NH diversity in a research network model

Cristel Pelsser
Internet Initiative Japan (IIJ)

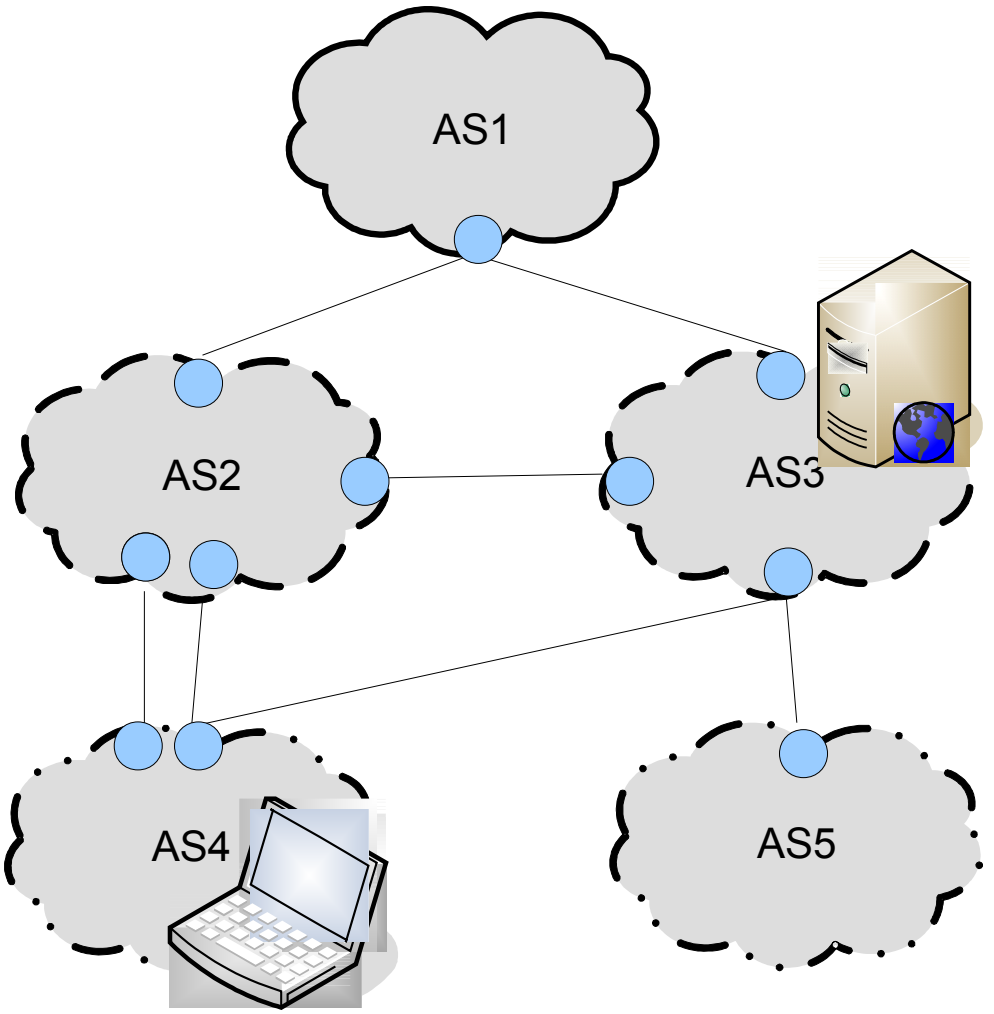
This work was done while I was on Post-Doc at NTT

Agenda

- Background
  - Brief introduction to BGP
  - Next-Hop (NH) diversity
  - iBGP topology design
- Evaluation with C-BGP
  - Why C-BGP
  - WIDE network model
  - C-BGP model
  - Evaluation

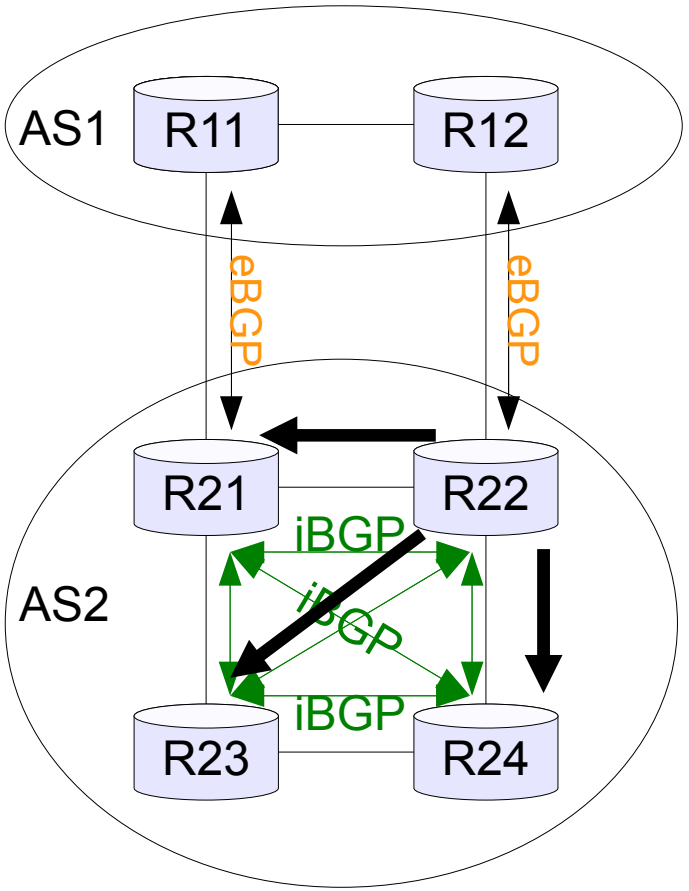
Interdomain routing with BGP

- The Internet is composed of **domains**
  - A domain is also called an Autonomous System (**AS**)
- **BGP** distributes **routes** for destinations outside a domain
  - i.e. Information on how to reach distant destinations
    - The Next-Hop (NH)
      - Router that knows how to forward the traffic to the destination
    - AS-path
    - …

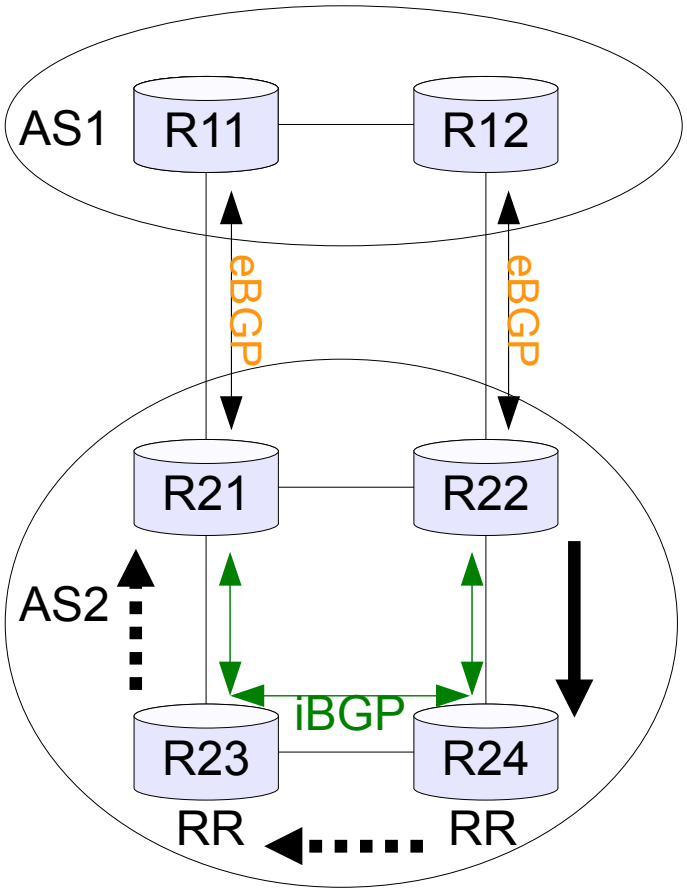


BGP route distribution

- iBGP full-mesh
  - $\frac{n * (n-1)}{2}$ sessions

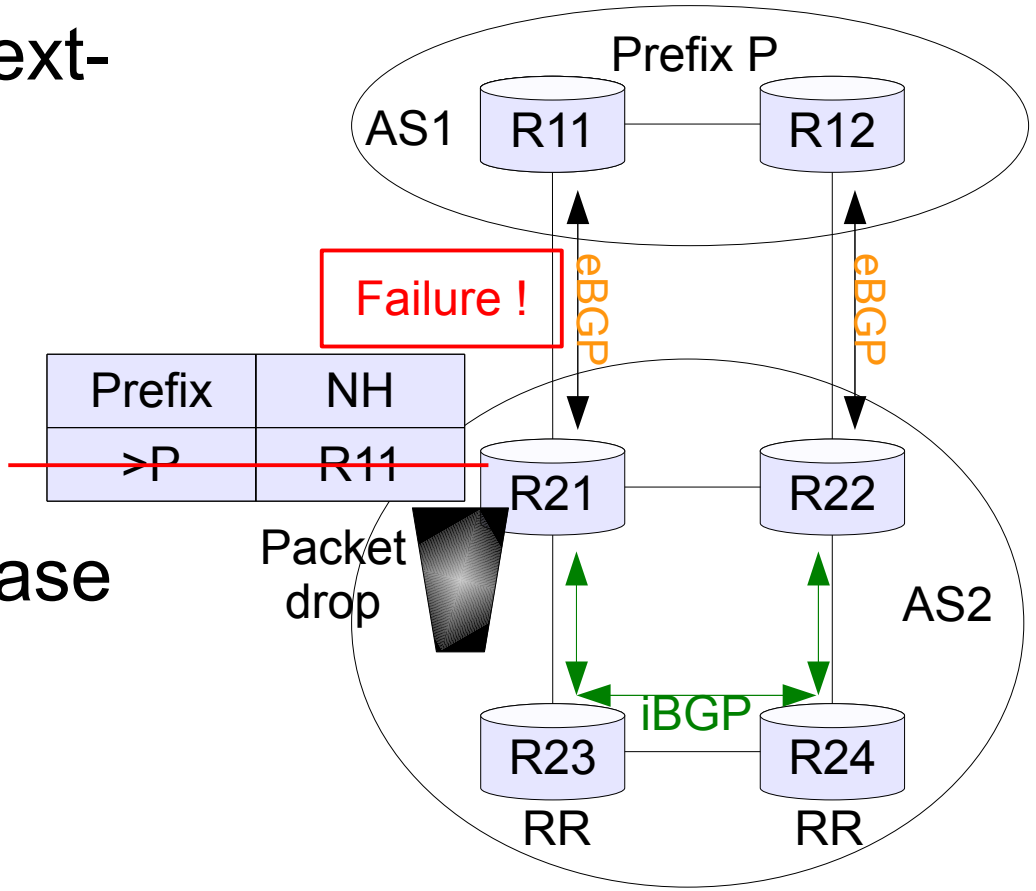


- Route-reflectors
  - less sessions → scalable



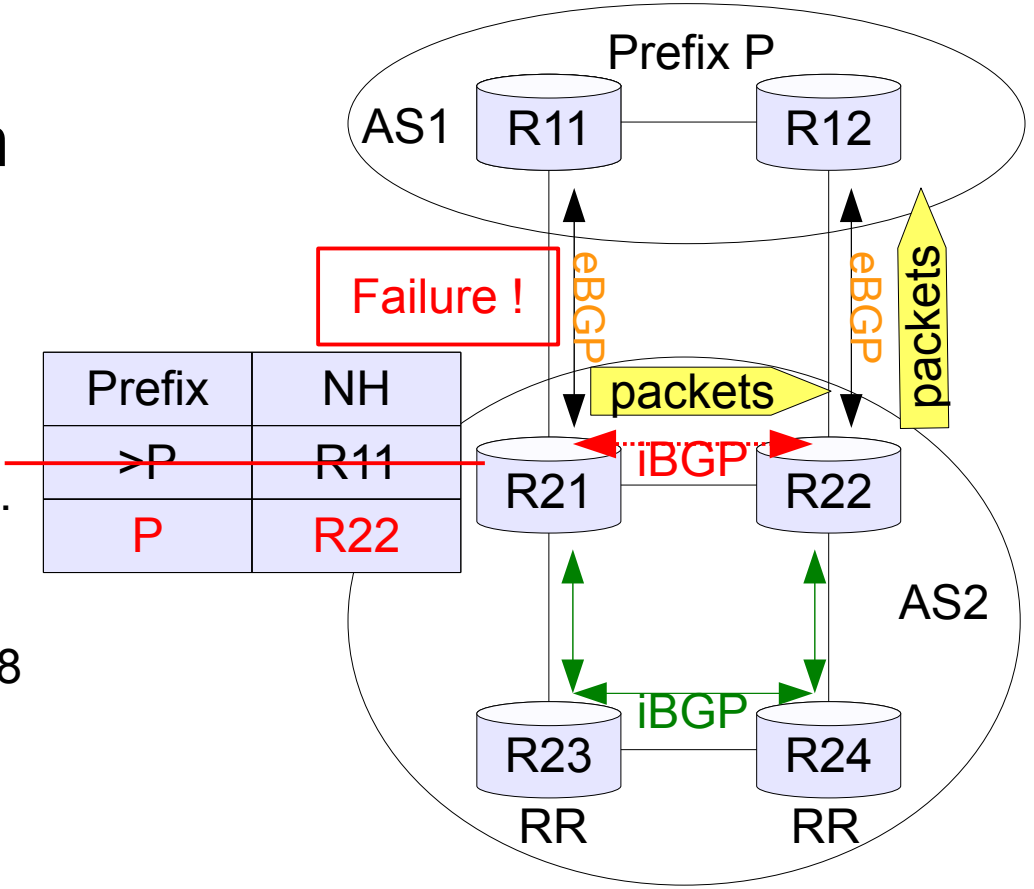
NH diversity

- Definition: each router knows at least two next-hops for each prefix
- Many sparse iBGP topologies do not ensure NH diversity
- Objective: Minimize connectivity loss in case of **failure** of inter-AS resources



NH diversity through iBGP design

- Without modifying BGP, NH diversity can be achieved by **adding iBGP sessions** [1]



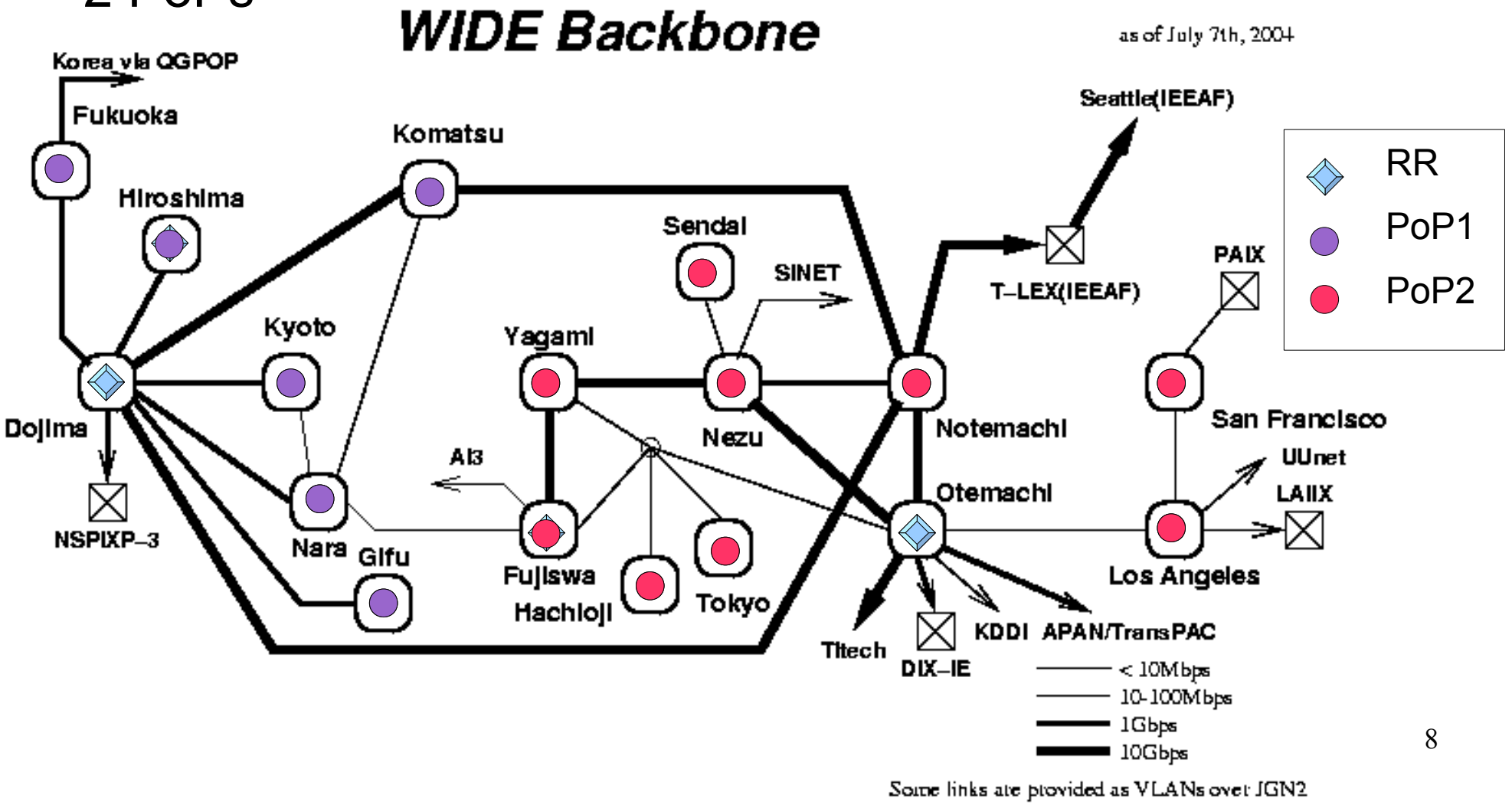
[1] C. Pelsser, T. Takeda, E. Oki and K. Shiomoto. *Improving Route Diversity through the Desing of iBGP Topologies*. ICC'08, May 2008

Evaluation: why C-BGP?

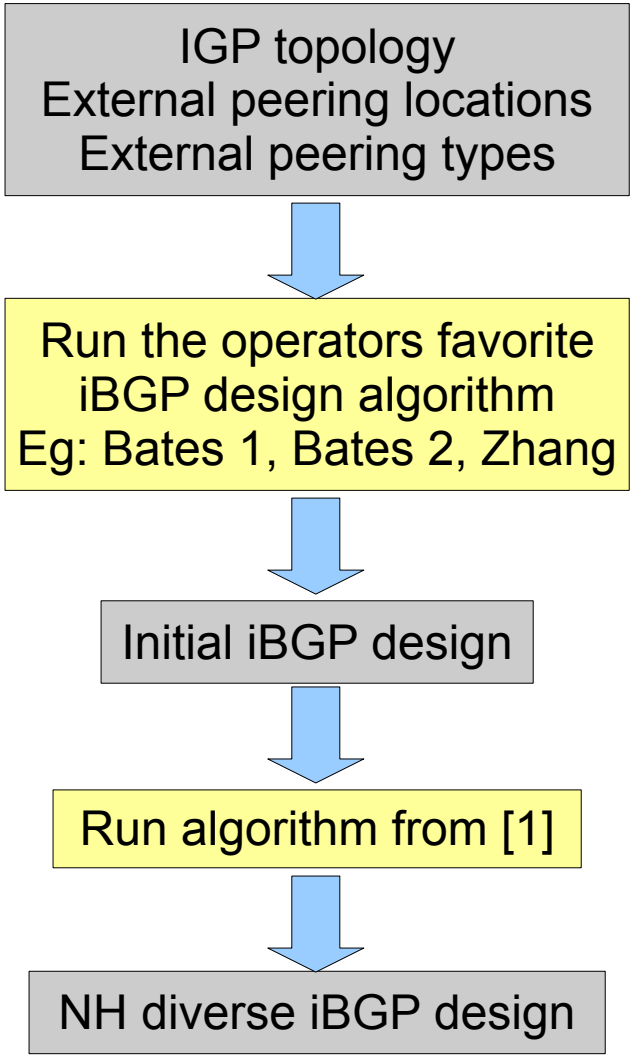
- C-BGP computes the BGP routes for each router in the AS
  - Check if a router has two NHs for a destination
- C-BGP supports
  - Route-reflection
  - Expressive input and output filters
- C-BGP is light-weight
- Great user support from Bruno ;)

WIDE network model

- 17 backbone nodes (public info, dates from 2004)
- 2 PoPs



Evaluation method



Inputs

IGP topology
External peering locations
External peering types

IGP topology

```
# wide backbone
# <node_id> <node_id> <igp_cost> <delay> <bdw>
DOMAIN=1
NAME=wide
router3 router1 1 1 2400
router1 router4 1 1 2400
router1 router5 1 1 2400
…
NETWORK=*
```

External peering types

```
# <provider AS name> <customer AS name> 1
# <peer AS name> <peer AS name> 0
wide ai3 1
asiares wide 0
asiacomm wide 1
wide qgpop 0
wide sinet 1
wide titech 1
usres wide 0
uscomm wide 1
```

External peering locations

```
#inter-AS connectivity
#<domain-name>:<router-id> <domain-name:router-id> (<igp-cost>) <delay> <bdw>
wide:router3 qgpop:koren1 1 1 2400
wide:router1 qgpop:koren1 1 1 2400
wide:router1 sinet:sinet1 1 1 2400
wide:router1 titech:titech1 1 1 2400
wide:router1 asiares:asiares1 1 1 2400
wide:router1 asiacomm:asiacomm1 1 1 2400
...
```

Initial iBGP topology design

Name	hierarchy	top-level full-mesh	PoP full-mesh	RR redundancy
Bates1	no	yes	yes	no
Bates2	no	yes	no	yes
Zhang	yes	yes	no	yes

Run the operators favorite iBGP design algorithm
Eg: Bates 1, Bates 2, Zhang

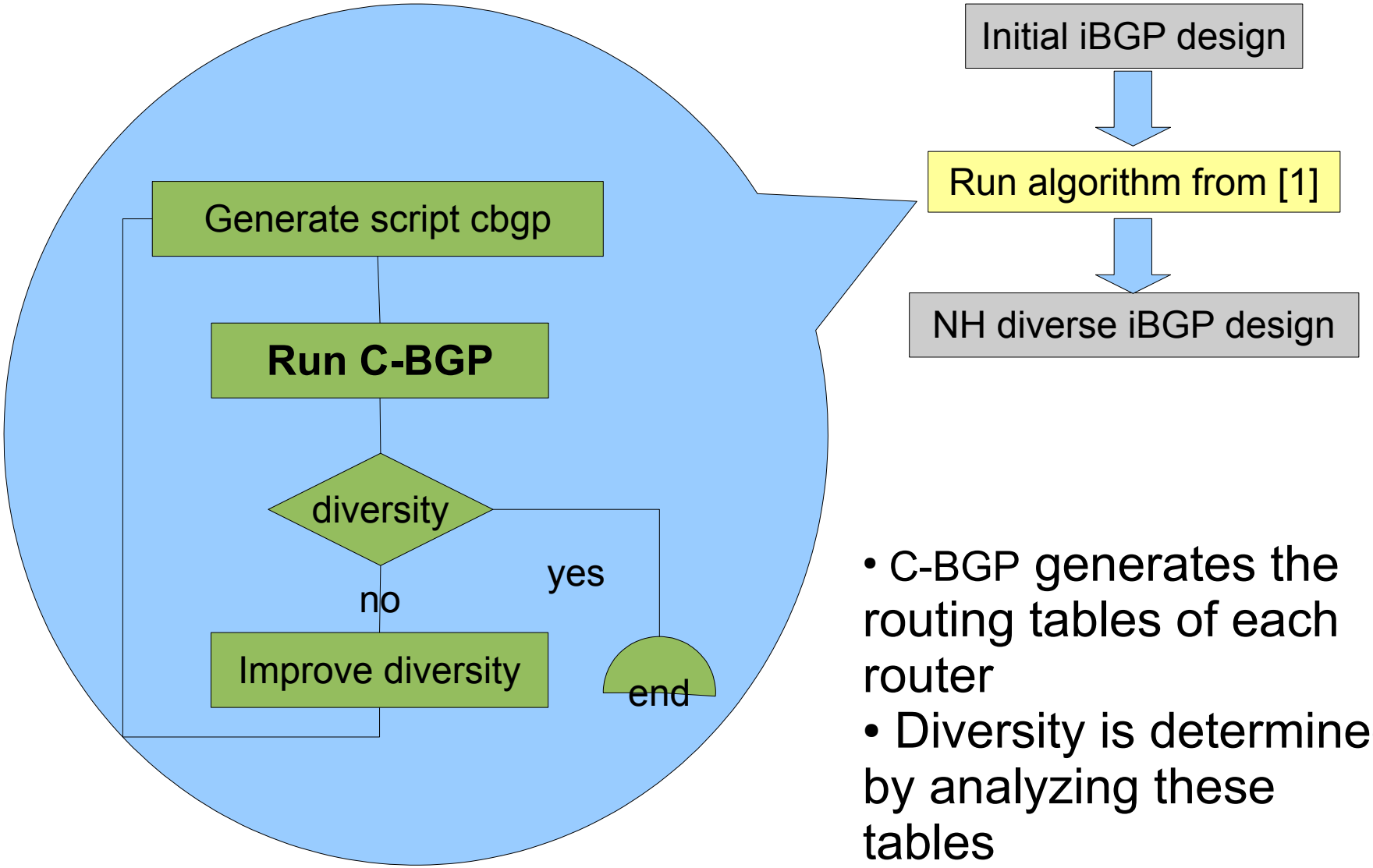
Initial iBGP design

Name	Number of PoPs	Nb RRs per PoP	Nb of top-level RRs	Number of nodes	Number of iBGP sessions
Bates 1	2	1	0	17	67 (49%)
Bates 2	2	2	0	17	32 (24%)
Zhang	17	2	4	51	100 (8%)

Initial iBGP design

```
#<Type> <value>
#<Type> ::= "ASN"|"session"|"cluster-id"
# if <Type> == "ASN", <value> ::= <ASN> ("full-mesh"|"RR")
# if <Type> == "session", <value> ::= ("over"|"up") <node_id> <node_id>
# if <Type> == "cluster-id", <value> ::= <node_id> <cluster-id> where <node_id> is the id of a RR
# -------------------------
ASN 1 RR
session up router3 router1
session up router4 router1
session up router5 router1
session up router6 router1
session up router7 router1
session up router8 router1
session over router3 router4
session over router3 router5
session over router3 router6
…
session over router1 router2
cluster-id router1 1
cluster-id router2 2
# -------------------------
ASN 2 full-mesh
# -------------------------
...
```

Run NH diversity algorithm



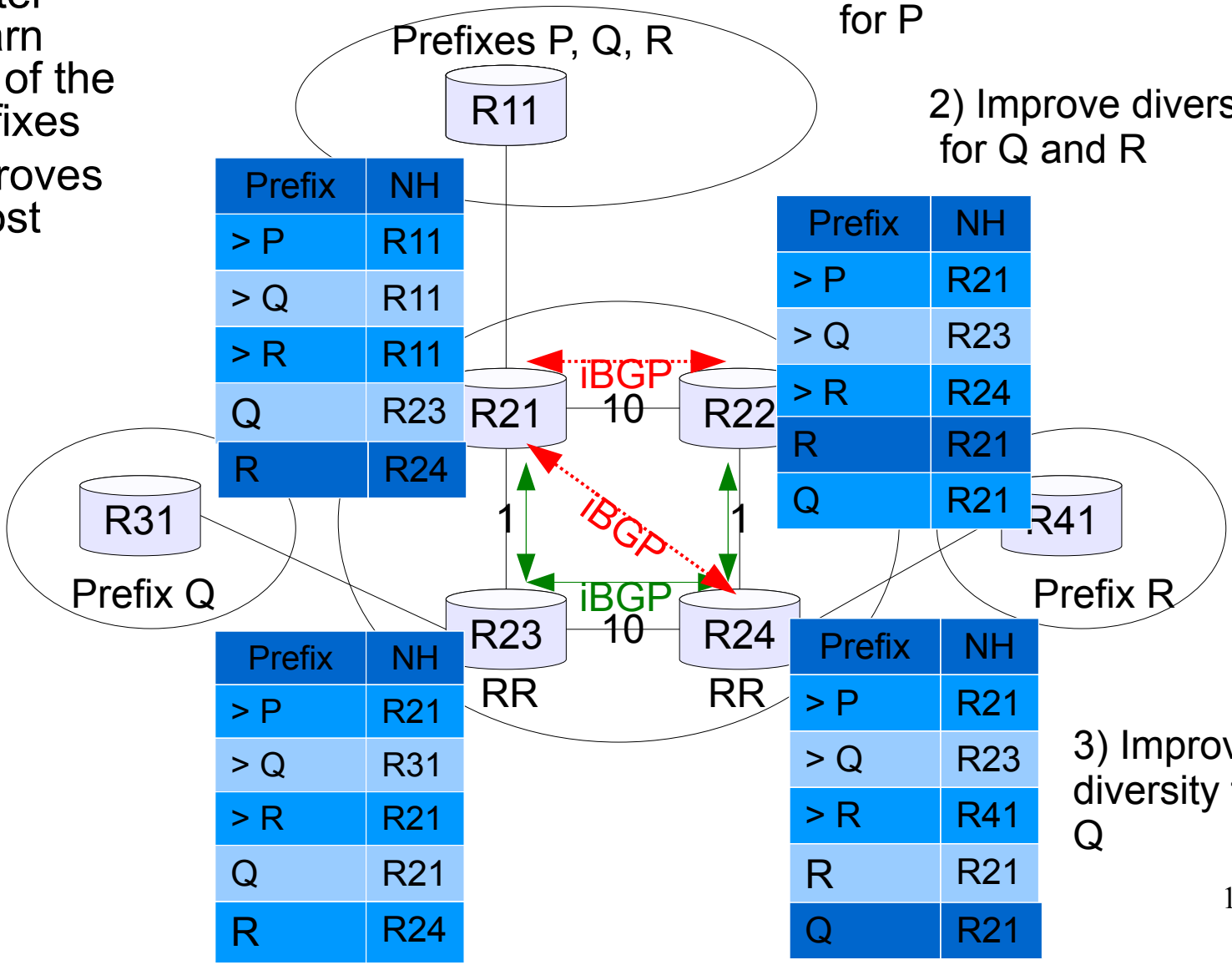
- C-BGP generates the routing tables of each router
- Diversity is determined by analyzing these tables

iBGP session topologies algorithm

- Add an iBGP session to AS Border Router (ASBR) that learn routes for most of the considered prefixes
- This ASBR improves diversity the most

1) Impossible to Increase diversity for P

2) Improve diversity for Q and R



3) Improve diversity for Q

Script C-BGP

Initial

```
bgp router 0.1.0.2
        …
        set cluster-id 1
        add peer 1 0.1.0.14
        peer 0.1.0.14 up
        add peer 1 0.1.0.6
        peer 0.1.0.6 rr-client
        peer 0.1.0.6 up
        add peer 1 0.1.0.4
        peer 0.1.0.4 rr-client
        peer 0.1.0.4 up
        ...
...
```

Normal sessions

Client sessions

NH diverse

```
bgp router 0.1.0.2
        ...
        set cluster-id 1
        add peer 1 0.1.0.14
        peer 0.1.0.14 up
        add peer 1 0.1.0.13
        peer 0.1.0.13 up
        add peer 1 0.1.0.11
        peer 0.1.0.11 up
        add peer 1 0.1.0.10
        peer 0.1.0.10 up
        …
        add peer 1 0.1.0.6
        peer 0.1.0.6 rr-client
        peer 0.1.0.6 up
        add peer 1 0.1.0.4
        peer 0.1.0.4 rr-client
        peer 0.1.0.4 up
...
```

C-BGP output

Adj-rib-in of router 0.1.0.6

Initial

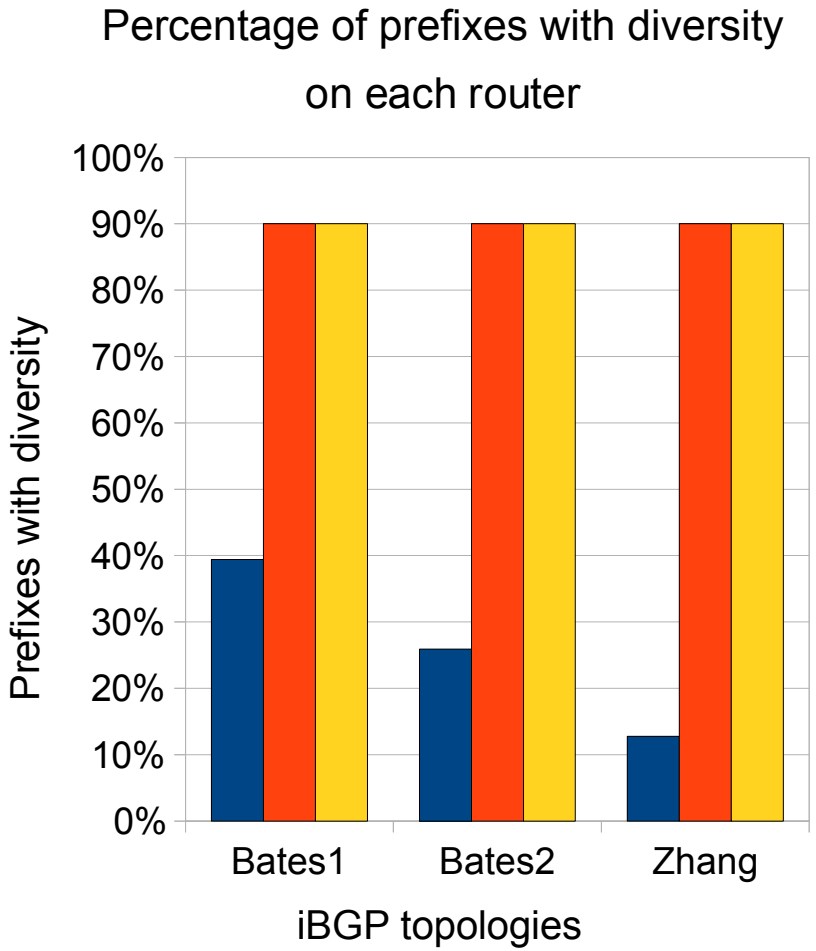
```
0.1.0.6
… <IGP routes removed>
*  1.0.0.0/8    0.1.0.1 0      4294967295     2      i
*> 1.0.0.0/8    0.1.0.2 0      4294967295     2      i
*> 2.0.0.0/8    0.1.0.2 0      4294967295     7      i
*> 3.0.0.0/8    0.1.0.9 0      4294967295     6      i
*> 4.0.0.0/8    0.1.0.2 0      4294967295     8      i
*> 5.0.0.0/8    0.1.0.2 0      4294967295     5      i
*> 6.0.0.0/8    0.1.0.8 0      4294967295     9      i
*> 7.0.0.0/8    0.1.0.2 0      4294967295     5      i
*> 8.0.0.0/8    0.1.0.2 0      4294967295     12     i
*> 9.0.0.0/8    0.1.0.8 0      4294967295     11     i
*> 10.0.0.0/8   0.1.0.2 0      4294967295     12     i
```

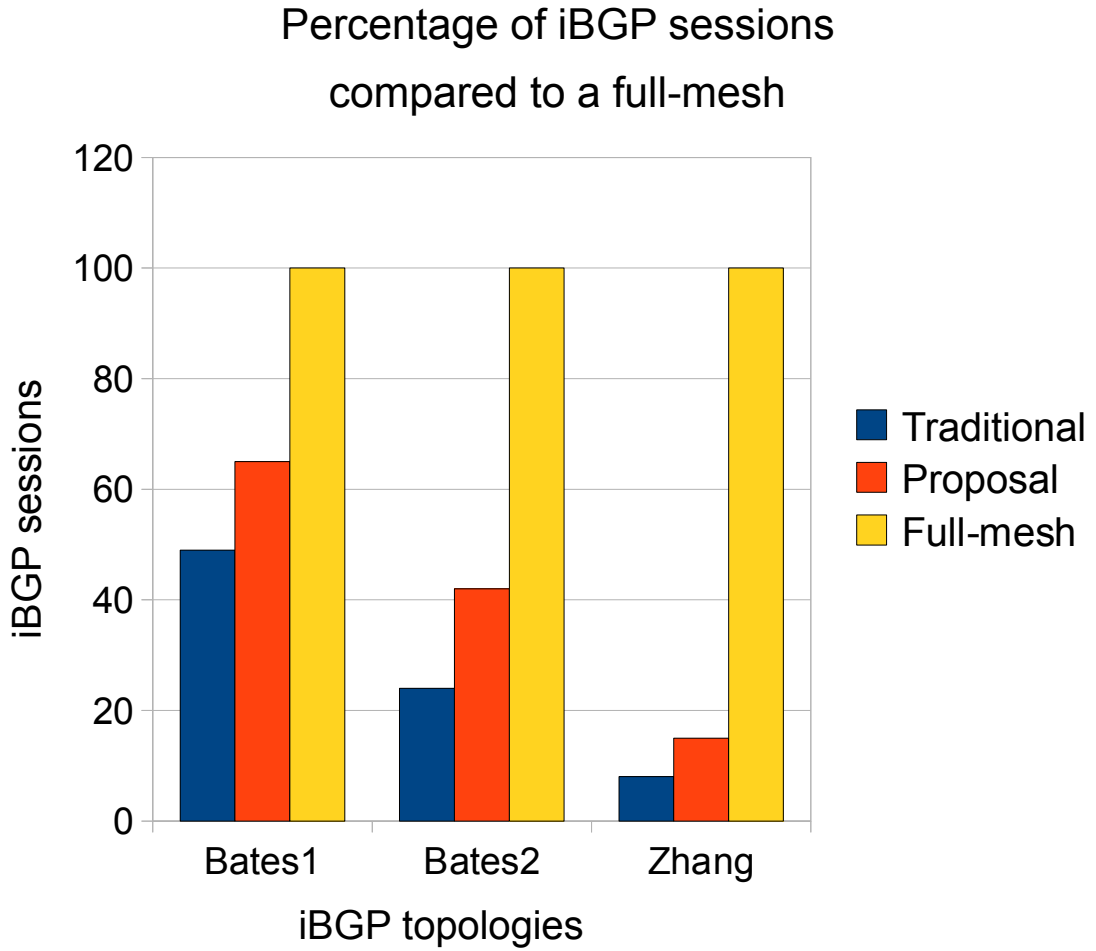
NH diverse

```
0.1.0.6
… <IGP routes removed>
*  1.0.0.0/8    0.1.0.1 ...
*> 1.0.0.0/8    0.1.0.2 ...
*> 2.0.0.0/8    0.1.0.2
*> 3.0.0.0/8    0.1.0.9
*> 4.0.0.0/8    0.1.0.2
*> 5.0.0.0/8    0.1.0.2
*> 6.0.0.0/8    0.1.0.8
*> 7.0.0.0/8    0.1.0.2
*> 8.0.0.0/8    0.1.0.2
*> 9.0.0.0/8    0.1.0.8
*> 10.0.0.0/8   0.1.0.2
*  1.0.0.0/8    0.1.0.14
*  2.0.0.0/8    0.1.0.14
*  3.0.0.0/8    0.1.0.9
*  4.0.0.0/8    0.1.0.14
*  5.0.0.0/8    0.1.0.14
*  6.0.0.0/8    0.1.0.8
...
```

Evaluation

- Same NH diversity level as in a full-mesh
- Low percentage of additional iBGP sessions





Conclusion

- C-BGP use case:

  Evaluation of a solution to reduce connectivity losses during the failure of inter-AS resources for a **research network**

- C-BGP provides route computation
  - Easy setup of the network to evaluate
    - Support of expressive filters
    - Route-reflection, ...
  - Output easy to parse
    - routing tables (best route, learned routes)
    - Traceroutes, ...