

BGP multi-homed stub

Wielodrogowość BGP

- Celem zajęć jest zapoznanie się z zasadą działania wielodrogowości protokołu BGP (ang. *multi-homed* BGP)
- W ramach zajęć zostaną uruchomione cztery rutery BGP
- Rutery będą wstępnie przygotowane do pracy (konfiguracja sieciowa, częściowa konfiguracja BGP)
- Podczas zajęć należy:
 - Skonfigurować łącze podstawowe (ang. *primary*) oraz zapasowe (ang. *backup*) z użyciem **prefix-list** oraz **route-map**
 - Zaznajomić się z przeznaczeniem atrybutów BGP: **Local Preference** (w tablicy BGP oznaczonego jako *LocPref*) oraz **MED** (w tablicy BGP oznaczonego jako *metric*)
 - Zrozumieć i sprawdzić działanie **external BGP** (eBGP) oraz **internal BGP** (iBGP)



Università degli Studi Roma Tre
Dipartimento di Informatica e Automazione
Computer Networks Research Group

netkit lab

bgp: multi-homed-stub

Version	2.0
Author(s)	G. Di Battista, M. Patrignani, M. Pizzonia, F. Ricci, M. Rimondini
E-mail	contact@netkit.org
Web	http://www.netkit.org/
Description	configuration of a multi-homed stub network with backup

Modified for the purpose of the IP Networks Lab

copyright notice

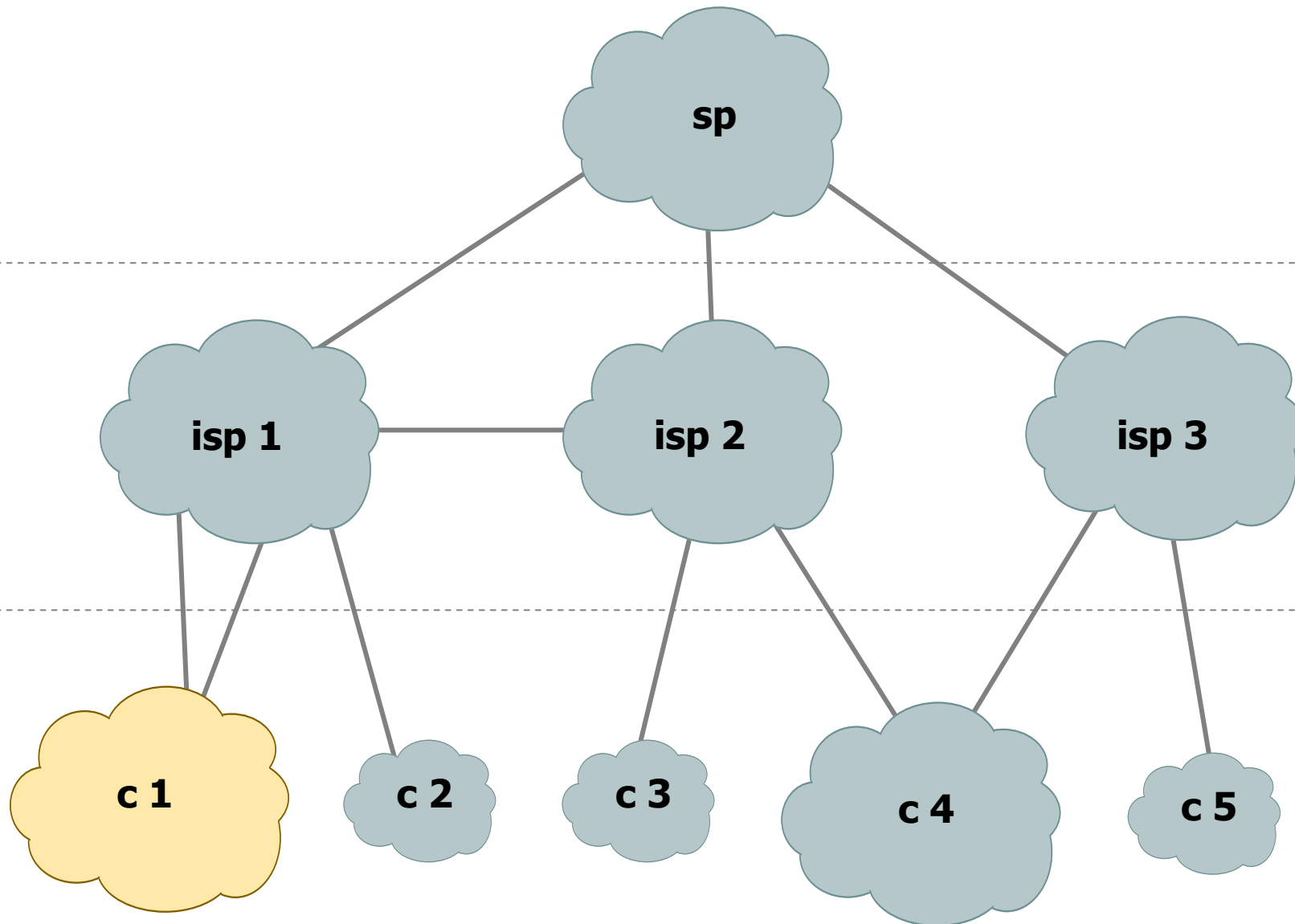
- All the pages/slides in this presentation, including but not limited to, images, photos, animations, videos, sounds, music, and text (hereby referred to as “material”) are protected by copyright.
- This material, with the exception of some multimedia elements licensed by other organizations, is property of the authors and/or organizations appearing in the first slide.
- This material, or its parts, can be reproduced and used for didactical purposes within universities and schools, provided that this happens for non-profit purposes.
- Information contained in this material cannot be used within network design projects or other products of any kind.
- Any other use is prohibited, unless explicitly authorized by the authors on the basis of an explicit agreement.
- The authors assume no responsibility about this material and provide this material “as is”, with no implicit or explicit warranty about the correctness and completeness of its contents, which may be subject to changes.
- This copyright notice must always be redistributed together with the material, or its portions.

multi-homed stub network

backbone

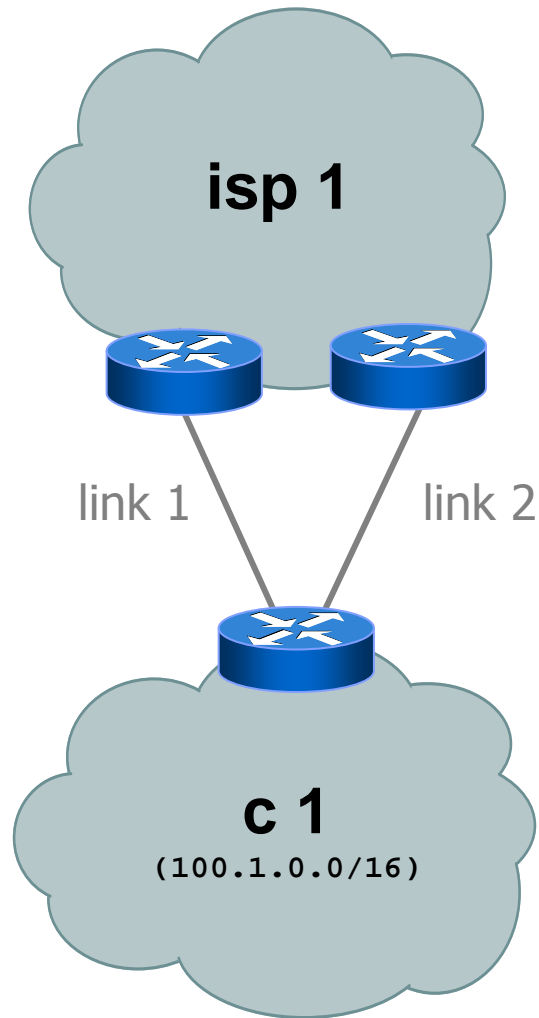
provider

customer

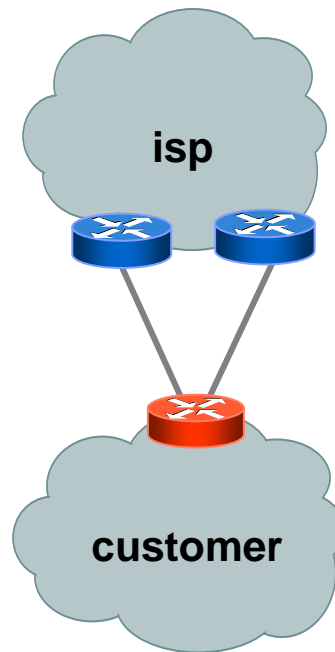


multi-homed stub network

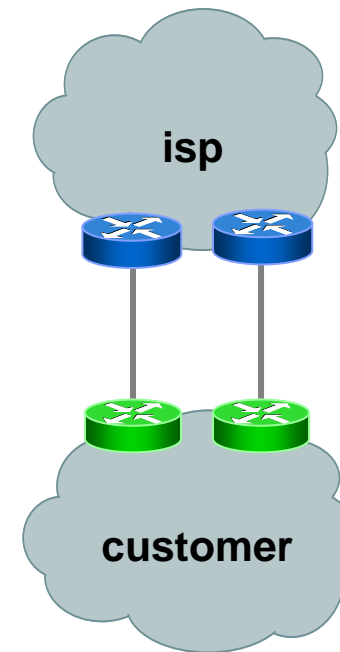
- two links to the same isp
- generally two routers of the customer as are involved



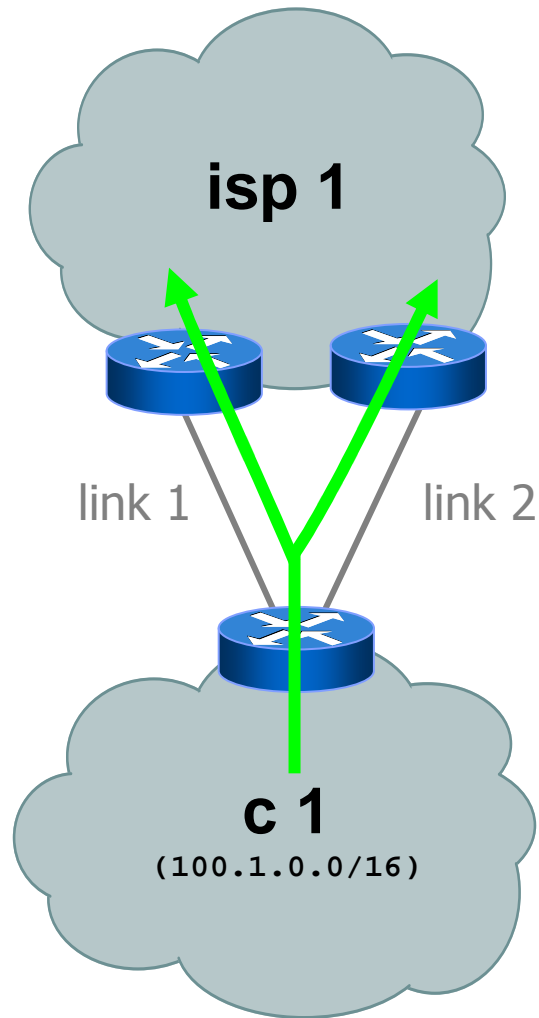
single point
of failure



augmented
redundancy

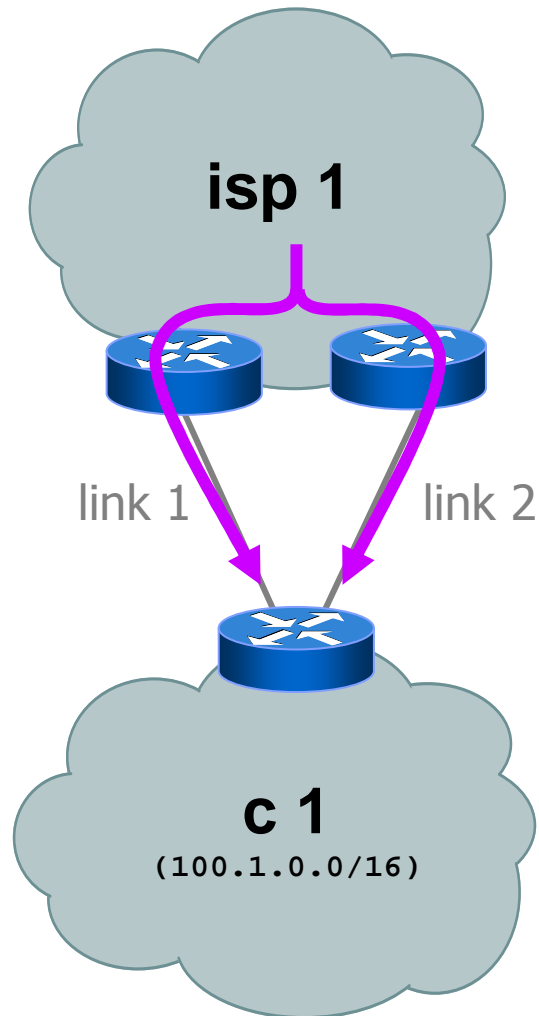


degrees of freedom



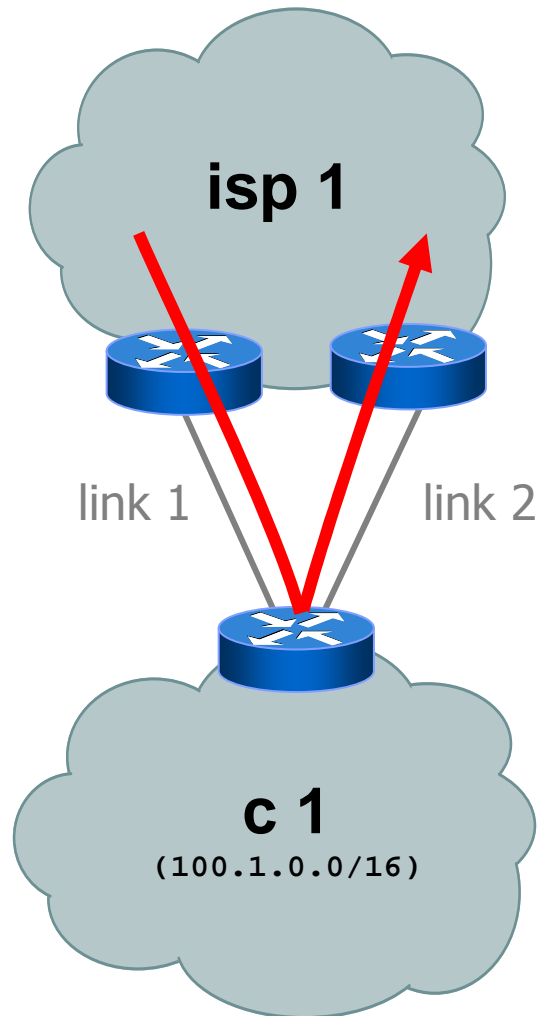
- an outbound packet may be sent through one of the two links in order to reach the internet

degrees of freedom



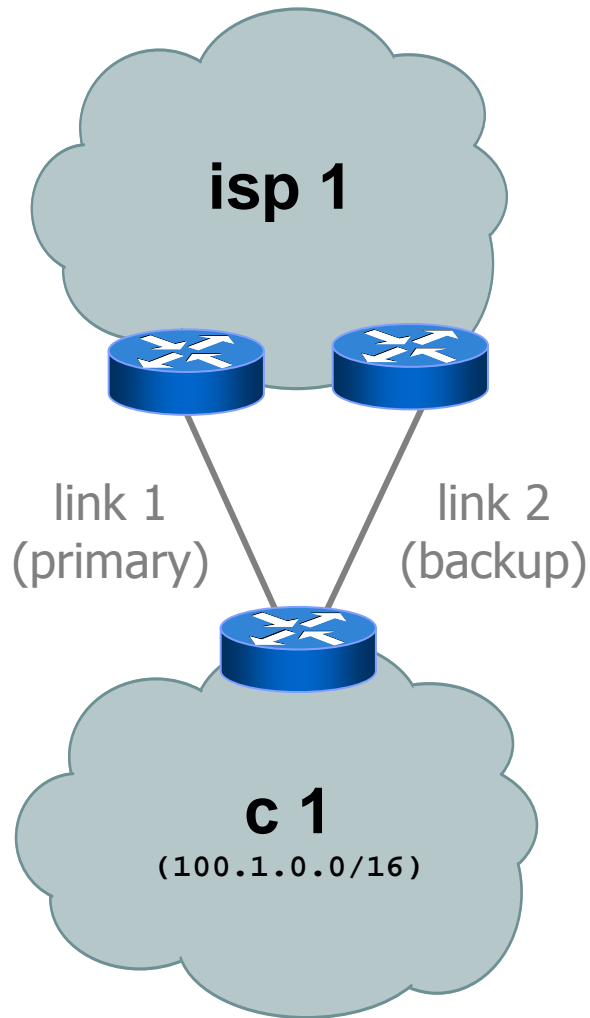
- an outbound packet may be sent through one of the two links in order to reach the internet
- an inbound packet may use any of the two links in order to reach the network

degrees of freedom



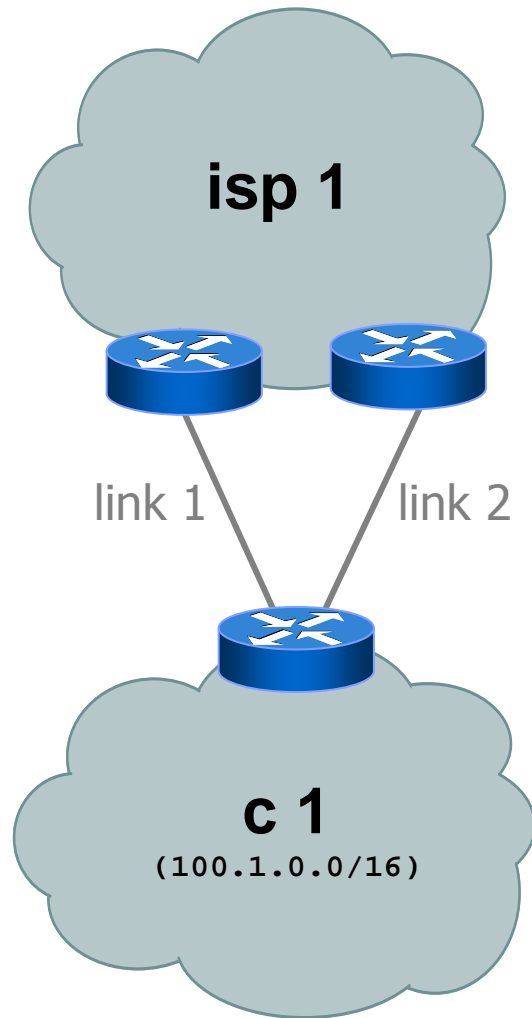
- an outbound packet may be sent through one of the two links in order to reach the internet
- an inbound packet may use any of the two links in order to reach the network
- an internet packet may traverse link 1 and link 2 (or vice versa)

desired policy: backup



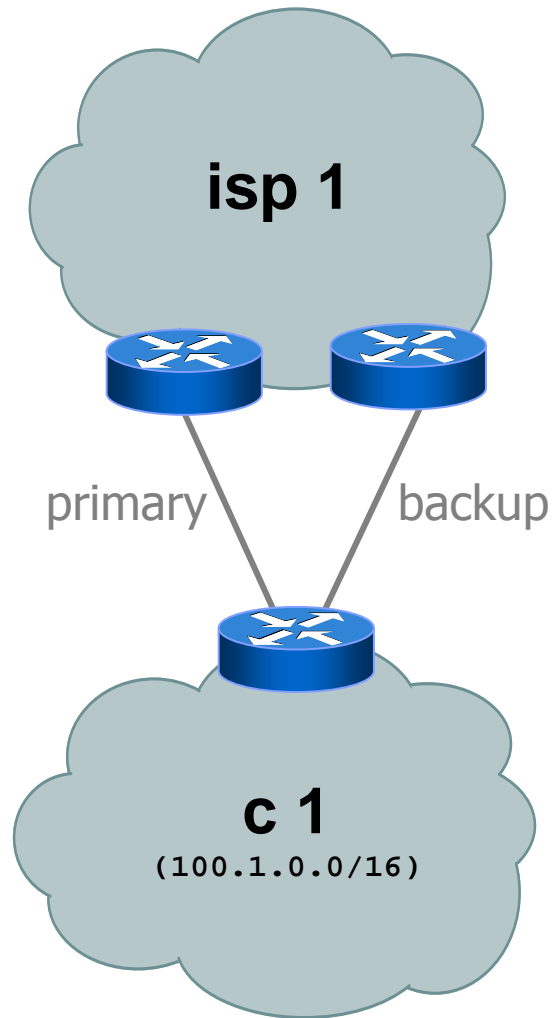
- rule out transit flows
- **inbound traffic:**
 - use link 1
 - use link 2 when link 1 is unavailable
- **outbound traffic:**
 - use link 1
 - use link 2 when link 1 is unavailable

alternatives to using bgp



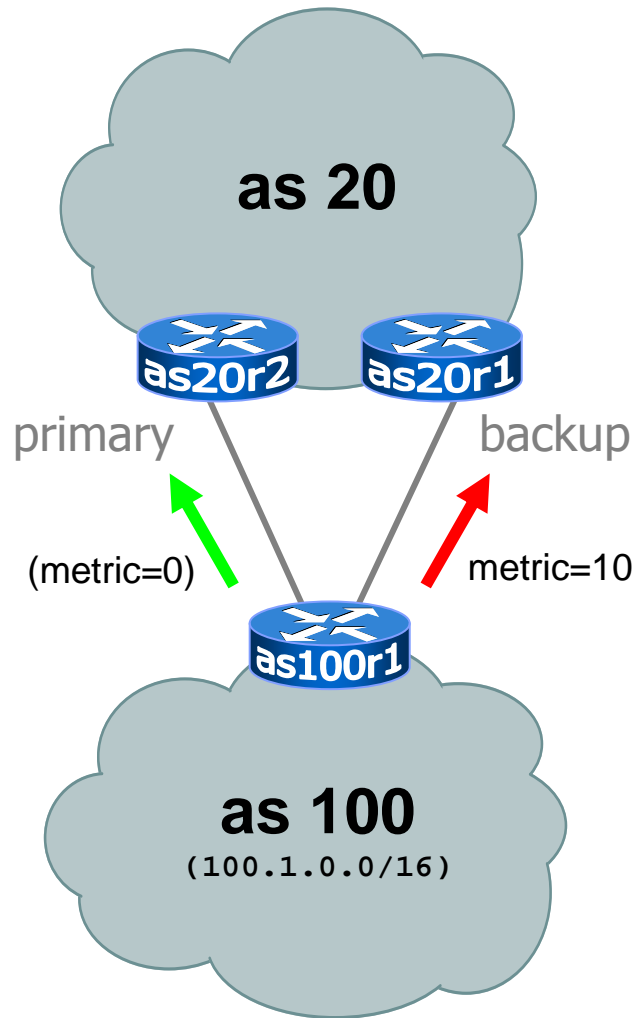
- using an igp (is-is, ospf, rip,...)
 - packets use link 1 or link 2 depending on the shortest path to customer c 1
 - there is no way to rule out transit packets when link 1 and link 2 are on the minimum path between a source and a destination
- using static routes
 - both the routers of the isp and the network have to be coherently configured by hand
 - there is no way to manage an automatic backup mechanism

using bgp



- announce /16 aggregate on each link
 - primary link makes standard announcements
 - backup link increases metric on outbound announcements, and reduces local-pref on inbound announcements
- when one link fails, the announcement of the /16 aggregate via the other link ensures continued connectivity

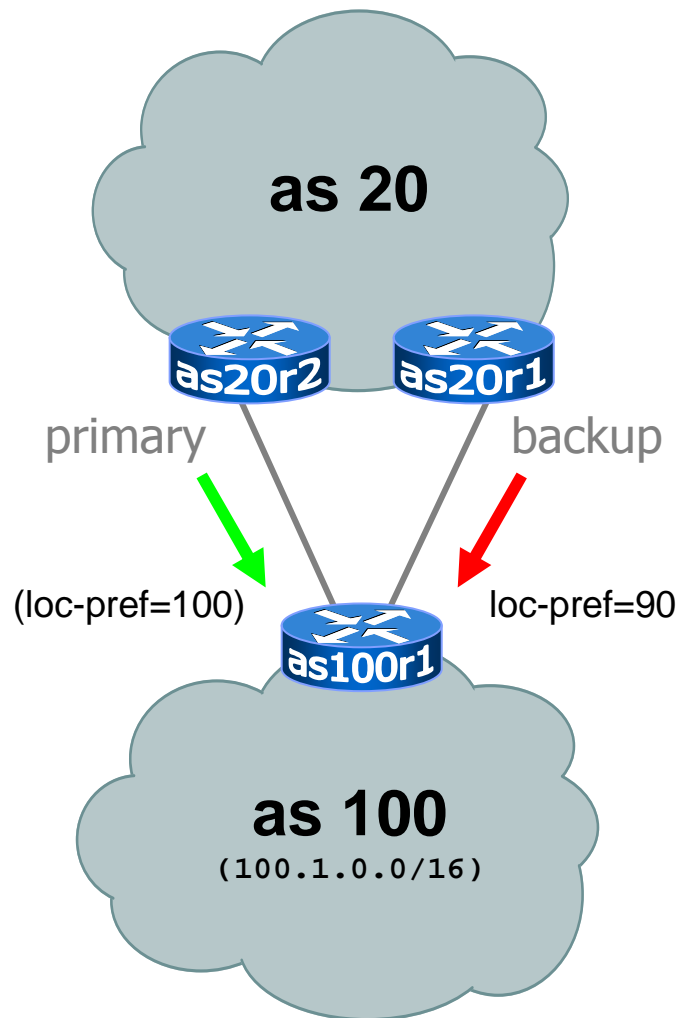
setting metric



- the value of the “**multi-exit-discriminator**” attribute is called “metric”
- upon receiving the same announcement with two different meds, the provider will (hopefully) adopt the one with the smaller one
- **MED is set on outgoing announcements and manages inbound traffic flows**
- metrics are comparable only among announcements coming from the same neighboring as

default value: 0

setting local-preference

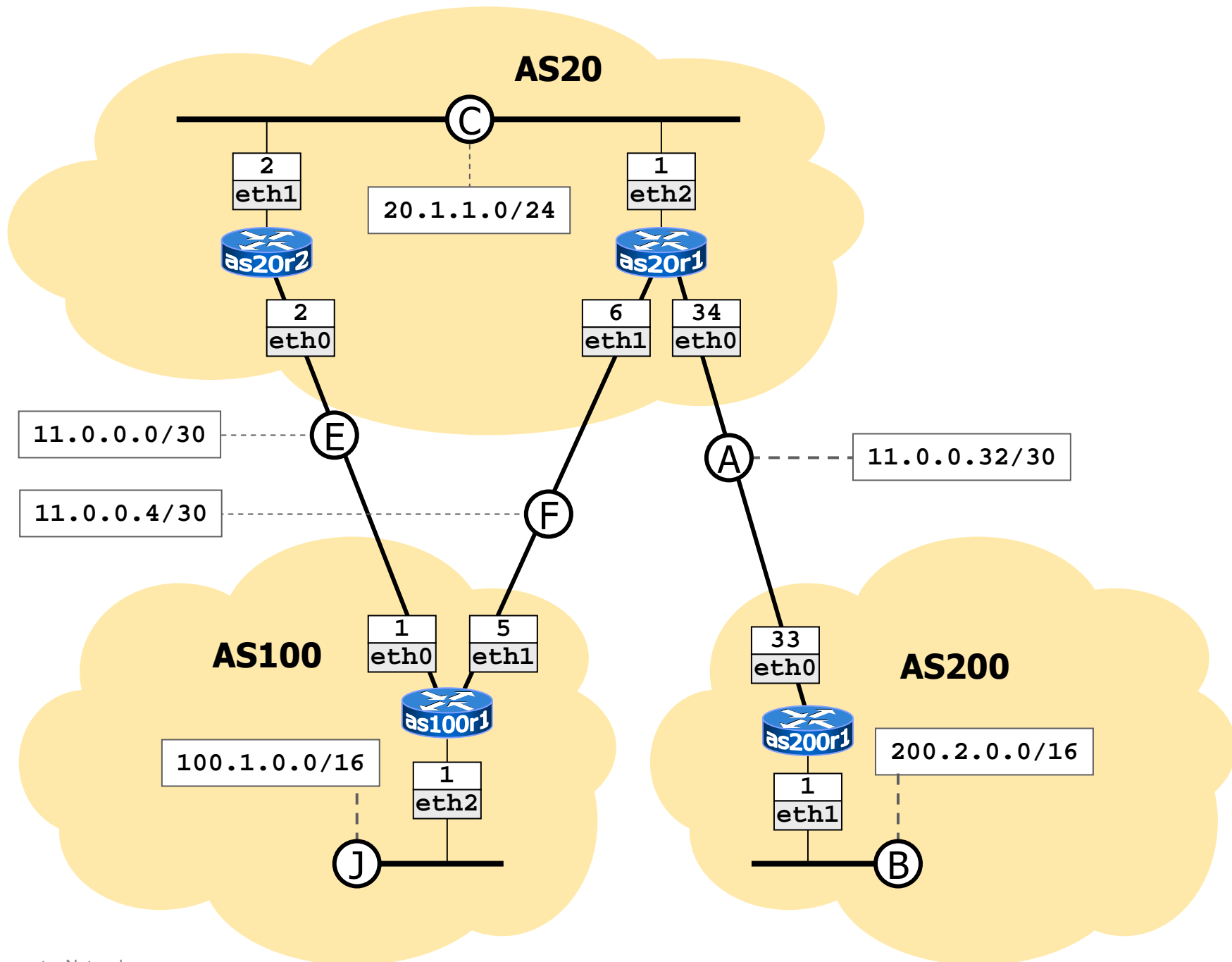


- the customer assigns a lower local-preference to the announcement coming from the backup peer
- the **local-preference** attribute is **checked before as-path length** in the route selection process
- **local-preference** applies to **incoming announcements** and manages **outbound traffic** flows

default value: 100

Lab Scenario Personalization

- Modify the default scenario in the following way:
 - change the **network B** IP address to **<LAB-ID>.2.0.0/16**, where LAB-ID is your personal ID assigned by the lab instructor
 - don't forget to update a **prefix-list** of **as20r1**
- **Note well:** from now-on
 - Command-line commands should reflect this change, therefore there can be differences in the outputs shown in the manual



multi-homed stub

- start the lab

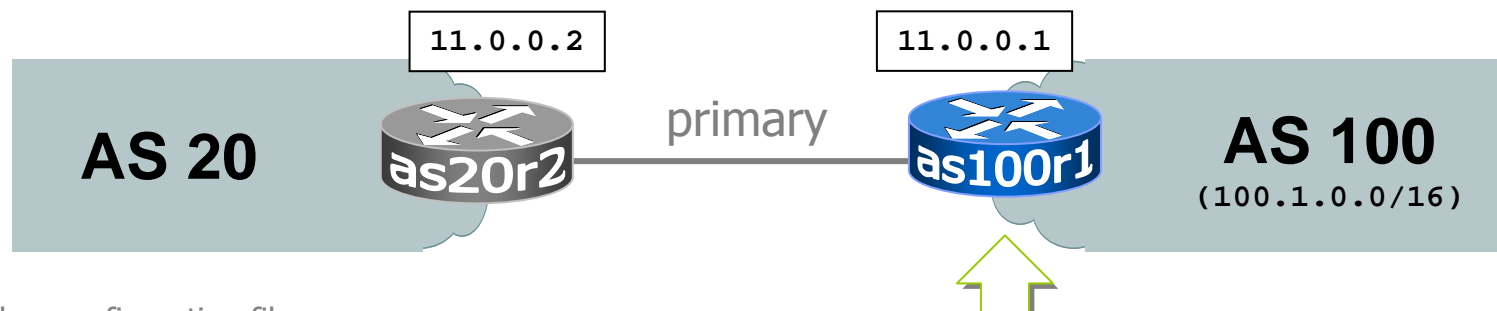
▼ host machine

```
user@localhost:~$ cd netkit-lab_bgp-multi-homed-stub  
user@localhost:~/netkit-lab_bgp-multi-homed-stub$ ./start
```


check existing configuration

- Check existing router configurations
 - What bgp neighbors as200r1 has?
 - What is the bgp configuration on as20r1 and as20r2?
 - What prefixes are allowed/denied by as20r1 and as20r2 routers?

configure as100r1 (primary link)



zebra configuration file

```
! router as100r1 (primary, customer side)
!
router bgp 100
network 100.1.0.0/16
!
neighbor 11.0.0.2 remote-as 20
neighbor 11.0.0.2 description Router as20r2 (primary)
neighbor 11.0.0.2 prefix-list mineOutOnly out
neighbor 11.0.0.2 prefix-list defaultIn in
!
... next slide
```

configure as100r1 (backup link)



zebra configuration file

```
!  
neighbor 11.0.0.6 remote-as 20  
neighbor 11.0.0.6 description Router as20r1 (backup)  
neighbor 11.0.0.6 prefix-list mineOutOnly out  
neighbor 11.0.0.6 route-map metricOut out  
neighbor 11.0.0.6 prefix-list defaultIn in  
neighbor 11.0.0.6 route-map localPrefIn in  
!  
... next slide
```

configure as100r1 cont'd

—zebra configuration file—

```
ip prefix-list mineOutOnly permit 100.1.0.0/16
!
ip prefix-list defaultIn permit 0.0.0.0/0
!
route-map metricOut permit 10
match ip address myAggregate
set metric 10
!
route-map localPrefIn permit 10
set local-preference 90
!
access-list myAggregate permit 100.1.0.0/16
```

multi-homed stub

- restart zebra daemon at as100r1
- ping as100r1 from as200r1

▼ **as200r1**

```
as200r1:~# ping 100.1.0.1
PING 100.1.0.1 (100.1.0.1) 56(84) bytes of data.
64 bytes from 100.1.0.1: icmp_seq=1 ttl=62 time=1.39 ms
64 bytes from 100.1.0.1: icmp_seq=2 ttl=62 time=1.88 ms

--- 100.1.0.1 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1022ms
rtt min/avg/max/mdev = 1.398/1.642/1.886/0.244 ms
```

- everything seems to work fine, but...

multi-homed stub

- there are strange things happening

▼ as200r1

```
as200r1:~# traceroute 100.1.0.1
traceroute to 100.1.0.1 (100.1.0.1), 64 hops max, 40 byte packets
 1  11.0.0.34 (11.0.0.34)  2 ms  1 ms  1 ms
 2  100.1.0.1 (100.1.0.1)  2 ms  2 ms  2 ms
```

- we set up the routing to prefer passing through as20r2! we are not traversing that router! why?
- even more strange:

▼ as100r1

```
as100r1:~# ping 200.2.0.1
PING 200.2.0.1 (200.2.0.1) 56(84) bytes of data.
From 11.0.0.2 icmp_seq=1 Destination Net Unreachable
From 11.0.0.2 icmp_seq=2 Destination Net Unreachable

--- 200.2.0.1 ping statistics ---
2 packets transmitted, 0 received, +2 errors, 100% packet loss,
time 999ms
```

multi-homed stub

- let us have a look at bgp

as20r1

```
as20r1:~# telnet localhost bgpd
```

```
as20r1# sh ip bgp
BGP table version is 0, local router ID is 20.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, R Removed
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop        Metric LocPrf Weight Path
*> 11.0.0.32/30    0.0.0.0          0         32768 i
* i100.1.0.0/16    11.0.0.1          0        100      0 100 i
*>                11.0.0.5         10         0 100 i
*> 200.2.0.0/16    11.0.0.33         0         0 200 i

Total number of prefixes 3
```

why is bgp choosing to pass through 11.0.0.5 rather than 11.0.0.1?

multi-homed stub

- the point of view of as20r2

```
as20r2
as20r2:~# telnet localhost bgpd
.....

as20r2# sh ip bgp
BGP table version is 0, local router ID is 20.1.1.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, R Removed
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network        Next Hop           Metric LocPrf Weight Path
*>i11.0.0.32/30    20.1.1.1             0     100      0  i
* i100.1.0.0/16    11.0.0.5            10     100      0 100  i
*>                11.0.0.1             0     100      0 100  i
* i200.2.0.0/16    11.0.0.33            0     100      0 200  i

Total number of prefixes 3
```

200.2.0.0 is in the table (just 1 entry) but is not selected as the best

multi-homed stub

- the configuration is wrong; ibgp and igp do not interplay properly in as20
 - no igp tells as20r1 how to reach next-hop 11.0.0.1
 - no igp tells as20r2 how to reach next-hop 11.0.0.33
 - since the **next-hops learned via ebgp are not reachable** (i.e., the recursive lookup fails), bgp does not use them
- notice that the ping from as200r1 to 100.1.0.1 works
 - forward path: 11.0.0.34, 11.0.0.5
 - backward path: 11.0.0.2, 20.1.1.1, 11.0.0.33 (have a look with a sniffer placed inside as20r1)

multi-homed stub

- how to fix?
- several possible solutions
 - activate rip in as20
 - add static routes in as20r1 and as20r2
 - ...
- the rip solution; on both as20r1 and as20r2 do:
 - configure rip (e.g. edit `/etc/zebra/ripd.conf`)
 - `router rip`
 - `network X.X.X.X/Y`
 - `redistribute connected`
 - activate rip (edit `/etc/zebra/daemons`)
 - restart zebra (`/etc/init.d/zebra restart`)

multi-homed stub

- how to check that it works?
 - perform a `show ip bgp` on all routers
 - check with `route` on all routers
 - perform `pings` and `tracert` from/to several sources/destinations
- example:

```
as100r1
as100r1:~# traceroute 200.2.0.1
traceroute to 200.2.0.1 (200.2.0.1), 64 hops max, 40 byte packets
 1  11.0.0.2 (11.0.0.2)  1 ms  2 ms  1 ms
 2  20.1.1.1 (20.1.1.1)  2 ms  2 ms  2 ms
 3  200.2.0.1 (200.2.0.1)  2 ms  2 ms  2 ms
```

as100r1 is reaching 200.2.0.1 via as20r2 (as it should)

multi-homed stub

- now shut down the primary connection on as100r1

```
as100r1
as100r1:~# telnet localhost bgpd
.....
User Access Verification

Password: zebra
bgpd> enable
Password: zebra
bgpd# configure terminal
bgpd(config)# router bgp 100
bgpd(config-router)# neighbor 11.0.0.2 shutdown
bgpd(config-router)# quit
bgpd(config)# quit
bgpd# write file
Configuration saved to /etc/zebra/bgpd.conf
bgpd# show ip bgp summary
BGP router identifier 100.1.0.1, local AS number 100
2 BGP AS-PATH entries
0 BGP community entries

Neighbor      V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down State/PfxRcd
11.0.0.2      4    20     12      21        0    0    0 00:00:22 Idle (Admin)
11.0.0.6      4    20     12      20        0    0    0 00:07:33 1

Total number of neighbors 2
bgpd#
```

multi-homed stub

- check the backup

```
as100r1
as100r1:~# route
Kernel IP routing table
Destination      Gateway          Genmask          Flags  Metric  Ref    Use  Iface
11.0.0.4         0.0.0.0          255.255.255.252  U      0        0      0    eth1
11.0.0.0         0.0.0.0          255.255.255.252  U      0        0      0    eth0
100.1.0.0        0.0.0.0          255.255.0.0      U      0        0      0    eth2
default         11.0.0.6         0.0.0.0          UG     0        0      0    eth1
```

```
as100r1
as100r1:~# traceroute 200.2.0.1
traceroute to 200.2.0.1 (200.2.0.1), 64 hops max, 40 byte packets
 1  11.0.0.6 (11.0.0.6)  2 ms  2 ms  2 ms
 2  200.2.0.1 (200.2.0.1)  2 ms  2 ms  1 ms
```

multi-homed stub

- restart the primary connection and check that the primary link is back

```
as100r1
as100r1:~# telnet localhost bgpd
.....
User Access Verification

Password: zebra
bgpd> enable
Password: zebra
bgpd# configure terminal
bgpd(config)# router bgp 100
bgpd(config-router)# no neighbor 11.0.0.2 shutdown
bgpd(config-router)# quit
bgpd(config)# quit
bgpd# write file
Configuration saved to /etc/zebra/bgpd.conf
bgpd# quit
Connection closed by foreign host.
as100r1:~# traceroute 200.2.0.1
traceroute to 200.2.0.1 (200.2.0.1), 64 hops max, 40 byte packets
 1  11.0.0.2 (11.0.0.2)  1 ms  1 ms  1 ms
 2  20.1.1.1 (20.1.1.1)  1 ms  2 ms  2 ms
 3  200.2.0.1 (200.2.0.1)  2 ms  2 ms  2 ms
as100r1:~#
```

Reporting

- Please deliver the following items to the UPEL system using your account
 1. A photocopy or a screenshot showing the output of the following command executed on router **as100r1**:
 - **Traceroute <Lab-ID>.2.0.1**