IP Multicast/Multipoint for IPTV (and beyond)

Toerless Eckert, eckert@cisco.com

Agenda

- Deployment strategy
  - Multicast/multipoint for IPTV or vice versa ?
  - Usage of IP multicast for further broadband svc.
- An architectural overview
- IP multicast primer
  - SSM for IPTV
- Transport network architecture
  - Native multicast, MPLS (mLDP, RSVP-TE P2MP)
  - VPNs and L2
- Resiliency
  - Source redundancy, fast convergence,
  - FRR, path separation, application side
- Path selection
- Admission control
- Channel changing
  - Join/leave latency, static/dynamic forwarding, acceleration

Broadband, IPTV and IP multicast
…where should the network go ..

Broadcast IPTV = IP multicast

…however transport network transits packets ..

- "Native IP multicast", MPLS, L2, optical

IP multicast sources:

- Encoder, Transcoder, Groomer, Ad-Splicer, …

IP multicast receivers:

- Transcoder, Groomer, Ad-Splicer, QAM, STB

IP == IPv6 (Japan) or IPv4 (RotW *rest of the world*)

- No address exhaustion issue (SSM)
- No/slow move to IPv6 for IPTV in RotW

Strategic direction
"Traditionalists"

- Build/extend a network for the application
- Application = traditional broadcast TV
- IP multicast for only broadcast IPTV
  - Not for any other services
- IPTV important, but why always TTM *sigh*
  - IPTV requirements in 2003 ??
- Result often
  - Make IPTV network look and behave as much as possible as legacy cable or other non-IP network.
  - Short term (initial phase) solutions view

Strategic direction
"IP multicast evangelists" (me)

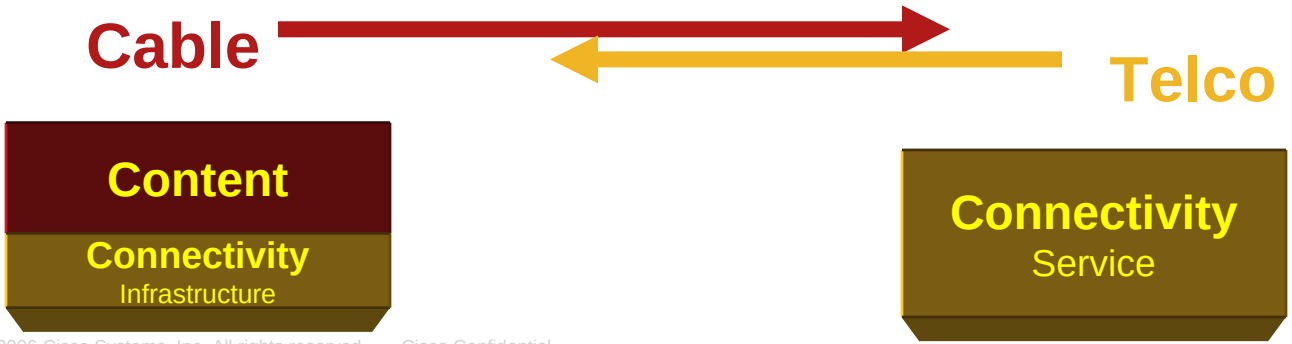
- To network operations branch
  - Add IP multicast to your network
  - Choose transport methods based on SLA and operational requirements/preferences
    - Native IP multicast, MPLS, L2, mix
  - Solution should minimize involvement in provisioning of individual applications/services
- To IPTV services branch
  - Start with traditional broadcast TV
  - Investigate extending IPTV and other (IP multicast) services
    - More RoI on investment

Strategic direction
more use / need for IP multicast

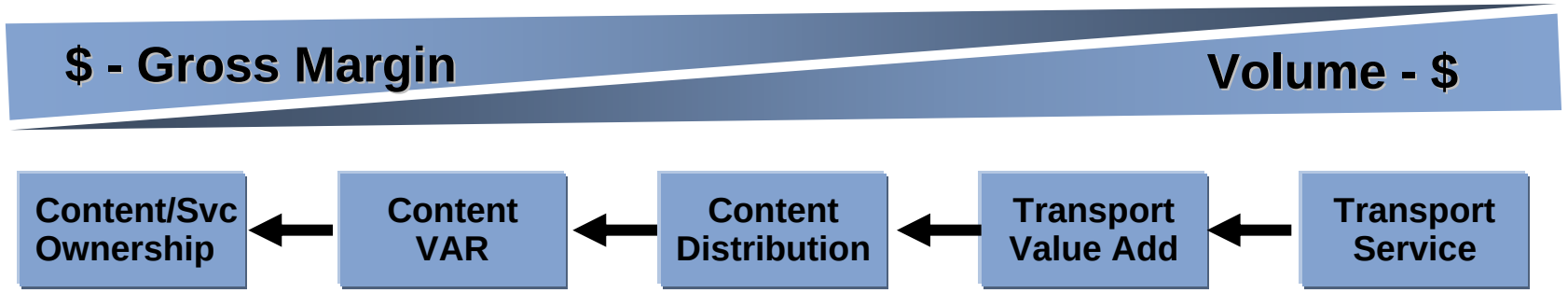
- Extending IPTV broadcast service
  - Dynamic redundancy (regional to national)
  - Variety of reach of transmission (src->rcvr)
    - Groomer/transcoders, Add-Splicers
  - Switched Digital Video, oversubscription
  - Wholesale, dynamic, international channels
- Other services
  - Commercial (MVPN)
  - Content preprovisioning to VoD server, STB
  - Multicast in Internet Service (eg: To PC)
  - Voice conferencing, gaming, surveilance, …

Strategic direction

- Driver: Broadband edge speed
  - Cable: Wideband DOCSIS (3.0) >= 100 Mbps
  - Telco: ADSL -> ADSL2 -> VDSL ~ 50 Mbps
  - FTTH (xPONS)
- Multitude of services (triple-play++) causes wide scale NGN design discussions
  - Traditional services will change, but slowly!



Business models
For transport and (content) services



Walled garden, All in one	*Content Management*	*Improved service to sol. provider*	*Improved service to sol. consumer*	*Service/Cont agnostic*
Exclusivity	Sol/Serv. Particip./ Operations	Extranet	QoS/SLA	Internet
Content and network ownership	Content acquisition/ mgmt/ conversion	Account Collect/ resell	Address privacy	Leased lines
		Transport VA for sol/cont provider	Security	ATM
			Multicast	Frame-relay
Telephony Excl. TV	IPTV/ Market Data	Market Data	VPN	"The Web"

Open for exploration: Pull model IPTV

- Previously known as “Internet Multicast”

  IPTV today: Push-model / walled-garden:
  Marketing in IPTV access provider decides viewers lineup!

- Viewer wants to decide what content to consume

  Web delivered pull model in 1995 – unicast only

  VoD Pull model IPTV quite hype – Youtube,… (HDTV ???)

  Live streaming/effective preprovisioning ? → IP multicast

- Variety of business mode/technical choices, eg:

  Free or add-priced option to Internet access

  Per-usage billing/revenue splitting
  (“900” number model – AP/CP)

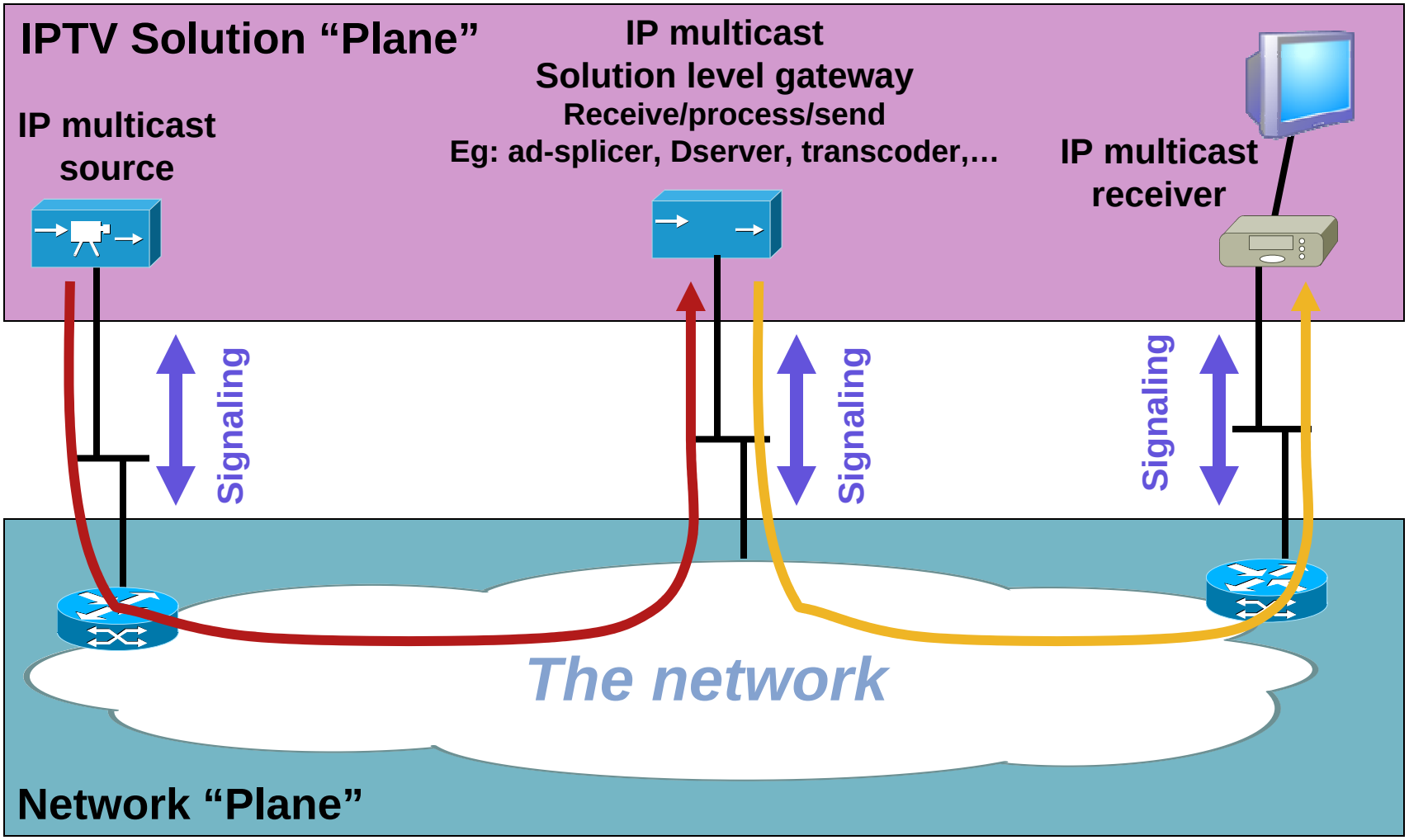
  Over-The-Top eg: from core ISP:

  AMT (automatic Multicast Tunneling)

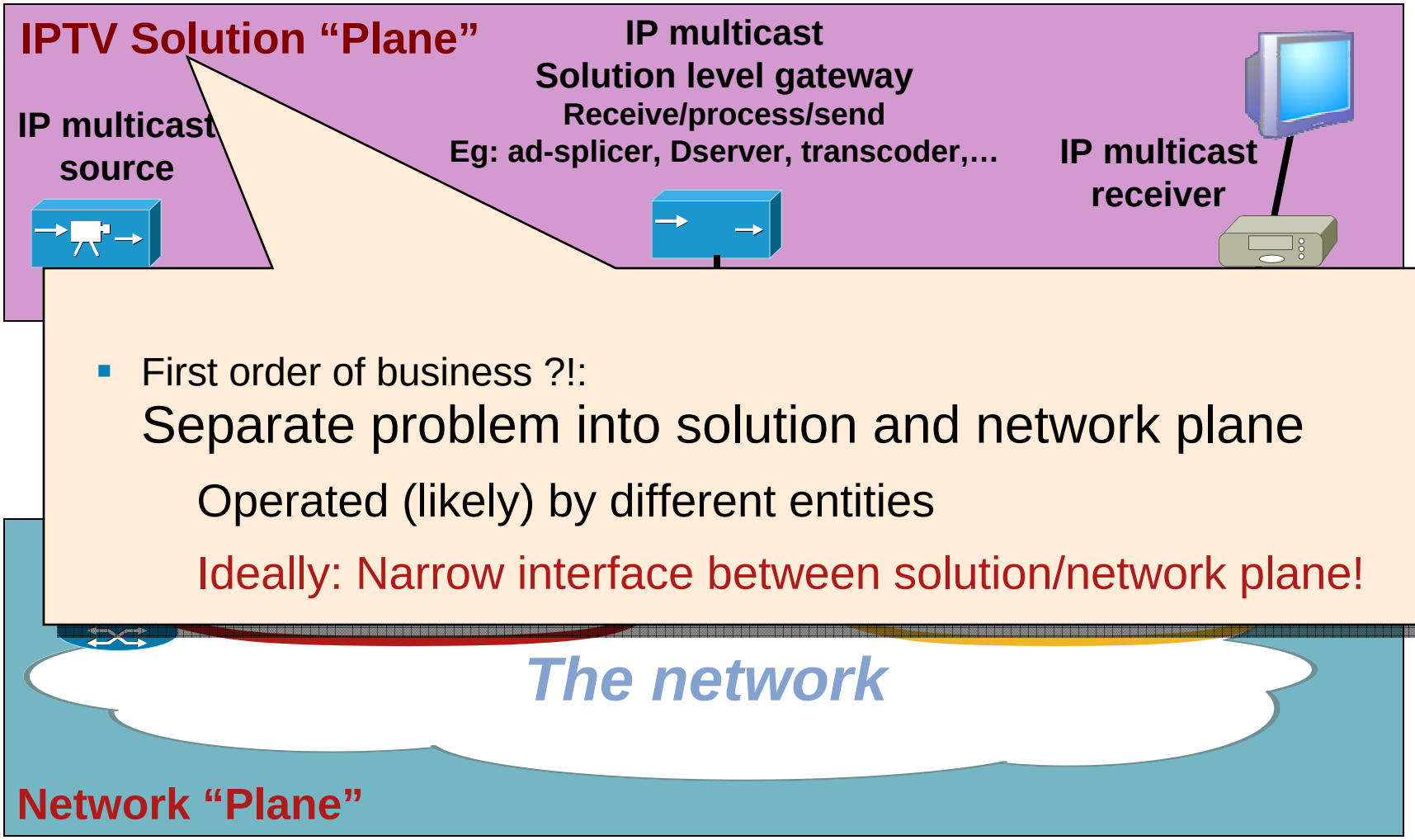
  Eg: Tunnel across non-multicast access provider/edge

An architectural overview

50,000 feet architecture

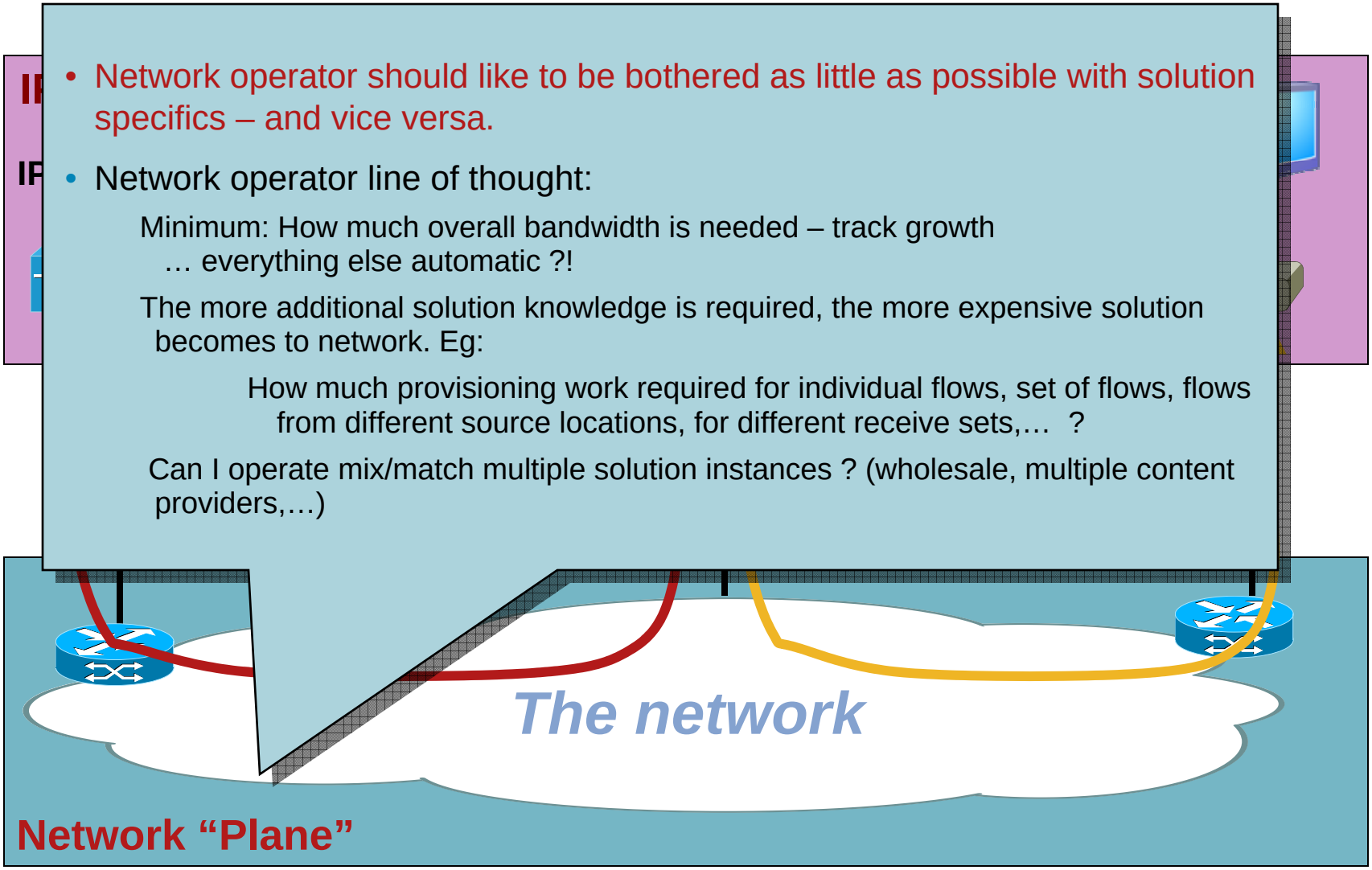


50,000 feet architecture

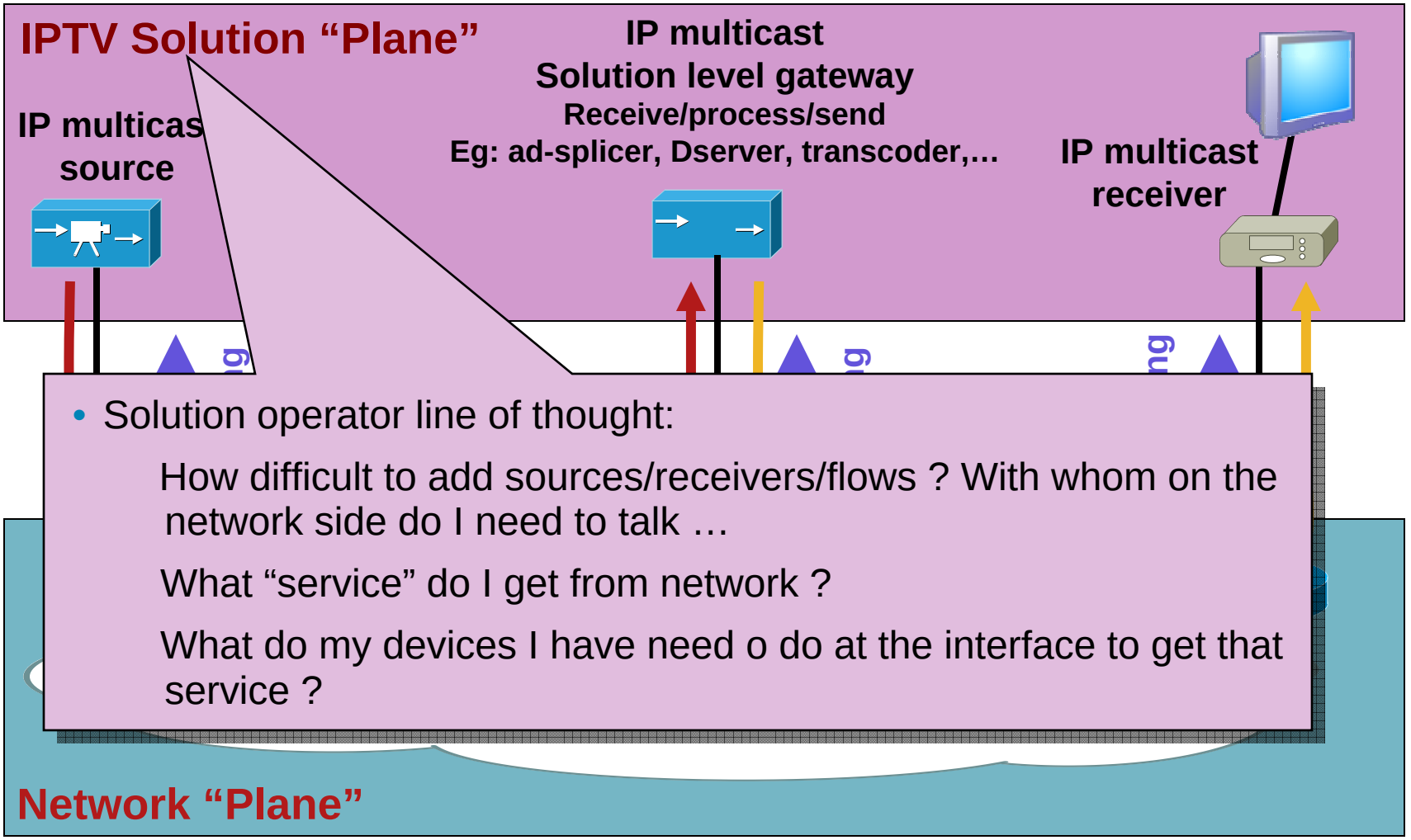


- First order of business ?!:
  Separate problem into solution and network plane
  - Operated (likely) by different entities
  - Ideally: Narrow interface between solution/network plane!

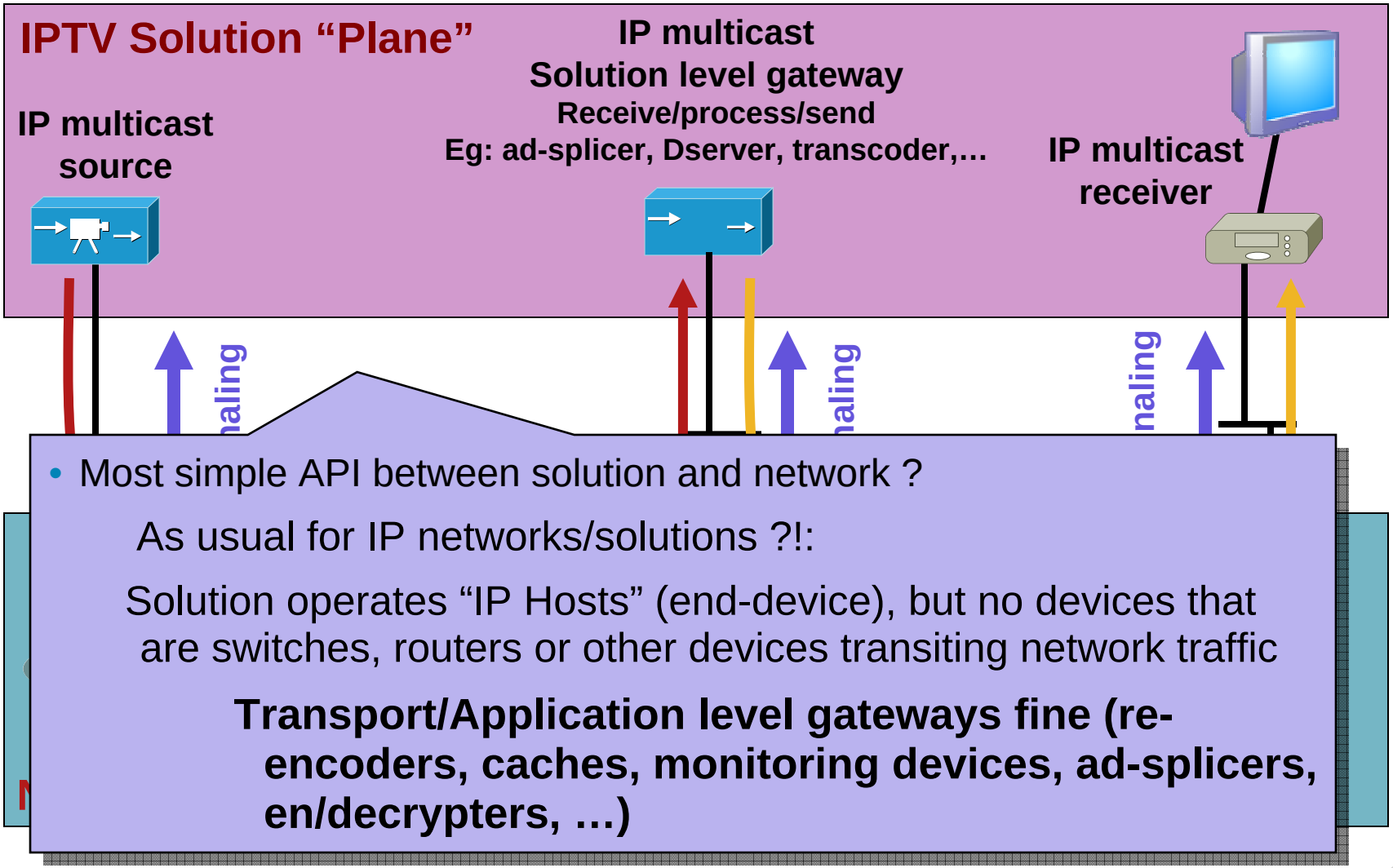
50,000 feet architecture



50,000 feet architecture



50,000 feet architecture



- Most simple API between solution and network ?

  As usual for IP networks/solutions ?!:

  Solution operates “IP Hosts” (end-device), but no devices that are switches, routers or other devices transiting network traffic

  Transport/Application level gateways fine (re-encoders, caches, monitoring devices, ad-splicers, en/decrypters, …)

Selling the network to the solution

What functions can the network provide (1) ?

- P2MP = SSM trees
  - Build trees from any individual source.
    - Inject everywhere, receive everywhere
  - Best join/prune latencies
    - *Warning: fast network join is not same as fast solution join*!
  - Largest #trees supported,…
  - No coordination of tree addresses (SSM channels)
  - No spoofing of traffic across the tree
- Redundancy
  - Anycast/Prioritycast for source redundancy
  - Path separation for solution layer dual-path-redundancy
    - for up to 0 packet loss during network failure

Selling the network to the solution

What functions can the network provide (2) ?

- Reliability
  - *Any DiffServ QoS class (as also IP unicast)*
  - Network failure recovery
    - Fast reconvergence (low 100 msec .. Sec)
    - Sub 50 msec (with protection and/or fast reroute)
    - *Warning:*
      Benefit of failure recovery highly solution dependent
    - Node-HA, APS, …
- Solution layer reliability
  - Path separation
  - FCC, other reliable multicast

Network sales promo to solution

What functions can the network provide (3) ?

- Admission control
  - Per-flow bandwidth based admission control
    - RSVP / workarounds (IGMP/mroute limits)
- More …:
  - (per subscriber) access control (eg: lineup), provisioning of subscriber policies, …
  - Accounting (Radius, Netflow, …)
  - Management, troubleshooting
  - Not further covered in this presentation
    Lots of product specifics

Requirements against solution
What API functions solution devices need:

- Mandatory:
  - SSM-tree building: IGMPv3/MLDv2 with SSM ‘joins’
    - receivers needs to know (S,G) channels to join
  - Send multicast packets with TTL > 1 ;-))
- Optional:
  - Signaling for source redundancy
  - Send/receive(mege) dual streams (for dual-path-redundancy)
  - RSVP – for admission control
- All host side signaling have workarounds that could be configured at network equipment, but the more of those are used, the more the network need to know exactly which multicast flows are required.
  - Eg: static building of multicast trees.

(Solution invisible)
functions within the network

- Preferred choice of transport:
  - IP (native multicast/PIM) or MPLS (mLDP and RSVP-TE P2MP)
- **Path selection**
  - (dual path) – exposed to solution
  - Cost optimization - why ?
  - Load-splitting:
    - ECMP: PIM and mLDP
    - Arbitrary: RSVP-TE (CSPF)
- **Preferred choice of virtualization**
  - L2VPN, L3VPN context – or why not…
- …not complete list

IP multicast primer (SSM)

… as required for IPTV…

Protocols and Services …and IP multicast

- multicast / multipoint *protocols*

  Between routers, switches, ..

  "Only of interest to network operator"

  PIM-SM, MSDP, (M)BGP, AutoRP, BSR, mLDP, RSVP-TE, …), IGPs (OSPF, ISIS), …

- multicast *services*

  How end-devices can use IP multicast

  "Of interest to network and service operator"

  ASM, SSM (and protocols "IGMP/MLD")

  Service operator just need to add SLA requirements!

IP multicast services

- ASM: “Any Source Multicast” (1990, rfc1112)
  - The “traditional IP multicast service”
  - Sources send packets to multicast groups
  - Receivers join groups, receive from any source
- SSM “source specific multicast” (since 2000, rfc4607)
  - Source Specific Multicast
  - Sources send packets to multicast groups
  - Receivers subscribe (S,G) channels, receive only traffic from S sent to G
  - *Primarily introduced (by IETF) for IPTV type services*
    - *Because of limitations of standard (protocol) model for ASM*

Standard protocol model for ASM

- What is the standard protocol model ?
- A1: MBone and DVMRP

  Please go back to your time machine and dial 1994

- A2: Native Multicast with:

  PIM-SM

  AutoRP, BSR or MSDP/Anycast-RP redundancy

  MSDP for Interdomain support

  Multiprotocol BGP for interdomain RPF selection

  Best available general purpose ASM protocol suite

  …but with issues

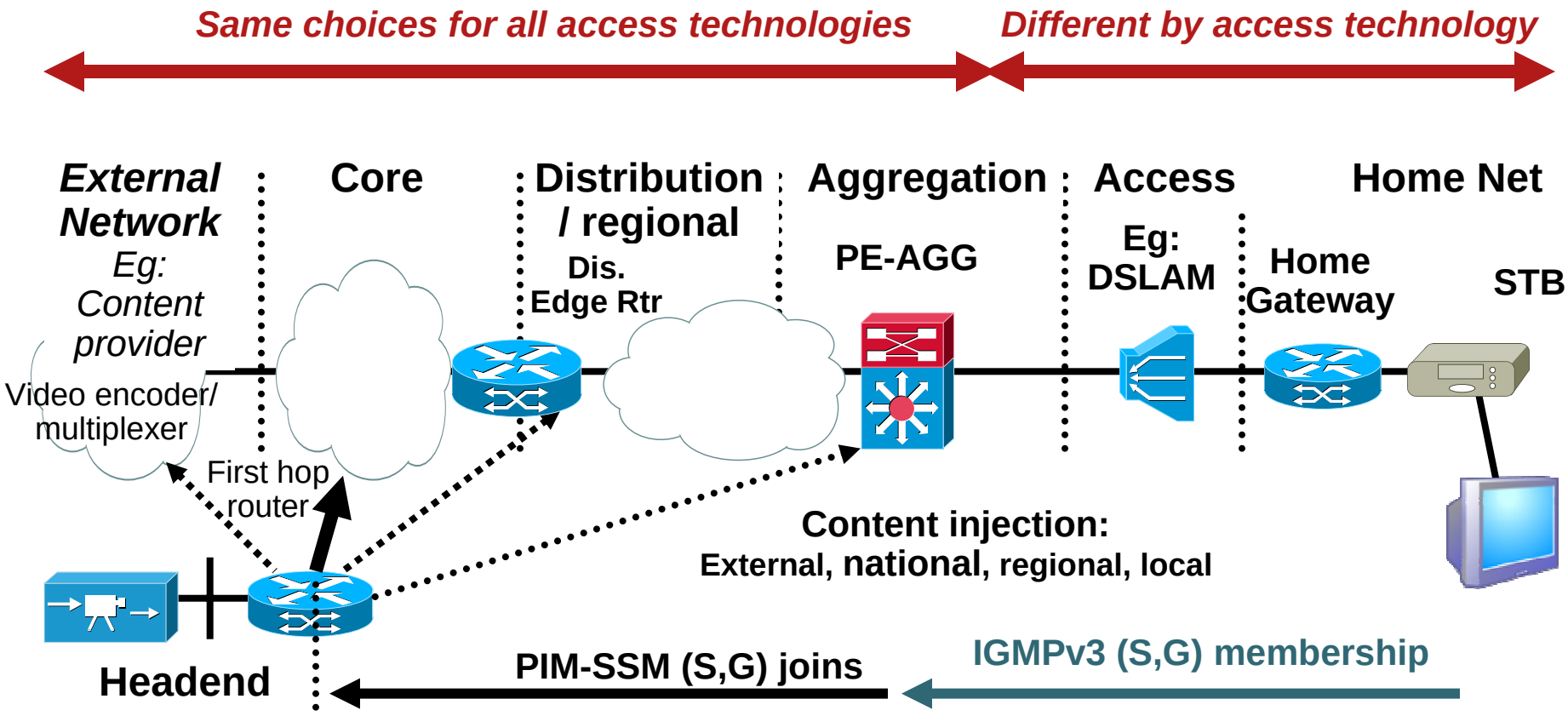
IP multicast services
Issues with ASM – resolved with SSM

- ASM
  - No traffic from unwanted sources
  - Address allocation (IPv4 only, not IPv6)
- Standard protocol suite
  - Complexity of protocol operations required
    - PIM-SM (RPT+SPT+Switchover), RP redundancy, announce, location
    - MSDP (RPF), BGP congruency,
    - Interactions with MPLS cores, bandwidth reservation, protection
  - Scalability, Speed of protocol operations (convergence)
    - RPT + SPT operations needed

Standard protocol model for SSM

- IETF
  - Receiver host to router (eg: IP-STB)
    - IGMPv3(IPv4) / MLDv2(IPv6) with (S,G) signaling
    - MUST be supported in host stack and host middleware (app)
  - Between routers
    - PIM-SSM == subset of PIM-SM for SSM (nothing new!)
    - IGMPv3 proxy routing / (snooping) on HAG, L2 access
    - Simple point to multipoint tree building == (S,G) SPTs only
- Cisco, (IETF ?)
  - Source redundancy (option – other options too)
    - Anycast/Prioritycast source addresses with eg: RIPv2 signaling

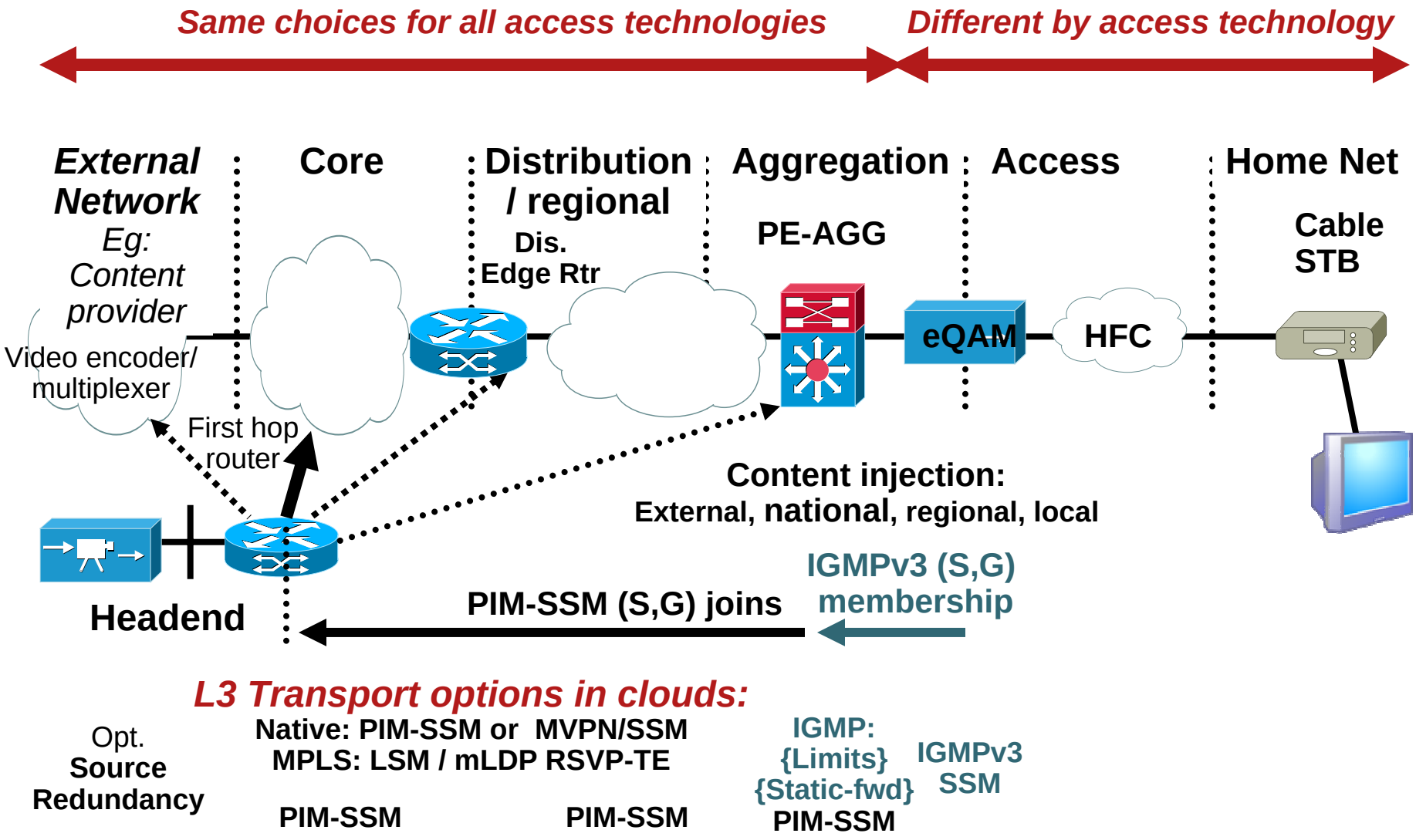
End-to-end protocol view



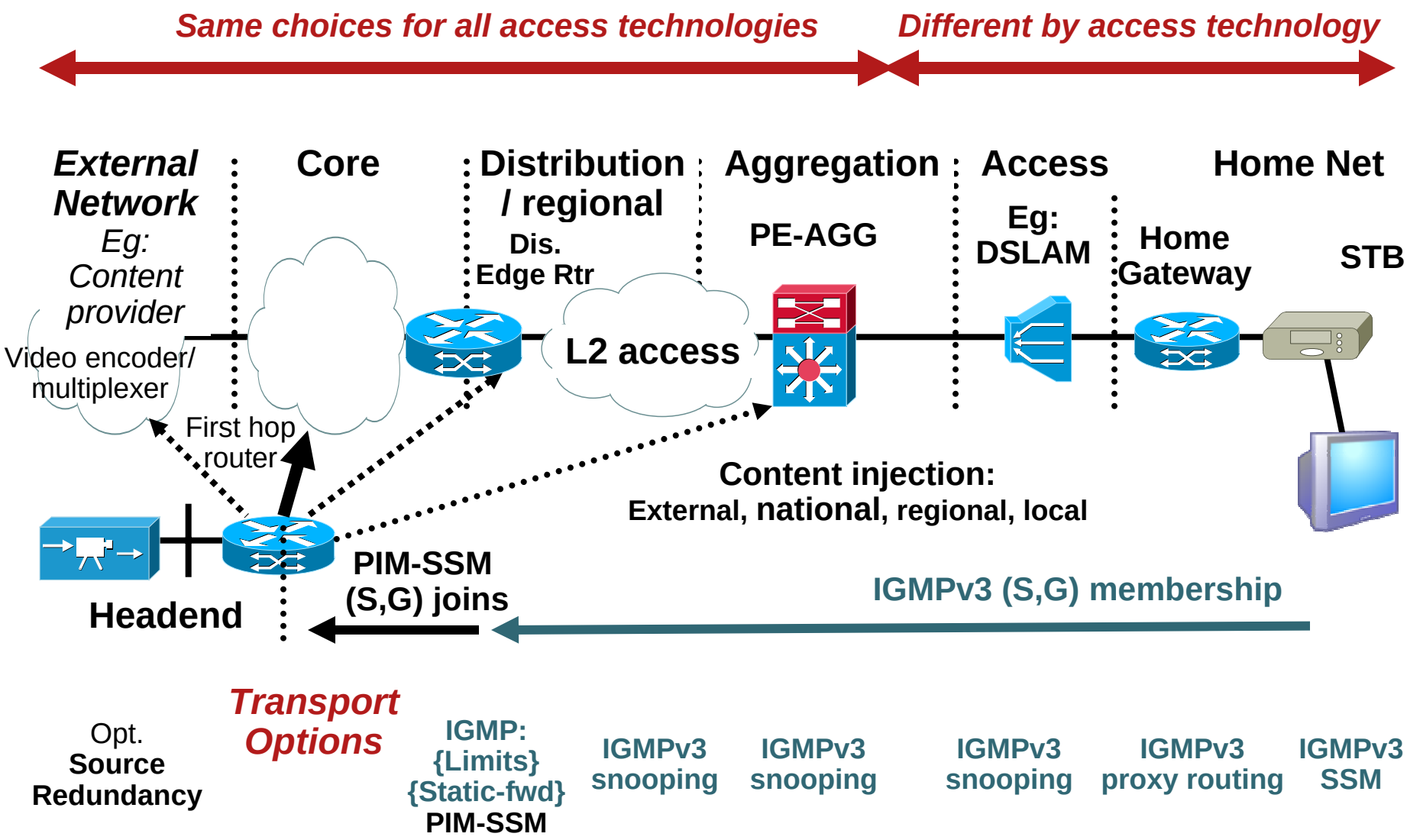
L3 Transport Options in clouds:

Opt. Source Redundancy	Native: PIM-SSM or MVPN/SSM MPLS: LSM / mLDP RSVP-TE		IGMP: {Limits} {Static-fwd} PIM-SSM	IGMPv3 snooping	IGMPv3 proxy routing	IGMPv3 SSM
	PIM-SSM	PIM-SSM				

End-to-end protocol view
digital (non DOCSIS) cable



End-to-end protocol view
with L2 distribution



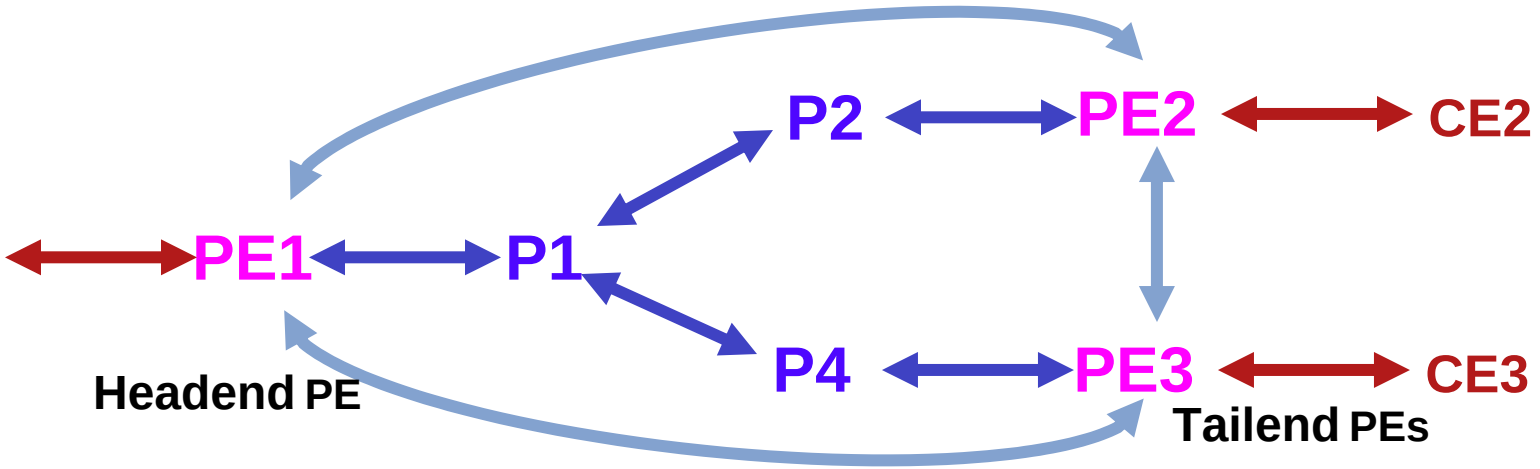
(Transit) Transport design alternatives

Transport architecture Overview

- Three main new directions / challenges
  All potentially orthogonal to each other (yikes).
  - In addition to IP multicast (PIM-SSM):
    - Support for MPLS multicast
      - Build P2MP (or also MP2MP) label switched delivery trees
      - RSVP-TE/P2MP and mLDP
  - Put traffic into a VPN context
    - As a method of service isolation / multiplexing
  - Using L2 vs. L3 on PE nodes
    - To “integrate” better into an L2 service model
    - (Mostly Telco) – claimed to be “simpler” service

Elements of transport architecture

- User-side protocols
  - IGMPv3 or PIM-SSM for IPTV
- Core Hop-by-Hop tree building
  - PIM-SSM (native multicast), mLDP, RSVP-TE (MPLS)
- PE functions
  - Put user traffic into VPN (L2, L3) or not (“native”)
  - Route (L3) or bridge traffic into core tree
  - Build core tree based on user side IGMP/PIM – or static
  - May require additional PE-PE signaling protocols



Combinations with L3 on PE
native IP multicast in core

- "Native IP multicast"

  Most simple: PE only uses uses PIM-SSM, no additional PE-PE signaling required.

  Use "RPF-Vector" for "BGP free core"

- "MVPN"

  Carries traffic across rfc2547 compatible L3 VPN.

  Can still rely on only PIM-SSM in core (option)

  Well established, widely deployed solution (at least dual vendor):

    Rfc2547 BGP (+ optional extensions)

    GRE encap on PE

    Default-MDT PE-PE signaling (I-PMSI)

    Other extensions: Inter-AS support, extranet, Data-MDT,…

Combinations with L3 on PE
with mLDP (possible futures)

- mLDP native
  - mLDP P2MP trees build pretty much like PIM-SSM trees
  - No additional PE-PE signaling required
    - Just standard IPv4 BGP on PE
- mLDP “Direct-MDT” in VPN context
  - Exactly like mLDP native! – just rfc2547 BGP
  - No “MVPN” or similar signaling required
- mLDP “MVPN”
  - Exactly like MVPN signaling
  - Just replaces PIM-SSM+GRE with mLDP

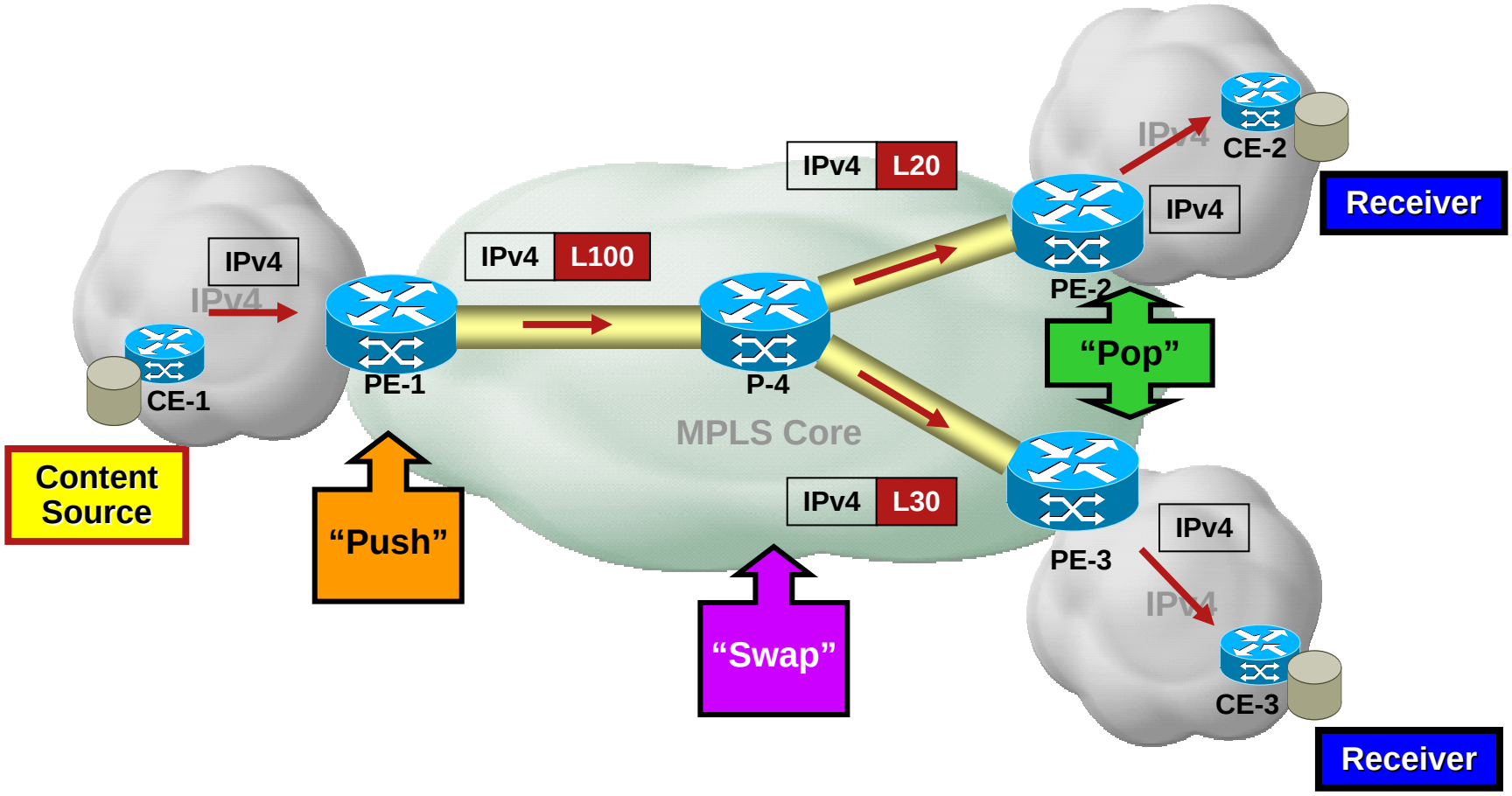
Combinations with L3 on PE
with RSVP-TE P2MP (possible futures)

- RSVP-TE P2MP static / native
  - Core trees statically provisioned:
    - Headend: configure each tree:
      - Set of tailend-PE
      - All IP multicast traffic that need to be passed into the tree.
      - *Remember message from sponsor (get involved in application provisioning)*
        But RSVP-TE looks much better in the resiliency section !!!
- RSVP-TE P2MP static in VPN context
  - Possible, some more per-VRF/VPN config
- RSVP-TE P2MP dynamic
  - TBD: MVPN or new PE-PE signaling (work in IETF, vendors)
  - Many customers opting for RSVP-TE do not even want this!

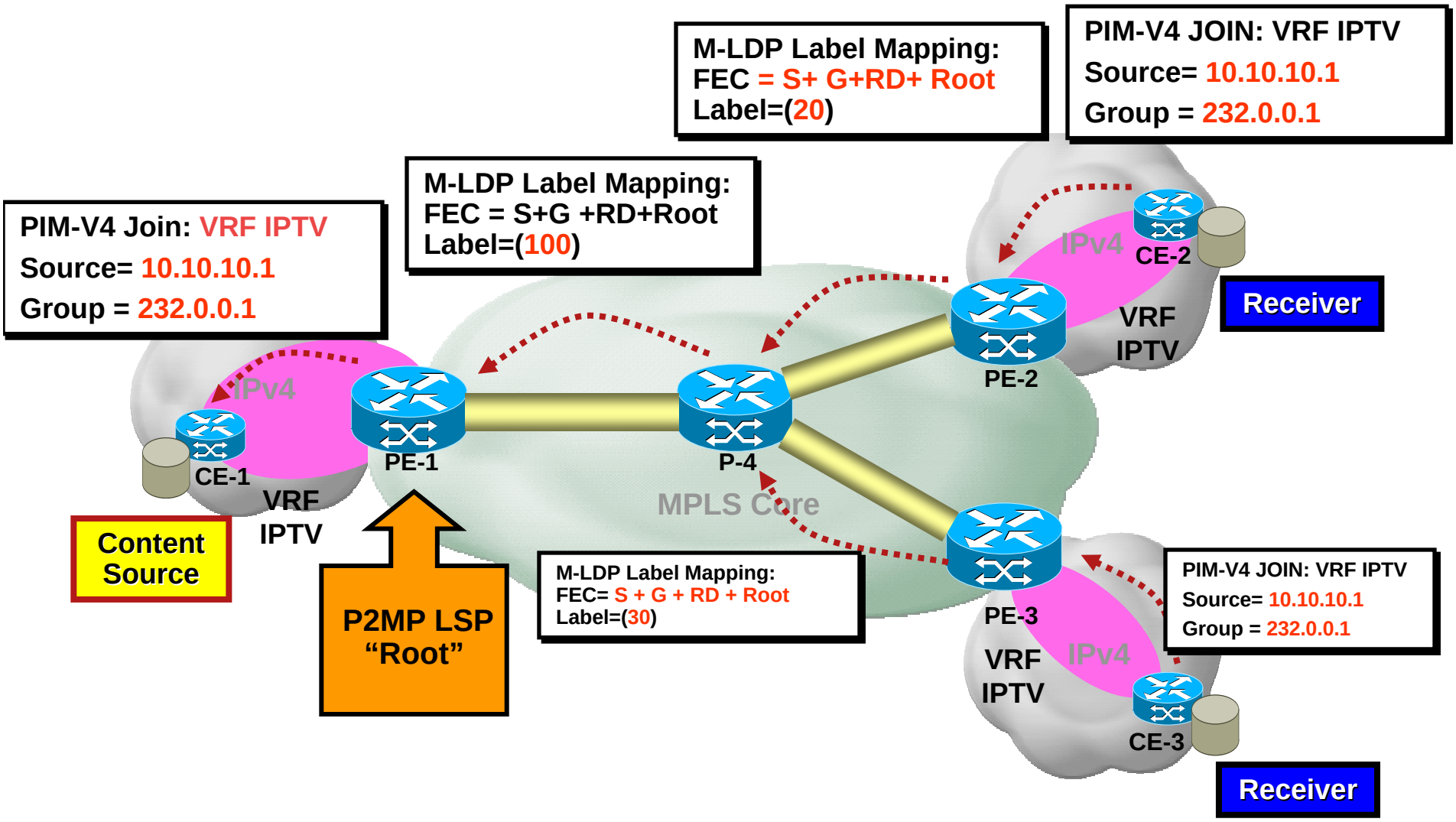
mLDP
Traffic forwarding

- Forwarding is the same for RSVP-TE P2MP and mLDP

  "just" signaling / protocols !



mLDP signaling
native and Direct-MDT



P2MP RSVP-TE
Signaling overview

1. Remember P2P RSVP-TE ?

   P2P LSP build from headend to tailend

2. P2MP RSVP-TE

   Headend builds equivalent of P2P RSVP-TE LSP
   and signals these “sub-LSPs” belong to one P2MP LSP

   P and PE nodes figure out that sub-LSPs belong to a tree and “merge” them:

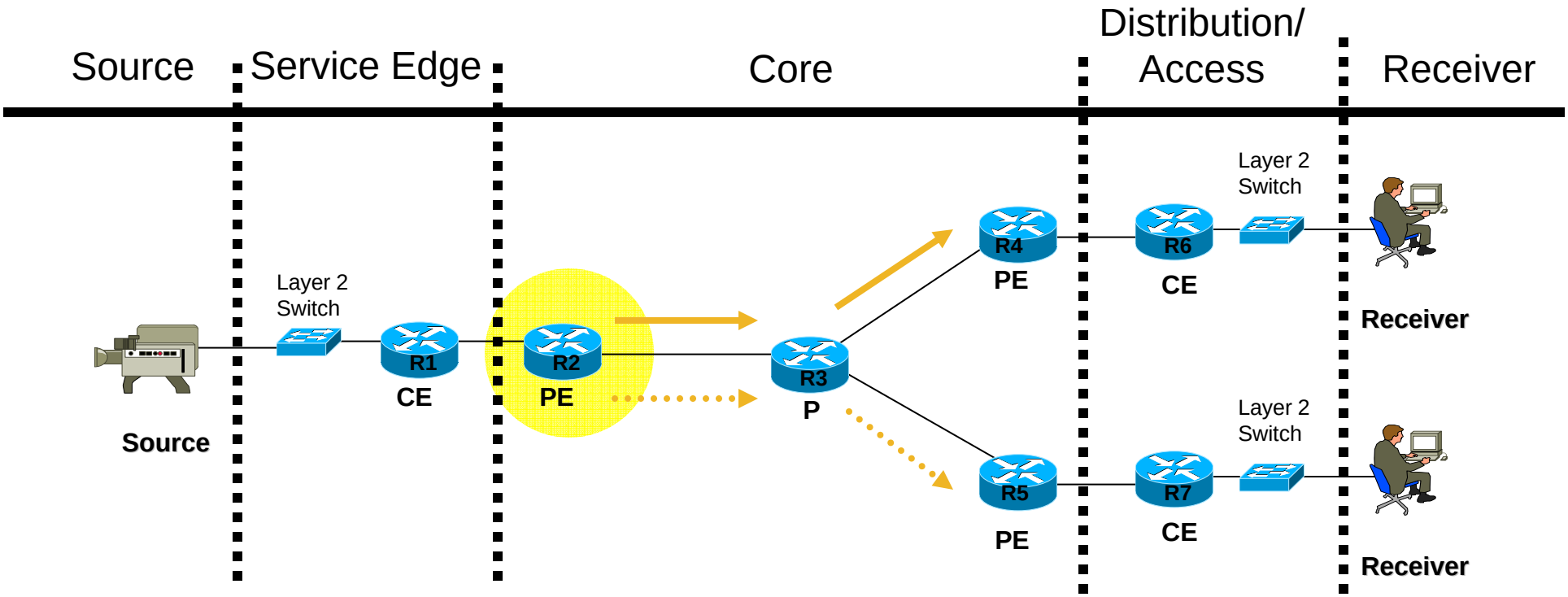
   Signal just a single label to the upstream for all sub-LSP of a P2MP LSP

- Everything else pretty much the same as P2P

  ERO, CSPF, link protection

  Node protection more difficult

P2MP RSVP-TE signaling details

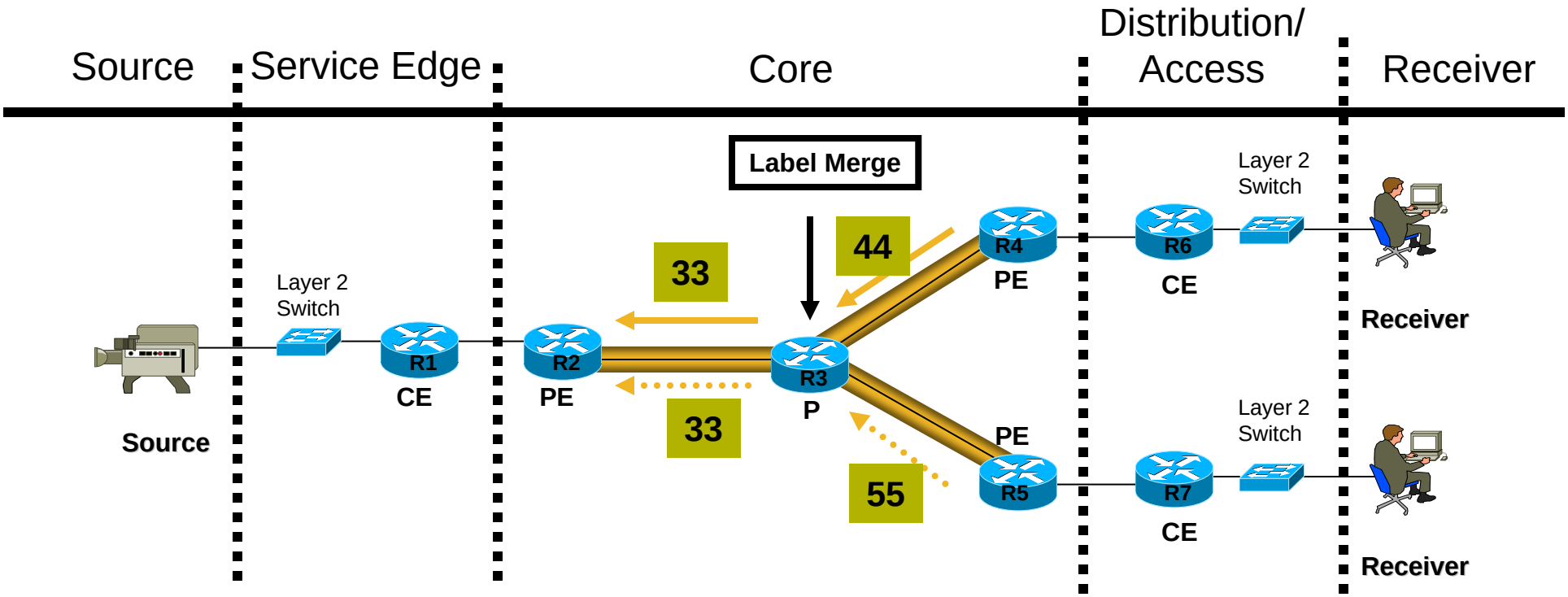


Headend sends one PATH message per destination

PATH Message : ERO -> R2-R3-R4

PATH Message : ERO -> R2-R3-R5

P2MP RSVP-TE signaling details



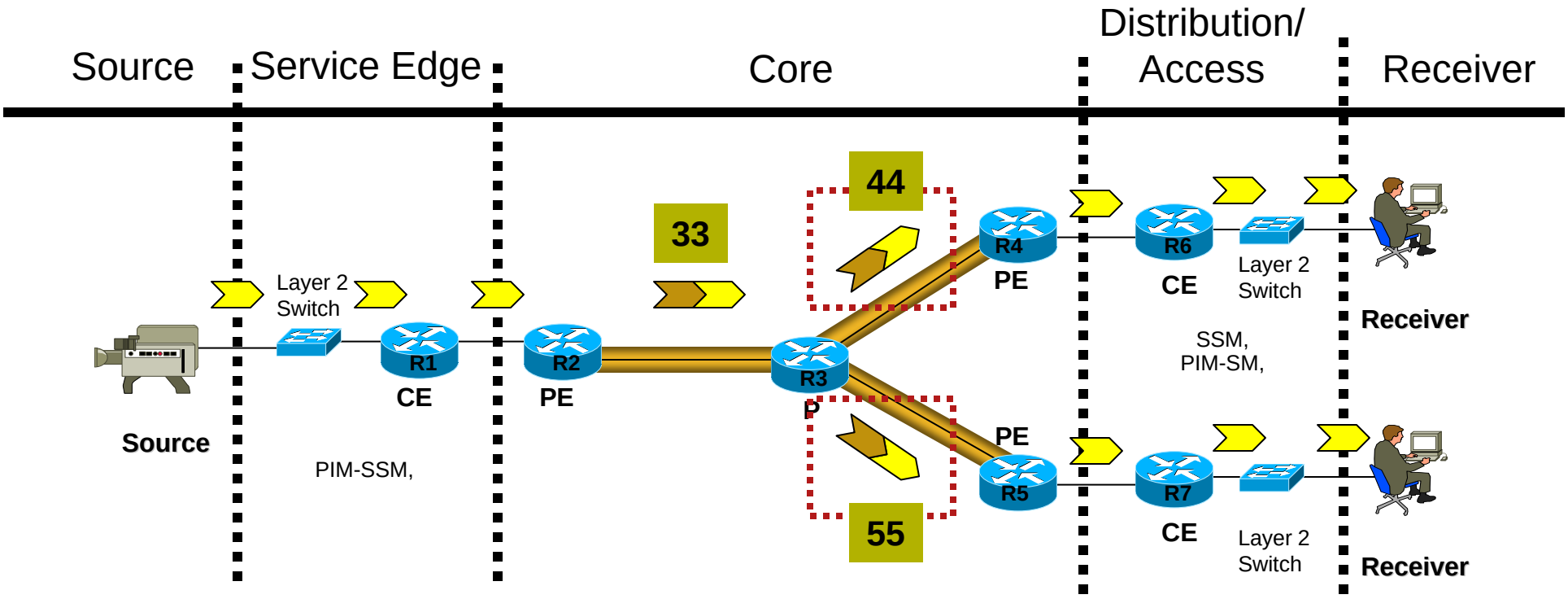
RESV Messages are sent by Tailend routers; Communicates labels & reserves BW on each link

RESV Msg Initiated by R4

RESV Msg Initiated by R5

55 Label Advertisement carries in the RESV Message

P2MP RSVP-TE Forwarding



No PHP ! Need label on tailend PE to identify tree

Multicast Packet

Labeled Packet

Comparison of core tree building

- Native PIM-SSM
  - Simple, 6 years availability, many deployments
  - Don’t confuse with complexity in PIM-SM!
- mLDP
  - Best approach for equivalent of PIM + label switching. All benefits of PIM (none of the old PIM-SM complexity), plus MPLS / LDP signaling benefits (signaling reliability, better VPN support,…)
- RSVP-TE P2MP
  - Strength in TE elements (ERO/CSPF + protection)
  - Headend based tree building opposite to PIM tree building
    - More signaling, more load on headend, dynamic tree issues, headend redundancy/failover
- Combining RSVP-TE P2MP + mLDP unresolved
  - Unicast: RSVP-TE between P nodes, LDP on PE to P.
  - Size of RSVP-TE P2MP trees can become quite large !

L2 on PE

- No P/PE L2 solution with P2MP trees
  - VPLS – full-mesh/hub&spoke P2P PS only
  - Non P/PE models: P are L2 swiches with protected pseudowires. IGMP/PIM snooping on P nodes.
- Futures (IETF/Vendors): Two main elements:
  - Define PE functions for L2 services (eg: VPLS++) with either core tree building (native, mLDP, RSVP-TE P2MP)
  - Discussion about PIM/IGMP snooping on PE:
    - Vendors: implement when customers ask
    - Presentor: Careful – if you need it, consider L3 on PE

Virtualization considerations

- (some) Customers ask for IPTV into (L2/L3)-VPN

  … because other services are also isolated amongst each other that way

  … to support “wholesale” in future (VPN for each SP / content-provider)

- Too complex ?

  Consider subs should be able to receive arbitrary subset of content from different SP/content-providers.

  L3VPN: Extranet. L2VPN… hmm

- Much simpler: just rely on implicit isolation of SSM – no need for virtualization to achieve isolation!

- (My) Mantra: If you already have VPN context for some unicast service and want to add multicast to it: fine

  But don’t create new VPN contexts just for multicast services.

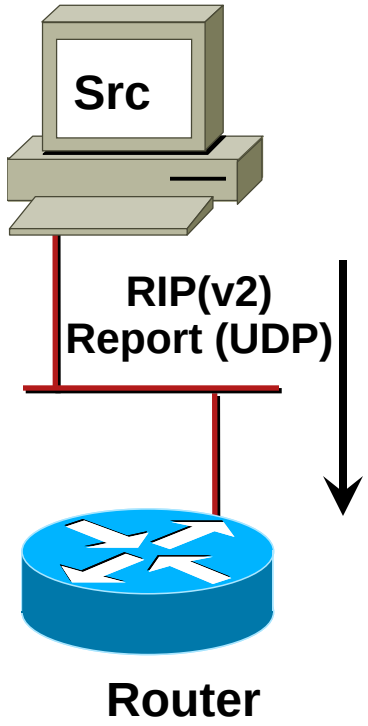
IP unicast tunneling – AMT
last but not least!

- AMT – Automatic Multicast Tunneling
  - Draft in IETF
  - UDP or GRE tunnel with automatic (anycast) discovery of headend router. Primarily for SSM-only to keep it simple.
  - Benefits over other tunnels (IPsec, L2TPv3, MobileIP,…)
    - Nothing really new, but:
      - As simple as possible, targeted to problem, consideration for NAT and can be implemented in App on PC (instead of OS – when UDP is used)
- Variety of target deployment cases
  - Core-SP to Home (HAG or PC or application in PC)
  - Access-SP to Home – eg: to overcome non-multicast DSL
  - In-Home only (to overcome multicast issues in WLAN)
  - …

Resiliency

Source Redundancy
Anycast/Prioritycast signaling

- Redundant sources announce Source Address via RIPv2
- Routers redistribute (with policy) into actual IGP
  - Easily done from IPTV middleware (UDP)
  - No protocol machinery required – only periodic announce packets.
  - Small periodicity for fast failure detection
    - Better: Use BFD between Router/Host too.
  - All routers support RIPv2, but not often used as real IGP:
    - Allows secure constrained config on routers



Source Redundancy
Anycast/Prioritycast policies

- Policies

  Anycast: clients connect to the closest instance of redundant IP address

  Prioritycast: clients connect to the highest-priority instance of the redundant IP address

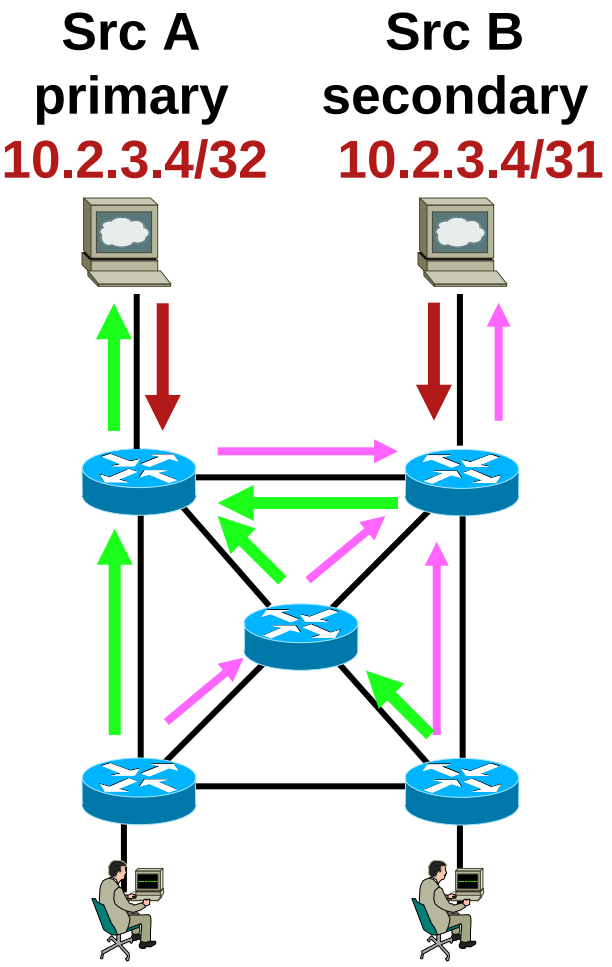
- Also used in other places

  Eg: PIM-SM and Bidir-PIM RP redundancy

- Policy simply determined by routing announcement and routing config

  Anycast well understood

  Prioritycast: engineer metrics of announcements or use different prefix length.



Example: prioritycast with Prefixlength annuncement

Source Redundancy
Anycast/Prioritycast benefits

- Subsecond failover possible
- Represent program channel as single (S,G)
  - SSM: single tree, no signaling, ASM: no RPT/SPT
- Move instances “freely” around the network
  - Most simply within IGP area
  - Regional to national encoder failover (BGP..) ?
- No vendor proprietary source sync proto required
- Per program, not only per-source-device failover
  - Use different source address per program

Multicast Fast Convergence

- IP multicast

  All failures / topology changes are corrected by re-converging the trees

  Re-convergence time is sum of:

  Failure detection time (only for failure cases)

  Unicast routing re-convergence time

  ~ #Multicast-trees PIM re-convergence time

  Possible

  ~ minimum of 200 msec initial

  ~ 500 ... 4000 trees convergence/sec (perf)

- Same behavior with mLDP !

Sub 50 msec solutions: Fast Reroute

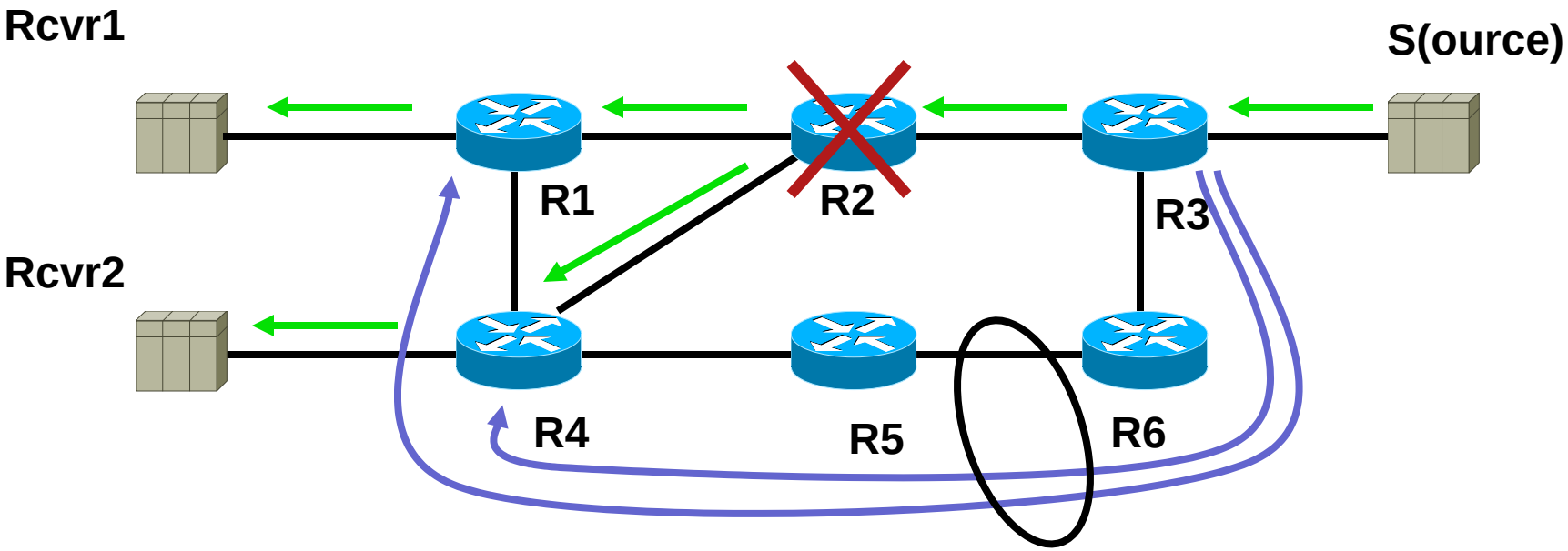
- Fast ReRoute
  - Targets sub 50msec interruption in covered cases
  - Pre-established (Link, Node) backup paths/(tunnels)
    - trunks/ECMP/LFA/NotVia/RSVP-TE/P2P
      - Only for link/node failures
  - Make before reconvergence
    - For everything else: link,linecard,node recovery, network (core) topology expansion/change
  - Not covered: Headend redundancy !
- RSVP-TE P2MP: All included (ietf) !
- Native IP multicast, mLDP
  - Nothing included
  - Vendors: Work on link protection extensions…
  - .. Or use existing – protected pseudowire instead of phys link

Multicast Node Protection
With p2p backup tunnels

- *If router with fan-out of N fails, N-times as much backup bandwidth as otherwise is needed.*

  Provisioning issue depending on topology !

- Some ideas to use multipoint backup to resolve this, but…
- Recommendation?: Rely on Node HA instead!!



Stream redundancy with path separation

Only solution that can guarantee 0 loss upon single network outages without adding latency

- Duplicate copies of multicast data
- Long-time use in finance market data feeds

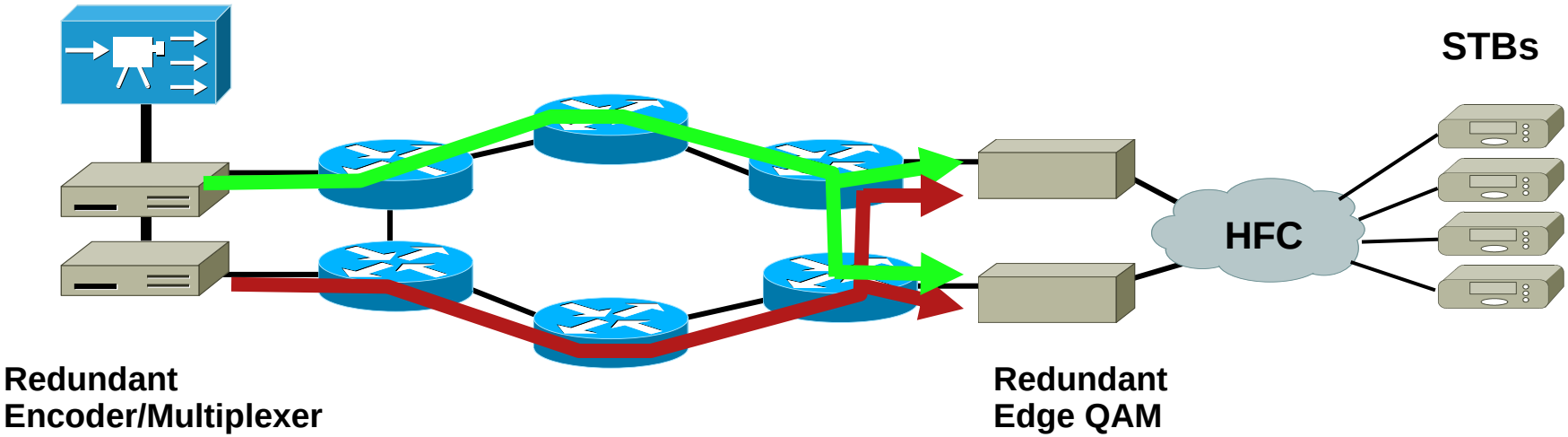
  Source and receiver hosts handle creation and elimination of duplicates

  Two networks built:

  No single network failure will impact both flows
- Also starting to see interest in broadcast industry – with more cost saving approaches

Stream redundancy with path separation

Candidate example from broadcast-TV in cable



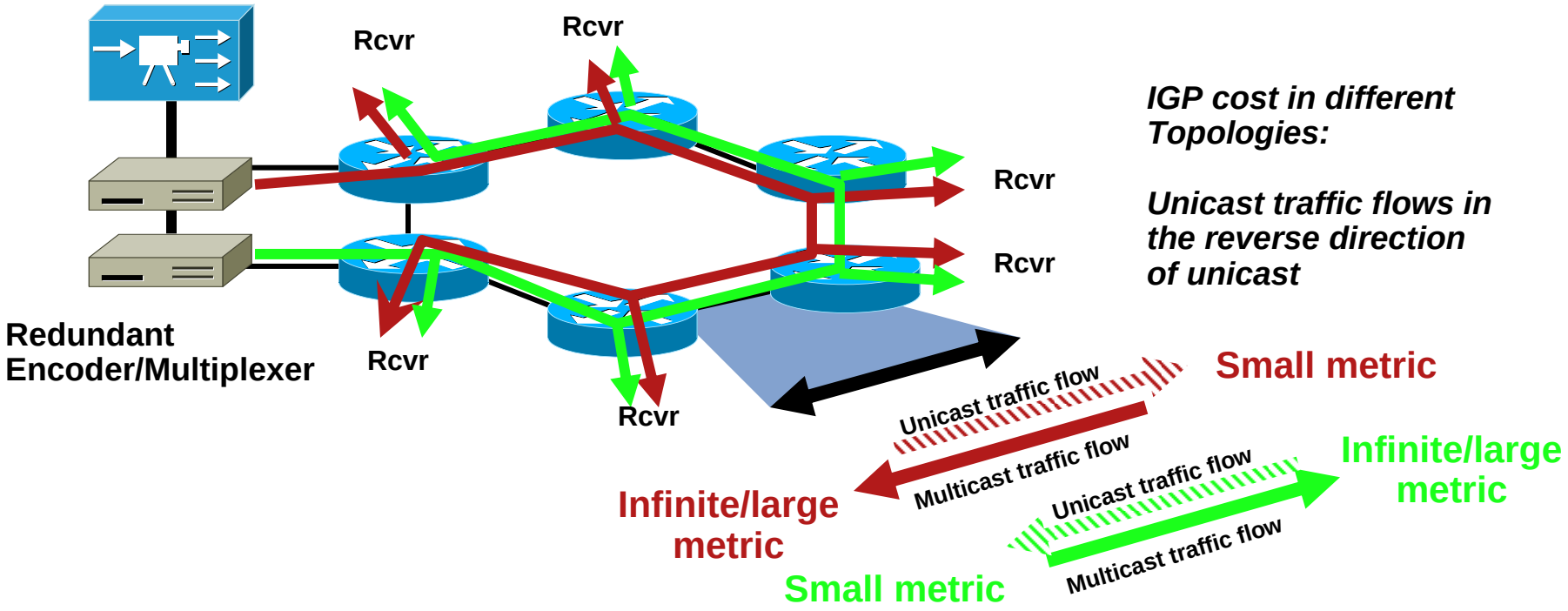
- Encoder/Multiplexers generate two copies of IP multicast flows
- Network uses methods of path separation

  Multiple IGP instances, topologies, two networks, VRF-lite, RSVP-TE, …
- Each receiver consumes both copies

  Remove duplicates by sequence numbers (eg: RTP sequence numbers).

  Any single failure in network: 0 packet loss. 0 added latency
- Same bandwidth allocation needed as in traditional SONET rings, but solution even better: 0 loss instead of <= 50 msec.

MT-technology for Stream redundancy with path separation - Details



Can share links for two copies in rings !

Use asymmetric metrics!

May need infinite metric if reconvergence is not wanted

Available in IS-IS. In draft only for OSPF

Application side resiliency

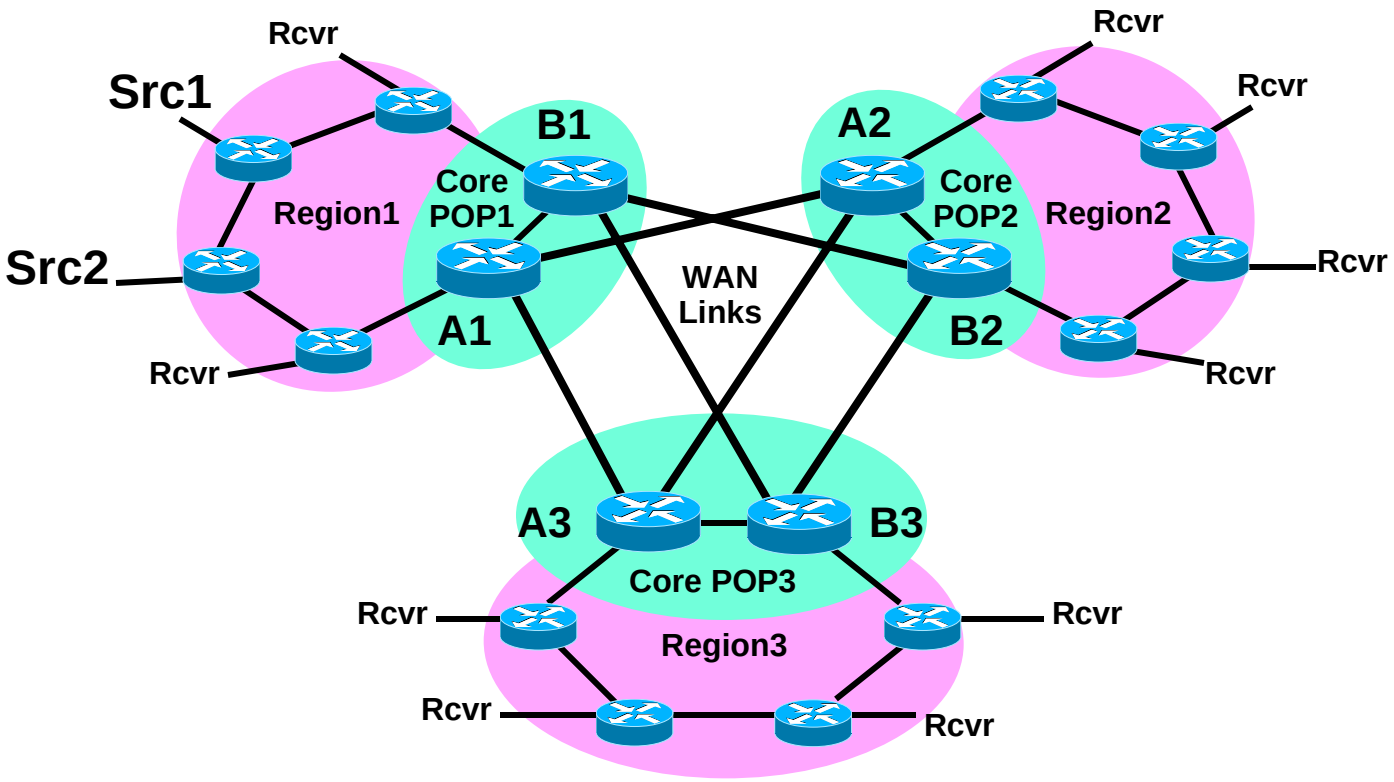
- FEC – Forward error correction
  - Compensate for statistical packet loss
  - Use existing FEC eg: for MPEG transport to overcome N msec (>= 50 msec) failures ?
  - Cover loss of N[t] introduces delay > N[t] !
- Retransmissions
  - Done eg: with vendor IPTV solutions – unicast retransmissions
    - Candidate large bursts of retransmissions!
    - Limit #retransmissions necessary
  - Multicast retransmissions (eg: PGM ?)
    - No broadcast IPTV solutions use this

Failure impact upon viewer experience

- *Very hard to measure and quantify*
- If I frames or frame-information is lost, impact will be for a whole GOP
  - GOP can be 250 msec (MPEG2) .. 10 sec (WM9)
- Encoding and intelligence of decoder to “hide” loss impact quality as well
- IPTV STB typically larger playout buffer than traditional non-IP STBs:
  - Loss can cause catch-up: no black picture, but just a jump in the motion.
- What loss is acceptable ?
  - Measured in #phone calls from complaining customers ?!

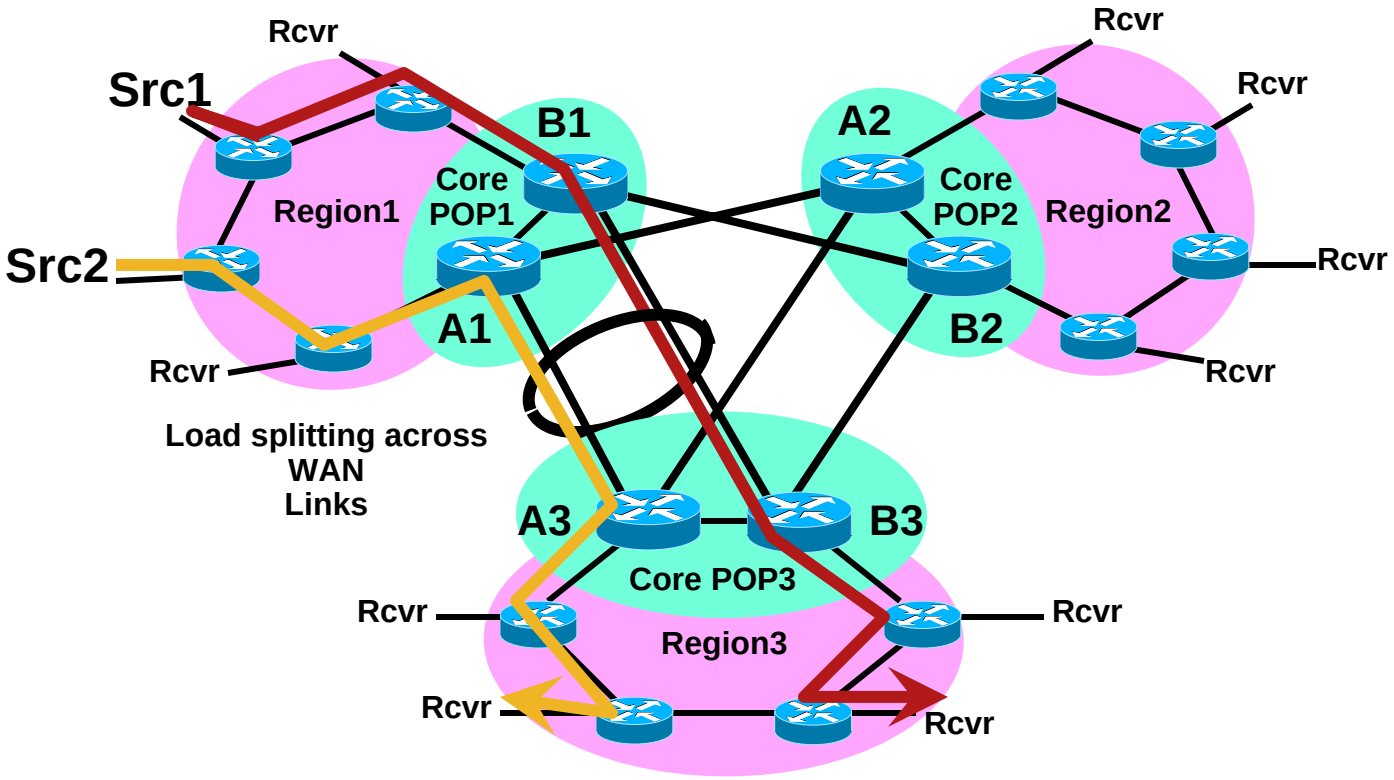
Path selection

Cost optimization example



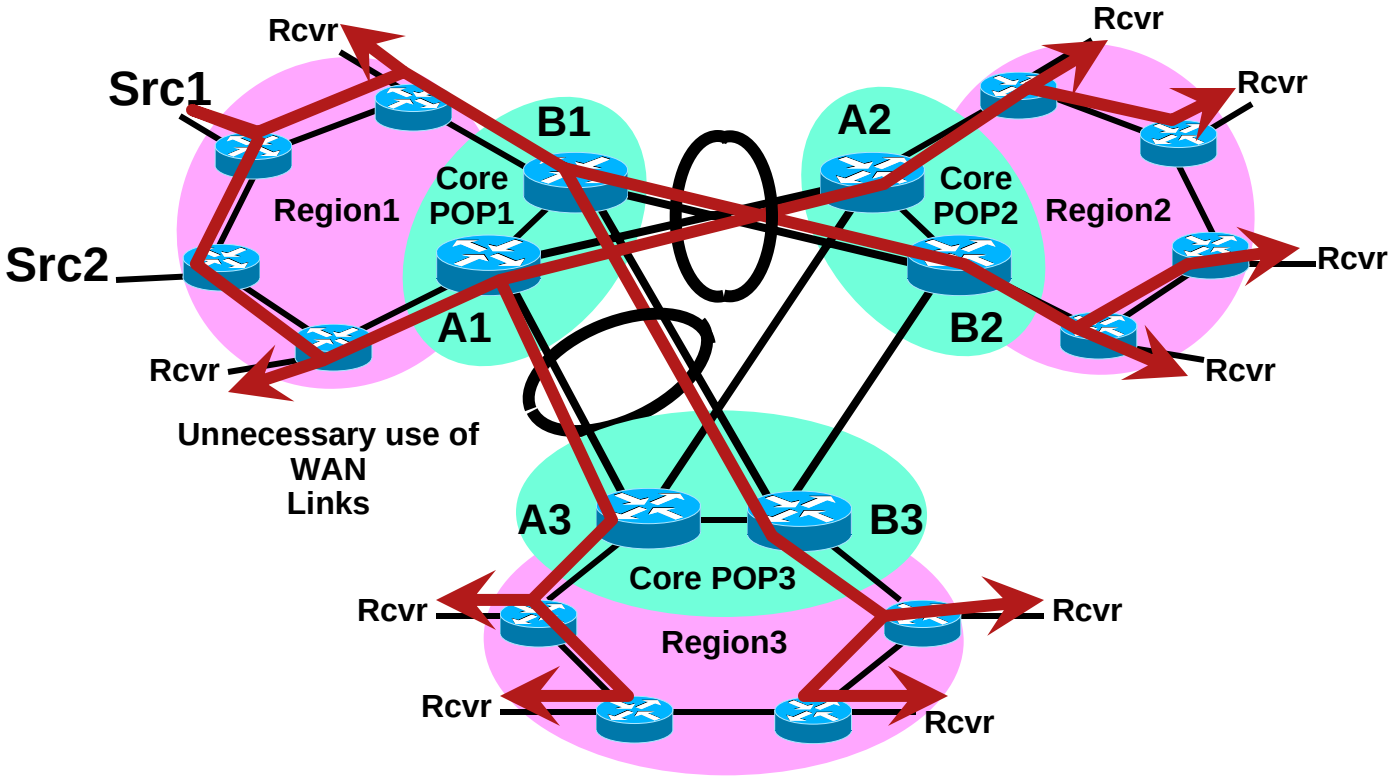
- Consider simplified example core/distribution network toplogy
- Core pops have redundant core routers, connectivity via (10Gbps) WAN links, redundant. Simple setup: A/B core routers, A/B links
- Regions use ring(s) for redundant connectivity

Cost optimization example (2)



- IGP metric are set to achieve good load distribution across redundant core.
  - Manual IGP metric setting and/or tools (Cariden)
  - Assume in the idealized topology cost of 1 on all links.
- Result: Unicast traffic is load split across redundant core links

Cost optimization example (3)

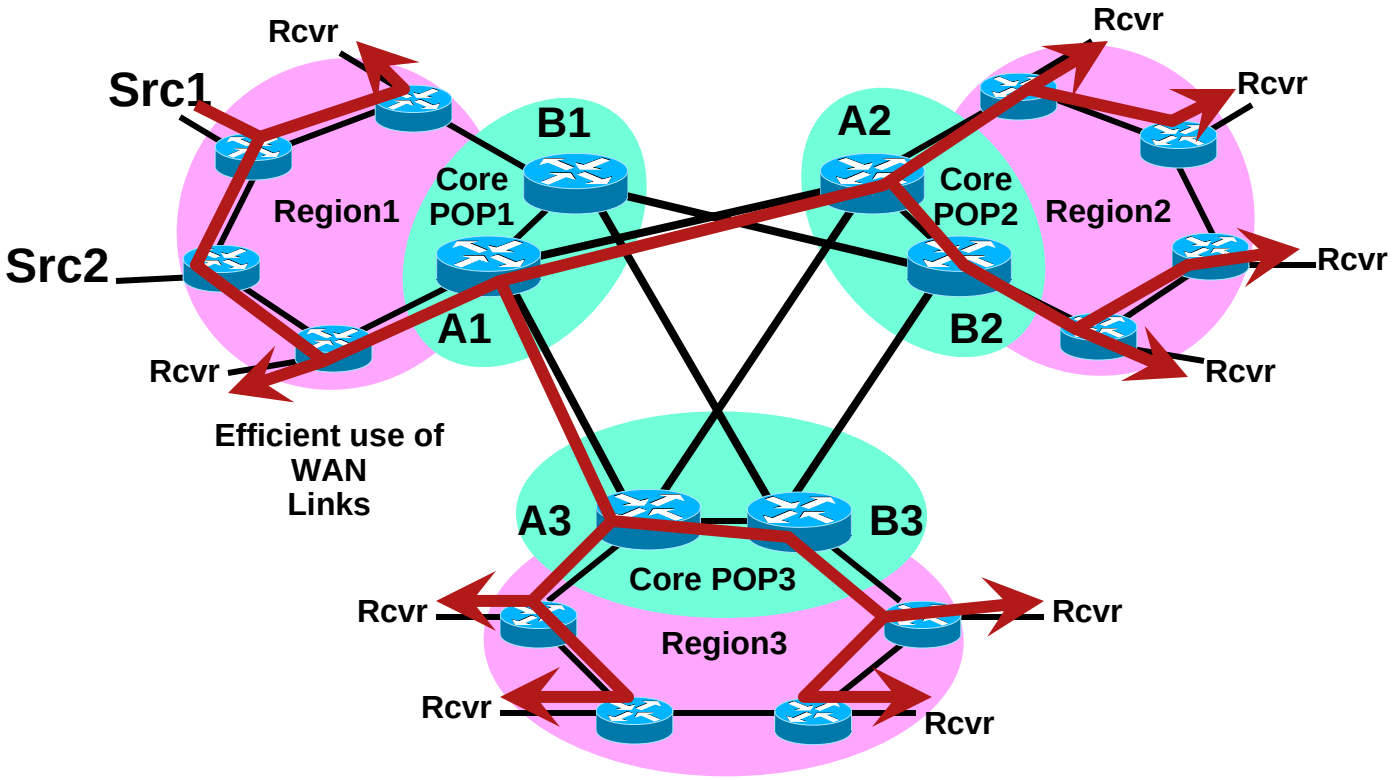


- The same metric good for unicast load splitting cause multicast traffic to go unnecessarily across both the A and B WAN links.

  10 Gbps WAN links, 1..2 Gbs multicast => 10..20% WAN waste (cost factor)

- Can not resolve problem well without multicast specific topology

MT-technology application 1
Cost optimization



- Simple? to minimize tree costs with a multicast specific topology

  Manual or tool based

  Example toplogy: make B links very expensive for multicast (cost 100), so they are only us as last resort (no A connectivity)

Metric engineering discussion

- On unicast side, IGP metric optimization competes with RSVP-TE (engineered paths). Tools like cariden support both
  - Comparison shows similar benefits (within 10%) http://www.nanog.org/mtg-0505/telkamp.html
- Comparison for multicast vs. P2MP RSVP ??
- Minimum cost tree (sum of all used link costs): Steiner tree.
  - Can be built by P2MP RSVP-TE/P2MP (ERO or CSPF)
    - Careful: stability in case of membership change
  - How much worse are topology with optimized metrics ?

IP multicast (and mLDP) ECMP (equal cost multipath)

- Per multicast tree selection of RPF-nexthop in case of ECMP

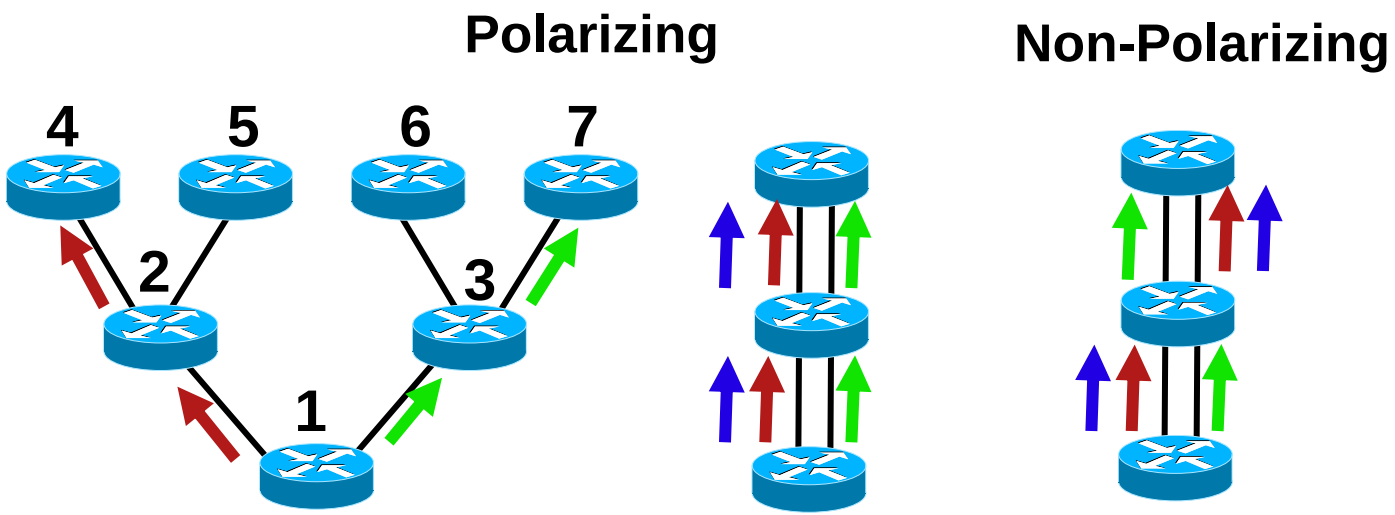
  Polarizing but consistent

  i = (hash(S) % n)

  Non-polarizing – stable in case of lin failures:

  i = i | max(hash(S, Nbr-i))

Given *1..n* (eg: 2) ECMPs, if all routers select the same neighbor *I* for a source S, then polarization may happen: A rtr2 will only be joined to by rtr1 for Sources that it's own ECMP would RPF to rtr4, but never to rtr5!



Multicast RPF Selection for different source addresses

Path selection review
RSVP-TE/P2MP

- CSPF/ERO “Traffic Engineering”
  (bandwidth, priority and affinity based path selection)
- Very powerful “can do everything we can think of”
- “Offline” management (ERO) most common

  Network provider incooperates “off-network” information about necessary multipoint trees
- “Online” / CSPF based path selection

  Products will support it …

  Ideal for single headends

  How much better than SPF without coordinated CSPF for multiple headends ?

  Network global CSPF calculation ultimate direction ?

Path selection review
PIM (native multicast) / mLDP

- Can not load split across non-equal-cost paths (from same sources to same set of receivers).
- Path engineering with topologies and ECMP:
- **ECMP**
  - best when multipoint traffic << link bandwidth (30%?)
  - Higher utilization deployments – special considerations (due to statistical chance of congestion)
- Topologies
  - Single incongruent topology – necessary/sufficient for cost opt. ?
  - Two topologies for path separation (dual-stream)
  - Could use more topologies for more functionality – eg: non-equal-cost load-splitting – but maintaining many topologies likely not less complex than RSVP-TE
  - *Note: MT-technology for multicast/multipoint only happens in control plane. No forwarding plane impact*

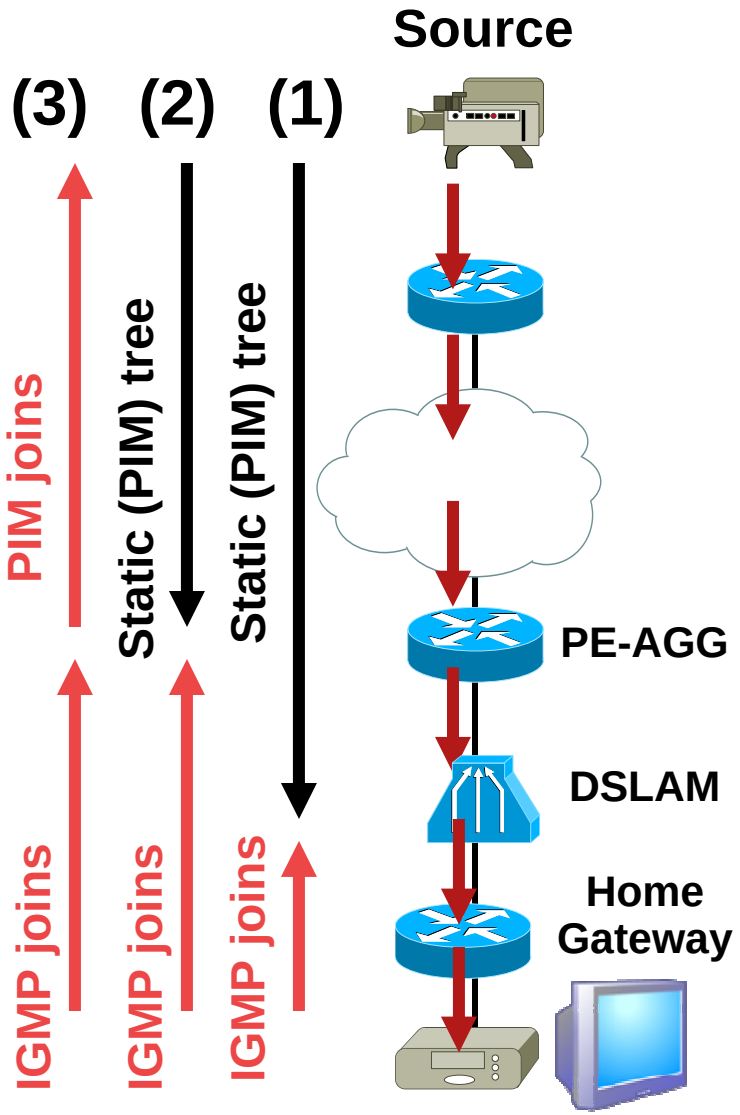
Channel changing

The problem (that should be obsolete)

- IGMPv2 leave latency !
- Example:
  - 4Mbps DSL link, 3.5 Mbps MPEG2
  - Can only receive one TV channel at a time
  - Leave latency on channel change complex (triggers IGMP queries from router/DSLAM) and long (spec default: 2 seconds)
- Resolved with IGMPv3/MLDv2
  - Ability for explicit tracking (vendor specific)
  - Can immediately stop forwarding upon leaves

Join-latency
Static vs. dynamic trees

1. “Broadcast Video”
   - static forwarding into DSLAM
   - Fear of join latency
   - History (ATM-DSLAM)
2. “Switched Digital Video”
   - Allow oversubscription of PE-AGG/DSLAM link
3. “Real Multicast”
   - dynamic tree building full path



Switched Digital Video

- Consider 500…1000 users on DSLAM
- Consider 300 available TV programs
- Run statistic how many TV programs are actually viewed in parallel
  - Numbers show often only 1/4 … 1/3 programs maximum needed (eg: from Cable networks)
  - Dynamic joining between DSLAM/PE-AGG can allow to fit traffic into typical 1Gbps budget
- Dynamic joining towards core ?
  - Todays offered content << #users aggregated -> worst case traffic will always flow.
  - More a provisioning issue – and when content expands well beyond current cable-TV models

Admission control

- Oversubscription (eg: PE-AGG/DSLAM link) raises question of admission control

  Real-time !
  One flow too many messes up everything

  Vendor-specific: Router/L2-Device local config for per-interface maximum# multicast flows

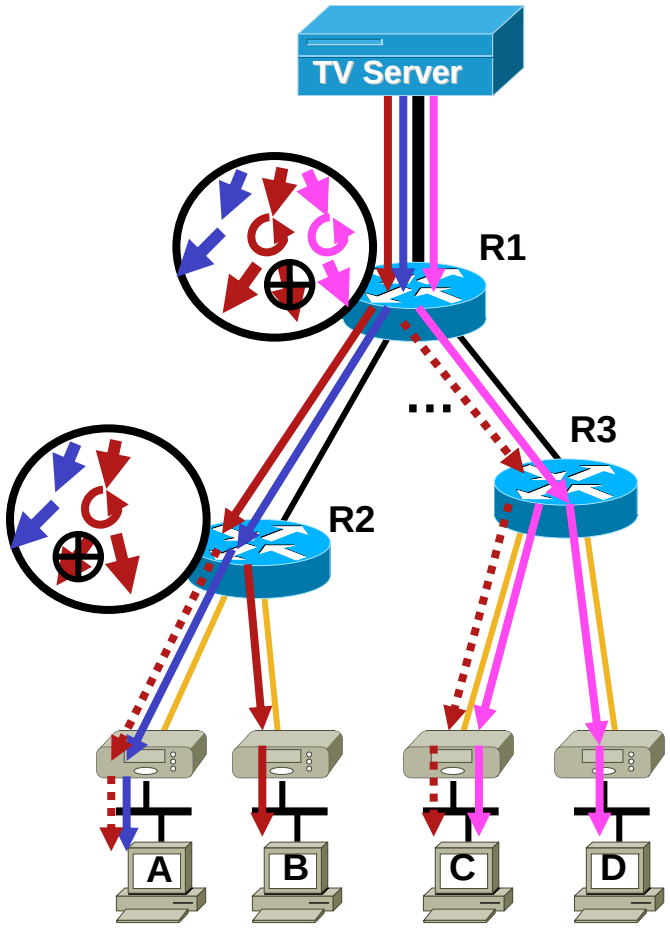
  With more varying bandwidth (2.. 20Mbps) of TV programming, this may need to become bandwidth aware

    Vendor specific: Local router CLI

  Revive RSVP for multicast admission control, please ?!?!

Multicast vs. Unicast
Application Side Difference – Intserv Admission Control

- Intserv:
  - per flow (admission) control
- Unicast:
  - Source side enforcement!
  - No need for network enforcement
- Multicast:
  - Network enforcement!
  - Need to block forwarding at replication points to individual branches!
- Mechanisms:
  - Vendor specific..
  - RSVP only standardized common unicast/multicast solution eg: limit VoD+Bcast



Join Latency

- Static forwarding (to PE-AGG, or DSLAM) often done to avoid join latency

  But other reasons too (policy, …)

- Bogus ?

  Join latency (PIM/IGMP) very low,
  eg: individual < 100 msec

  Relevant: worst-case zapping performance

  Joins stop at first router/switch in tree that already forwards tree

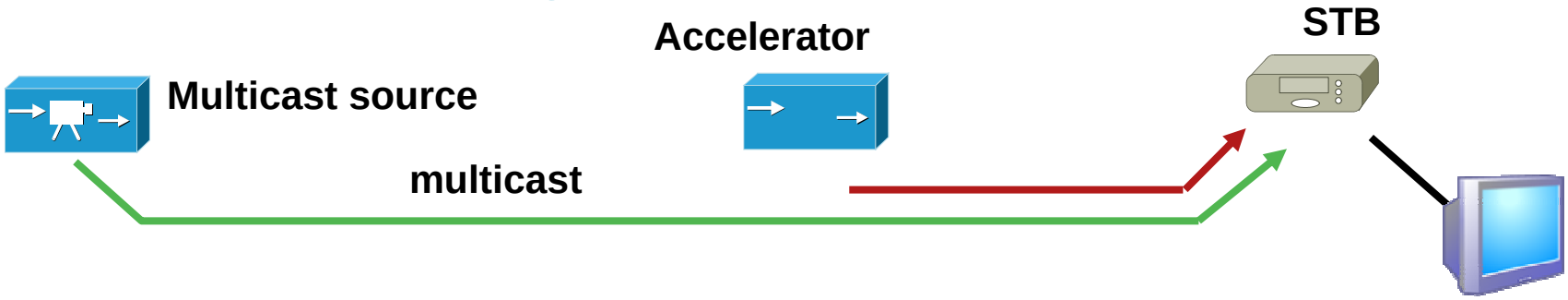
  Probability for joins to go beyond PE-AGG very low !

  If you zap to a channel and it takes ¼ sec more: You are the first guy watching this channel in a vicinity of eg: 50,000 people. Are you sure you want to watch this lame program ?

GOP size and channel changing

- GOP size of N seconds causes channel change latency >= N seconds
  - Can not start decoding before next I-frame
- Need/should-have channel change acceleration for GOP sizes > 0.5 sec ?
- Unclear (to me)
  - How much bandwidth is saved in different codecs by raising GOP size (same quality)
  - Eg: WM9/AV ~ 2.5 Mbps -> GOP size 3 sec
    - What bandwidth with 0.5 sec GOP size ?? Really 4 Mbps ?

Channel change acceleration



- Idea: add an IPTV channel change acceleration element
  Eg: “Instant Channel Change” in Microsoft TV IPTV edition

- Takeaway
  MPEG broadcast/multicast alone can not provide fast channel change
  IGMP join latency irrelevantly small compared to MPEG
  With IP network at least flexible solutions are possible
  As opposed to non-IP (eg: digital cable/satellite)
  Value vs. cost ? (just small gop-size ?)
  Acceleration could even hide whatever small IGMP join latency exists