateway y next-hop NO coinciden

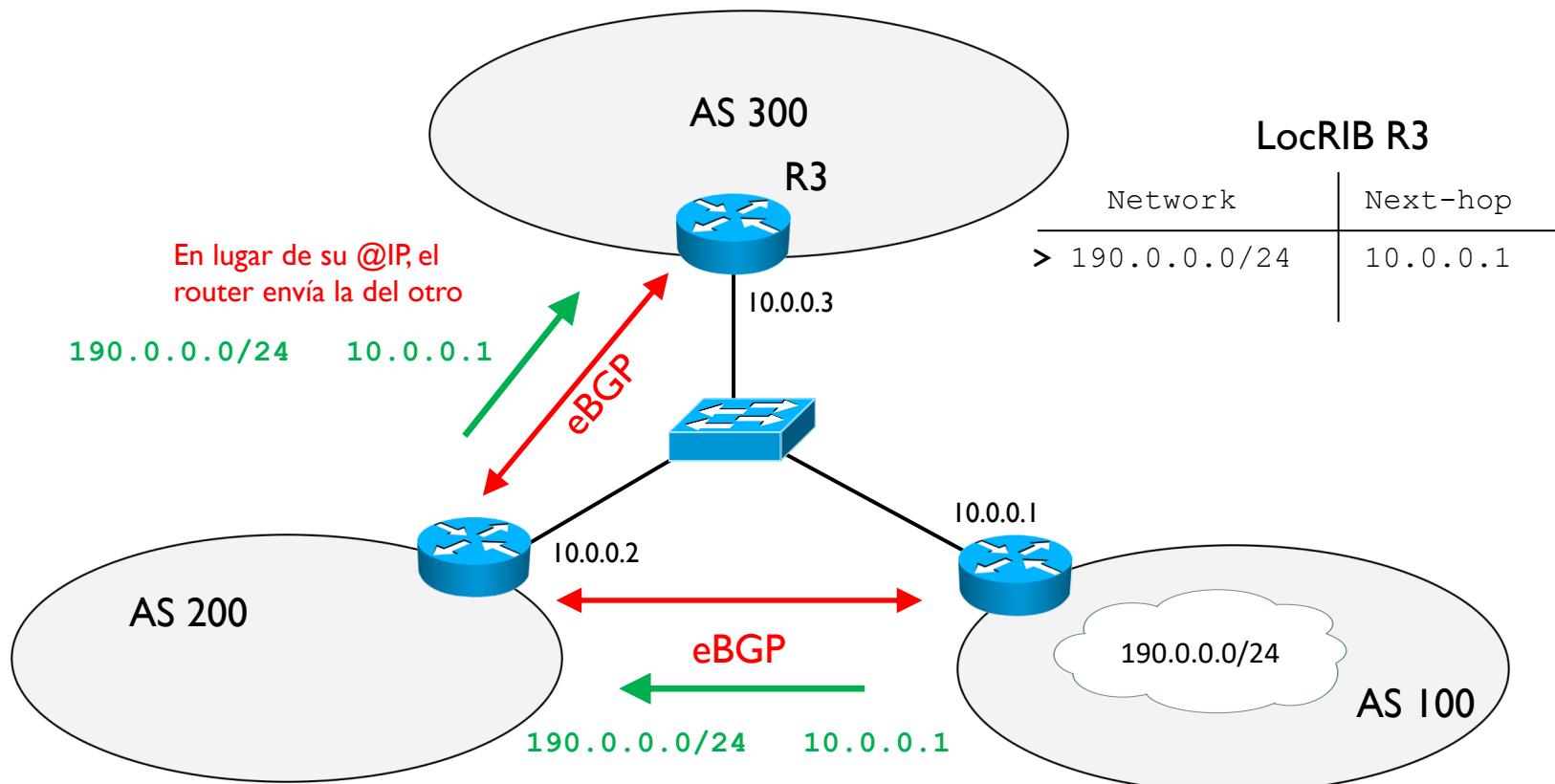
4. Next-hop third-party

- ▶ Se usa para evitar procesar información inútilmente



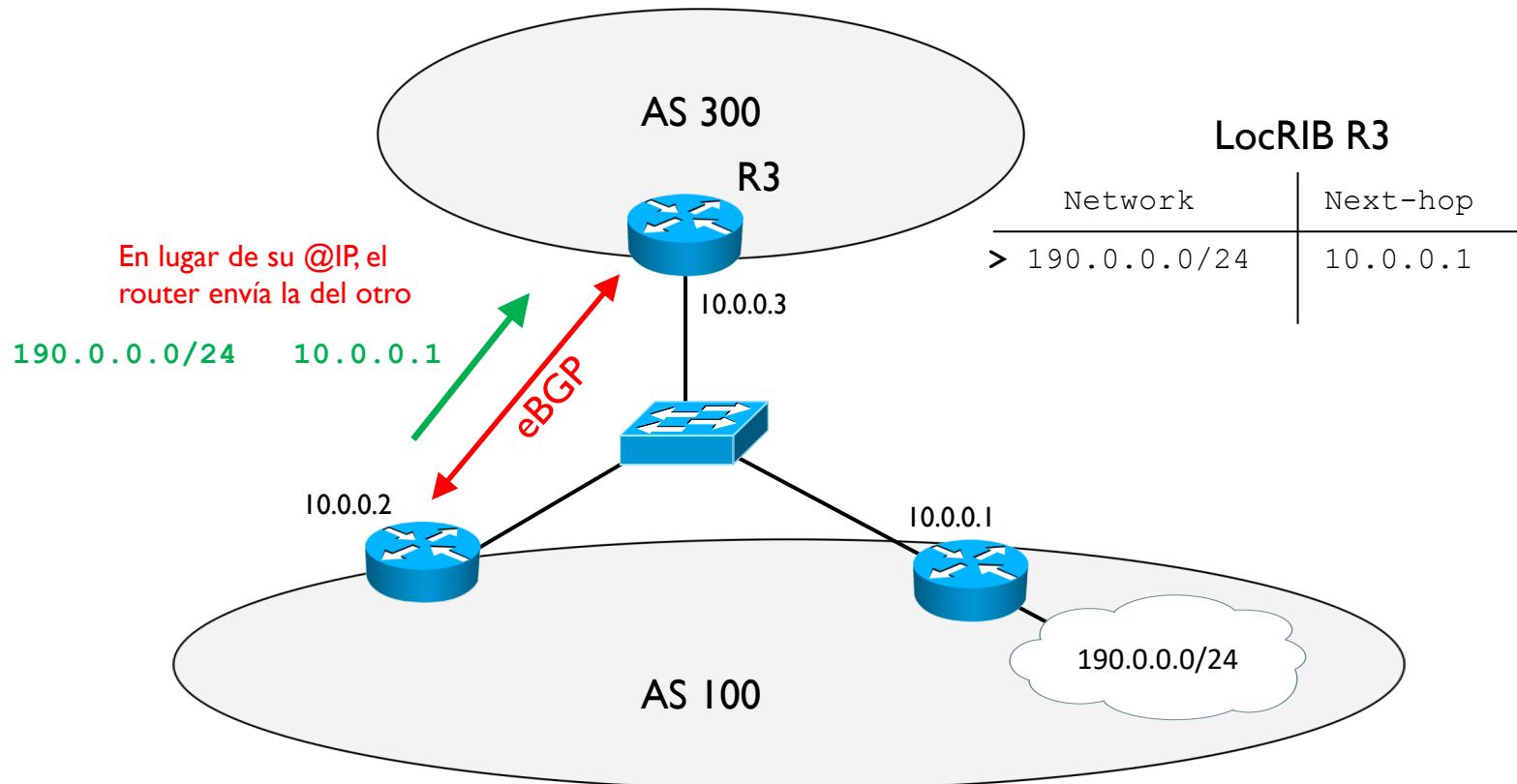
4. Next-hop third-party

- ▶ Se usa para evitar procesar información inútilmente



4. Next-hop backdoor

- ▶ Se usa para desacoplar el router BGP del que procesa datagramas



4. Atributos: ORIGIN

- ▶ Obligatorio e histórico
- ▶ Determina como se ha aprendido un prefijo (el origen del prefijo)
 - ▶ IGP: aprendido a partir de un encaminamiento interno dinámico como RIP o OSPF
 - ▶ EGP: aprendido del protocolo EGP (protocolo externo para inter-AS que se usaba al principio conjuntamente con BGP, hoy en día ya no se usa)
 - ▶ Incompleto: aprendido de otro protocolo o el origen no se quiere anunciar (generalmente usado cuando se ha aprendido un prefijo de una ruta estática)
- ▶ En CISCO, suele aparecer en la tabla Loc_RIB al final del AS-path
 - ▶ IGP: se indica con i
 - ▶ EGP: se indica con e
 - ▶ Incompleto: se indica con ?



4. Atributos: AGGREGATOR

- ▶ Opcional
- ▶ Cuando un router BGP agrega prefijos, se puede usar este atributo para indicar en que AS se ha hecho esta agregación y que router RID lo ha hecho
- ▶ Es optativo notificar esta atributo, es decir se puede hacer la agregación y no usar el atributo para notificarlo a los demás routers



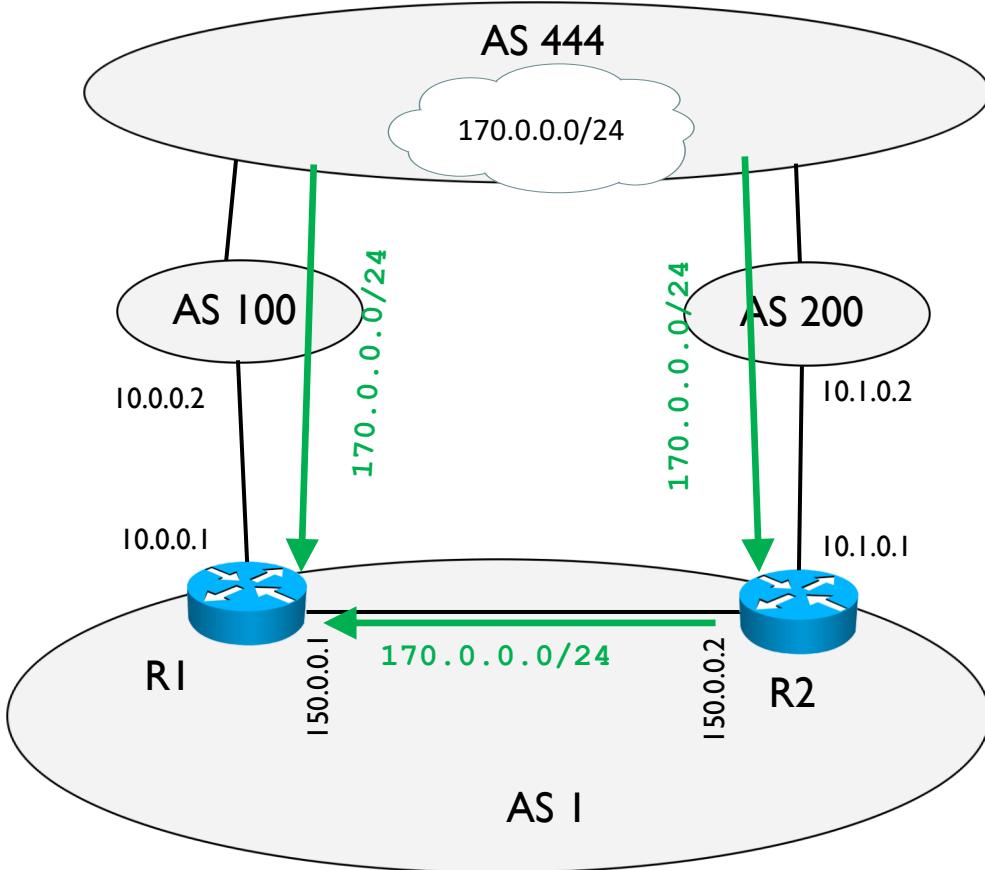
4. Atributos: LOCAL PREFERENCE

- ▶ Opcional
- ▶ Si no se usa, tiene 100 como valor por defecto
- ▶ Se usa para manipular la selección del mejor camino
- ▶ Se elige la ruta con el local preference más alto
- ▶ Los local preference tienen significado local, es decir interno al AS
 - ▶ No se anuncian por eBGP
 - ▶ Se anuncian por iBGP



4. Atributos: LOCAL PREFERENCE

▶ Ejemplo

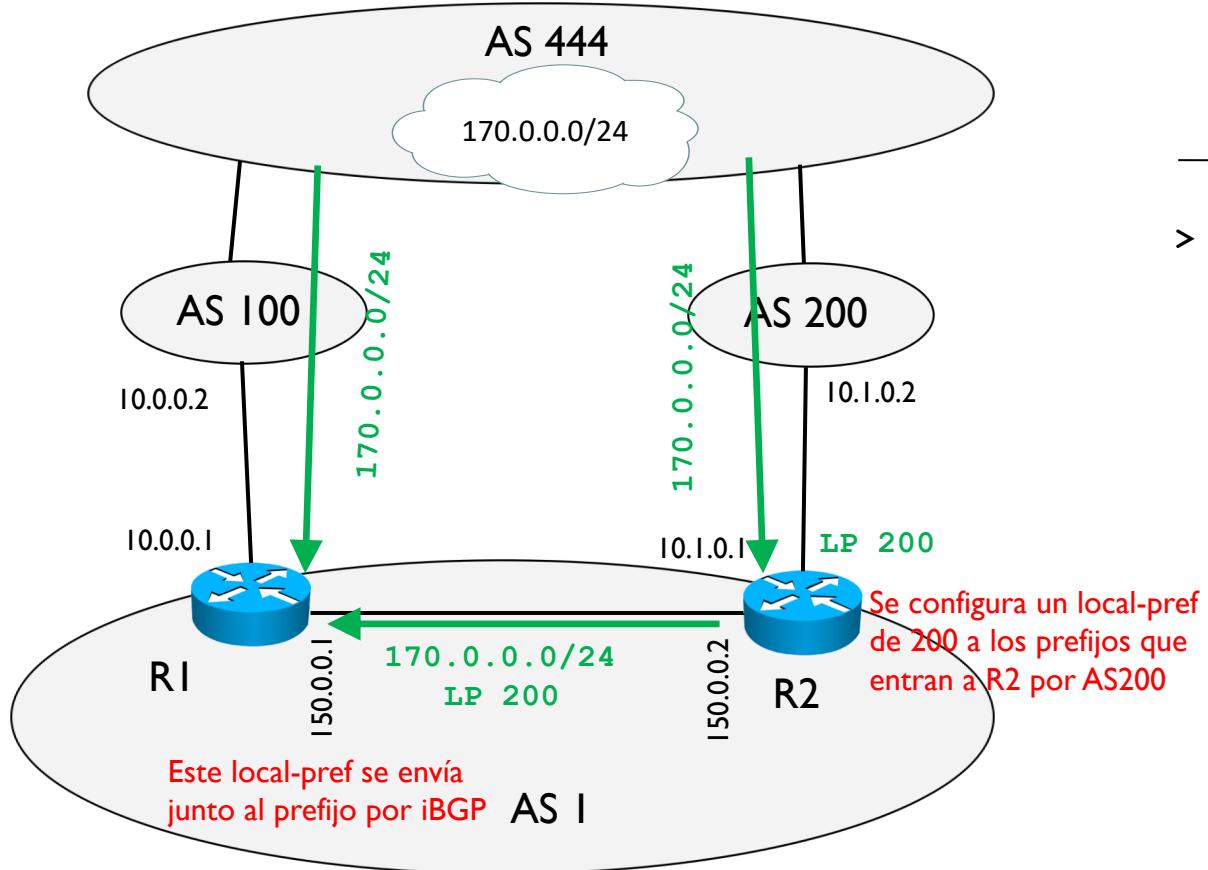


LocRIB RI		
Network	Next-hop	Local-pref
170.0.0.0/24	10.0.0.2	100
	10.1.0.2	100

¿Cuál es mejor?

4. Atributos: LOCAL PREFERENCE

Ejemplo



LocRIB RI

Network	Next-hop	Local-pref
170.0.0.0/24	10.0.0.2	100
	10.1.0.2	200

R1 elige la ruta por R2 que va al AS200

Routing table RI

Network	Gateway
170.0.0.0/24	150.0.0.2

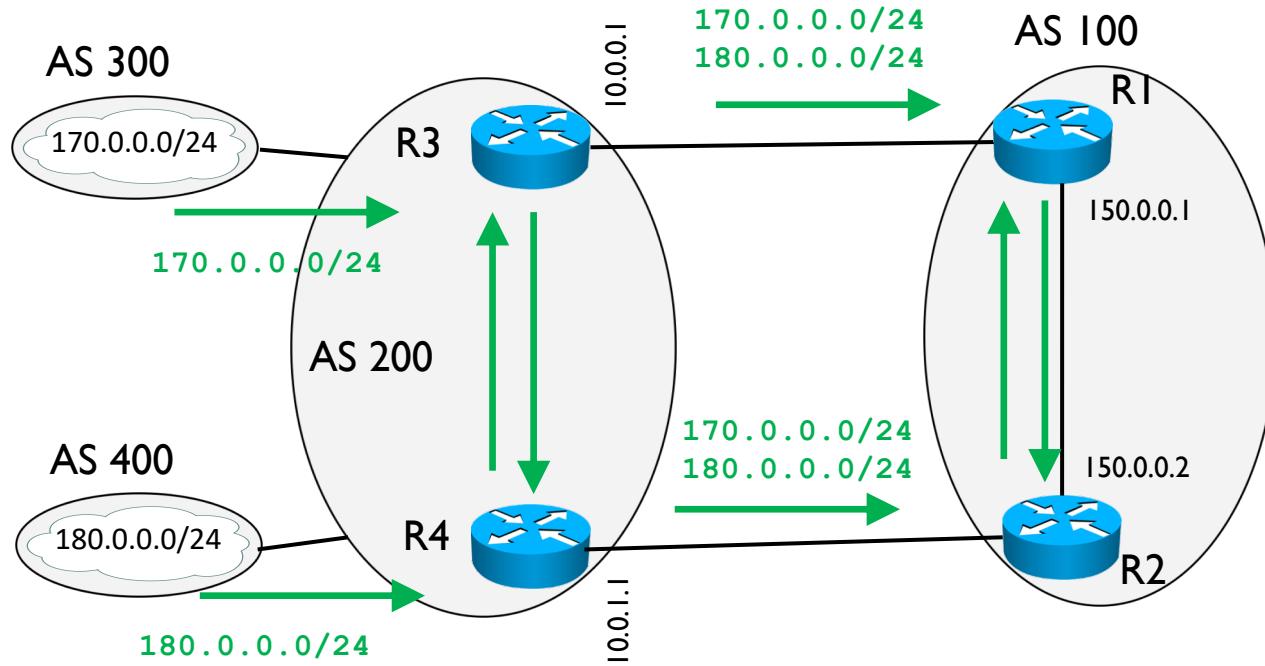
4. Atributos: MULTI EXIT DISCRIMINATOR (MED)

- ▶ Opcional
- ▶ Si no se usa, tiene 0 como valor por defecto
- ▶ Su valor se llama metric
- ▶ Se elige la ruta con el metric más bajo
- ▶ Un AS puede indicar al AS vecino cual enlace sería mejor usar entre varios disponibles
 - ▶ Libertad por parte del vecino de usar o rechazar esta sugerencia
- ▶ Los metric solo se transmiten entre dos AS vecinos



4. Atributos: MULTI EXIT DISCRIMINATOR (MED)

▶ Ejemplo

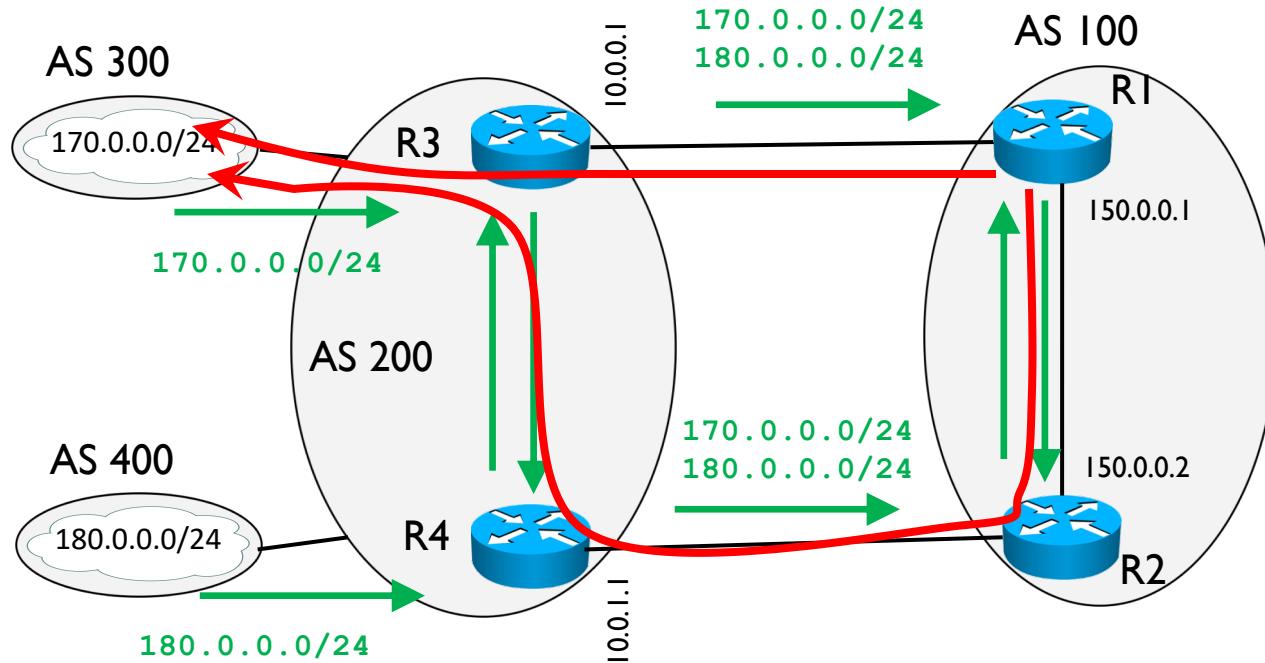


Network	Next-hop	metric
170.0.0.0/24	10.0.0.1	0
	10.0.1.1	0
180.0.0.0/24	10.0.0.1	0
	10.0.1.1	0

LocRIB RI

4. Atributos: MULTI EXIT DISCRIMINATOR (MED)

Ejemplo

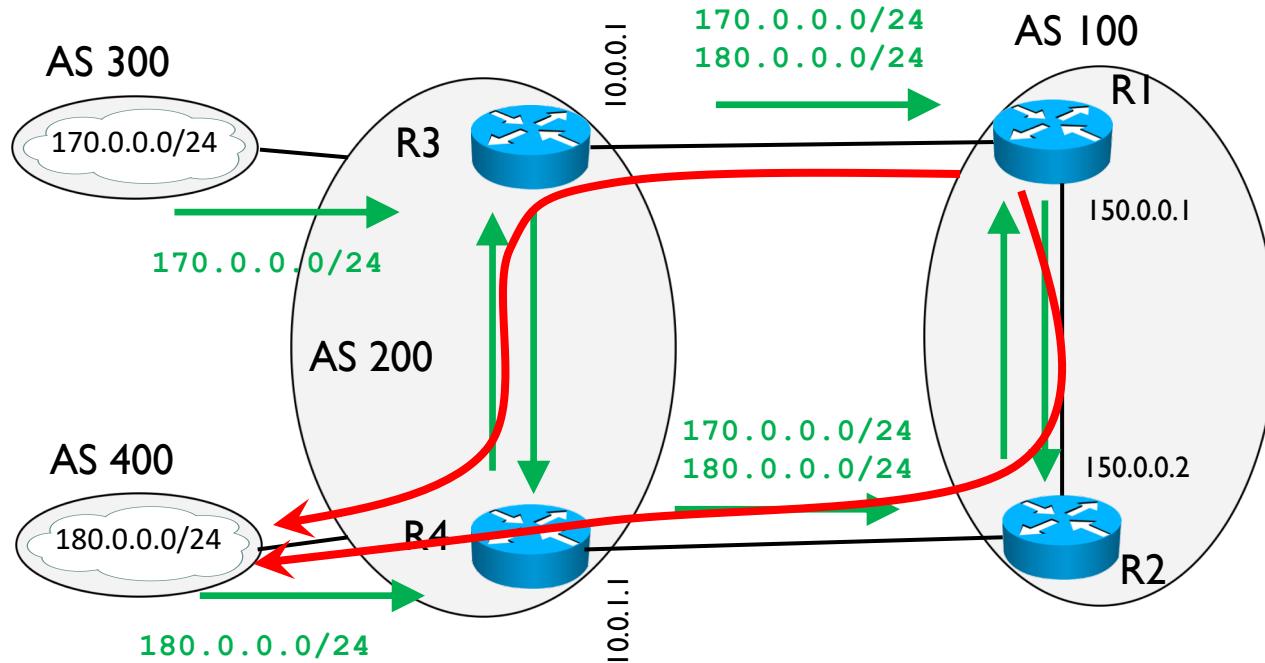


Network	Next-hop	metric
170.0.0.0/24	10.0.0.1	0
	10.0.1.1	0
180.0.0.0/24	10.0.0.1	0
	10.0.1.1	0

¿Cuál es mejor para llegar a 170.0.0.0/24?

4. Atributos: MULTI EXIT DISCRIMINATOR (MED)

Ejemplo

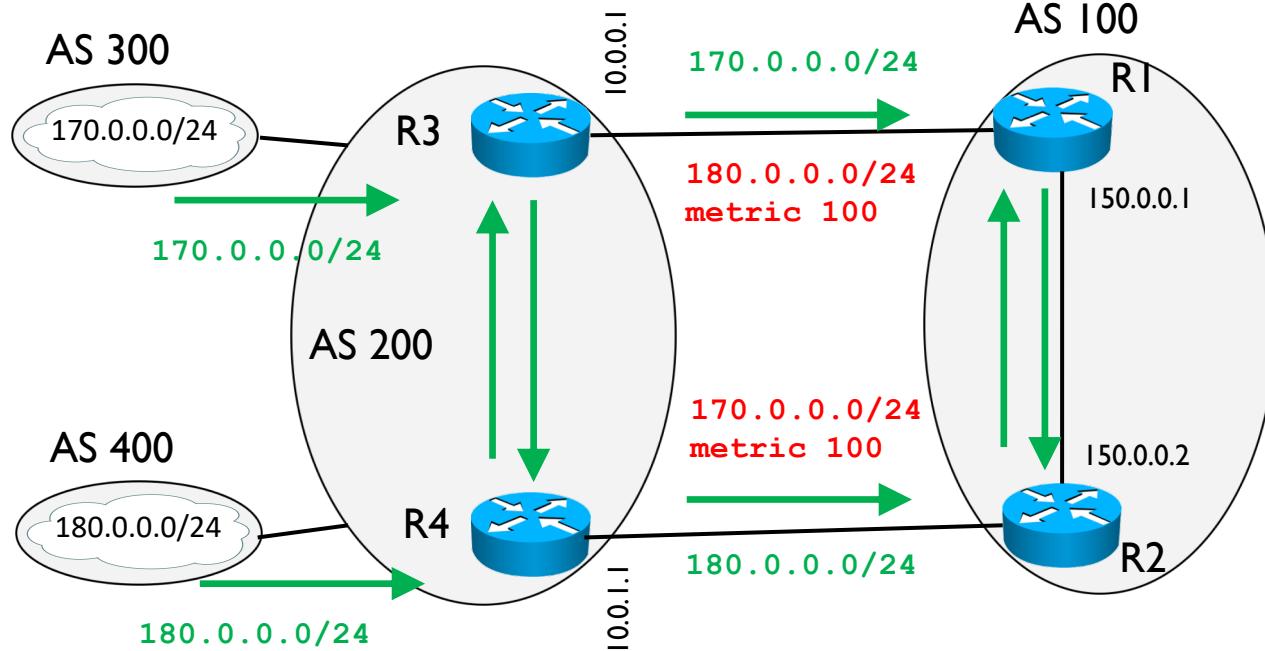


Network	Next-hop	metric
170.0.0.0/24	10.0.0.1	0
	10.0.1.1	0
180.0.0.0/24	10.0.0.1	0
	10.0.1.1	0

¿Cuál es mejor para
llegar a 180.0.0.0/24?

4. Atributos: MULTI EXIT DISCRIMINATOR (MED)

▶ Ejemplo



El AS200 puede sugerir al AS100 las mejores rutas

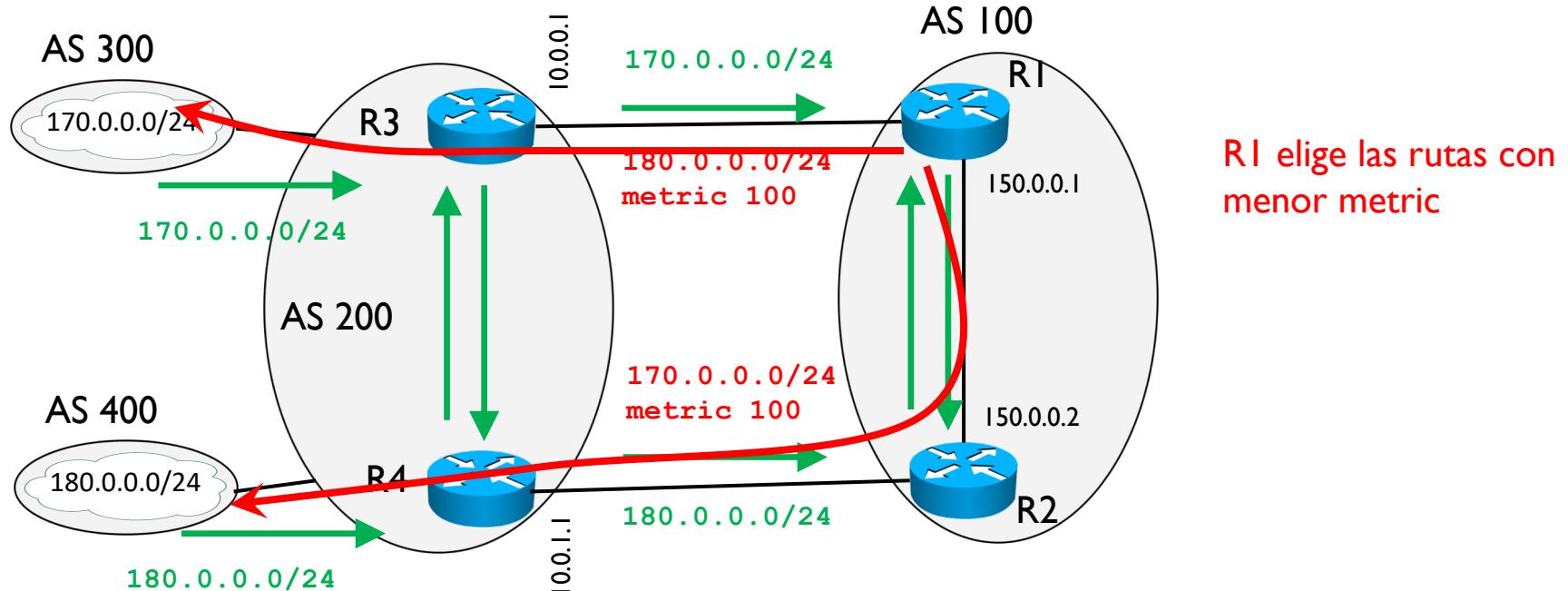
En R3, se configura un metric de 100 al prefijo 180.0.0.0/24 de salida hacia R1

En R4, se configura un metric de 100 al prefijo 170.0.0.0/24 de salida hacia R2

Network	Next-hop	metric
> 170.0.0.0/24	10.0.0.1	0
	10.0.1.1	100
180.0.0.0/24	10.0.0.1	100
>	10.0.1.1	0

4. Atributos: MULTI EXIT DISCRIMINATOR (MED)

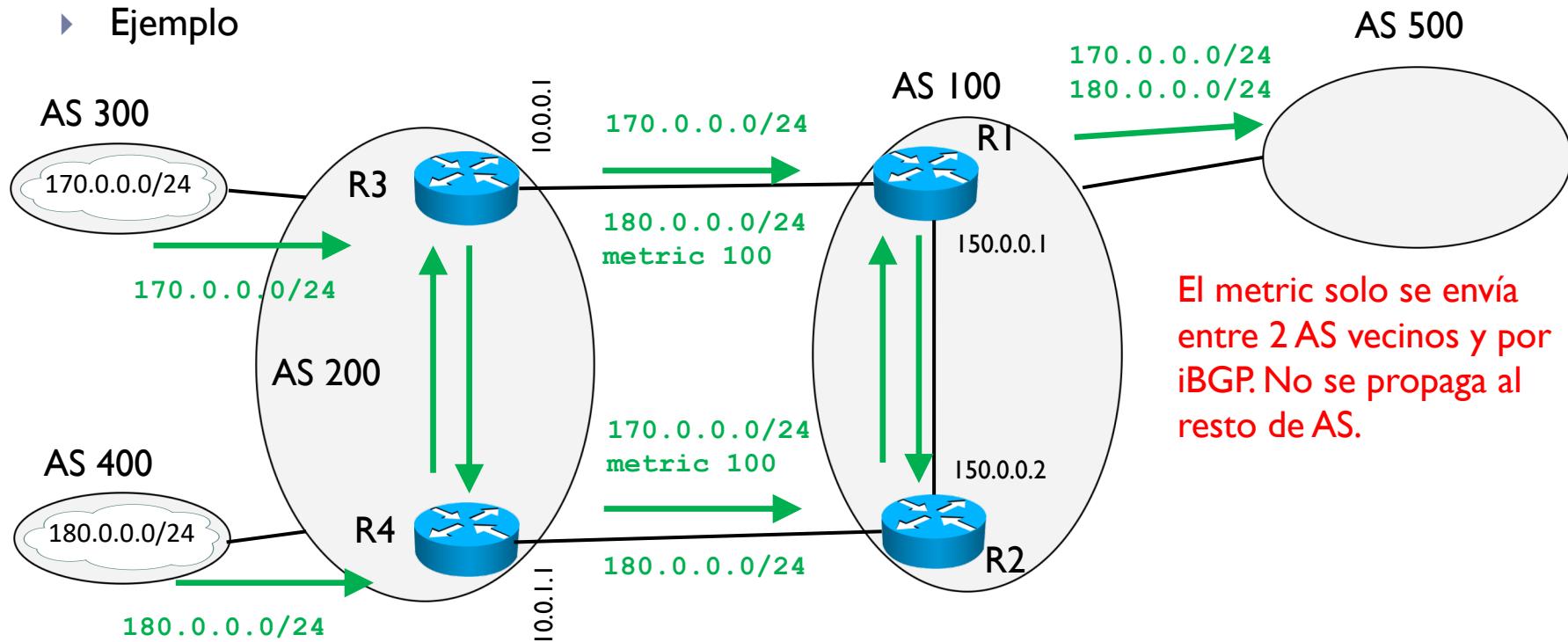
Ejemplo



LocRIB RI			Routing table R1		
Network	Next-hop	metric	Network	Gateway	
> 170.0.0.0/24	10.0.0.1	0	170.0.0.0/24	10.0.0.1	
	10.0.1.1	100	180.0.0.0/24	150.0.0.2	
180.0.0.0/24	10.0.0.1	100			
>	10.0.1.1	0			

4. Atributos: MULTI EXIT DISCRIMINATOR (MED)

Ejemplo



Network	Next-hop	metric
> 170.0.0.0/24	10.0.0.1	0
	10.0.1.1	100
180.0.0.0/24	10.0.0.1	100
>	10.0.1.1	0



Xarxes de computadors II