Reconsidering the Interdomain Routing Architecture

Olivier Bonaventure

Dept. Computing Science and Engineering
Université catholique de Louvain (Belgium)
http://www.info.ucl.ac.be/~obo

March 17th, 2007

Outline

- id/locator split and map/encaps

- More on the locators

- How to deal with link failures

- How to support traffic engineering

Two different namespaces

- ## Identifiers
  - Not used inside headers of packets forwarded through core Internet
  - Assigned to devices that terminate transport-level connections (hosts, ...)

- ## Locators
  - Used inside headers of packets forwarded through core Internet
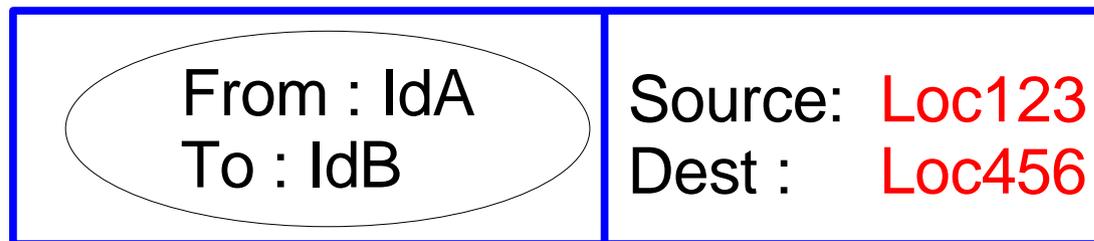 - Should have topological meaning to aid aggregation

The map and encaps paradigm

- How to forward data in next Internet ?

From : IdA
To : IdB

- First operation : map

IdA <=> Loc123
IdB <=> Loc456

- Second operation : encapsulate

From : IdA
To : IdB

Source: Loc123
Dest :   Loc456

Outline

- id/locator split and map/encaps

- More on the locators

- How to deal with link failures

- How to support traffic engineering

Where should the locators be placed ?

- ## On the endsystem
  - e.g. Shim6, HIP, ...

- ## On border routers
  - e.g. GSE

- ## On routers
  - e.g. LISP

- ## On proxies
  - e.g. Proxied-shim6, proxied-HIP, ...

New architecture should
not dictate the exact location of locators.

How many locators should be associated to an identifier ?

- **Exactly one**
  - Not sufficient

- **Exactly N**
  - How to select the appropriate value for N ?

- **One or more**
  - The architecture should be flexible enough to allow
    - Each identifier to be associated to different numbers of locators.
    - the number of locators associated to an identifier to change with time

Why using multiple locators ?

- Reducing the FIB size or the BGP churn rate is not the selling point.
  - Selling point is offering new or better services

- For redundancy
  - If one locator becomes suddenly unreachable, the others might still continue to be reachable
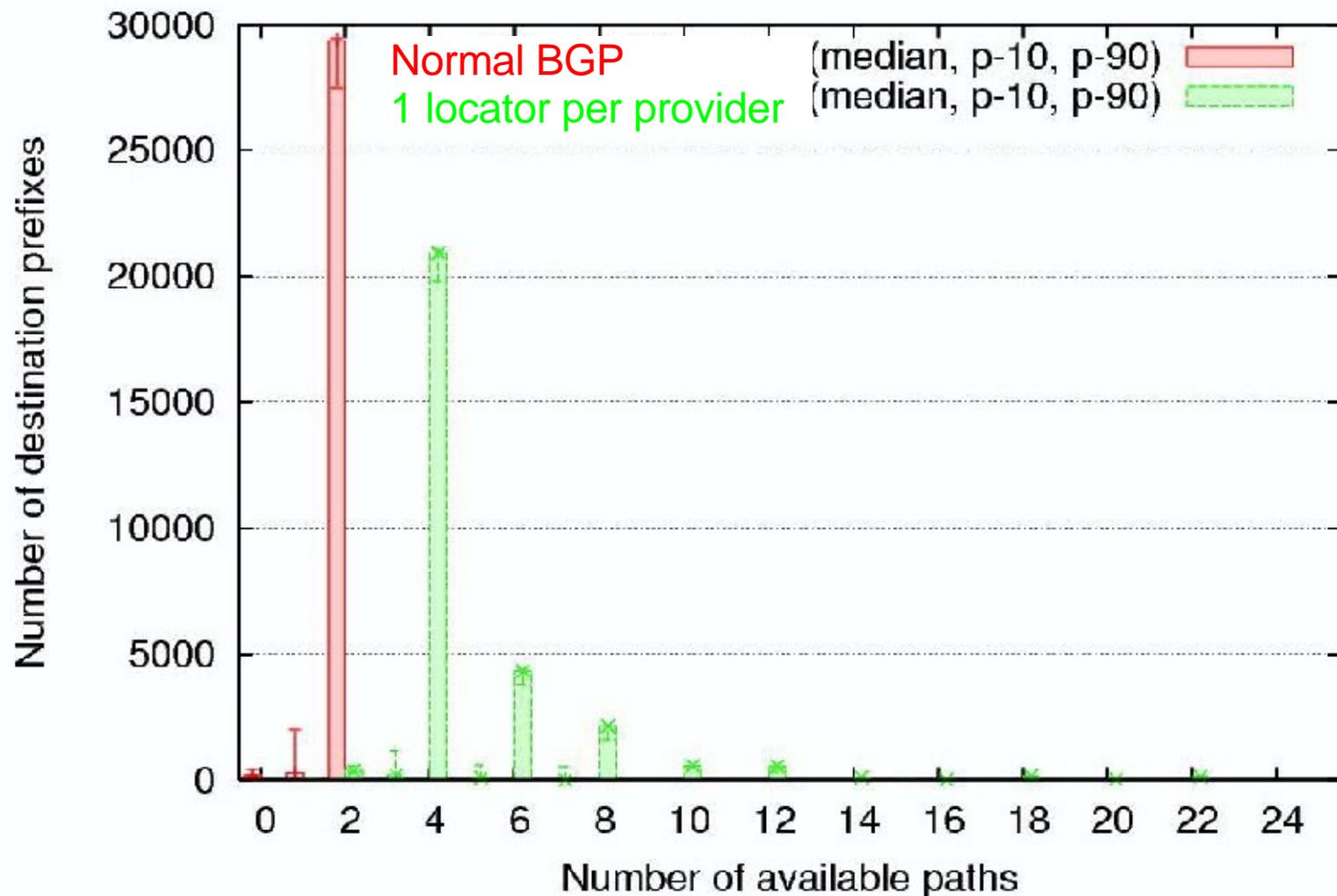
Why using multiple locators ? (2)

- To increase the number of available paths
  - Simulations based on BGP routing tables from RIPE RC00
    - Collect BGP table dump and identify
      - Peers announcing full BGP routing table (about 30 transit)
      - multi-homed stub Ases (about 6000)

    - For each pair of simulated transit
      - Simulate a new dual-homed stub AS connected to the two considered transit providers
      - Compute BGP routing table to determine AS-Paths available to reach each multihomed stub AS when using
        - Normal IPv4 multihoming
        - map/encaps with one locator assigned to each multihomed stub by each of its provider
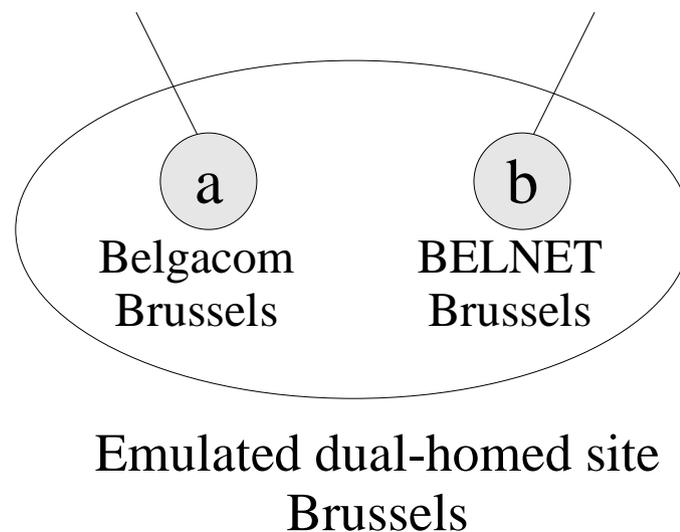
Why using multiple locators (3)

- ## Simulation results

© O. Bonaventure, 2007

Why using multiple locators ? (4)

- To improve end-to-end performance
  - Simulations using the RIPE TTM data set
    - accurate one-way delay measurements between 120 RIPE testboxes

  - 13 emulated multihomed sites assuming each receives one locator per provider
    - 10 dual-homed
    - one 3-homed
    - one 4-homed
    - one 8-homed

a
Belgacom
Brussels

b
BELNET
Brussels

Emulated dual-homed site
Brussels

- ## Simulation results
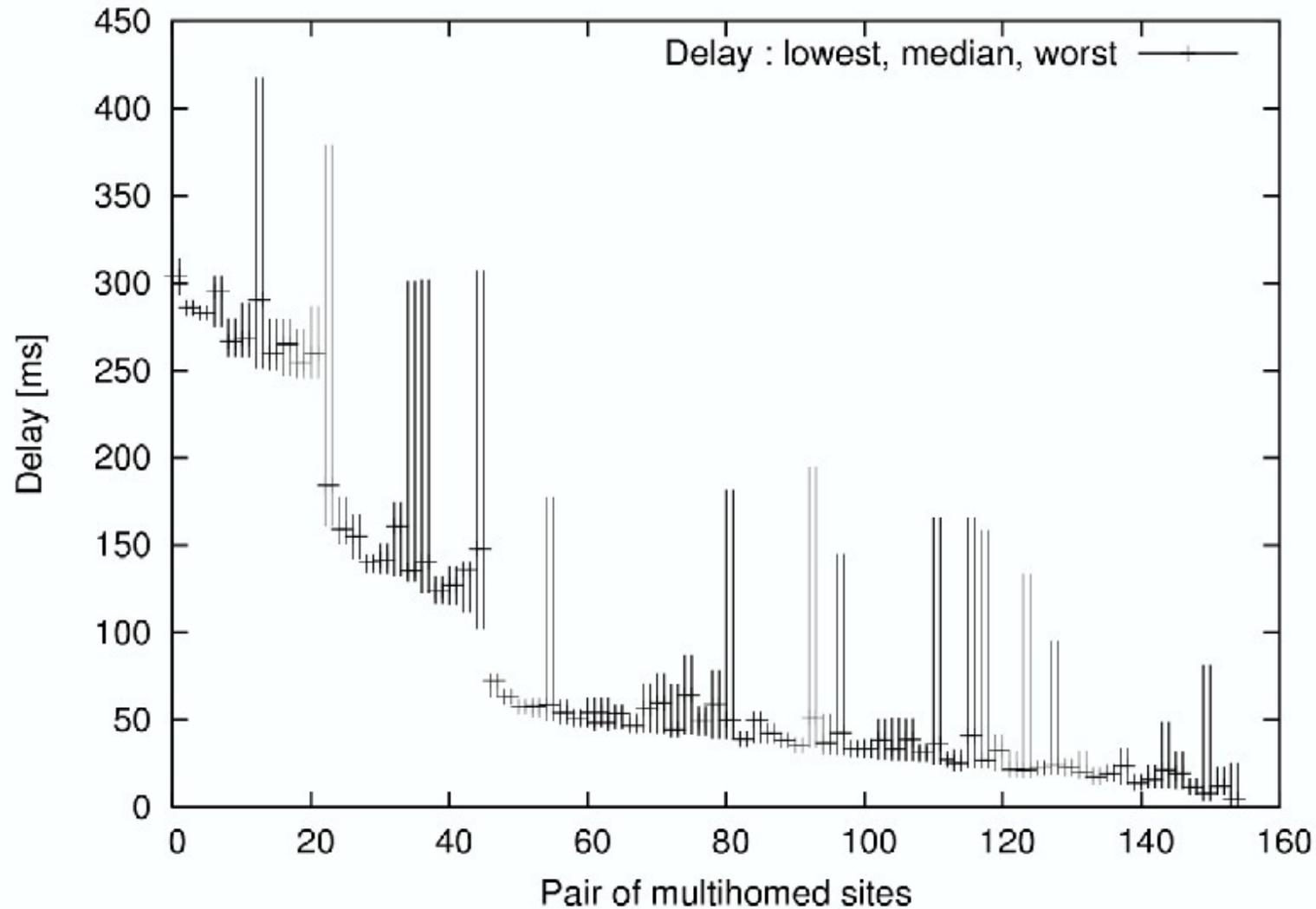
Assignment of locators

- How should locators be assigned to participating systems ?

    - Manually
        - Nice to start doing experiments
        - But experience with IPv4 shows that manual address assignments are cumbersome to manage and make renumbering almost impossible

    - Automatically
        - More complex but will payoff in the long term
            - If IP address numbering was automated, renumbering would probably have been much more easier
        - Automatic mechanism is required to allow a system to dynamically obtain its current set of locators

Locator lifetime

- **What should be the lifetime of a locator ?**

  - **Indefinite as IPv4 address allocations today**
    - Not necessarily the best approach
    - Systems using locators should be prepared to
      - Loose temporarily one assigned locator
      - Loose permanently one assigned locator

  - **Limited in time**
    - A lifetime should be associated to each locator assignment
    - The locator assignment mechanism should allow to
      - Renew for some time a locator assignment
      - Deprecate temporarily a previously assigned locator
      - Deprecate permanently a previously assigned locator
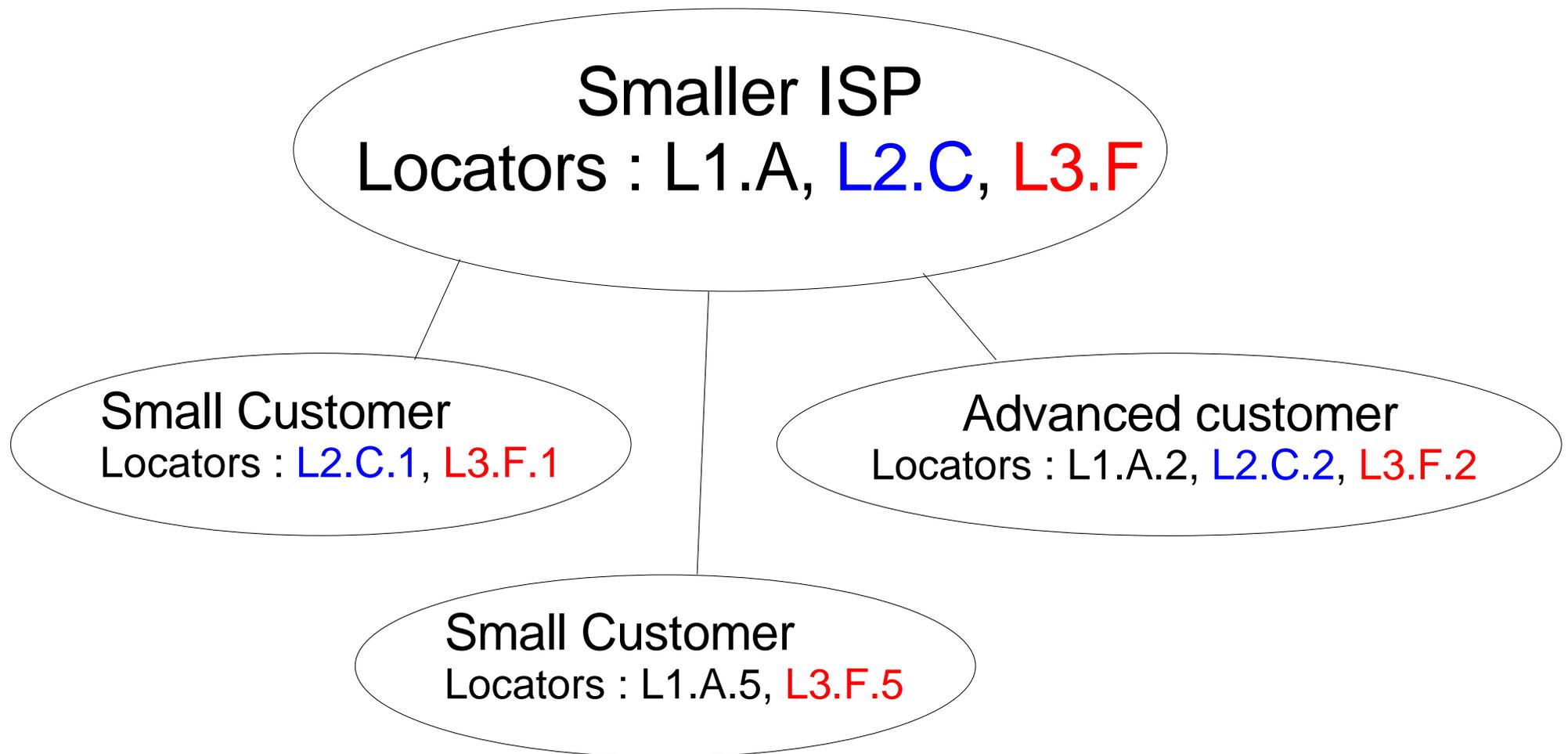
How should a provider assign locators to its customers ?

- **Large provider with a single locator block**
  - **Principle**
    - ◆ Assign subset of locator block to each customer
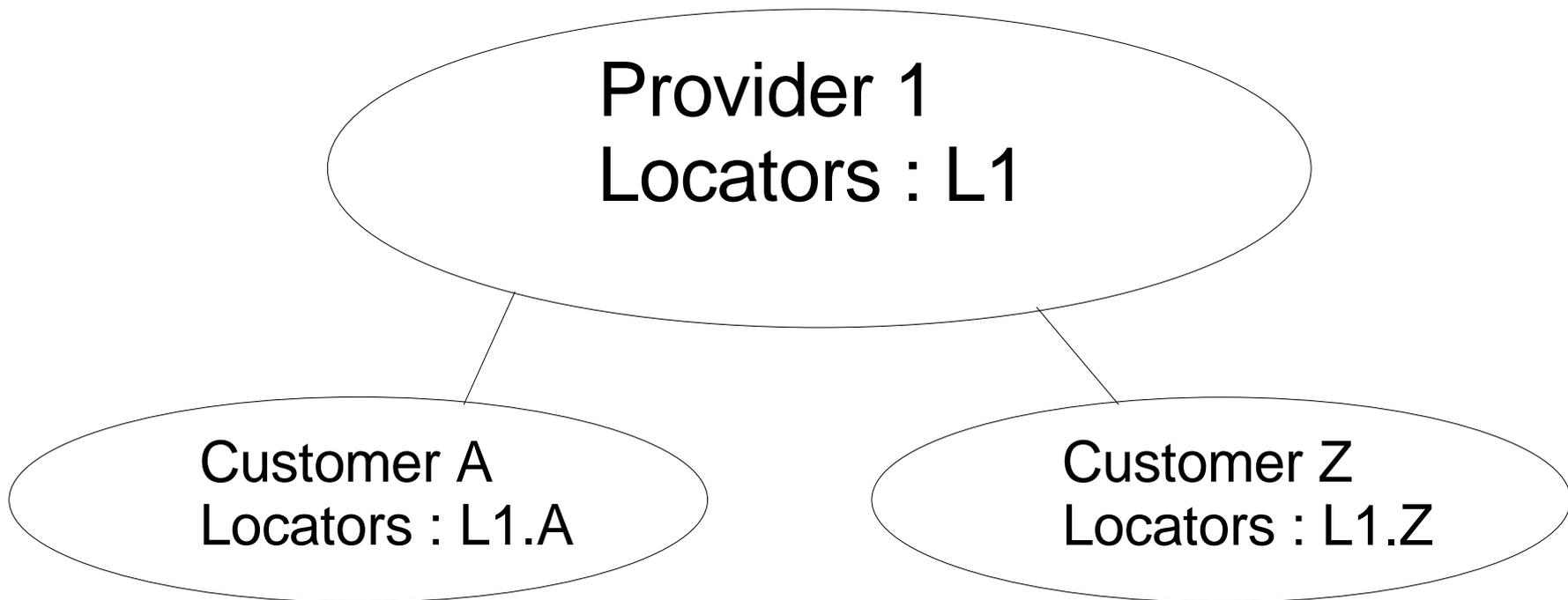
How should a provider assign locators to its customers ? (2)

- Smaller ISP customer of 3 larger ISPs
  - Freedom in locators assignment

Smaller ISP
Locators : L1.A, L2.C, L3.F

Small Customer
Locators : L2.C.1, L3.F.1

Advanced customer
Locators : L1.A.2, L2.C.2, L3.F.2

Small Customer
Locators : L1.A.5, L3.F.5

Outline

- id/locator split and map/encaps

- More on the locators

- How to deal with link failures

- How to support traffic engineering
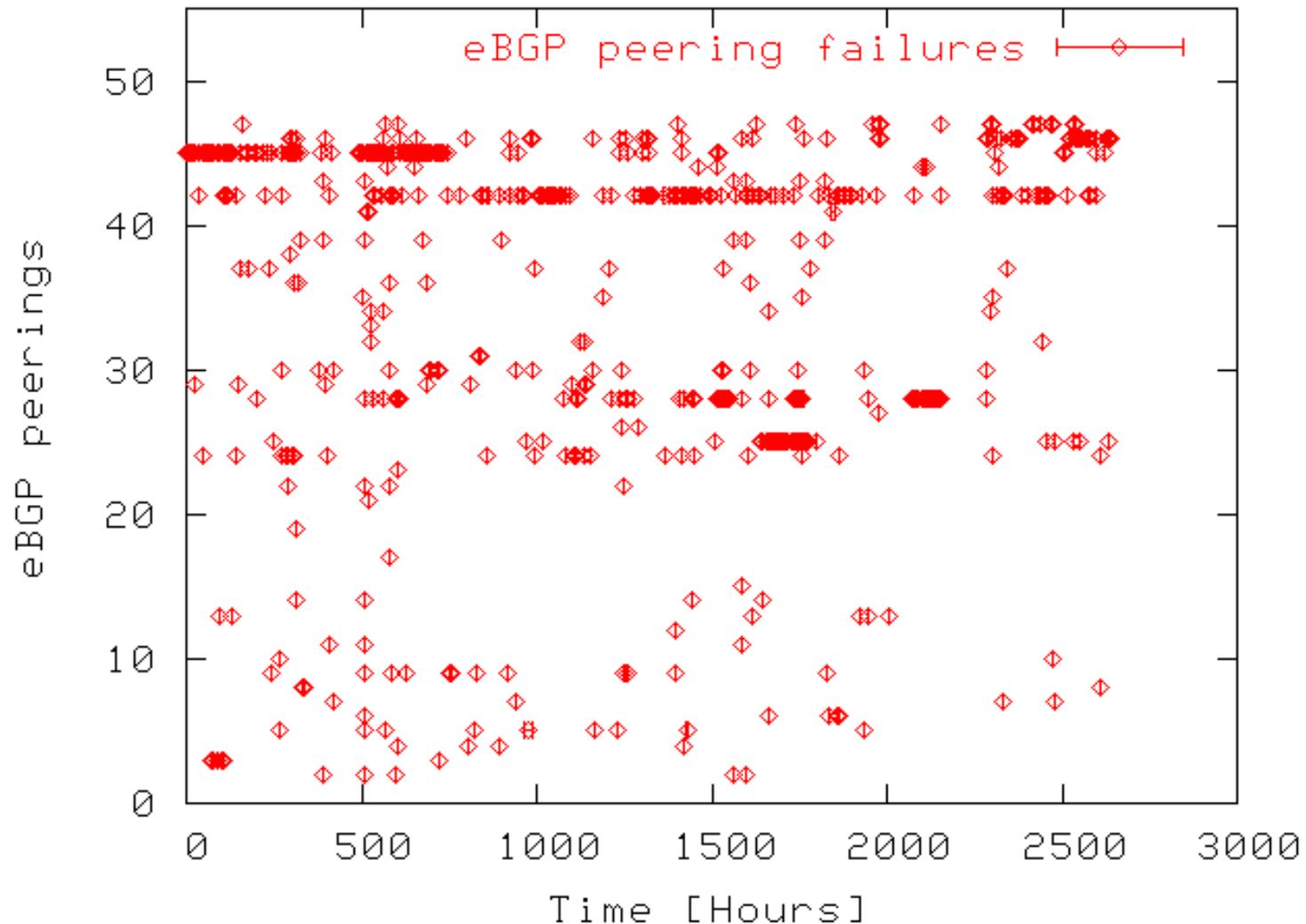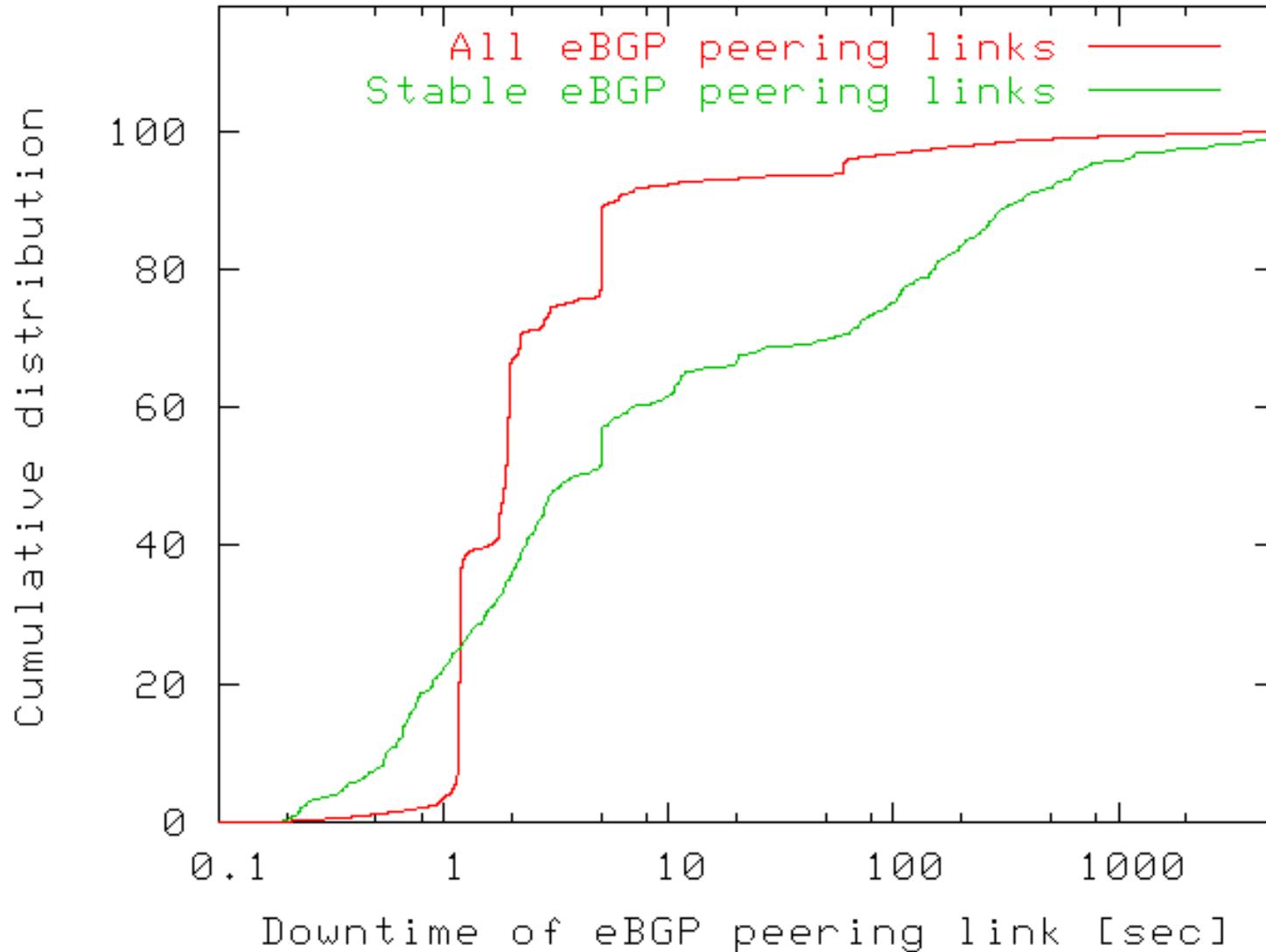
The peering link failure dilemma

- What should happen when a peering link fails ?

  - Customers expect fast recovery to protect their important packets

  - Most interdomain routers do not want to be informed about every remote link failure

Frequency and duration of peering link failures

- One example from a transit AS

Duration of the peering link failures

How to react locally to failures ?

- **Three steps solution**

  1. Provider router activates a tunnel to alternate router to reroute packets affected by failure

  2. Inform mapping mechanism of the failure to possibly stop advertising mappings with the affected locators

  3. If the failure lasts long, deprecate temporarily the affected locators until peering link is up again
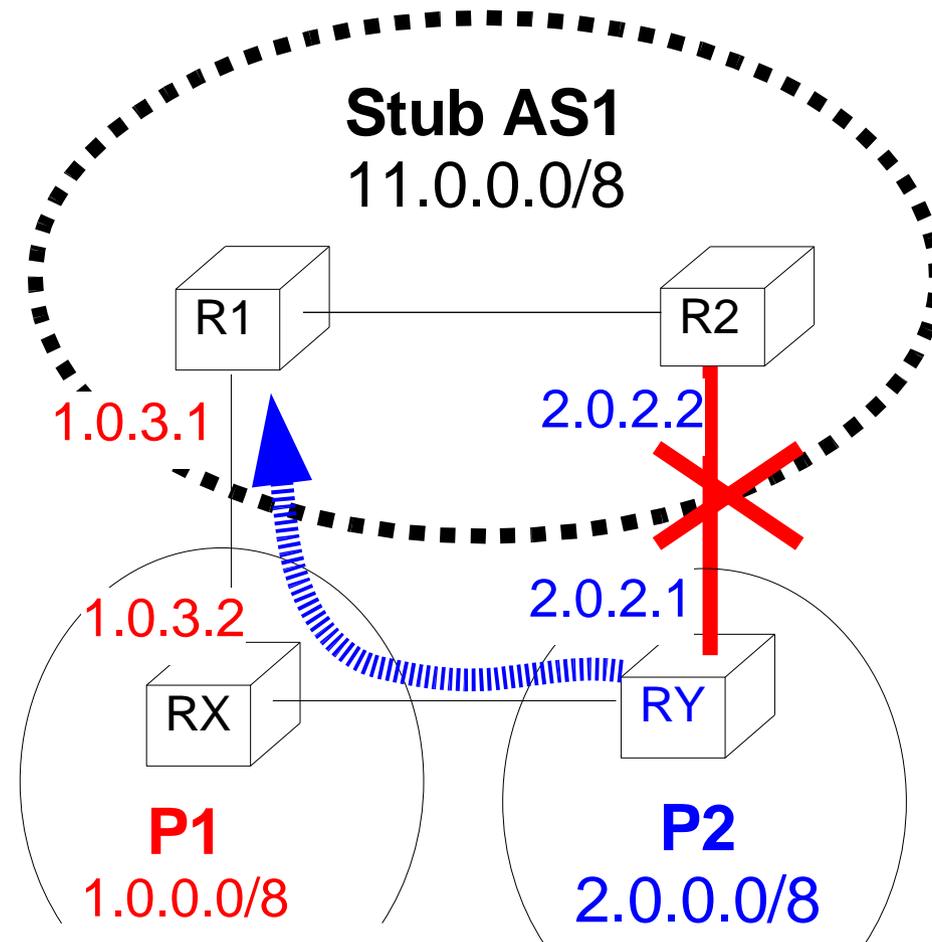
Rerouting packets through tunnels

- ## First case : dual connected AS

Upstream AS

Packets towards R2

R1 — R2

12.0.0.0/8

Downstream AS
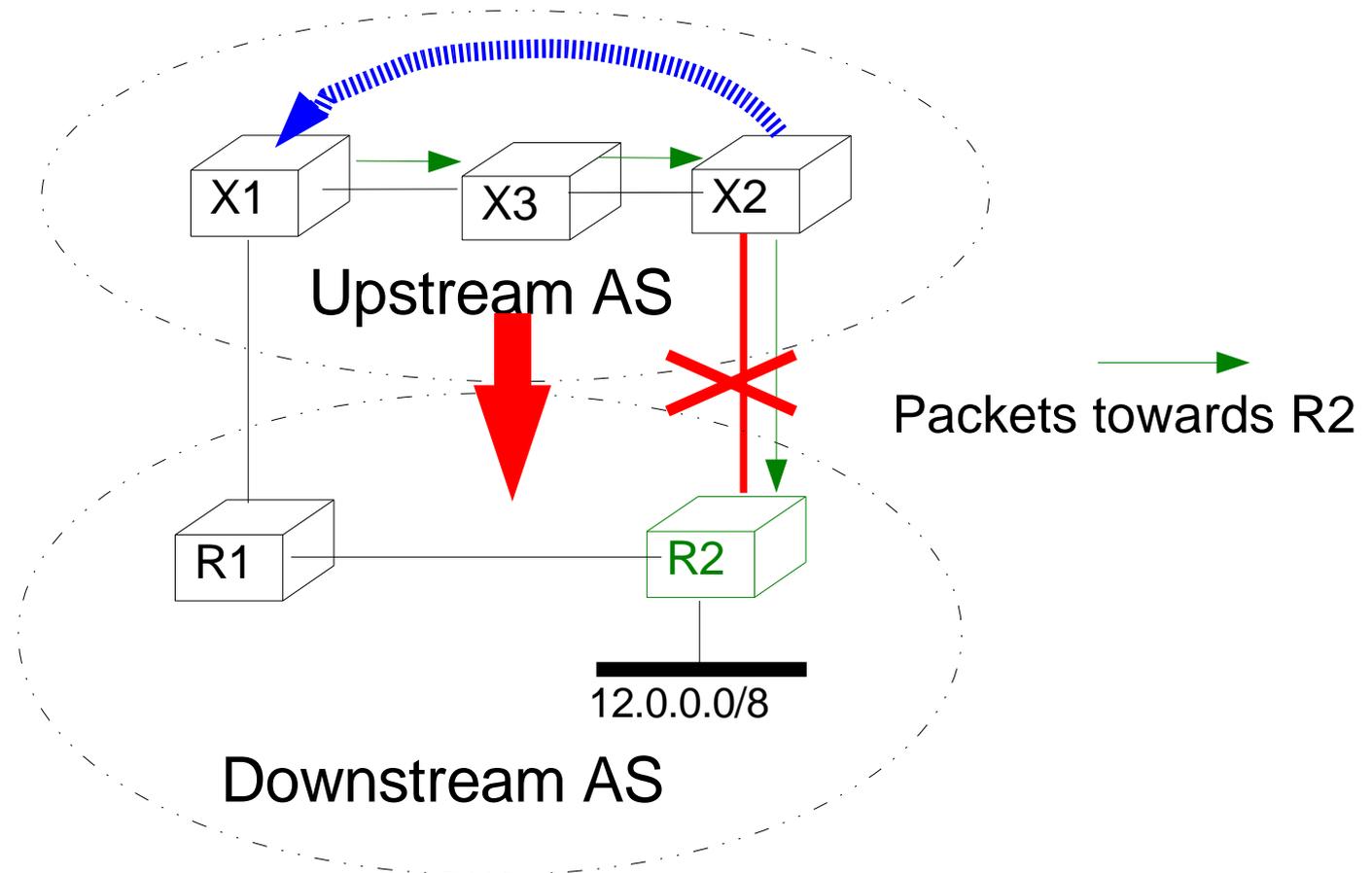
Everything can be handled by the provider as
an added value service for its customers

Rerouting packets through tunnels (2)

- Second case : stub connected to 2 providers

Stub AS1
11.0.0.0/8

R1 — R2

1.0.3.1   2.0.2.2

1.0.3.2   2.0.2.1

RX   RY

P1
1.0.0.0/8

P2
2.0.0.0/8

Can be deployed without any cooperation between providers

Outline

- id/locator split and map/encaps

- More on the locators

- How to deal with link failures

- How to support traffic engineering

How to do traffic engineering ?

- What are the traffic engineering objectives ?

  - Use better paths (delay, bandwidth,...) to reach specific prefixes

  - Load-balance traffic (statically or dynamically) among provider links

  - ...

How to do traffic engineering ? (2)

- Two possible mechanisms that do not require any change in route advertisements
  - Support incoming and outgoing traffic engineering

  - Provider changes the locators assigned to its customers
    - ◆ e.g. Load balance traffic over different provider links based on capacity planning data or projections

  - Mapping mechanism changes the identifiers-locators associations
    - ◆ Should be able to react faster than locator deprecation

Conclusion

- Selling point
  - Added value services, not FIB size or BGP churn

- Requirements for locators
  - New architecture should not dictate the exact location of locators
  - It should be possible to automatically and dynamically assign locators to participating systems
  - Locators should have limited lifetime but be renewable
  - Link failures and traffic engineering can be solved

Acknowledgements

- Based on joint work with Bruno Quoitin, Pierre François, Cédric de Launois, Clarence Filsfils
  - supported by IST AGAVE project

- Additional information may be found in
  - C. de Launois, S. Uhlig, and O. Bonaventure. *Scalable route selection for IPv6 multihomed sites*. Networking 2005
  - O. Bonaventure, C. de Launois, B. Quoitin, M. Yannuzzi, Improving the quality of interdomain paths, unpublished, 2005
  - C. de Launois, B. Quoitin, and O. Bonaventure. *Leveraging network performances with IPv6 multihoming and multiple provider-dependent aggregatable prefixes*. QoSIP 2005
  - B. Quoitin and O. Bonaventure. *A cooperative approach to interdomain traffic engineering*.NGI 2005, April 18-20th 2005
  - O. Bonaventure, C. Filsfils, P. François, Achieving sub 50-msec recovery upon BGP peering link failures, Conext 2005
  - O. Bonaventure, Reconsidering the Internet Routing Architecture, unpublished manuscript, March 2007
  - Most are available from http://totem.info.ucl.ac.be