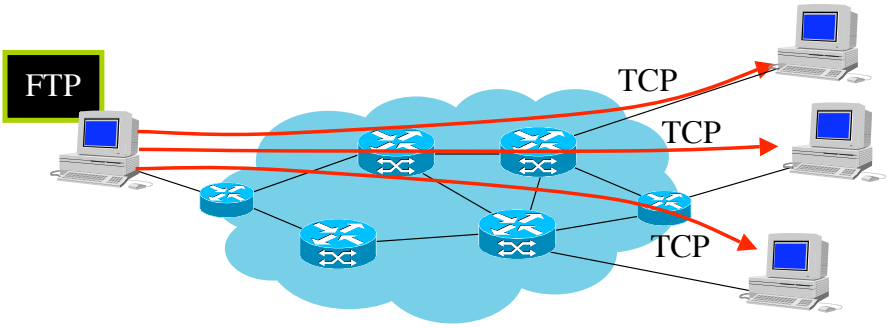
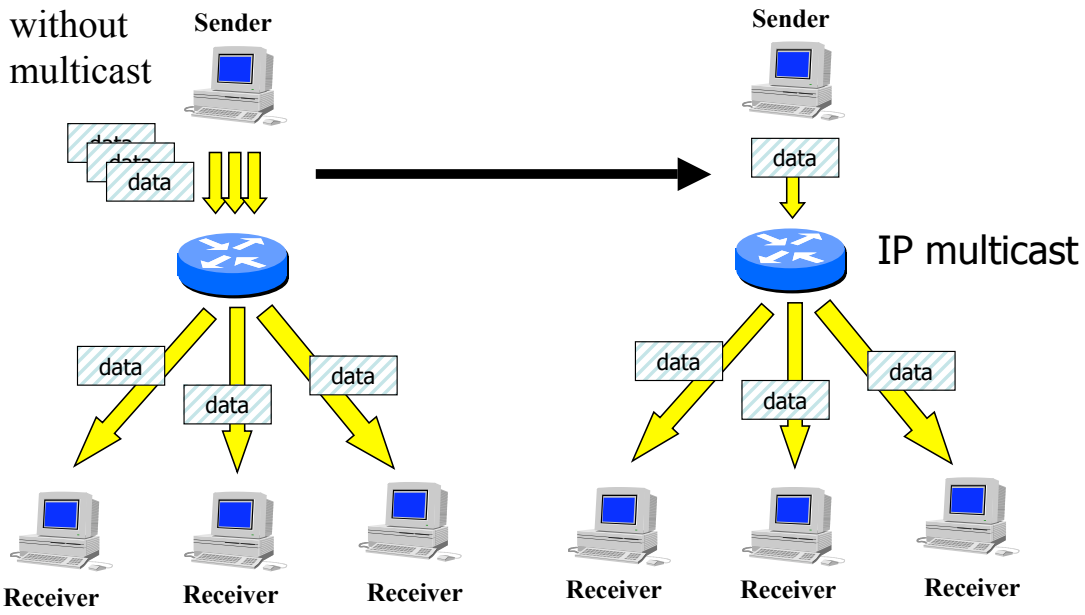
NEW CHAPTER

Unicast, the current (Internet) communication model



- There are applications that naturally need multi-destination communication model
  - Collaborative works
  - Visio-conferencing
  - Software distribution
  - Video-on-Demand
  - Virtual Reality
  - Distributed Simulation

From unicast to multicast

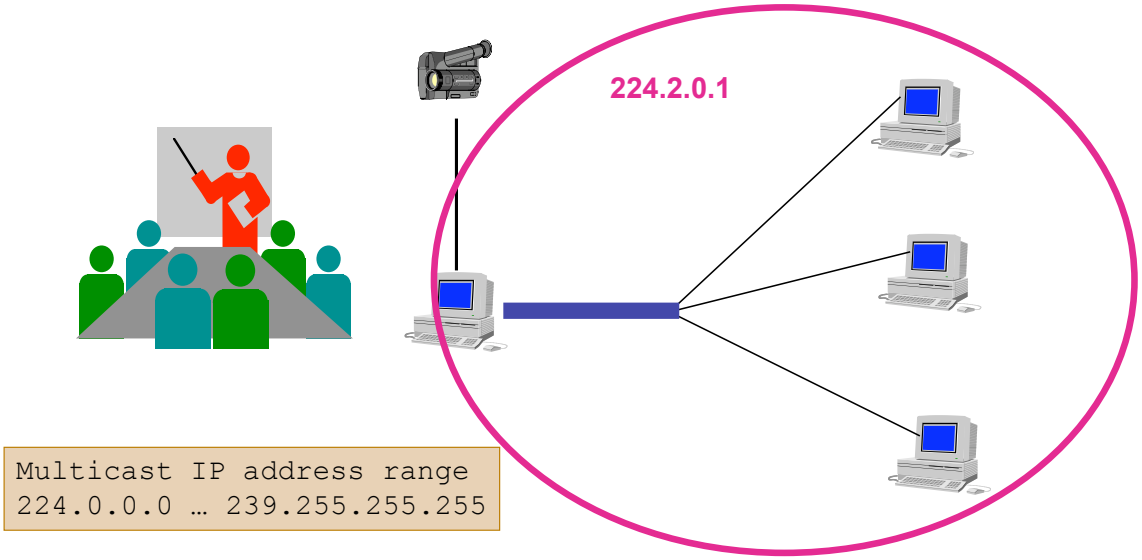


A very simple example in figures

- File replication (PUSH) with ftp
  - 10MBytes file
  - 1 source, n receivers (replication sites)
  - 512KBits/s upstream access
  - n=100
    - T_x= 4.55 hours
  - n=1000
    - T_x= 1 day 21 hours 30 mins!

Multicast in example

The user's perspective

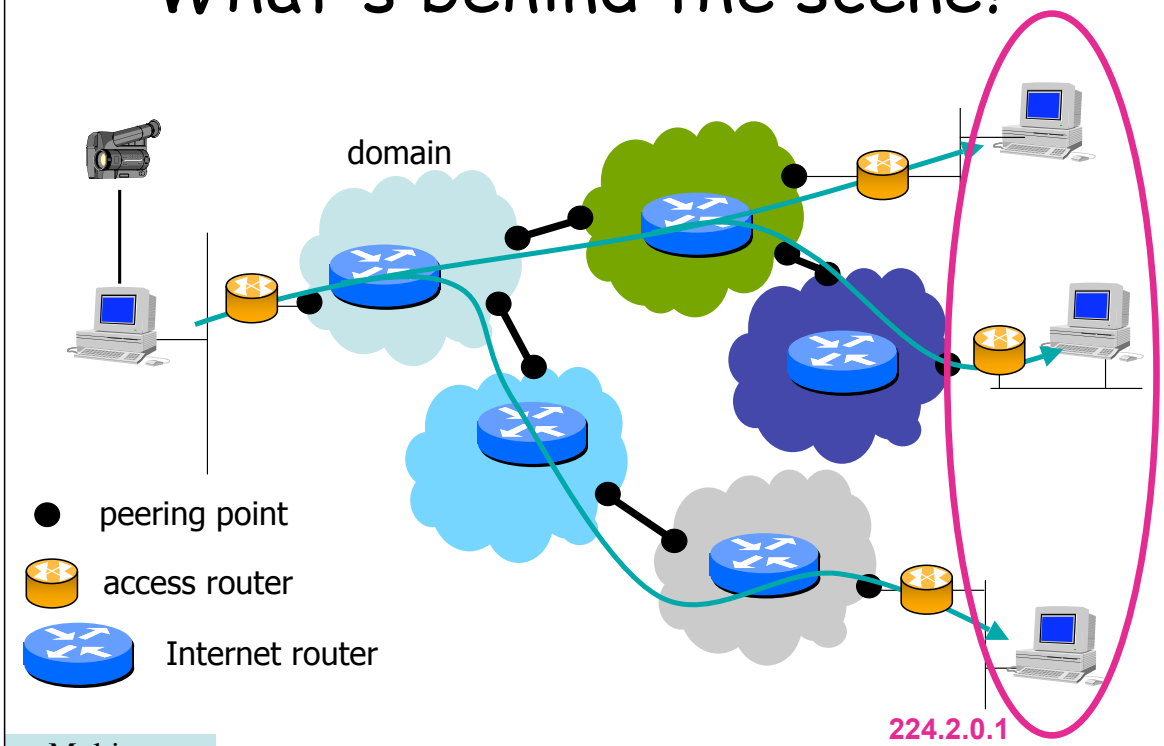


Multicast IP address range
224.0.0.0 … 239.255.255.255

Multicast address group 224.2.0.1

from UREC, http://www.urec.fr

What's behind the scene?



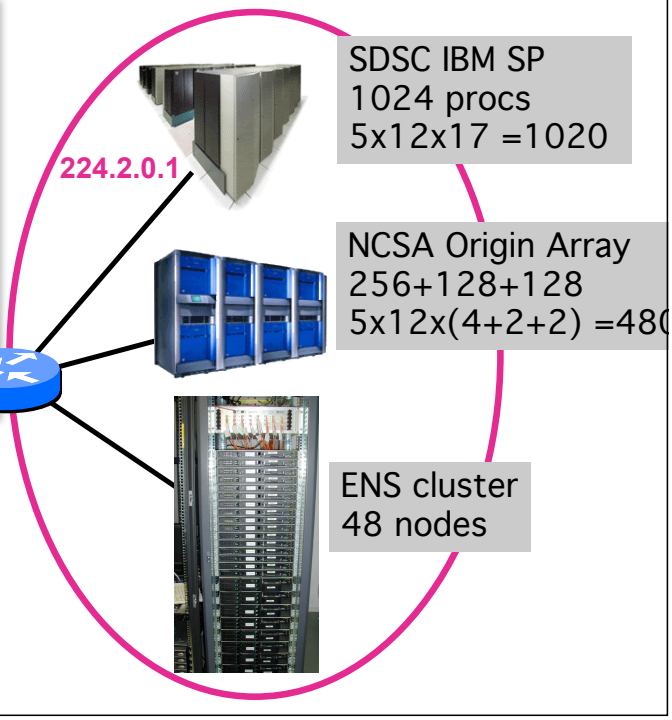
Ex: Reliable multicast on grids

Data replications

Code & data transfers, interactive job submissions

Data communications for distributed applications (collective & gather operations, sync. barrier)

Databases, directories services



IP multicast TODO list

- ✓ Receivers must be able to subscribe to groups, need group management facilities
- ✓ A communication tree must be built from the source to the receivers
- ✓ Branching points in the tree must keep multicast state information
- ✓ Inter-domain routing must be reconsidered for multicast traffic
- ✓ Need to consider non-multicast clouds

The challenges of multicast

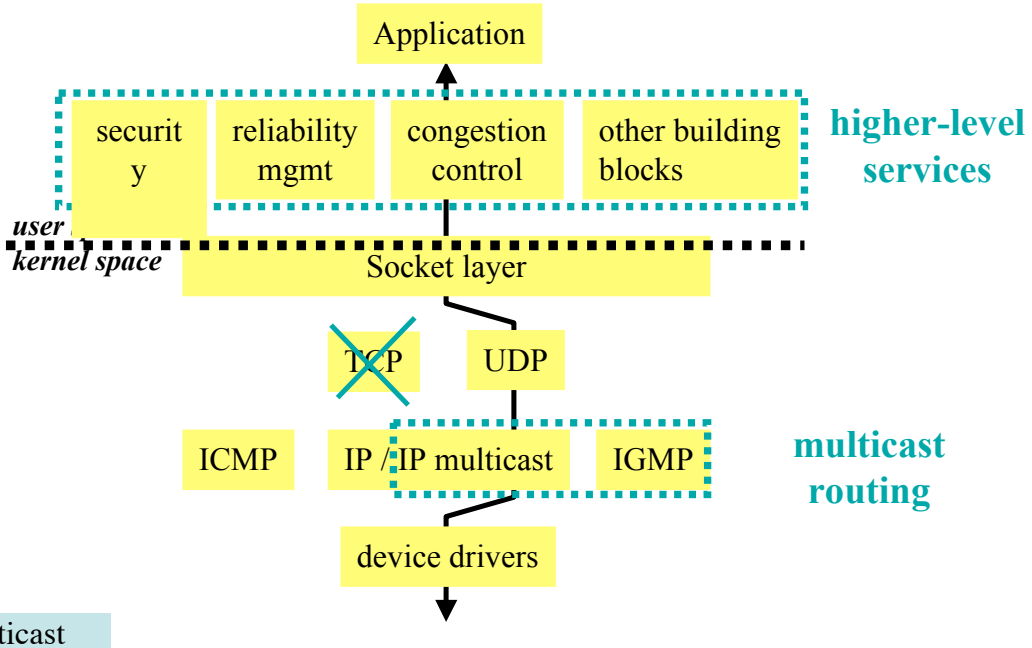
SCALABILITY - SECURITY - TCP Friendliness - MANAGEMENT

SCALABILITY

SCALABILITY

SCALABILI

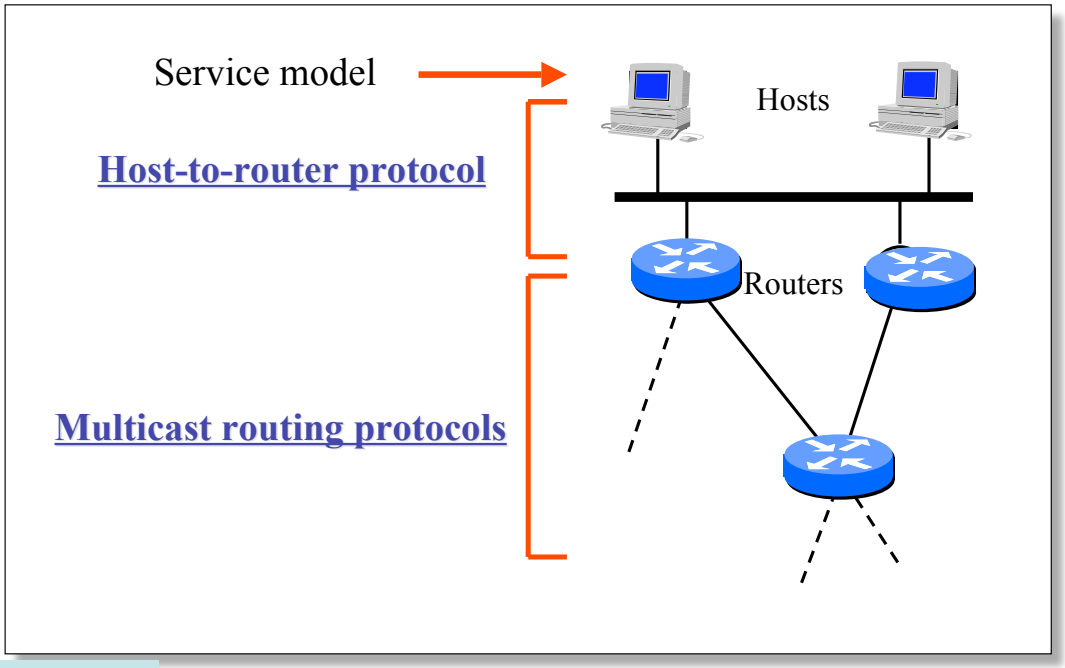
Multicast and the TCP/IP layered model



The two sides of IP multicast

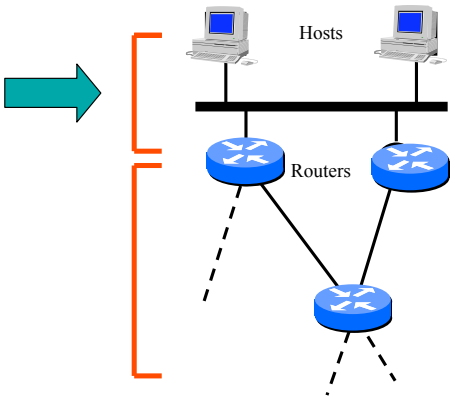
- local-area multicast
  - use the potential diffusion capabilities of the physical layer (e.g. Ethernet)
  - efficient and straightforward
- wide-area multicast
  - requires to go through multicast routers, use IGMP/multicast routing/...(e.g. DVMRP, PIM-DM, PIM-SM, PIM-SSM, MSDP, MBGP, BGMP, MOSPF, etc.)
  - routing in the same administrative domain is simple and efficient
  - inter-domain routing is complex, not fully operational

IP Multicast Architecture



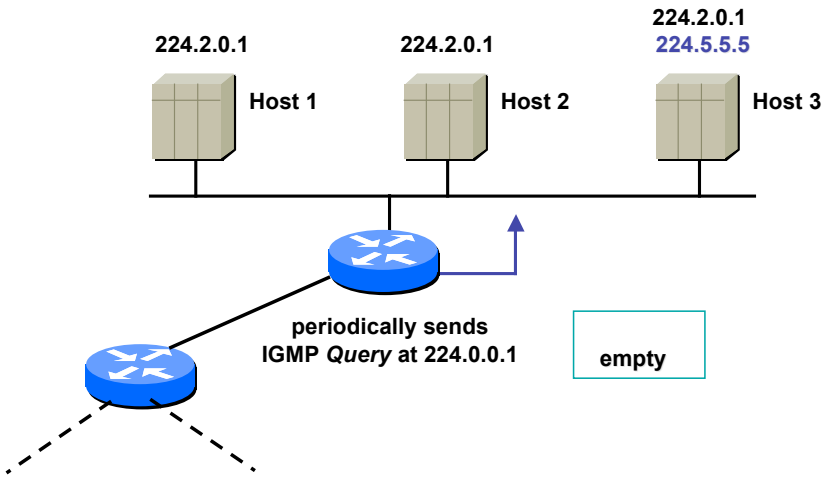
Internet Group Management Protocol (RFC 1112)

- IGMP: "signaling" protocol to establish, maintain, remove groups on a subnet.
- Objective: keep router up-to-date with group membership of entire LAN
- Each host keeps track of which mcast groups are subscribed to
  - Socket API informs IGMP process of all joins



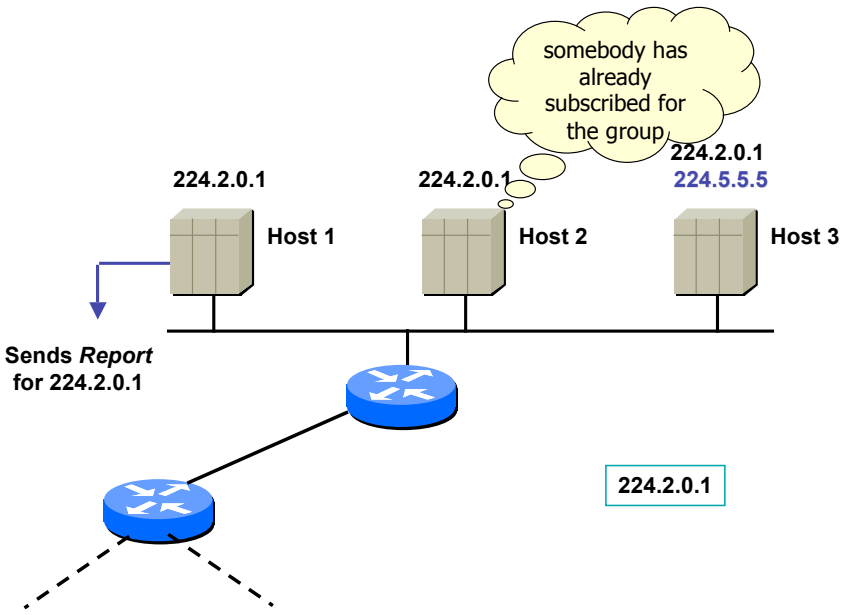
Routers need not know who all the members are, only that members exist

IGMP: subscribe to a group (1)

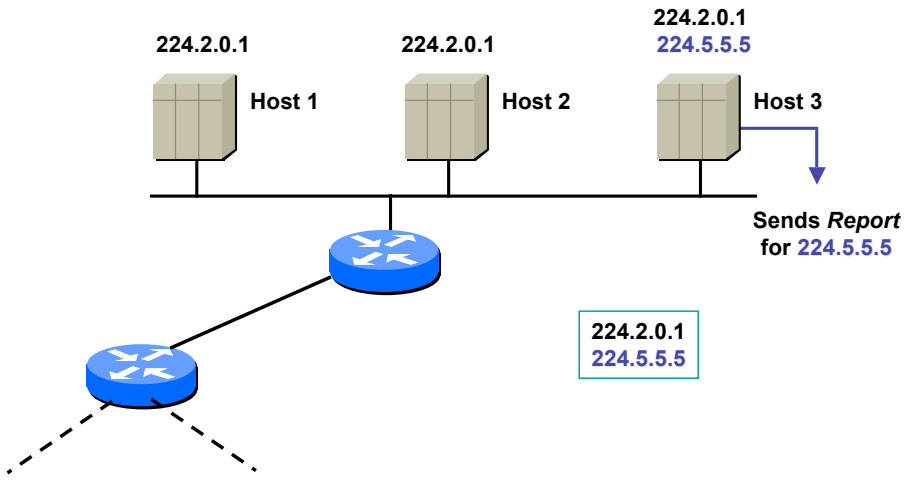


224.0.0.1 reach all multicast host on the subnet

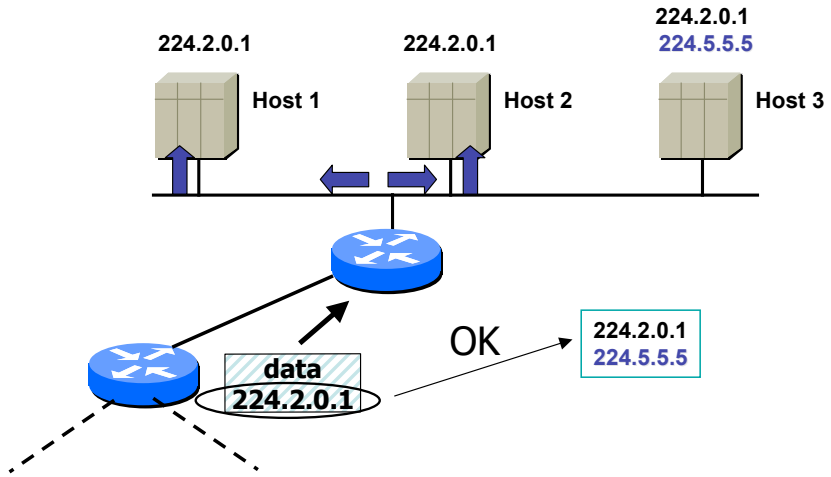
IGMP: subscribe to a group (2)



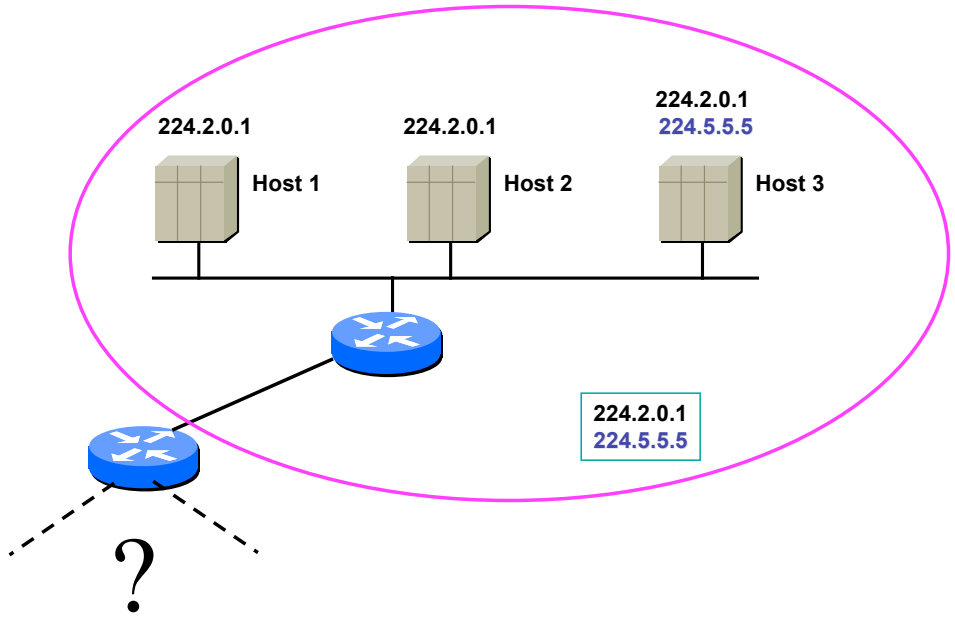
IGMP: subscribe to a group (3)



Data distribution example



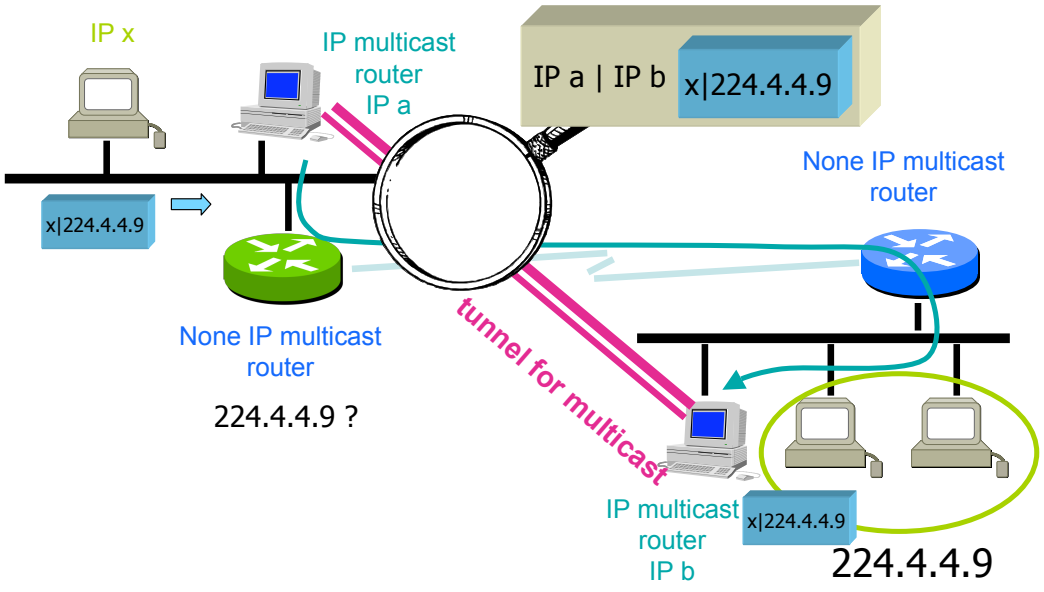
OK, now I can express local interest, so what?



IP multicast routing

- Find a tree (dedicated, shared) between the source(s) and the receivers
- Dense Mode
  - Assume that there are many many receivers willing to get multicast traffic
- Sparse Mode
  - Assume that the number of receivers is small. Require an explicite query from the receivers.

Multicast tunnelling illustrated

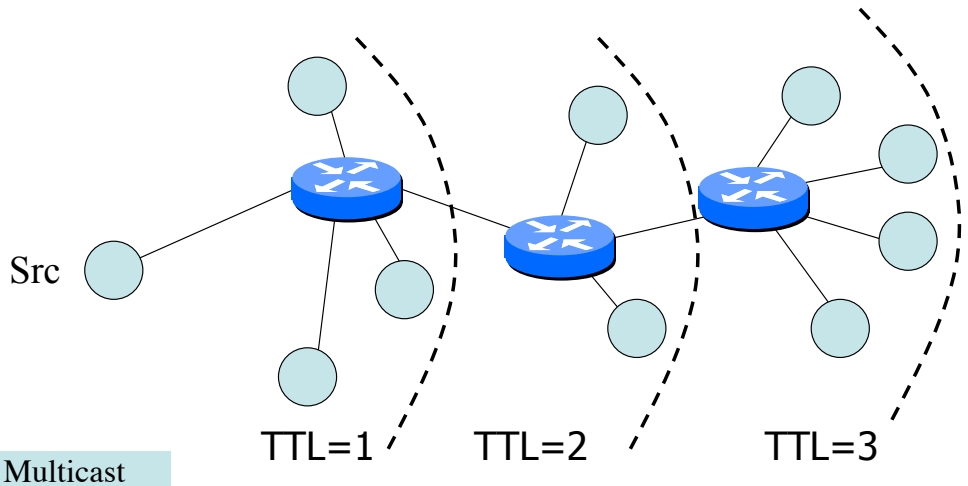


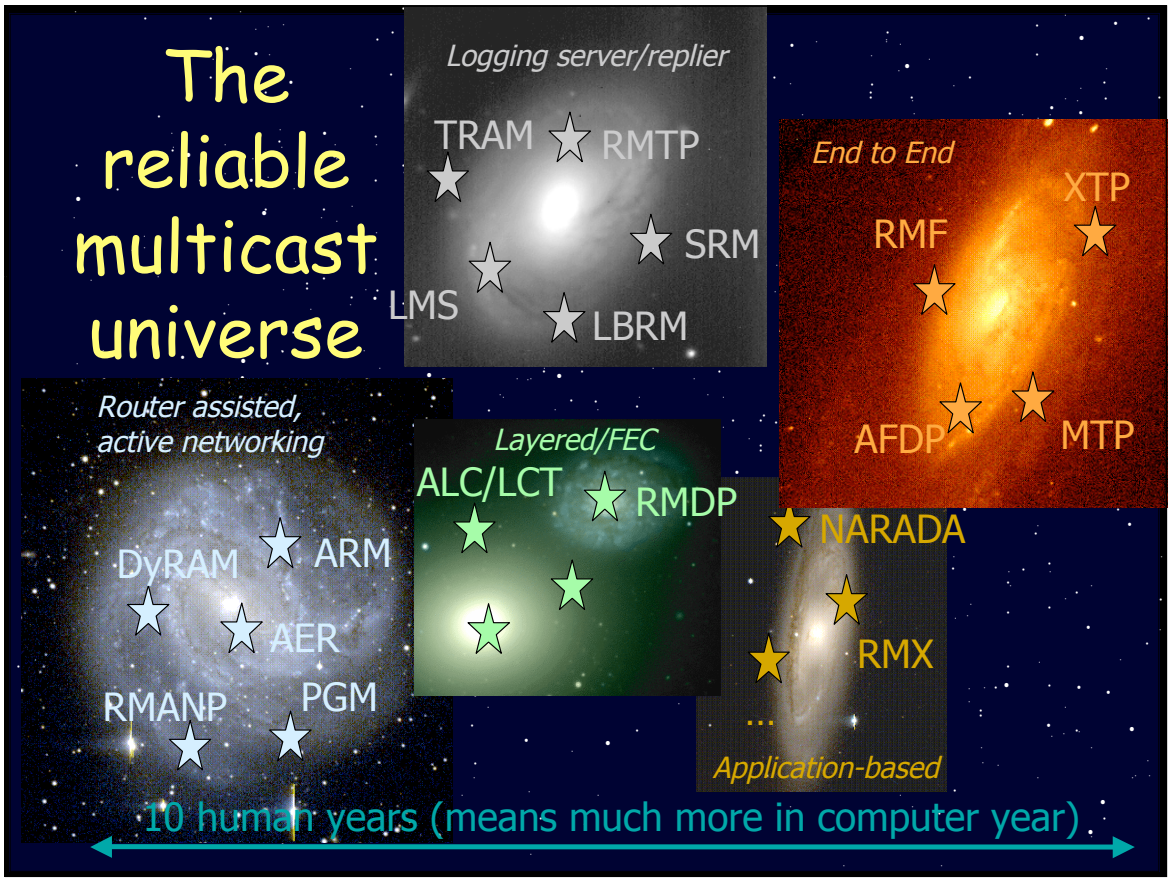
Reliability Models

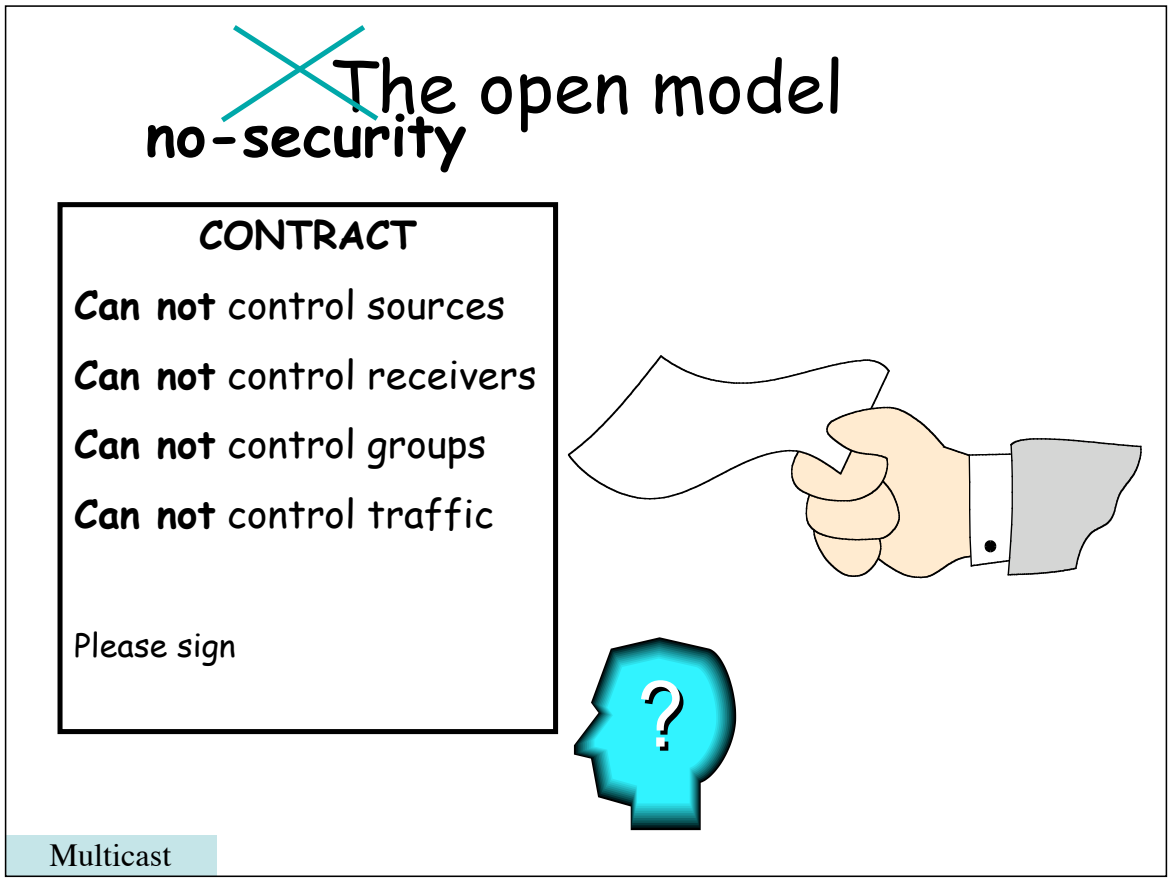
- Reliability => requires redundancy to recover from uncertain loss or other failure modes.
- Two types of redundancy:
  - Spatial redundancy: independent backup copies
    - Forward error correction (FEC) codes
    - Problem: requires huge overhead, since the FEC is also part of the packet(s) it cannot recover from erasure of all packets
  - Temporal redundancy: retransmit if packets lost/error
    - Lazy: trades off response time for reliability
    - Design of status reports and retransmission optimization important

Simple TTL-scoped of repairs

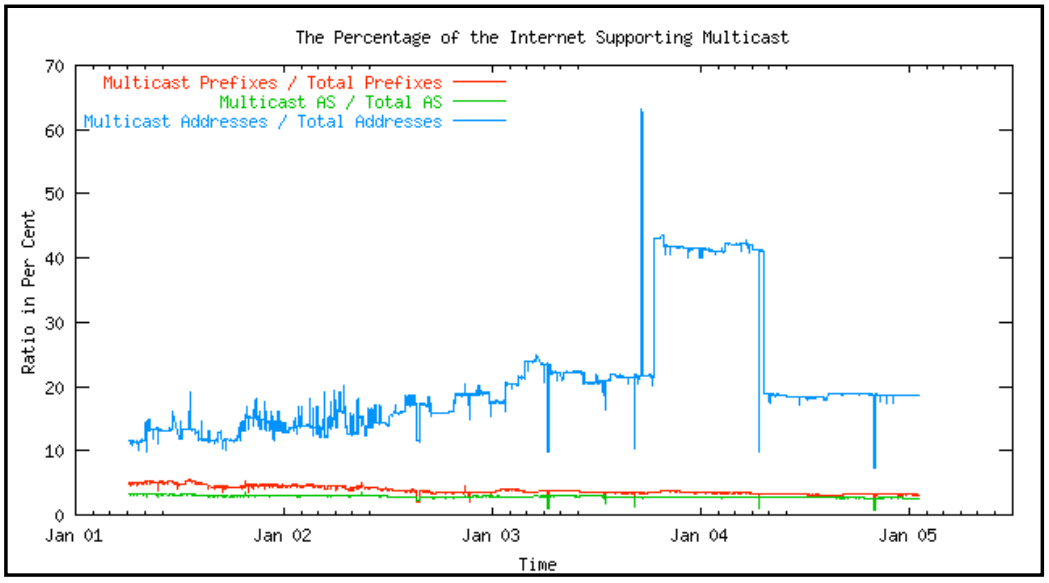
- use the TTL field of IP packets to limit the scope of the repair packet




The reliable multicast universe
Logging server/replier
TRAM
RMTP
SRM
LMS
LBRM
End to End
XTP
RMF
AFDP
MTP
Router assisted, active networking
DyRAM
ARM
AER
RMANP
PGM
Layered/FEC
ALC/LCT
RMDP
NARADA
RMX
...
Application-based
10 human years (means much more in computer year)


no-security
The open model
CONTRACT
Can not control sources
Can not control receivers
Can not control groups
Can not control traffic
Please sign
?
Multicast

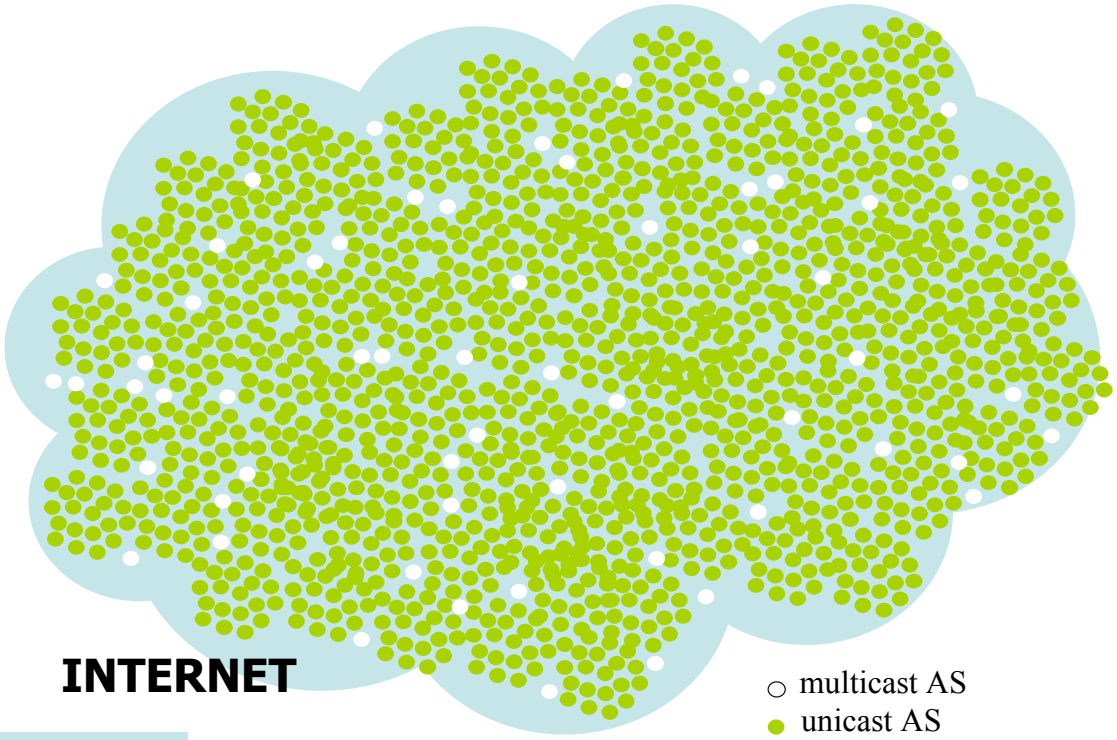
Relative Size of the Multicast Enabled Internet



source www.multicasttech.com/status

Status?

The gap in images



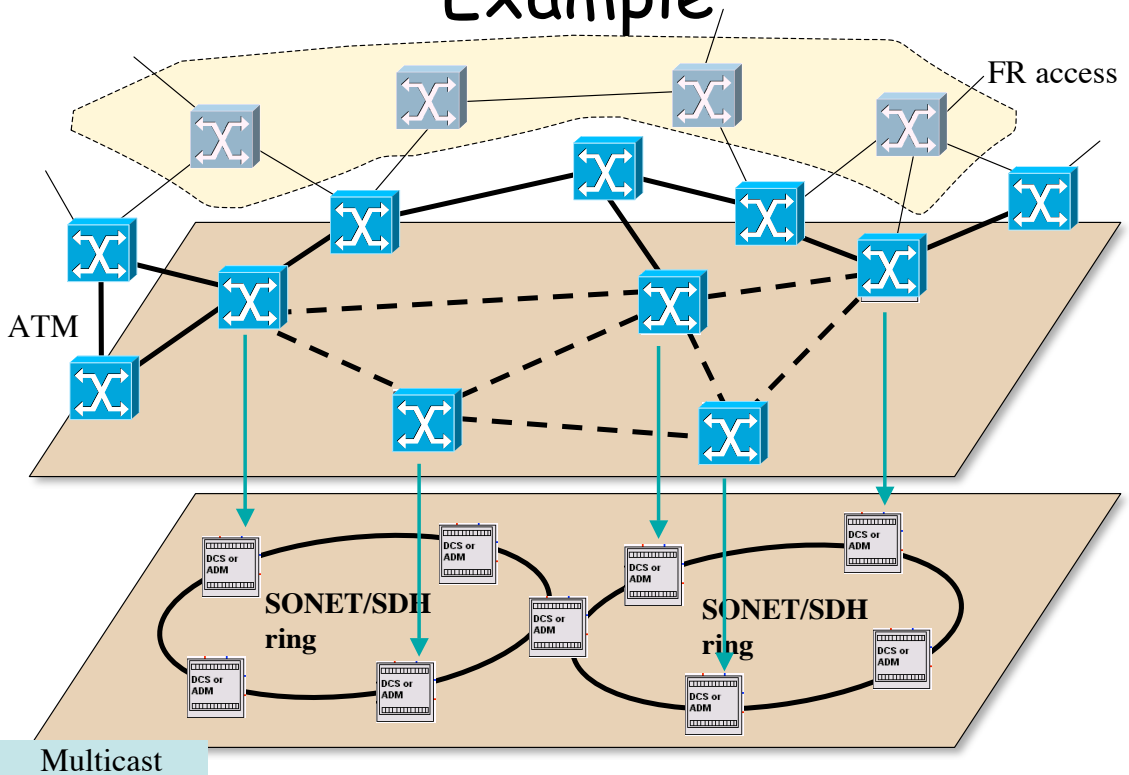
Multicast

The future of multicast

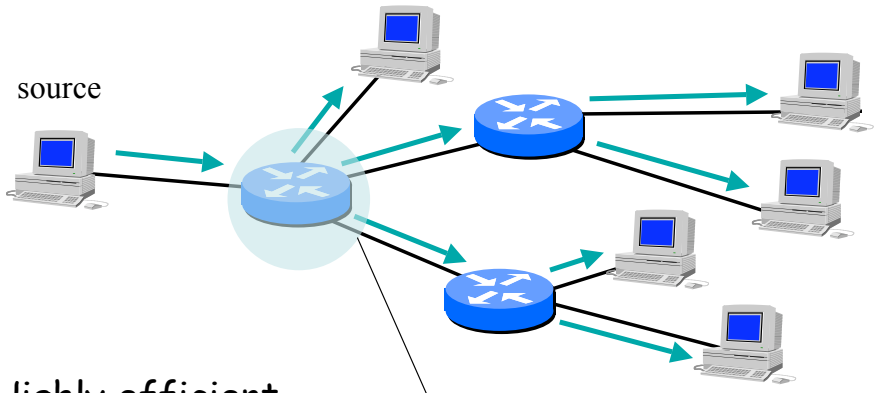
Overlay networks

- 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



Native IP multicast

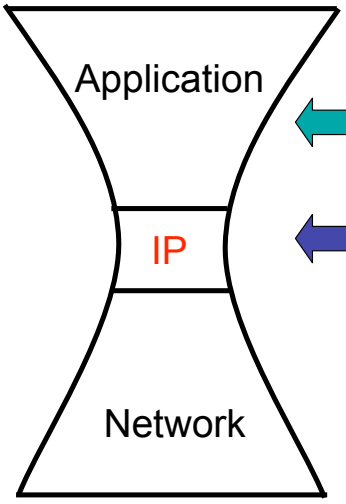


- Highly efficient
- Good delay

BUT …

Additional features in routers are critical to multicast deployment

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?

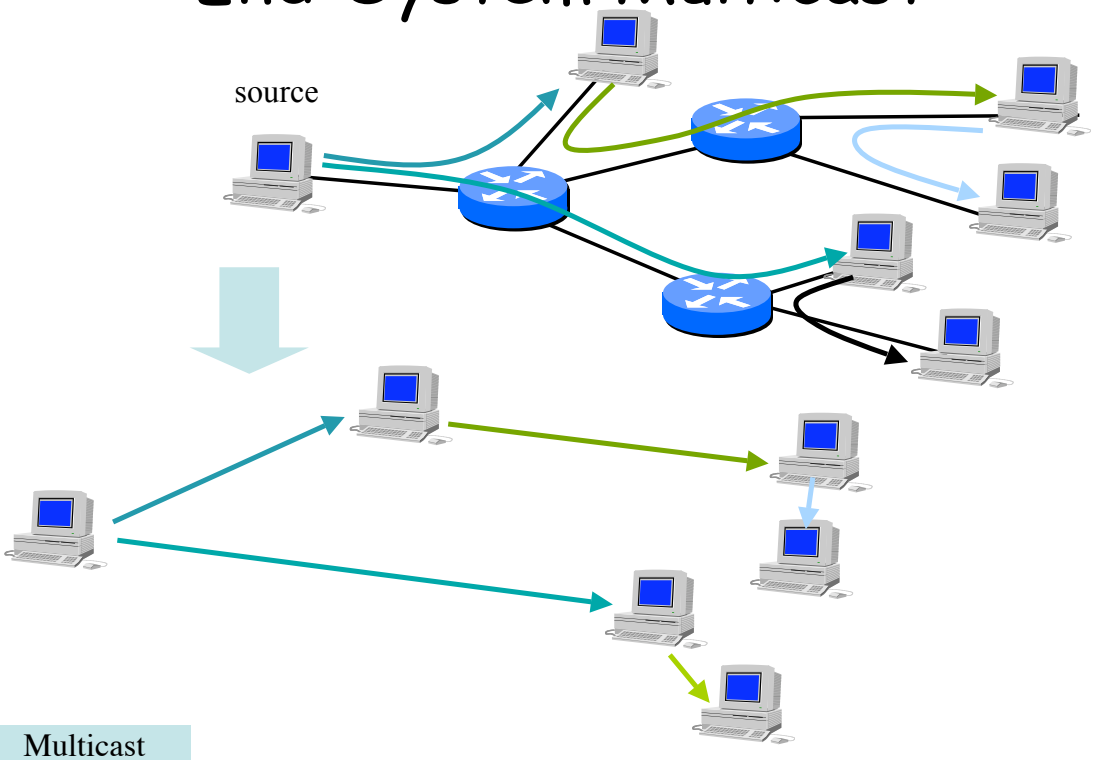
A: ISP's reluctant to turn on IP Multicast

Similar to P2P communication

- Peer-to-peer communication models use end-systems to implement advanced file sharing/system features
  - Naspter
  - Gnutella
  - CHORDS
  - PASTRY
  - …
- Multicast on overlays mainly use end-systems to implement multicast-related features: group management, routing, duplication engine…

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

End-System Multicast

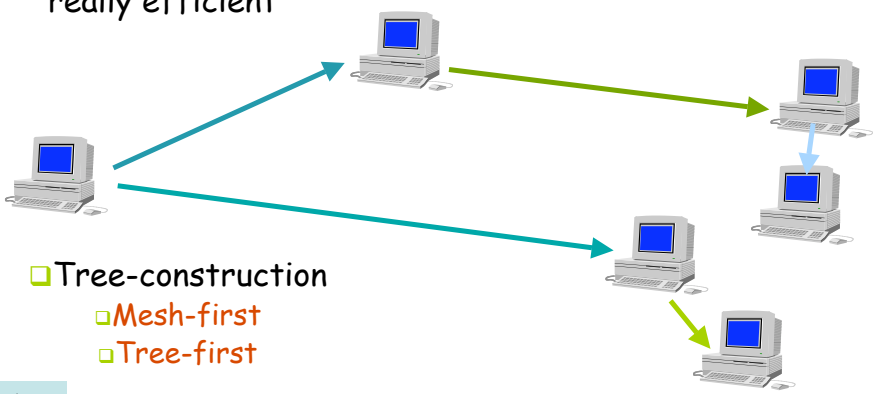


Pros and cons of end-system multicast

- 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
- …

Conclusions

- There's a lot more technologies going on that have impact on computational science
  - Pure optical networks, broadband wireless
  - Peer-to-Peer, Overlays
  - Web services…
- The future will be all connected, all IP, anytime, anywhere, for more…