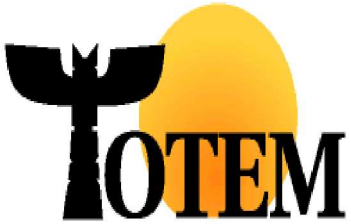
UCL

Interdomain routing with BGP4

Part 1/5



Olivier Bonaventure

Department of Computing Science and Engineering
Université catholique de Louvain (UCL)
Place Sainte-Barbe, 2, B-1348, Louvain-la-Neuve (Belgium)

Email : Bonaventure@info.ucl.ac.be
URL : *http://www.info.ucl.ac.be/people/OBO*

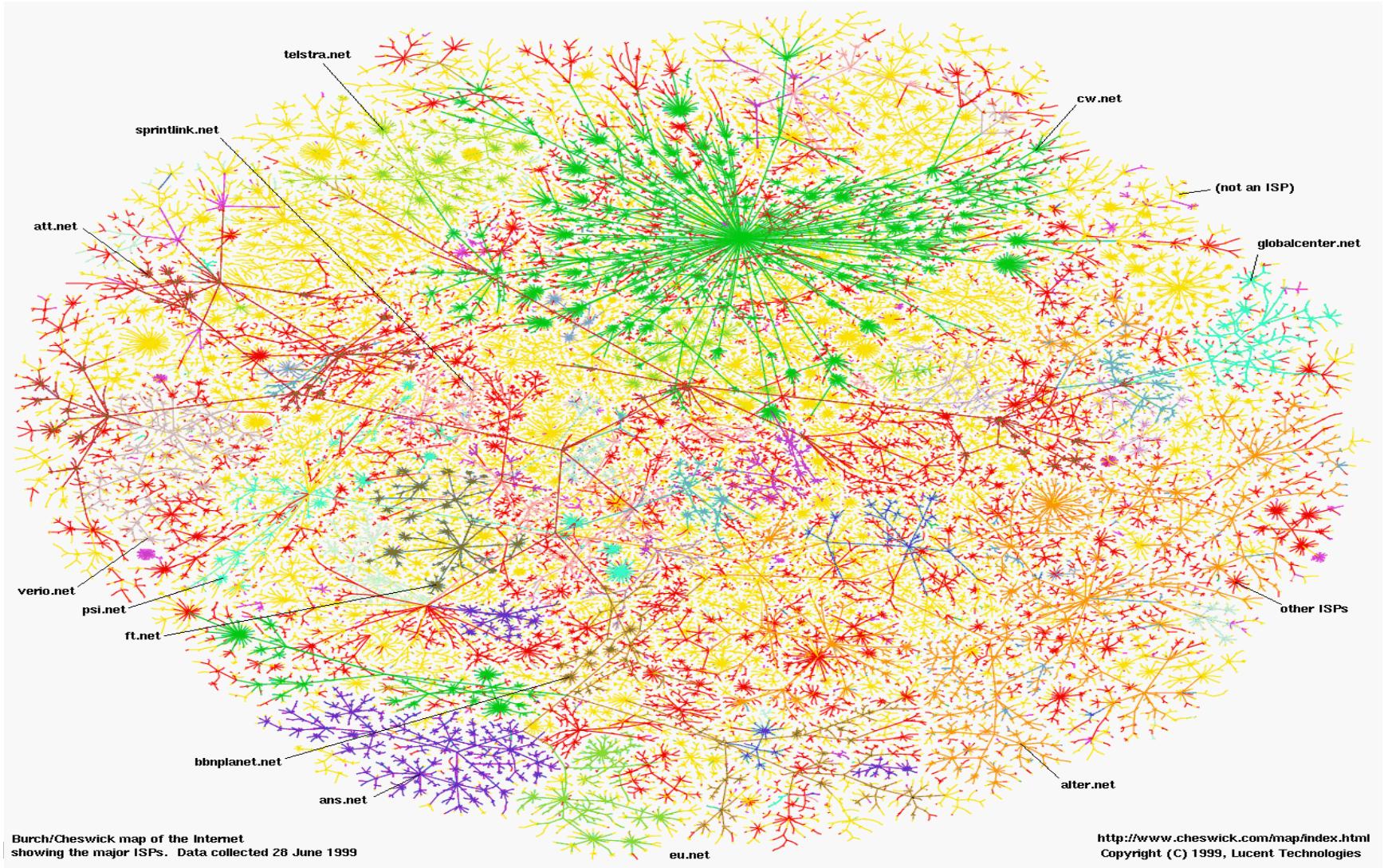
INGI
Département d'ingénierie informatique

Outline

- Organization of the global Internet
  - Example of domains
  - Intradomain routing
- BGP basics
- BGP in large networks
- Interdomain traffic engineering with BGP
- BGP-based Virtual Private Networks

How to route IP packets in the global Internet ?

A map of the global Internet in 2000 ...

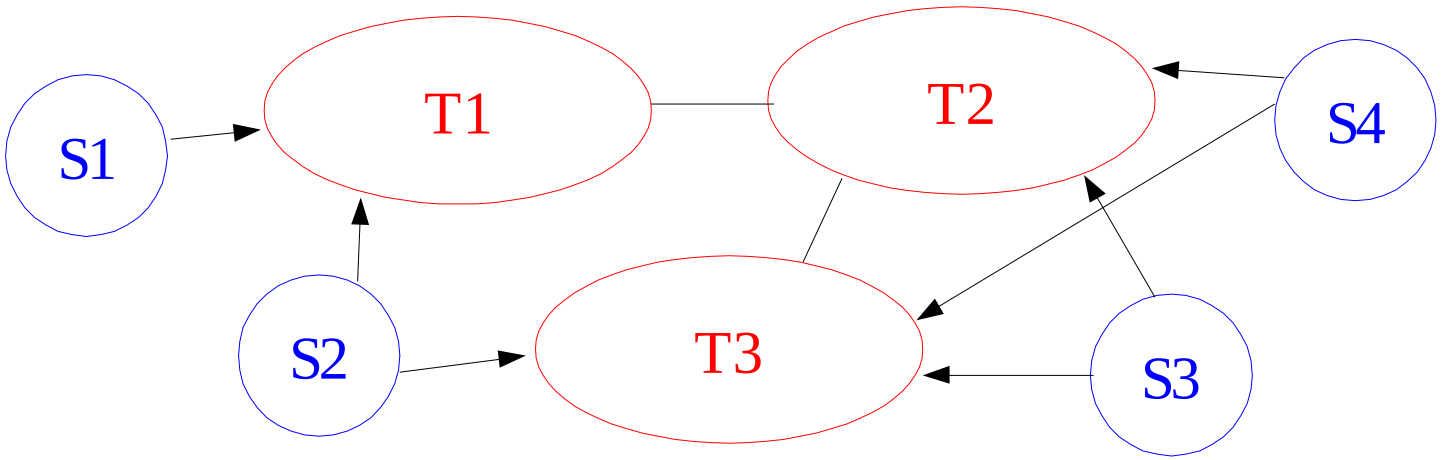


Organization of the Internet

- Internet is composed of more than 10.000 autonomous routing domains
  - A domain is a set of routers, links, hosts and local area networks under the same administrative control
    - A domain can be very large...
      - AS568: SUMNET-AS DISO-UNRRA contains 73154560 IP addresses
    - A domain can be very small...
      - AS2111: IST-ATRIUM TE Experiment a single PC running Linux...
  - Domains are interconnected in various ways
    - The interconnection of all domains should in theory allow packets to be sent anywhere
    - Usually a packet will need to cross a few ASes to reach its destination

Types of domains

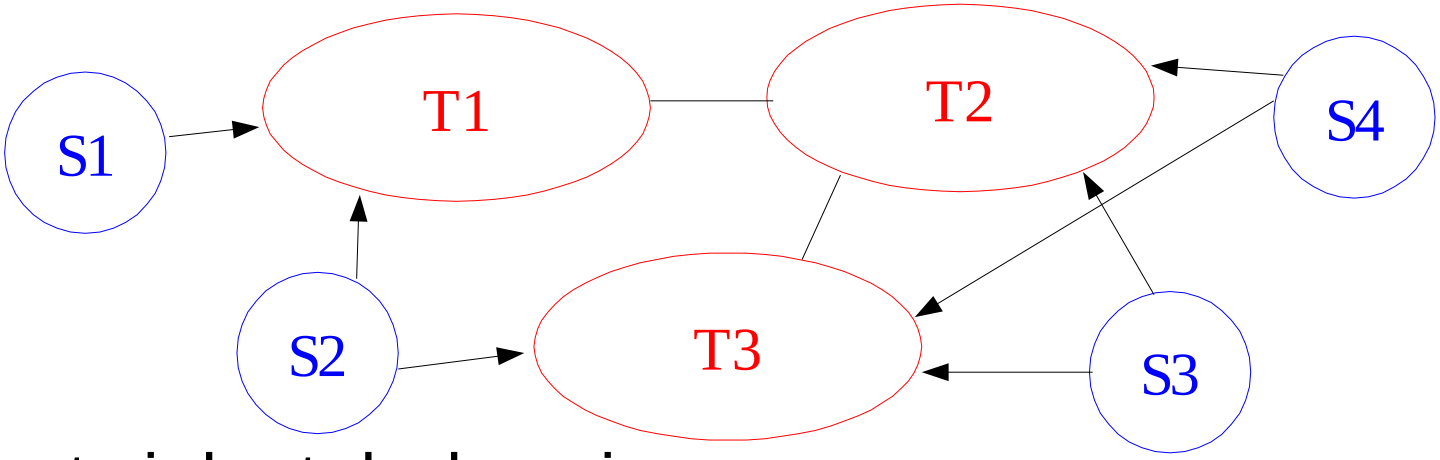
- Transit domain
  - A transit domain allows external domains to use its own infrastructure to send packets to other domains



- Examples
  - UUNet, OpenTransit, GEANT, Internet2, RENATER, EQUANT, BT, Telia, Level3,...

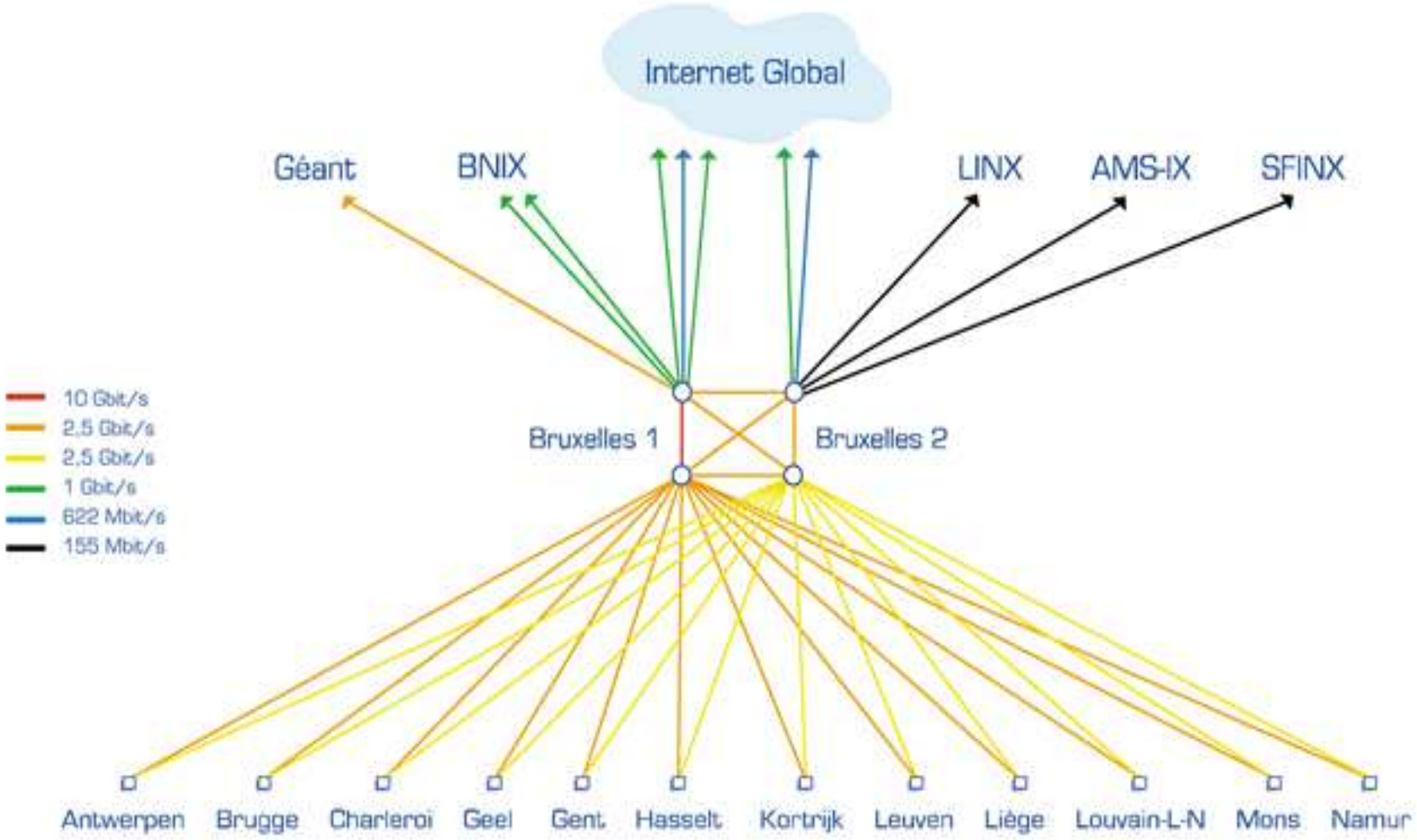
Types of domains (2)

- Stub domain
  - A stub domain does not allow external domains to use its infrastructure to send packets to other domains
    - A stub is connected to at least one transit domain
      - Single-homed stub : connected to one transit domain
      - Dual-homed stub : connected to two transit domains

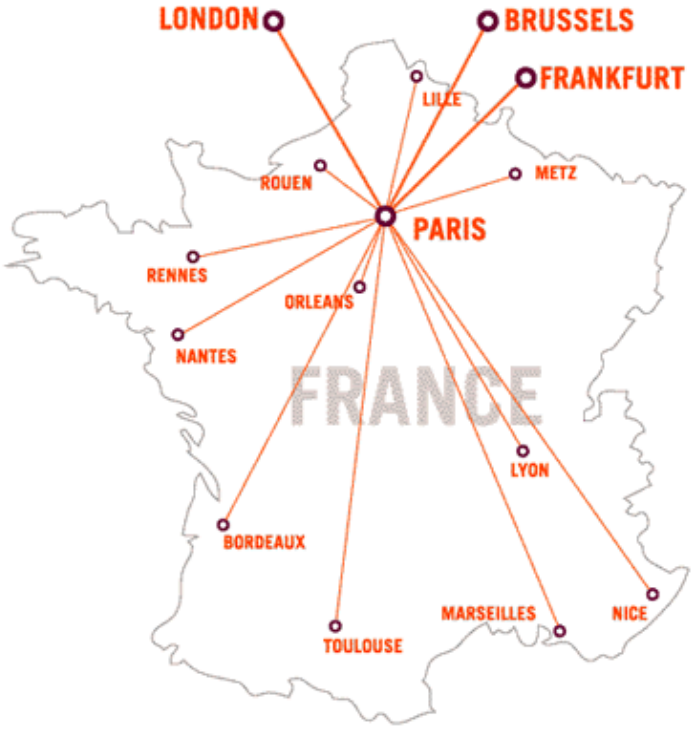


  - Content-rich stub domain
    - Large web servers : Yahoo, Google, MSN, TF1, BBC,...
  - Access-rich stub domain
    - ISPs providing Internet access via CATV, ADSL, ...

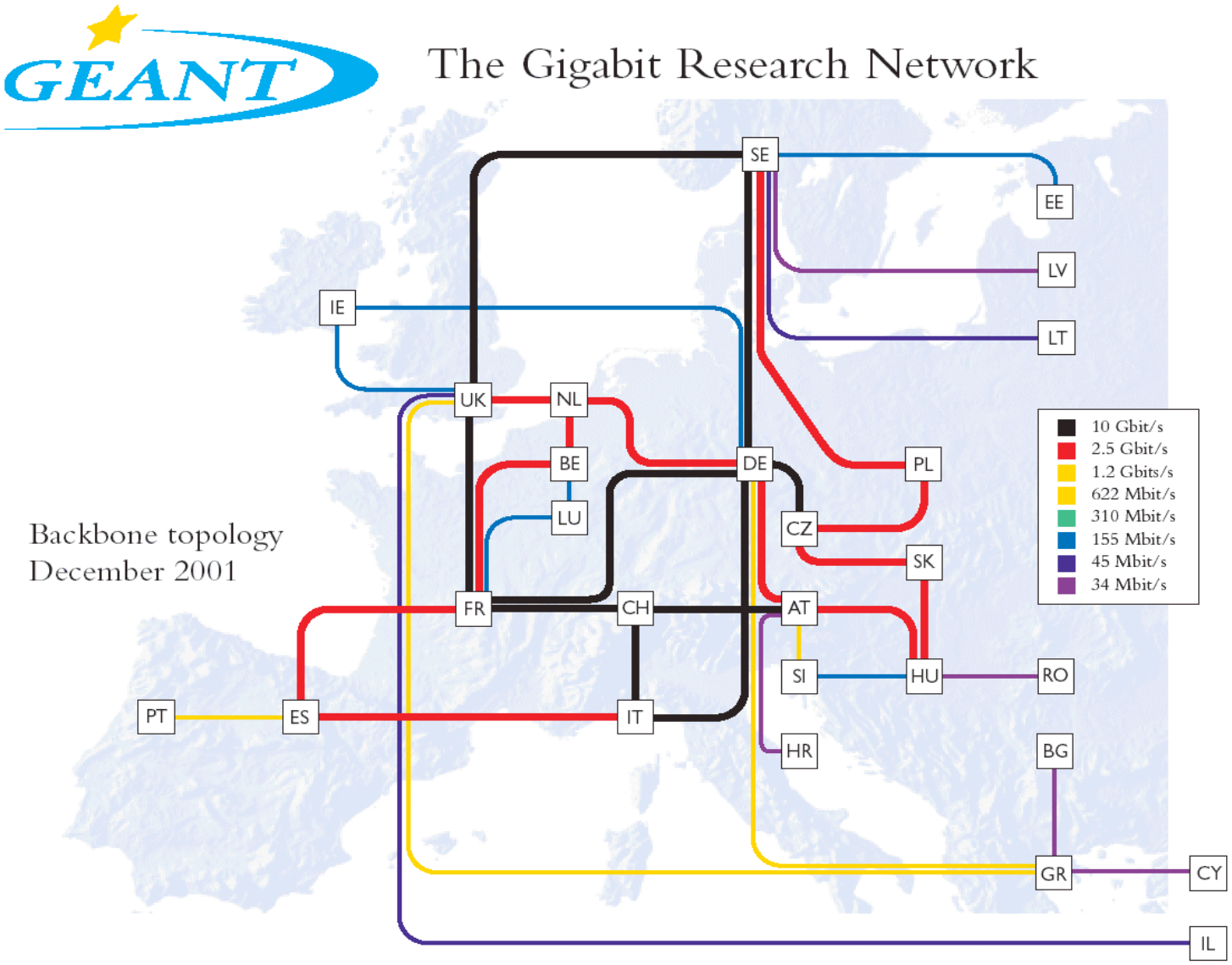
A Stub domain : Belnet



A transit domain : Easynet

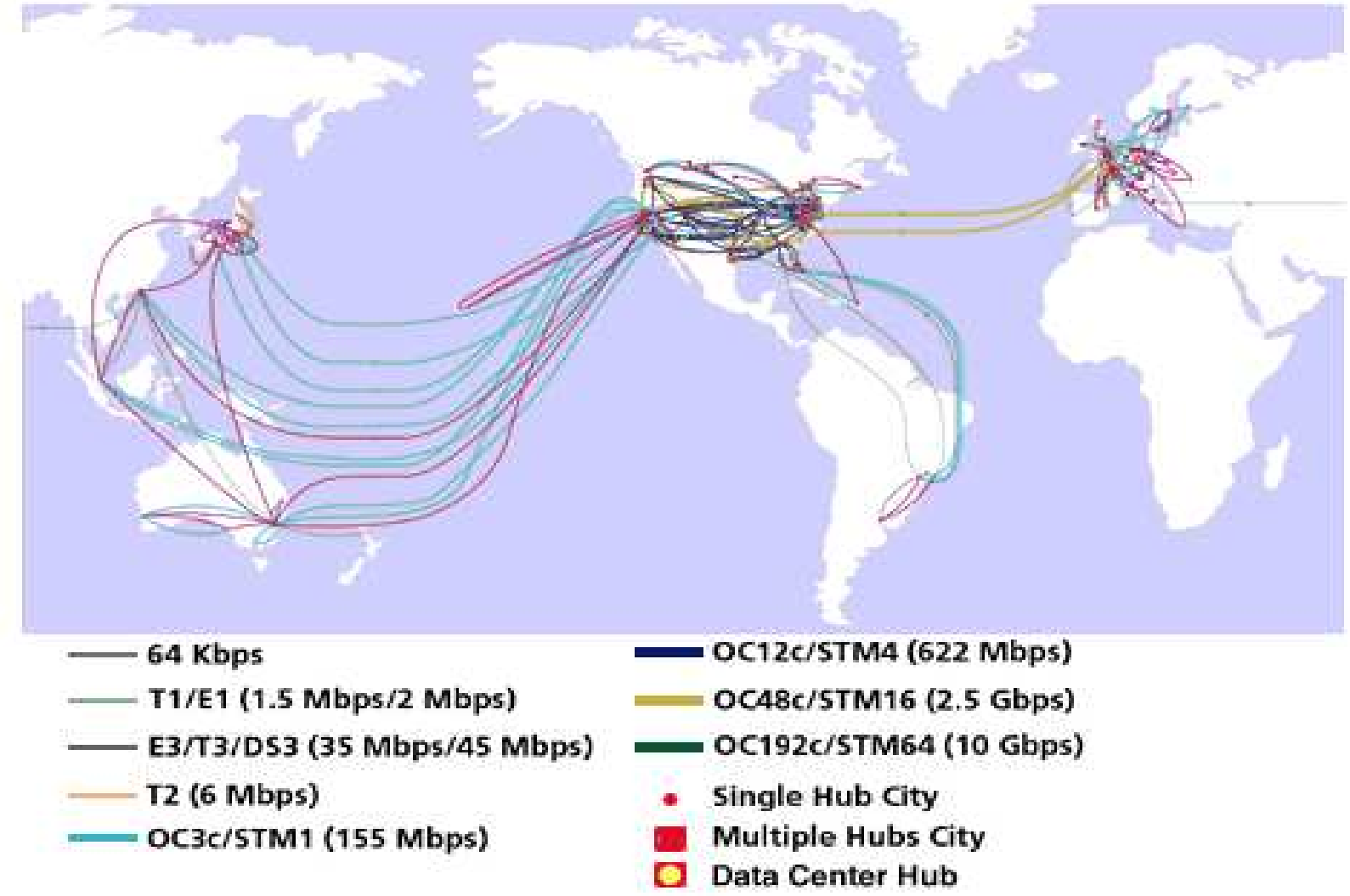


A transit domain : GEANT



A transit domain : BT/IGnite

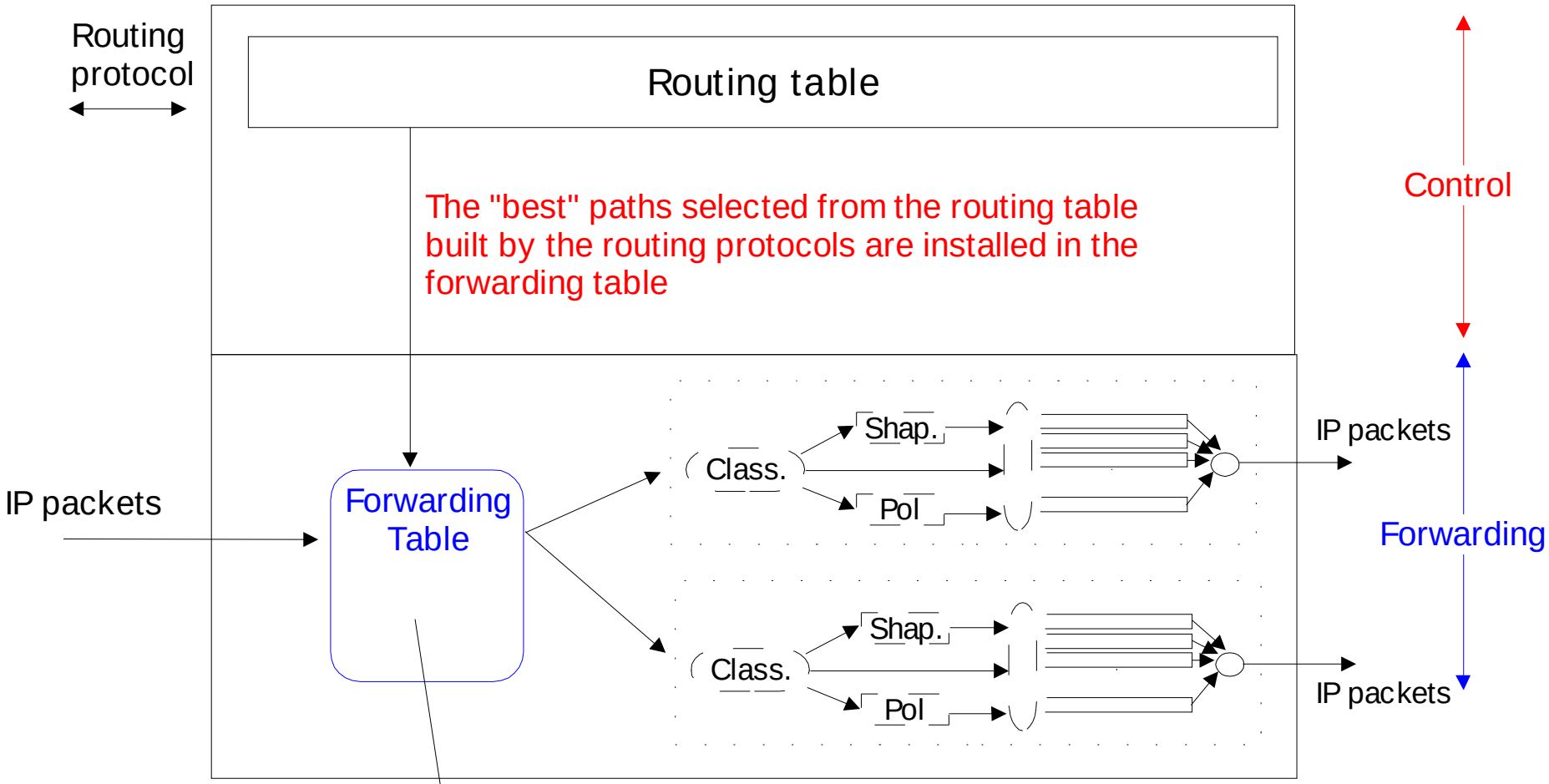
A large transit domain : UUNet



Outline

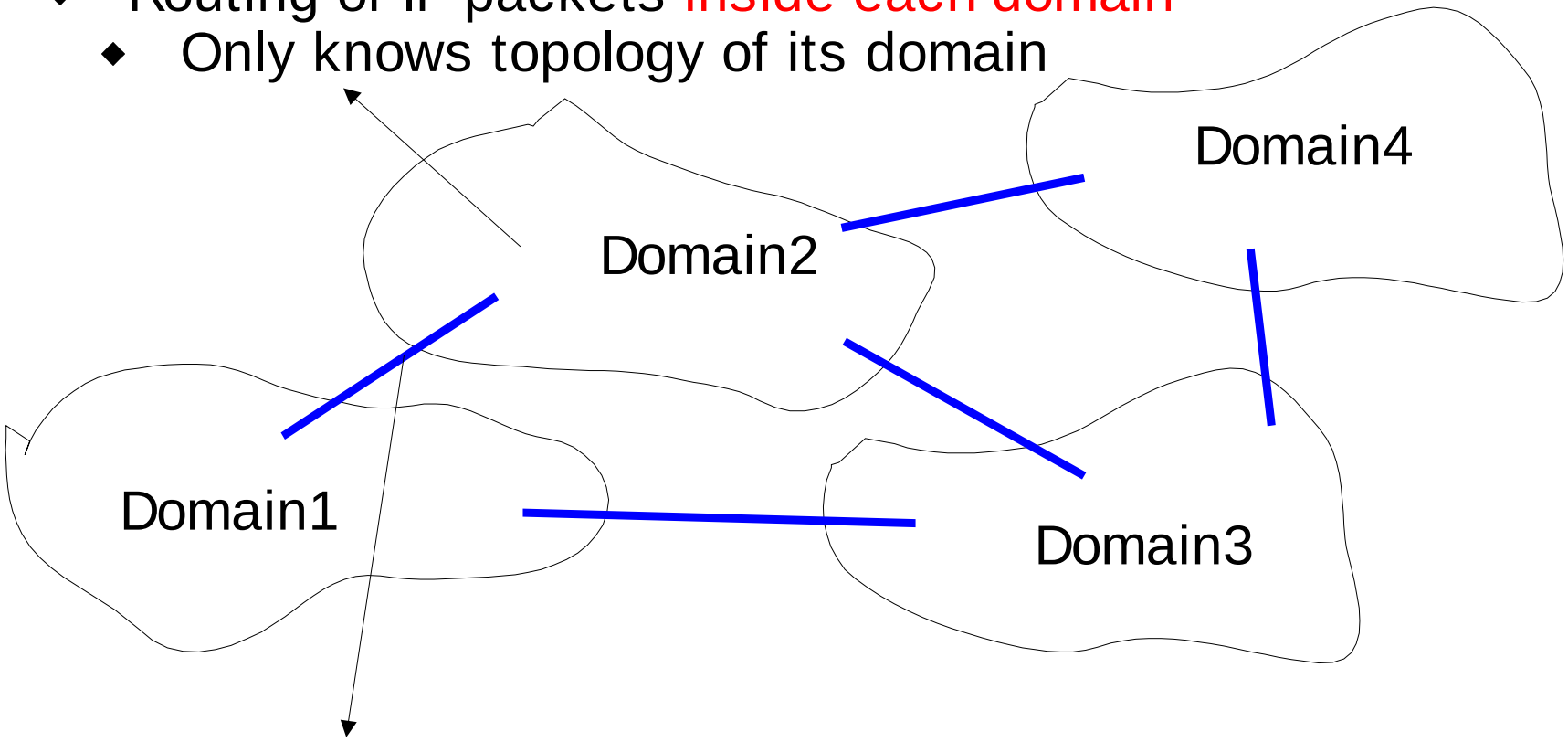
- Organization of the global Internet
  - Example of domains
  - Intradomain routing
- BGP basics
- BGP in large networks
- Interdomain traffic engineering with BGP
- BGP-based Virtual Private Networks

Architecture of a normal IP router



Internet routing

- Interior Gateway Protocol (IGP)
  - Routing of IP packets inside each domain
    - Only knows topology of its domain



- Exterior Gateway Protocol (EGP)
  - Routing of IP packets between domains
    - Each domain is considered as a blackbox

Intradomain routing

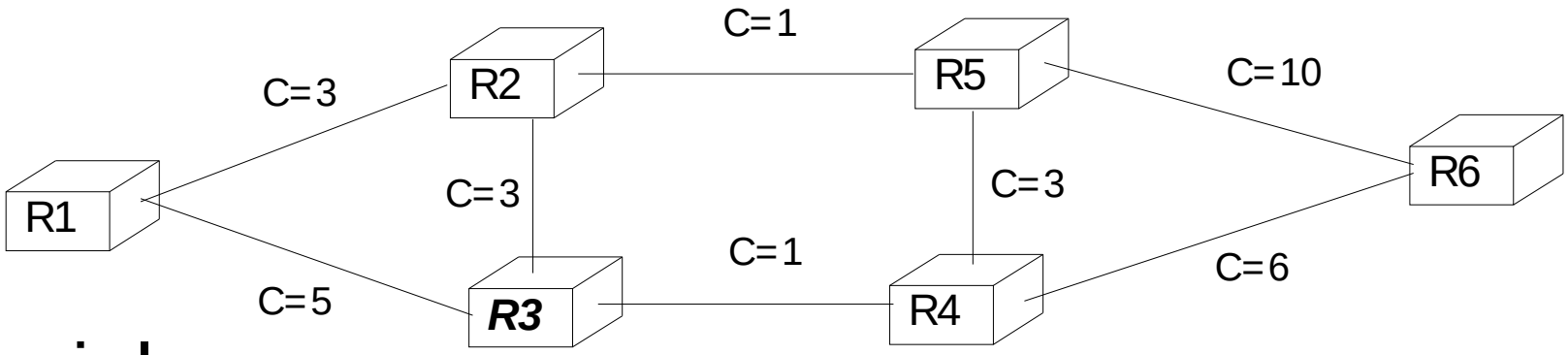
- Goal
  - Allow routers to transmit IP packets along the best path towards their destination
    - best usually means the shortest path
      - Shortest measured in seconds or as number of hops
    - sometimes best means the less loaded path
  - Allow to find alternate routes in case of failures

- Behavior
  - All routers exchange routing information
    - Each domain router can obtain routing information for the whole domain
    - The network operator or the routing protocol selects the cost of each link

Three types of Interior Gateway Protocols

- Static routing
  - Only useful in very small domains
- Distance vector routing
  - Routing Information Protocol (RIP)
    - Still widely used in small domains despite its limitations
- Link-state routing
  - Open Shortest Path First (OSPF)
    - Widely used in enterprise networks
  - Intermediate System- Intermediate-System (IS-IS)
    - Widely used by ISPs

Distance vector routing

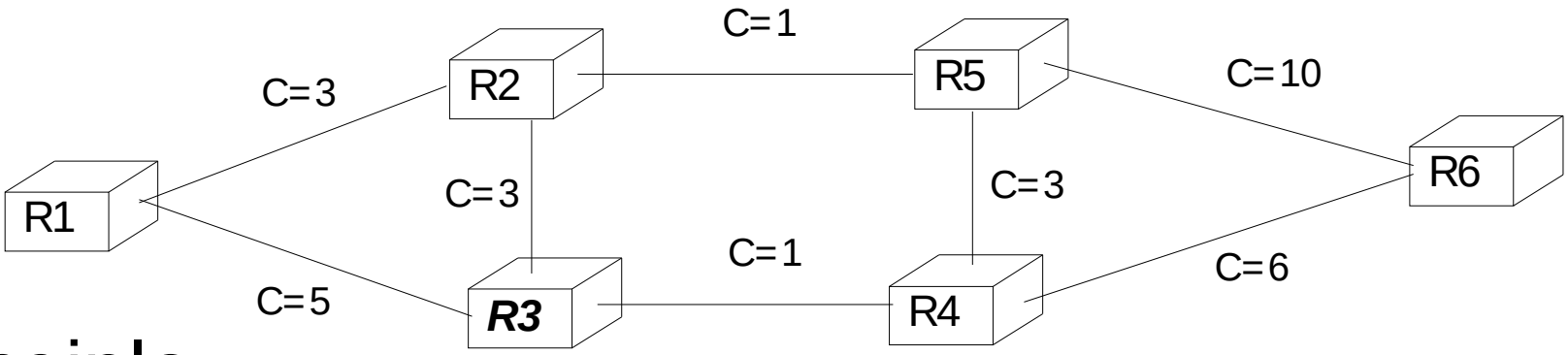


- Principle
  - Router configuration
    - Cost associated with each link
  - Each router sends <u>periodically</u> a distance vector containing, for each known prefix, :
    1. The IP prefix
    2. The distance between itself and the destination
       - The distance vector is a summary of the router's routing table
  - Each router receives its neighbor's distance vectors and builds its routing table based on those vectors

Issues with distance vector routing

- How to deal with link failures ?
  - Routers should send their distance vector when they detect the failure of one of their links
- How to avoid the count-to-infinity problem ?
  - Utilize a non-redundant star shaped network
  - Limit the maximum distance between routers
    - For RIP, ∞=16 !
  - Split horizon
    - Router A does not advertise to router B the routes for which it sends packets via router B
  - Split horizon with Poison reverse

Link state routing



- Principle
  - Each router builds link state packet containing its local topology
    - Link state packets are created at regular intervals and when the local topology changes
  - Link state packet is reliably flooded to all routers inside the domain
  - Each router knows the complete domain topology
  - Computes routing tables by using Dijkstra
    - The best path is the path with the smallest cost

IP forwarding

- Usually
  - Forwarding table contains, for each prefix
    - The prefix
    - The best path (outgoing interface) to reach this prefix
- Sometimes
  - Forwarding table contains, for each prefix
    - The prefix
    - N equal cost paths to reach this prefix
      - A first path (outgoing interface) to reach this prefix
      - A second path (outgoing interface) to reach this prefix
      - A third path (outgoing interface) to reach this prefix
      - ...
  - A load balancing mechanism is used to send the IP packets over the N available paths

Load balancing algorithms

- Simple solution
  - Round-Robin or variants to dispatch packets on a per packet basis
  - Advantages
    - easy to implement since number of paths is small
    - traffic will be divided over the equal cost paths on a per packet basis
      - each path will carry the same amount of traffic
  - Drawbacks
    - two packets from the same TCP connection may be sent on different paths and thus be reordered
      - TCP performance can be affected by reordering

Load balancing algorithms (2)

- How to perform load balancing without maintaining state for each TCP connection ?
  - Principle
    - concatenate IP src, IP dest, IP protocol, Src port, and Dest port from the IP packet inside a bit string
      - bitstring = [IP src:IP dest:IP protocol:Src port:Dest port]
    - compute path = Hash(bitstring) mod P
      - hash function should be easy to implement and should produce very different numbers for close bitstring values
      - candidate hash functions are CRC, checksum, ...
  - Advantages
    - all packets from TCP connection sent on same path
    - traffic to a server will be divided over the links
  - Drawback
    - does not work well if a few TCP connections carry a large fraction of the total traffic

Summary

- Types of domains
  - Transit domain
  - Stub domain
- Intradomain routing
  - Selects the best route towards each destination based on one metric
    - Static routing
    - Distance vector routing
    - Link-state routing
  - Load balancing methods allow to place several paths in the forwarding table