Towards Multicast Traffic Engineering?

Christian JACQUENET
christian.jacquenet@orange-ft.com

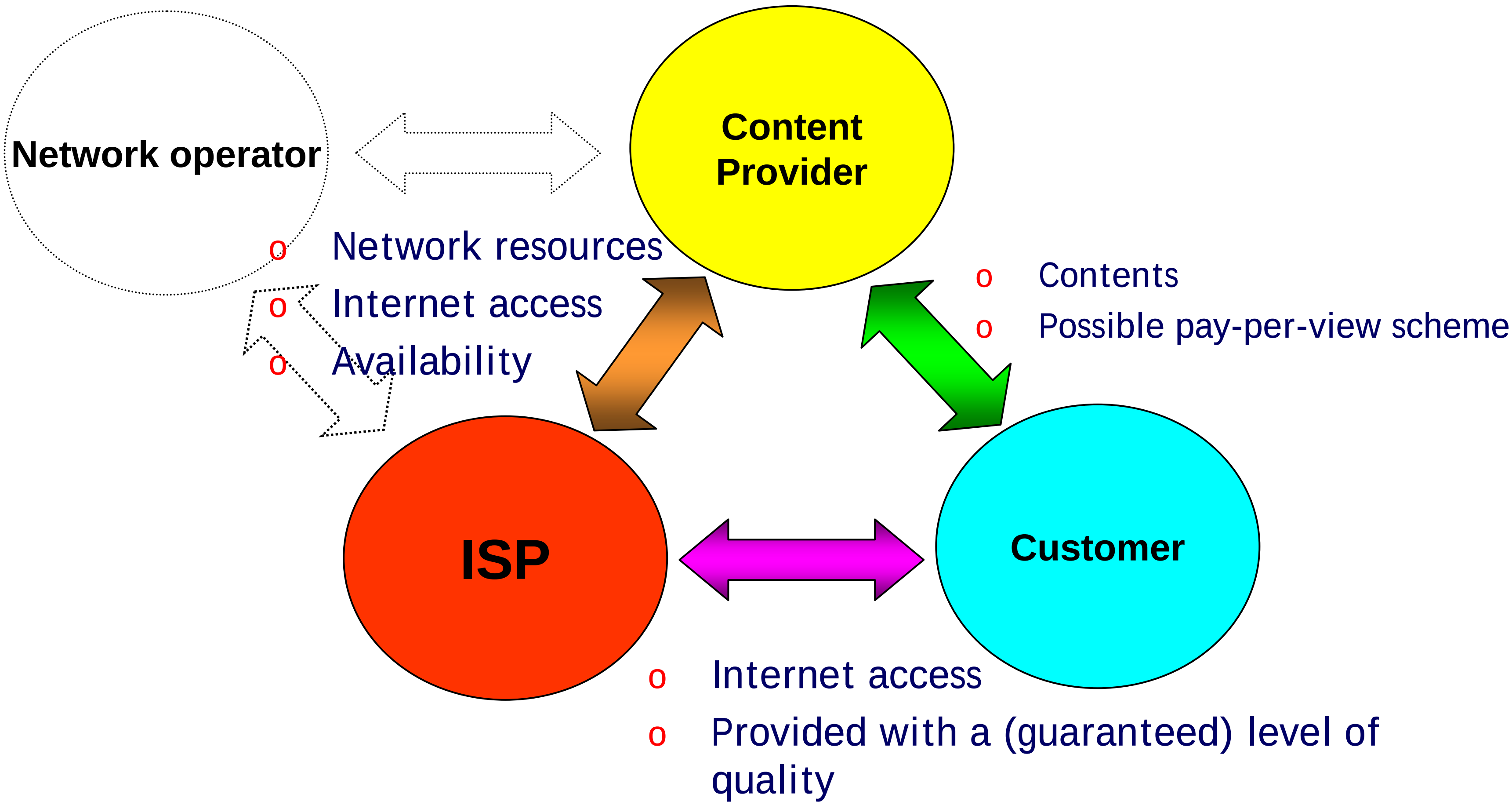
ITU-T IPTV Global Technical Workshop
Seoul, Korea, 12-13 October 2006

Outline

- Context and motivation
- IP multicast reminder
- Approaches to multicast-inferred traffic engineering
  - DiffServ and MPLS-TE capabilities
  - Design considerations
- Conclusion

ITU-T

Some Actors of Multimedia Services



Sample Figures

- Around 200 Gbytes of traffic per VOD server per week
- MPEG2-encoded TV traffic at ~4 Mbit/s per channel
  - Access any TV channel in less than 2s
  - Packet loss rate should not exceed 10^{-7}
  - A three 9's+ (99,97%) availability
- MPEG4-encoded HD TV is underway

Why Consider IP Multicast?

- Resource optimization
  - Moving from n x P2P communications to 1 x P2MP communication
    - Data replication is performed as closely to the receivers as possible
- Dynamics
  - Finer management of zapping behaviors
    - Tradeoff between IGP performance and state maintenance

Application Spectrum

- Tree-based multicast network design is somewhat application-dependent

1-to-n Group Communication — p-to-n Group Communication

SSM — PIM-SM — Bi-Dir PIM

Multicast Tree Taxonomy

- Rendezvous Point Trees (RPT), *a.k.a.* "shared trees"
  - A router of the multicast network is the root of the tree
  - Preferred design in case of numerous sources
    - Optimizes state maintenance in participating routers
- Shortest Path Trees (SPT), *a.k.a.* "source trees"
  - The source is the root of the tree
  - Preferred design for path optimization purposes
    - Assumes a few sources

PIM-SM Basics

- Requires a Rendezvous Point router to connect sources with receivers
  - By means of the dynamic establishment and maintenance of multicast distribution trees
  - Uses either RPT or SPT distribution trees
- Explicit Join model
  - Source sends Register messages towards the RP router
  - DR routers send Join messages towards the RP to connect receivers
  - Assumes no one wants traffic unless explicitly requested

ITU-T

QoS Issues

- IP multicast provides no means to:
  - Take into account the receivers' access capabilities
  - Provide some guarantees about the resources' availability
  - Gracefully handle congestion occurrences
- Reliability of multicast traffic forecasts is hard to achieve because of:
  - The dynamics of multicast
  - The behavior of receivers

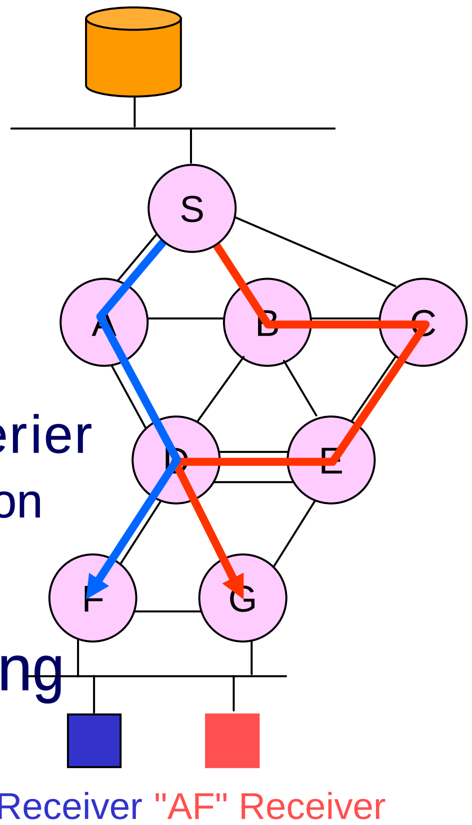
ITU-T

Resource Optimization

- Notify multicast sources
  - When they should start and stop transmitting
    - According to IGMP(v3)-based interest expressed by potential receivers
  - Make the subscription procedure evolve
    - For a better usage of the bandwidth resources
    - So as to enforce a subscription policy on a timely basis, for example
- Investigate traffic engineering techniques
  - To address path restoration issues
  - To comply with strict QoS guarantees
    - *E.g.* TV broadcasting services

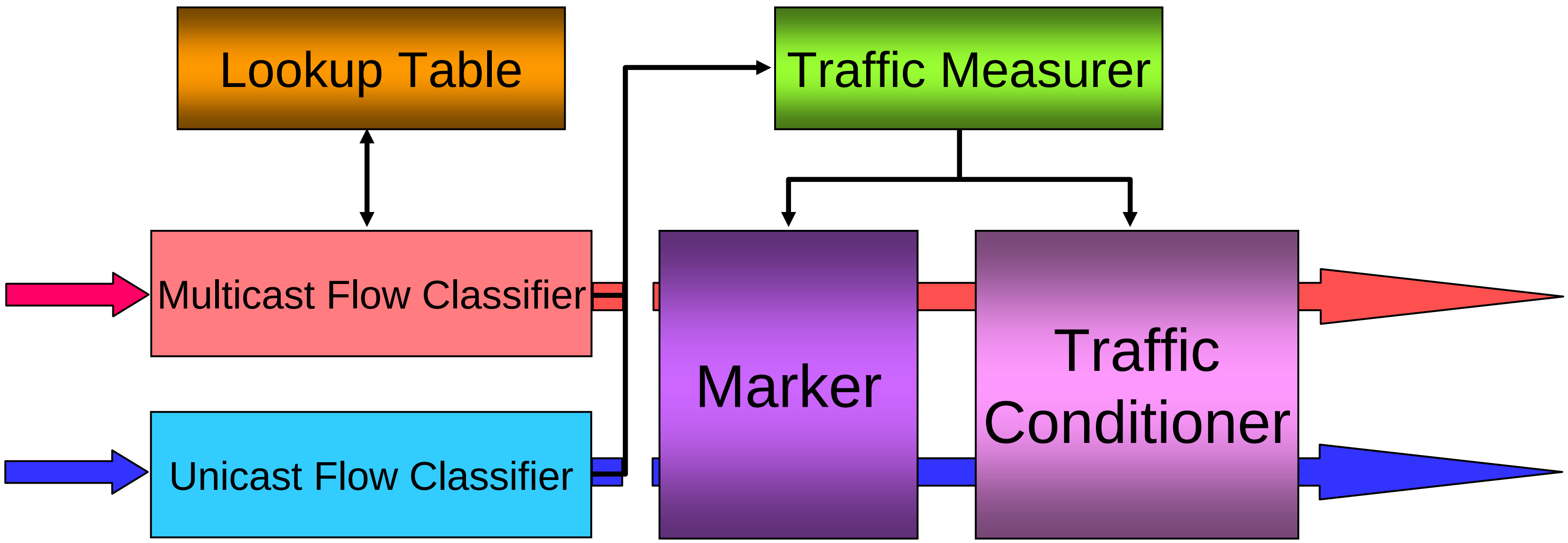
DiffServ-Inferred Distribution Trees

- Keeps the receiver-initiated principle
  - Adds signaling capability
    - Based upon IGMP traffic
    - To notify leaf routers about access rate conditions
  - Marking is enforced by aggregation devices (*e.g.* BRAS)
    - Information is processed by the IGMP Querier
      - As per the corresponding PHB configuration
- Mandates QoS-based routing
  - To preserve the QoS information during RPF checks
- Yields "colored" distribution trees
  - Core region is hopefully shared



Multicast Traffic Conditioning

- Remains a local treatment
  - Indicates a router how to behave in case of congestion
- No strict guarantee



DiffServ-Inferred Tables

- When multicast traffic enters the DiffServ domain:
  - Processing done by the access router
  - Then enforces the traffic conditioning policy
    - Multicast traffic volume can be estimated as many unicast flows as receivers

Group Address	PHB	Weight
233.12.144.2	EF	1
	AF_1	3
	AF_2	2
233.12.144.4	EF	1
	AF_2	2
233.12.144.25	AF_1	1
	AF_2	2

Investigating Traffic Engineering

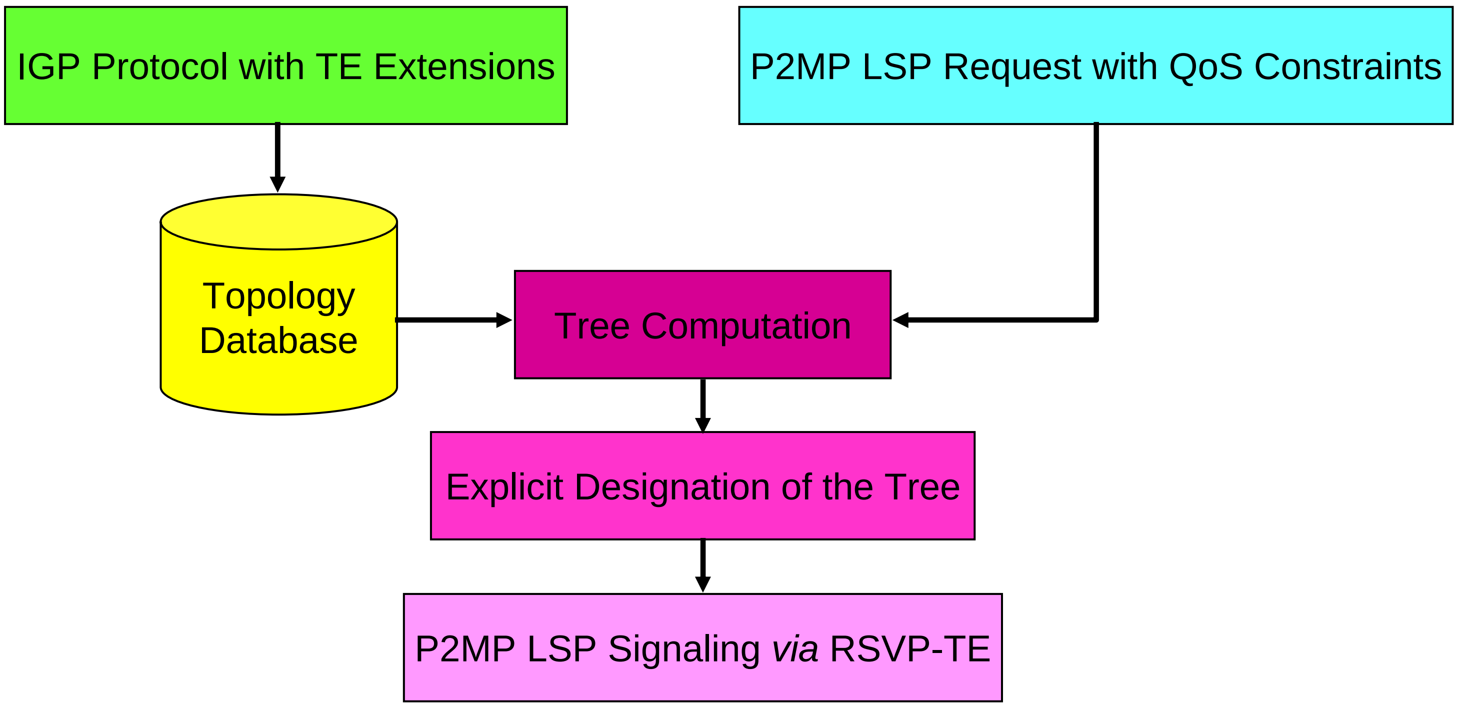
- Compute and select paths to accommodate:
  - Network-wide QoS requirements
  - Application-specific constraints
    - One-way transit delay, inter-packet delay variation, packet loss rate, *etc*.
    - Restoration capabilities
  - Network planning considerations
- Provision resources accordingly
  - Towards automated configuration processes
- Make sure the TE policy is not CPU-intensive
  - Switching performances of the participating devices
  - Volume of solicitations per unit of time

MPLS as the Natural Candidate

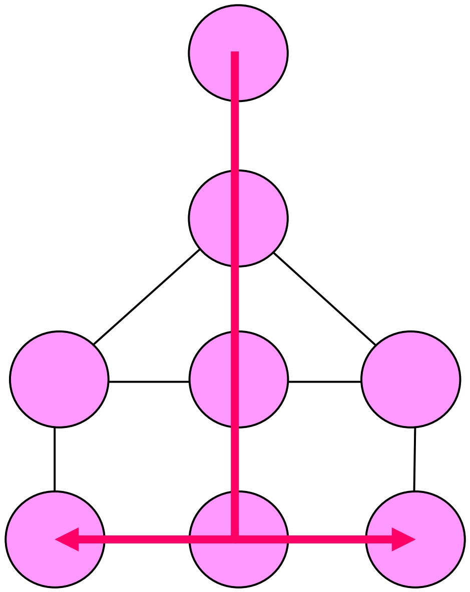
- Implicit and attractive traffic engineering capabilities
  - IGP-inferred LSP path computation
  - TE extensions of RSVP with no scalability issues
  - Fast recovery mechanisms for both link and node protection
- Commercial and operational implementations
  - Mostly for Fast Re-Route purposes
  - Claimed deployments of DiffServ-aware MPLS TE

Basic Principles of Point-to-Multipoint Traffic-Engineered LSP Paths

- The IGP conveys information about routers' capabilities
- P2MP LSP paths are established as per cost considerations

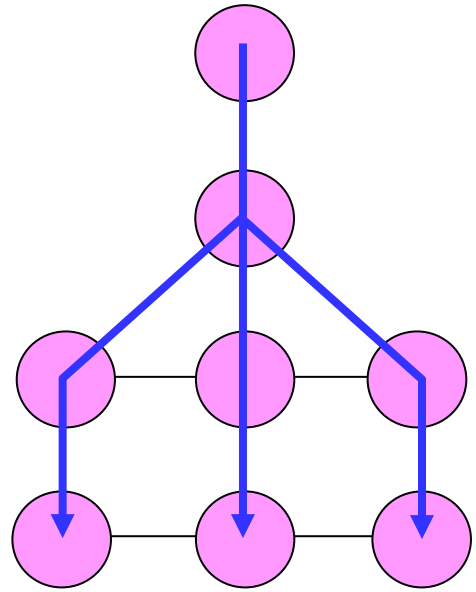


Lowest Cost Trees *vs*. Shortest Path Trees



Steiner tree (low cost):

1. Cost = 5
2. Path cost = 4



Shortest Path Tree

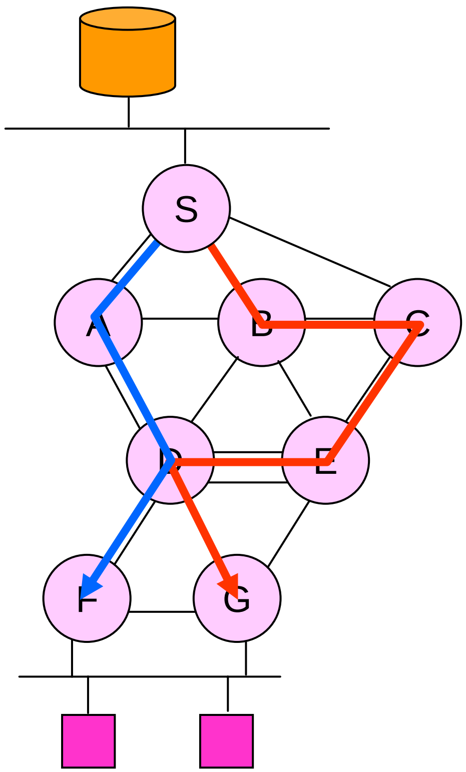
1. Cost = 7
2. Path cost = 3

Tree Computation Options

- "Root" router-centric dynamic computation
  - Requires an adapted CSPF algorithm
  - May be CPU-intensive and yield performance issues
- Dynamic computation performed by an external entity
  - Notion of Path Computation Element (PCE)
    - Downloads the subsequent configuration information to the "root" router
    - Suggests a hopefully automated procedure
  - May yield robustness issues in case of frequent solicitations

P2MP LSP Establishment

- RSVP-TE-based:
  - "Source" router sends RSVP_PATH message that includes:
    - Sub-LSPs towards F and G
  - A sends RSVP_PATH message with sub-LSP towards F
  - B sends RSVP_PATH message with sub-LSP towards G
  - *Etc*.
- Leaf routers (F, G, D, E,...) send back RSVP_RESV messages to "source" router (S)
  - A and B send a single RSVP_RESV message towards S that compiles all the sub-LSP-specific information



State Maintenance

- Technically, the number of RSVP states grows as the group subscription dynamics
  - Yields scalability and performance issues
- Practically, RSVP states could be de-composed into "sub-states":
  - Each sub-state corresponds to a subset of branches and leaves
  - Each sub-state can be dynamically refreshed independently
  - The global state is updated as per the grafting process

ITU-T

One P2MP LSP per Service

- Yields as many tree structures as (S, G_i) pairs
  - Dramatically increases the number of states to maintain
  - Addresses finer resource optimization concerns
- Well-suited for application-specific QoS requirements
  - *E.g*. VoD service and a few TV channels to broadcast
- Poorly adapted to massive deployments
  - *E.g.* several bouquets with hundreds of TV channels
  - Inevitably raises operational and scalability concerns

One Single P2MP LSP per Source

- Yields the maintenance of a single (*, *) Forwarding Equivalence Class (FEC) per source
  - Dramatically reduces the number of states to maintain
  - Network resource optimization is poorly addressed
- Comparable to RPT tree structures in IP multicast
- Well-suited for massive deployments
  - *E.g*. several bouquets with hundreds of TV channels
- Poorly adapted to application-specific QoS requirements
  - No means to enforce traffic prioritization within the tree structure

(S, G_j) Channel Aggregation into a Given P2MP Tree Structure

- A FEC corresponds to a set of (S, G_j) pairs
- Presumably an attractive tradeoff, where:
  - Several bouquets of hundreds of TV channels need to be broadcast
  - Service subscription policy relies upon a thematic typology
    - Movie channels, documentary channels, *etc*.

Pending Issues

- RSVP-TE is not suited for high dynamics
  - Context of frequent leaf addition or removal of the tree structure
- RSVP-TE raises scalability issues when the number of PE routers is high
- RSVP-TE requires more states than the LDP approach

So, What's Next?

- "Triple-play" services are key drivers for the:
  - Design and enforcement of QoS policies in access regions
    - Never-ending over-provisioning is not an option
  - Deployment of multicast-inferred transmission schemes
    - TV broadcasting is becoming more than just another hype
- DiffServ mechanisms remain an attractive option to deal with (probable) congestion
- Multicast-inferred MPLS TE capabilities are exciting, assuming:
  - Dynamic computation and configuration features
  - An adapted multicast packet forwarding scheme
    - To better address dynamics of subscription schemes