

# finding the path

Matsuzaki 'maz' Yoshinobu

<maz@ij.ad.jp>

# keys for today

1. understand your network
2. cooperation

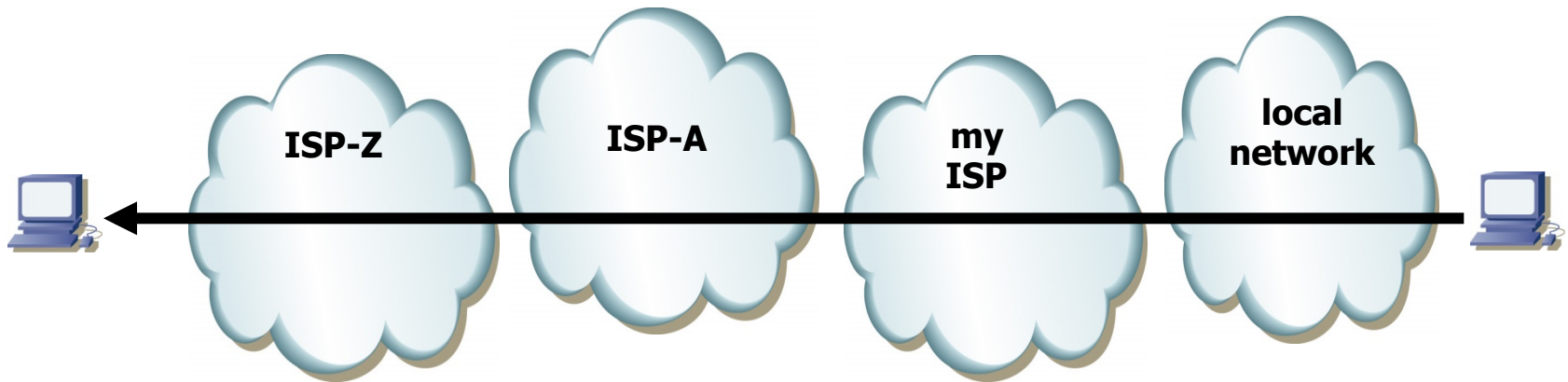
# key: understand your network

- you should be able to fix your network
  - not by guessing
  - not by assuming
  - not by random practice
- based on understanding and knowledge

# case study

- what should we consider today, in case your customer faces a reachability issue?
- you need to find the path that the customer used

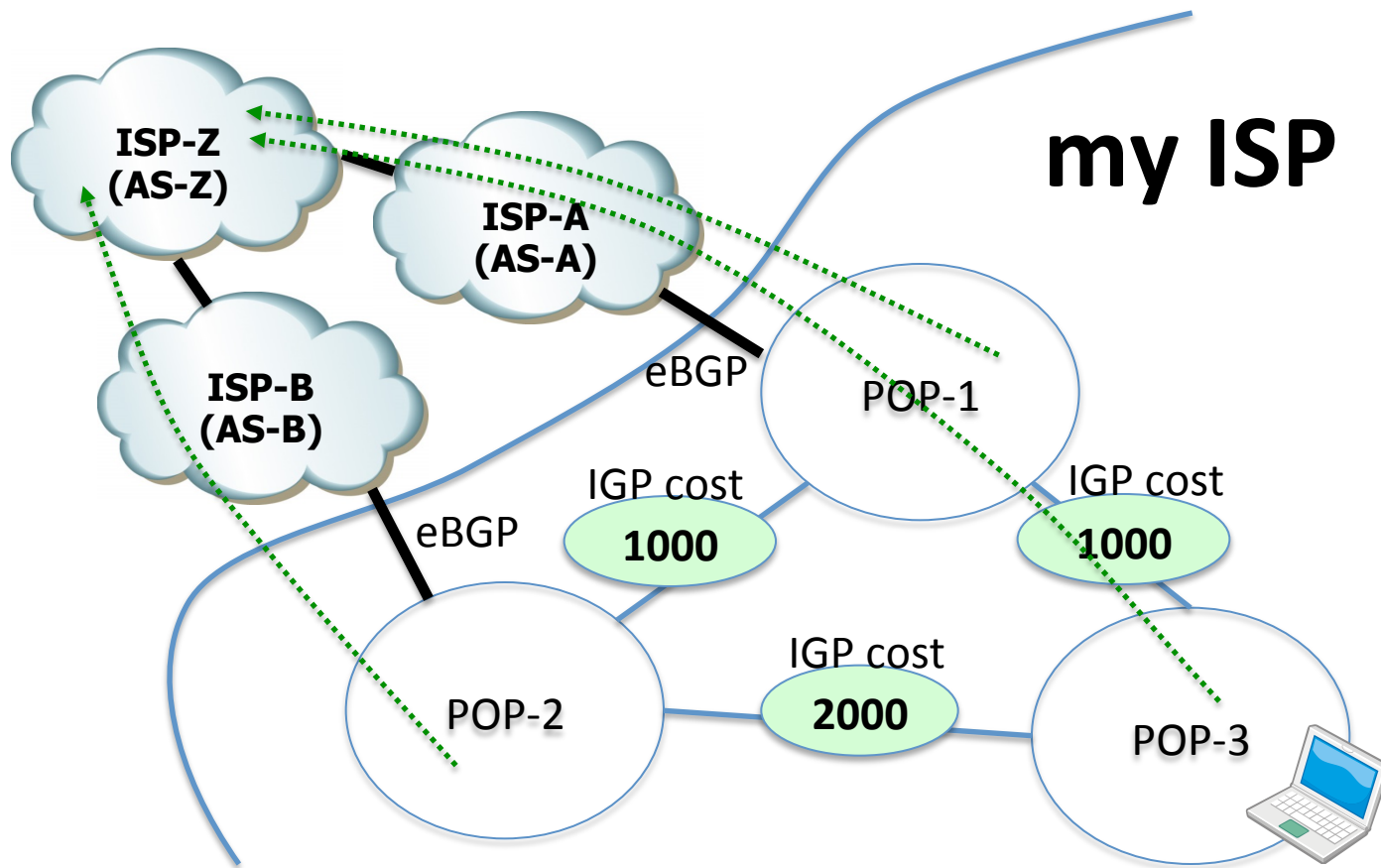
# step1: finding the outgoing path



- you can 'presume' outgoing path
  - BGP, traceroute
- users might use different path
  - Source network and destination IP address is the key

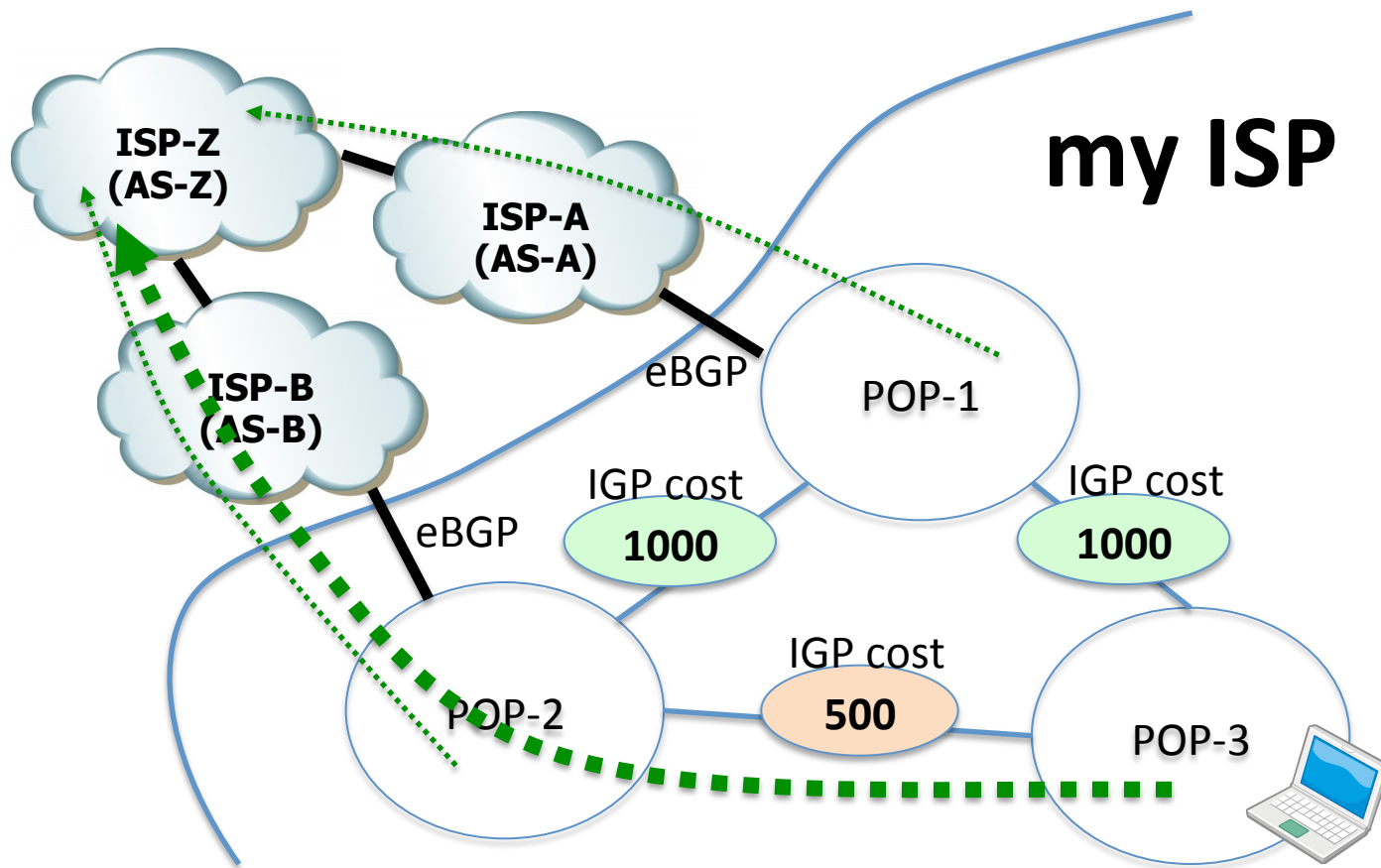
# Multiple exit points

- BGP prefers the 'closest exit'



# Multiple exit points and IGP

- IGP sometimes matters



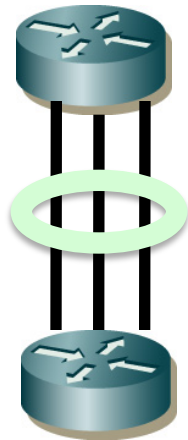
# multiple paths

- more bandwidth, but cost effectively
  - currently 10x10Gbps is cheaper than a 100Gbps
  - just wanted 4Gbps rather than installing 10Gbps
- Layer2
  - link aggregation
- Layer3
  - mutlipath
    - IGP equal cost multipath
    - BGP multipath



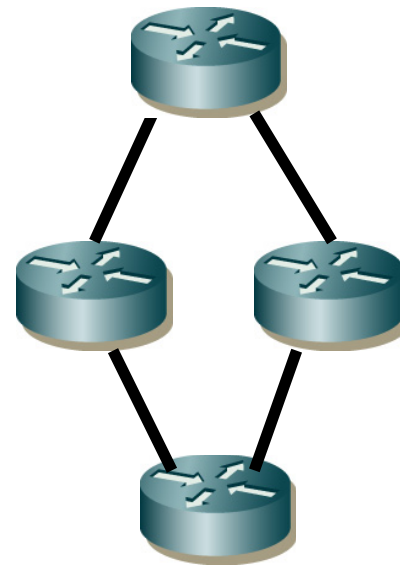
# using multi links at the same time

- Layer2



- lose whole links in case of router failure
- link down policy is usually configurable

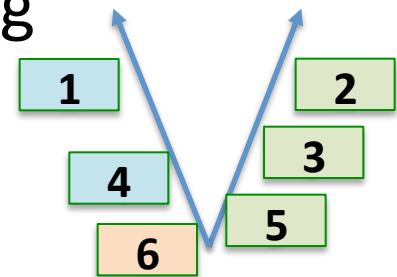
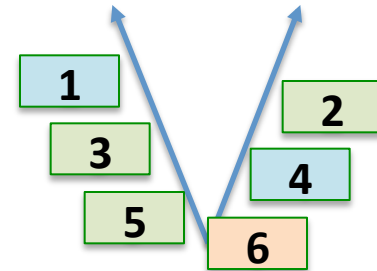
- Layer3



- more redundancy
- not many traffic engineering choices in case of failure

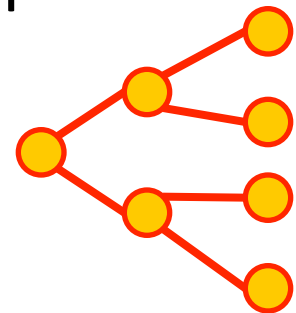
# load-sharing method

- per packet load-sharing
  - good for equal load-sharing
  - possibility of mis-ordering
    - may break communications
- per flow load-sharing (recommended)
  - good for packet ordering/consistent delays
  - possibility of unbalanced load-sharing



# keys for per flow load-sharing

- flow
  - src/dst ip, protocol, src/dst port
- salt
  - to get efficient balance in multi stage multipath
  - A device generates a salt during bootup
- $\text{hash}(\text{flow}, \text{salt}) \rightarrow \text{outgoing link}$



# multipath even in an ISP

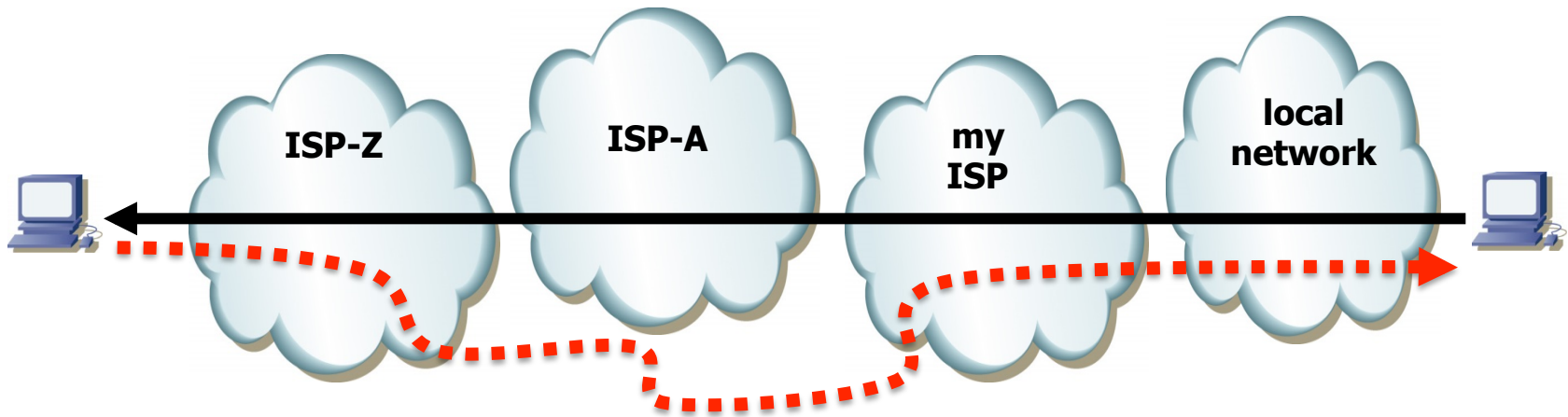
```
%tracert -q1 ftp.iij.ad.jp
 1 202.32.157.2 (202.32.157.2) 0.783 ms
    1 path
 2 210.130.161.90 (210.130.161.90) 0.965 ms
    2 paths
 3 tky006bb01.IIJ.Net (58.138.120.21) 0.976 ms
    1 path (+1backup)
 4 tky001bf01.IIJ.Net (58.138.82.53) 1.470 ms
    6 paths
 5 tky009bf01.IIJ.Net (58.138.80.45) 1.473 ms
    2 paths
 6 osk004bf00.IIJ.Net (58.138.98.126) 10.968 ms
    4 paths
 7 osk004bb11.IIJ.Net (58.138.82.170) 11.468 ms
    1 path (+1backup)
 8 osk004agr00.IIJ.Net (58.138.106.218) 9.970 ms
    1 path (+1backup)
 9 nas200.ftp.pub.2iij.net (202.232.140.170) 12.967 ms !Z
```

96paths  
(excluding backup paths)

# key: cooperation

- most parts of Internet are operated by others
  - out of control
  - each of us is just a part of the Internet
- we need to cooperate to fix a problem
  - by sharing information
  - by exchanging knowledge

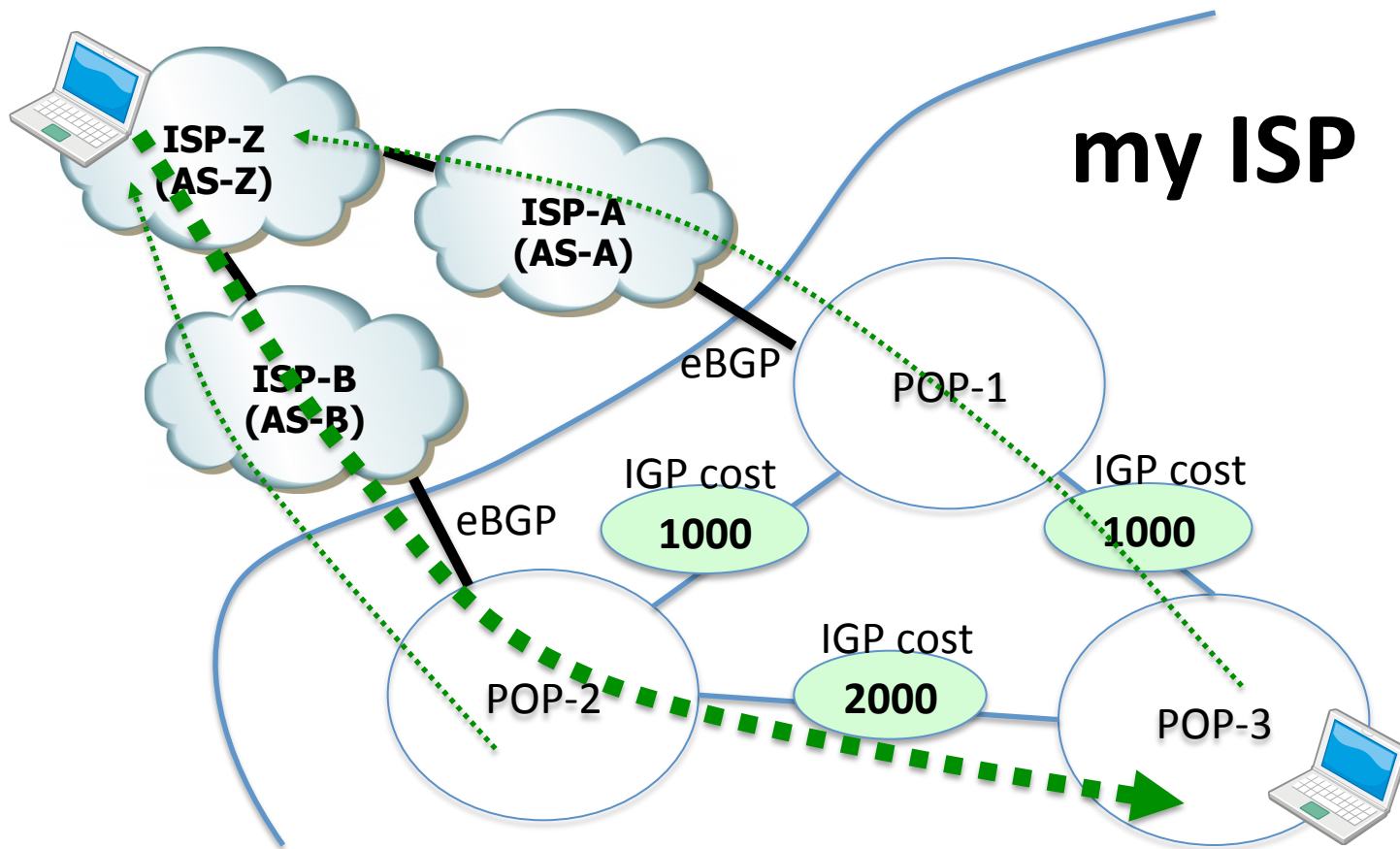
# step2: finding the return path



- almost no clue
  - BGP does not tell much about the reverse path
- difficult to guess
  - other networks' business relationships

# return path from intermediate nodes may vary

- it always happens



# looking glass

- Route Views Project
  - <http://www.routeviews.org/>
  - telnet and generic router CUI
- RIPE Routing Information Service (RIS)
  - <http://www.ripe.net/data-tools/stats/ris/>
  - web UI
- and many others...



# routeviews

```
$ telnet route-views.routeviews.org  
Trying 2001:468:d01:33::80df:3367...  
Connected to route-views.routeviews.org.  
Escape character is '^]'.
```

\$ telnet route-views.routeviews.org

```
*****
```

```
Oregon Exchange BGP Route Viewer  
route-views.oregon-ix.net / route-views.routeviews.org
```

```
route views data is archived on http://archive.routeviews.org
```

```
This hardware is part of a grant from Cisco Systems.  
Please contact help@routeviews.org if you have questions or  
comments about this service, its use, or if you might be able to  
contribute your view.
```

```
This router has views of the full routing tables from several ASes.  
The list of ASes is documented under "Current Participants" on  
http://www.routeviews.org/.
```

```
*****
```

```
route-views.routeviews.org is now using AAA for logins. Login with  
username "rviews". See http://routeviews.org/aaa.html
```

```
*****
```

```
User Access Verification
```

```
Username: rviews  
route-views>
```

Username: rviews

route-views.routeviews.org is now using AAA for logins. Login with username "rviews". See <http://routeviews.org/aaa.html>

# checking routes to SANOG website

```
route-views>show ip bgp 149.20.54.61
BGP routing table entry for 149.20.0.0/16, version 244610
Paths: (31 available, best #26, table Default-IP-Routing-Table)
Not advertised to any peer
286 6939 1280
 134.222.87.1 from 134.222.87.1 (134.222.85.99)
  Origin IGP, metric 650, localpref 100, valid, external
  Community: 286:18 286:19 286:28 286:29 286:800 286:888 286:3031
3303 1280
 164.128.32.11 from 164.128.32.11 (138.187.128.158)
  Origin IGP, localpref 100, valid, external
  Community: 3303:1004 3303:1005 3303:3061
16150 6939 1280
 217.75.96.60 from 217.75.96.60 (217.75.96.60)
  Origin IGP, metric 0, localpref 100, valid, external
  Community: 16150:63392 16150:65415
1221 4637 1280
 203.62.252.186 from 203.62.252.186 (203.62.252.186)
  Origin IGP, localpref 100, valid, external
3333 3356 1280
 193.0.0.56 from 193.0.0.56 (193.0.0.56)
  Origin IGP, localpref 100, valid, external
3257 1280
 89.149.178.10 from 89.149.178.10 (213.200.87.91)
  Origin IGP, metric 10, localpref 100, valid, external
  Community: 3257:4580 3257:5010
3267 6939 1280
 194.85.40.15 from 194.85.40.15 (193.232.80.7)
  Origin IGP, localpref 100, valid, external
2497 1280
 202.232.0.2 from 202.232.0.2 (202.232.0.2)
  Origin IGP, localpref 100, valid, external
:
```

```
route-views>show ip bgp 149.20.54.61
```

# RIPE RIS looking glass

The screenshot shows the RIPEstat website interface. The browser address bar displays <https://stat.ripe.net/widget/looking-glass#w.resource=149.20.54.61>. The main navigation bar includes links for Internet Coordination, Data & Tools, LIR Services, and RIPE Community. Below this is a search bar with the text "Search Site" and a "Search" button. A secondary navigation bar contains links for RIPE Database, Statistics, RIPE Labs, DNS, RIPE Atlas, RIPEstat, and Developer Documentation. At the bottom of this bar are links for RIPEstat Home, About RIPEstat, Documentation, Use Cases, and Login.

You are here: Home > [Data & Tools](#) > RIPEstat > looking-glass

The screenshot shows the "BGP Looking Glass (149.20.54.61)" widget. At the top, there is a text input field with the placeholder "Reload this widget by entering a resource here" and a refresh icon. Below this is an "Advanced Settings" section with a gear icon. The main content area displays a hierarchical view of network information:

- 13 RRCs see 112 peers announcing 149.20.54.61 originated by AS1280. [EXPAND EVERYTHING]
- RRC00 in Amsterdam, Netherlands sees 1 ASN originating 149.20.0.0/16. (AS1280)
- AS1280 is seen as the origin by 14 peers.
- 213.200.87.254 is announcing route AS3257 AS1280.

A detailed view of the route announcement is shown below:

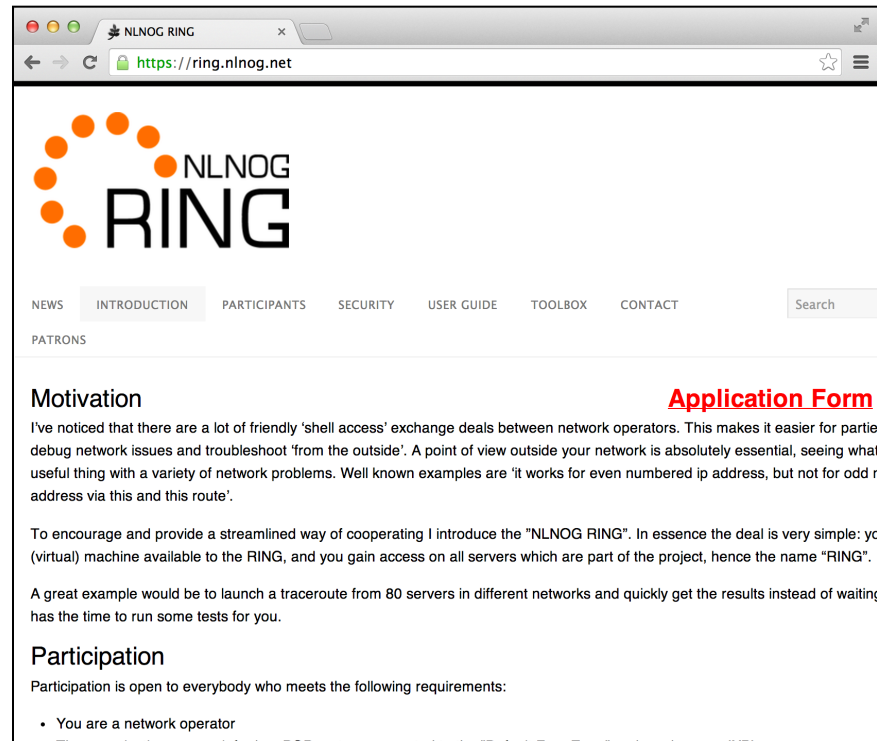
```
213.200.87.254 from 213.200.87.254 (213.200.87.91)
Origin IGP, metric 10, localpref 100, valid, external,
best
Community: 3257:4580 3257:5010
Last update: Sat Jun 21 08:06:17 2014
```

The "Quick Links" sidebar contains the following links:

- Looking For Abuse Information >
- FAQ >
- Feedback >
- Documentation >
- Mobile Version >
- Widget List >

# NLNOG RING – <https://ring.nlnog.net/>

- ‘shell access’ exchange
  - ssh
- usual debug tools
  - ping, traceroute
- useful tools
  - ring-ping
- many participants
  - 259ASNs



# RING servers



# ring-ping

```
iiij@iiij01:~$ ring-ping -v www.sanog.org
```

```
telus01:      55.784
hibernia01:   63.766
icannndns02:  77.196
hibernia02:   153.499
cybercom01:   195.004
telecifyfi01: 175.362
fullsave01:  167.548
funet01:     186.218
selectel01:  199.938
hivane01:    158.808
itps01:      135.921
cybervers01:  9.258
ntt01:       53.480
bogalnet01:  198.277
viatel02:    154.113
nynex01:     161.549
viatel01:    156.817
businessconnect01: 159.376
iucc01:      225.642
mknetzdienste01: 164.888
:
```

```
www.sanog.org - 249 servers: 152ms average
www.sanog.org - unreachable via: occaid01
```

```
iiij@iiij01:~$ ring-ping -v www.sanog.org
```

```
www.sanog.org - 249 servers: 152ms average
www.sanog.org - unreachable via: occaid01
```

# summary

- The internet is composed of simple things
- We have very limited views of the internet
  - traceroute, pings, routing tables
  - many parts of the network are hidden
- what we need:
  - better understanding
  - more cooperation