Internet Routing Protocols
Lecture 01 & 02

Advanced Systems Topics

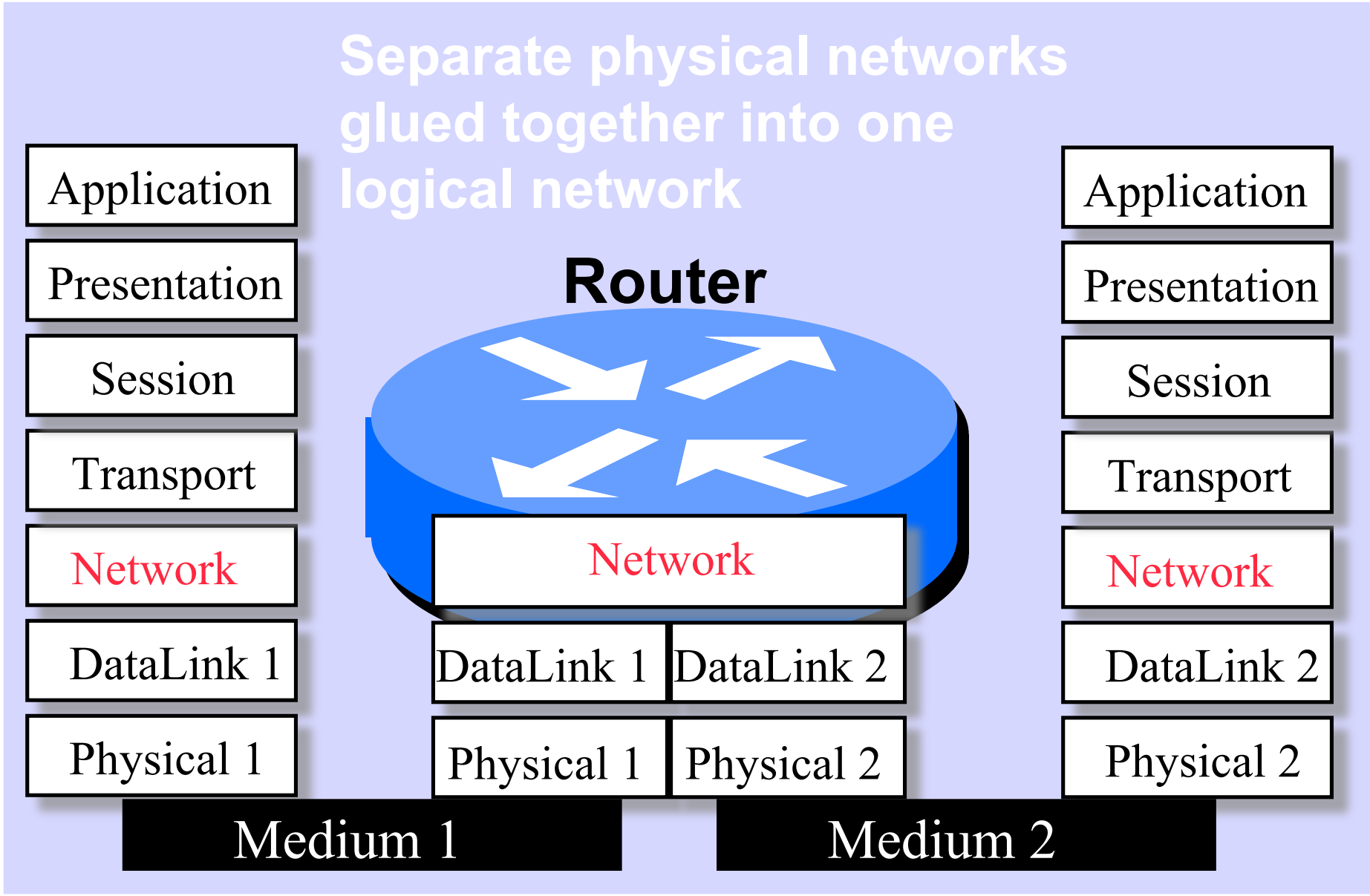
Lent Term, 2010

Timothy G. Griffin
Computer Lab
Cambridge UK

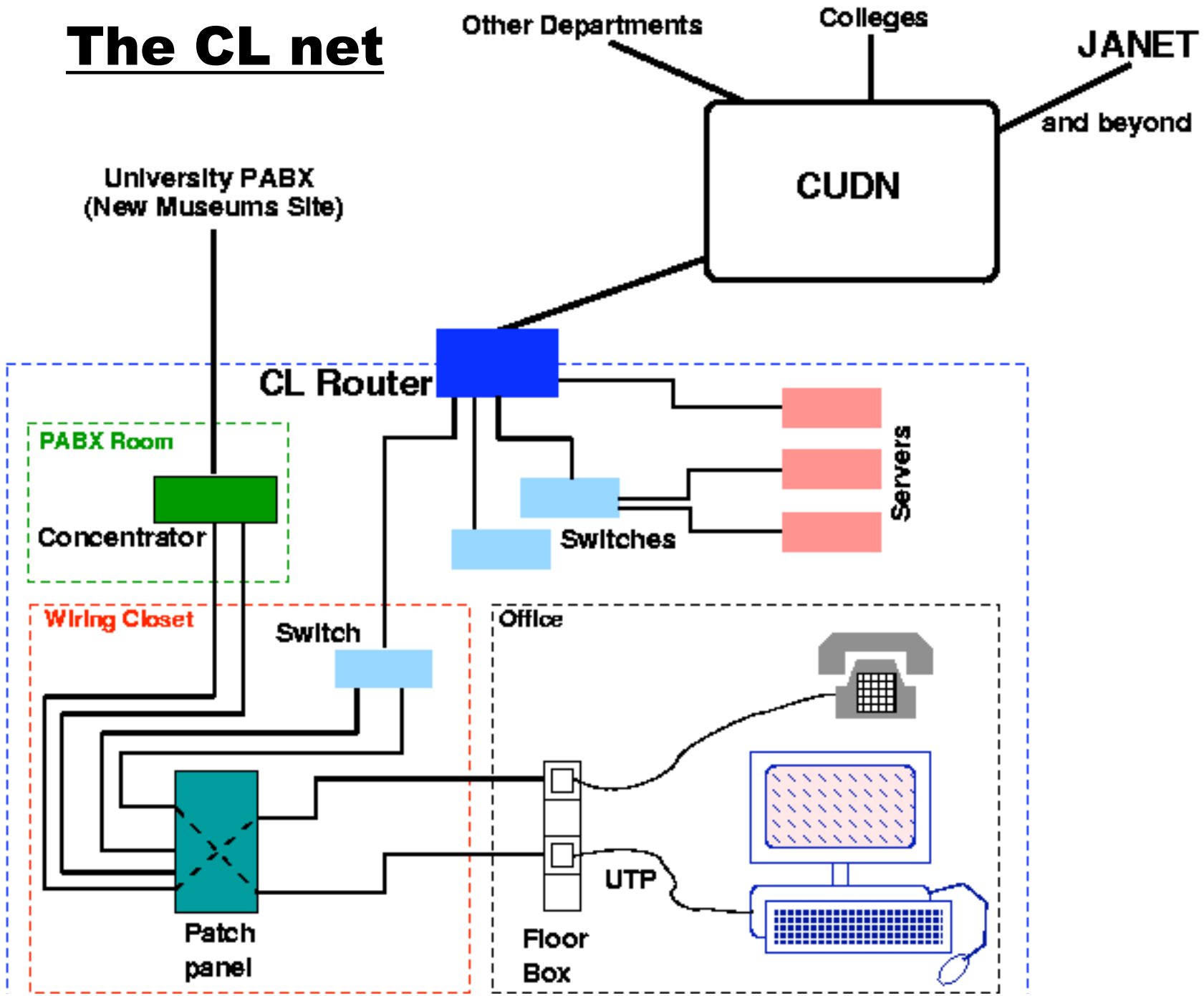
Internet Routing Outline

- **Lecture 1 : Inter-domain routing architecture, the Border Gateway Protocol (BGP)**
- **Lecture 2: More BGP.**
- **Lecture 3 : BGP traffic engineering and protocol dynamics**
- **Lecture 5 : Locator/ID split to the rescue?**
- **Lecture 6 : How has the global Internet changed in the last 10 years?**

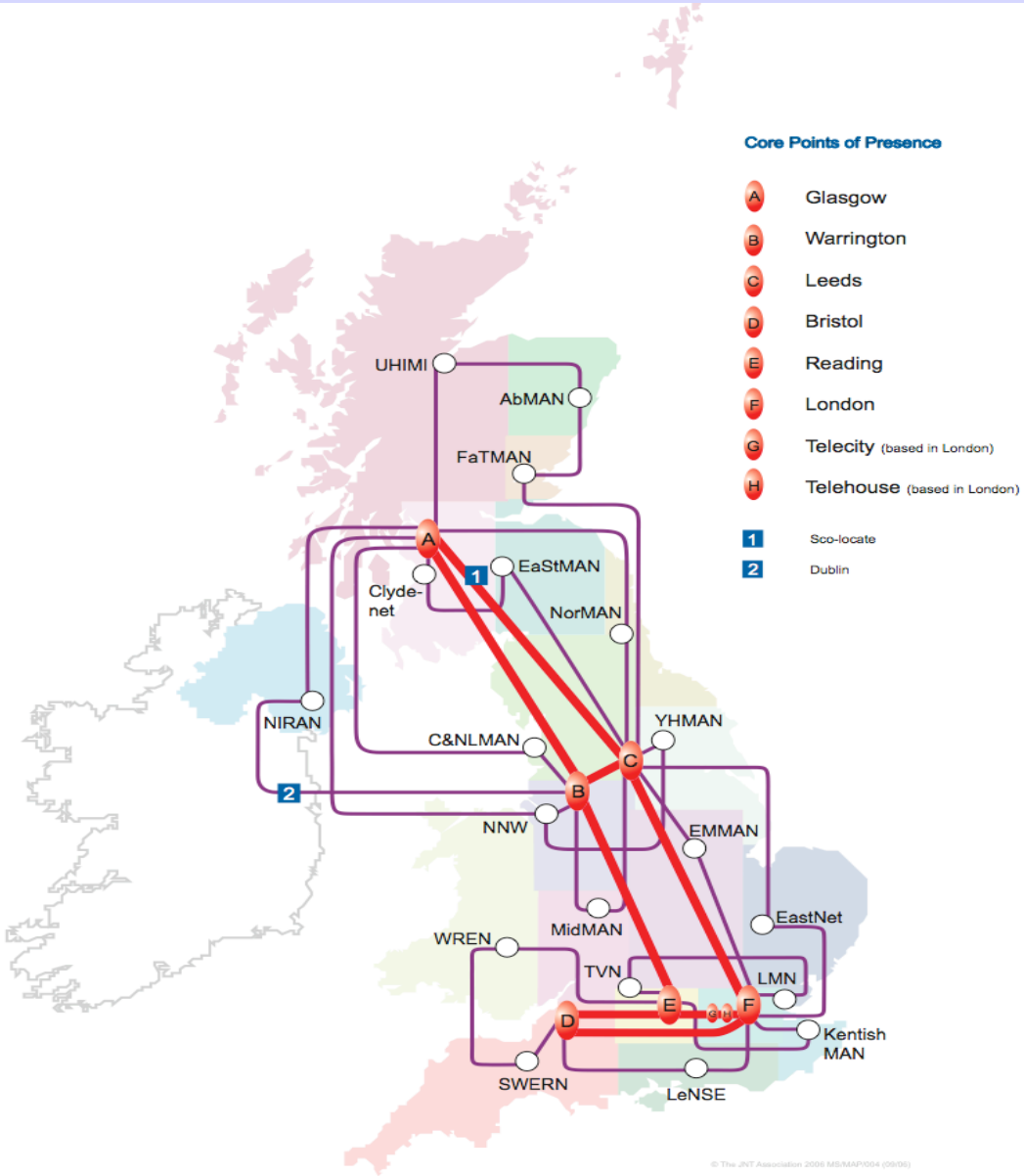
IP is a Network Layer Protocol



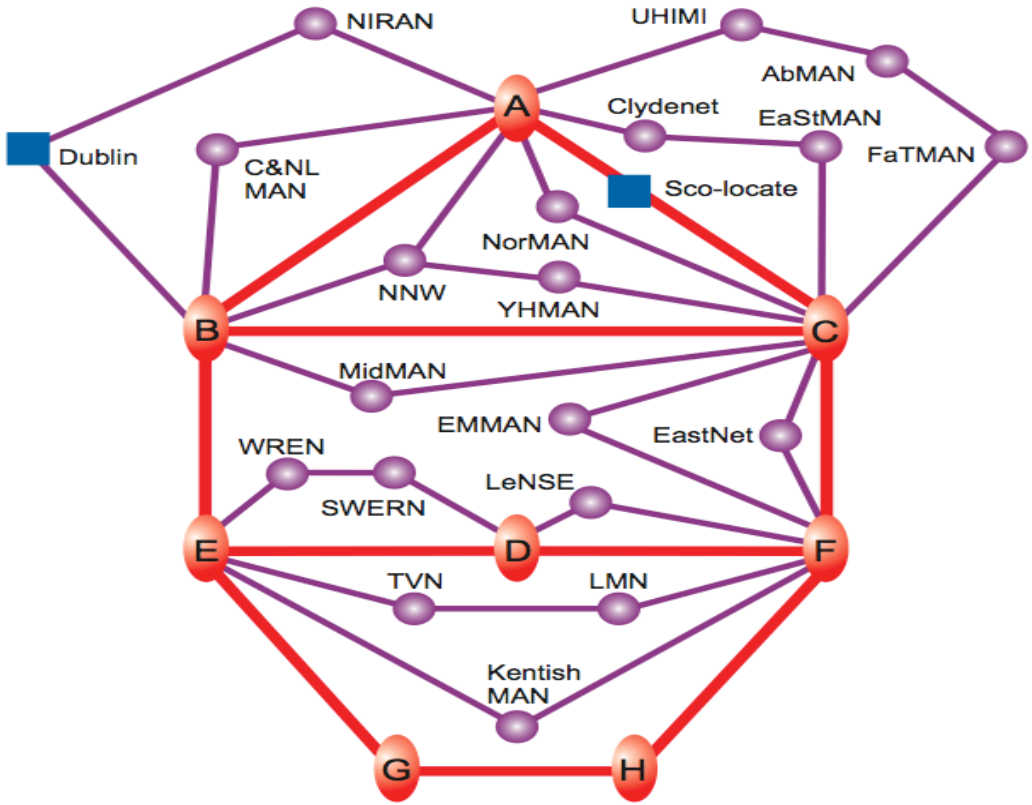
The CL net


Other Departments
Colleges
JANET
and beyond
CUDN
University PABX
(New Museums Site)
CL Router
PABX Room
Concentrator
Switches
Servers
Wiring Closet
Switch
Office
Patch
panel
UTP
Floor
Box

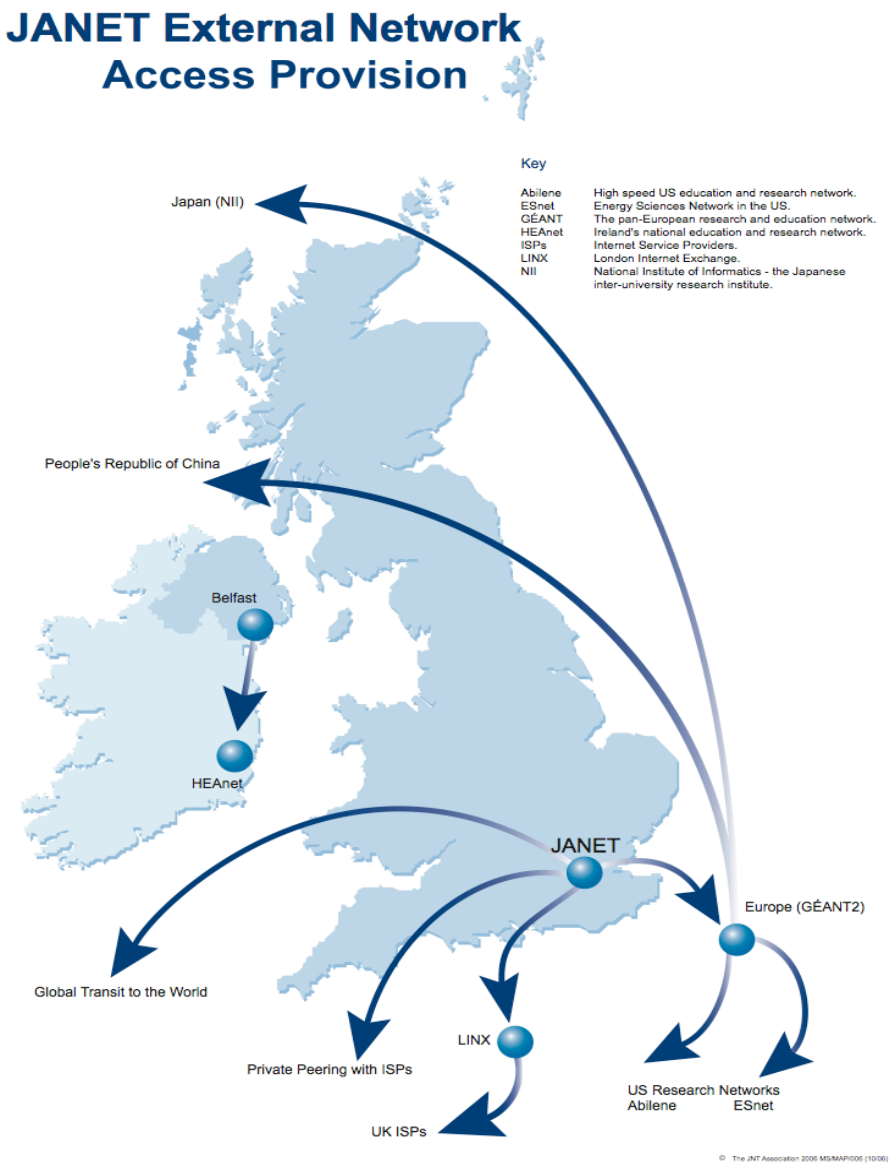
JANET

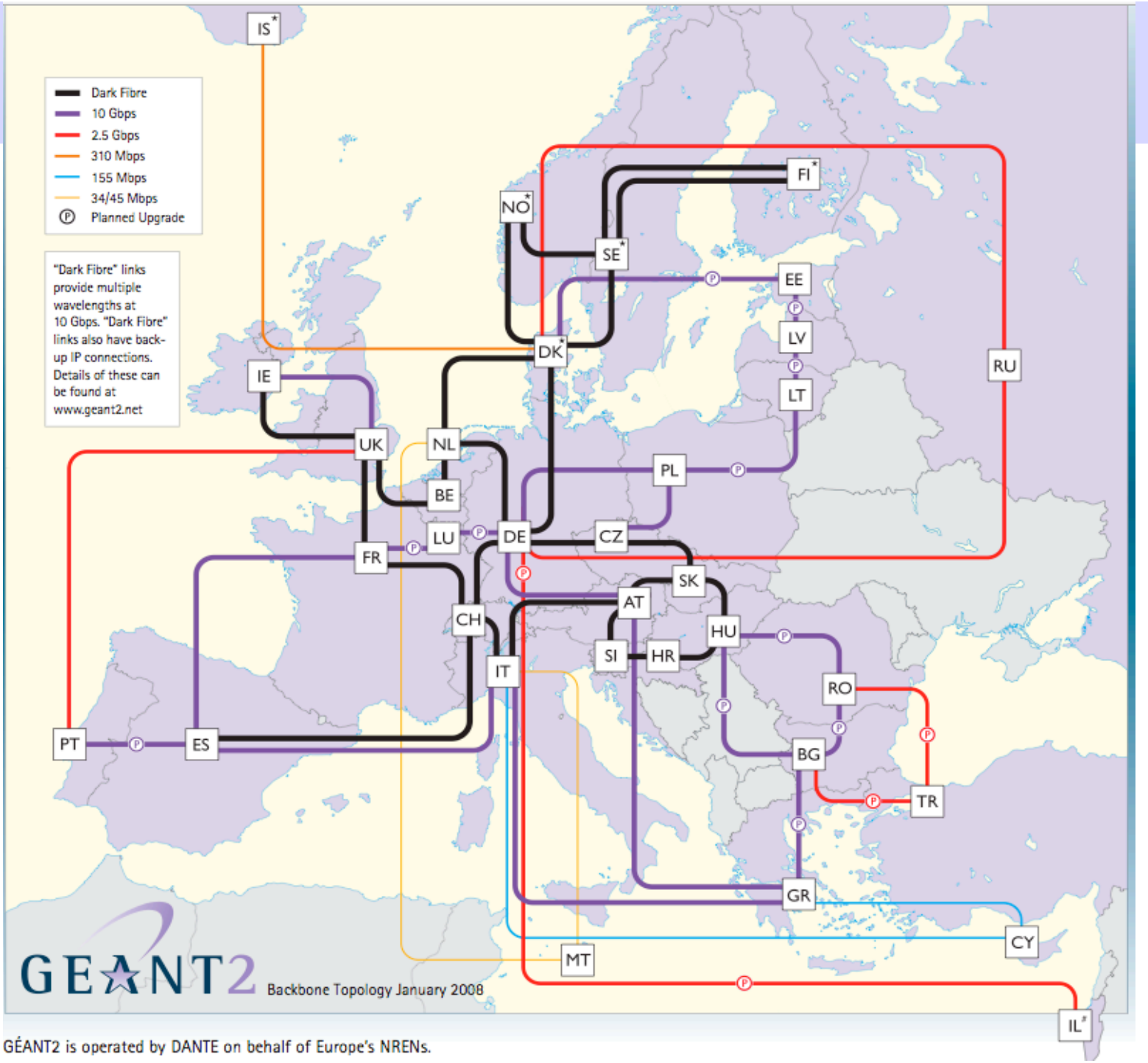

Core Points of Presence
A Glasgow
B Warrington
C Leeds
D Bristol
E Reading
F London
G Telecity (based in London)
H Telehouse (based in London)
1 Sco-locate
2 Dublin
UHIMI
AbMAN
FaTMAN
A
EaStMAN
1
Clyde-net
NorMAN
NIRAN
C&NLMAN
YHMAN
C
B
2
NNW
EMMAN
EastNet
MidMAN
WREN
TVN
LMN
E
F
D
Kentish MAN
SWERN
LeNSE

JANET Design

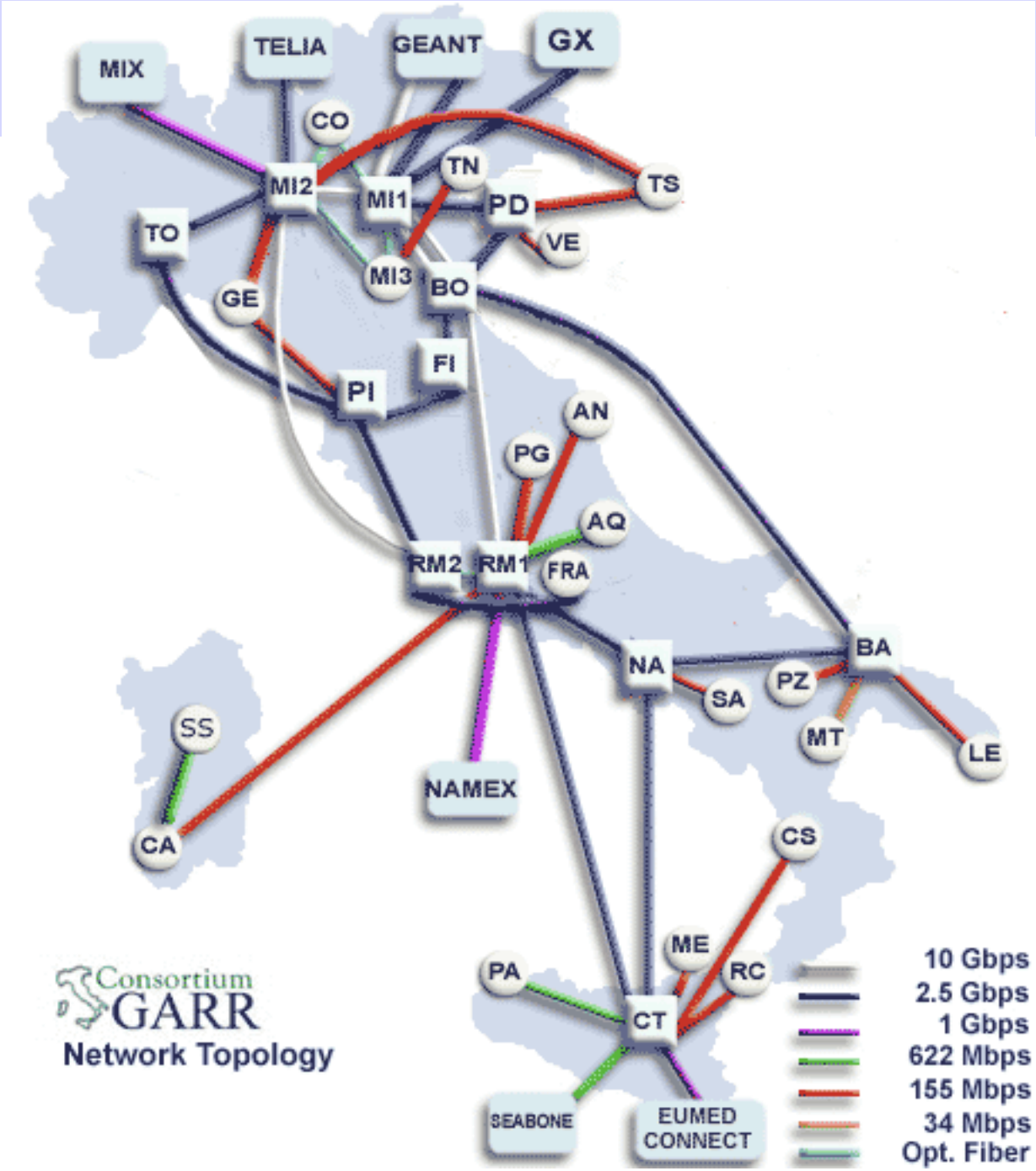


JANET and the Internet

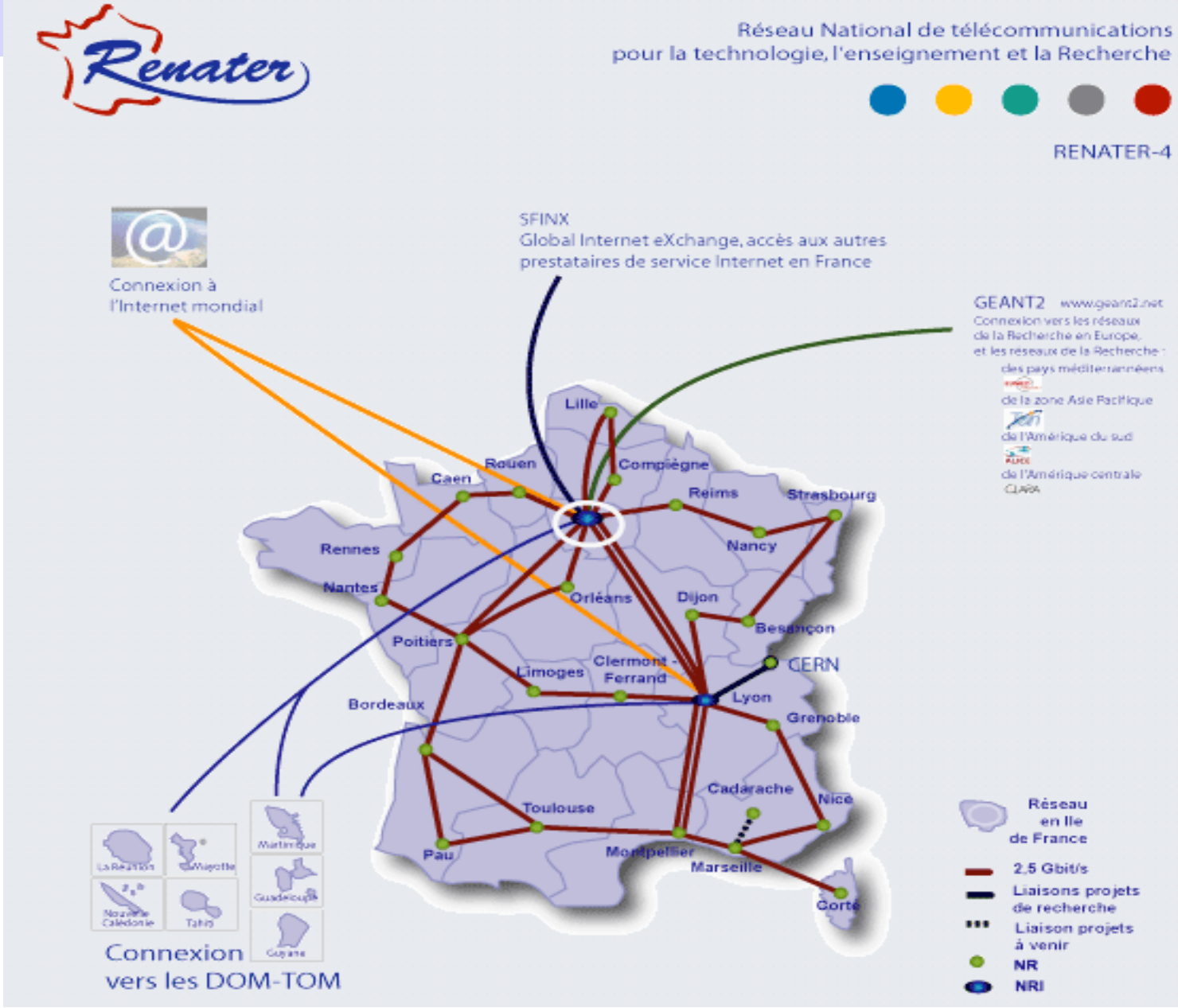




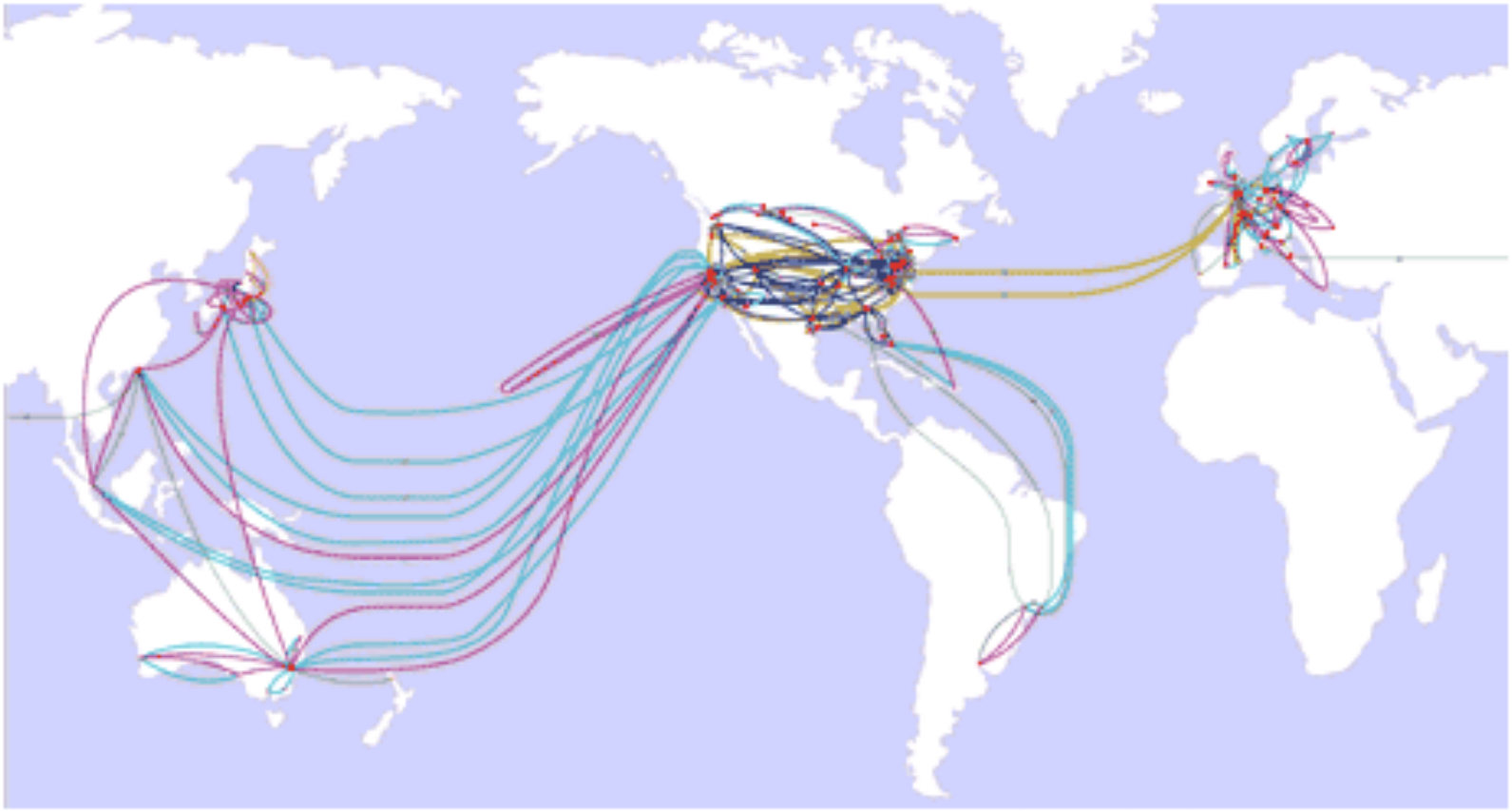
GÉANT2 is operated by DANTE on behalf of Europe's NRENs.


MIX
TELIA
GEANT
GX
CO
MI2
MI1
TN
TS
PD
VE
TO
MI3
BO
GE
FI
PI
AN
PG
AQ
RM2
RM1
FRA
NA
BA
PZ
SA
MT
LE
SS
NAMEX
CA
CS
ME
RC
PA
CT
SEABONE
EUMED CONNECT
Consortium GARR
Network Topology
10 Gbps
2.5 Gbps
1 Gbps
622 Mbps
155 Mbps
34 Mbps
Opt. Fiber

RENATER-4 is deployed since september 2005

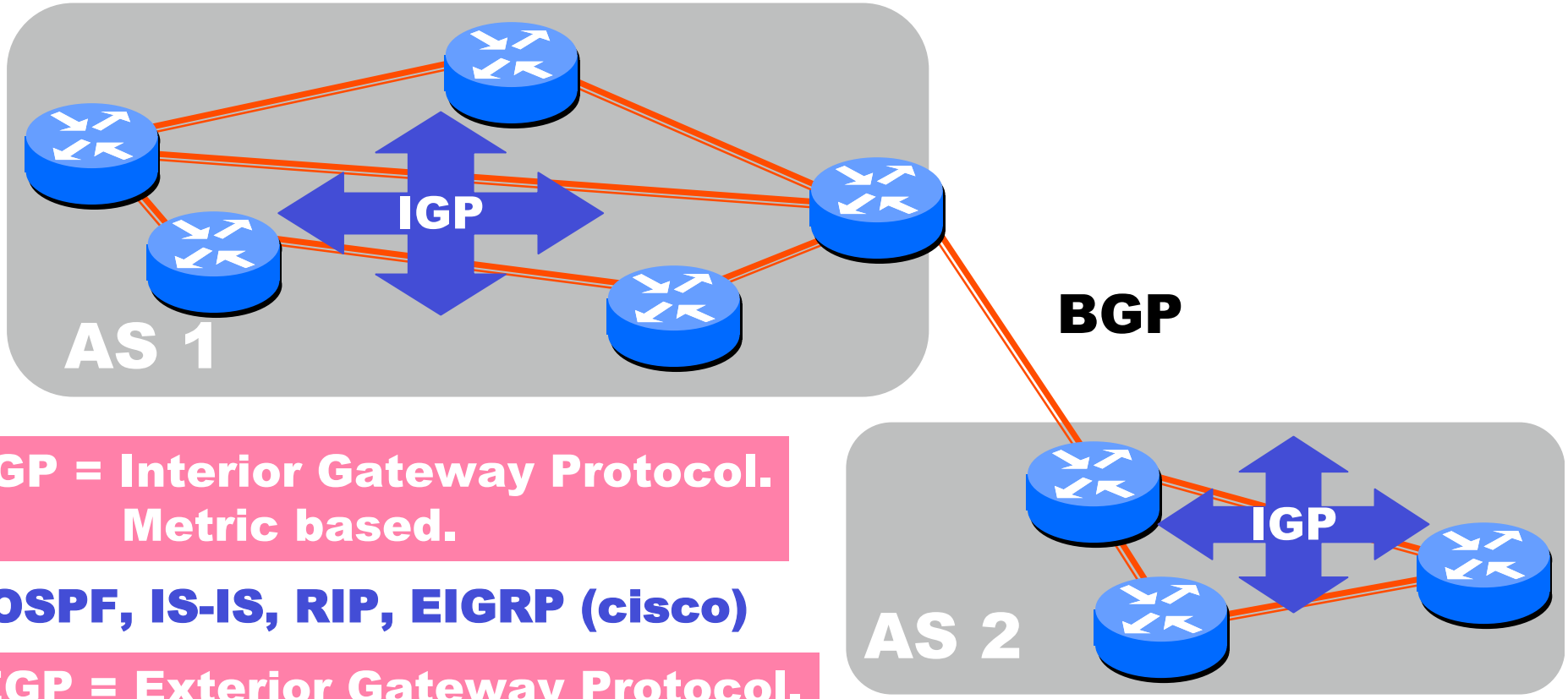


WorldCom (UUNet)



64 Kbps
T1/E1 (1.5 Mbps/2 Mbps)
E3/T3/DS3 (35 Mbps/45 Mbps)
T2 (6 Mbps)
OC3c/STM1 (155 Mbps)
OC12c/STM4 (622 Mbps)
OC48c/STM16 (2.5 Gbps)
OC192c/STM64 (10 Gbps)
Single Hub City
Multiple Hubs City
Data Center Hub

Architecture of Dynamic Routing



IGP = Interior Gateway Protocol.
Metric based.

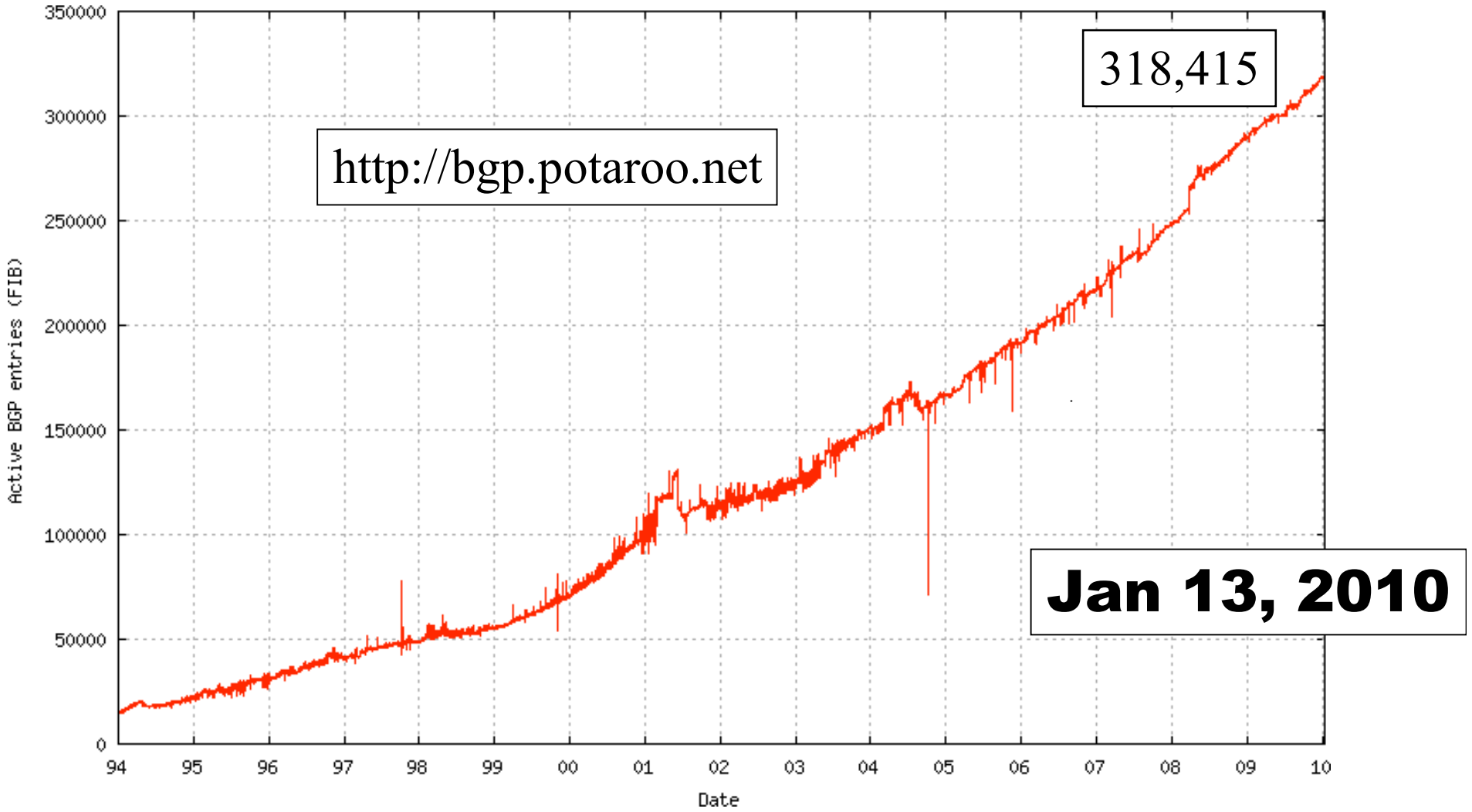
OSPF, IS-IS, RIP, EIGRP (cisco)

EGP = Exterior Gateway Protocol.
Policy Based.

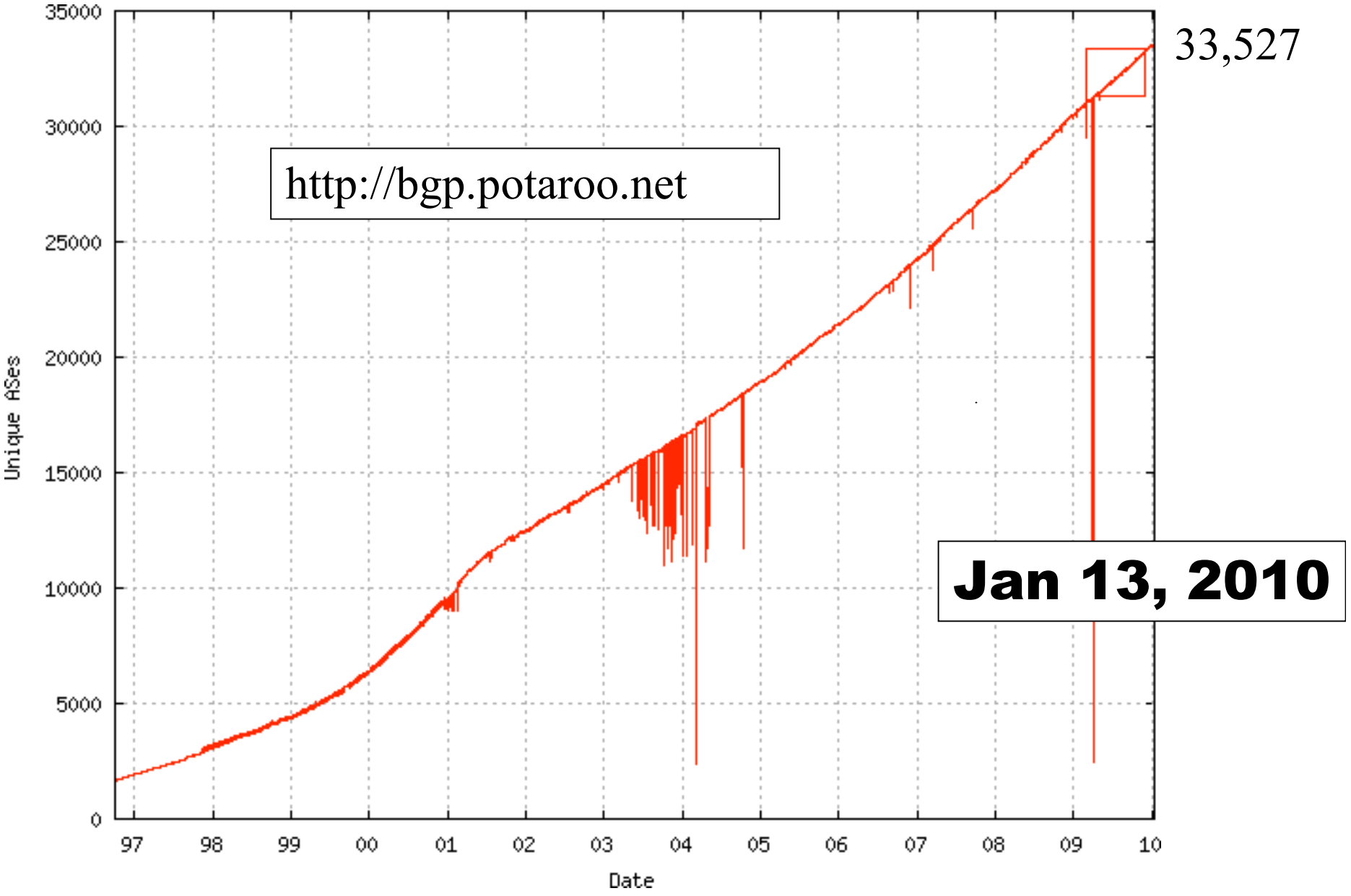
Only one: BGP

The Routing Domain of BGP is the entire Internet

How many prefixes are used today?



How many ASNs are used today?

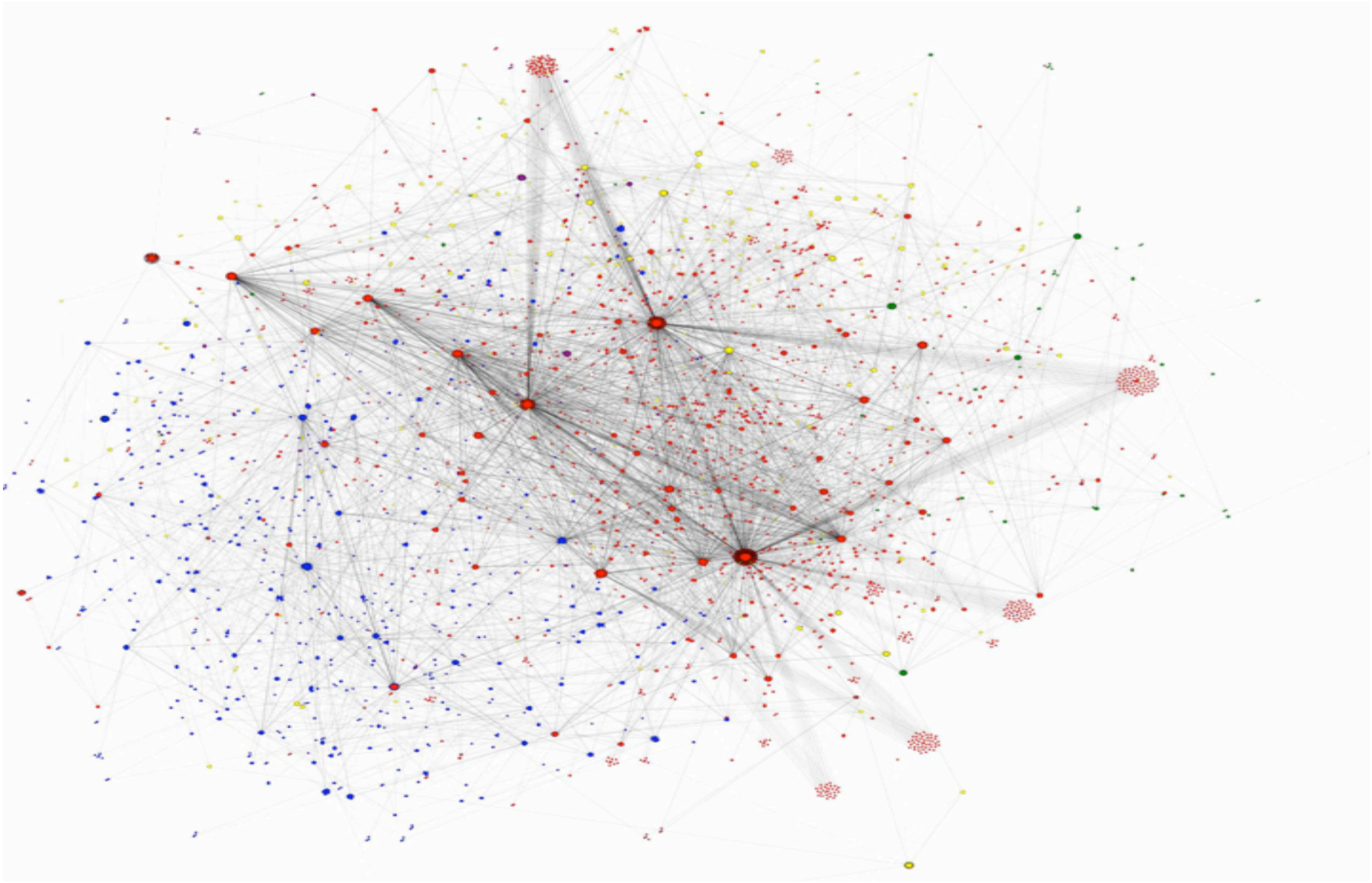


The connectivity of ASNs is hard to visualize

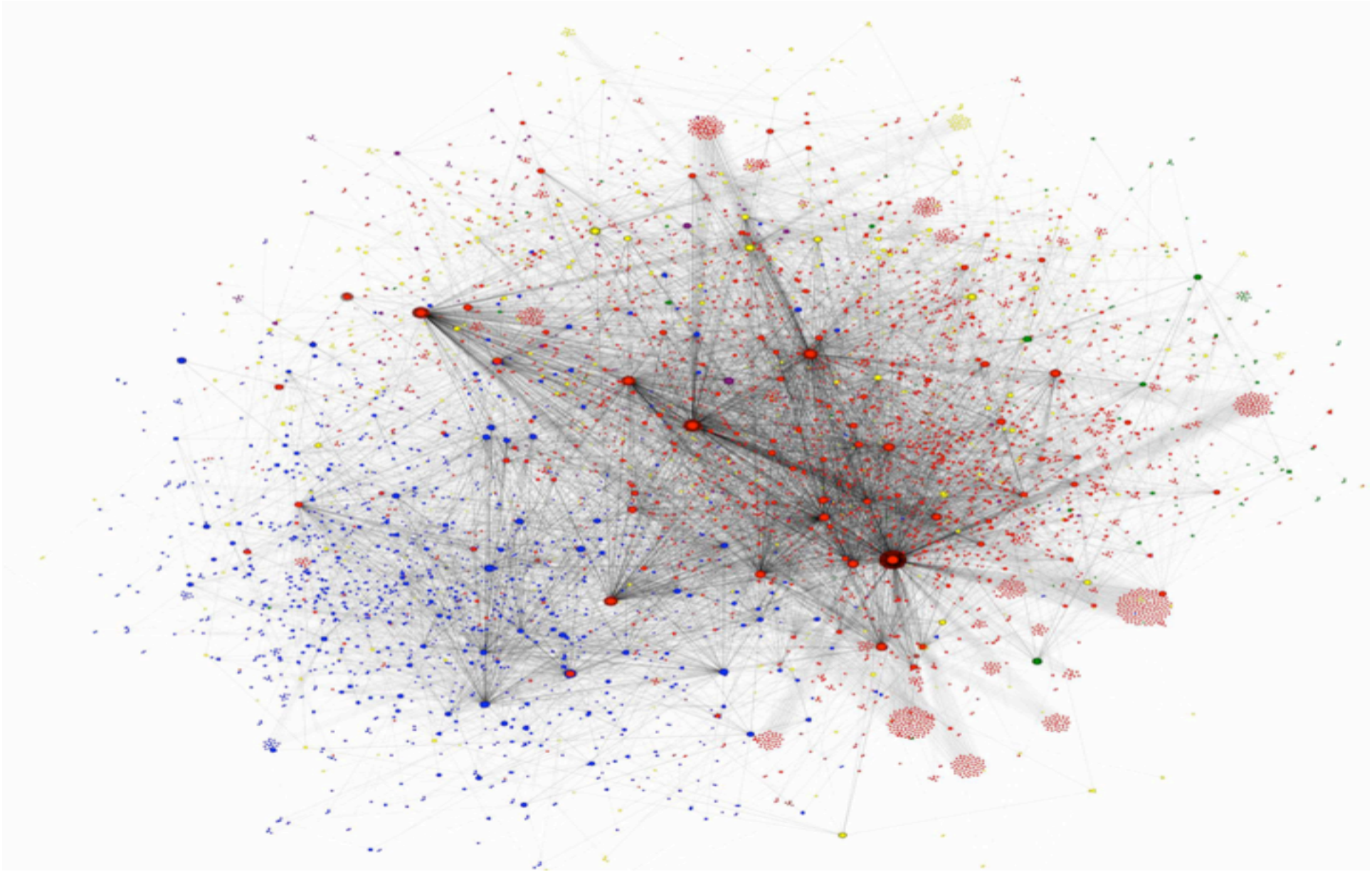
- **The graph is huge.**
- **Transit and stub networks.**
- **How can this be displayed in a meaningful way? and protocol dynamics**
- **My favorite approach:**

Visualizing Internet Evolution on the Autonomous Systems Level
Boitmanis, Krists and Brandes, Ulrik and Pich, Christian (2008)

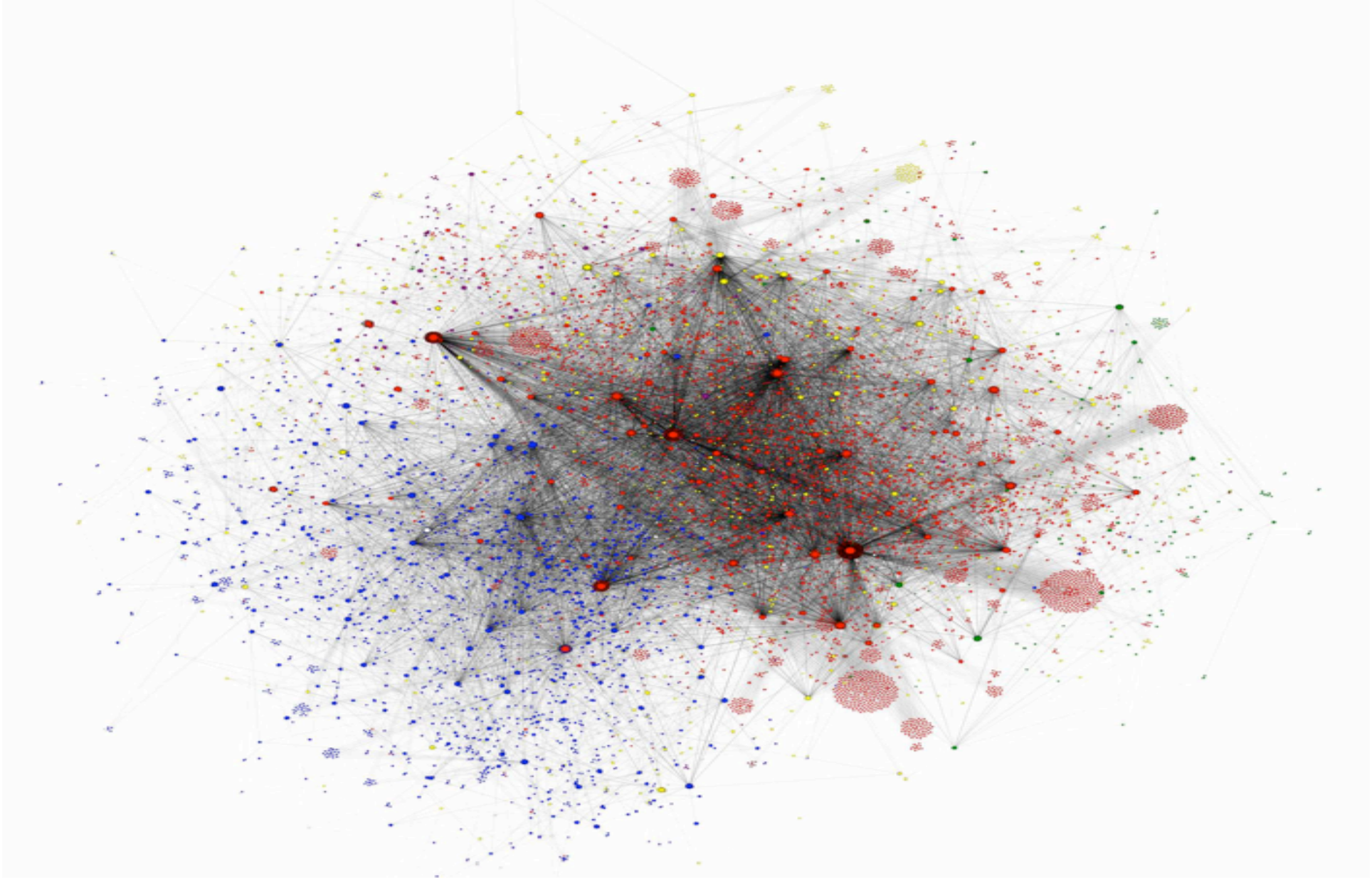
1998



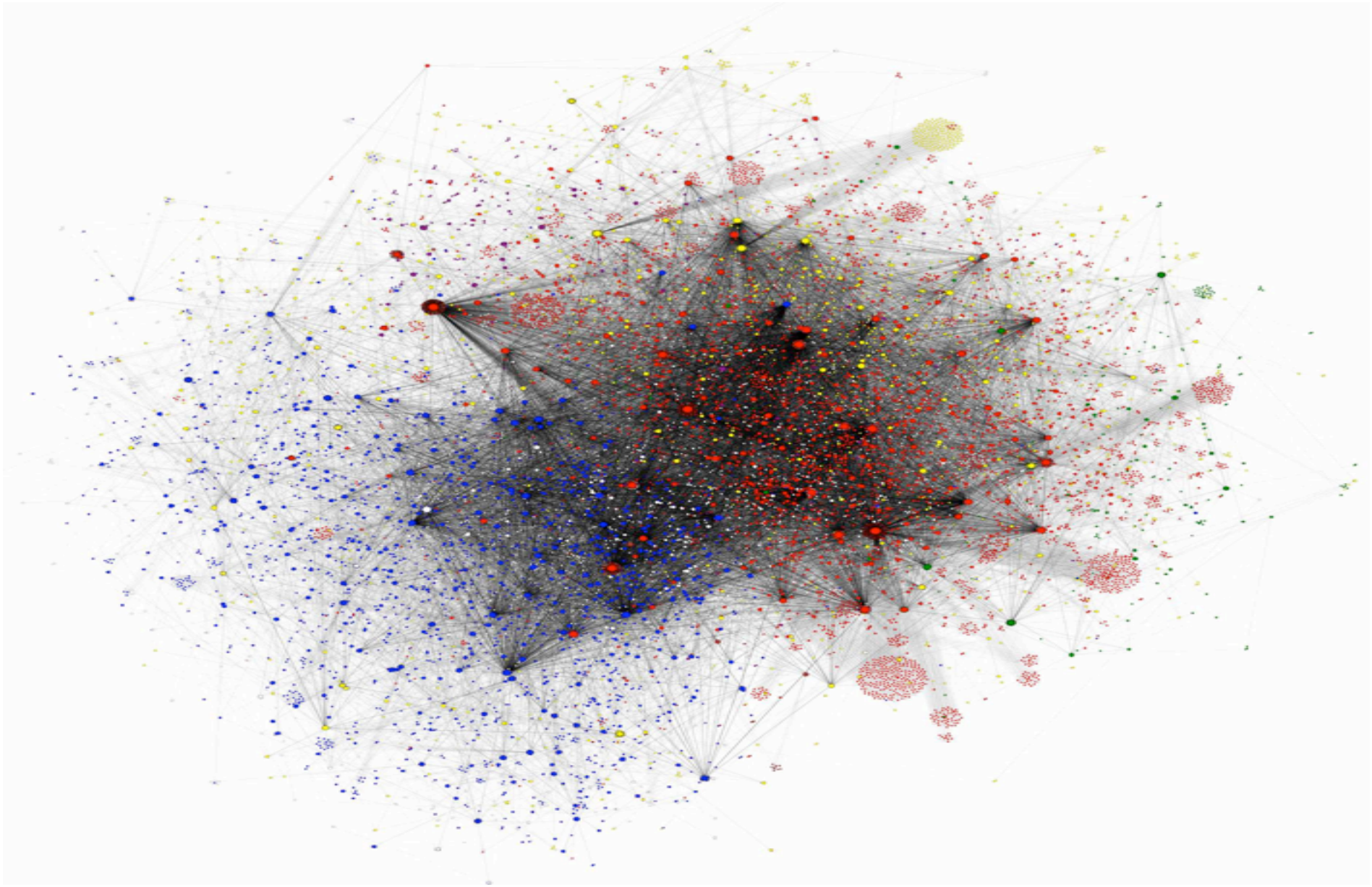
2000



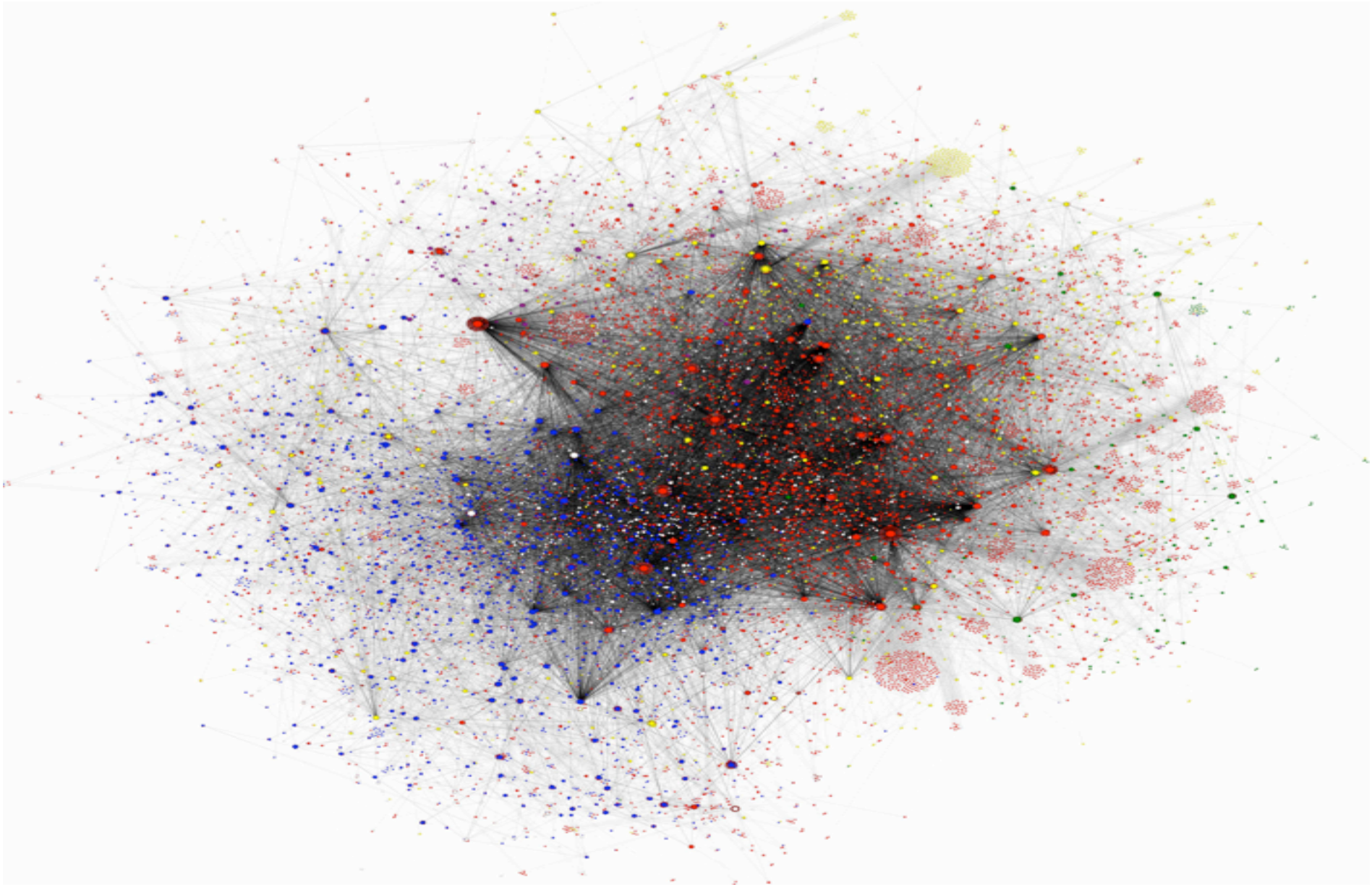
2002



2004



2006



Technology of Distributed Routing

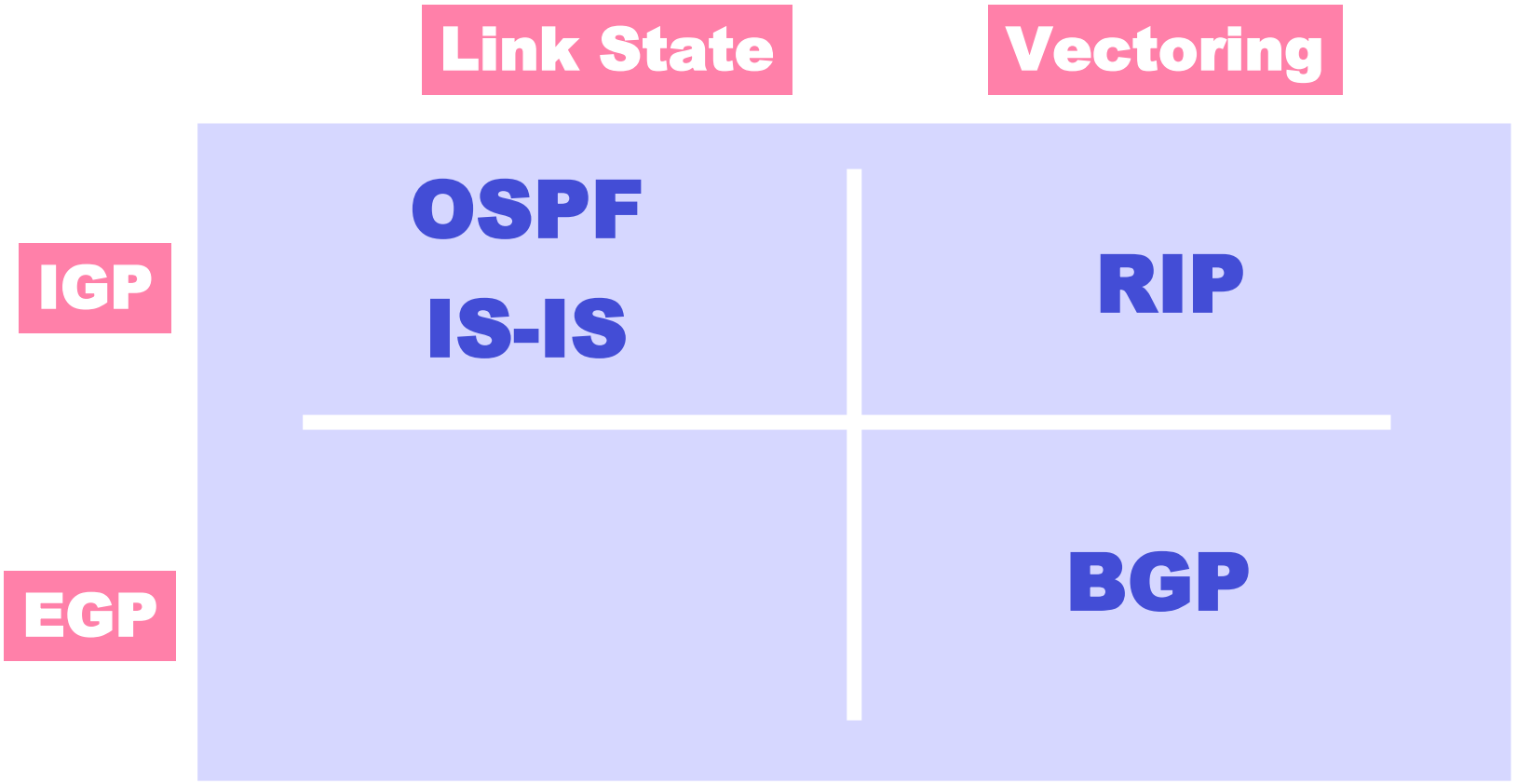
Link State

- Topology information is flooded within the routing domain
- Best end-to-end paths are computed locally at each router.
- Best end-to-end paths determine next-hops.
- Based on minimizing some notion of distance
- Works only if policy is shared and uniform
- Examples: OSPF, IS-IS

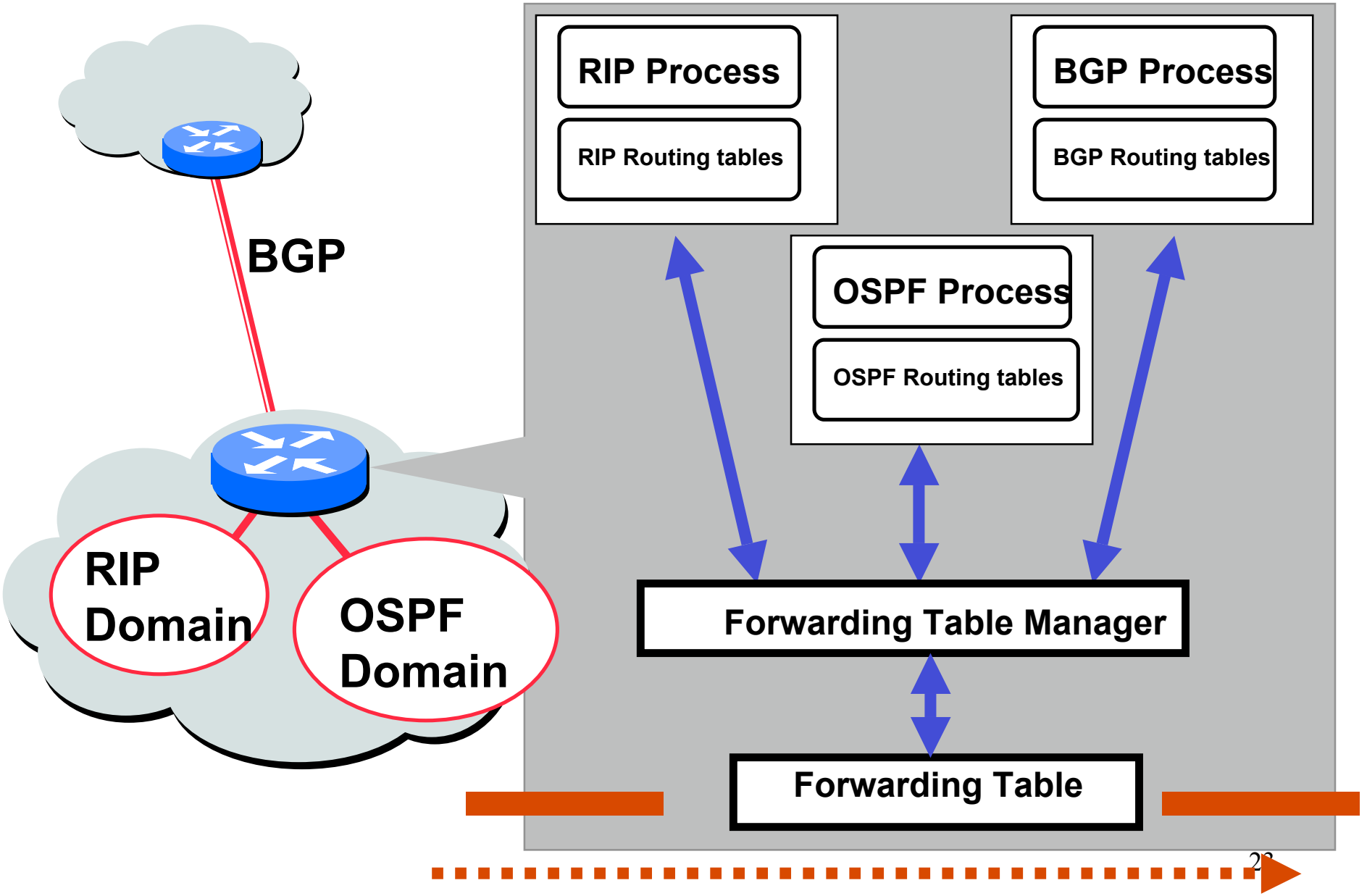
Vectoring

- Each router knows little about network topology
- Only best next-hops are chosen by each router for each destination network.
- Best end-to-end paths result from composition of all next-hop choices
- Does not require any notion of distance
- Does not require uniform policies at all routers
- Examples: RIP, BGP

The Gang of Four



Happy Packets: The Internet Does Not Exist Only to Populated Routing Tables



Before We Go Any Further ...

IP ROUTING PROTOCOLS DO NOT DYNAMICALLY ROUTE AROUND NETWORK CONGESTION

- **IP traffic can be very bursty**
- **Dynamic adjustments in routing typically operate more slowly than fluctuations in traffic load**
- **Dynamically adapting routing to account for traffic load can lead to wild, unstable oscillations of routing system**

Autonomous Routing Domains

A collection of physical networks glued together using IP, that have a unified administrative routing policy.

- **Campus networks**
- **Corporate networks**
- **ISP Internal networks**
- **…**

Autonomous Systems (ASes)

An autonomous system is an autonomous routing domain that has been assigned an Autonomous System Number (ASN).

… the administration of an AS appears to other ASes to have a single coherent interior routing plan and presents a consistent picture of what networks are reachable through it.

RFC 1930: Guidelines for creation, selection, and registration of an Autonomous System

AS Numbers (ASNs)

ASNs are 16 bit values (soon to be 32 bits)

64512 through 65535 are "private"

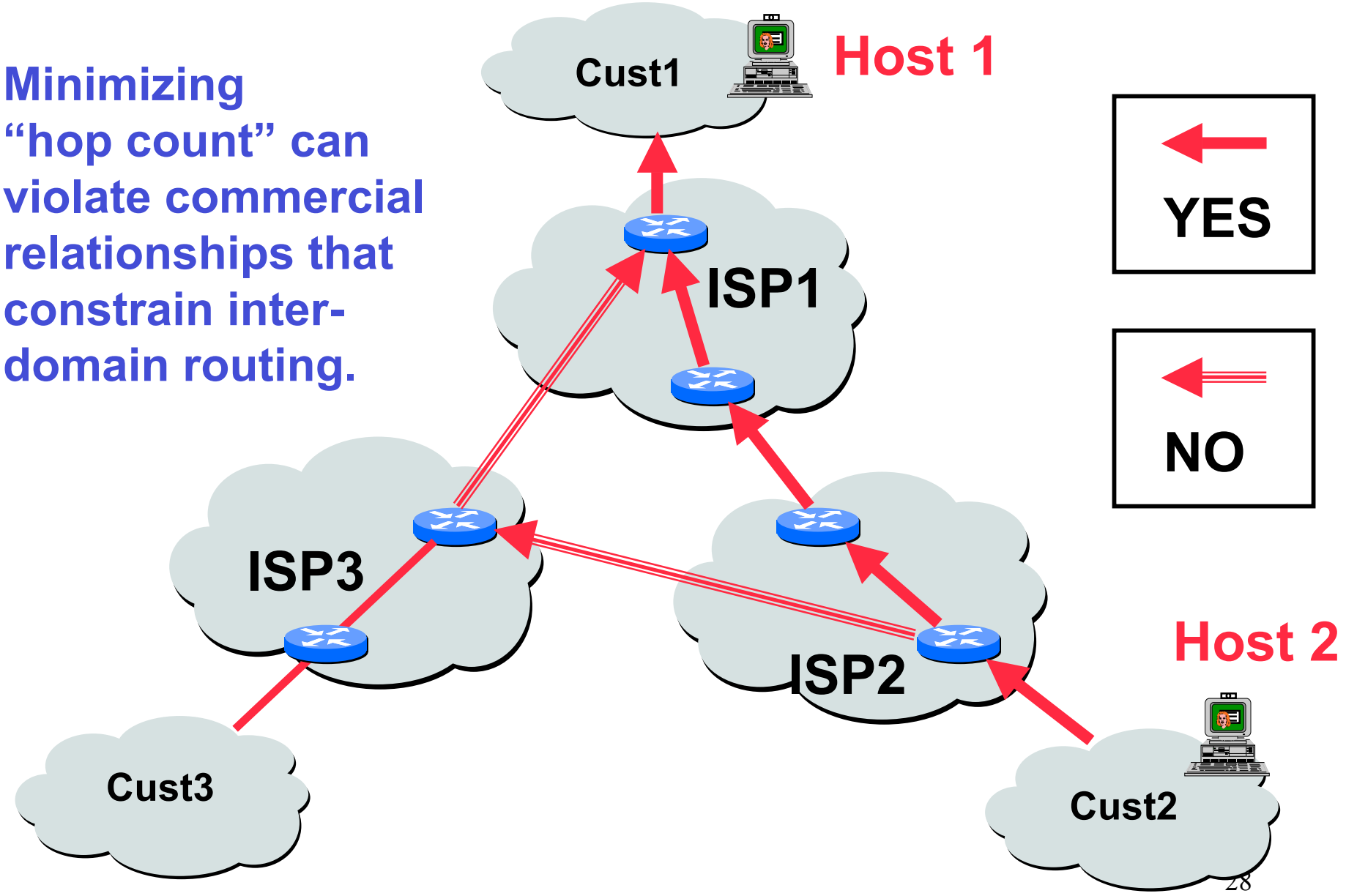
Currently nearly 30,000 in use.

- **JANET: 786**
- **MIT: 3**
- **Harvard: 11**
- **UC San Diego: 7377**
- **AT&T: 7018, 6341, 5074, ...**
- **UUNET: 701, 702, 284, 12199, ...**
- **Sprint: 1239, 1240, 6211, 6242, ...**
- **...**

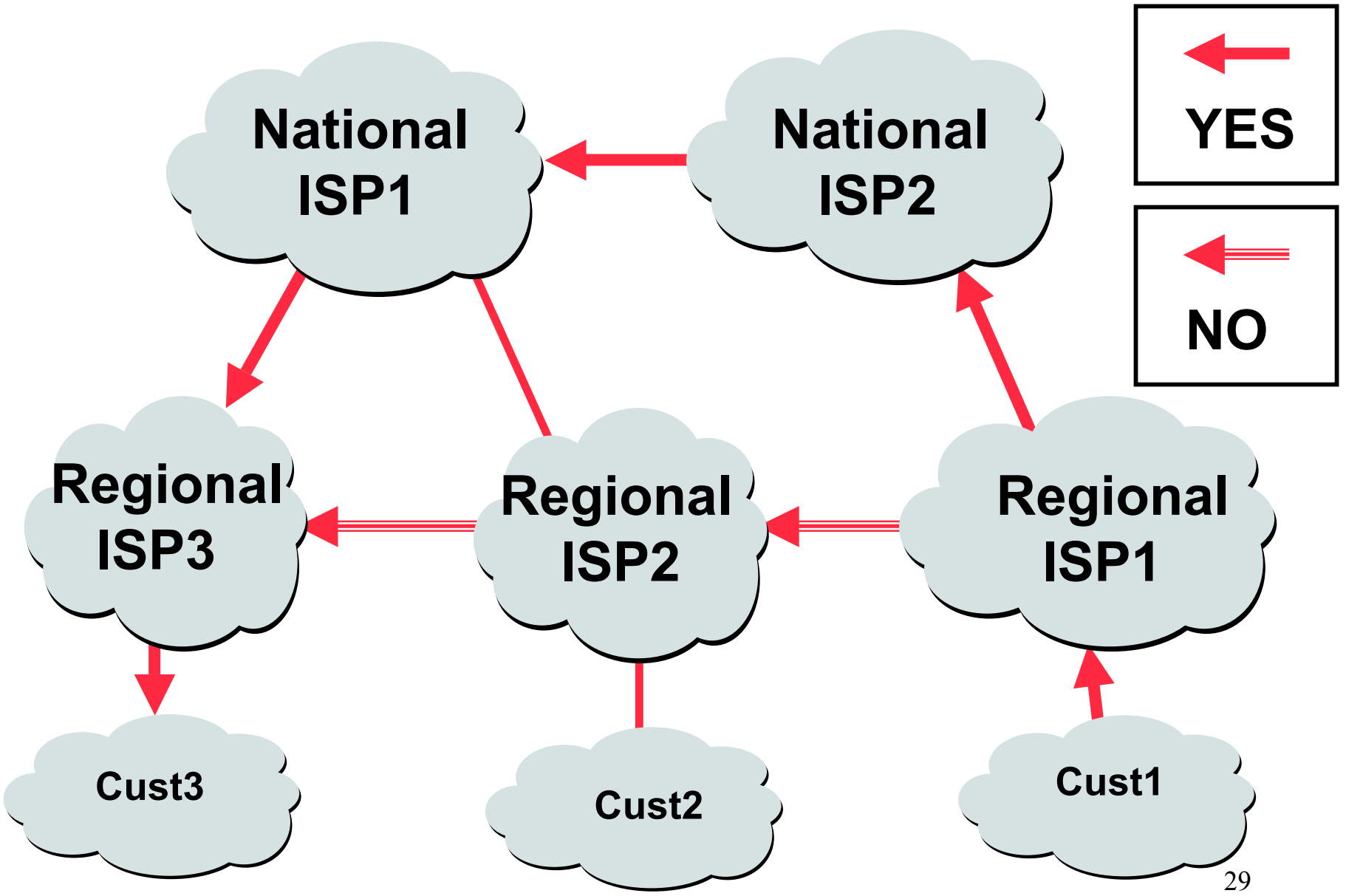
ASNs represent units of routing policy

Policy-Based vs. Distance-Based Routing?

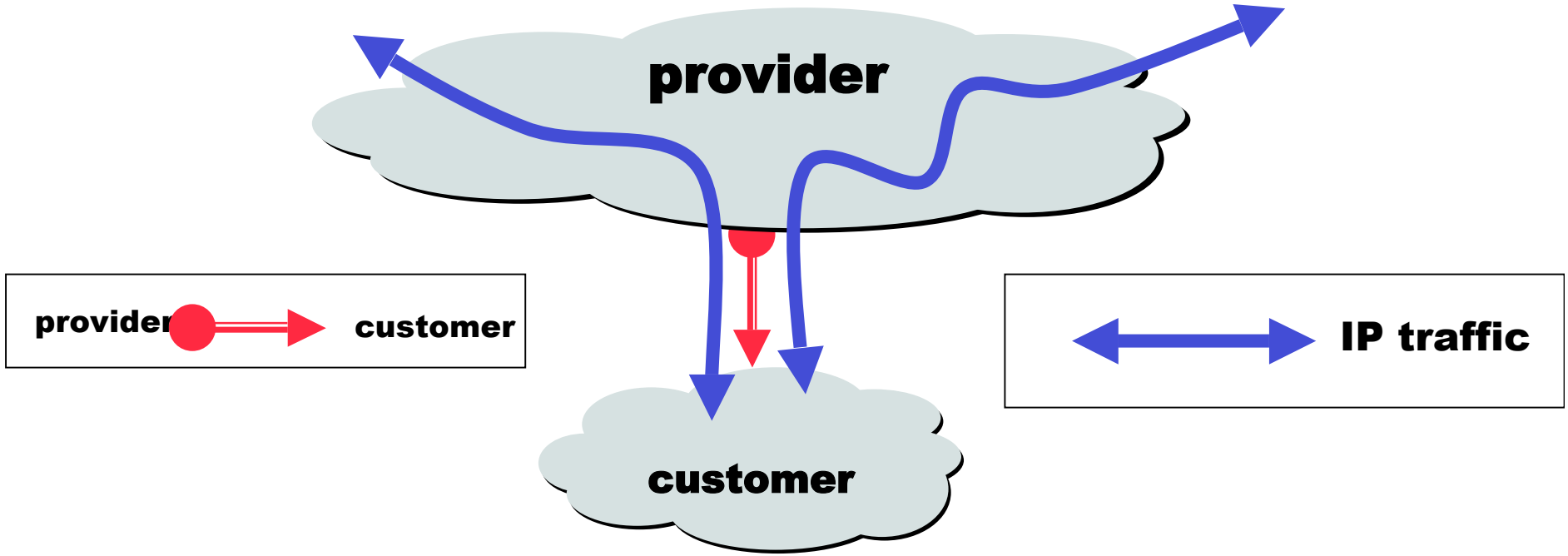
Minimizing "hop count" can violate commercial relationships that constrain inter-domain routing.



Why not minimize “AS hop count”?

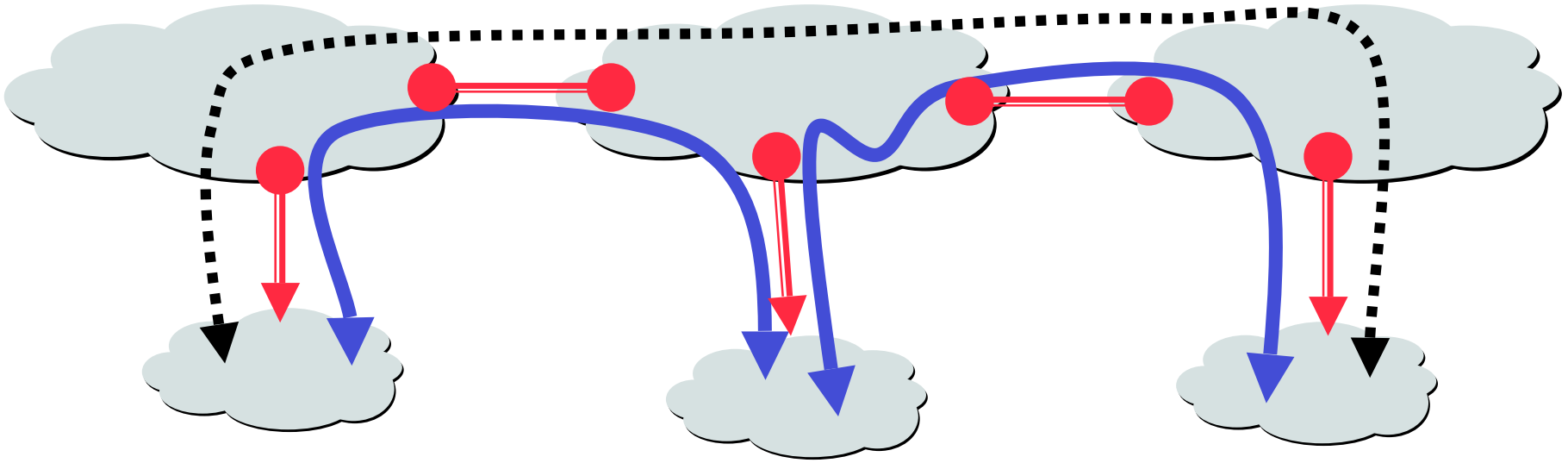


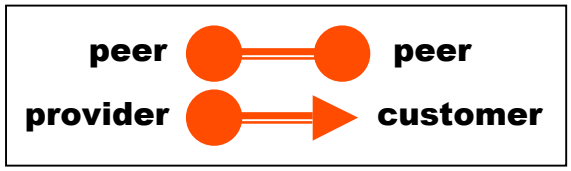
Customers and Providers



Customer pays provider for access to the Internet

The “Peering” Relationship





traffic allowed

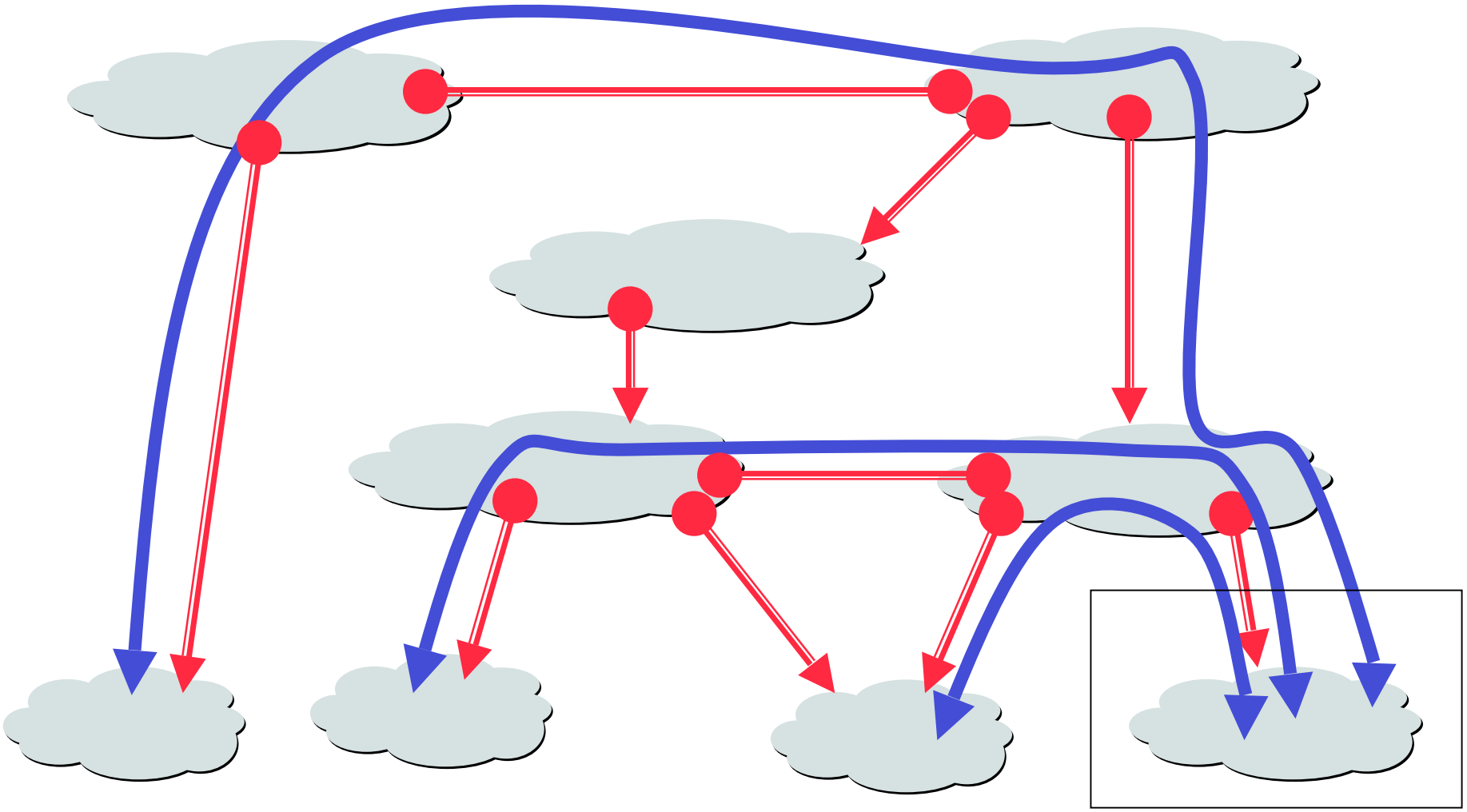
traffic NOT allowed

Peers provide transit between their respective customers

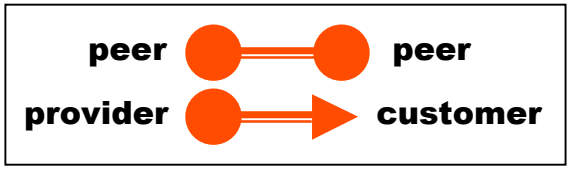
Peers do not provide transit between peers

Peers (often) do not exchange $$$

Peering Provides Shortcuts



Peering also allows connectivity between the customers of “Tier 1” providers.



Peering Wars

Peer

- Reduces upstream transit costs
- Can increase end-to-end performance
- May be the only way to connect your customers to some part of the Internet ("Tier 1")

Don't Peer

- You would rather have customers
- Peers are usually your competition
- Peering relationships may require periodic renegotiation

Peering struggles are by far the most contentious issues in the ISP world!

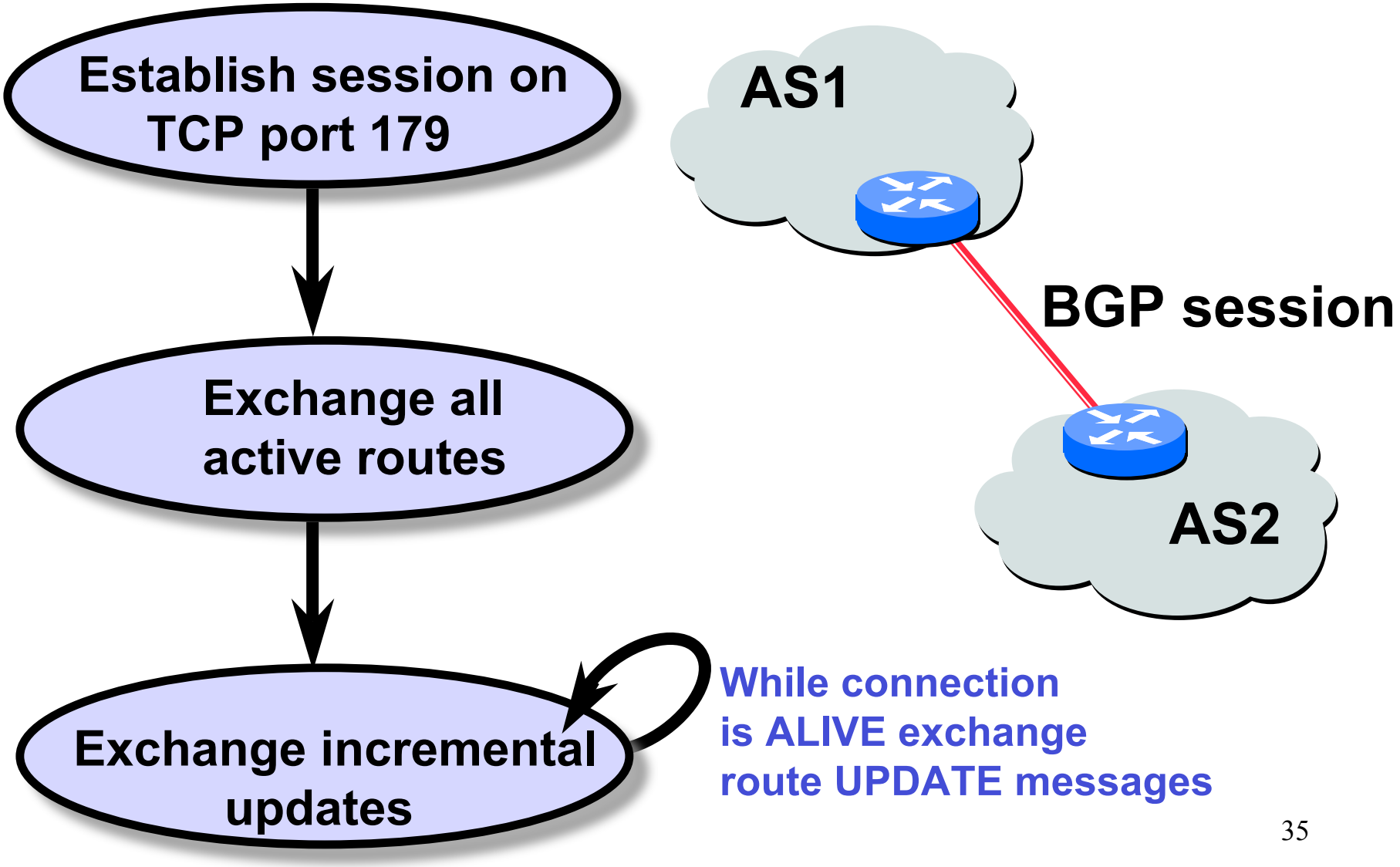
Peering agreements are often confidential.

BGP-4

- **BGP** = **B**order **G**ateway **P**rotocol
- Is a **Policy-Based** routing protocol
- Is the **de facto EGP** of today's global Internet
- Relatively simple protocol, but configuration is complex and the entire world can see, and be impacted by, your mistakes.

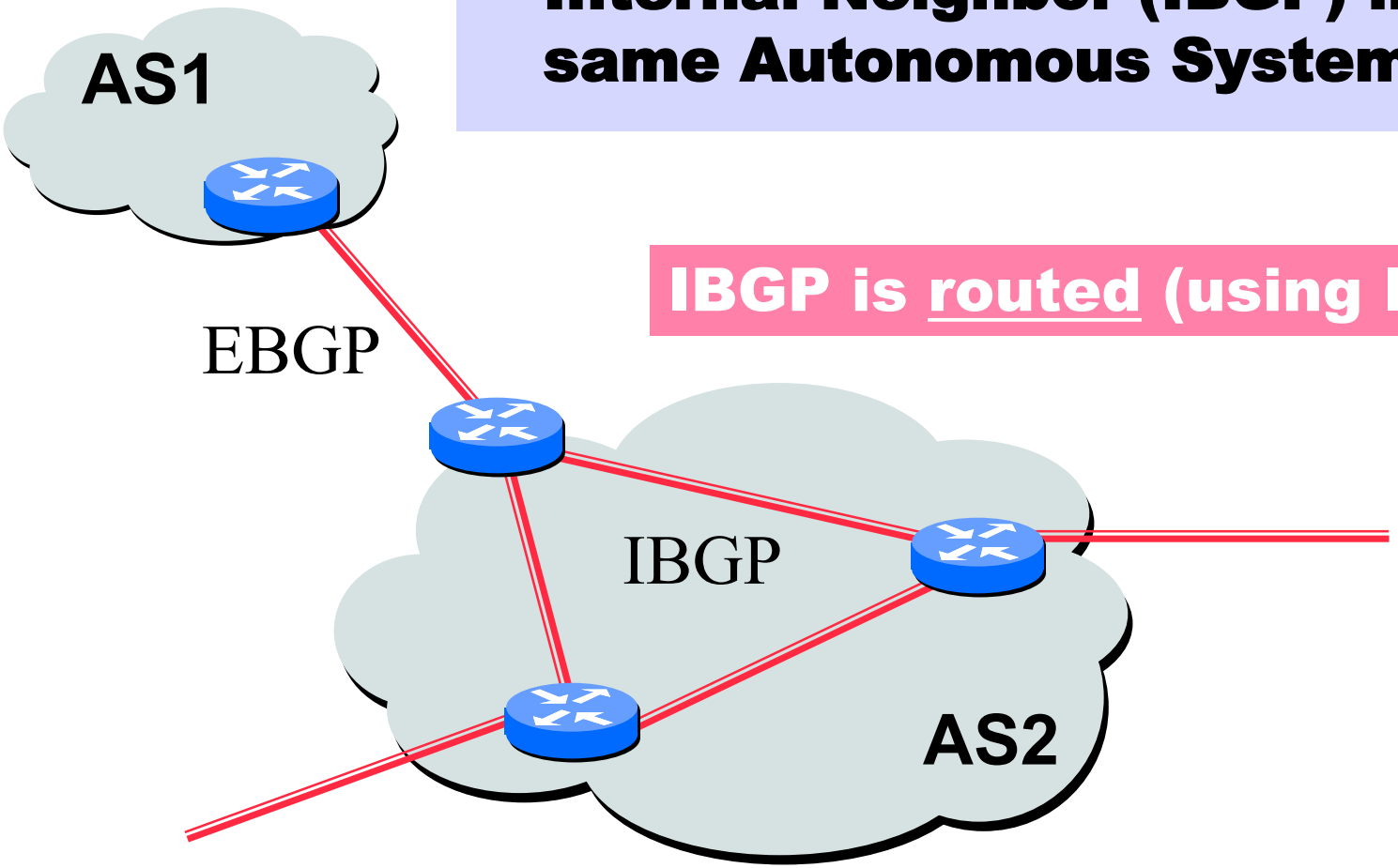
- **1989 : BGP-1 [RFC 1105]**
  - **Replacement for EGP (1984, RFC 904)**
- **1990 : BGP-2 [RFC 1163]**
- **1991 : BGP-3 [RFC 1267]**
- **1995 : BGP-4 [RFC 1771]**
  - **Support for Classless Interdomain Routing (CIDR)**
- **2006 : BGP-4 [RFC 4271]**

BGP Operations (Simplified)



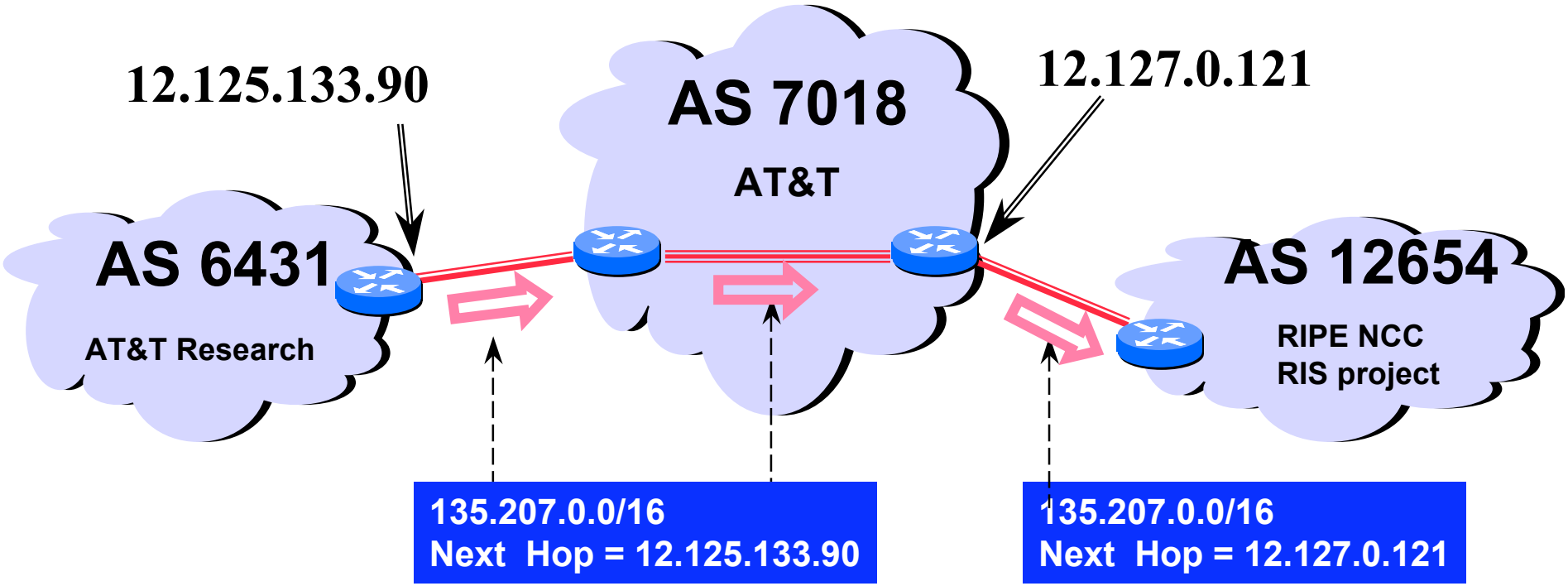
Two Types of BGP Sessions

- **External Neighbor (EBGP) in a different Autonomous Systems**
- **Internal Neighbor (IBGP) in the same Autonomous System**



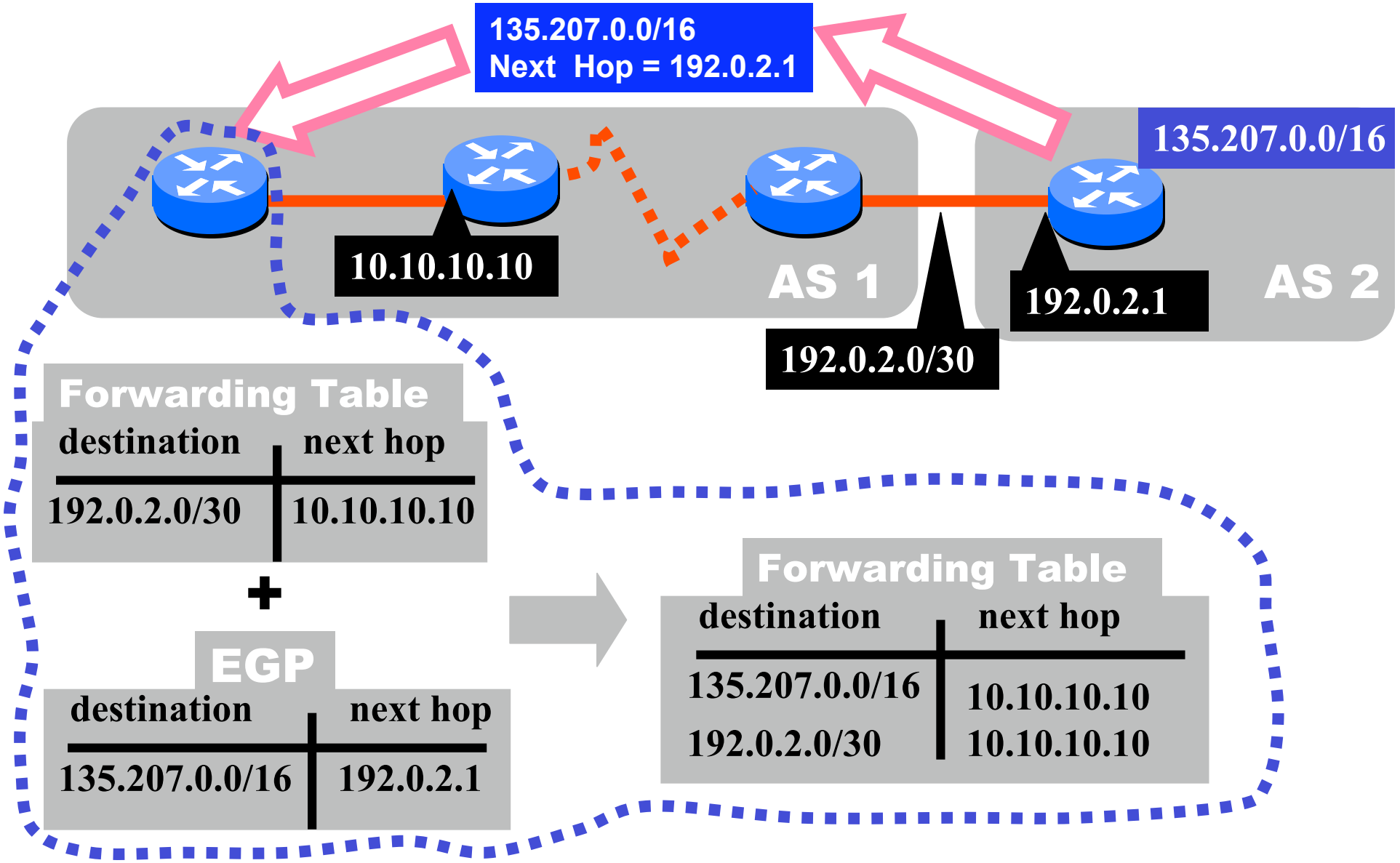
IBGP is routed (using IGP!)

BGP Next Hop Attribute



Every time a route announcement crosses an AS boundary, the Next Hop attribute is changed to the IP address of the border router that announced the route.

Join EGP with IGP For Connectivity



Four Types of BGP Messages

- **Open** : Establish a peering session.
- **Keep Alive** : Handshake at regular intervals.
- **Notification** : Shuts down a peering session.
- **Update** : Announcing new routes or withdrawing previously announced routes.

announcement
=
prefix + attributes values

BGP Attributes

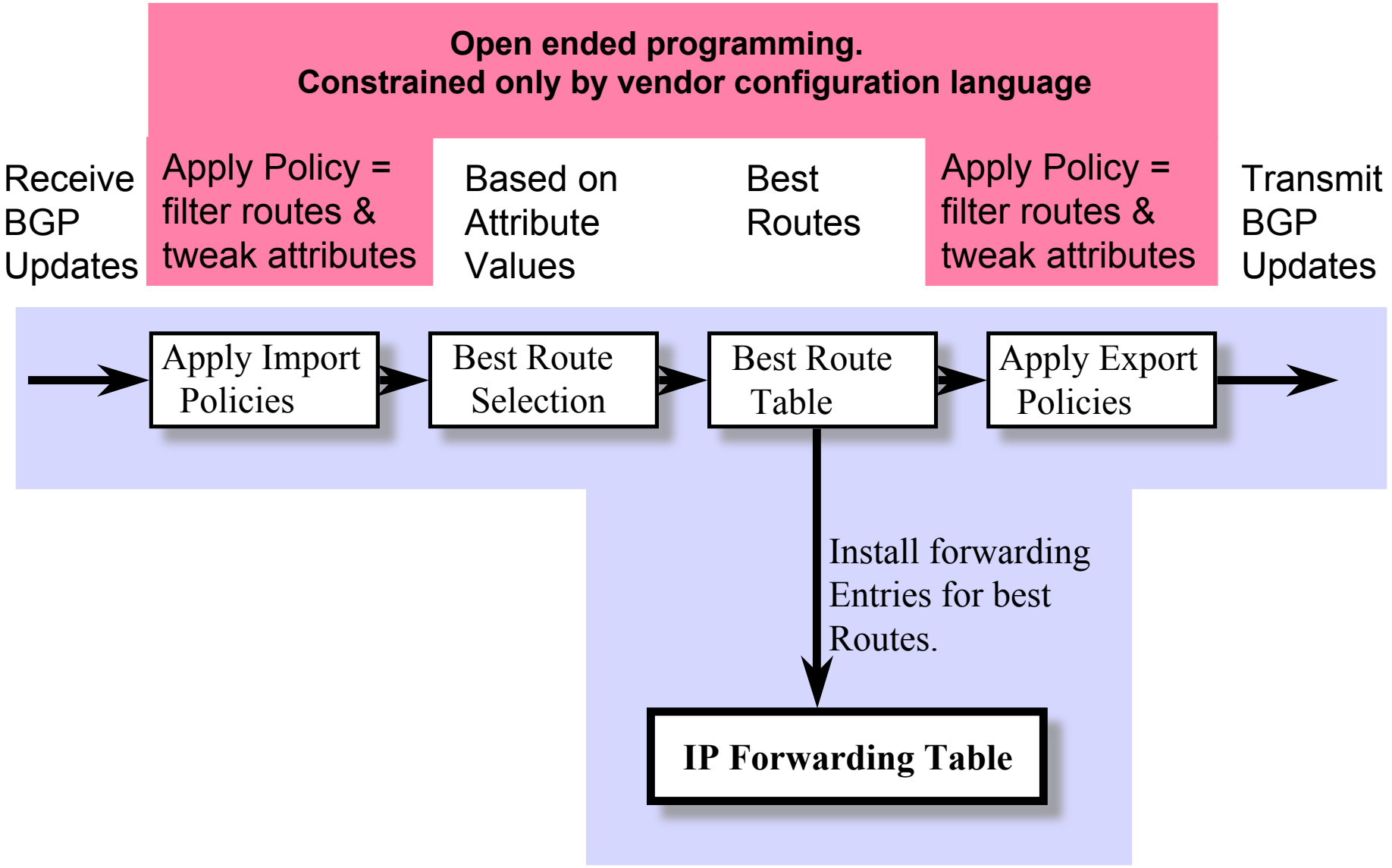
Value	Code	Reference
1	ORIGIN	[RFC1771]
2	AS_PATH	[RFC1771]
3	NEXT_HOP	[RFC1771]
4	MULTI_EXIT_DISC	[RFC1771]
5	LOCAL_PREF	[RFC1771]
6	ATOMIC_AGGREGATE	[RFC1771]
7	AGGREGATOR	[RFC1771]
8	COMMUNITY	[RFC1997]
9	ORIGINATOR_ID	[RFC2796]
10	CLUSTER_LIST	[RFC2796]
11	DPA	[Chen]
12	ADVERTISER	[RFC1863]
13	RCID_PATH / CLUSTER_ID	[RFC1863]
14	MP_REACH_NLRI	[RFC2283]
15	MP_UNREACH_NLRI	[RFC2283]
16	EXTENDED COMMUNITIES	[Rosen]
...		
255	reserved for development	

From IANA: http://www.iana.org/assignments/bgp-parameters

Most important attributes

Not all attributes need to be present in every announcement

BGP Route Processing



Route Selection Summary

Highest Local Preference — Enforce relationships

Shortest ASPATH

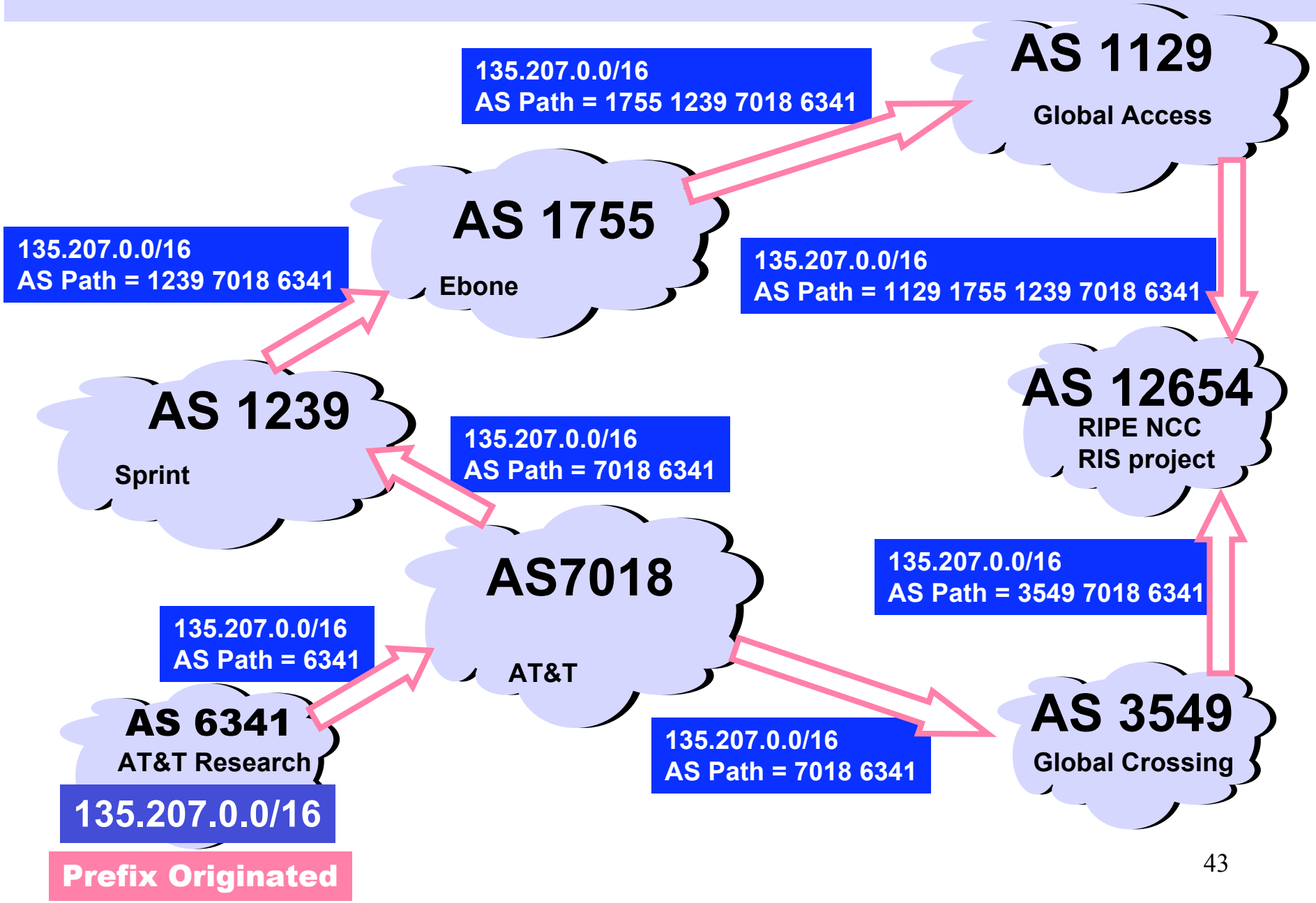
Lowest MED

i-BGP < e-BGP

Lowest IGP cost to BGP egress — traffic engineering

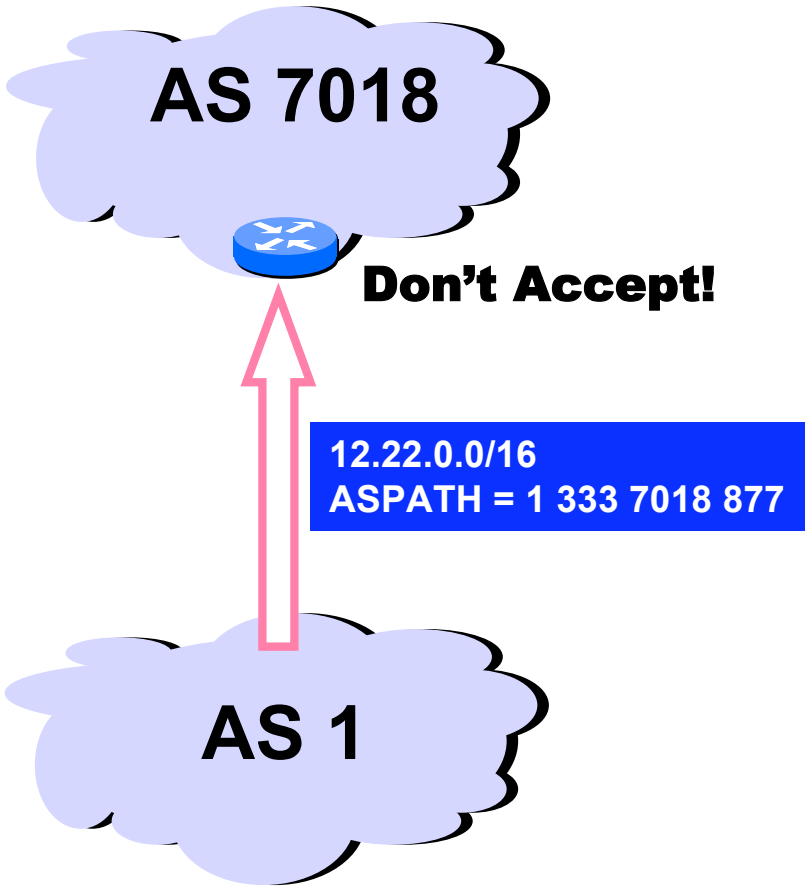
Lowest router ID — Throw up hands and break ties

ASPATH Attribute



Interdomain Loop Prevention

BGP at AS YYY will never accept a route with ASPATH containing YYY.



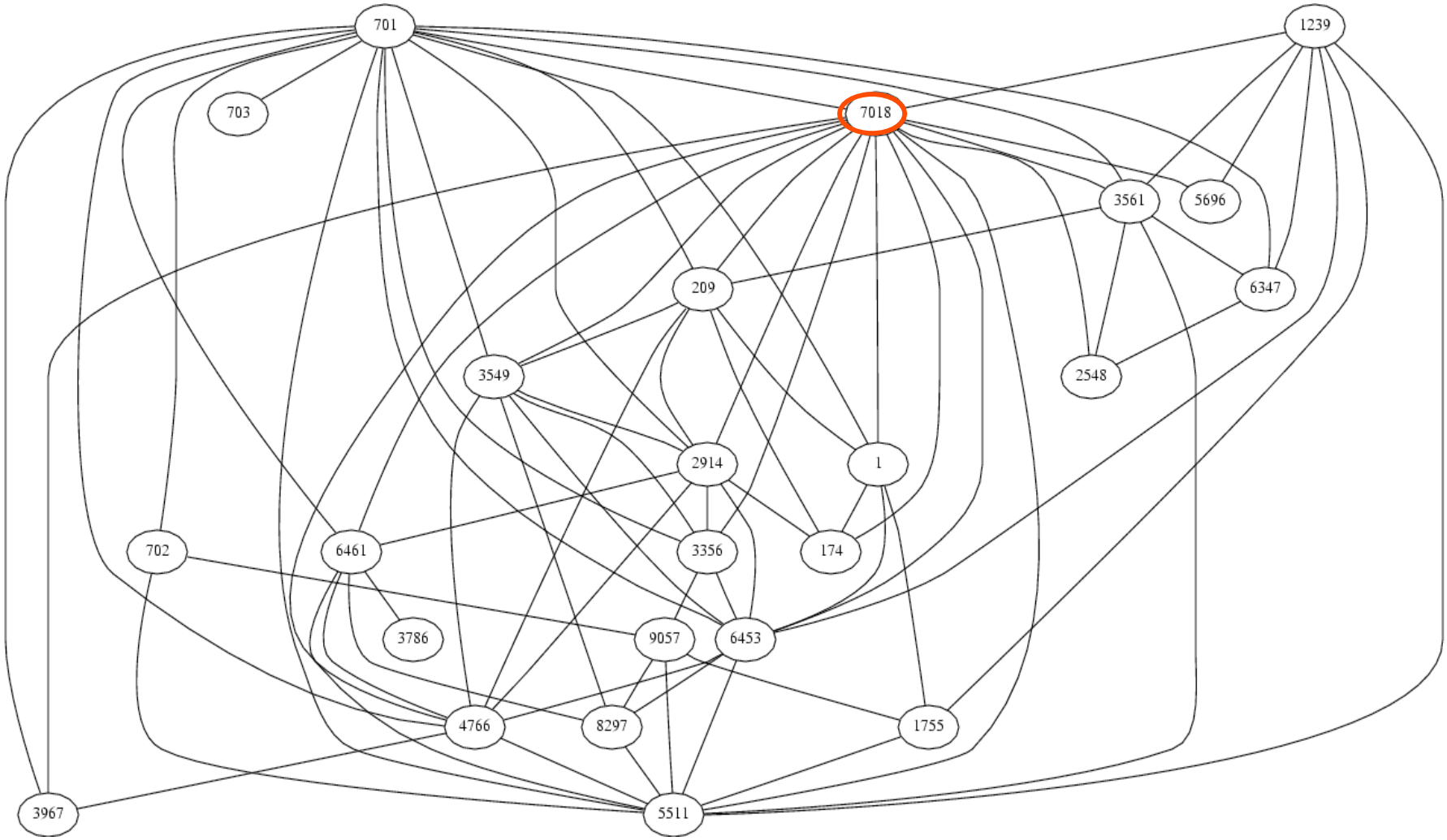
BGP Routing Tables

```
show ip bgp
BGP table version is 0, local router ID is 203.119.0.116
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, R Removed
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 0.0.0.0          193.0.4.28                             0 12654 34225 1299 i
*  3.0.0.0          193.0.4.28                             0 12654 7018 701 703 80 i
*>                  203.50.0.33                            0 65056 4637 703 80 i
*                   202.12.29.79                           0 4608 1221 4637 703 80 i
*  4.0.0.0          193.0.4.28                             0 12654 7018 3356 i
*>                  203.50.0.33                            0 65056 4637 3356 i
*                   202.12.29.79                           0 4608 1221 4637 3356 i
*  4.0.0.0/9        193.0.4.28                             0 12654 7018 3356 i
*>                  203.50.0.33                            0 65056 4637 3356 i
*                   202.12.29.79                           0 4608 1221 4637 3356 i
*  4.23.112.0/24    193.0.4.28                             0 12654 7018 174 21889 i
*>                  203.50.0.33                            0 65056 4637 174 21889 i
*                   202.12.29.79                           0 4608 1221 4637 174 21889 i
*  4.23.113.0/24    193.0.4.28                             0 12654 7018 174 21889 i
*>                  203.50.0.33                            0 65056 4637 174 21889 i
*                   202.12.29.79                           0 4608 1221 4637 174 21889 i
*  4.23.114.0/24    193.0.4.28                             0 12654 7018 174 21889 i
*>                  203.50.0.33                            0 65056 4637 174 21889 i
*                   202.12.29.79                           0 4608 1221 4637 174 21889 i
*  4.36.116.0/23    193.0.4.28                             0 12654 7018 174 21889 i
*>                  203.50.0.33                            0 65056 4637 174 21889 i
*                   202.12.29.79                           0 4608 1221 4637 174 21889 i
*  4.36.116.0/24    193.0.4.28                             0 12654 7018 174 21889 i
*>                  203.50.0.33                            0 65056 4637 174 21889 i
*                   202.12.29.79                           0 4608 1221 4637 174 21889 i
*  4.36.117.0/24    193.0.4.28                             0 12654 7018 174 21889 i
*>                  203.50.0.33                            0 65056 4637 174 21889 i
*                   202.12.29.79                           0 4608 1221 4637 174 21889 i
*  4.36.118.0/24    193.0.4.28                             0 12654 7018 174 21889 i
*>                  203.50.0.33                            0 65056 4637 174 21889 i
*                   202.12.29.79                           0 4608 1221 4637 174 21889 i
*> 4.78.22.0/23     193.0.4.28                             0 12654 3257 19151 13909 13909 13909 13909 13909 13909 13909 13909 13909 13909 13909 i
*                   203.50.0.33                            0 65056 4637 1299 1239 19151 13909 13909 13909 13909 13909 13909 13909 13909 13909 13909
*                   202.12.29.79                           0 4608 1221 4637 1299 1239 19151 13909 13909 13909 13909 13909 13909 13909 13909 13909 13
*> 4.78.56.0/23     193.0.4.28                             0 12654 3257 19151 13909 13909 13909 13909 13909 13909 13909 13909 13909 13909 13909 i
*                   203.50.0.33                            0 65056 4637 1299 1239 19151 13909 13909 13909 13909 13909 13909 13909 13909 13909 13909
*                   202.12.29.79                           0 4608 1221 4637 1299 1239 19151 13909 13909 13909 13909 13909 13909 13909 13909 13909 13
*  4.79.181.0/24    193.0.4.28                             0 12654 3741 10310 14780 i
*>                  203.50.0.33                            0 65056 4637 10310 14780 i
*                   202.12.29.79                           0 4608 1221 4637 10310 14780 i
```

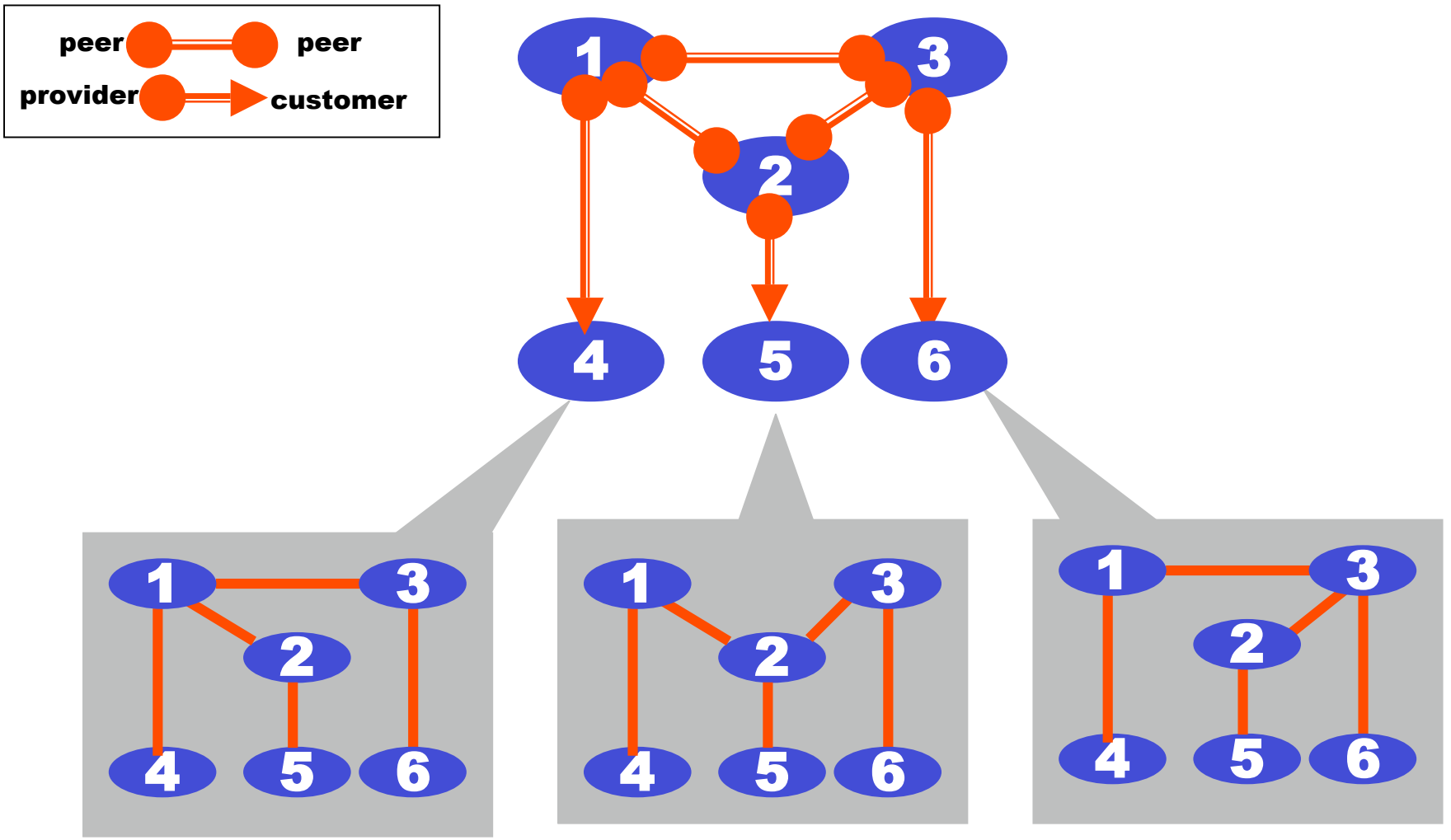
Thanks to Geoff Huston.
http://bgp.potaroo.net on Feb 1, 2008

AS Graphs Can Be Fun



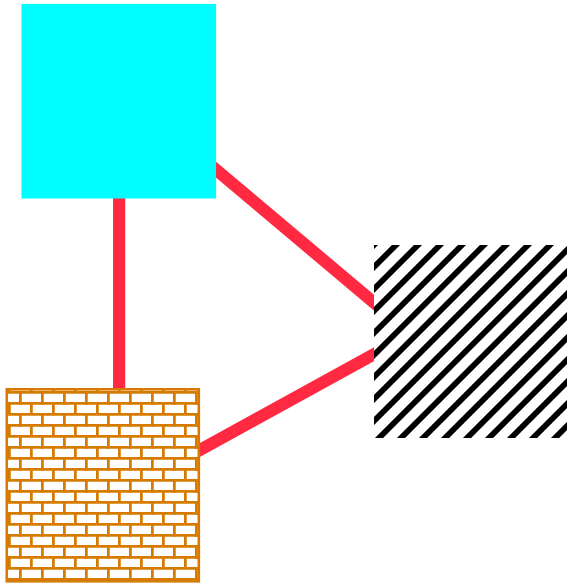
The subgraph showing all ASes that have more than 100 neighbors in full graph of 11,158 nodes. July 6, 2001. Point of view: AT&T route-server

AS Graphs Depend on Point of View

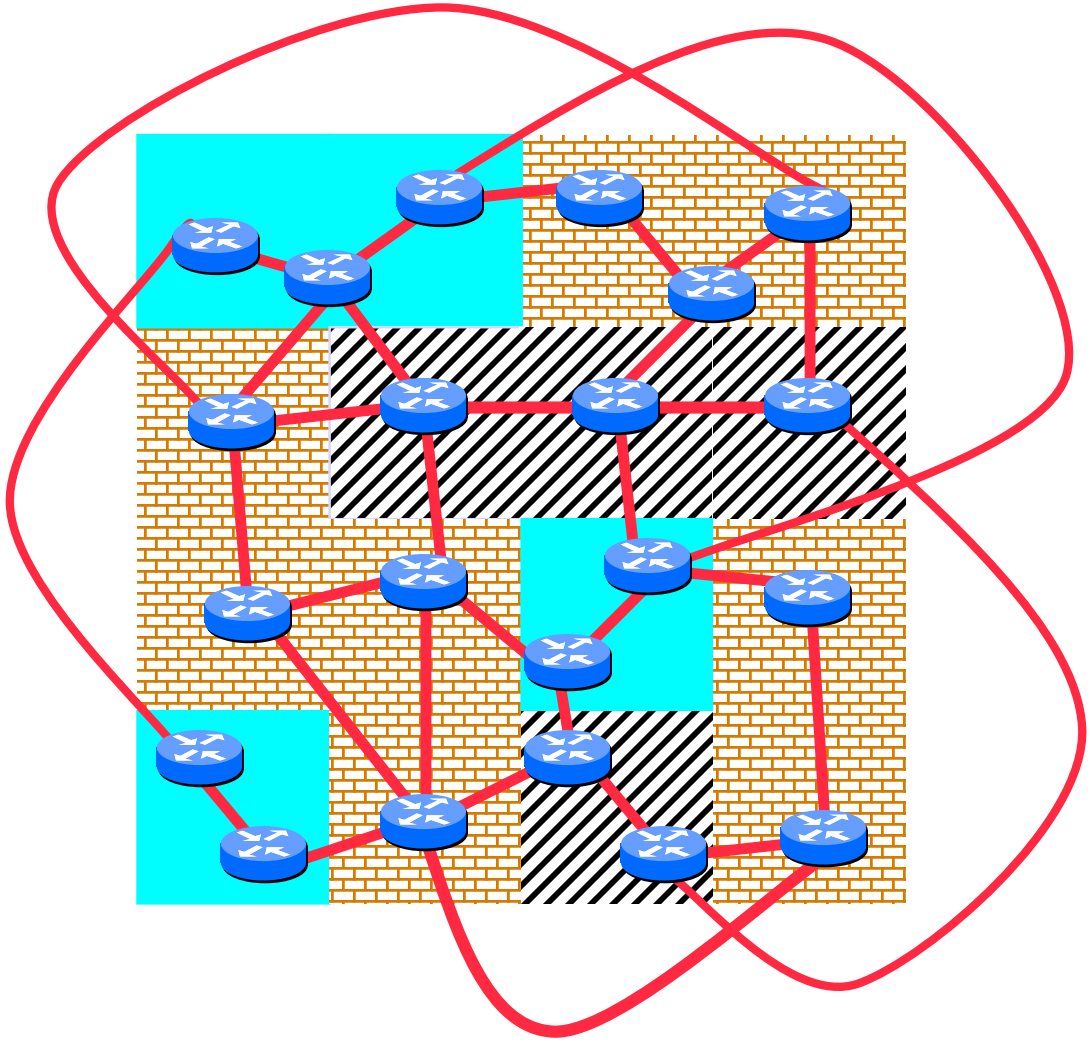


AS Graphs Do Not Show "Topology"!

BGP was designed to throw away information!



The AS graph may look like this.



Reality may be closer to this…