Evolution in Protocols and Network Design for Multicast Support

Tiziana Ferrari

Tiziana.Ferrari@cnaf.infn.it

Istituto Nazionale di Fisica Nucleare

INFN - CNAF

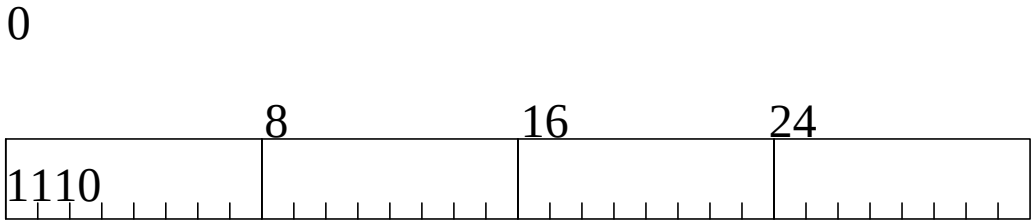
Outline

- Multicast
  - overview
  - Tunnelling vs native multicast
  - Multicast forwarding
- MSDP
- MBGP
- Implementation on GARR-B

Multicast

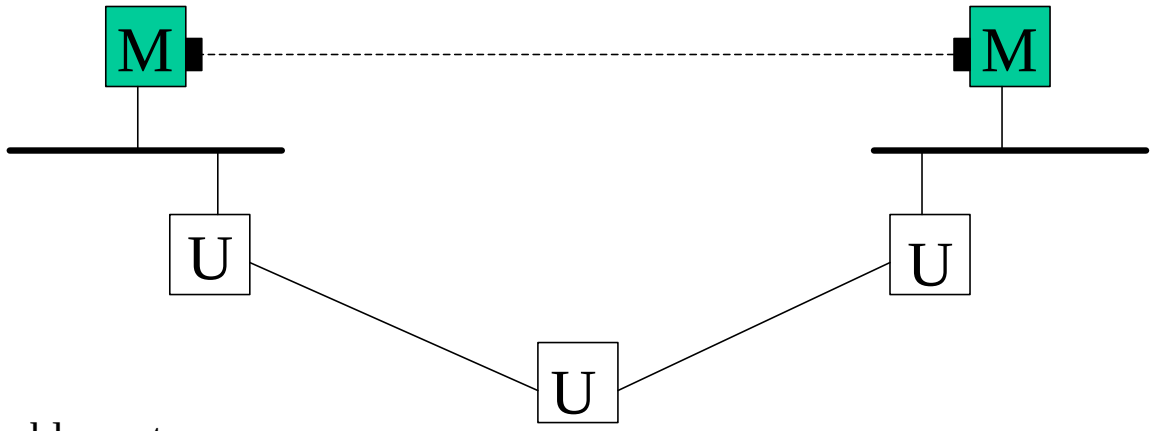
IP multicast: *transmission of a single data unit to multiple destinations (multicast group) which can be anywhere in the Internet, may join and leave dynamically*

- *Problem*: efficient utilization of transmission resources (e.g. link capacity); no duplication of datagrams
- *Applications*:
  - videoconferencing among multiple participants
  - mailing-list / news / software distribution
  - updates of replicated file systems or databases
- *Class D IP addresses*



Tunneling vs native multicast

- Tunnelling: mcast backbone is a virtual internet overlaid on the existing unicast infrastructure (mcast network = subset of unicast infrastructure)
- Native multicast: one-to-one correspondence with the unicast infrastructure
- *tunnels*: multicast virtual connections to bypass non multicast-capable routers

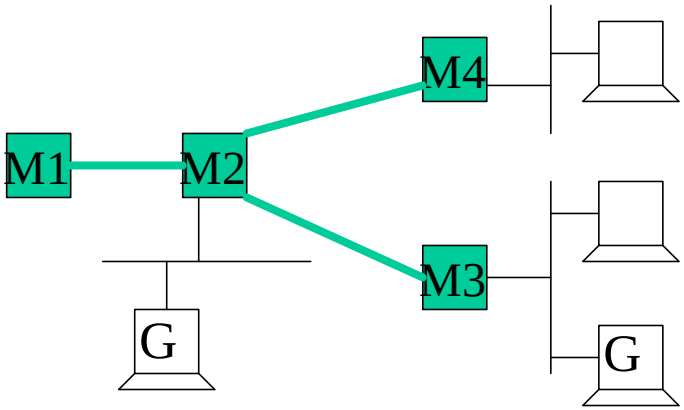


M Multicast-capable router

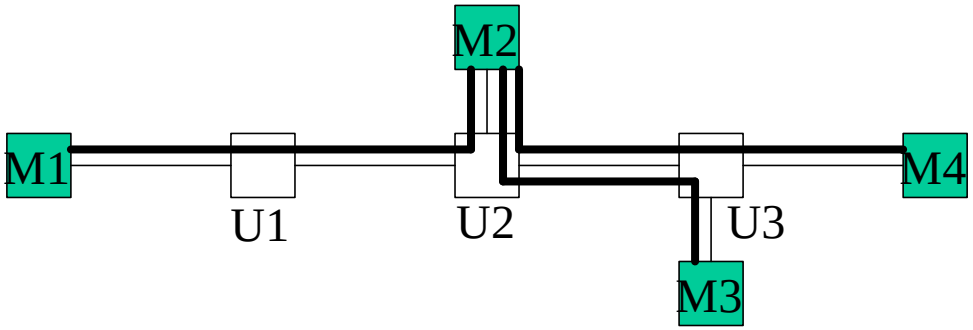
U Non multicast-capable router

Multicast functions

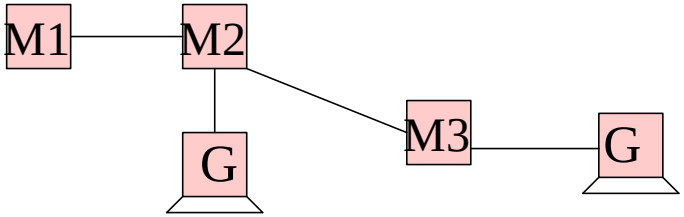
- Construction of the multicast tree for delivery of mcast packets
- multicast packet forwarding: unicast routing (optional, needed if the multicast topology does not match with the unicast topology)



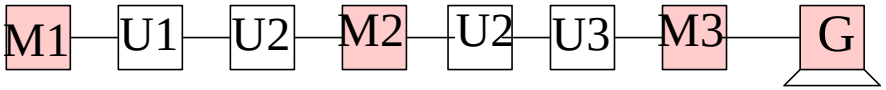
Multicast Topology (a)



Unicast Topology (b)



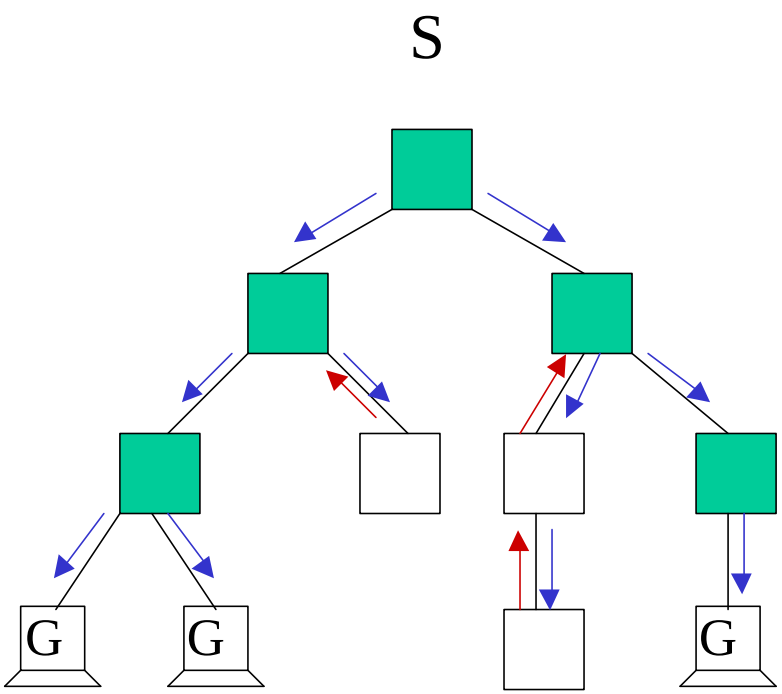
Multicast Tree(G) (c)



Multicast Packet Forwarding (d)

Approaches to Multicast Forwarding: 1. Spanning tree

- Truncated Reverse Path Forwarding (TRPF):

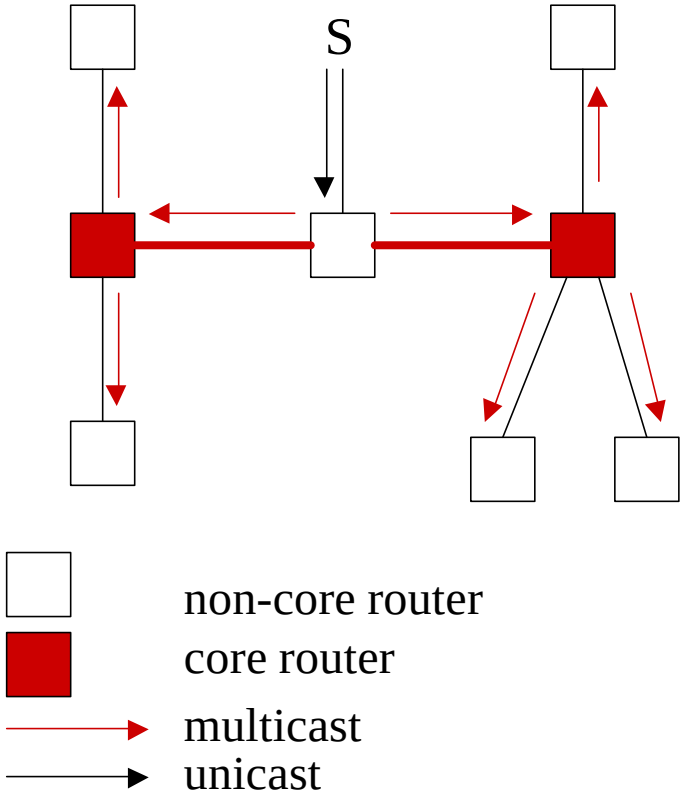


flooding

pruning

- multicast spanning tree's links are truncated if any leaf sub-network has no group members (IGMP) – pruning
- Pruning propagates upwards to the source
- Periodic flooding, poor scalability
- 1 source-rooted tree for each active sender
- DVMRP

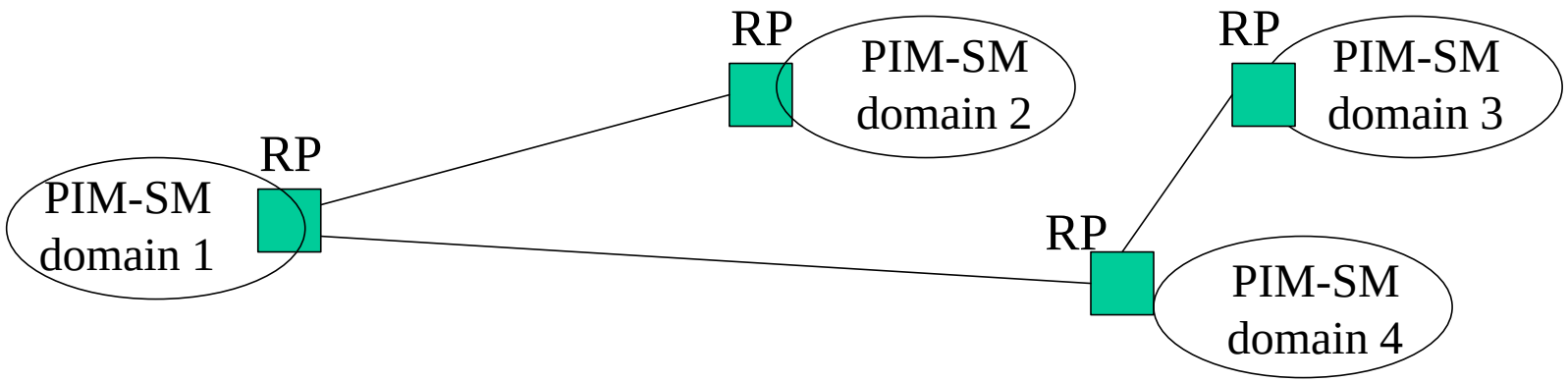
Approaches to Multicast Forwarding: 2. Core Based Trees (CBT)



- CBT: a single shared tree for any source transmitting to a given group G
- advantage: reduced amount of state information about (S,G) and forwarding interfaces
- single or multiple routers (fixed) define the core of the delivery tree
- hosts willing to be member of a group issue a *join* message forwarded towards the core
- scalability, no flooding
- PIM

MSDP

- Goal: discovery of multicast sources in different PIM-SM domains or within a single domain when multiple RPs for the same group are in use
- Overview:
  - connects multiple PIM-SM domains together
  - a domain uses its own independent RP(s)
  - no global source state is not required -no caching of Source Active (SA) messages- thanks to periodic transmission
  - TCP-based peering
  - one or more MDSP peers per domain
  - normal tree construction procedure applied in a inter-domain environment
  - MSDP peering congruent with BGP peering



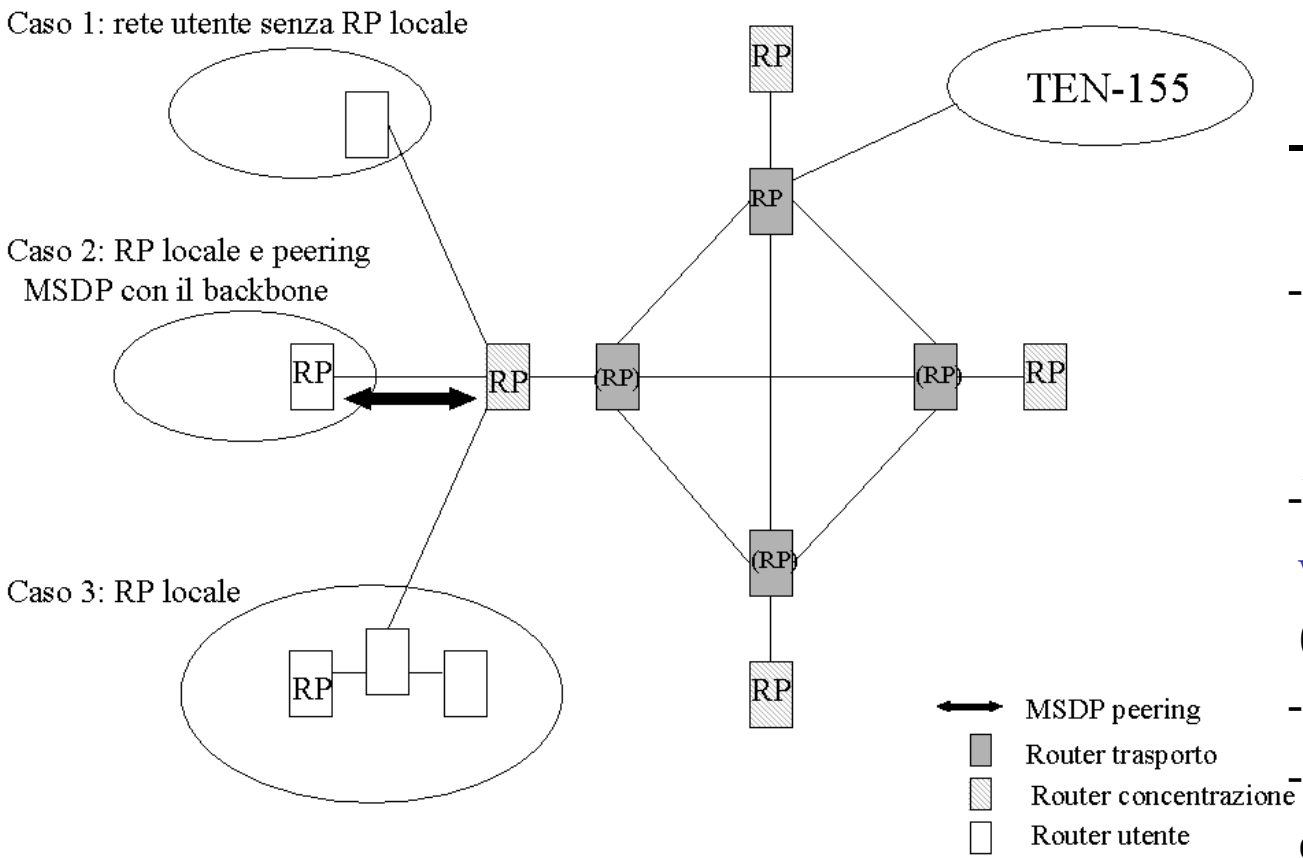
MBGP (RFC 2283)

- RFC 2283: Multiprotocol Extensions for BGP-4
- goal:
  - multicast routing policy throughout the internet
  - connection of within and between BGP autonomous systems
  - unicast topology different from multicast topology
- MBGP carries multicast routes: unicast + multicast (to build multicast data forwarding trees)
- NLRI: Network Layer Reachability Information
- exchange of multicast traffic only in a restricted set of points or deployment of links dedicated to multicast
- different unicast and multicast policies (e.g. in congruent unicast and mcast networks)
- MBGP mcast routing table used for RPF lookup
- restriction: MBGP clouds cannot be connected through BGP clouds

Multicast Routing

- Backbone:
  - PIM-SM domain
  - PIM-SM enabled on each router (RT, RC): native multicast
  - each RC is also an internal RP, MSDP peering between the RPs of the backbone
  - RT connected to national/international providers providing with multicast feed also need to be configured as RPs. They only announce GARR-B sources (GARR-B does not provide multicast transit to non-GARR domains)
  - backbone RPs serve the whole group range (by default)
  - automatic discovery of RPs
  - leaf PIM-SM domains can deploy RC as RP or deploy an internal RP (recommended solution for traffic optimization with multicast scoping)

RP Distribution



-PIM-SM native mcast domain (RTs, RCs)

-RC= internal RP, MSDP peering between backbone RPs

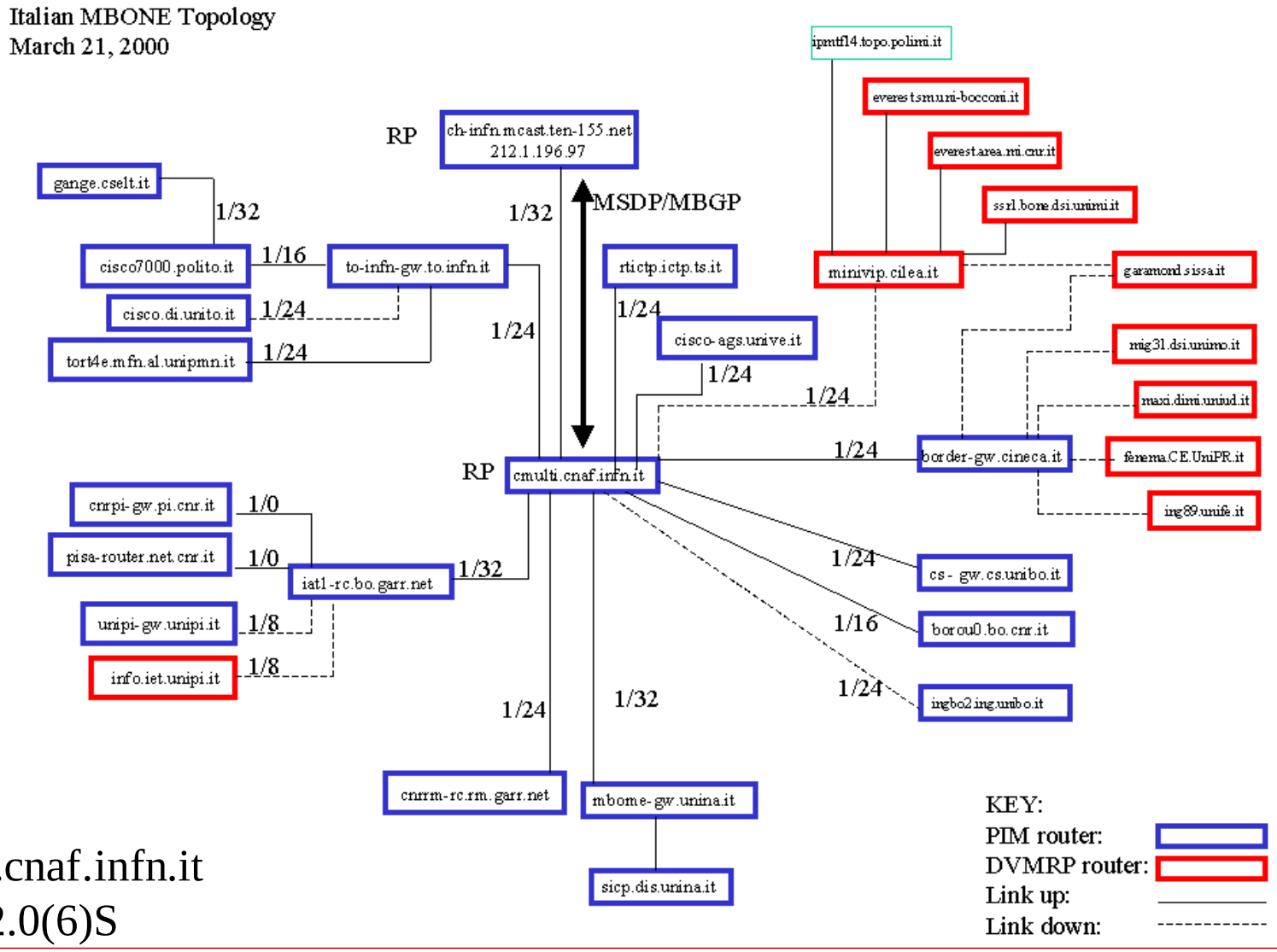
-backbone RPs serve the whole group range (by default)

-automatic discovery of RPs

-leaf PIM-SM domains deploy RC or internal RPs

-mcas MBGPt peering at RTs

Italian Multicast Infrastructure



cmulti.cnaf.infn.it
IOS 12.0(6)S

European Multicast Infrastructure