

Prácticas recomendadas de BGP



Configurando BGP



Dónde empezar?

IOS – Prácticas recomendadas

- Debería empezar con los siguientes comandos BGP como plantilla básica:

```
router bgp 64511  
  bgp deterministic-med  
  distance bgp 200 200 200  
  no synchronization  
  no auto-summary
```

← ASN público

← Distancia de egp e igp igual

- Si tiene “peers” que no sean unicast IPv4
 - no bgp default ipv4-unicast
 - También muy importante y mandatorio

Prácticas recomendadas en Cisco IOS

- ❑ BGP en Cisco IOS es **permisivo** por defecto
- ❑ Configurar BGP sin utilizar filtros significa que:
 - Todas las mejores trayectorias del enrutador local se pasan a los vecinos
 - Todas las rutas anunciadas por los vecinos son recibidas por el enrutador local
 - Puede tener consecuencias desastrosas
- ❑ **Lo mejor es asegurarse de que cada vecino eBGP tenga filtros de entrada y salida:**

```
router bgp 64511
  neighbour 1.2.3.4 remote-as 64510
  neighbour 1.2.3.4 prefix-list as64510-in in
  neighbour 1.2.3.4 prefix-list as64510-out out
```

Para qué es bueno BGP?

Para qué no es bueno un IGP?

BGP versus OSPF/ISIS

- Protocolos de enrutamiento internos (IGPs)
 - Ejemplos son ISIS y OSPF
 - Para transportar direcciones de **infraestructura**
 - **NO** para transportar prefijos de Internet o de clientes
 - La meta de diseño es **minimizar** el número de prefijos IGP para mejorar la convergencia rápida y la sostenibilidad

BGP versus OSPF/ISIS

- ❑ BGP utilizado internamente (iBGP) and externamente (eBGP)
- ❑ iBGP para transportar:
 - Algunos/todos los prefijos de Internet a través de la dorsal
 - Prefijos de los clientes
- ❑ eBGP para:
 - Intercambiar prefijos con otros ASs
 - Implementar políticas de enrutamiento

BGP versus OSPF/ISIS

□ NUNCA:

- Distribuya prefijos BGP dentro de un IGP
- Distribuya prefijos de un IGP dentro de BGP
- Utilice un IGP para transportar prefijos de clientes

□ **SU RED NO SERÁ SOSTENIBLE**

Agregación



Agregación

- Agregación significa anunciar el bloque IP recibido del RIR a los otros ASs conectados a su red
- Los sub-prefijos de este bloque pueden ser:
 - Usados internamente en la red
 - Anunciados a otros ASs para ayudar con *multihoming*
- Desafortunadamente demasiada gente todavía piensa en clases C, resultando en la proliferación de /24s en la tabla de enrutamiento de Internet

Configurando Agregación– Cisco IOS

- ❑ ISP tiene bloque 101.10.0.0/19
- ❑ Para poner en BGP como agregado:
 - `router bgp 64511`
 - `network 101.10.0.0 mask 255.255.224.0`
 - `ip route 101.10.0.0 255.255.224.0 null0`
- ❑ La ruta estática es una ruta “pull up”
 - Los prefijos más específicos dentro de este bloque aseguran la conectividad a los clientes del ISP
 - Búsqueda de “longest match”

Agregación

- ❑ El bloque de direcciones se debe anunciar a Internet como un agregado
- ❑ Los sub-prefijos del bloque no deberían ser anunciados a la Internet a menos que se esté haciendo ingeniería de tráfico
- ❑ La agregación debe ser generada internamente
 - No en las fronteras de la red!

Anunciando un agregado – Cisco IOS

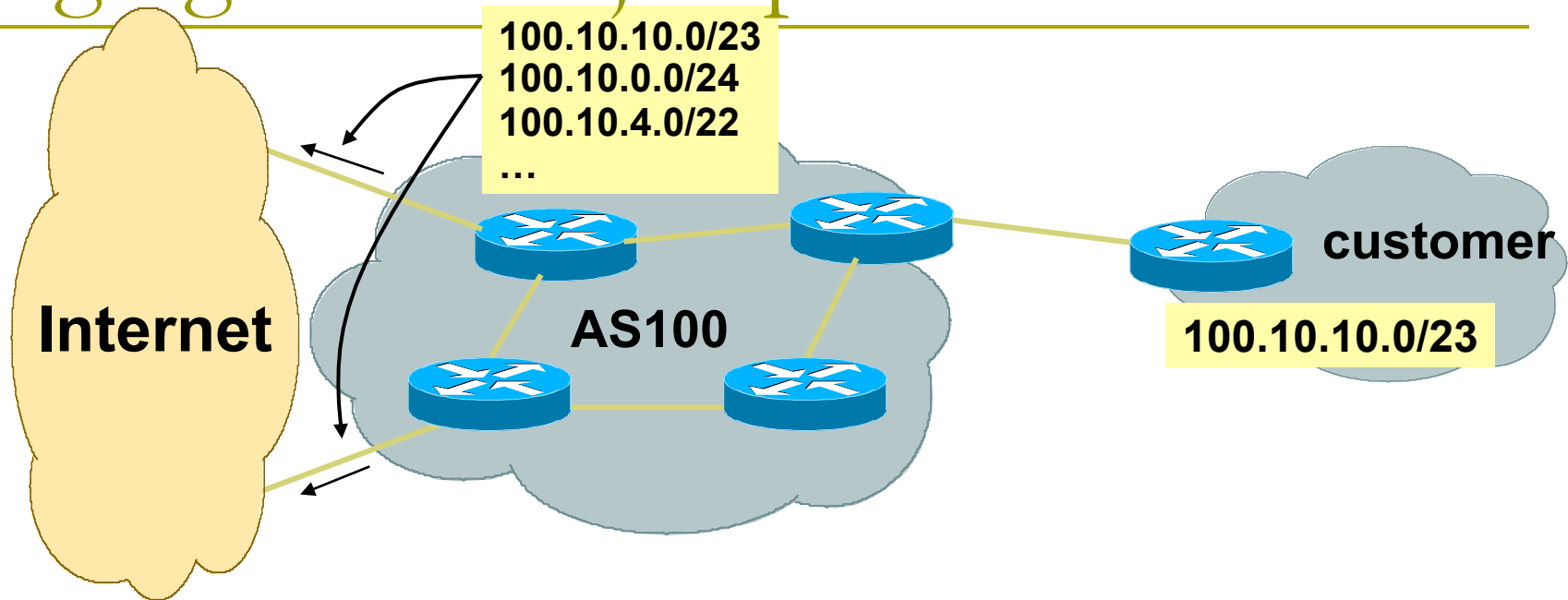
□ Ejemplo de configuración

```
router bgp 64511
  network 101.10.0.0 mask 255.255.224.0
  neighbor 102.102.10.1 remote-as 101
  neighbor 102.102.10.1 prefix-list out-filter out
  !
ip route 101.10.0.0 255.255.224.0 null0
  !
ip prefix-list out-filter permit 101.10.0.0/19
ip prefix-list out-filter deny 0.0.0.0/0 le 32
```

Anunciando un Agregado

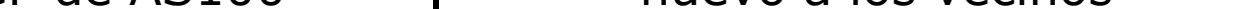
- ❑ Los ISPs que no efectúan agregación tienen mala reputación en Internet
- ❑ Los RIRs publican sus tamaños mínimos de asignación:
 - Entre /20 y /22 dependiendo del RIR
 - Tamaños diferentes para distintos bloques
- ❑ No hay ninguna razón válida para anunciar prefijos más específicos que /22 en la Internet
 - PERO actualmente hay (Mayo 2011) >185000 /24s!
- ❑ PERO: APNIC cambió (Oct 2010) su tamaño de asignación mínimo a /24
 - El agotamiento de IPv4 está empezando a tener impacto

Agregación– Ejemplo

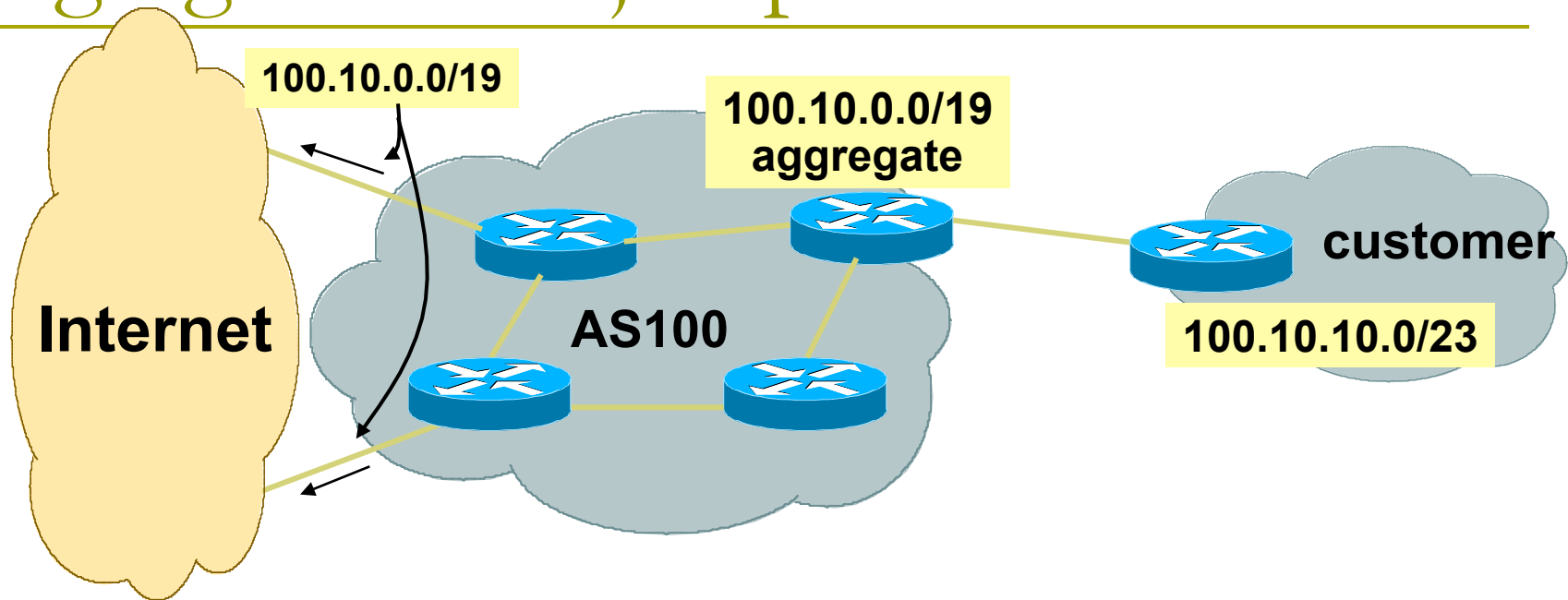


- ❑ Cliente tiene una red /23 asignada desde el bloque /19 de AS100
- ❑ AS100 anuncia a Internet las redes de sus clientes individualmente

Agregación– Mal ejemplo

- El enlace del cliente se cae
 - Su red /23 se hace inalcanzable
 - El prefijo /23 es retirado en el iBGP de AS100
 - El ISP no agrega su bloque /19
 - La retirada del /23 se anuncia a los vecinos
 - Comienza a viajar en cadena a través de Internet
 - Se carga innecesariamente a los enrutadores dorsales de Internet
- 
- El enlace del cliente retorna
 - Su /23 es nuevamente visible
 - El /23 se anuncia de nuevo a los vecinos
 - Viaja en cadena a través de Internet
 - Más carga en los enrutadores de Internet
 - Algunos ISPs suprimen los “flaps”
 - Puede tomar 10-20 minutos para que Internet sea visible
 - ¿Dónde está la Calidad de Servicio???

Agregación– Ejemplo



- ❑ El cliente tiene una red /23 asignada desde el bloque /19 de AS100
- ❑ AS100 anuncia el /19 agregado al Internet

Agregación – Buen ejemplo

- El enlace del cliente se cae
 - Su /23 se hace inalcanzable
 - /23 es retirado del iBGP de AS100
 - /19 agregado todavía está siendo anunciado
 - No hay retardos de propagación de BGP
 - No hay supresión por otros ISPs
- 
- El enlace del cliente retorna
 - Su red /23 es nuevamente visible
 - El /23 es re-inyectado en el iBGP de AS100
 - Todo el Internet es visible inmediatamente
 - El cliente tiene percepción de calidad de servicio

Agregación – Resumen

- ❑ El buen ejemplo es lo que todo el mundo debería hacer!
 - Añade estabilidad a Internet
 - Reduce el tamaño de la tabla de rutas
 - Reduce la ida y venida de prefijos
 - Mejora la calidad de servicio **para todos**
- ❑ El mal ejemplo es lo que demasiada gente todavía hace!
 - Por qué? Ignorancia?
 - Pereza?

Separación de iBGP y eBGP

- ❑ Muchos ISPs todavía no entienden la importancia de separar iBGP de eBGP
 - iBGP es para transportar los prefijos de los clientes
 - eBGP es para anunciar agregados a Internet y para ingeniería de tráfico
- ❑ **No** haga ingeniería de tráfico con los prefijos originados por iBGP de los clientes
 - Conduce a inestabilidad similar a la explicada en el mal ejemplo anterior
 - A pesar de que se anuncia un agregado, un subprefijo que rebota conllevará inestabilidad para el cliente en cuestión
- ❑ **Genere los prefijos para ingeniería de tráfico en el enrutador de frontera**

Internet hoy (Septiembre 2011)

□ Estadísticas actuales de la tabla de rutas

■ Entradas en la tabla BGP	376330
■ Prefijos después de máxima agregación	220899
■ Prefijos únicos en Internet	176062
■ Prefijos menores que la asignación de RIR	148138
■ Número de /24s anunciados	185829
■ ASs en uso	38961

Esfuerzos para mejorar la agregación

□ El “CIDR Report”

- Iniciado y operado durante años por Tony Bates
- Ahora combinado con el análisis de enrutamiento de Geoff Huston
 - www.cidr-report.org
 - (Abarca ambas tablas IPv4 e IPv6)
- Los resultados son enviados por e-mail cada semana a la mayoría de las listas de operadores en el mundo
- Lista los “top 30” proveedores que podrían hacer un mejor trabajo

□ Recomendación del grupo de trabajo de Enrutamiento de RIPE

- RIPE-399 — www.ripe.net/ripe/docs/ripe-399.html

Esfuerzos para mejorar la agregación

El “CIDR Report”

- ❑ También calcula el tamaño de la tabla de enrutamiento asumiendo que los ISPs efectúen agregación óptima
- ❑ El sitio web permite búsquedas y cálculos de agregación a efectuar por AS
 - Herramienta flexible y poderosa para ayudar a los ISPs
 - La meta es mostrar cómo una mayor eficiencia en términos del tamaño de la tabla de BGP se pueden obtener sin pérdida de información de enrutamiento y políticas
 - Muestra qué tipos de agregación a nivel de AS origen se pueden efectuar y el potencial beneficio de tales acciones
 - Cuestiona muy eficazmente la excusa de la ingeniería de tráfico

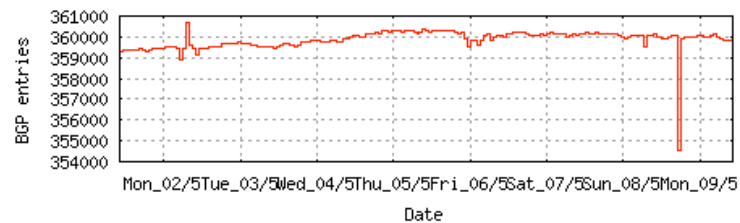
A list of advertisements of address blocks and Autonomous System numbers where there is no matching allocation data.

Status Summary

Table History

Date	Prefixes	CIDR Aggregated
02-05-11	359477	210568
03-05-11	359727	210958
04-05-11	359799	211320
05-05-11	360266	210489
06-05-11	359540	210993
07-05-11	360101	211009
08-05-11	359995	211087
09-05-11	360076	211042

Plot: [BGP Table Size](#)



AS Summary

37589	Number of ASes in routing system
15852	Number of ASes announcing only one prefix
3646	Largest number of prefixes announced by an AS
	AS6389 : BELLSOUTH-NET-BLK - BellSouth.net Inc.
110377472	Largest address span announced by an AS (/32s)
	AS4134 : CHINANET-BACKBONE No.31,Jin-rong Street

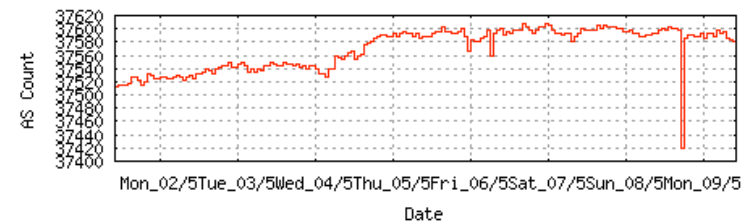
Plot: [AS count](#)

Plot: [Average announcements per origin AS](#)

Report: [ASes ordered by originating address span](#)

Report: [ASes ordered by transit address span](#)

Report: [Autonomous System number-to-name](#) mapping (from Registry WHOIS data)



Aggregation Summary

Announced Prefixes

Rank	AS	Type	Originate	Addr Space (pfx)	Transit	Addr space (pfx)	Description
131	AS4755	ORG+TRN	Originate:	3621120 /10.21	Transit:	9484544 /8.82	TATACOMM-AS TATA Communications formerly VSNL is Leader

Aggregation Suggestions

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation possibilities.

Rank	AS	AS Name	Current	Withdw	Aggte	Annce	Redctn	%
9	AS4755	TATACOMM-AS TATA Communications formerly VSNL	1461	1142	54	373	1088	74.47%

Prefix	AS Path	Aggregation Suggestion
14.140.0.0/14	4777 2516 6453 4755	
14.140.0.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.4.0/23	4608 1221 4637 6453 4755	
14.140.6.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.16.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.20.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.24.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.32.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.40.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.48.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.56.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.64.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.72.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.80.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.82.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.84.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.88.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
14.140.254.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 14.140.0.0/14 4777 2516 6453 4755
49.32.0.0/12	4777 2516 6453 4755	
59.151.144.0/22	4608 1221 4637 6453 4755	
59.160.0.0/16	4777 2516 6453 4755	
59.160.0.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.4.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.5.0/24	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.8.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.11.0/24	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.12.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.15.0/24	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.16.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.24.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.24.0/24	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.32.0/21	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.34.0/24	4608 1221 4637 6453 4755	
59.160.38.0/24	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.44.0/22	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755
59.160.46.0/23	4777 2516 6453 4755	- Withdrawn - matching aggregate 59.160.0.0/16 4777 2516 6453 4755

Announced Prefixes

Rank	AS	Type	Originate	Addr Space (pfx)	Transit	Addr space (pfx)	Description
168	AS18566		ORG+TRN Originate:	2647296 /10.66	Transit:	1024 /22.00	COVAD - Covad Communications Co.

Aggregation Suggestions

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation possibilities.

Rank	AS	AS Name	Current	Withdw	Aggte	Annce	Redctn	%
8	AS18566	COVAD - Covad Communications Co.	1782	1394	271	659	1123	63.02%

Prefix	AS Path	Aggregation Suggestion
64.81.16.0/22	4777 2516 4565 18566	
64.81.22.0/24	4777 2516 4565 18566	
64.81.24.0/21	4777 2516 4565 18566	+ Announce - aggregate of 64.81.24.0/22 (4777 2516 4565 18566) and 64.81.28.0/22 (4777 2516 4565 18566)
64.81.24.0/22	4777 2516 4565 18566	- Withdrawn - aggregated with 64.81.28.0/22 (4777 2516 4565 18566)
64.81.28.0/22	4777 2516 4565 18566	- Withdrawn - aggregated with 64.81.24.0/22 (4777 2516 4565 18566)
64.81.32.0/19	4777 2516 4565 18566	+ Announce - aggregate of 64.81.32.0/20 (4777 2516 4565 18566) and 64.81.48.0/20 (4777 2516 4565 18566)
64.81.32.0/20	4777 2516 4565 18566	- Withdrawn - aggregated with 64.81.48.0/20 (4777 2516 4565 18566)
64.81.32.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.33.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.34.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.35.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.36.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.37.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.38.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.39.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.40.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.44.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.48.0/20	4777 2516 4565 18566	- Withdrawn - aggregated with 64.81.32.0/20 (4777 2516 4565 18566)
64.81.48.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.49.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.50.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.51.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.52.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.53.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.54.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.55.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.56.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.57.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.58.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.59.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.60.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.61.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 4565 18566
64.81.64.0/19	4777 2516 4565 18566	+ Announce - aggregate of 64.81.64.0/20 (4777 2516 4565 18566) and 64.81.80.0/20 (4777 2516 4565 18566)
64.81.64.0/20	4777 2516 4565 18566	- Withdrawn - aggregated with 64.81.80.0/20 (4777 2516 4565 18566)
64.81.64.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.64.0/20 4777 2516 4565 18566
64.81.65.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.64.0/20 4777 2516 4565 18566

Importancia de la Agregación

- ❑ Tamaño de la tabla de rutas
 - La memoria de los enrutadores no es ya tanto problema como fue en los años 90
 - Se puede comprar un enrutador que soporte más de un millón de prefijos
- ❑ Convergencia del sistema de enrutamiento
 - Esto es un problema
 - Una tabla mayor toma más tiempo para procesar
 - Las actualizaciones de BGP toman más tiempo
 - El reporte de inestabilidad de BGP sigue la actividad del sistema de enrutamiento BGP
 - bgpupdates.potaroo.net/instability/bgpupd.html

The BGP Instability Report

The BGP Instability Report is updated daily. This report was generated on 09 May 2011 06:14 (UTC+1000)

50 Most active ASes for the past 7 days

RANK	ASN	UPDs	%	Prefixes	UPDs/Prefix	AS NAME
1	9829	42709	2.60%	1039	41.11	BSNL-NIB National Internet Backbone
2	19743	33117	2.01%	7	4731.00	
3	17974	26180	1.59%	1834	14.27	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia
4	14434	17436	1.06%	68	256.41	
5	32528	16930	1.03%	8	2116.25	ABBOTT Abbot Labs
6	21826	16840	1.02%	306	55.03	Internet Cable Plus C. A.
7	24560	15062	0.92%	1169	12.88	AIRTELBROADBAND-AS-AP Bharti Airtel Ltd., Telemedia Services
8	35819	14744	0.90%	411	35.87	MOBILY-AS Etihad Etisalat Company (Mobily)
9	35931	14220	0.86%	6	2370.00	ARCHIPELAGO - ARCHIPELAGO HOLDINGS INC
10	4274	13591	0.83%	81	167.79	ERX-AU-NET Assumption University
11	9299	13112	0.80%	1316	9.96	IPG-AS-AP Philippine Long Distance Telephone Company
12	44609	12813	0.78%	3	4271.00	FNA Fars News Agency Cultural Arts Institute
13	14420	12143	0.74%	667	18.21	CORPORACION NACIONAL DE TELECOMUNICACIONES - CNT EP
14	6458	12022	0.73%	302	39.81	Telgua
15	33475	11795	0.72%	215	54.86	RSN-1 - RockSolid Network, Inc.
16	11492	11776	0.72%	1268	9.29	CABLEONE - CABLE ONE, INC.
17	1660	11317	0.69%	79	143.25	ANS-CORP-NY - ANS Communications
18	8151	10312	0.63%	1383	7.46	Uninet S.A. de C.V.
19	24534	10156	0.62%	4	2539.00	TRANSHYBRID-AS-ID PT. Transhybrid Communication
20	45595	9832	0.60%	362	27.16	PKTELECOM-AS-PK Pakistan Telecom Company Limited
21	27738	9761	0.59%	339	28.79	Ecuadortelecom S.A.
22	24757	9573	0.58%	52	184.10	EthioNet-AS
23	9498	9544	0.58%	801	11.92	BBIL-AP BHARTI Airtel Ltd.
24	3454	8975	0.55%	8	1121.88	Universidad Autonoma de Nuevo Leon
25	7491	7916	0.48%	98	80.78	PI-PH-AS-AP PI-PHILIPINES

50 Most active Prefixes for the past 7 days

RANK	PREFIX	UPDs	%	Origin AS -- AS NAME
1	200.23.202.0/24	8713	0.50%	3454 -- Universidad Autonoma de Nuevo Leon
2	130.36.35.0/24	8463	0.49%	32528 -- ABBOTT Abbot Labs
3	130.36.34.0/24	8460	0.49%	32528 -- ABBOTT Abbot Labs
4	202.92.235.0/24	8077	0.46%	9498 -- BBIL-AP BHARTI Airtel Ltd.
5	63.211.68.0/22	7654	0.44%	35931 -- ARCHIPELAGO - ARCHIPELAGO HOLDINGS INC
6	198.140.43.0/24	6538	0.38%	35931 -- ARCHIPELAGO - ARCHIPELAGO HOLDINGS INC
7	65.122.196.0/24	6494	0.37%	19743 --
8	178.22.72.0/21	6414	0.37%	44609 -- FNA Fars News Agency Cultural Arts Institute
9	178.22.79.0/24	6391	0.37%	44609 -- FNA Fars News Agency Cultural Arts Institute
10	221.121.96.0/19	5462	0.31%	7491 -- PI-PH-AS-AP PI-PHILIPINES
11	72.164.144.0/24	5330	0.31%	19743 --
12	66.238.91.0/24	5323	0.31%	19743 --
13	66.89.98.0/24	5323	0.31%	19743 --
14	65.162.204.0/24	5322	0.31%	19743 --
15	65.163.182.0/24	5322	0.31%	19743 --
16	64.43.0.0/16	4617	0.27%	18704 -- T-SYSTEMS-NA - T-Systems North America, Inc.
17	66.248.160.0/22	4362	0.25%	14434 --
18	66.248.170.0/23	4362	0.25%	14434 --
19	66.248.172.0/23	4301	0.25%	14434 --
20	66.248.168.0/24	4297	0.25%	14434 --
21	68.65.152.0/22	3648	0.21%	11915 -- TELWEST-NETWORK-SVCS-STATIC - TEL WEST COMMUNICATIONS LLC
22	202.153.174.0/24	3437	0.20%	17408 -- ABOVE-AS-AP AboveNet Communications Taiwan
23	208.54.82.0/24	3093	0.18%	701 -- UUNET - MCI Communications Services, Inc. d/b/a Verizon Business
24	58.147.191.0/24	2539	0.15%	24534 -- TRANSHYBRID-AS-ID PT. Transhybrid Communication
25	58.147.185.0/24	2539	0.15%	24534 -- TRANSHYBRID-AS-ID PT. Transhybrid Communication
26	58.147.184.0/24	2539	0.15%	24534 -- TRANSHYBRID-AS-ID PT. Transhybrid Communication
27	58.147.188.0/24	2539	0.15%	24534 -- TRANSHYBRID-AS-ID PT. Transhybrid Communication
28	65.181.192.0/23	2200	0.13%	11492 -- CABLEONE - CABLE ONE, INC.
29	213.55.75.0/24	2126	0.12%	24757 -- EthioNet-AS
30	213.55.74.0/24	2122	0.12%	24757 -- EthioNet-AS

Recibiendo Prefijos



Recibiendo Prefijos

- ❑ Hay tres escenarios para recibir prefijos de otros ASs
 - Un cliente con BGP
 - Un “Peer” con BGP
 - Un proveedor con BGP
- ❑ Cada uno tiene requerimientos de filtrado y deben ser considerados individualmente

Recibiendo Prefijos:

De los clientes

- El ISP sólo debe aceptar prefijos que han sido asignados al cliente
- Si el ISP ha asignado un bloque al cliente, entonces el cliente TIENE la legitimidad para anunciarlo hacia el ISP
- Si el ISP NO ha asignado un bloque al cliente, entonces:
 - Revisar las bases de datos de los 5 RIRs para ver si este bloque ha sido asignado al cliente
 - **whois -h whois.apnic.net x.x.x.0/24**

Recibiendo prefijos: De los clientes

- Ejemplo de uso de WHOIS para verificar si el cliente puede anunciar un prefijo:

```
$ whois -h whois.apnic.net 202.12.29.0
inetnum:          202.12.28.0 - 202.12.29.255
netname:          APNIC-AP
descr:            Asia Pacific Network Information Centre
descr:            Regional Internet Registry for the Asia-Pacific
descr:            6 Cordelia Street
descr:            South Brisbane, QLD 4101
descr:            Australia
country:          AU
admin-c:          AIC1-AP
tech-c:           NO4-AP
mnt-by:           APNIC-HM
mnt-irt:           IRT-APNIC-AP
changed:          hm-changed@apnic.net
status:           ASSIGNED PORTABLE
changed:          hm-changed@apnic.net 20110309
source:           APNIC
```

Portable – means its an assignment to the customer, the customer can announce it to you



Recibiendo Prefijos: De los Clientes

- Example use of whois to check if customer is entitled to announce address space:

```
$ whois -h whois.ripe.net 193.128.0.0
inetnum:          193.128.0.0 - 193.133.255.255
netname:          UK-PIPEX-193-128-133
descr:           Verizon UK Limited
country:          GB
org:              ORG-UA24-RIPE
admin-c:          WERT1-RIPE
tech-c:           UPHM1-RIPE
status:           ALLOCATED UNSPECIFIED
remarks:          Please send abuse notification to abuse@uk.uu.net
mnt-by:           RIPE-NCC-HM-MNT
mnt-lower:        AS1849-MNT
mnt-routes:       AS1849-MNT
mnt-routes:       WCOM-EMEA-RICE-MNT
mnt-irt:          IRT-MCI-GB
source:           RIPE # Filtered
```

ALLOCATED – means that this is Provider Aggregatable address space and can only be announced by the ISP holding the allocation (in this case Verizon UK)

Recibiendo prefijos del cliente: Cisco IOS

- Por ejemplo:
 - Cliente tiene 100.50.0.0/20
 - Sólo anunciar hacia ASs superiores (upstream)
 - ASs superiores sólo deben aceptar este prefijo
- Configuración en el AS superior

```
router bgp 100
  neighbor 102.102.10.1 remote-as 101
  neighbor 102.102.10.1 prefix-list customer in
!
ip prefix-list customer permit 100.50.0.0/20
```

Recibiendo Prefijos: Desde los Peers

- Un Peer es otra red con la cual usted acuerda intercambiar prefijos originados por usted hacia la Internet
 - Los prefijos a aceptar de un peer son únicamente los que el peer ha indicado que va a anunciar
 - Y viceversa

Recibiendo Prefijos: Desde los Peers

- Acordar qué se va anunciar mutuamente:
 - Intercambio de e-mail como parte del acuerdo de intercambio, y subsecuentes actualizaciones
 -
 - Uso del Internet Routing Registry y herramientas de configuración como IRRToolSet

www.isc.org/sw/IRRToolSet/

Recibiendo prefijos de un Peer: Cisco IOS

- ❑ Por ejemplo:
 - El Peer tiene 220.50.0.0/16, 61.237.64.0/18 y 81.250.128.0/17
- ❑ Configuración en el enrutador local

```
router bgp 100
  neighbor 102.102.10.1 remote-as 101
  neighbor 102.102.10.1 prefix-list my-peer in
!
ip prefix-list my-peer permit 220.50.0.0/16
ip prefix-list my-peer permit 61.237.64.0/18
ip prefix-list my-peer permit 81.250.128.0/17
ip prefix-list my-peer deny 0.0.0.0/0 le 32
```

Recibiendo Prefijos:

Desde Proveedor de Tránsito

- ❑ El proveedor de tránsito es un ISP al cual usted paga para recibir tránsito hacia **TODA** la Internet
- ❑ Recibir prefijos de éste no es recomendable a menos que haya una buena razón
 - Ingeniería de tráfico
- ❑ Pedir al proveedor de tránsito que:
 - Origine una ruta por defecto
 -
 - Anuncie un prefijo que usted pueda utilizar por defecto

Recibiendo Prefijos: Desde Proveedor de Tránsito

- ❑ Configuración del enrutador de nivel inferior (downstream)

```
router bgp 100
  network 101.10.0.0 mask 255.255.224.0
  neighbor 101.5.7.1 remote-as 101
  neighbor 101.5.7.1 prefix-list infilter in
  neighbor 101.5.7.1 prefix-list outfilter out
!
ip prefix-list infilter permit 0.0.0.0/0
!
ip prefix-list outfilter permit 101.10.0.0/19
```


Recibiendo Prefijos: Desde Proveedor de Tránsito

❑ Configuración del enrutador superior

```
router bgp 101
  neighbor 101.5.7.2 remote-as 100
  neighbor 101.5.7.2 default-originate
  neighbor 101.5.7.2 prefix-list cust-in in
  neighbor 101.5.7.2 prefix-list cust-out out
!
ip prefix-list cust-in permit 101.10.0.0/19
!
ip prefix-list cust-out permit 0.0.0.0/0
```

Recibiendo Prefijos:

Desde Proveedor de Tránsito

- ❑ Si fuese necesario recibir prefijos de cualquier proveedor, debe proceder con cuidado.
 - No acepte una ruta por defecto (a menos que la necesite)
 - No acepte sus propios prefijos
- ❑ Para IPv4:
 - No acepte prefijos privados (RFC1918) y prefijos de uso especial:
<http://www.rfc-editor.org/rfc/rfc5735.txt>
 - No acepte prefijos más específicos de /24 (?)
- ❑ Para IPv6:
 - No acepte prefijos de uso especial:
<http://www.rfc-editor.org/rfc/rfc5156.txt>
 - No aceptar prefijos más específicos de /48 (?)

Recibiendo Prefijos: Desde Proveedor de Tránsito

- ❑ Revise la lista de “bogons” del grupo Cymru
www.team-cymru.org/Services/Bogons/http.html
- ❑ Para IPv4 también consulte:
datatracker.ietf.org/doc/draft-vegoda-no-more-unallocated-slash8s
- ❑ Para IPv6 también consulte:
www.space.net/~gert/RIPE/ipv6-filters.html
- ❑ Servidor de rutas Bogon:
www.team-cymru.org/Services/Bogons/routeserver.html
 - Proporciona una fuente BGP (IPv4 y/o IPv6) con bloques que no deben aparecer en la tabla de rutas

Recibiendo prefijos IPv4

```
router bgp 100
 network 101.10.0.0 mask 255.255.224.0
 neighbor 101.5.7.1 remote-as 101
 neighbor 101.5.7.1 prefix-list in-filter in
!
ip prefix-list in-filter deny 0.0.0.0/0                ! default
ip prefix-list in-filter deny 0.0.0.0/8 le 32
ip prefix-list in-filter deny 10.0.0.0/8 le 32
ip prefix-list in-filter deny 101.10.0.0/19 le 32       ! Block local prefix
ip prefix-list in-filter deny 127.0.0.0/8 le 32
ip prefix-list in-filter deny 169.254.0.0/16 le 32     ! Auto-config
ip prefix-list in-filter deny 172.16.0.0/12 le 32
ip prefix-list in-filter deny 192.0.2.0/24 le 32       ! TEST1
ip prefix-list in-filter deny 192.168.0.0/16 le 32
ip prefix-list in-filter deny 198.18.0.0/15 le 32      ! Benchmarking
ip prefix-list in-filter deny 198.51.100.0/24 le 32    ! TEST2
ip prefix-list in-filter deny 203.0.113.0/24 le 32     ! TEST3
ip prefix-list in-filter deny 224.0.0.0/3 le 32        ! Block multicast
ip prefix-list in-filter deny 0.0.0.0/0 ge 25          ! Block prefixes >/
24
ip prefix-list in-filter permit 0.0.0.0/0 le 32
```

Recibiendo Prefijos IPv6

```
router bgp 100
  network 2020:3030::/32
  neighbor 2020:3030::1 remote-as 101
  neighbor 2020:3030::1 prefix-list v6in-filter in
  !
  ipv6 prefix-list v6in-filter deny ::/0                                ! Default
  ipv6 prefix-list v6in-filter deny ::/8 le 128
  ipv6 prefix-list v6in-filter permit 2001::/32                        ! Teredo
  ipv6 prefix-list v6in-filter deny 2001::/32 le 128
  ipv6 prefix-list v6in-filter deny 2001:db8::/32 le 128              !
  Documentation
  ipv6 prefix-list v6in-filter permit 2002::/16                        ! 6to4
  ipv6 prefix-list v6in-filter deny 2002::/16 le 128
  ipv6 prefix-list v6in-filter deny 2020:3030::/32 le 128              ! Local Prefix
  ipv6 prefix-list v6in-filter deny 3ffe::/16 le 128                  ! Old 6bone
  ipv6 prefix-list v6in-filter deny fc00::/7 le 128                   ! Unique Local
  ipv6 prefix-list v6in-filter deny fe80::/10 le 128                  ! Link Local
  ipv6 prefix-list v6in-filter deny ff00::/8 le 128                   ! Multicast
  ipv6 prefix-list v6in-filter permit 2000::/3 le 48                  ! Global Unicast Block
  ipv6 prefix-list v6in-filter deny ::/0 le 128
```

Recibiendo Prefijos

- ❑ Poner atención a los prefijos recibidos de los clientes, peers y proveedores de tránsito ayuda con:
 - La integridad de su propia red
 - La integridad de la Internet
- ❑ Es la responsabilidad de todos los ISPs ser buenos ciudadanos de la red

Prefijos en iBGP



Inyectar prefijos en iBGP

- ❑ Utilice iBGP para llevar los prefijos de los clientes
 - No use IGP
- ❑ Apunte ruta estática a la interfaz del cliente
- ❑ Use el comando network de BGP
- ❑ Mientras la ruta estática exista (la interfaz esté activa), el prefijo estará en BGP

Configuración del Enrutador: comando network

□ Ejemplo:

```
interface loopback 0
  ip address 215.17.3.1 255.255.255.255
!
interface Serial 5/0
  ip unnumbered loopback 0
  ip verify unicast reverse-path
!
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
  network 215.34.10.0 mask 255.255.252.0
```

Inyectando prefijos en iBGP

- ❑ Un rebote de la interfaz resultará en la retirada y re-anuncio del prefijo
 - use `"ip route...permanent"`
- ❑ Muchos ISPs redistribuyen las rutas estáticas en BGP en vez de usar el comando `network`
 - Sólo haga esto si tiene una buena razón

Configuración del Enrutador: redistribute static

□ Ejemplo:

```
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100
  redistribute static route-map static-to-bgp
<snip>
!
route-map static-to-bgp permit 10
  match ip address prefix-list ISP-block
  set origin igp
<snip>
!
ip prefix-list ISP-block permit 215.34.10.0/22 le 30
```

Inyectando prefijos en iBGP

- Route-map ISP-block puede usarse para muchas cosas:
 - Asignar comunidades y otros atributos
 - Asignar el código de origen al IGP, etc
- Tenga cuidado con las prefix-lists y los route-maps
 - La ausencia de uno u otro significa que todos los prefijos enrutados estáticamente van a parar a iBGP

Haciendo la red sostenible



Cómo dejar de transportar
todos los prefijos en iBGP

Por qué usar BGP en vez de IGP?

□ Los IGP tienen limitaciones:

- Mientras más información de enrutamiento haya en la red:
 - Actualizaciones periódicas/inundaciones “sobrecarga”
 - Tiempos de convergencia altos
 - Afecta al núcleo (core) de red primero
- Definición de políticas
 - No es fácil de lograr

Preparando la Red

- ❑ Queremos desplegar BGP ahora...
- ❑ Es necesario un ASN
- ❑ Si se planea tener multihoming con varios ISPs en el futuro, necesita un ASN público:
 - Obténgalo del ISP, si es miembro de un RIR, o
 - Solicite usted mismo al RIR una asignación individual, o
 - Pregunte a un ISP que sea miembro de un RIR, o
 - **Hágase miembro del RIR y obtenga también su propio bloque IP (esta opción muy recomendada)!**

Preparando la Red

Suposiciones iniciales

- La red no está utilizando BGP en este momento
 - Una ruta estática única apuntando al ISP
- La red no tiene ningún IGP en este momento
 - Rutas estáticas alrededor de toda la red

Preparando la Red

Primer Paso: IGP

- ❑ Decídase por un IGP: OSPF o ISIS ☺
- ❑ Asigne interfaces loopback y direcciones /32 a cada enrutador que será parte del IGP
 - La Loopback se usa como ID para OSPF y BGP
 - Utilizado en iBGP y origen de rutas
- ❑ Despliegue el IGP (ej. OSPF)
 - El IGP se puede implementar SIN NINGUN IMPACTO sobre las rutas estáticas existentes
 - ej. Distancia OSPF es 110; Distancia estática es 1
 - **Smallest distance wins**

Preparando la Red

IGP (cont.)

- ❑ Sea prudente desplegando el IGP – Mantenga la base de datos de link state al mínimo!
 - Los loopbacks de enrutadores van en el IGP
 - Los enlaces WAN punto-a-punto van en el IGP
 - (De hecho, cualquier enlace donde se ejecute el IGP debe incluirse en el IGP)
 - Resuma rutas en las fronteras de área/nivel (si es posible) – Ej. Piense en su plan de direccionamiento

Preparando la Red

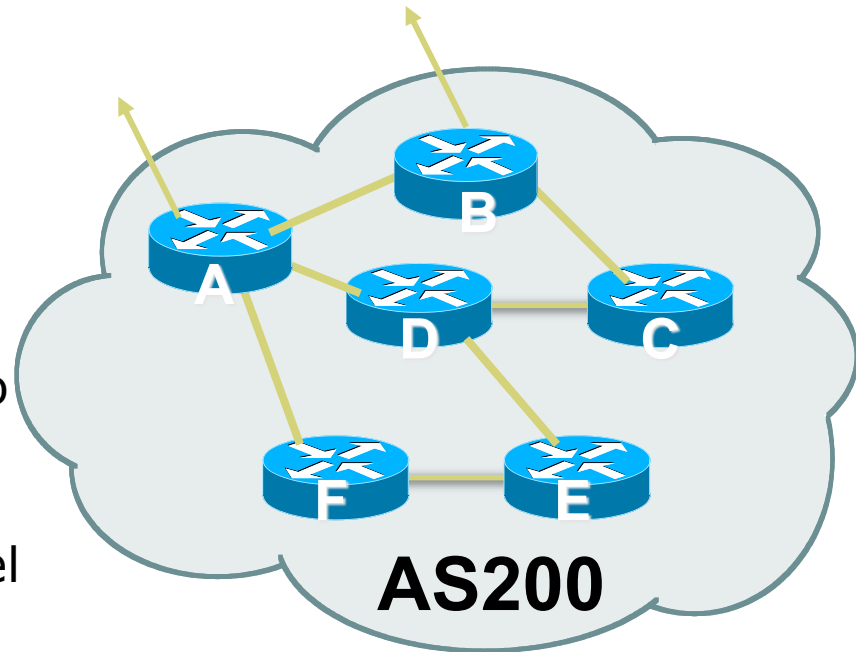
IGP (cont.)

- ❑ Las rutas que no deben ir en el IGP incluyen:
 - Rangos de asignación dinámica (DSL/Cable/Dial-up)
 - Los enlaces punto a punto con los clientes
 - ❑ (Use next-hop-self in iBGP para que éstos NO tengan que ir en el IGP)
 - Espacio de direcciones de clientes
 - Cualquier cosa que no esté en la diapositiva anterior

Preparando la Red

Segundo Paso: iBGP

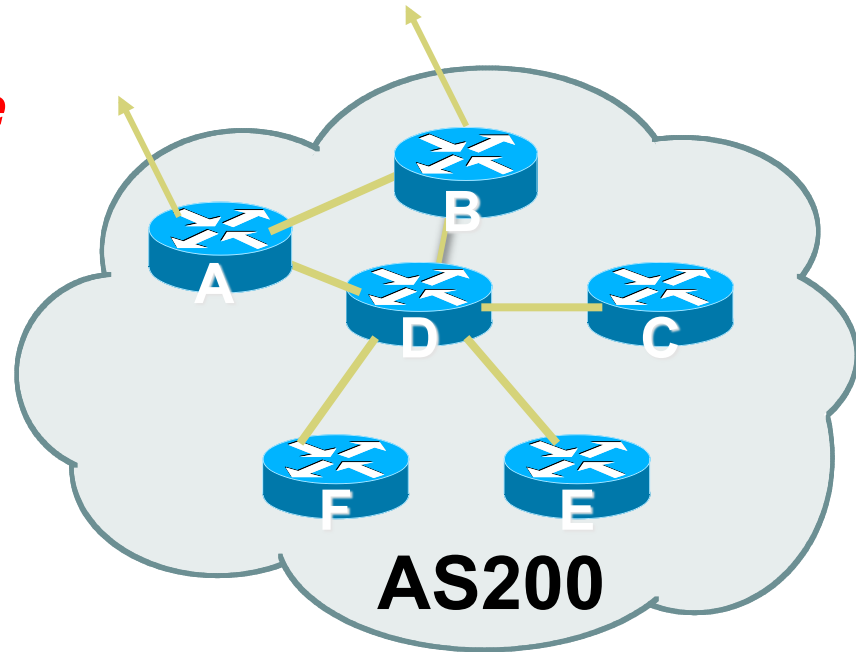
- ❑ El segundo paso es configurar la red local para usar iBGP
- ❑ iBGP puede ejecutarse en:
 - Todos los enrutadores, o
 - Un subconjunto de los enrutadores, o
 - Sólo en la frontera con el proveedor
- ❑ **iBGP debe ejecutarse en todos los enrutadores que estén en la trayectoria entre conexiones externas**



Preparando la Red

Segundo Paso: iBGP (Trayectoria de Tránsito)

- ❑ ***iBGP must run on all routers which are in the transit path between external connections***
- ❑ Enrutadores C, E y F no están en la trayectoria del tránsito
 - Rutas estáticas o IGP son suficientes
- ❑ El enrutador D está en la trayectoria del tránsito
 - Deben formar una malla completa iBGP, o de lo contrario habrá bucles de enrutamiento



Preparando la Red

Capas

- La redes de proveedores típicas tienen tres capas:
 - Núcleo (core)– la dorsal, generalmente en la trayectoria del tránsito
 - Distribución– la zona media, capa de agregación de los puntos de presencia (PoPs)
 - Agregación– la frontera, los dispositivos que conectan a los clientes

Preparando la Red

Capa de Agregación

- ❑ iBGP es opcional
 - Muchos ISPs corren iBGP aquí, ya sea con enrutamiento parcial (más común) o enrutamiento completo (menos común)
 - Enrutamiento completo no necesario a menos que los clientes quieran todas las rutas
 - Enrutamiento parcial es más barato/simple, usualmente consiste de prefijos internos y, opcionalmente, prefijos externos para ayudar en el balanceo de carga
 - ❑ Comunidades y peer-groups hace esto más fácil
- ❑ Muchos dispositivos de agregación no pueden correr iBGP
 - Rutas estáticas desde dispositivos de distribución para rangos de direcciones
 - IGP para la mejor salida

Preparando la Red

Capa de Distribución

- Generalmente corre iBGP
 - Enrutamiento parcial o completo (como en la capa de agregación)
- Pero no tiene que correr iBGP
 - IGP se puede usar para llevar los prefijos de los clientes (no es sostenible)
 - IGP se utiliza para determinar la salida más cercana
- Las redes que planean crecer mucho deberían usar iBGP desde el primer día
 - Migrar después es más trabajo
 - No hay costo agregado en utilizar iBGP

Preparando la Red

Capa Núcleo

- ❑ El núcleo de la red es generalmente la trayectoria de tránsito
- ❑ iBGP necesario entre dispositivos del núcleo
 - Rutas completas o parciales:
 - ❑ ISPs de tránsito llevan rutas completas
 - ❑ ISPs en las fronteras llevan rutas parciales
- ❑ La capa núcleo incluye los enrutadores de frontera

Preparando la Red

Implementación de iBGP

Decidir sobre lo siguiente:

- ❑ Mejor política de iBGP
 - Rutas completas por todas partes, o parciales, o una mezcla?
- ❑ Técnica de escalabilidad iBGP
 - Política basada en comunidades?
 - Route-reflectors?
 - Técnicas como peer groups y peer templates?

Preparando la Red iBGP

□ Desplegar iBGP:

- Paso 1: Crear una malla completa entre los enrutadores seleccionados para iBGP
 - Asegurarse de que la distancia iBGP es mayor que la distancia del IGP (generalmente lo es)

- Paso 2: Instalar los prefijos de los “clientes” en iBGP

Revisar! Todavía funciona la red?

- Paso 3: Sacar con cuidado las rutas estáticas que están ahora en el IGP y en iBGP

Revisar! Todavía funciona la red?

- Paso 4: Continuar con el despliegue de eBGP

Preparando la Red

Implementación de iBGP

Instalar los prefijos de "clientes" en iBGP?

- ❑ Espacio de direcciones asignado a clientes
 - Combinación de Comando 'network' y rutas estáticas
 - Utilizar comunidad única para identificar las asignaciones de los clientes
- ❑ Enlaces punto a punto con clientes
 - Usar "redistribute connected" con filtros que permitan únicamente a los enlaces punto a punto entrar en iBGP
 - Utilizar comunidad única para identificar los enlaces punto a punto (sólo para el sistema de monitoreo)
- ❑ LANs locales y rangos de asignación dinámica
 - Usar simplemente el comando network
 - Utilizar comunidad única para identificar estas redes

Preparando la Red

Implementación de iBGP

Quitar cuidadosamente las rutas estáticas?

- ❑ Trabaje con un enrutador a la vez:
 - Revisar que la ruta estática a un destino en particular también es aprendida por iBGP
 - Si es así, quite la ruta estática
 - Si no, averigüe por qué y arréglole
 - (Recuerde mirar en la RIB, no en la FIB!)
- ❑ Continúe con el enrutador siguiente, hasta que todo el PoP esté listo
- ❑ Continúe con el siguiente PoP, y así sucesivamente hasta que la red sólo dependa del IGP y el iBGP que ha instalado

Preparando la Red

Completando

- ❑ Los pasos anteriores no son para hacer todo en un día
 - Cada uno puede ser realizado en períodos de mantenimiento separados, por ejemplo:
 - Paso 1 en semana 1
 - Paso 2 en semana 2
 - Paso 3 en semana 3
 - Y así sucesivamente
 - Y con buena planificación, no tendrá ningún impacto sobre el servicio

Preparando la Red

Resumen de la Configuración

- ❑ Sólo las redes necesarias están en el IGP
- ❑ Las redes de los clientes están ahora en iBGP
 - iBGP desplegado alrededor de la dorsal
 - Rutas completas o parciales
- ❑ La distancia BGP es mayor que la de cualquier IGP
- ❑ Ahora está listo para desplegar eBGP

Prácticas Recomendadas de BGP

