N E W C H A P T E R

Part III
«Where are we now?»

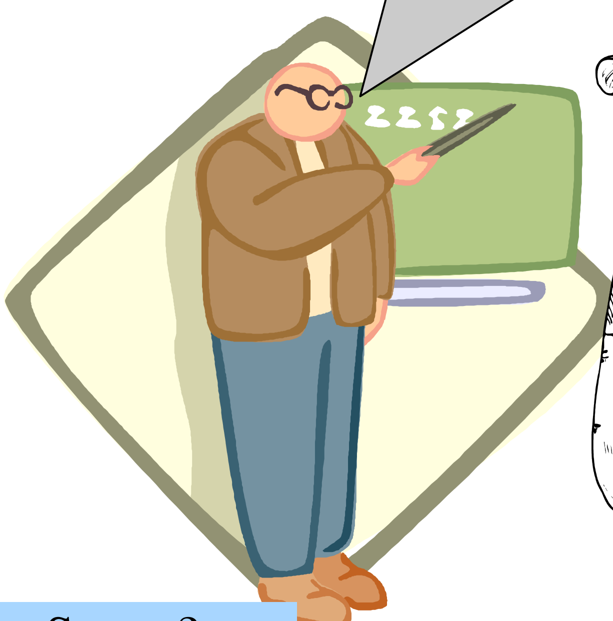
Status and deployment of multicast technologies

Status?

Academics vs Users

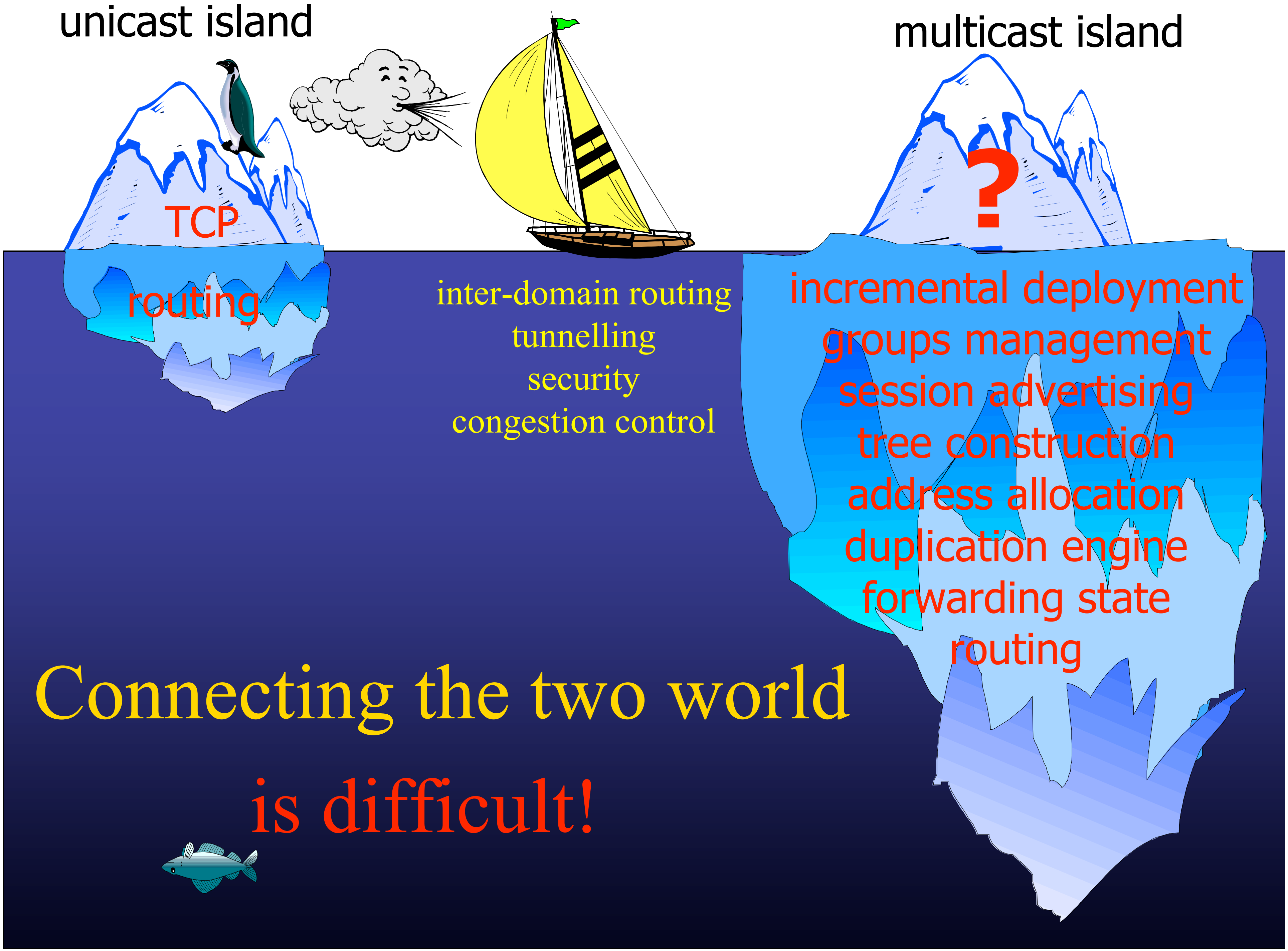
Multicast has been around for more than a decade, and we've proposed many protocols!

Yes, but very few real applications have been deployed on the Internet!

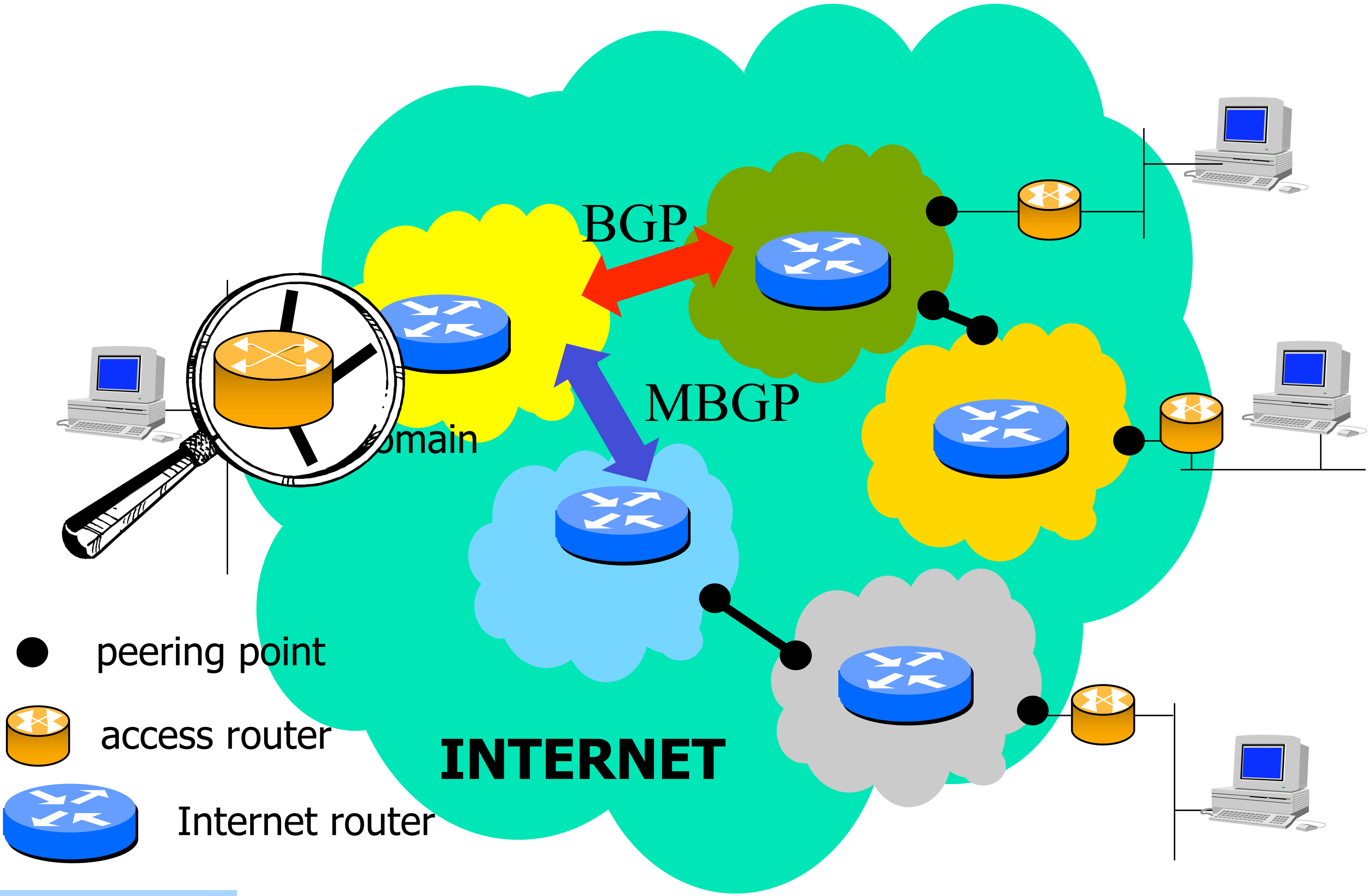


SRM, DVMRP
CBT, RMTP,
LMS, MOSPF,
MBGP, PIM-DM,
MSDP, IGMP,
RPM, HBH,
LBRM,
DyRAM...

Status?

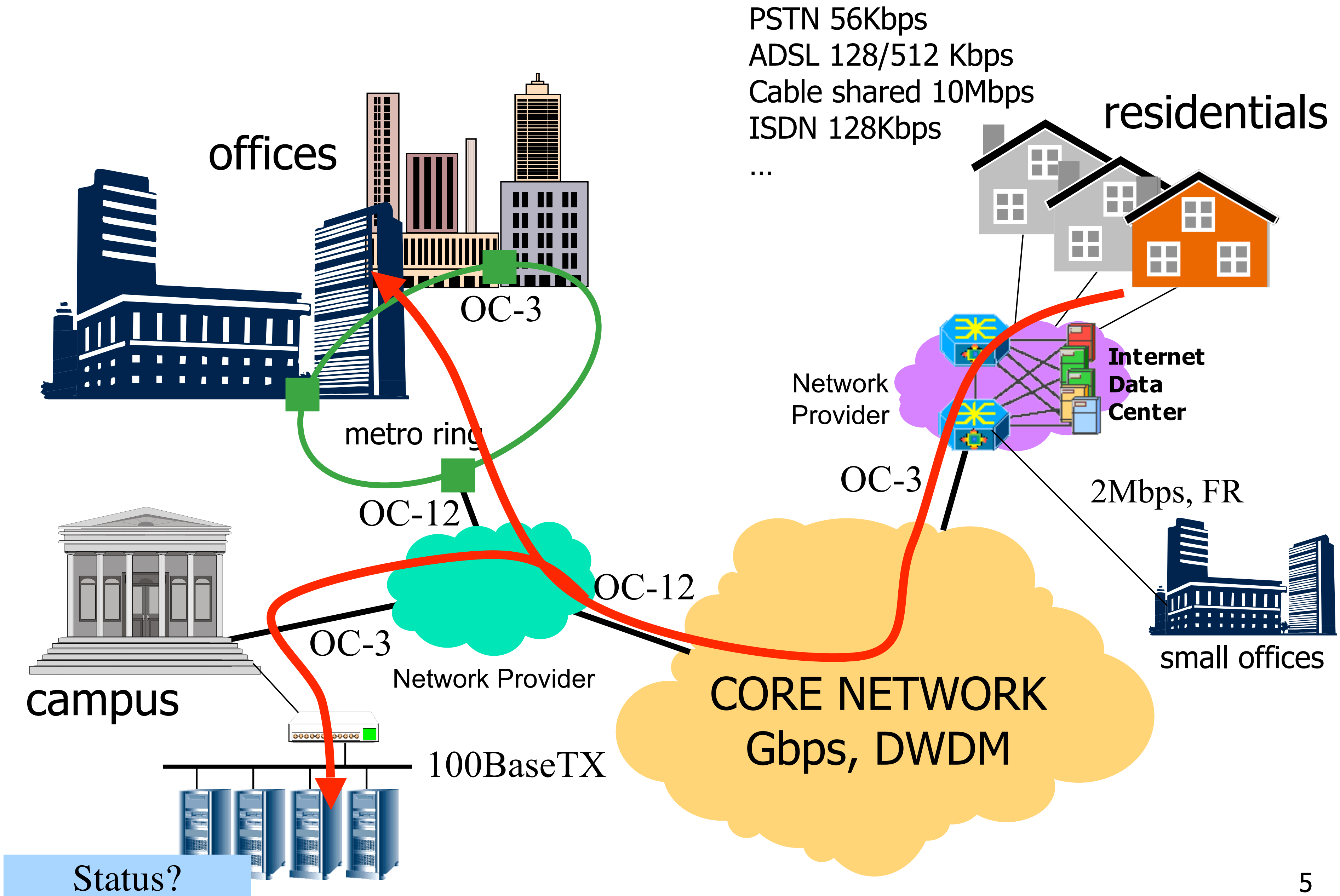

unicast island
TCP
routing
inter-domain routing
tunnelling
security
congestion control
multicast island
?
incremental deployment
groups management
session advertising
tree construction
address allocation
duplication engine
forwarding state
routing
Connecting the two world
is difficult!

Inter-domain agreement



Status?

Users' accesses



Links heterogeneity

- Backbone links
  - optical fibers
  - 2.5 to 160 Gbps with DWDM techniques
- End-user access
  - 9.6Kbps (GSM) to 2Mbps (UMTS) V.90 56Kbps modem on twisted pair
  - 64Kbps to 1930Kbps ISDN access
  - 128Kbps to 2Mbps with xDSL modem
  - 1Mbps to 10Mbps Cable-modem
  - 155Mbps to 2.5Gbps SONET/SDH

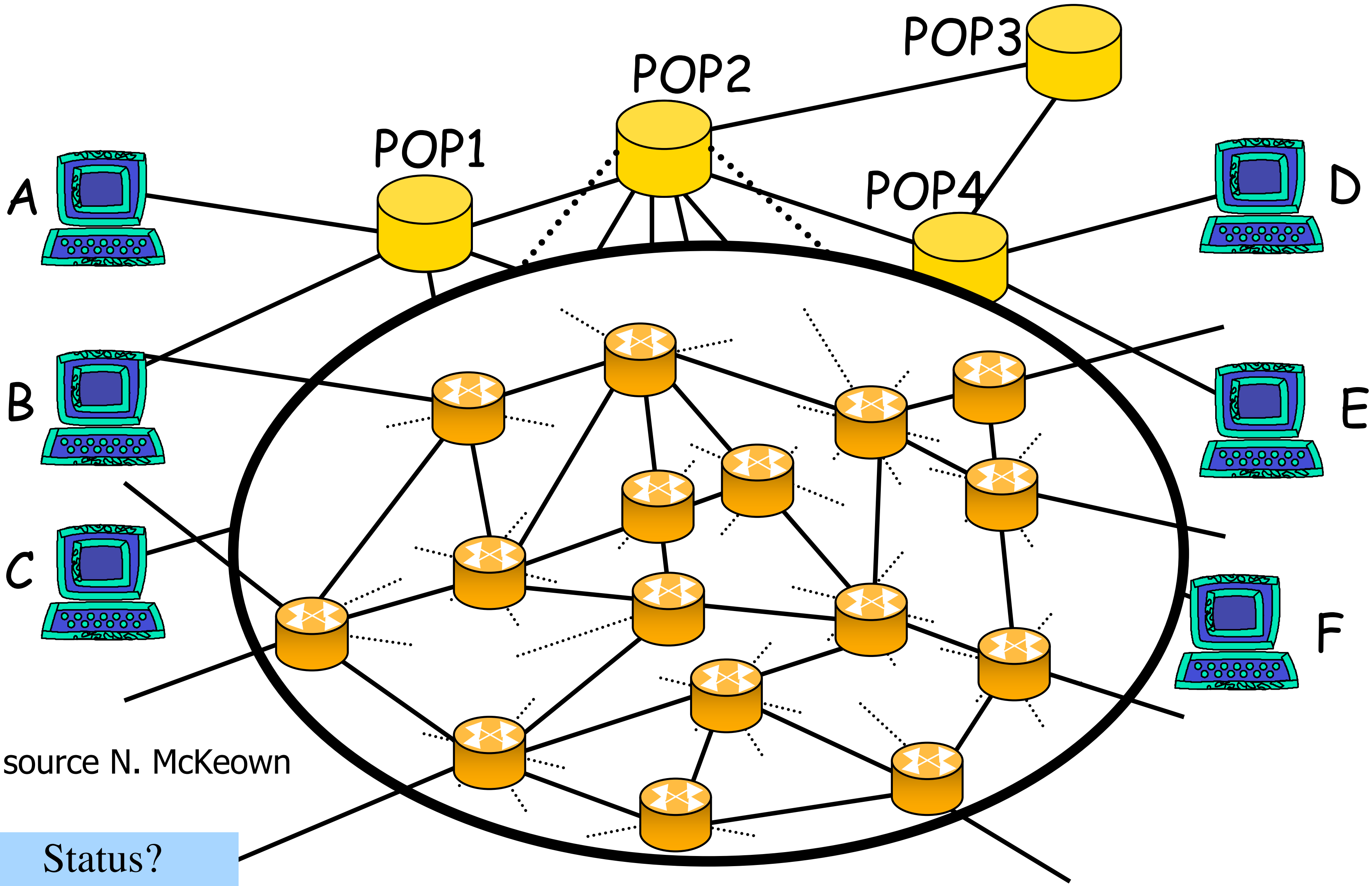
Status?

Internet routers: key elements of internetworking

- Routers
  - run routing protocols and build routing table,
  - receive data packets and perform relaying,
  - may have to consider Quality of Service constraints for scheduling packets,
  - are highly optimized for packet forwarding functions.

Status?

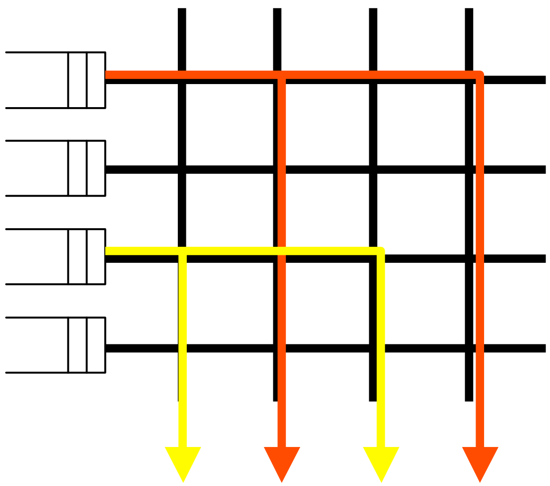
Multicast in Points of Presence



source N. McKeown

Status?

Multicast, a threat for high-performance routers!



Status?

The ~~open~~ model no-security

CONTRACT

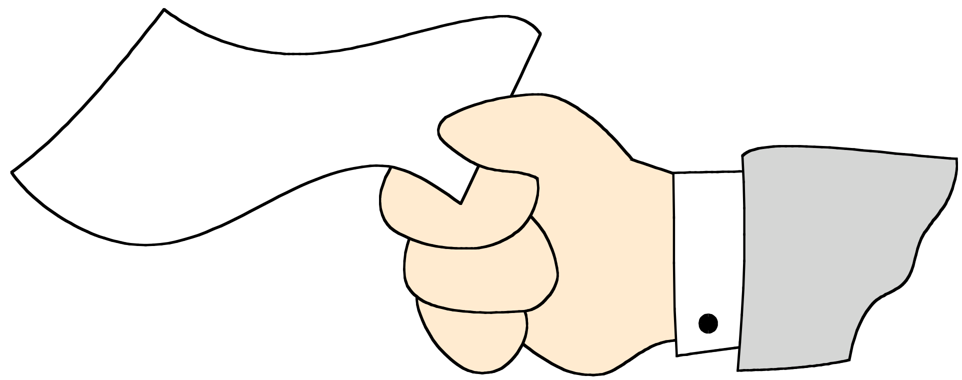
Can not control sources

Can not control receivers

Can not control groups

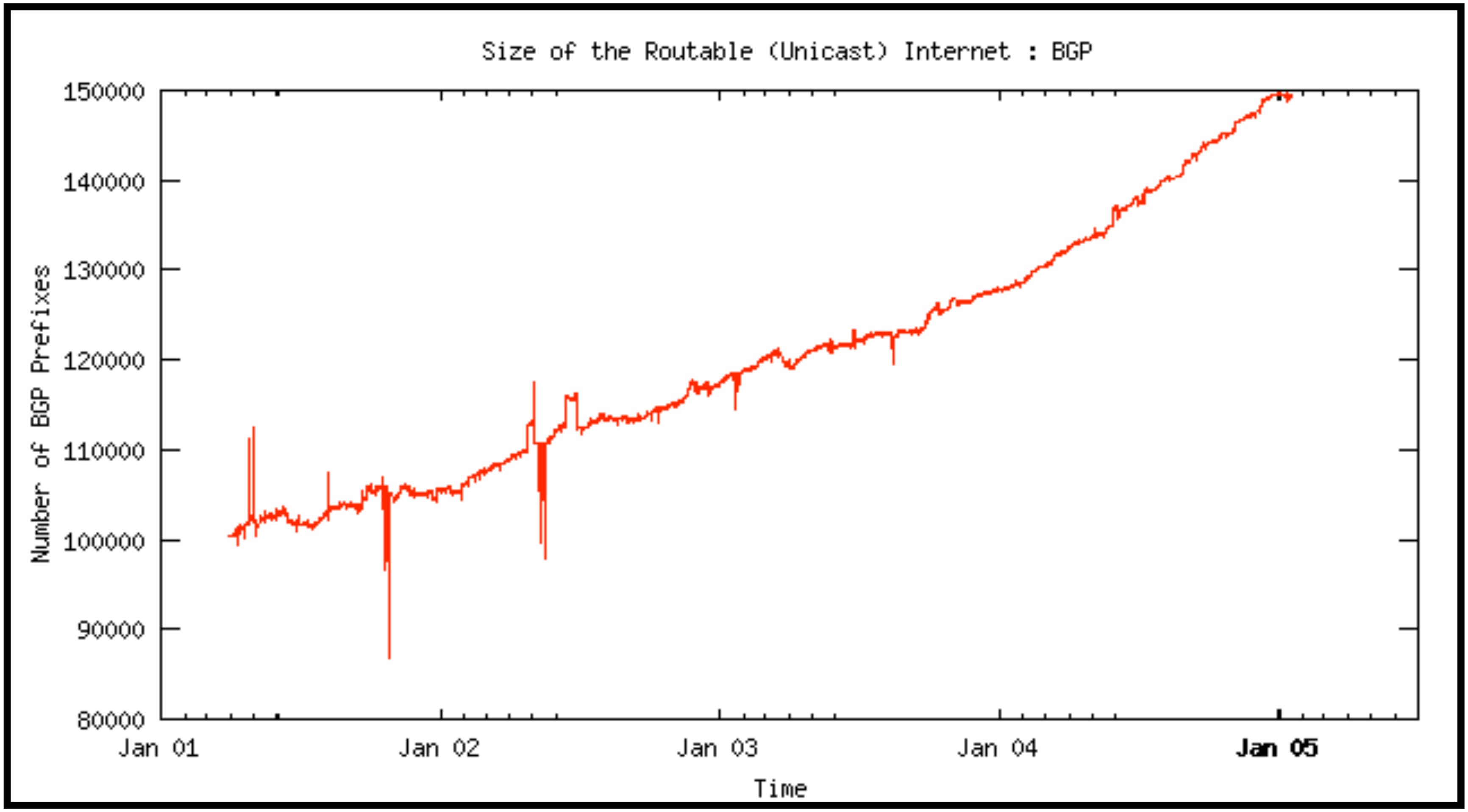
Can not control traffic

Please sign



Status?

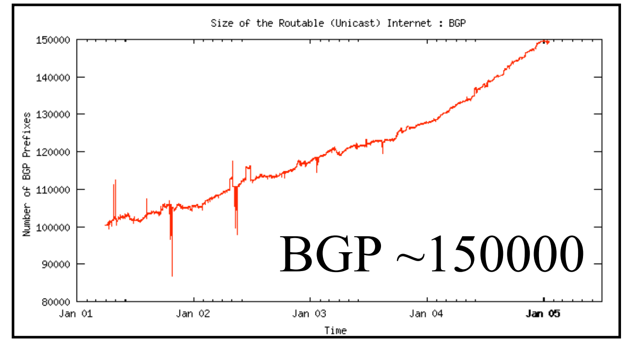
BGP table size

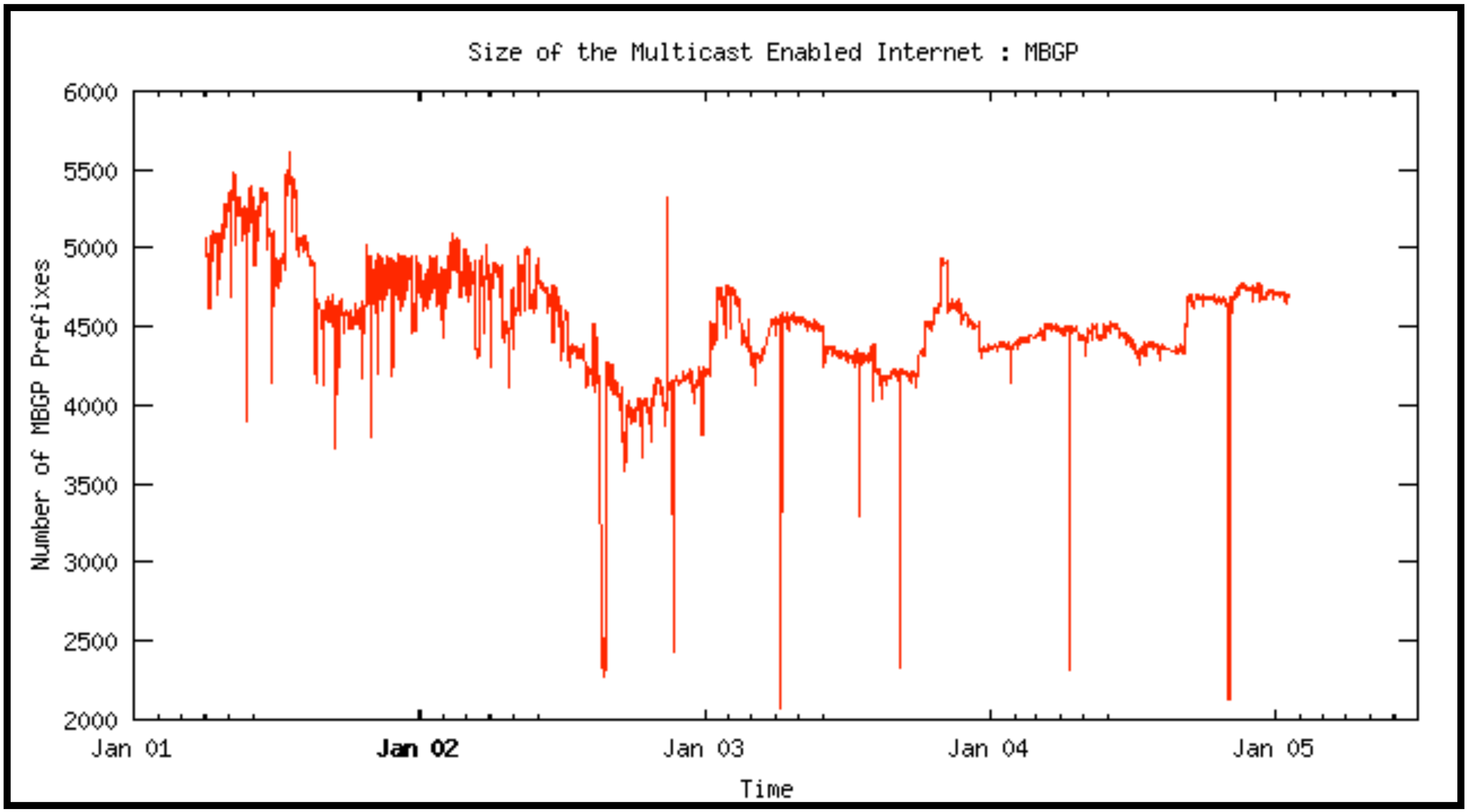


source www.multicasttech.com/status

Status?

MBGP table size

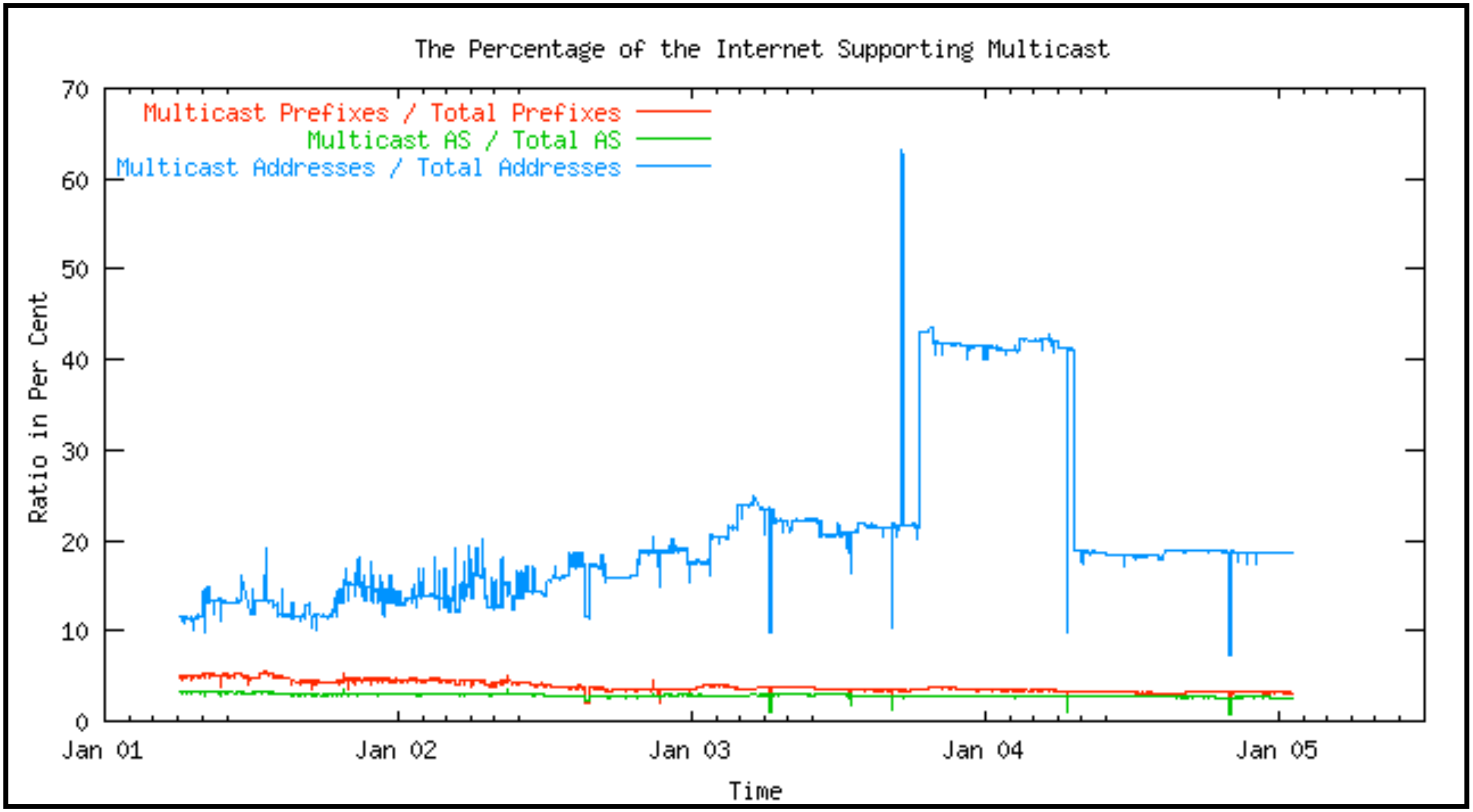




Status?

source www.multicasttech.com/status

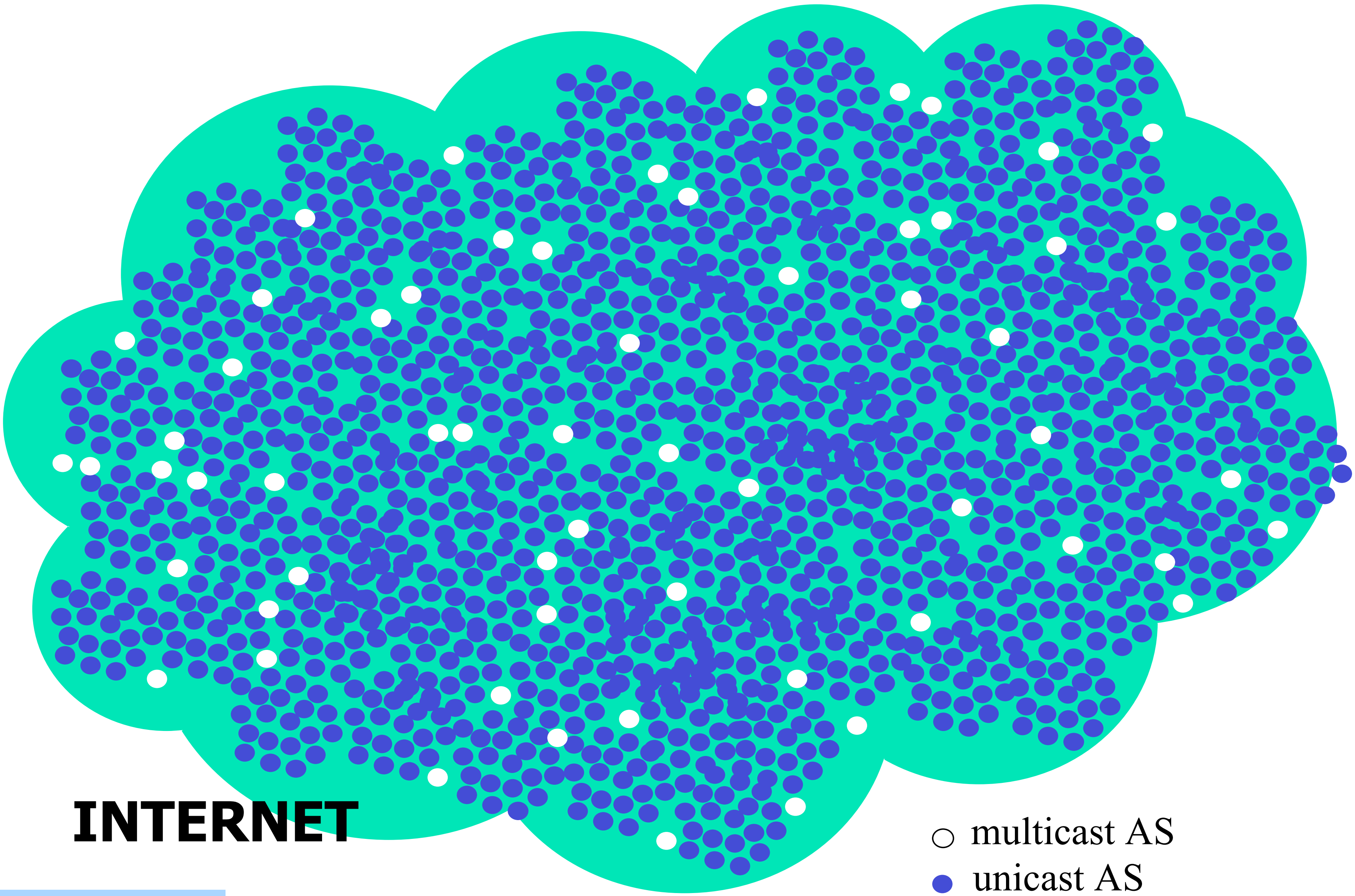
Relative Size of the Multicast Enabled Internet



source www.multicasttech.com/status

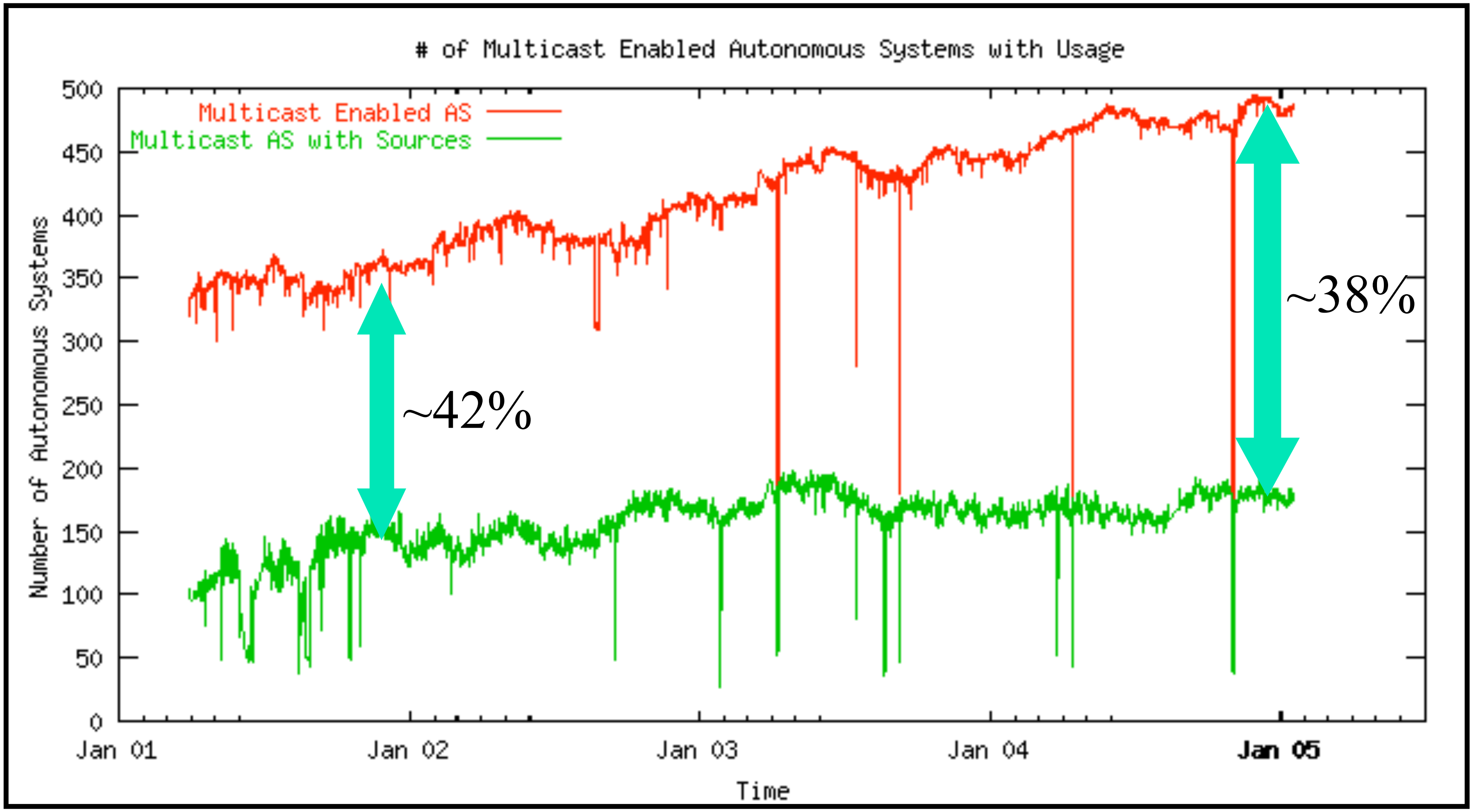
Status?

The gap in images



Status?

Autonomous Systems in the Multicast Enabled Internet: Totals and Those With Active Sources



source www.multicasttech.com/status

Status?

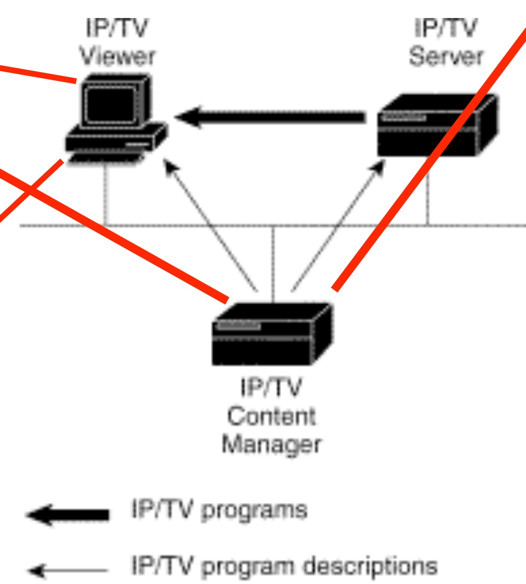
Selection of other commercial/prototype products

- CISCO IP/TV, CISCO IP/VC
- XtremeCast from mPulse
- Digital Fountain
- Multicast Monitor
- much more
  - RendezVous, Freephone,
  - MASH, CMT, MultiMon, NTE
  - MPOLL, MLC, MFTP

Status?

CISCO IP/TV,

- Usages
  - Training, Busines
    Corporate Comm
    Learning, Videoc



Status?

XtremeCast from mPulse

- Usage
  - Used by financial firms for stock quotes broadcasting
  - Chat server
- Reliable multicast implementation with the JRMS library (©SUN)
- http://www.mpulsetech.com/prod/xcast.htm

Status?

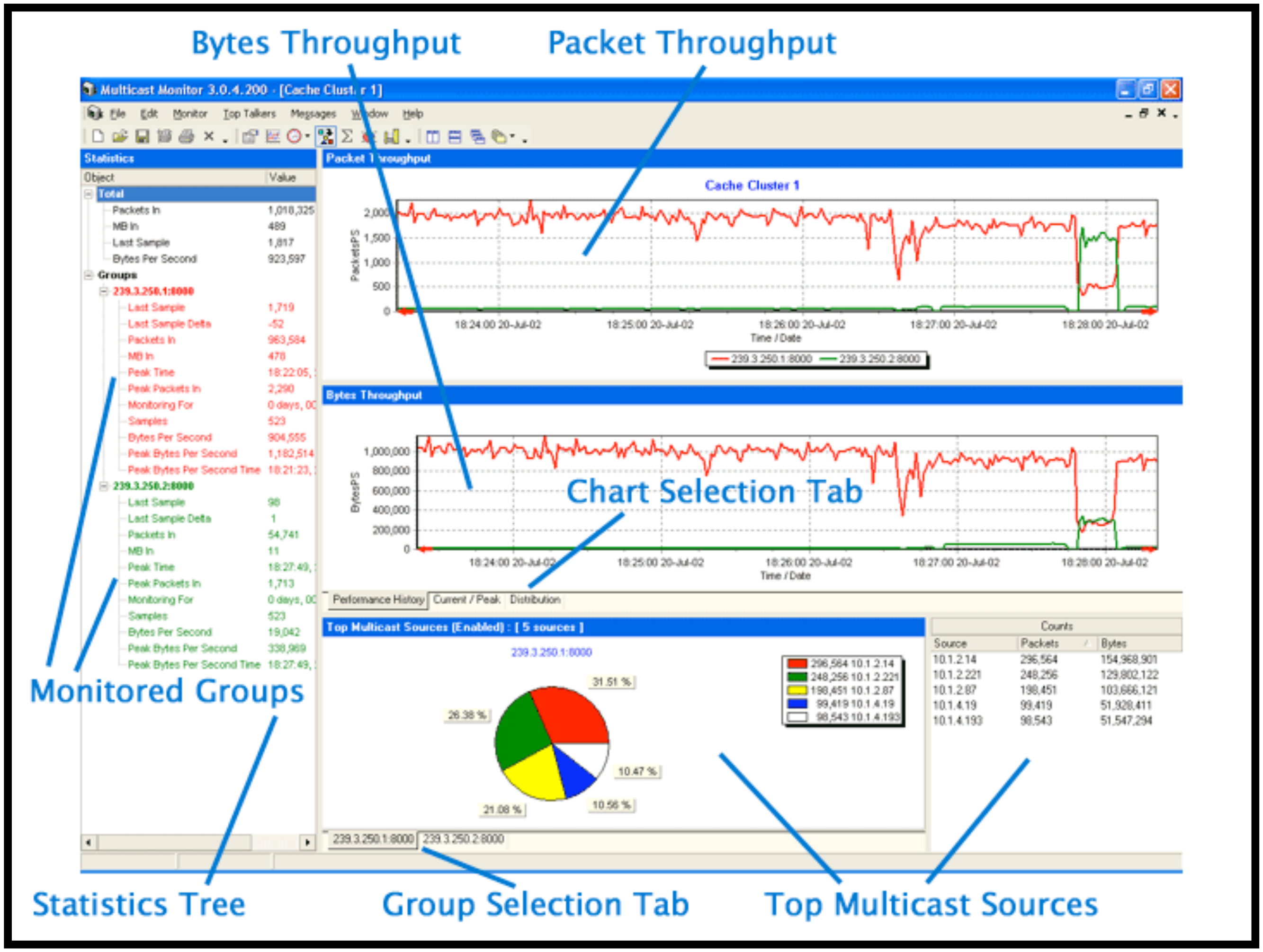
Digital Fountain products

- Implement ALC/LCT/WEBRC and rely on two highly efficient large block FEC codecs
  - http://www.digitalfountain.com
  - high implication in the IETF RMT standardization process

Status?

Multicast Monitor

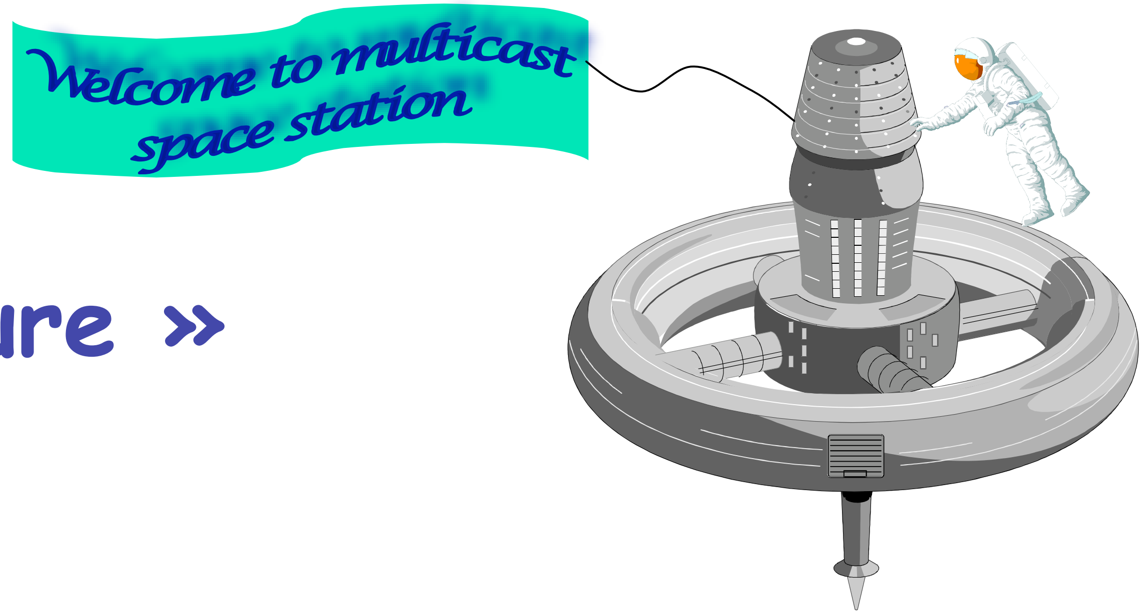
monitor multicast traffic in the entreprise network



Status?

N E W C H A P T E R

Part IV
« The Future »



BGMP & MASC

IPv6

Multicast and IP-MPLS networks

Multicast and Overlays networks

Future of inter-domain routing

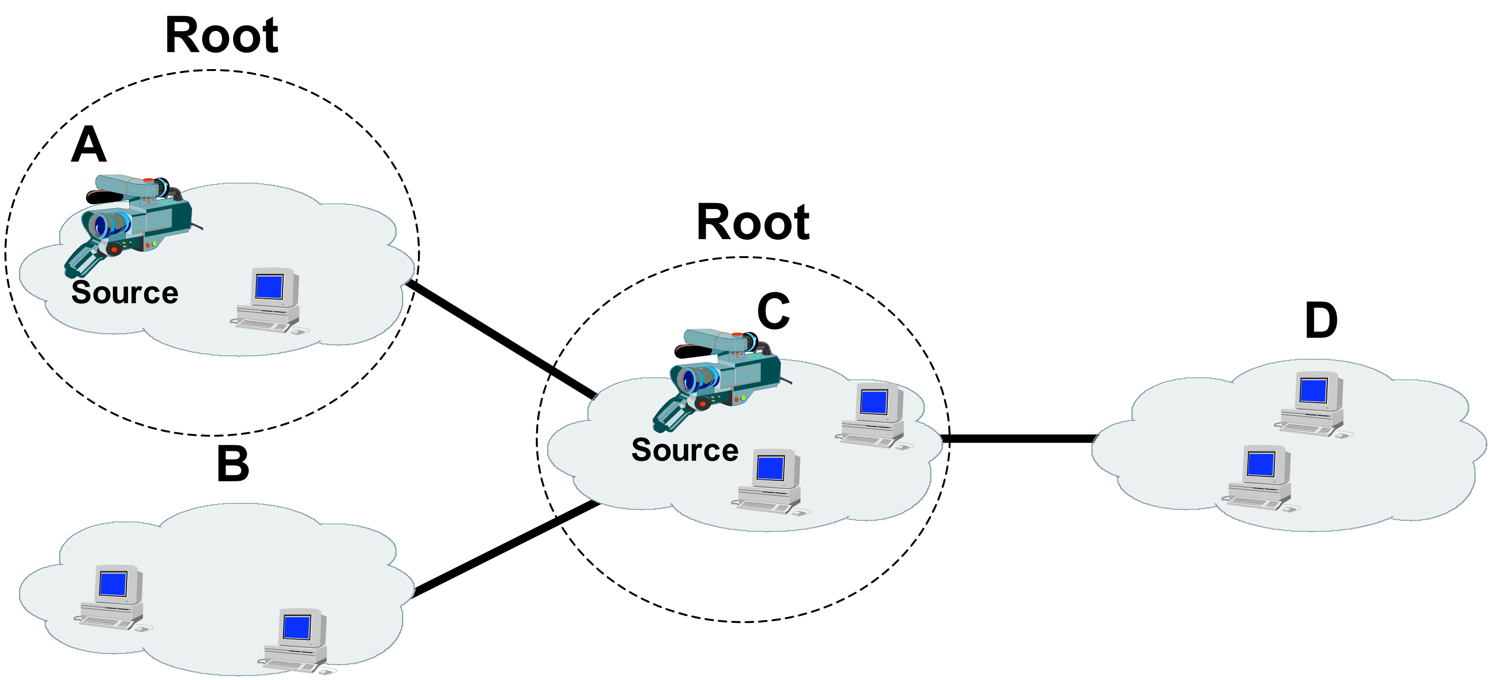
- PIM-SM/MBGP/MSDP is currently deployed and operational
- Longer-term solutions are being investigated
- Border Gateway Multicast Protocol is one of those
  - Should scale to Internet-size
  - Generalizes the concept of rendez-vous point

BGMP

- Border Gateway Multicast Protocol
  - Use a PIM-like method between domains
  - BGMP builds a bidirectional shared tree of domains for a group
  - A root domain is defined for each multicast group G
    - Rendez-vous point mechanism at the domain level
  - Runs in routers that border a multicast routing domain
  - Joins and prunes travel accross domains

How to define the root domain?

- The belief is that no matter the type of session, one domain will always be the logical choice for the root domain



Need a mechanism for strict multicast address allocation!

MASC

- Multicast Address-Set Claim allocates multicast addresses
  - At the domain level
  - Within a domain
  - Between hosts and the networks
- Each domain would obtain (from a top-server) a range of multicast addresses that it would manage for lower-level servers (MAAS)

GLOP, RFC 2770

- Multicast addresses are assigned base on the AS number
  - 233/8 address space is used for GLOP
  - The 16-bit number of the AS number will be concatenated

```
+0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|------233------|----------16 bits AS-----------|--local bits---|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

  - Thus giving 256 multicast addresses per AS

MASC vs GLOP

- GLOP is much simpler but…
- MASC is more scalable!
- However, more class D addresses could be used for GLOP.
- GLOP does not speficy how multicast addresses will be allocated within a domain
- MASC is more hierarchical

Part IV
« The Future »

IPv6

BGMP & MASC

IPv6

Multicast and IP-MPLS networks

Multicast and Overlays networks

Multicast and IPv6

- IPv6 multicast addresses (RFC 2373) are distinguished from unicast addresses by the value of the high-order octet of the addresses: a value of 0xFF (binary 11111111) identifies an address as a multicast address
  - FF02:0:0:0:0:0:0:1 for all Nodes Address
  - FF02:0:0:0:0:0:0:4 for all DVMRP routers
  - ...
- IPv6 adds mobility
- Multicast for mobile users should be considered

IPv6 multicast protocol suite

- Multicast Listener Discovery replaces the IGMP protocol. Current version is MLDv2 (allows SSM, equivalent to IGMPv3)
- MLD messages are carried in ICMPv6 packets
- PIM-SM & PIM-SSM remain the same
- MBGP remains the same, uses address extension to handle seemlessly IPv6 addresses
- No MSDP for the moment: not scalable enough. Other solutions are investigated

Part IV
« The Future »

BGMP & MASC

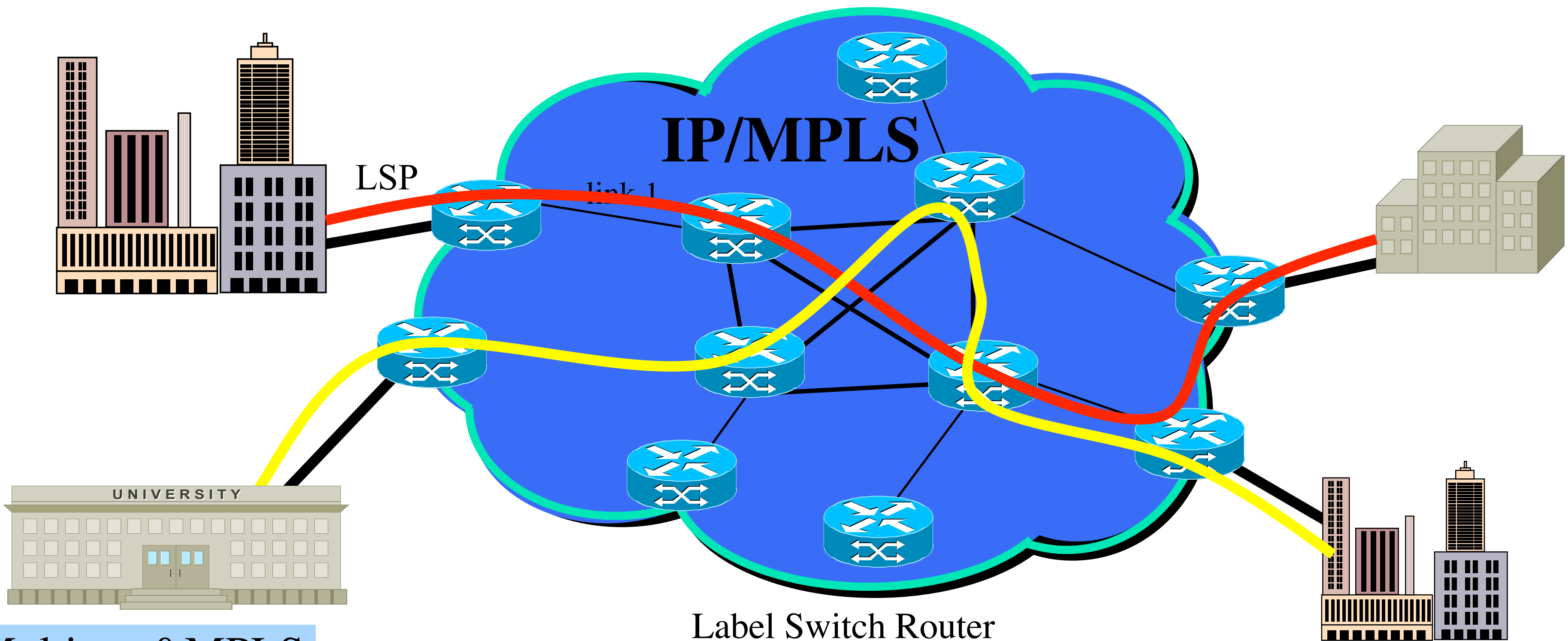
IPv6

Multicast and IP-MPLS networks

Multicast and Overlays networks

MPLS

- Multi-Protocol Label Switching
  - Used to create virtual circuits in IP networks
  - Offers traffic engineering features that make it an attractive technology for many telcos and ISPs.



MPLS is used for…

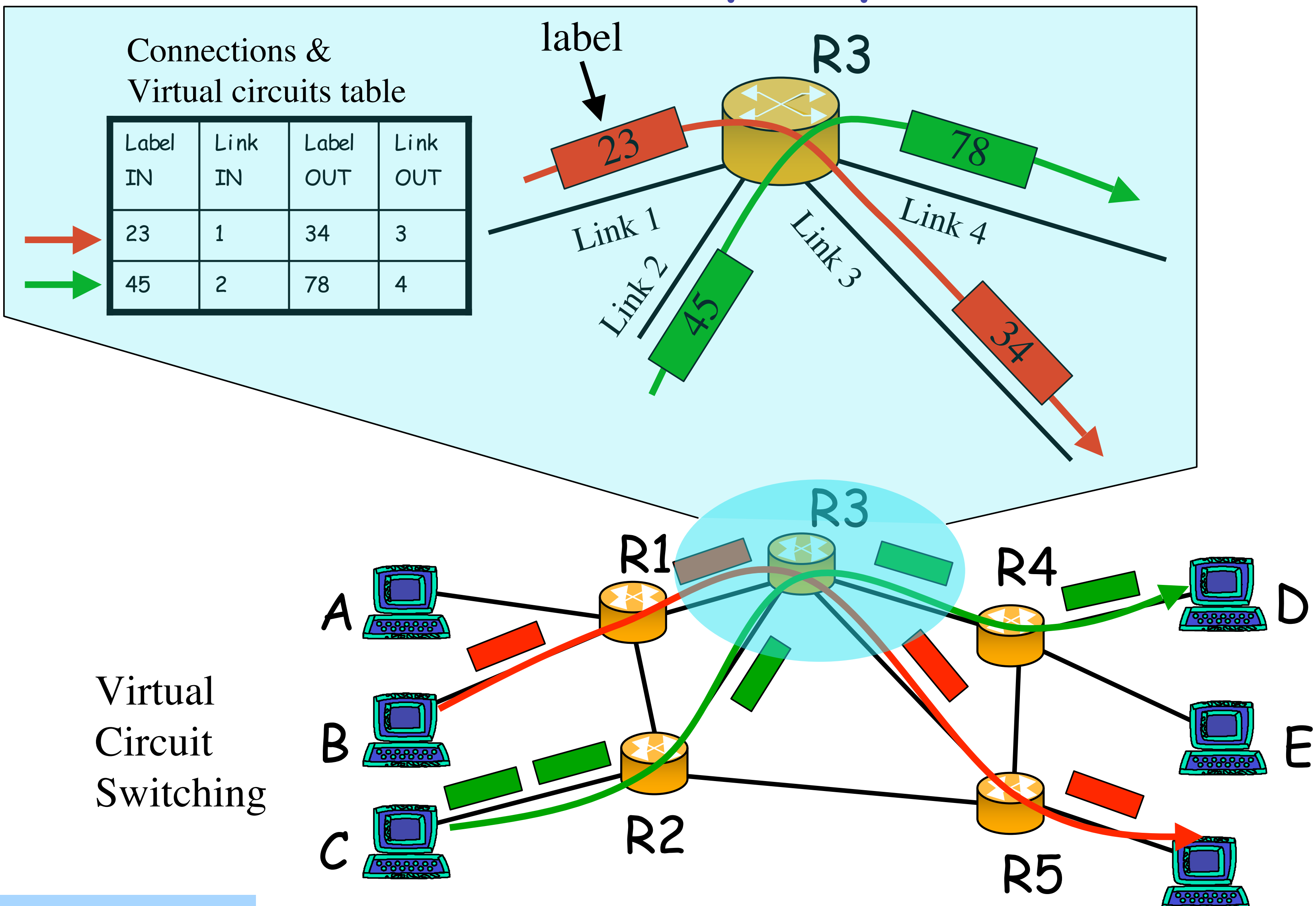
- Virtual Private Networks (VPN)
- Dynamic bandwidth provisioning
- Traffic Engineering
- Quality of Service
- Optical networks with (G)MPLS
- …

Multicast on MPLS networks

- Is a concern because all operators' IP networks may be running MPLS in a very near future
- MPLS and multicast are in the different layers: L2 for MPLS, L3 for multicast
- MPLS routers include 2 separate components
  - Control
    - use standard router protocols in L3 to exchange information with other routers to build and maintain a forwarding table
  - Forwarding
    - Search the forwarding table to make a routing decision for each packet (based on labels)

Review of MPLS operation

Virtual circuit principles



Label IN	Link IN	Label OUT	Link OUT
23	1	34	3
45	2	78	4

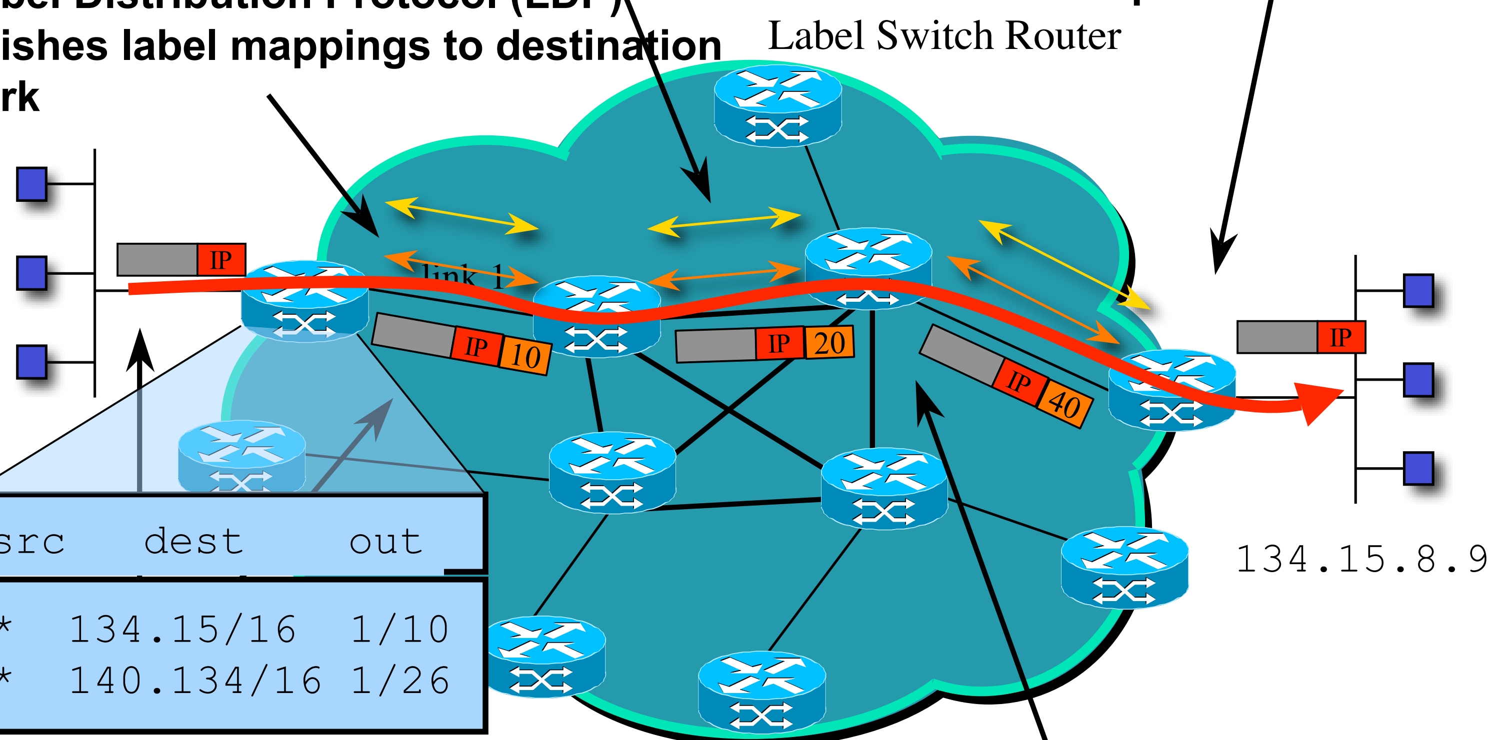
Virtual
Circuit
Switching

Review of MPLS operations (2)

1a. Routing protocols (e.g. OSPF-TE, IS-IS-TE) exchange reachability to destination networks

4. LSR at egress removes label and delivers packet

1b. Label Distribution Protocol (LDP) establishes label mappings to destination network



src	dest	out
*	134.15/16	1/10
*	140.134/16	1/26

2. Ingress LSR receives packet and "label"s packets

Source Yi Lin, modified C. Pham

3. LSR forwards packets using label switching

Multicast on MPLS networks (con't)

- MPLS sets mainly point-to-point LSP (i.e. a virtual circuit) in the core network
  - Multicast needs at least point-to-multipoint
- Existing routing protocols use flood/prune mechanism to build the tree
  - Flood/prune mechanism is costly to support in a virtual circuit approach
- Multicast routing protocols usually use Reverse Path Forwarding (RPF) or other incoming interface check to determine if the packet received belongs to a particular multicast group.
  - In MPLS, multicast tree should be built on a per-interface basis by combining label value and incoming interface.

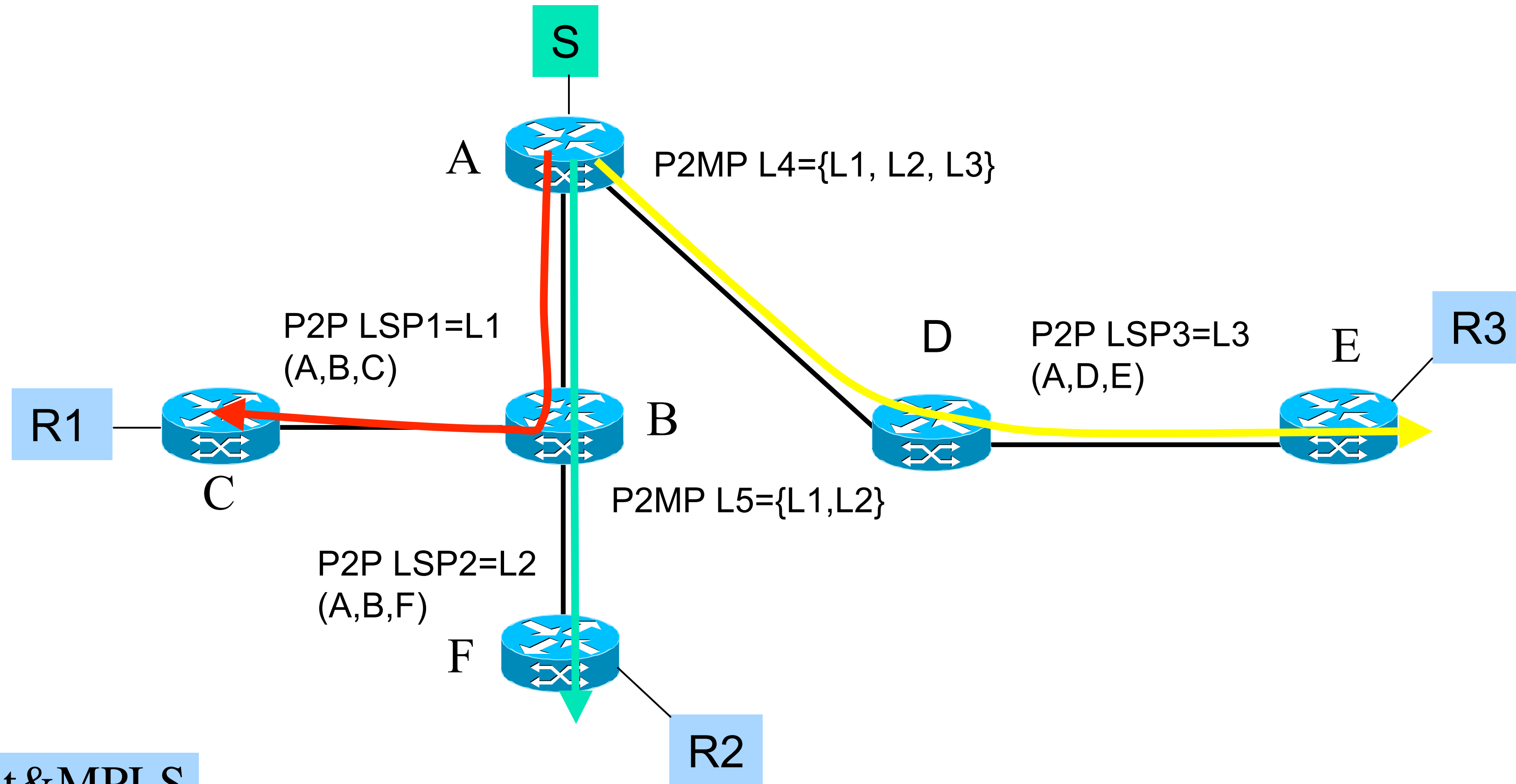
P2MP LSP (work in progress)

draft-raggarwa-mpls-rsvp-te-p2mp-01.txt

- The problem is to introduce multicast functionality in the MPLS data plane
  - Optimize the data plane for high volume multicast
  - No need to optimize the control plane for multicast
- P2MP is done in the data plane
- Control plane uses P2P LSPs as building blocks

P2MP LSP (con't)

- P2MP LSP is setup by merging individual P2P LSPs (called sub-LSP) in the network
  - Most solutions use merging in the data plane
  - MPLS multicast label mappings are setup at the merge nodes



Multicast label assignment

- There are 3 ways to initiate label assignment
  - topology-driven
  - request-driven
  - traffic-driven
- Topology-driven
  - When MPLS is used to transmit unicast traffic, Label Switching Path (LSP) is usually triggered by the network topology. In this case LSP already exists before traffic is transmitted.
  - If topology-driven is applied to multicast, L3 tree needs to be mapped to L2 tree. MPLS-capable routers also have to maintain multicast tree.

Multicast label assignment (con't)

- Traffic-driven
  - only sets up LSP to branches with traffic.
  - consumes fewer labels than topology-driven approach. This may take a longer setup time of LSP, but is better for the longer life span multicast group members.
- Request-driven
  - For explicit multicast members joining/leaving protocols, such as PIM-SM and CBT, join/prune messages can be used to trigger LSP.
  - The drawback is that multicast routing tree has to be constructed twice in L3 and in L2.

Multicast label assignment (cont.)

- Label distribution can be achieved by dedicated protocols, e.g. Label Distribution Protocol (LDP) or RSVP-TE, or by piggybacking on routing protocols.
- Some problems in an MPLS multicast network
  - mixed forwarding
  - co-existence of SPT and RPT
    - Setting up a source specific LSP is a solution in PIM-SM.

Part IV
« The Future »

IPv6

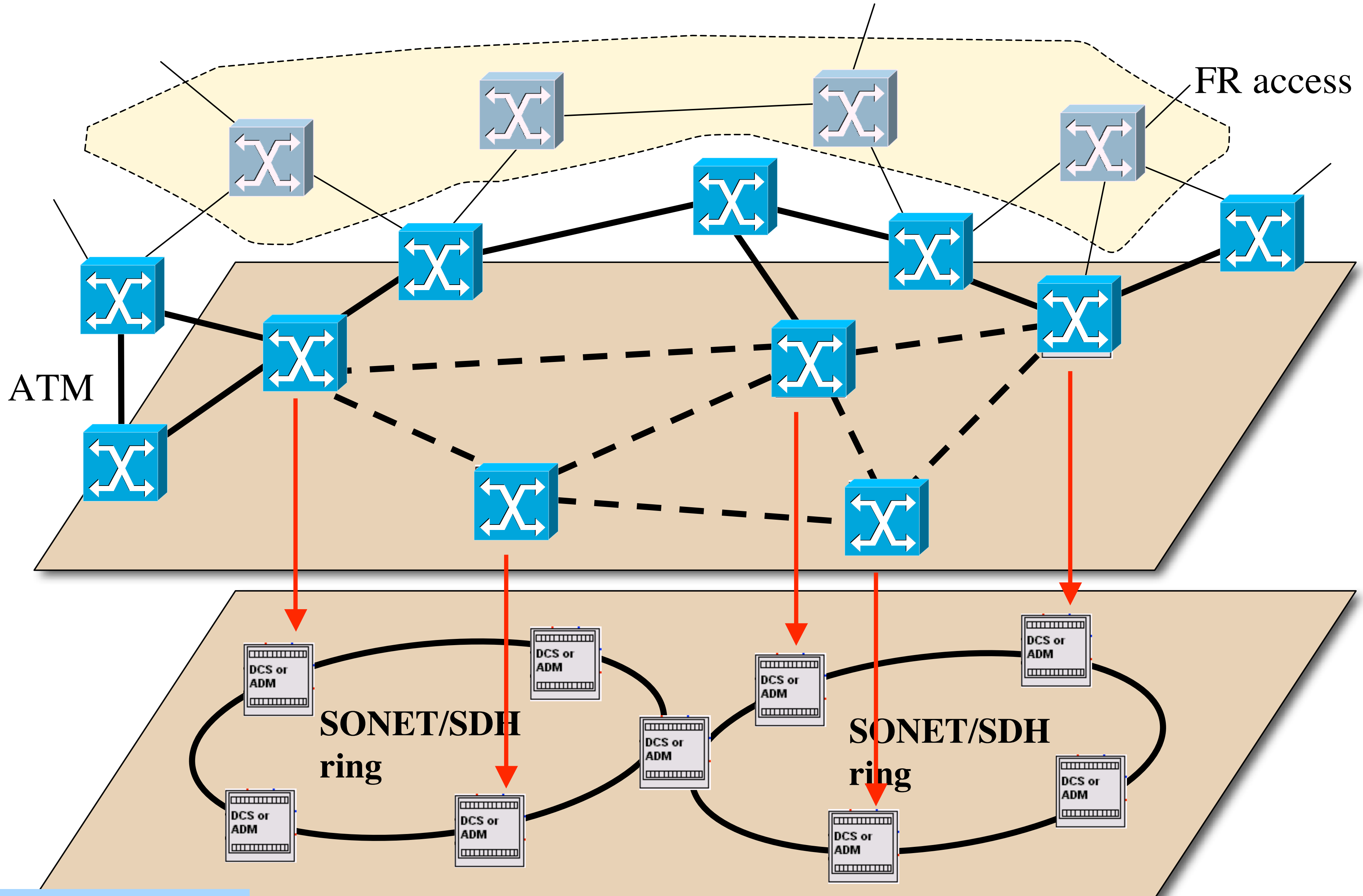
Multicast and IP-MPLS networks

Multicast and Overlays networks

Overlay networks

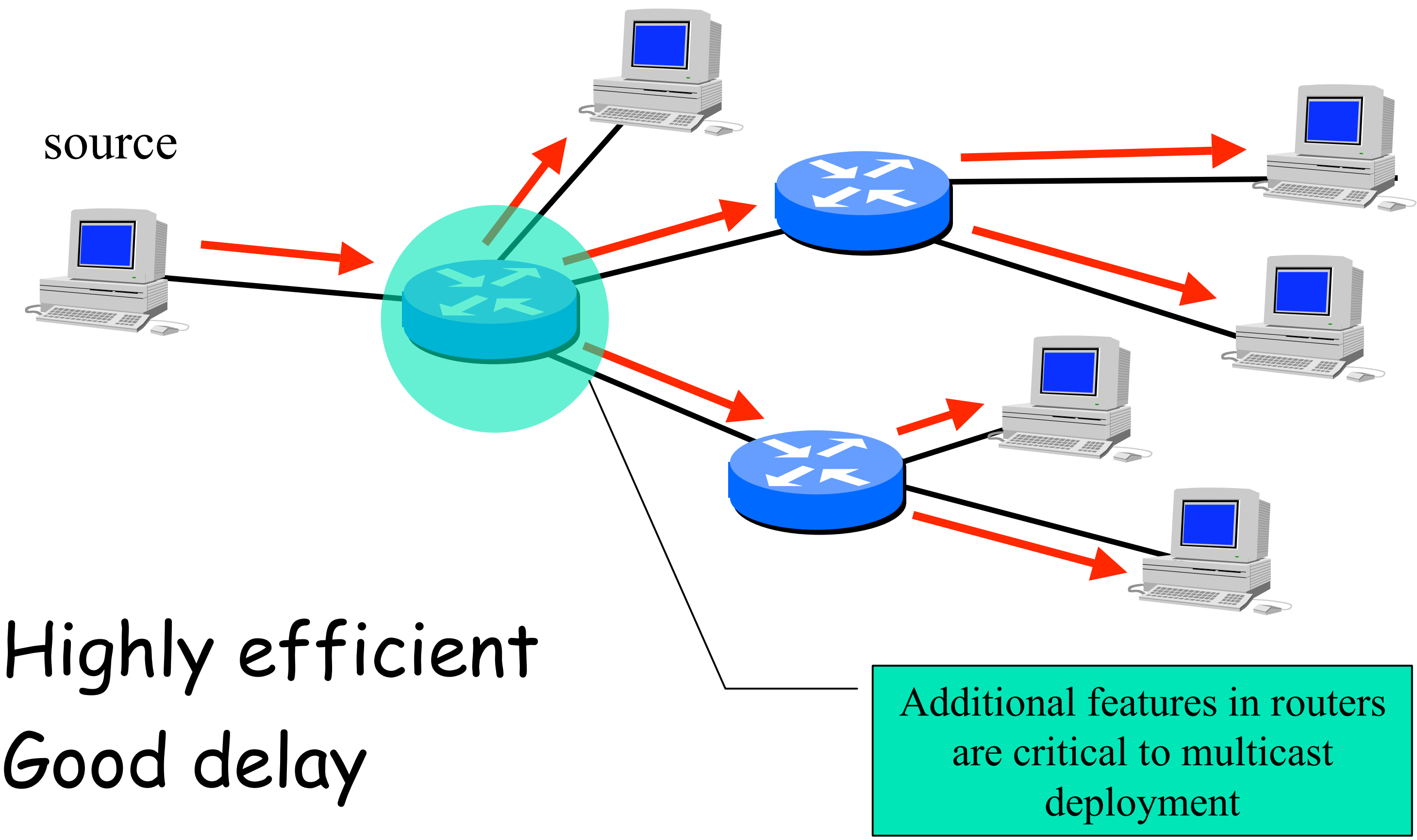
- An overlay network
  - is a network built on top of one or more existing networks
  - adds an additional layer of indirection/virtualization
  - changes properties in one or more areas of underlying network
- Alternative
  - change an existing network layer

Example



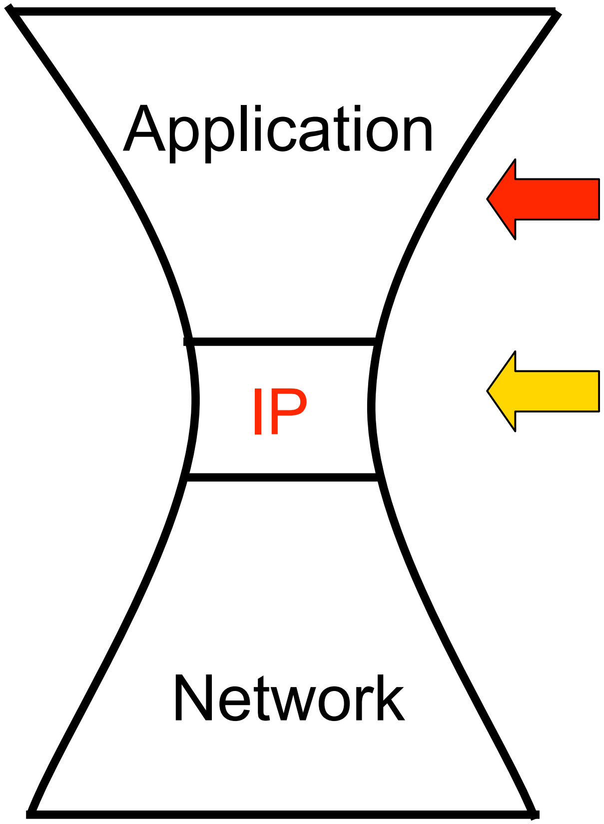
Overlays

Review of native IP Multicast



- Highly efficient
- Good delay

At which layer should multicast be implemented?



Internet architecture

Why not be independant from the network/ISP?

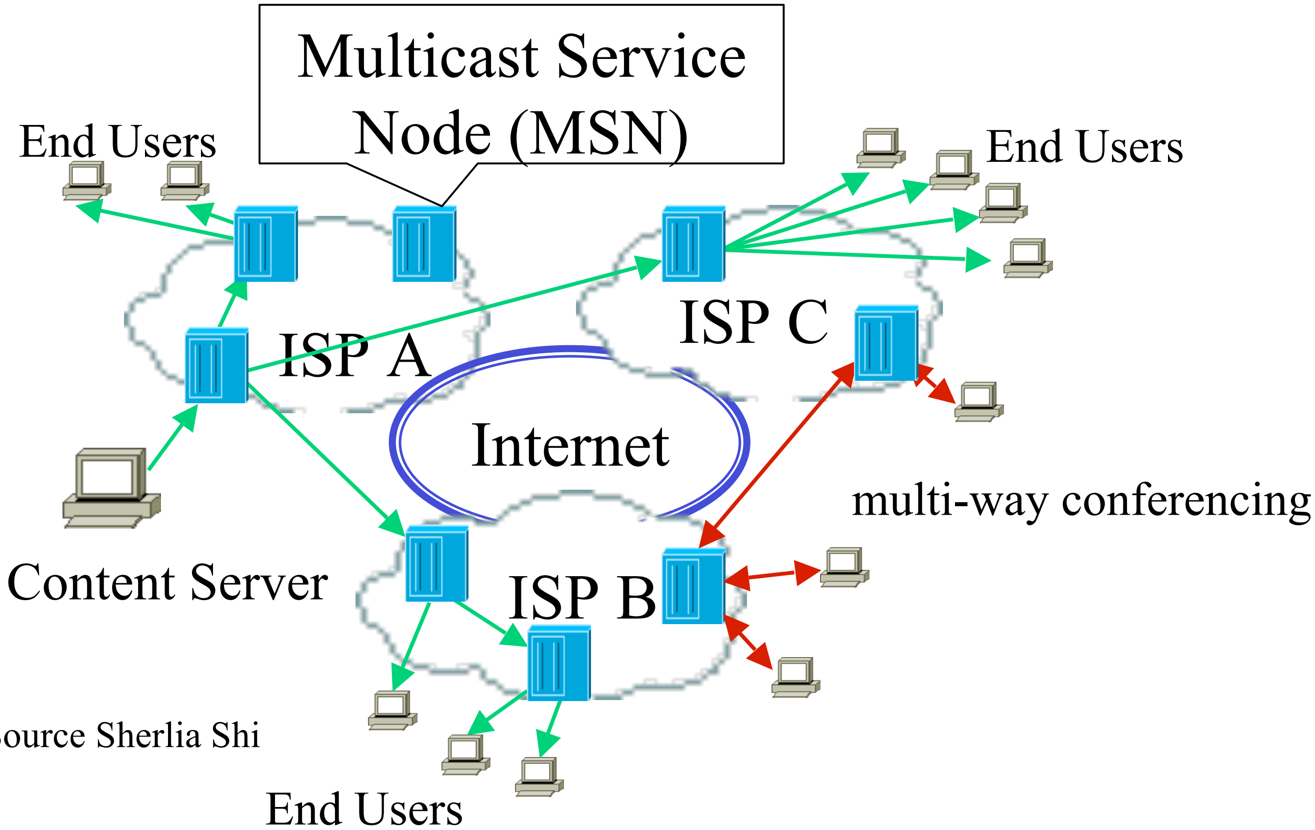
Q: Why has IP Multicast not become popular?

A: ISP's reluctant to turn on IP Multicast

Other problems with IP multicast

- Scales poorly with number of groups
  - A router must maintain state for every group that traverses it
- Supporting higher level functionality is difficult
  - IP Multicast: best-effort multi-point delivery service
  - Reliability and congestion control for IP Multicast complicated
    - Scalable, end-to-end approach for heterogeneous receivers is very difficult
    - Hop-by-hop approach requires more state and processing in routers

Overlays for multicast: example



Source Sherlia Shi

Can go further!

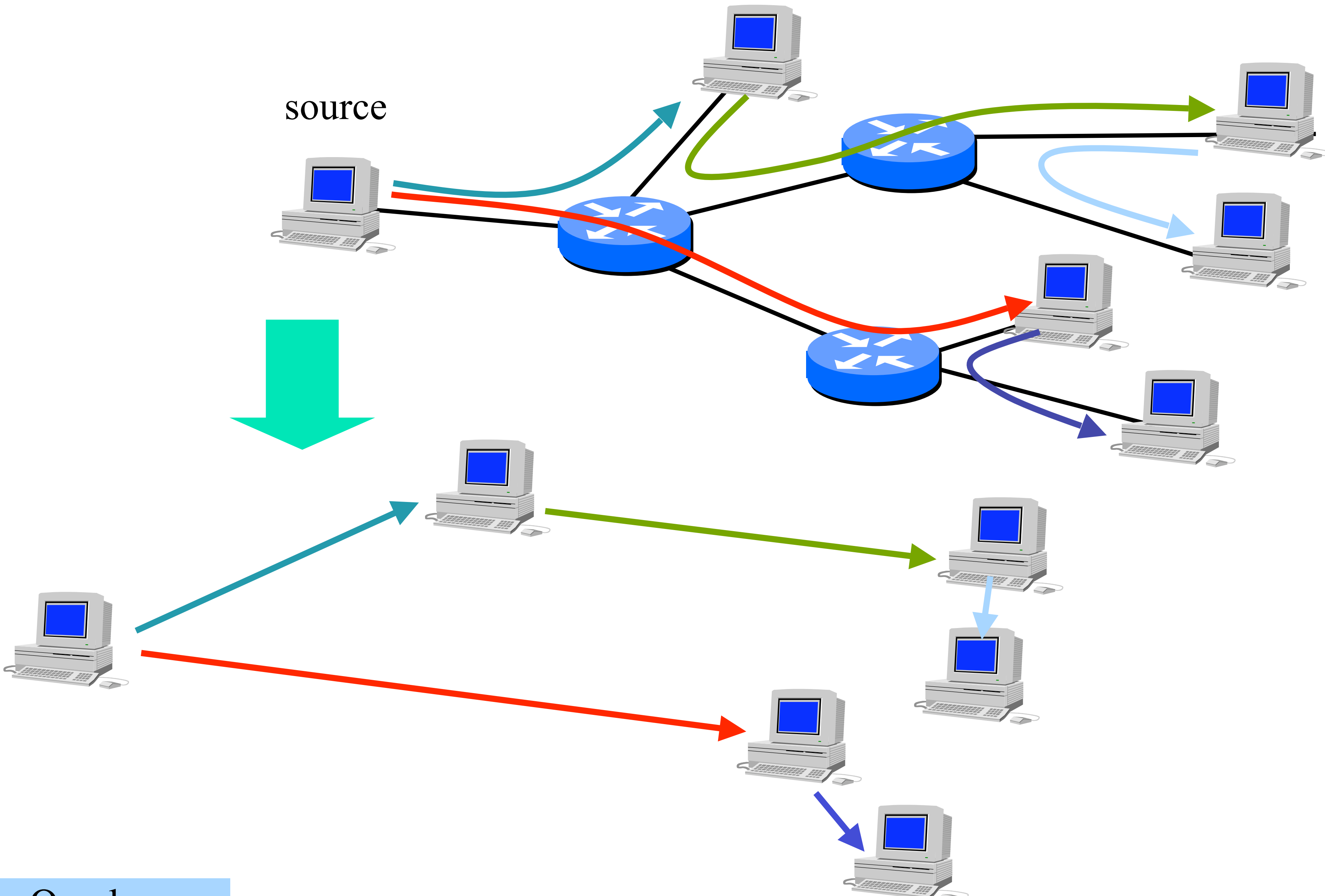
Similar to peer-to-peer comm.

- Peer-to-peer communication models use end-systems to implement advanced file sharing/system features
  - Naspter
  - Gnutella
  - CHORDS
  - PASTRY
  - …

Overlay multicast
End-system multicast
Host-based multicast
Application-level/layer multicast

- Multicast on overlays mainly use end-systems to implement multicast-related features: group management, routing, duplication engine…

End-System Multicast

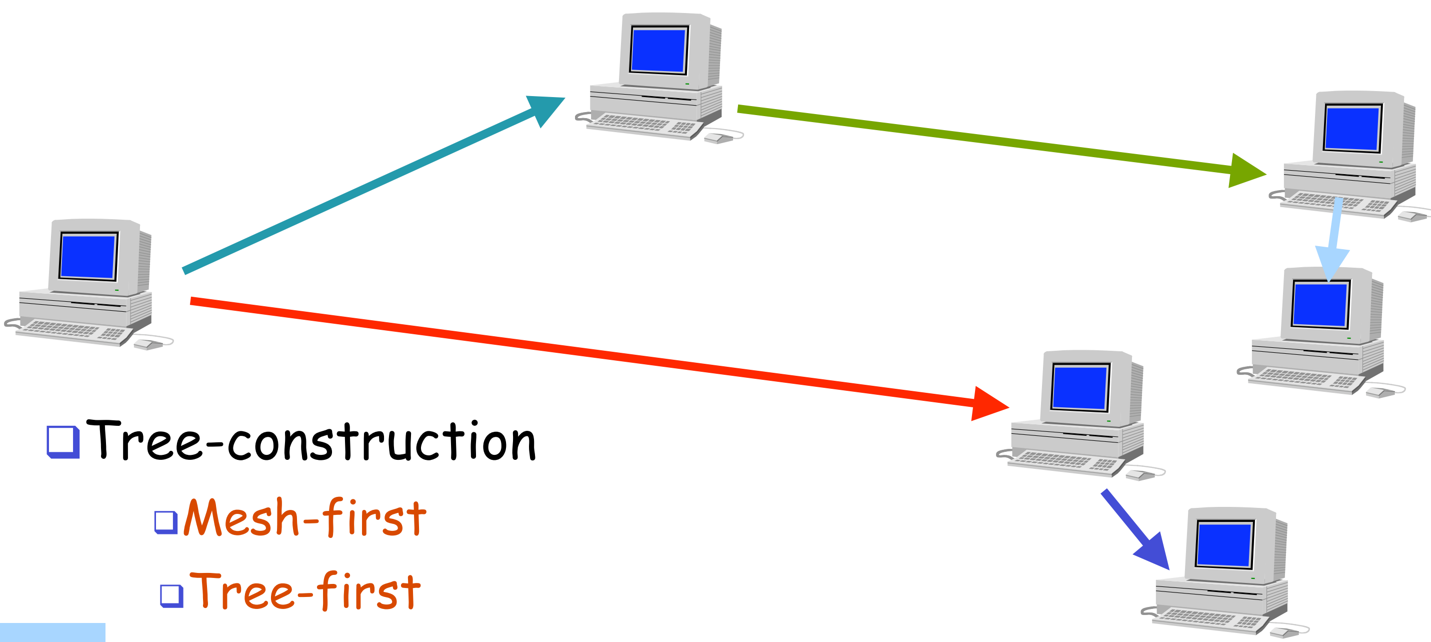


Pros and cons of end-system multicast

- Pros
  - Quick deployment
  - All multicast state in end systems
  - Computation at forwarding points simplifies support for higher level functionality: data packet cache, msg aggregation, congestion control…
- Cons
  - Higher cost of data replication (bandwidth waste)
  - Higher delay: if every body use it on the Internet, what will happen?
  - Can not scale to thousands of node (who needs it?)

Core problem: tree construction

- Well-known optimization problem: can vary width or depth?
  - According to link bandwidth/usage
- However, on the Internet, the tree
  - Must be closely matched to real network topology to be really efficient



- Tree-construction
  - Mesh-first
  - Tree-first

End-system multicast design space

- The tree can be dynamically built with several constraints/heuristics
  - Node's degree
  - Node's utilization
  - Node's geographic position (landmark)
  - Link bandwidth
  - Link delay
  - …

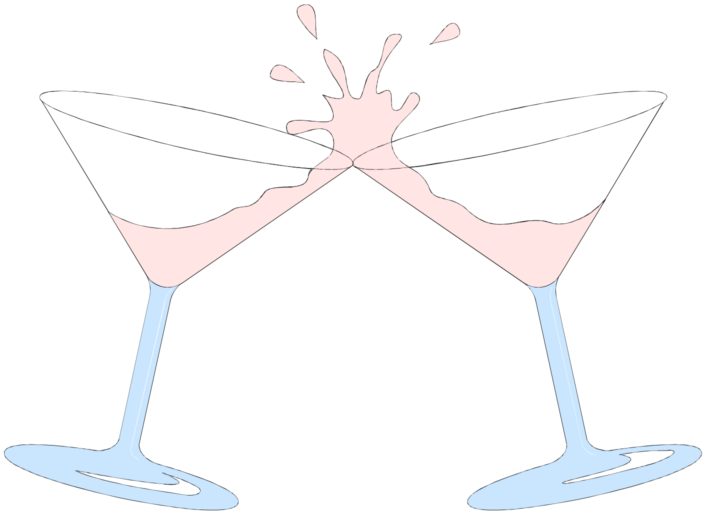
End-systems multicast projects

- NARADA (mesh-first)
- OVERCAST (tree-first, bandwidth)
- SCATTERCAST (tree-first, delay)
- YOID
- YallCast (tree-first)
- HMTP (tree-first)
- OMNI
- ...

N
E
W

C
H
A
P
T
E
R

Conclusions



Conclusions (1)

- Multicast: a technology with high potential…
  - … but also awfully complex !
- Technology starts to be mature:
  - problems are well known and some protocols are already standardized (ALC family)
  - ACK/NACK protocols are on the way to standardization (takes more time as problems are tougher)
  - congestion control (and fairness) is a real concern for large scale deployment
  - does not prevent the use of private reliable multicast solutions

Conclusions (2)

- Deployment is mainly driven by academic networks…
  - where are the killing applications ?
  - video and popular content distribution to clients… yes
  - high performance computing over datagrids… yes
- Where should we go?
  - More specific models (i.e. SSM),
  - More security, more control
  - More "individual" initiatives (end-system multicast)?