A Cooperative Approach to Interdomain Traffic Engineering

Bruno Quoitin
(bqu@info.ucl.ac.be)
Computer Science and Engineering Dept.
Université Catholique de Louvain
Belgium

part of the TOTEM project
http://totem.info.ucl.ac.be

Agenda

1. Introduction
2. Interdomain Path Diversity
3. Interdomain Path Selection
4. Cooperative Interdomain TE
    1. Load-balancing
    2. Delay improvement
5. Conclusion

1. Introduction

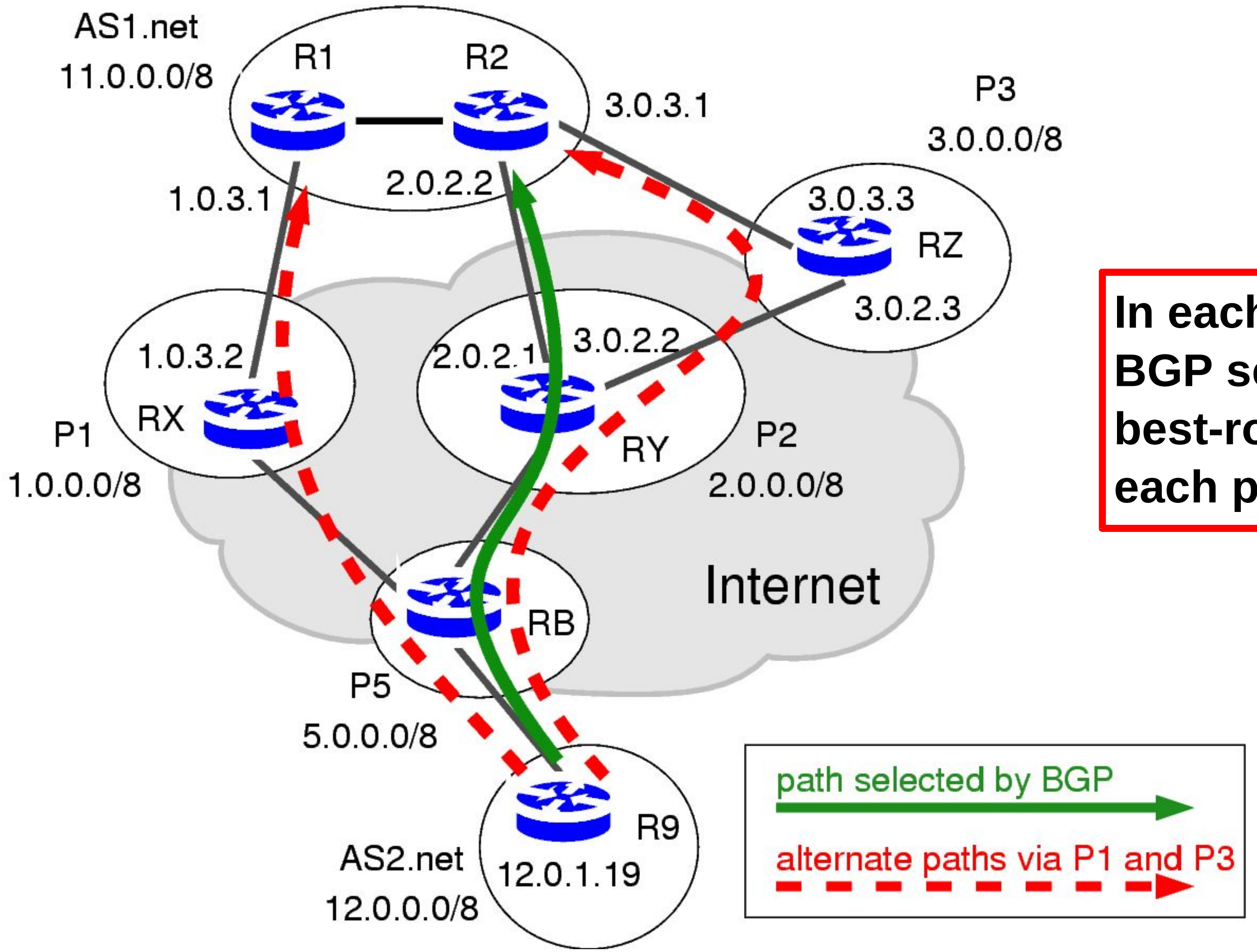
Introduction

- **Internet Structure**
  - Transit ASes vs **Stub ASes** (85%)
  - Trend to multihome for reliability and performance reasons
  - Multihoming brings new needs
    - Handle **traffic imbalance** on multiple access links
    - Chose better paths in terms of **delay/bandwidth** (if available)
    - Better manage the **cost** of Internet access
  - BGP imposes limitations:
    - Low path diversity + difficult control of route selection

2. Interdomain Path Diversity

Interdomain Path diversity (1)

- **BGP reduces path diversity**



In each router, BGP selects a single best-route towards each prefix.

Interdomain Path diversity (2)

What if we could select the ingress router of a destination stub ?

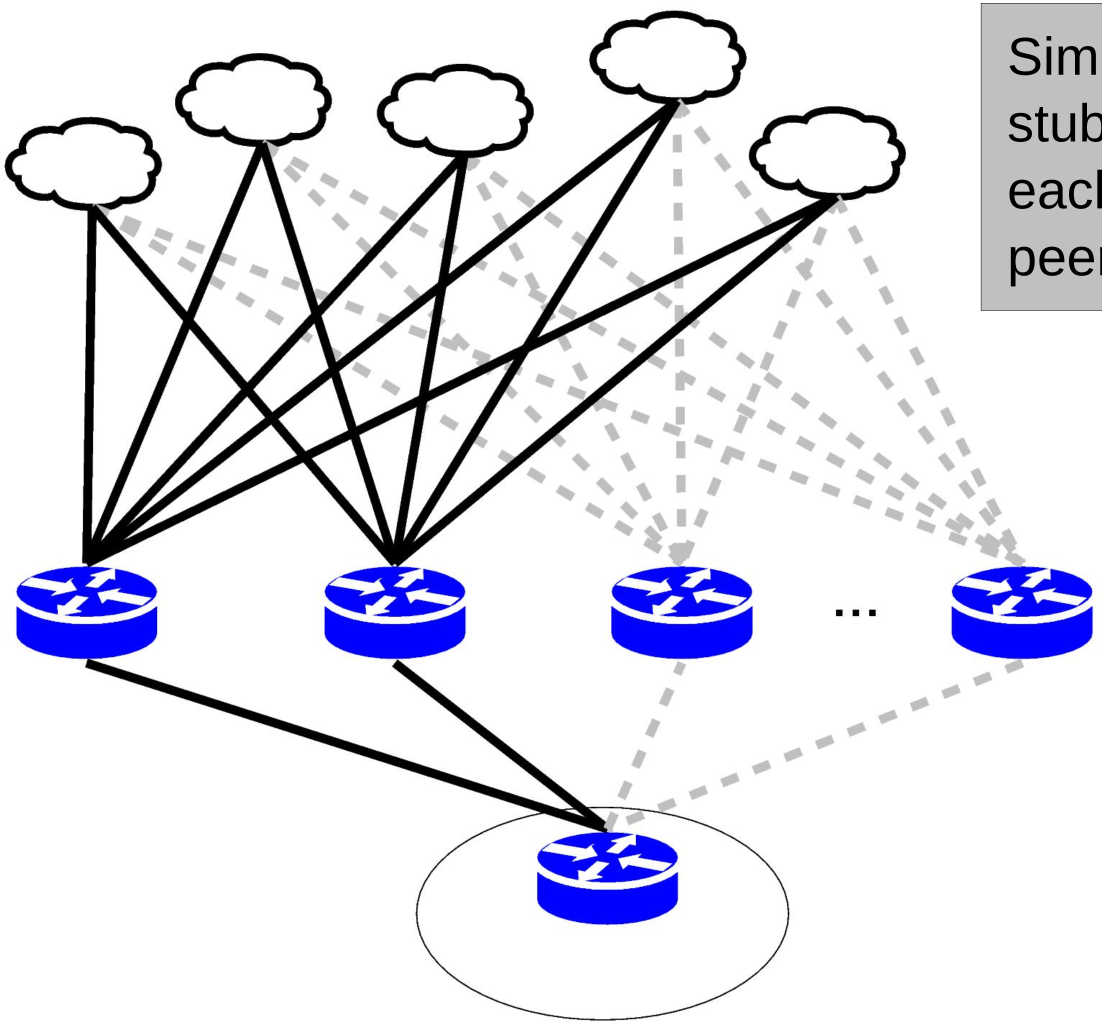
- **Measurement:**
  - Based on RouteViews (RV) archives
  - December 1st, 2004
  - 34 BGP peers
  - 5,750,380 BGP routes collected
  - 29,575 destination prefixes originated by 6,402 multihomed stubs

Interdomain Path diversity (3)

29.575 multi-homed destination prefixes

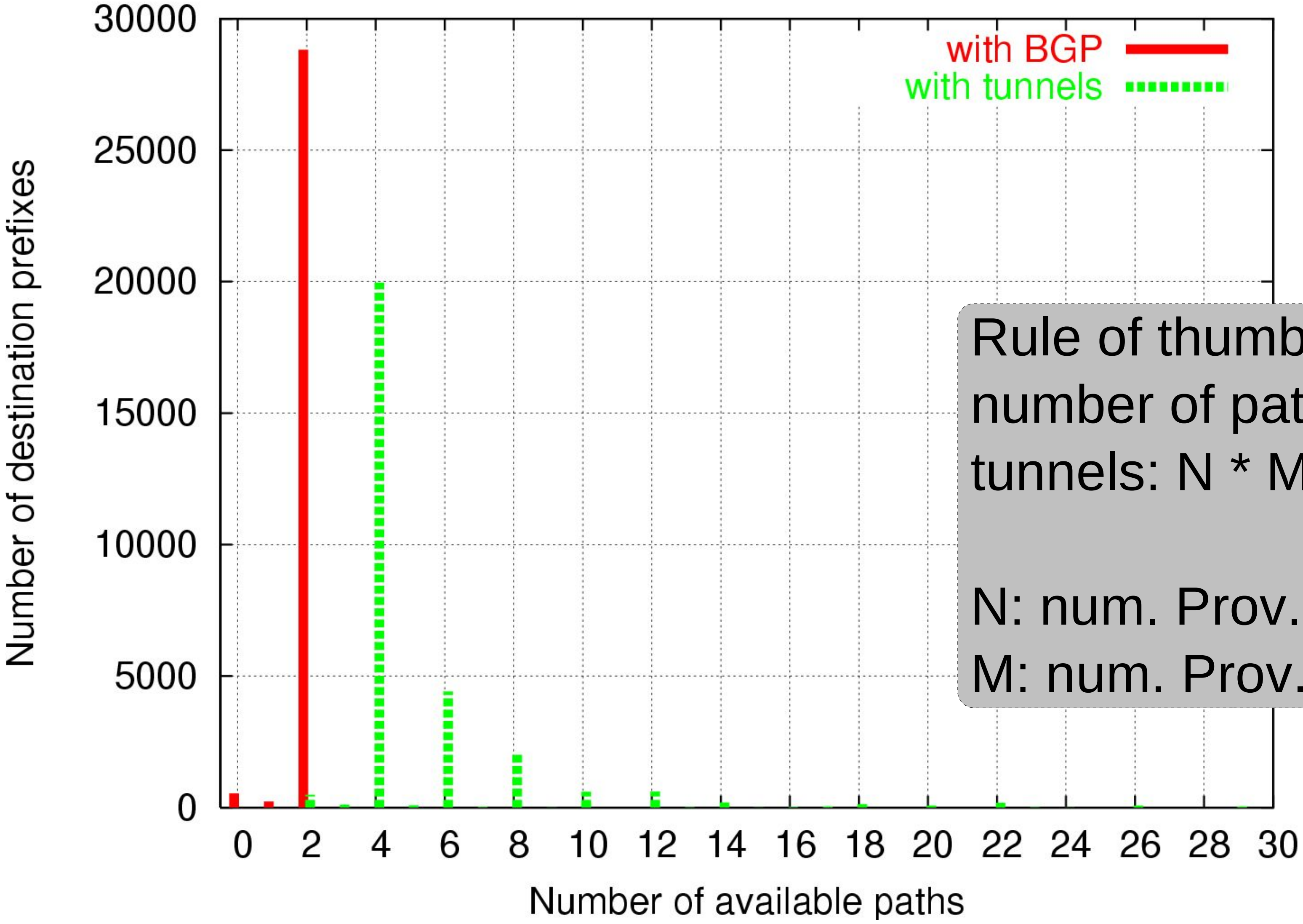
34 RV peers

Simulated dual-homed stub



Simulate dual-homed stubs connected to each pair of RouteViews peers (~ 500 pairs).

Interdomain Path diversity (4)

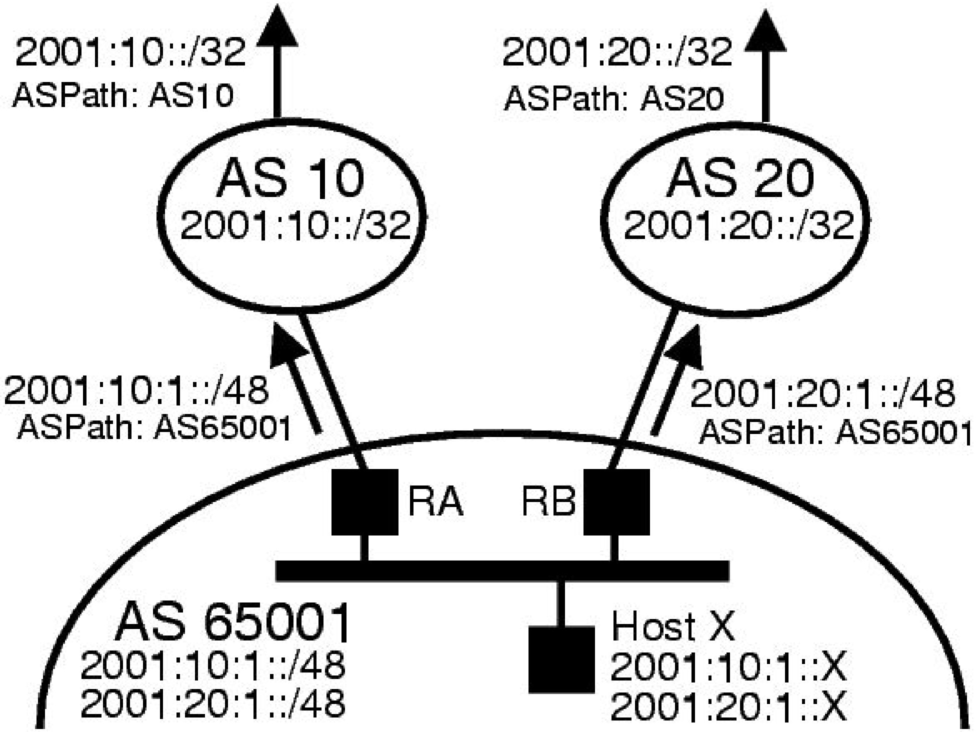


Rule of thumb for the number of paths with tunnels: N * M

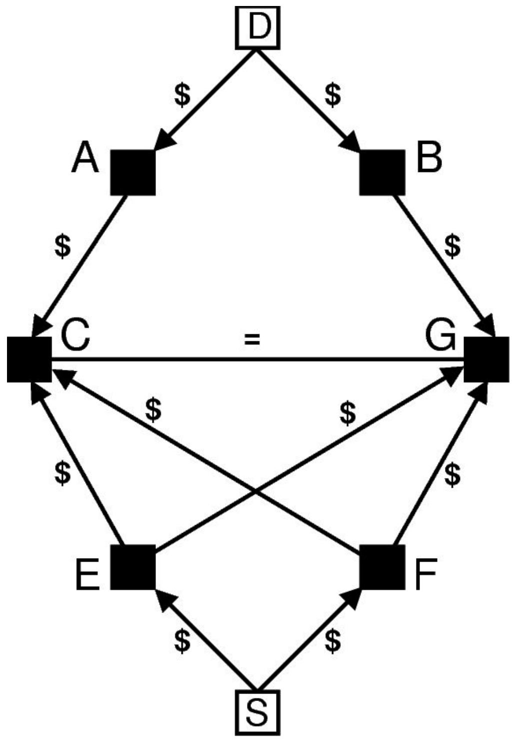
N: num. Prov. Source
M: num. Prov. dest

Interdomain Path diversity (5)

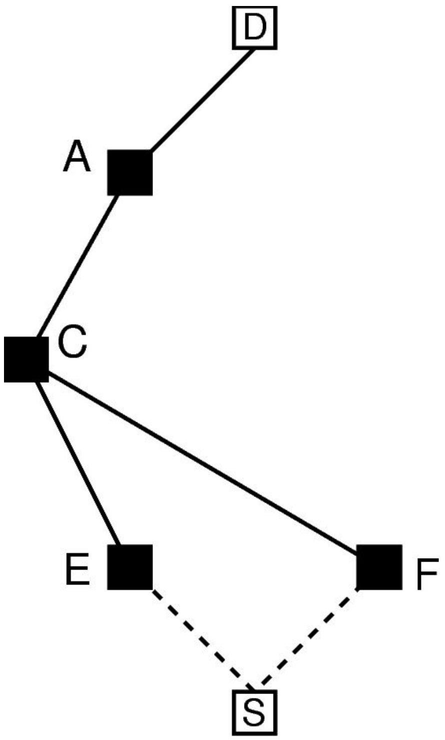
- **Leveraging Internet path diversity with Ipv6**
  - With IPv6, possible to reach destinations through any of the destination's providers



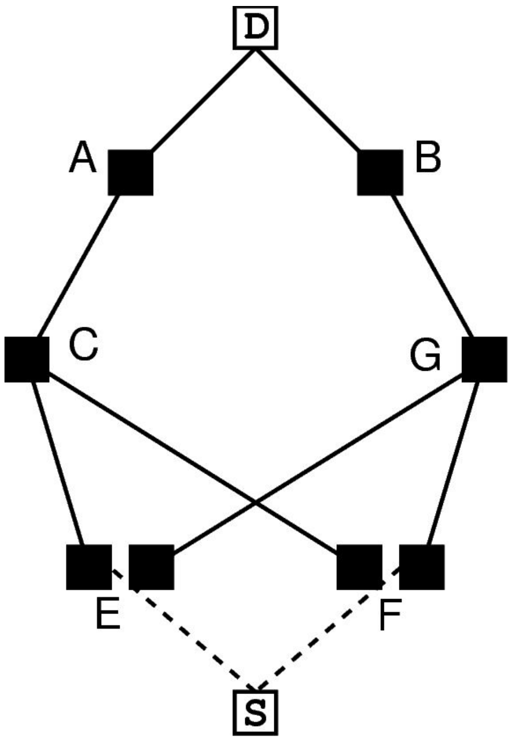
Interdomain Path diversity (6)



Topology



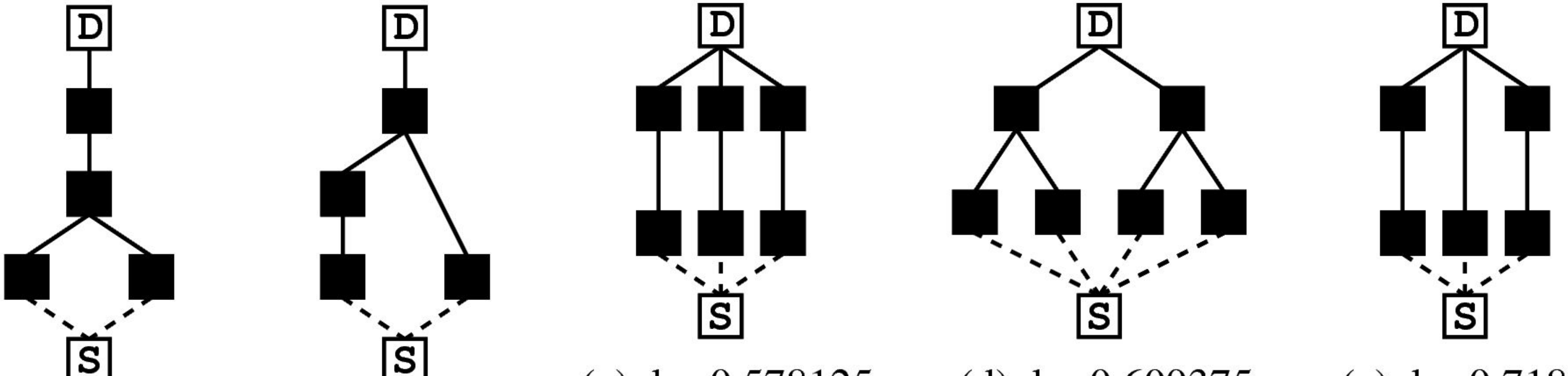
IPv4 tree



IPv6 tree

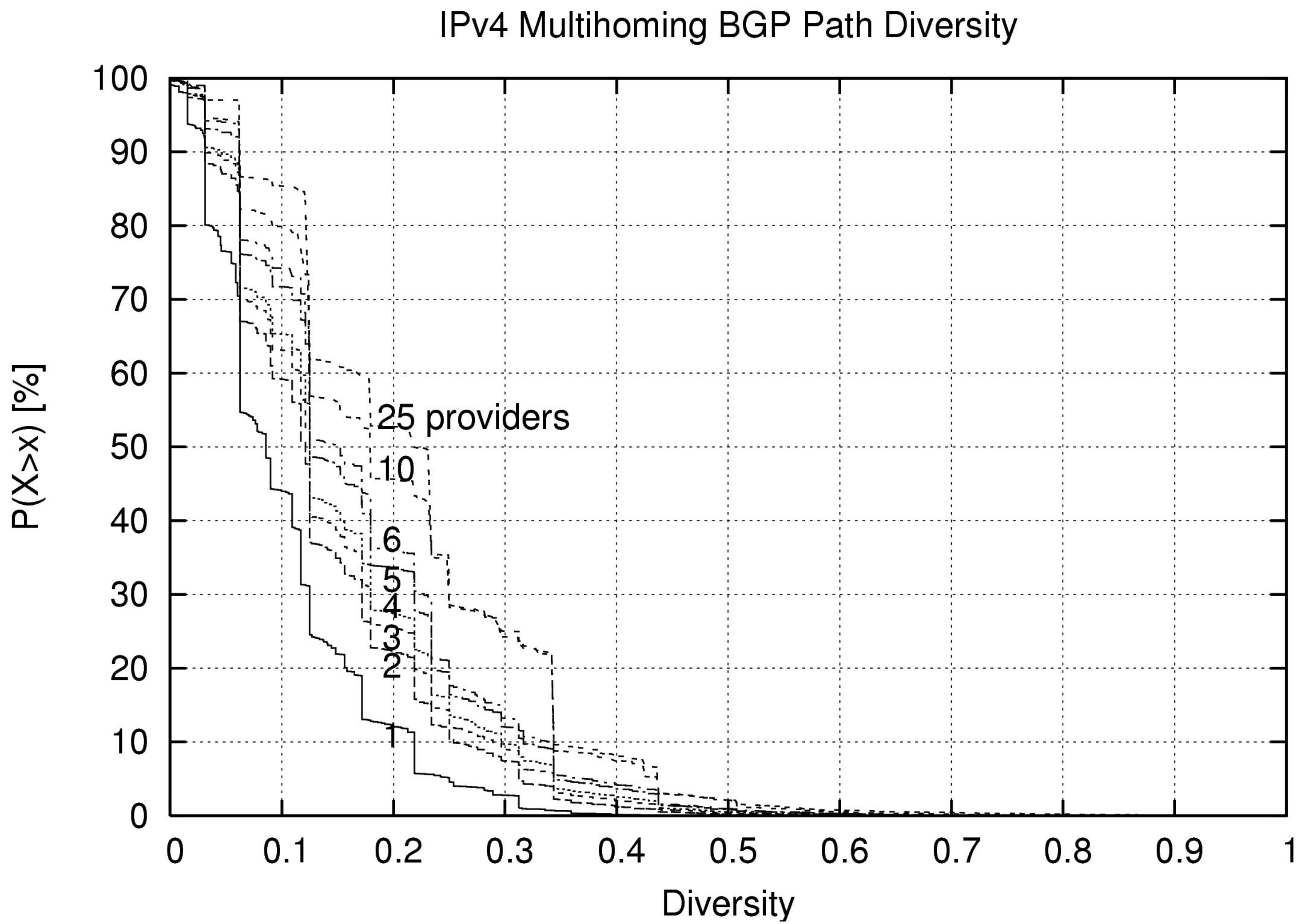
Interdomain Path diversity (7)

- **New path-diversity metric**
  - Probability of failure of a link
  - Probability of failure of a tree
  - Simulations on topology inferred by [Subramanian et al], using C-BGP and policies

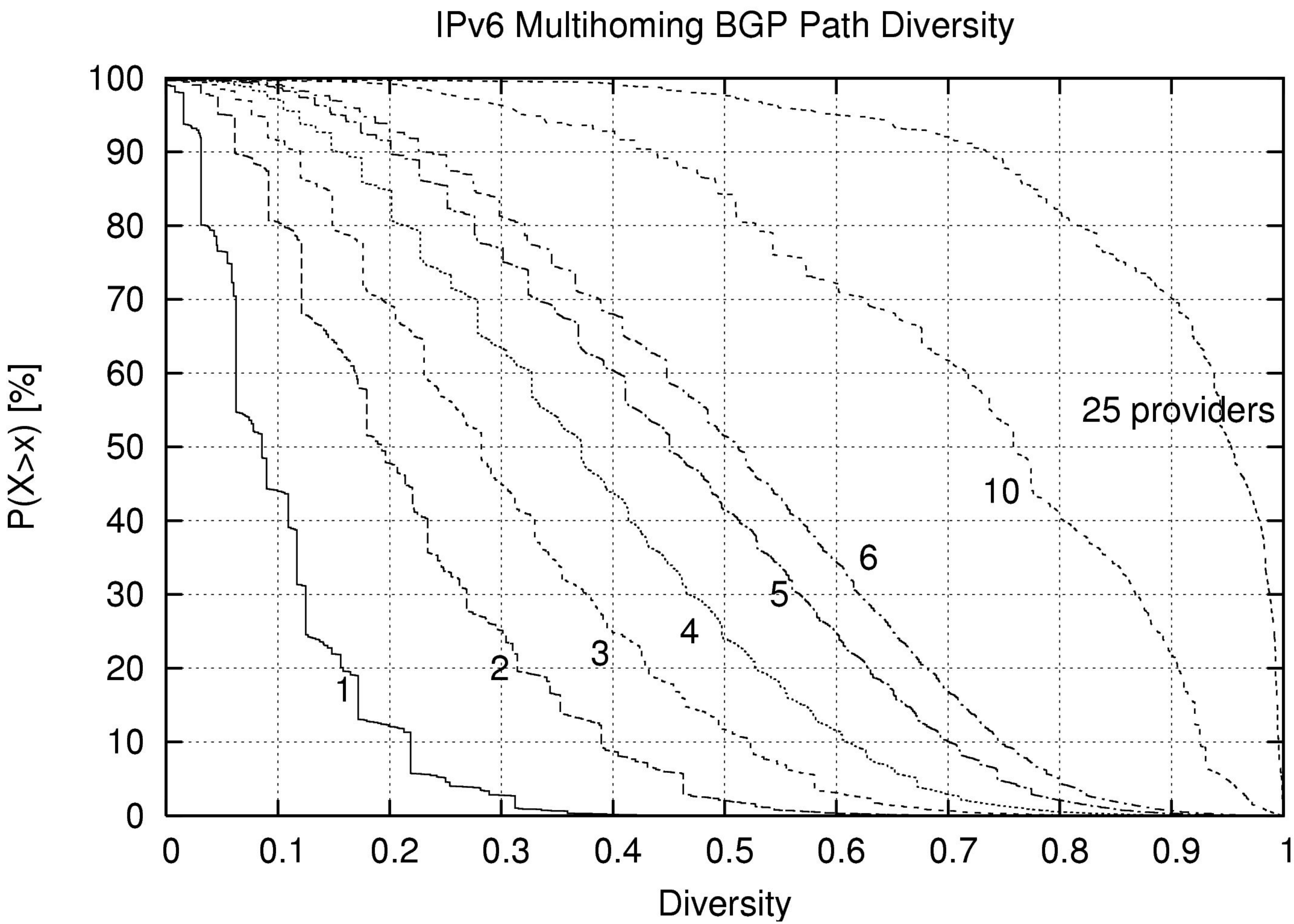


(a) d = 0.1875 (b) d = 0.3125 (c) d = 0.578125 (d) d = 0.609375 (e) d = 0.71875

Interdomain Path diversity (8)



Interdomain Path diversity (9)



3. Interdomain Path Selection

Interdomain Path Selection (1)

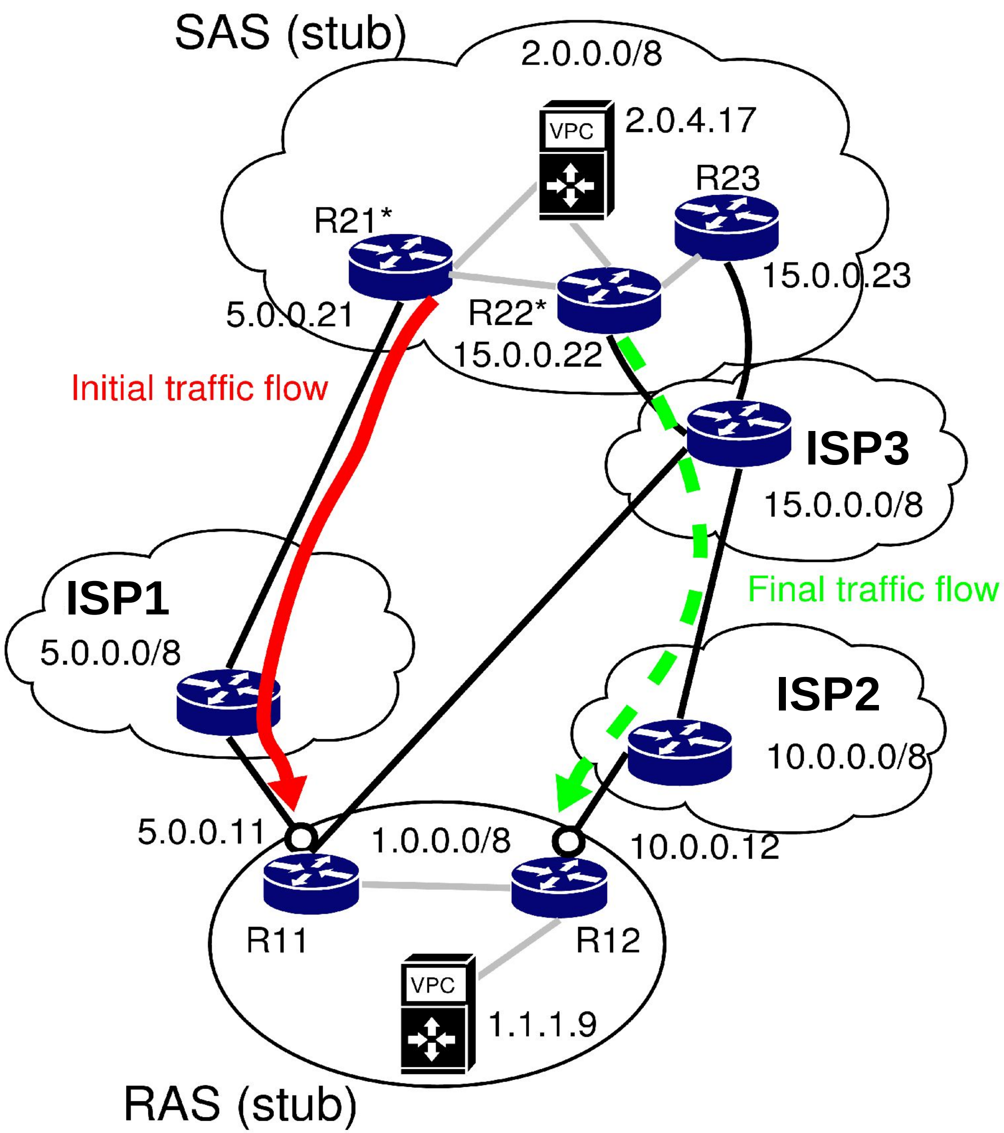
- **Path selection is difficult with BGP**
  - BGP designed to provide reachability and policy routing, not for route control.
  - Control of outbound paths:
    - Possible: local decision.
  - Control of inbound paths:
    - Difficult/impossible: requires controlling the routing decisions taken by other ASes

Interdomain Path Selection (2)

- **Current techniques/proposals**
  - BGP tweakings
    - More specific prefixes, Selective announcements, AS-path prepending, Communities
    - RIB growth, coarse, non-deterministic :-(
  - New interdomain Routing Arch.
    - OPCA, ...
    - Need to change all the 200.000+ BGP routers running the Internet. Unlikely to be deployed :-(

Can we change the current scheme in order to better control the inbound paths ?

Interdomain Path Selection (3)



Legend

SAS = Source AS

RAS = Requester AS

VPC = Virtual Peering Controller

Interdomain Path Selection (4)

- **Virtual Peerings**
  - Tunnels: GRE, IPSec, MPLS, ...
  - Encapsulation/decapsulation at line rate
  - Control the access link used to reach the destination in a **deterministic** way
  - BGP or DNS-based signalling
    - Few/no changes required to routers
    - Makes possible an **incremental deployment**

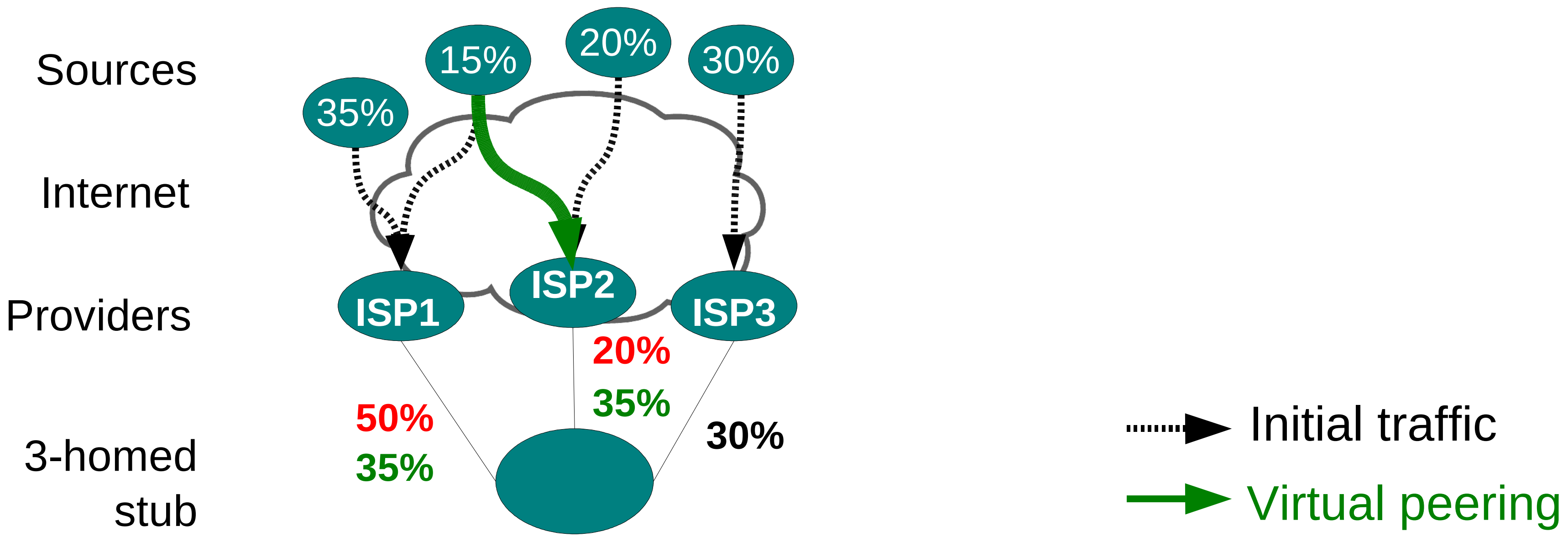
4. Cooperative Interdomain Traffic Engineering

Cooperative Interdomain TE (1)

- **Cooperation**
  - Exchange information between ASes in order to select the paths to be used
  - Objectives:
    1. **Load-balancing**: decrease probability of congestion on ingress links
    2. **Improving delay/bandwidth**
    3. Decreasing cost of Internet access
  - Currently focus on stubs
    - Represent 85% of Internet ASes

Load-balancing (1)

- **1. Load-balancing**
  - Inbound traffic load unevenly distributed
  - Request some sources to use a Virtual Peering
  - Combinatorial problem:
    - ~17.000 traffic sources, up to 25 ingress links



Load-balancing (2)

- **Evaluation methodology**
  (1) Input:
  - Internet topology with business relationships
  - Traffic demand
  - Stubs to evaluate

  (2) Compute interdomain paths from each AS towards destination stub
  (3) Weight paths with traffic
  (4) Compute load of the stub's ingress links
  (5) Optimize
  (6) Setup Virtual Peerings

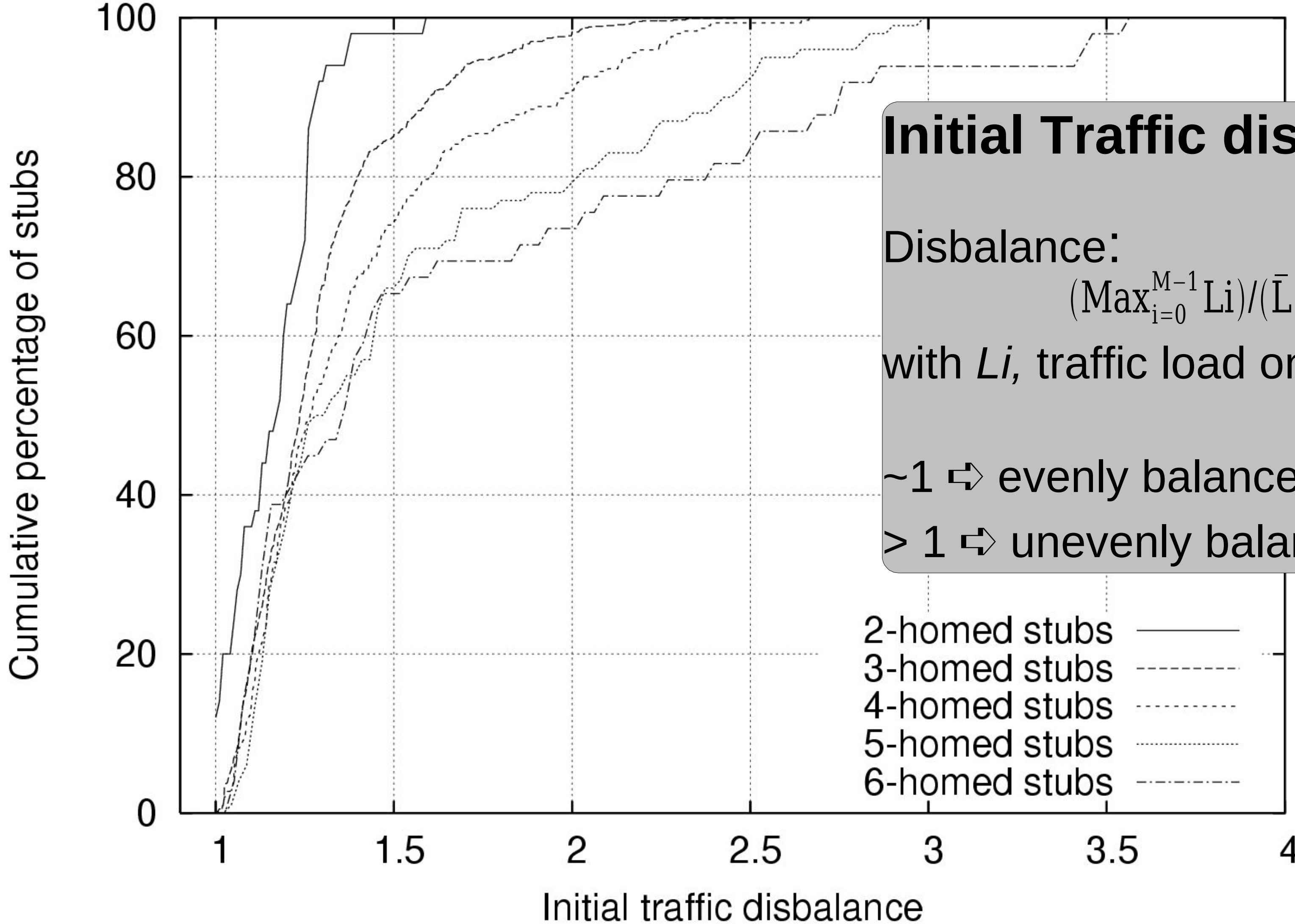
Load-balancing (3)

- **Internet topology**
  - Inferred by [Subramanian et al] on February 10th, 2004: 16,921 ASes and 37,271 interdomain links
  - Relationships: customer/provider/peer
- **Traffic model**
  - heavy-tailed, Weibull (α=0.5)
  - 95% traffic sent by ~1000 sources
  - Similar to interdomain traffic distribution described in the literature (Feamster, Uhlig)
- **BGP model**
  - One router per AS
  - Policies: selective export rules (avoid transit) + preferences (prefer cheaper routes)

Load-balancing (4)

- **C-BGP: routing solver**
  - Computes outcome of BGP route selection
  - Features:
    - IGP model
    - iBGP & eBGP
    - versatile policies
    - complete BGP decision process
    - Route-reflection
  - Large scale simulations
    - more than 30.000 BGP routers
  - Open source, LGPL, publicly available
    http://cbgp.info.ucl.ac.be

Load-balancing (5)



Initial Traffic distrib.

Disbalance:

$$(\mathrm{Max}_{i=0}^{M-1} \mathrm{Li})/(\bar{\mathrm{L}})$$

with *Li,* traffic load on link *i*

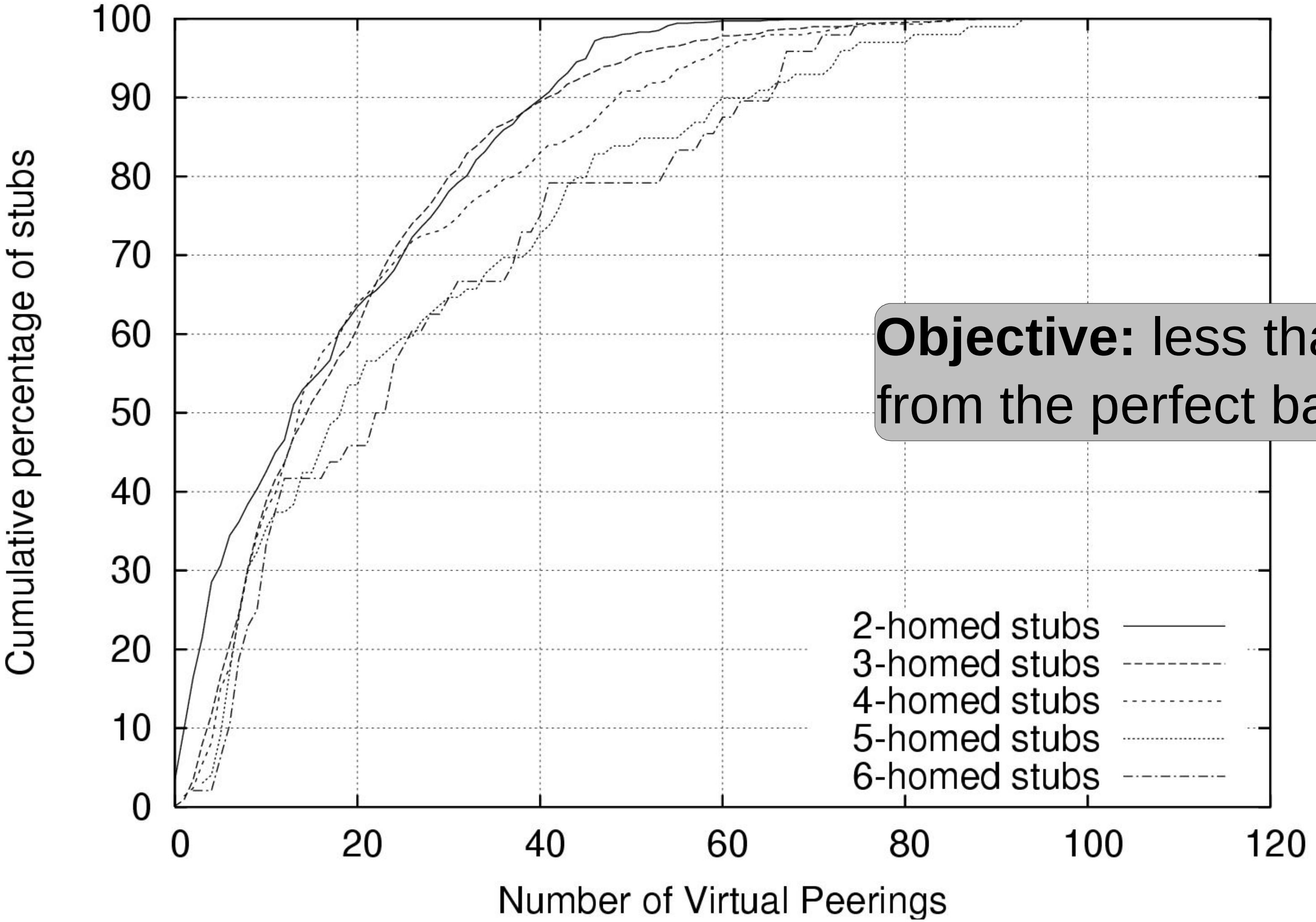
~1 ➪ evenly balanced

> 1 ➪ unevenly balanced

Load-balancing (6)

- **Load-balancing**
  - Method: solve allocation problem using simple Evolutionnary Algorithm
    - Population: an individual represents the allocation of sources on ingress links
    - Mutation: changes the inbound path used by one source so that the traffic from this source enters through another ingress link
    - Objective: minimize disbalance $\Sigma_{i=0}^{M-1}(L_i - 1/M)^2$ where Li is the traffic load on link i, $0 \leq Li \leq 1$

Load-balancing (7)

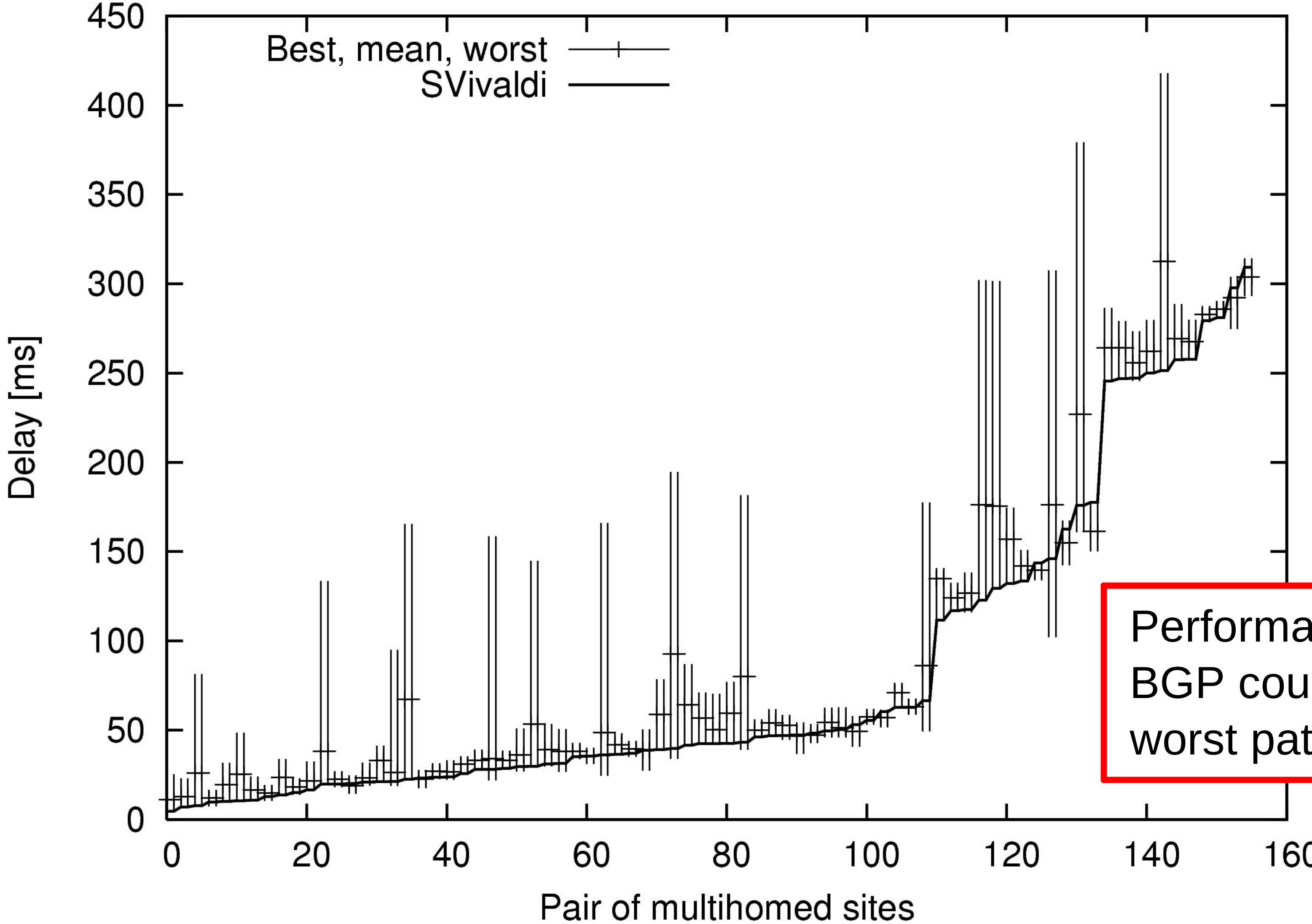


Objective: less than 1% from the perfect balance

Delay Improvement (1)

- **2. Delay improvement**
  - Let's select the path with lowest delay!..
  - How to select the “right path” without probing every path ?
    - Using stable synthetic coordinates computed in a distributed manner (SVivaldi)
    - Needs to probe only a few neighbors
    - Publish coordinates in DNS
  - Evaluation
    - Simulate 13 multi-homed sites
    - based on RIPE NCC RTT dataset

Delay Improvement (2)



Performance-blind BGP could select worst paths

5. Conclusion

Conclusion

- **Virtual Peerings**
  - Alternative to BGP tweakings / Overlays
  - **Deterministic** and **scalable** control of inbound paths
- **Cooperative Interdomain TE**
  - Load-balancing of inbound traffic is possible **at a reasonnable cost** (typically, less than 40 tunnels for 80% of stubs)
  - Delay improvement without probing of every possible path (using SVivaldi and the DNS)

References

- **Leveraging network performances with IPv6 multihoming and multiple provider-dependent aggregatable prefixes**
  C. de Launois, B. Quoitin, and O. Bonaventure. In QoSIP 2005, February 2005.
- **A cooperative approach to interdomain traffic engineering**
  B. Quoitin and O. Bonaventure. In NGI 2005, April 2005.
- **C-BGP, an efficient BGP simulator**
  B. Quoitin. http://cbgp.info.ucl.ac.be
- **Improving the quality of interdomain paths by using IP tunnels and the DNS**
  O. Bonaventure, C. de Launois, B. Quoitin and M. Yanuzzi.
  Technical report, 2005.
- **Characterizing the Internet hierarchy from multiple vantage points**
  L.Subramanian, S. Agarwal, J. Rexford and R. H. Katz. Infocom. June 2002.
- **Scalable Route Selection for Ipv6 Multihomed Sites**
  C. de Launois, S. Uhlig and O. Bonaventure. Networking 2005.