



RFC1998

- Informational RFC
- Describes how to implement loadsharing and backup on multiple inter-AS links

BGP communities used to determine local preference in upstream's network

- Gives control to the customer
- Simplifies upstream's configuration simplifies network operation!

RFC1998

• Community values defined to have particular meanings:

ASx:100	set local pref 100	preferred route	
ASx:90	set local pref 90	backup route if dualhomed on ASx	
ASx:80	set local pref 80	main link is to another ISP with same AS path length	
ASx:70	set local pref 70	main link is to another ISP	

RFC1998

```
Sample Customer Router Configuration
router bgp 107
neighbor x.x.x.x remote-as 109
neighbor x.x.x.x description Backup ISP
neighbor x.x.x.x route-map config-community out
neighbor x.x.x.x send-community
!
ip as-path access-list 20 permit ^$
ip as-path access-list 20 deny .*
!
route-map config-community permit 10
match as-path 20
set community 109:90
```

RFC1998

Sample ISP Router Configuration

- ! Homed to another ISP
- ip community-list 70 permit 109:70
- ! Homed to another ISP with equal ASPATH length
- ip community-list 80 permit 109:80
- ! Customer backup routes
- ip community-list 90 permit 109:90
- !
- route-map set-customer-local-pref permit 10
- match community 70
- set local-preference 70

RFC1998
 Sample ISP Router Configuration
route-map set-customer-local-pref permit 20
match community 80
set local-preference 80
!
route-map set-customer-local-pref permit 30
match community 90
set local-preference 90
!
route-map set-customer-local-pref permit 40
set local-preference 100

RFC1998

Supporting RFC1998

many ISPs do, more should

check AS object in the Internet Routing Registry

if you do, insert comment in AS object in the IRR





Two links to the same ISP (one as backup only)

- Announce /19 aggregate on each link
 - primary link makes standard announcement

backup link sends community

• When one link fails, the announcement of the /19 aggregate via the other link ensures continued connectivity

Two links to the same ISP (one as backup only)

Router A Configuration

```
router bgp 65534
network 221.10.0.0 mask 255.255.224.0
neighbor 222.222.10.2 remote-as 109
neighbor 222.222.10.2 description RouterC
neighbor 222.222.10.2 prefix-list aggregate out
neighbor 222.222.10.2 prefix-list default in
!
ip prefix-list aggregate permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
```

Two links to the same ISP (one as backup only) Bouter B Configuration router bgp 6534 network 221.10.0.0 mask 255.255.224.0 neighbor 222.222.10.6 remote-as 109 neighbor 222.222.10.6 description RouterD neighbor 222.222.10.6 send-community neighbor 222.222.10.6 prefix-list aggregate out neighbor 222.222.10.6 route-map routerD-out out

neighbor 222.222.10.6 prefix-list default in

neighbor 222.222.10.6 route-map routerD-in in

ı.

Two links to the same ISP (one as backup only)

ip prefix-list aggregate permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
route-map routerD-out permit 10
match ip address prefix-list aggregate
set community 109:90
route-map routerD-out permit 20
!
route-map routerD-in permit 10

set local-preference 90

Two links to the same ISP (one as backup only)

Router C Configuration (main link)

router bgp 109

neighbor 222.222.10.1 remote-as 65534
neighbor 222.222.10.1 default-originate
neighbor 222.222.10.1 prefix-list Customer in
neighbor 222.222.10.1 prefix-list default out
!

ip prefix-list Customer permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0

www.cisco.com

Dublic State State

www.cisco.com

Two links to the same ISP (one as backup only)

```
ip prefix-list Customer permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
ip community-list 90 permit 109:90
!
<snip>
route-map bgp-cust-in permit 30
match community 90
set local-preference 90
route-map bgp-cust-in permit 40
```

set local-preference 100







Two links to different ISPs (one as backup only)

Announce /19 aggregate on each link

main link makes sends community 109:100 - this sets local pref in AS109 to 100

backup link sends community 108:80 - this sets local pref in AS108 to 80

 When one link fails, the announcement of the /19 aggregate via the other link ensures continued connectivity

Two links to different ISPs (one as backup only)

- Note that this assumes that AS109 and AS108 are interconnected
- If they are not, AS path length "stuffing" has to be used too

but that can be done on a per community basis also

custom additions to RFC1998

Two links to different ISPs (one as backup only)

```
Router A Configuration
```

```
network 221.10.0.0 mask 255.255.224.0
neighbor 222.222.10.1 remote-as 109
neighbor 222.222.10.1 prefix-list aggregate out
neighbor 222.222.10.1 prefix-list default in
i 
i 
ip prefix-list aggregate permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
```

```
route-map routerC-out permit 10
```

```
set community 109:100
```

Two links to different ISPs (one as backup only)

Router B Configuration

```
router bgp 107
network 221.10.0.0 mask 255.255.224.0
neighbor 220.1.5.1 remote-as 108
neighbor 220.1.5.1 prefix-list aggregate out
neighbor 220.1.5.1 route-map routerD-out out
neighbor 220.1.5.1 prefix-list default in
neighbor 220.1.5.1 route-map routerD-in in
..next slide
```

Two links to different ISPs (one as backup only)

ip prefix-list aggregate permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
route-map routerD-out permit 10

set community 108:80
!
route-map routerD-in permit 10

set local-preference 80

Two links to different ISPs (one as backup only)

Router D

sees path from router B with community 108:80 set - sets local preference to 80

sees path from peering with AS109 - default local preference is 100

local-pref comes before AS Path length

highest local-pref wins

traffic for AS107 is sent to AS109

Two links to different ISPs (one as backup only)

 Router D Only requires RFC1998 configuration no per customer configuration scalability!



Two links to different ISPs (one as backup only)

 If AS107 wants to make the link to AS108 the main link

sends community 108:100 to router C sends community 109:80 to router B

 AS108 and AS109 NOC intervention not required

Summary

- Communities are fun!
- And they are extremely powerful tools
- Think about community policies
- Supporting extensive community usage makes customer configuration easy
- Watch out for routing loops!





ISP Transit Issues

- Only announce default to your BGP customers unless they need more prefixes
- Only accept the prefixes which your customer is entitled to originate
- If your customer hasn't told you he is providing transit, don't accept anything else

ISP Transit Issues

Many mistakes are made on the Internet today due to incomplete understanding of how to configure BGP for transit



ISP Transit

- AS107 and AS109 are stub/customer ASes of AS108
 - they may have their own peerings with other ASes
 - minimal routing table desired
 - minimum complexity required





ISP Transit • Router B Configuration router bgp 108 neighbor 222.222.10.1 remote-as 107 neighbor 222.222.10.1 default-originate

neighbor 222.222.10.1 prefix-list Customer107 in neighbor 222.222.10.1 prefix-list default out

ip prefix-list Customer107 permit 221.10.0.0/19
ip prefix-list default permit 0.0.0.0/0

 Router B announces default to Router A, only accepts customer /19

-

ISP Transit

Router C Configuration

router bgp 108

neighbor 222.222.20.1 remote-as 109

neighbor 222.222.20.1 default-originate

neighbor 222.222.20.1 prefix-list Customer109 in neighbor 222.222.20.1 prefix-list default out

-

ip prefix-list Customer109 permit 219.0.0.0/19
ip prefix-list default permit 0.0.0.0/0

 Router C announces default to Router D, only accepts customer /19

ISP Transit

Router D Configuration

router bgp 109

network 219.0.0.0 mask 255.255.224.0

neighbor 222.222.20.2 remote-as 108

neighbor 222.222.20.2 prefix-list upstream out neighbor 222.222.20.2 prefix-list default in

l

ip prefix-list default permit 0.0.0.0/0
ip prefix-list upstream permit 219.0.0.0/19

www.cisco.com

ip route 219.0.0.0 255.255.224.0 null0

ISP Transit

• This is simple case:

if AS107 or AS109 get another address block, it requires AS108 and their own filters to be changed

some ISP transit provider are better skilled at doing this than others!

May not scale if they are frequently adding new prefixes







ISP Transit					
 Router A Configuration 					
router bgp 107					
network 221.10.0.0 mask 255.255.224.0					
neighbor 222.222.10.2 remote-as 108					
neighbor 222.222.10.2 prefix-list upstream-out out					
neighbor 222.222.10.2 filter-list 5 out					
neighbor 222.222.10.2 prefix-list upstream-in in					
!					
ip route 221.10.0.0 255.255.224.0 null0 250					
!					
next slide					
ISPEXP Workshops 0 286, Gass Systems, Inc. WWW.cisco.com	46				

ISP Transit
ip as-path access-list 5 permit ^\$
ip as-path access-list 5 permit ^(100_)+\$ ip as-path access-list 5 permit ^101\$
ip as-path access-list 5 permit ^102\$
ip as-path access-list 5 deny ^104_
! next slide

	ISP Transit	
	! Outbound Martian prefixes to be blocked to eBGP peers	
	ip prefix-list upstream-out deny 0.0.0.0/8 le 32	
	ip prefix-list upstream-out deny 10.0.0.0/8 le 32	
	ip prefix-list upstream-out deny 127.0.0.0/8 le 32	
	ip prefix-list upstream-out deny 169.254.0.0/16 le 32	
	ip prefix-list upstream-out deny 172.16.0.0/12 le 32	
	ip prefix-list upstream-out deny 192.0.2.0/24 le 32	
	ip prefix-list upstream-out deny 192.168.0.0/16 le 32	
	ip prefix-list upstream-out deny 224.0.0.0/3 le 32	
	ip prefix-list upstream-out deny 0.0.0.0/0 ge 25	
	! Extra prefixes	
	ip prefix-list upstream-out deny 221.10.0.0/19 ge 20	
	ip prefix-list upstream-out permit 0.0.0.0/0 le 32	
ISP/IXP Workshops	next slide © 2006, Chun Sydman, Iac. www.cisco.com 48	

ISP Transit						
	!]	Inbound Marti	lan prefixes	to be	blocked from eBGP peers	
	ip	prefix-list	upstream-in	deny	0.0.0.0/8 le 32	
	ip	prefix-list	upstream-in	deny	10.0.0/8 le 32	
	ip	prefix-list	upstream-in	deny	127.0.0.0/8 le 32	
	ip	prefix-list	upstream-in	deny	169.254.0.0/16 le 32	
	ip	prefix-list	upstream-in	deny	172.16.0.0/12 le 32	
	ip	prefix-list	upstream-in	deny	192.0.2.0/24 le 32	
	ip	prefix-list	upstream-in	deny	192.168.0.0/16 le 32	
	ip	prefix-list	upstream-in	deny	224.0.0.0/3 le 32	
	ip	prefix-list	upstream-in	deny	0.0.0/0 ge 25	
	! 1	Extra prefixe	es			
	ip	prefix-list	upstream-in	deny	221.10.0.0/19 le 32	
	ip	prefix-list	upstream-in	permi	t 0.0.0.0/0 le 32	
ISP/IXP Workshops	1	00 ClaroSystems Inc.	www.ci	sco.com		49

ISP Transit
 Router C Configuration

router bgp 108

neighbor 222.222.20.1 remote-as 109
neighbor 222.222.20.1 default-originate
neighbor 222.222.20.1 prefix-list Customer109 in
neighbor 222.222.20.1 prefix-list default out

ip prefix-list Customer109 permit 219.0.0.0/19
ip prefix-list default permit 0.0.0.0/0

 Router C announces default to Router D, only accepts customer /19

• www.cit

Spectrum prediction of the prediction of the

ISP Transit

AS107 only hears AS108 and AS109 prefixes

inbound AS path filter on Router A is optional, but good practice (never trust a peer)

DSUA prefix-list filters are mandatory on all Internet peerings



ISP Transit

- AS107 and AS109 are stub/customer ASes of AS108
 - AS107 has many customers with their own ASes

AS104 doesn't get announced to AS108

AS108 provides transit between AS107 and AS109

Same example as previously but using communities



ISP Transit ISP Transit Router A configuration is greatly **Router A Configuration** simplified router bgp 107 network 221.10.0.0 mask 255.255.224.0 route-map setcomm all prefixes to be announced to neighbor 222.222.10.2 remote-as 108 upstream are marked with community neighbor 222.222.10.2 prefix-list upstream-out out 107:5100 neighbor 222.222.10.2 route-map to-AS108 out neighbor 222.222.10.2 prefix-list upstream-in in route-map on outbound peering 1 implements community policy ip route 221.10.0.0 255.255.224.0 null0 250 **DSUA** prefix-lists still required ..next slide www.cisco.com

ISP Transit

```
ip community-list 5 permit 107:5100
! Set community on local prefixes
route-map setcomm permit 10
 set community 107:5100
!
route-map to-AS108 permit 10
match community 5
upstream-in and upstream-out prefix-lists are
```

1

```
the same as in the previous example
```

ISP Transit

www.cisco.com

router bg	p 107	
neighbor	x.x.x.x	remote-as 100
neighbor	x.x.x.x	default-originate
neighbor	x.x.x.x	prefix-list customer100 in
neighbor	x.x.x.x	route-map bgp-cust-in in
neighbor	x.x.x.x	prefix-list default out
neighbor	x.x.x.x	remote-as 101
neighbor	x.x.x.x	default-originate
neighbor	x.x.x.x	prefix-list customer101 in
neighbor	x.x.x.x	route-map bgp-cust-in in
neighbor	x.x.x.x	prefix-list default out

ISP Transit

neighbor s.s.s.s remote-as 104 neighbor s.s.s.s default-originate neighbor s.s.s.s prefix-list customer104 in neighbor s.s.s.s route-map no-transit in neighbor s.s.s.s prefix-list default out ! ! Set community on eEGP customers announced to AS108 route-map bgp-cust-in permit 10 set community 107:5100 route-map no-transit permit 10 set community 107:5199 Notice that AS104 peering has no route-map to

set the community policy

ISP Transit

- AS107 only announces the community 107:5100 to AS108
- Notice how Router E tags the prefixes to be announced to AS108 with community 107:5100
- More efficient to manage than using filter lists



Why Collocate Overseas?

- Hard to re-terminate transoceanic circuit in case of "issues" with upstream ISP
- No Quality of Service
- No Control over infrastructure
- No Monitoring of link performance

Collocation Overseas

 Many AP region ISPs collocate equipment in the US

install their own router(s) and other hardware (servers, caches,...)

establish peering relationships with US NSPs and domestic ISPs

buy facilities management services

usually hardware maintenance, installation management

Collocation Overseas

 Many AP region ISPs collocate equipment in the US

US domestic circuits are "cheap"

- Easy to change your upstream
- Easy to have multiple upstreams

Easy to implement QoS related features, service differentiation, etc...



	Collocation
• Comr	non Scenario:
AS1	07 has collocate space in the US
AS1 prov	08 and AS109 are transit riders for AS107
AS1 exch	07 is also present at the local hange point for regional peers



Collocation

• AS107

Router A is dedicated to peering at local IXP

Router G is dedicated to links with the transit providers

Router H is dedicated to the transoceanic link





Collocation Router A Configuration	Collocation Router A Configuration
router bgp 107	neighbor 220.5.10.4 remote-as 110
neighbor ixp-peers peer-group	neighbor 222.5.10.4 peer-group ixp-peers
neighbor ixp-peers soft-reconfiguration in	neighbor 222.5.10.4 prefix-list peer110 in
neighbor ixp-peers prefix-list myprefixes out	neighbor 220.5.10.5 remote-as 111
neighbor 221.0.0.2 remote-as 107	neighbor 222.5.10.5 peer-group ixp-peers
neighbor 221.0.0.2 description Router G - Upstream Peers	neighbor 222.5.10.5 prefix-list peer111 in
neighbor 221.0.0.2 update-source loopback 0	neighbor 220.5.10.3 remote-as 112
neighbor 221.0.0.3 remote-as 107	neighbor 222.5.10.3 peer-group ixp-peers
neighbor 221.0.0.3 description Router H - transpacific router	neighbor 222.5.10.3 prefix-list peer112 in
neighbor 221.0.0.3 update-source loopback 0	1
neighbor 221.0.0.4 remote-as 107	ip prefix-list myprefixes permit 221.10.0.0/19
neighbor 221.0.0.4 description Router at HQ	<pre>ip prefix-list peer110 permit 222.12.0.0/19</pre>
neighbor 221.0.0.4 update-source loopback 0	ip prefix-list peer111 permit 222.18.128.0/19
next slide	ip prefix-list peer112 permit 222.1.32.0/19
ISP/IXP Workshops 0 2000, Cheo Systems, Inc. WWW.CİSCO.COM 73	ISP/IXP Workshops 0 2009, Cheer Systems, Inc. WWW.CİSCO.COM

Collocation **Router G Configuration**

Collocation **Router A Configuration**

 Router A does NOT originate AS107's prefix block

if router is disconnected from AS107 either locally or across the ocean, announcement could cause blackhole

 Prefix-list filtering is the minimum required

usually include AS path filtering too

interface loopback 0

```
description Peering Router Loopback
ip address 221.0.0.2 255.255.255.255
```

```
interface fastethernet 0/0
description Crossover 100Mbps Connection to Router A
 ip address 221.0.10.1 255.255.255.252
```

no ip directed-broadcast

- no ip proxy-arp
- no ip redirects

..next slide

1

Collocation **Router G Configuration**

> interface hssi 1/0 description T3 link to BigISP ip address 222.0.0.2 255.255.255.252 no ip directed-broadcast no ip proxy-arp no ip redirects

interface hssi 2/0

- description T3 link to MegaISP ip address 218.6.0.2 255.255.255.252
- no ip directed-broadcast
- no ip proxy-arp
- no ip redirects
- ..next slide

Collocation **Router G Configuration**

router bgp 107 neighbor 221.0.0.1 remote-as 107 neighbor 221.0.0.1 description Router A - US Local IXP neighbor 221.0.0.1 update-source loopback 0 neighbor 221.0.0.1 prefix-list myprefixes out neighbor 221.0.0.3 remote-as 107 neighbor 221.0.0.3 description Router H - transpacific router neighbor 221.0.0.3 update-source loopback 0 neighbor 221.0.0.4 remote-as 107 neighbor 221.0.0.4 description Router at HO neighbor 221.0.0.4 update-source loopback 0 ..next slide

Collocation Router G Configuration

neighbor 222.0.0.1 remote-as 108 neighbor 222.0.0.1 prefix-list myprefixes out neighbor 222.0.0.1 prefix-list rfc1918-dsua in neighbor 218.6.0.1 remote-as 109 neighbor 218.6.0.1 prefix-list myprefixes out neighbor 218.6.0.1 prefix-list rfc1918-dsua in

ip prefix-list myprefixes permit 221.10.0.0/19

Collocation Router G Configuration

- Router G accepts full BGP prefixes from both AS108 and AS109
- Router G announces AS107 prefix to upstreams
- Simple Example policy may also be required for loadsharing etc



Collocation Router H Configuration

router bgp 107

neighbor 221.0.0.1 remote-as 107 neighbor 221.0.0.1 description Router A - US Local IXP neighbor 221.0.0.1 update-source loopback 0

neighbor 221.0.0.2 remote-as 107

neighbor 221.0.0.2 description Router G - peering router

- neighbor 221.0.0.2 update-source loopback 0
- neighbor 221.0.0.4 remote-as 107
- neighbor 221.0.0.4 description Router at HQ
- neighbor 221.0.0.4 update-source loopback 0

Collocation Router H Configuration

- Router H is dedicated to transoceanic link
 - part of ISP core iBGP mesh
- More complex configuration likely CAR, RED, etc
- More complex links likely

e.g satellite uplink for low revenue latency insensitive traffic

Collocation

- Richer interconnectivity possible
- Better redundancy possible
- Overall advantage control!

Summary

- Advanced Community Usage
- Transit
- Collocation
- Any questions?