Designing a Carrier Class TV Broadcast Network using P2MP MPLS-TE

Jean-Louis Le Roux, Orange France Telecom Group

jeanlouis.leroux@orange-ftgroup.com

Outline

- Point-to-Multipoint MPLS-TE: Overview
- Case Study: Orange TV Service
  - IP TV service growth and requirements
  - Various deployment scenarios
  - Various protection approaches
- Ongoing studies
- Closing remarks

P2MP MPLS-TE Overview

- P2MP MPLS-TE = Extensions of P2P MPLS TE mechanisms for setting up explicitly routed P2MP LSPs (aka P2MP TE-LSPs)
  - Relies upon P2MP RSVP-TE, an extension to RSVP-TE for trees
  - See RFC 4875
- Properties: Multicast Traffic Engineering and Fast Recovery
  - Allows setting up minimum cost trees (MCT, aka Steiner trees)
    => bandwidth savings
  - Multicast admission control: P2MP Resources reservation
  - Multicast fast recovery: P2MP Fast Reroute => sub-50ms recovery
- Applications: Multicast services with High bandwidth & Availability requirements
  - TV Broadcasting
  - Some Multicast VPN services

P2MP MPLS-TE Theory of operations

A P2MP TE-LSP from R1 to {R3, R4, R5}

1: P2MP TE-LSP configuration on R1

Destinations: R3, R4, R5

TE constraint: Bandwidth = 2Gbps

2: Tree Computation on R1

Taking into account TE constraints

R1-R2-R3

R1-R2-R6-R4

R1-R2-R7-R5

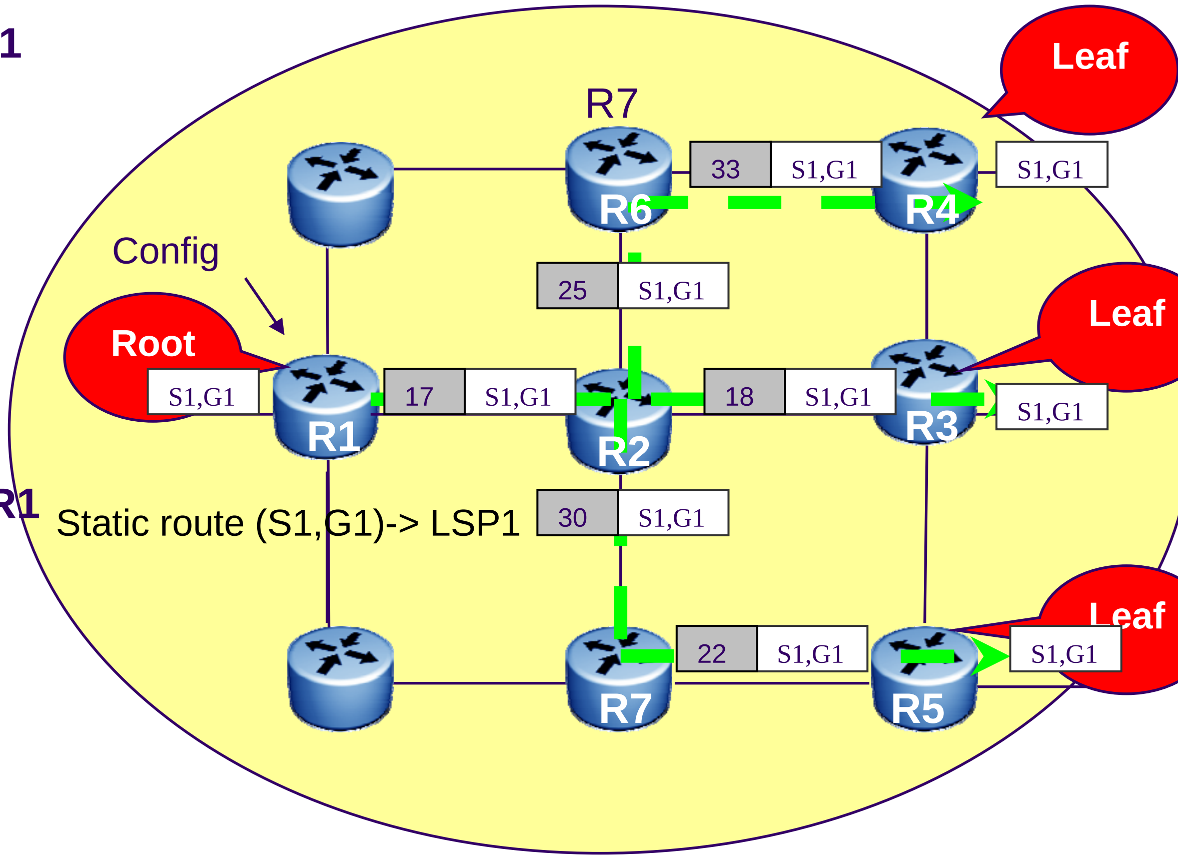
3: P2MP TE-LSP Setup initiated by R1

P2MP RSVP-TE along the computed paths

Explicit routing, label distribution, bandwidth reservation

4: LSP Utilization

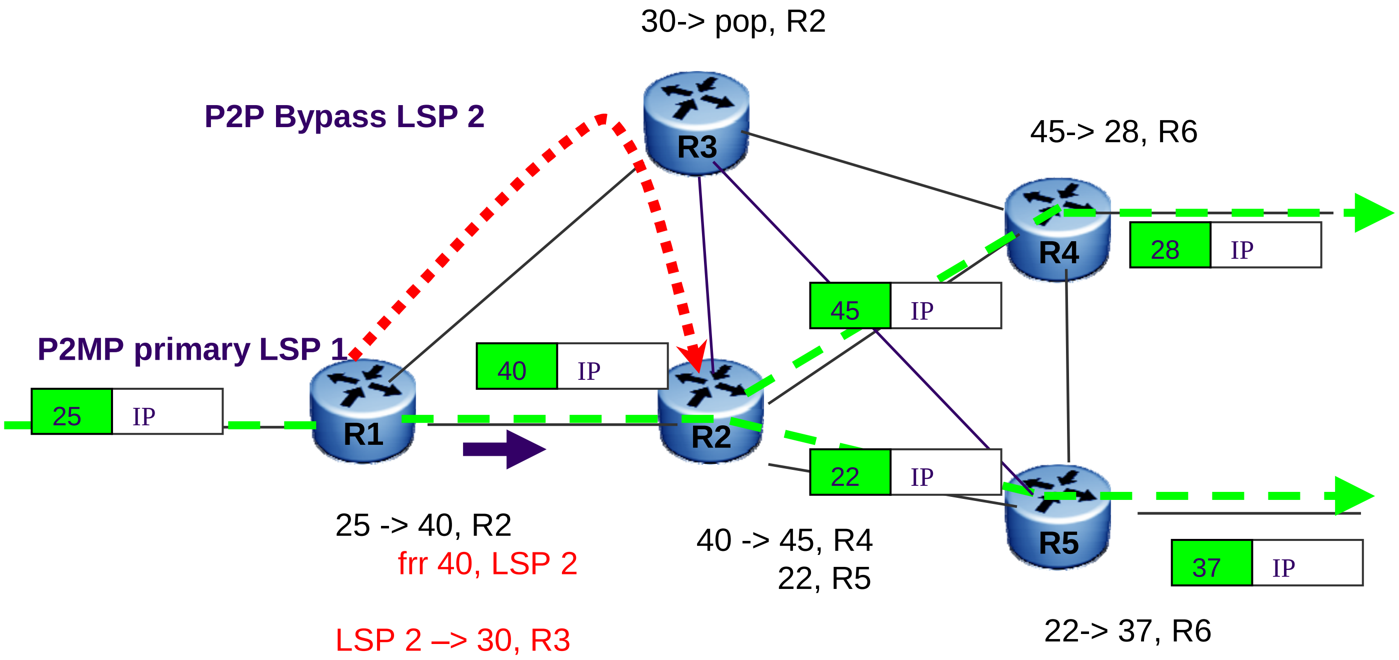
Example: Static IP multicast route



Root Initiated LSP Setu

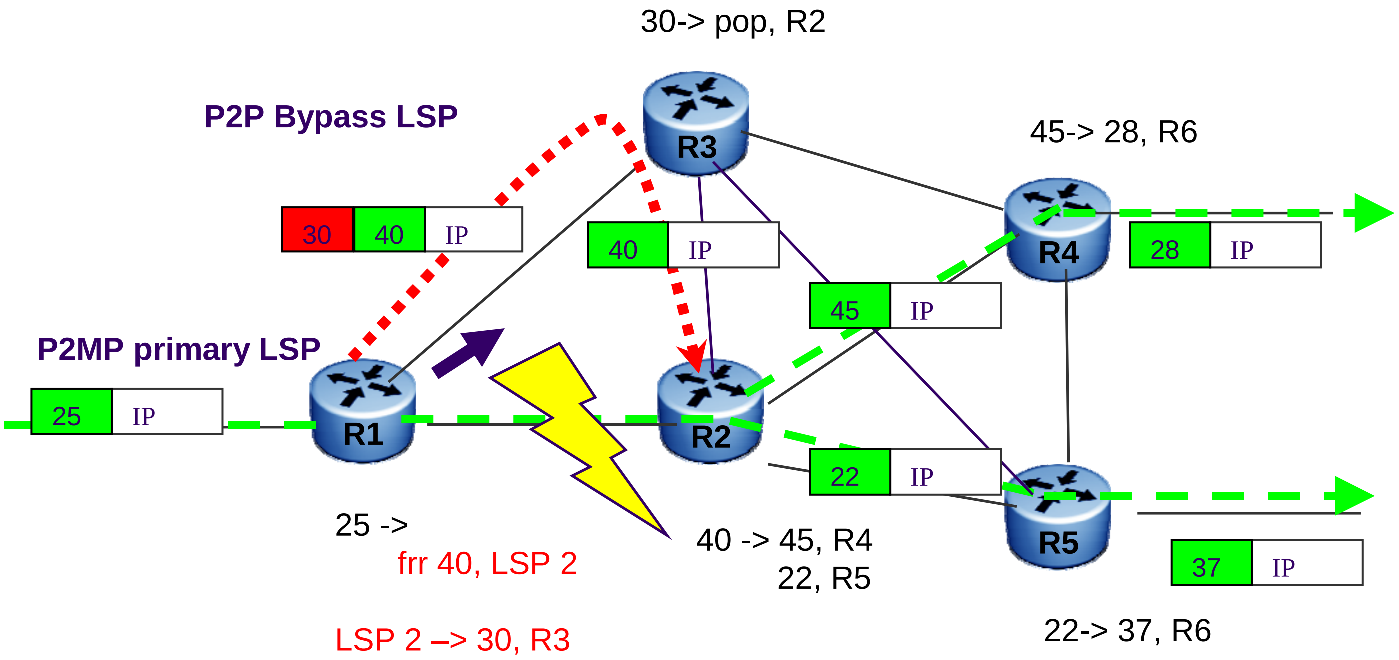
Explicit routing

P2MP MPLS-TE Fast Reroute



- P2MP primary LSP protected by a Link protection P2P Bypass LSP
  - Reuse existing FRR mechanisms
- Upon link failure all traffic is rerouted onto the Bypass LSP
- Sub-50 ms recovery
  - Failure detection + MPLS Table update on the PLR

P2MP MPLS-TE Fast Reroute



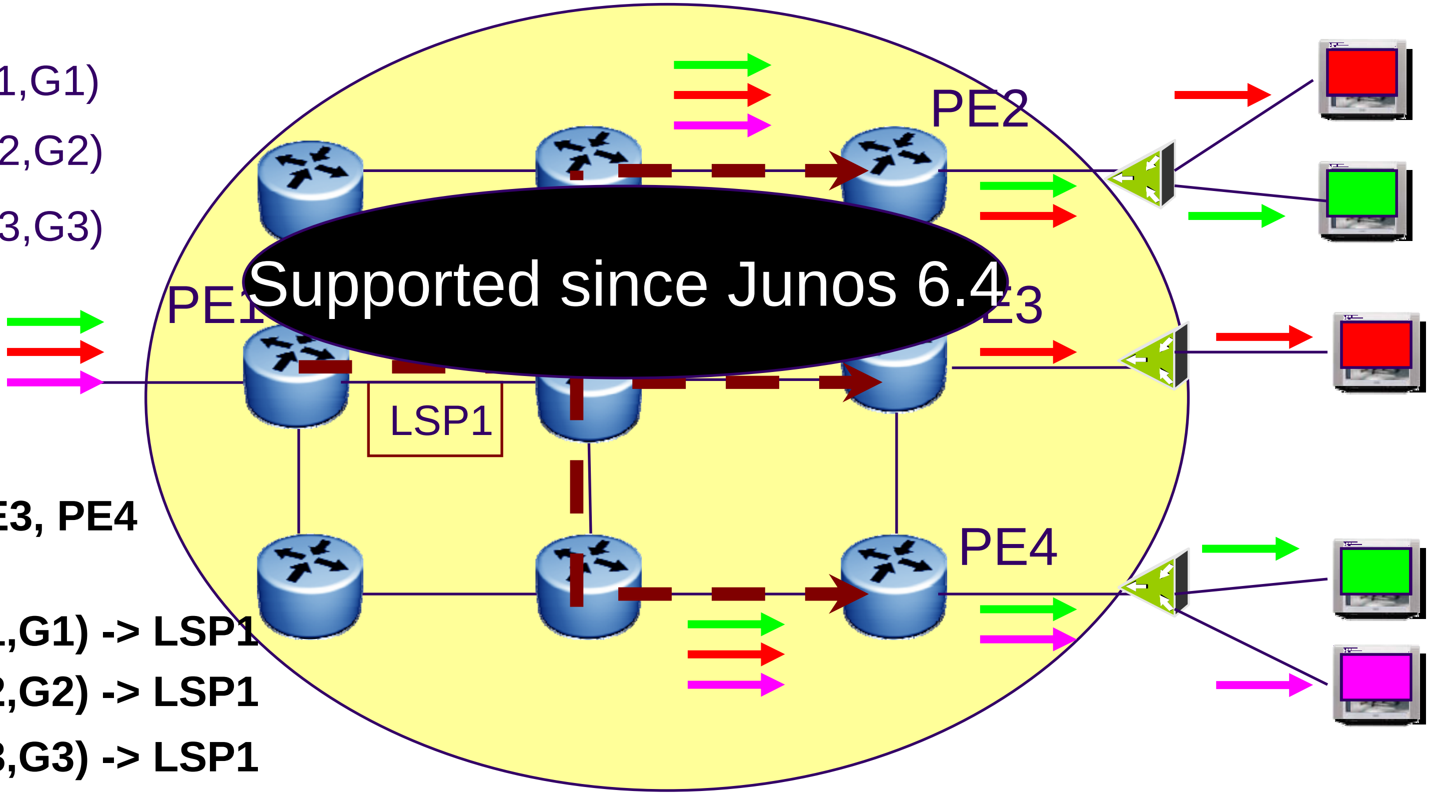
- P2MP primary LSP protected by a Link protection P2P Bypass LSP
  - Reuse existing FRR mechanisms
- Upon link failure all traffic is rerouted onto the Bypass LSP
- Sub-50 ms recovery
  - Failure detection + MPLS Table update on the PLR

P2MP LSP Utilisation: Static mode

🟥 Channel 1 (S1,G1)

🟩 Channel 2 (S2,G2)

🟪 Channel 3 (S3,G3)



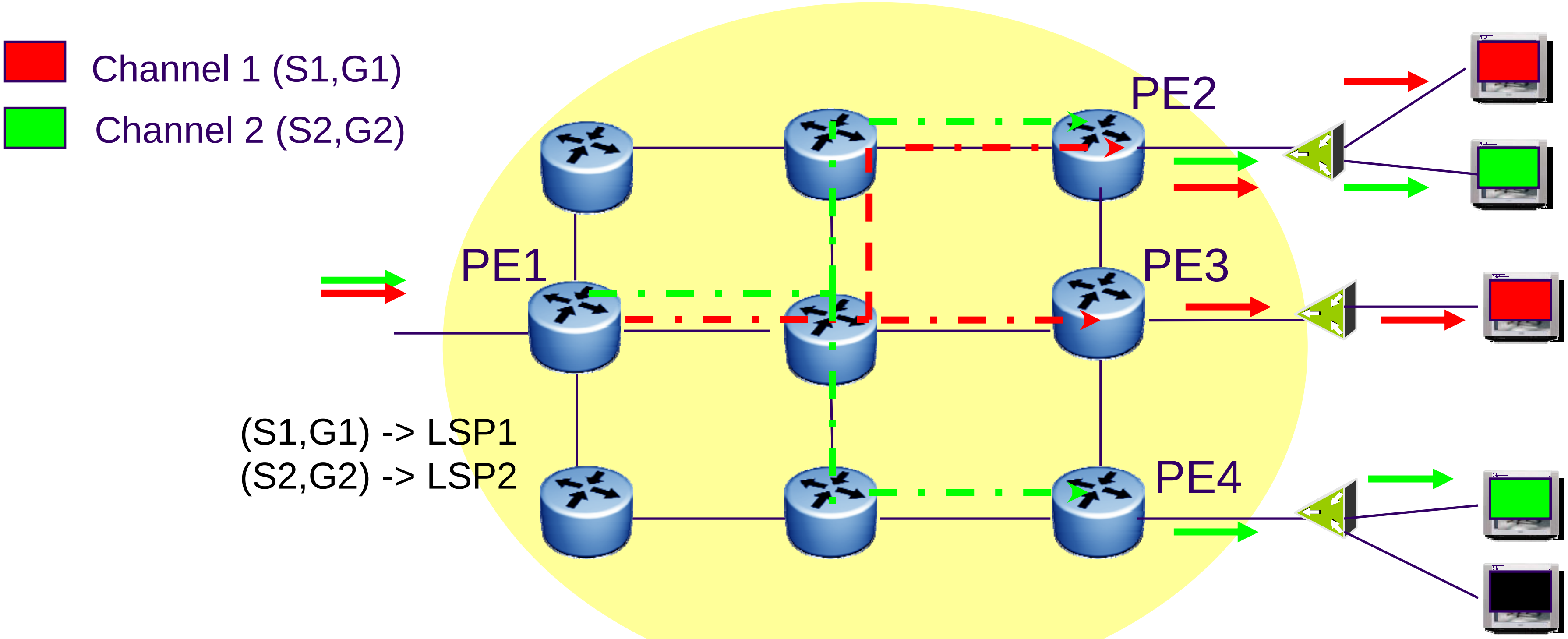
Supported since Junos 6.4

P2MP LSP1
Leaf: PE2, PE3, PE4

Static route (S1,G1) -> LSP1
Static route (S2,G2) -> LSP1
Static route (S3,G3) -> LSP1

- Fixes leaves => No dynamic Leaf addition/removal upon multicast receiver activity
- Static Routing => multicast traffic is statically routed within P2MP LSPs
- Traffic transported to PEs with no receivers => potential bandwidth wastings
- Well Suited when leaf PEs aggregate a lot of receivers

P2MP TE-LSP Utilization: Dynamic mode



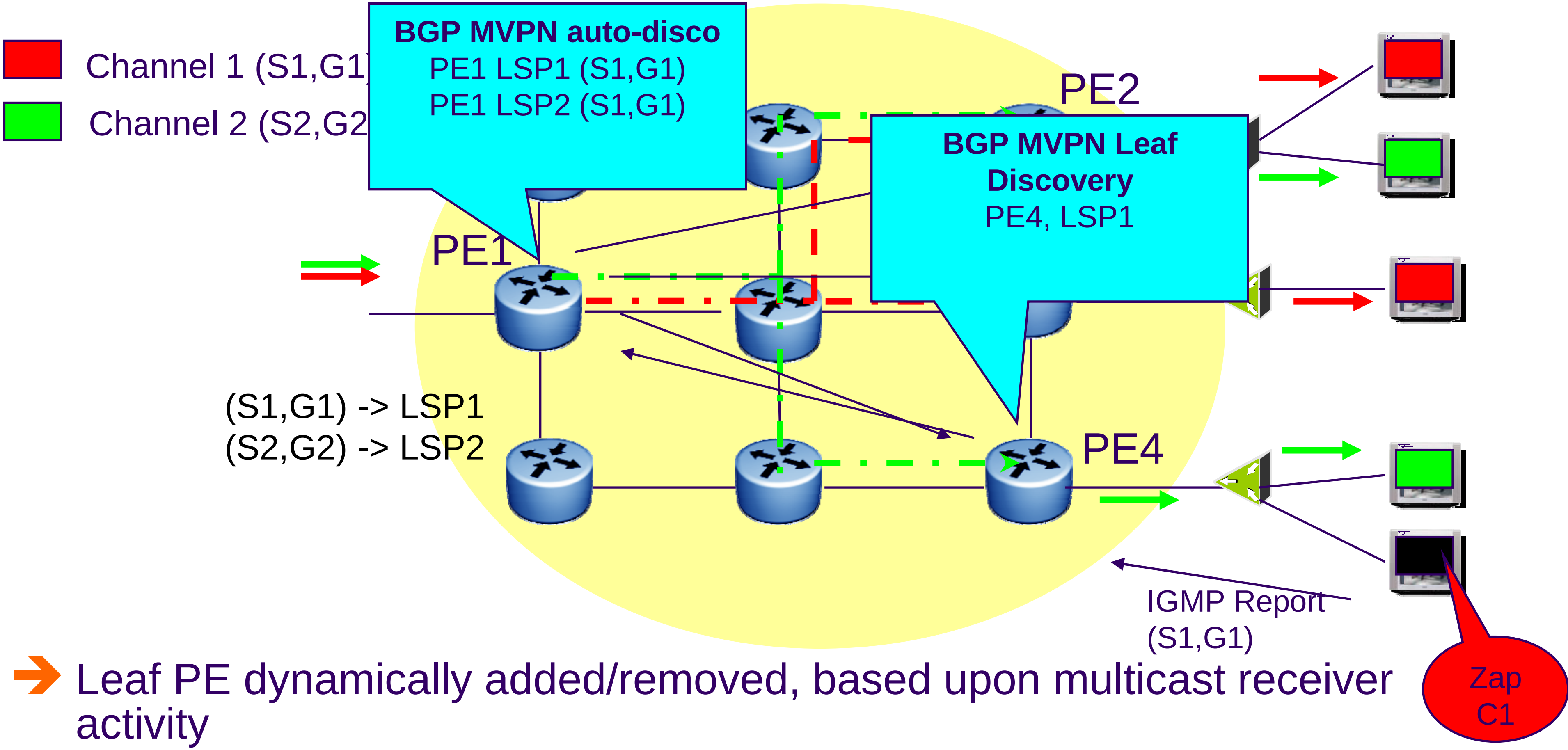
🟥 Channel 1 (S1,G1)

🟩 Channel 2 (S2,G2)

(S1,G1) -> LSP1
(S2,G2) -> LSP2

- Leaf PE dynamically added/removed, based upon multicast receiver activity
- One tree per channel that cover leaf PEs with receivers
- Mcast Traffic Dynamically routed in P2MP LSP
- Relies on BGP extensions for NGEN Multicast VPN
  - See draft-ietf-l3vpn-2547bis-mcast and draft-ietf-l3vpn-2547bis-mcast-bgp
- Well suited when Leaf PEs do not aggregate a lot of receivers

P2MP TE-LSP Utilization: Dynamic mode



🟥 Channel 1 (S1,G1

🟩 Channel 2 (S2,G2

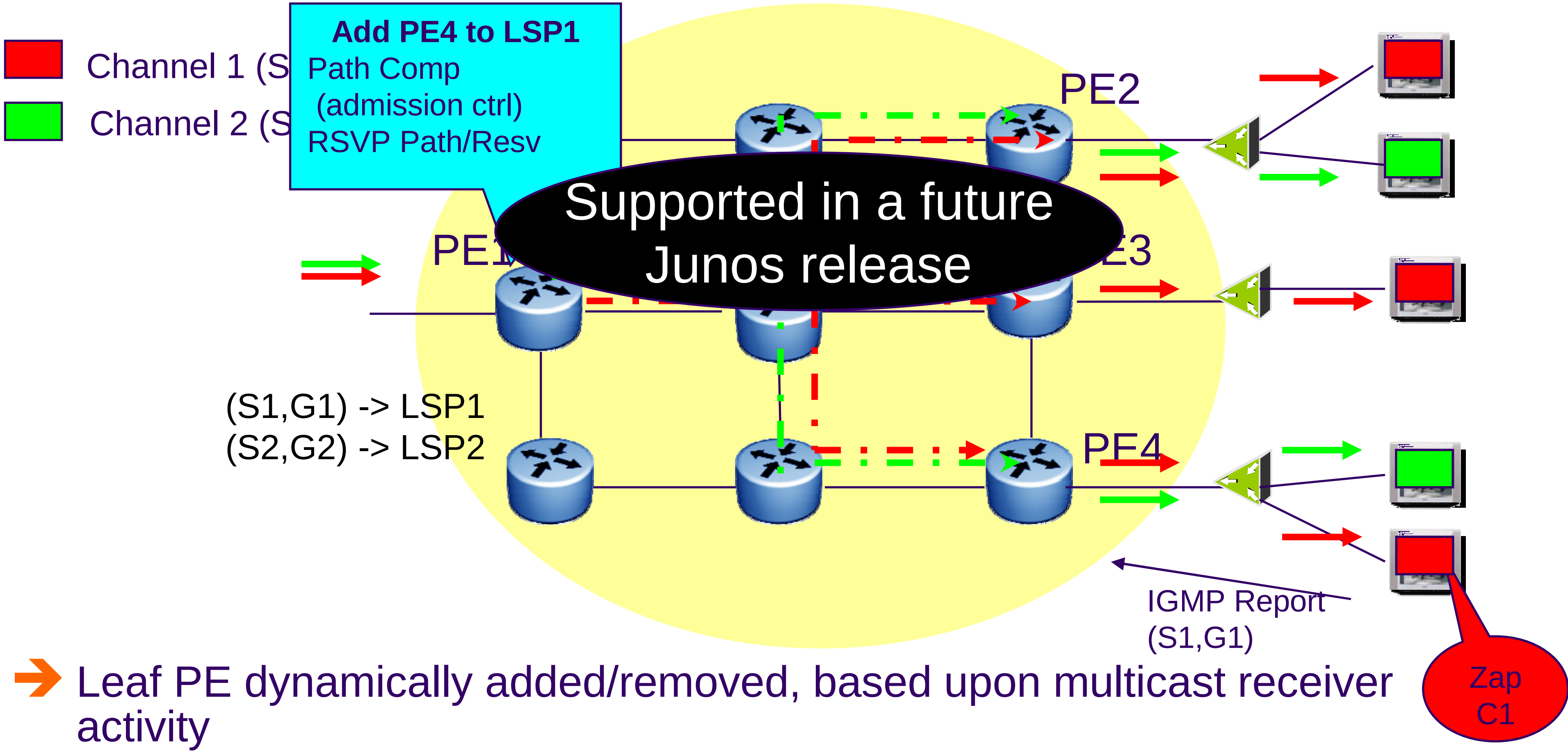
BGP MVPN auto-disco
PE1 LSP1 (S1,G1)
PE1 LSP2 (S1,G1)

BGP MVPN Leaf Discovery
PE4, LSP1

(S1,G1) -> LSP1
(S2,G2) -> LSP2

- Leaf PE dynamically added/removed, based upon multicast receiver activity
- One tree per channel that cover leaf PEs with receivers
- Mcast Traffic Dynamically routed in P2MP LSP
- Relies on BGP extensions for NGEN Multicast VPN
  - See draft-ietf-l3vpn-2547bis-mcast and draft-ietf-l3vpn-2547bis-mcast-bgp
- Well suited when Leaf PEs do not aggregate a lot of receivers

P2MP TE-LSP Utilization: Dynamic mode



🟥 Channel 1 (S

🟩 Channel 2 (S

Add PE4 to LSP1
Path Comp
(admission ctrl)
RSVP Path/Resv

Supported in a future Junos release

(S1,G1) -> LSP1
(S2,G2) -> LSP2

IGMP Report
(S1,G1)

Zap
C1

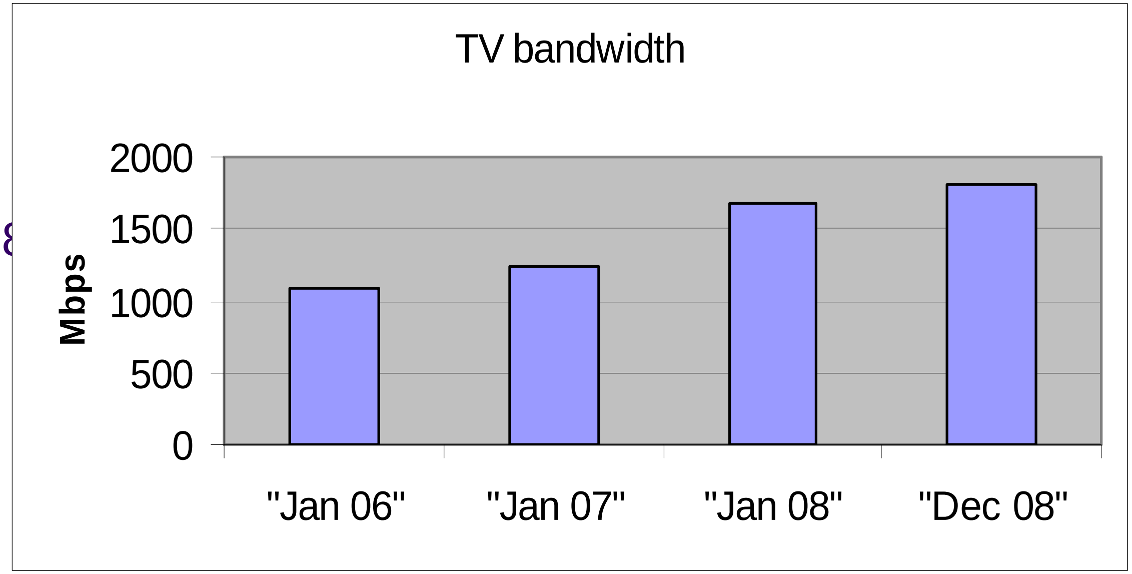
- Leaf PE dynamically added/removed, based upon multicast receiver activity
- One tree per channel that cover leaf PEs with receivers
- Mcast Traffic Dynamically routed in P2MP LSP
- Relies on BGP extensions for NGEN Multicast VPN
  - See draft-ietf-l3vpn-2547bis-mcast and draft-ietf-l3vpn-2547bis-mcast-bgp
- Well suited when Leaf PEs do not aggregate a lot of receivers

Case Study: Orange TV services

Orange IPTV Service Growth *(French Market)*

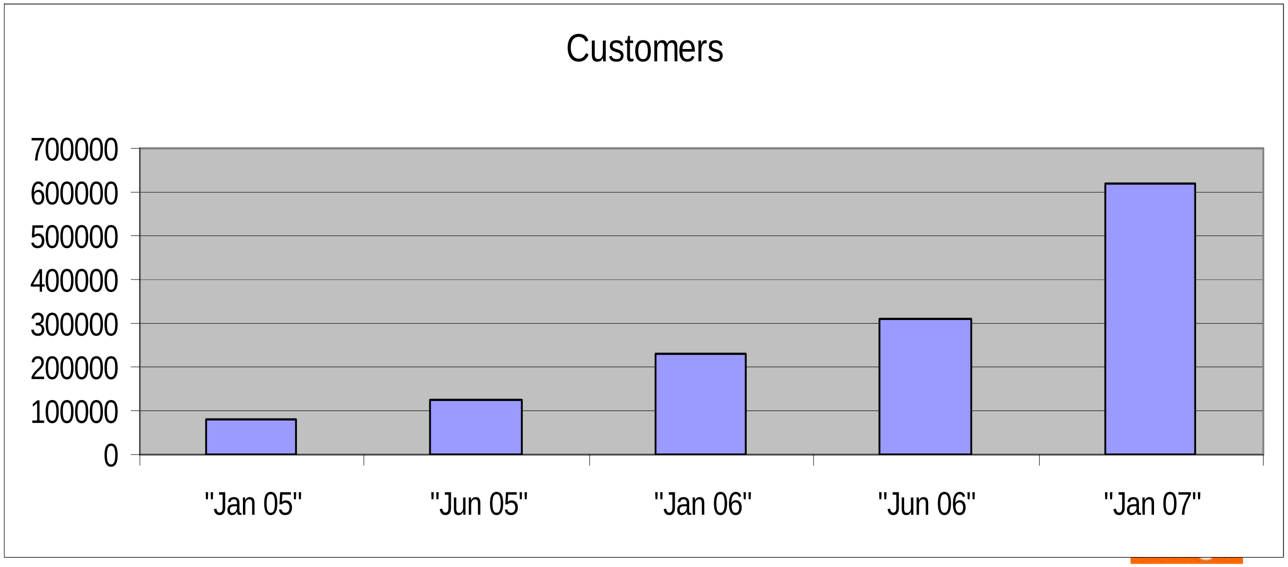
Total Bandwidth

- 1.8G Traffic at the end 2008
  - 470 LD channels (3Mbps)
  - 30 HD channels (12Mbps)

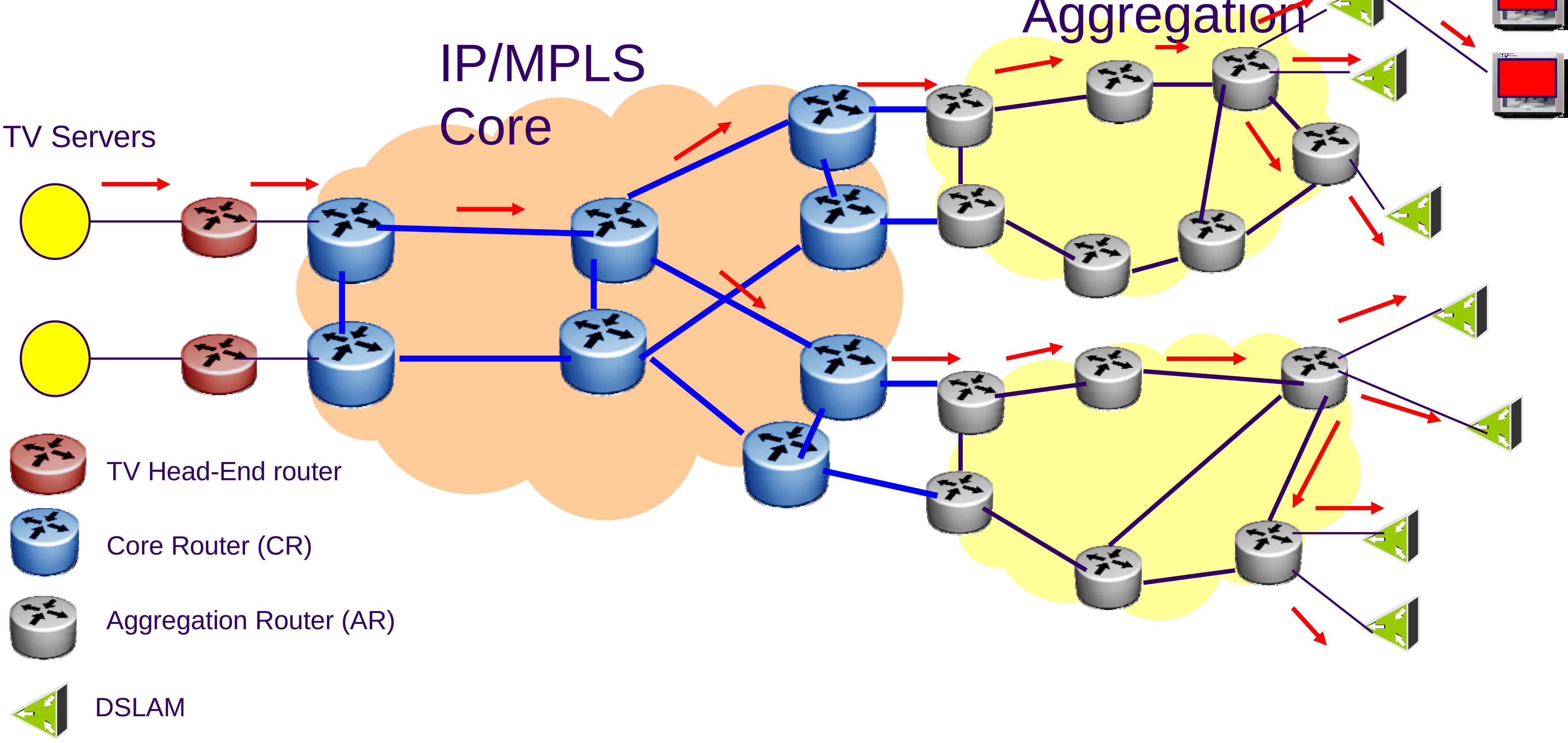


Customers

- > 1 million in Q2 2007



IPTV Transport



- IPTV transport used to rely on ATM P2MP PVC
- Now Orange is migrating TV to IP/MPLS Core & Agg networks
  - Drivers = Convergence; Need for higher bandwidth; need for dynamicity

Requirements for Carrier Class TV Broadcast Service Provider Side

- Resources optimization: Need for minimum cost trees
  - Required only in well meshed topology
- Dynamicity and Admission Control
  - Dynamicity and Admission control allow significant bandwidth savings and CAPEX reduction
    - A channel is transported only if there is a receiver
    - A requested channel is transported only if there are resources
      - Finer dimensioning
  - Useful in the Aggregation: Examples: Less than 80% channels requested in primary rings, less than 50% in secondary rings and on DSLAM links
  - Useless in the Core (near 100% channel requested)
- OAM: Need for fast fault detection/isolation, tree tracing, etc.

Requirements for Carrier Class TV Broadcast Customer Side

- Resiliency: Need to minimize the impact on images upon link or node failure:
  - Short term target: sub-1s
  - Mid term target: Sub-100ms

- QoS: Need for high QoS guarantees (packet loss, jitter)

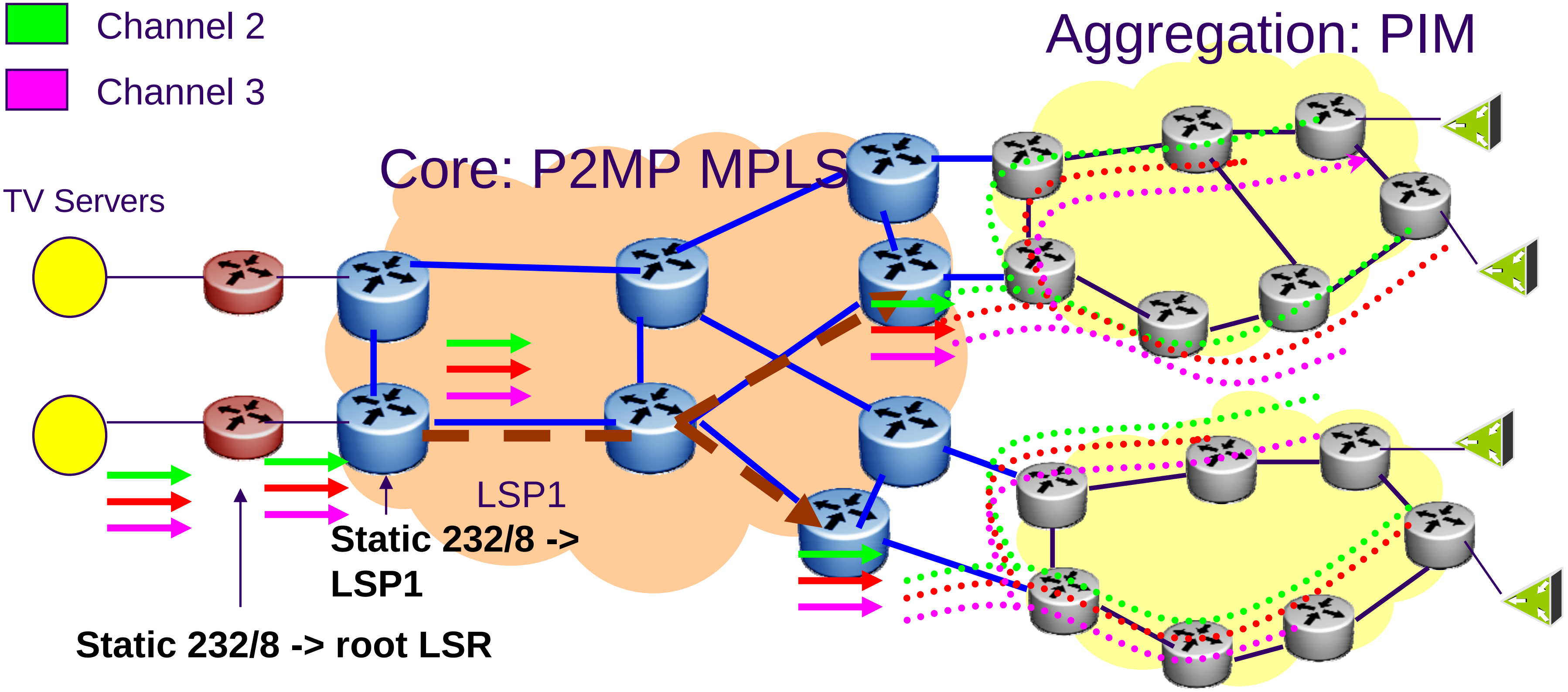
- P2MP MPLS-TE fits in well with all these requirements

Deployment scenarii for TV Broadcasting

- Three P2MP MPLS based scenarios under study
  - Scenario 1: P2MP MPLS in the Core and PIM in the Agg
  - Scenario 2: Contiguous P2MP MPLS in Core + Agg
  - Scenario 3: Non contiguous P2MP MPLS in Core + Agg

Scenario 1: MPLS Core & PIM Metro

🟥 Channel 1
🟩 Channel 2
🟪 Channel 3



- ➔ One "static" P2MP LSP that carries all channels from the Root down to all Aggregation PoPs: Static Broadcasting in the core
- ➔ IP Multicast (PIM) in the Aggregation, for dynamicity
- ➔ Keep stability and simplicity in the core

Scenario 2: MPLS Core + Metro

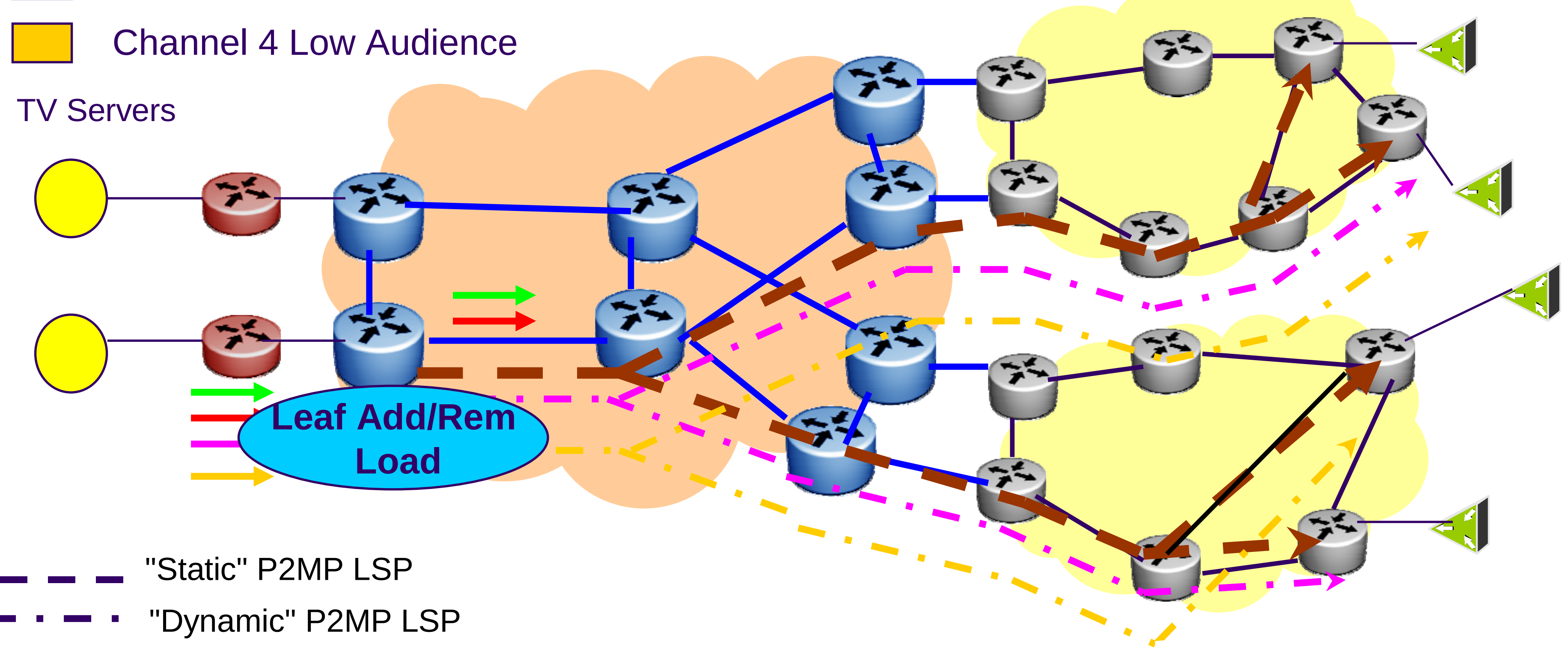
🟥 Channel 1 High Audience

🟩 Channel 2 High Audience

🟪 Channel 3 Low Audience

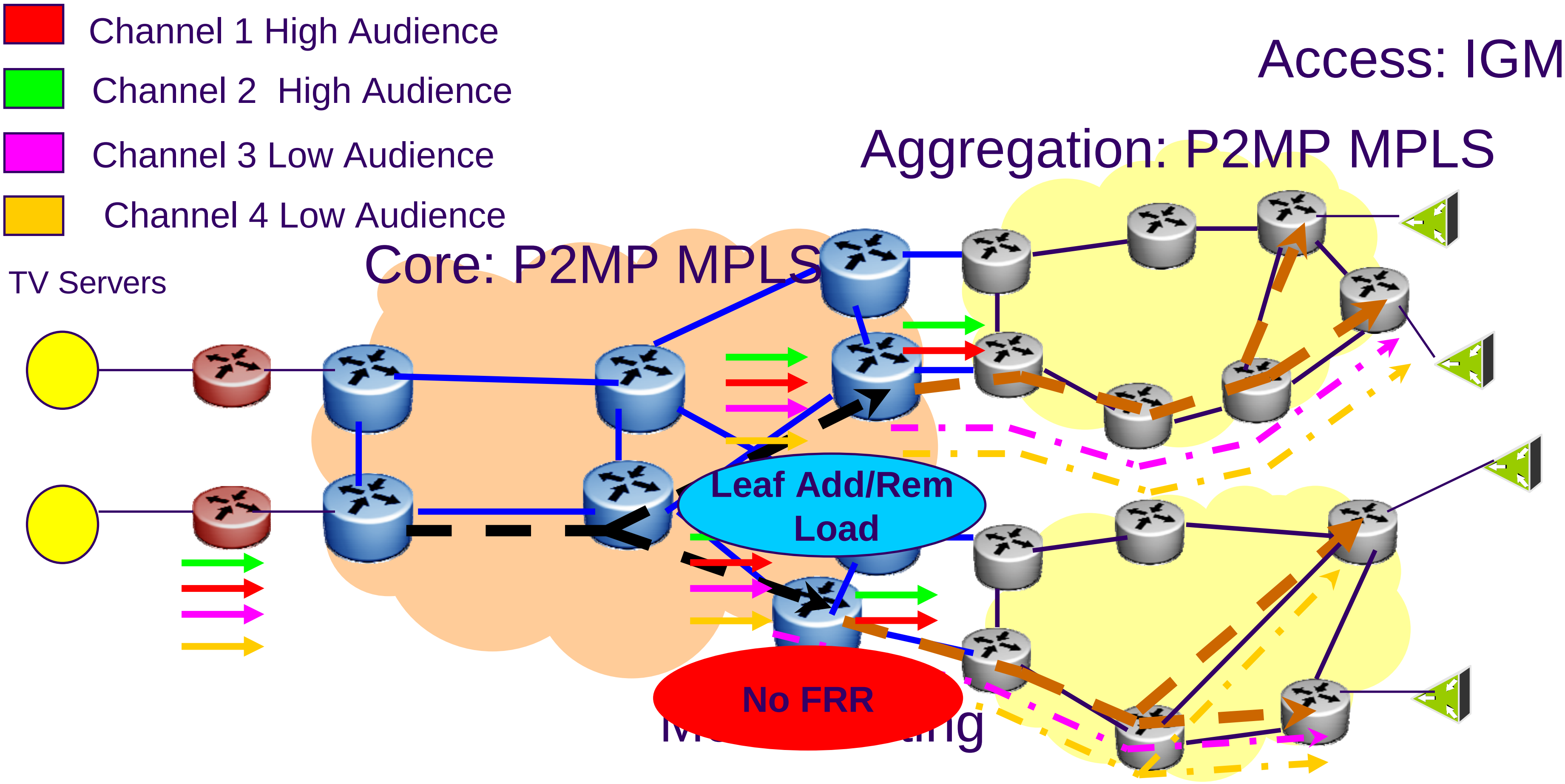
🟨 Channel 4 Low Audience

Core + Agg: Contiguous P2MP MPLS



- A set of "static" and "dynamic" P2MP LSPs from Root Router down to Agg Routers
  - High audience channels: A single static LSP that broadcasts all channels to all agg routers
  - Low audience channels: One dynamic P2MP LSP per channel to agg routers with receivers
- Potential Scalability Issue: All leaf add/remove load on the Root LSR

Scenario 3: MPLS Core + MPLS Metro



🟥 Channel 1 High Audience

🟩 Channel 2 High Audience

🟪 Channel 3 Low Audience

🟨 Channel 4 Low Audience

- Core: One "static" P2MP LSP that carries all traffic down to all Border Routers
- Metro: "Static" P2MP LSP for high audience, "dynamic" P2MP LSPs for low audience
- IP Multicast routing (static, or MVPN) between core and aggregation
- Scales better than S2: Leaf add/remove load is distributed on Border routers
- But no fast protection against failure of Border Routers

Analysis

	Dynamic Leaf addition/removal	**Admission Control**	**Tree optimization**	**Fast Reroute**
IP Multicast (PIM)	OK	NOK (only local AC)	NOK	NOK
P2MP MPLS-TE	OK with NGEN MVPN but not supported today	OK with NGEN MVPN but not supported today	OK	OK

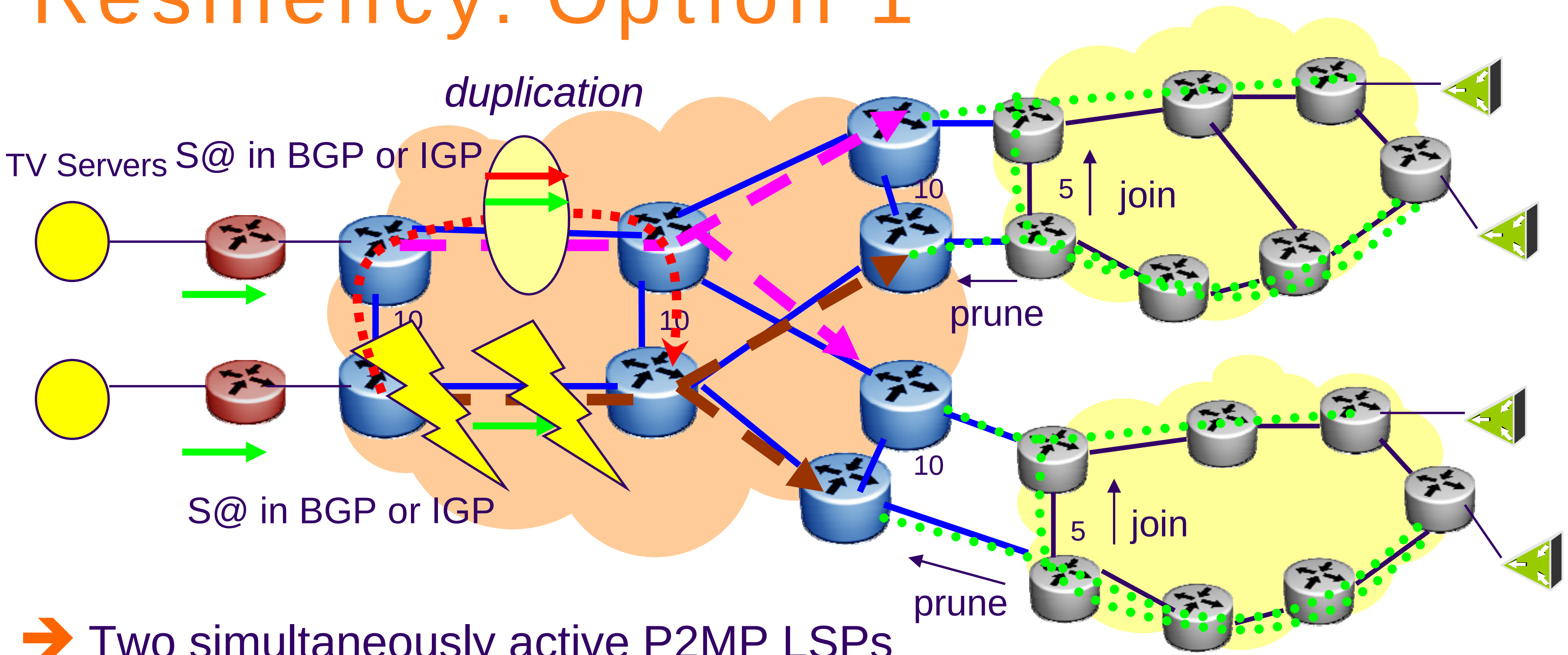
- P2MP MPLS-TE: Advance TE features. Current Lack of dynamicity that restricts its usage to the core only
- IP Multicast PIM: No advanced TE, but dynamicity
- Relevant design at short term = Scenario 1
  - Combine P2MP MPLS-TE in the core and IP Multicast in the aggreagtion
- When "dynamic" P2MP LSPs available migrate to scenario 2 or 3
  - Driver between scenarios 2 and 3 will rely upon scalability considerations

Recovery scenarios

Resiliency

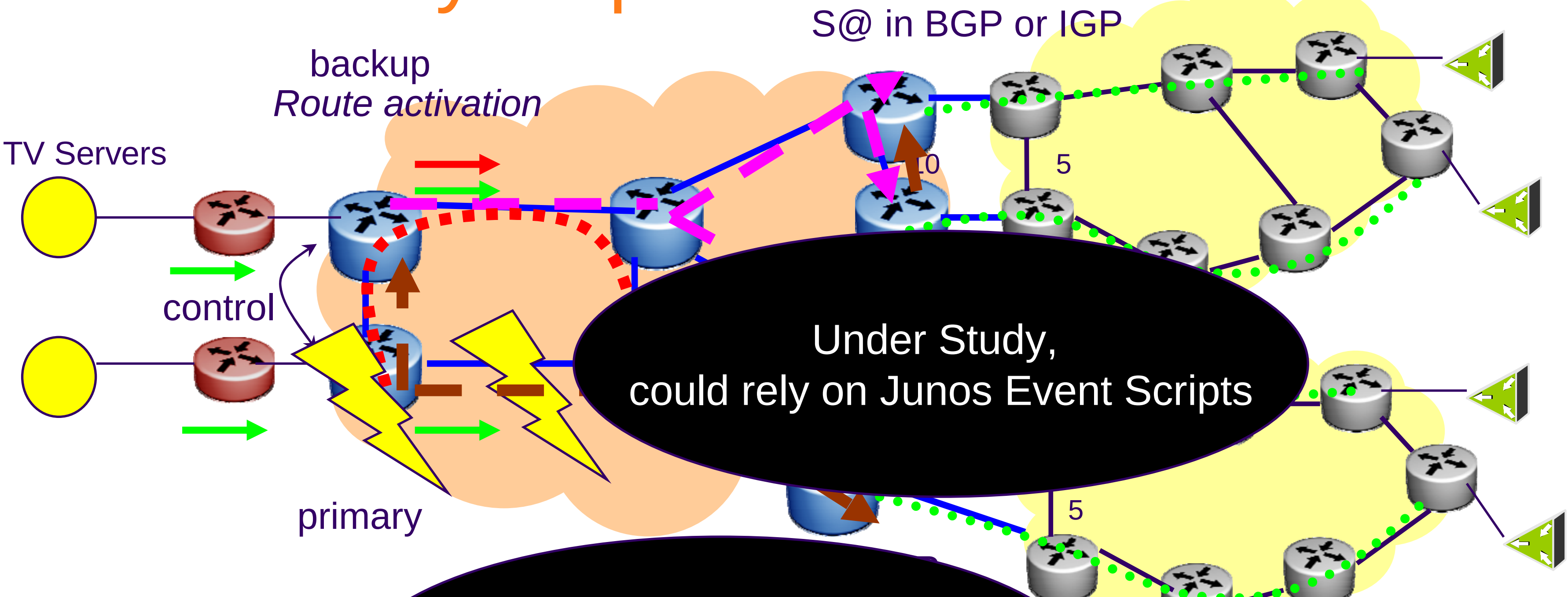
- Redundant TV Head-Ends: Traffic Transmitted twice with same source address, to two distinct core Routers
- Redundant Core: Odd and Even networks
- Several Approaches for P2MP LSP redundancy
  - Option 1: Two Trees serving distinct leaf PEs and simultaneously active
  - Option 2: Two Trees serving all leaf PEs and not simultaneously active
  - Option 3: Two Trees serving all leaves and simultaneously active
- A lot of options not covered here

Resiliency: Option 1



- Two simultaneously active P2MP LSPs
  - One P2MP LSP in the Odd network serving Odd Leaf LSRs
  - One P2MP LSP in the even network serving Even Leaf LSRs
- The same TV source @ is distributed by the two head-end Routers in IGP or BGP
- PIM rerouting triggered within the aggregation upon any failure in the core and at the edge (Head-End Router, Root LSR and Leaf LSR)
  - OK if odd-even transition cost lower in agg than in core (require metric tuning)
- Fast Reroute can be activated but lead to twice the traffic on some links

Resiliency: Option 2

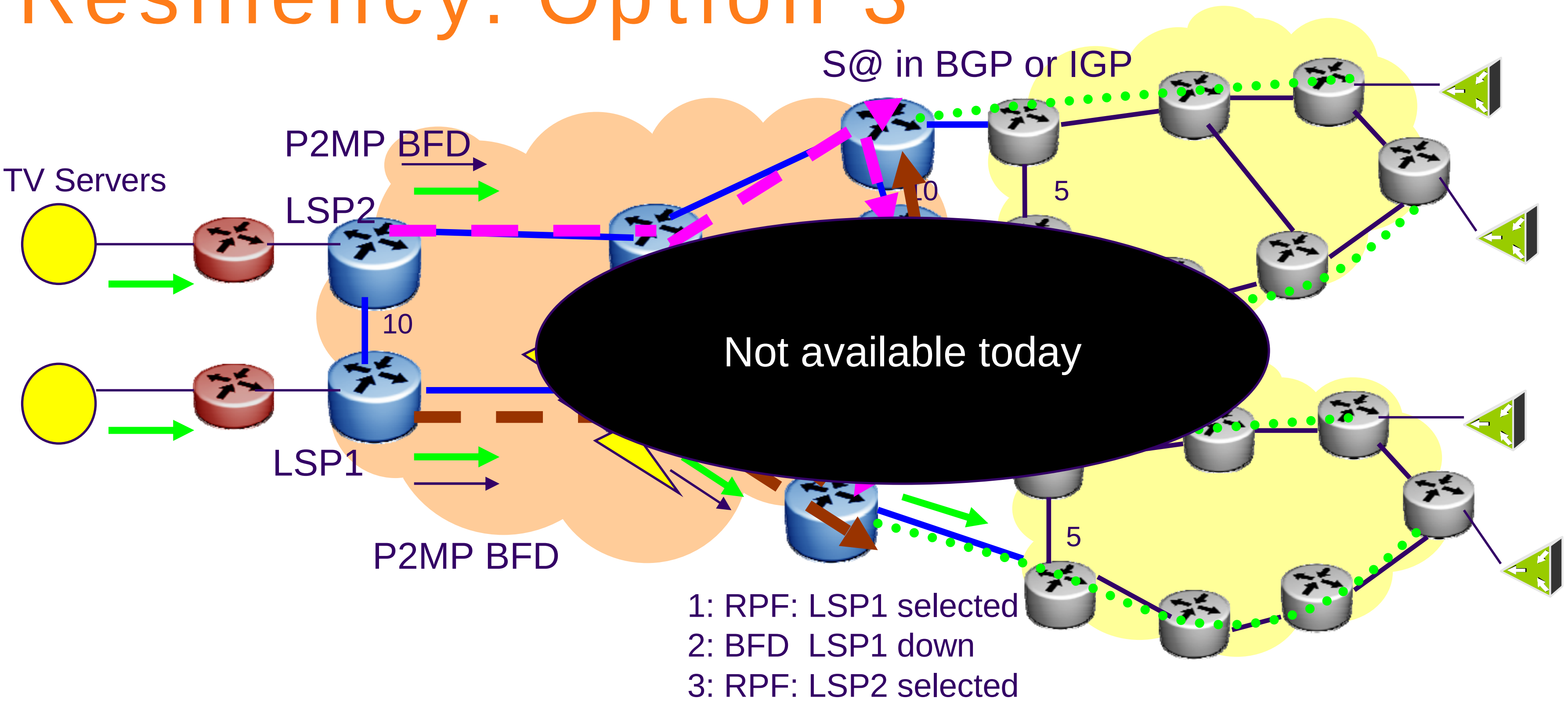


Under Study,
could rely on Junos Event Scripts

ntegration with MVPN will also provide
a switchover upon root LSR failure

- Two non simul
  - One P2MP L en Leaf LSRs
  - One P2MP LS d Odd Leaf LSRs
- LSP Rerouting, and/or ctivated
- The TV source @ is distributed in aggregation by Leaf LSRs, in IGP or BGP
- PIM rerouting triggered only upon Leaf LSR failure and Aggregation failure
- Control of LSP Activity: When the Backup Root LSR detects that the source is no longer reachable via the Primary Root LSR, it activates the backup Tree

Resiliency: Option 3



- Two simultaneously active P2MP LSPs
  - One P2MP LSP in the Odd network serving Odd and Even Leaf LSRs
  - One P2MP LSP in the Even network serving Even and Odd Leaf LSRs
- A Leaf LSR receives the traffic twice (on the two LSPs)
  - RPF check on LSP interfaces to select reception on a single LSP (PHP deactivation)
  - P2MP BFD running in the two P2MP LSPs
  - LSP failure detected thanks to P2MP BFD => RPF updated to the other LSP

Resiliency: Analysis

	Fast Recovery	*Optimality*	*Required features*
Option 1	-	+ (no - (FRR) (Twice the traffic)	
Option 2	+ + (core) - (Root) (FRR in the core, Root Switchover)	+ + Traffic only on odd in nominal case	Root Switchover control feature
Option 3	+ Depends on the BFD interval and RPF update	+ Traffic on odd and even in nominal case	P2MP BFD and LSP based RPF

Orange Short term option

Ongoing deployment in FT Group

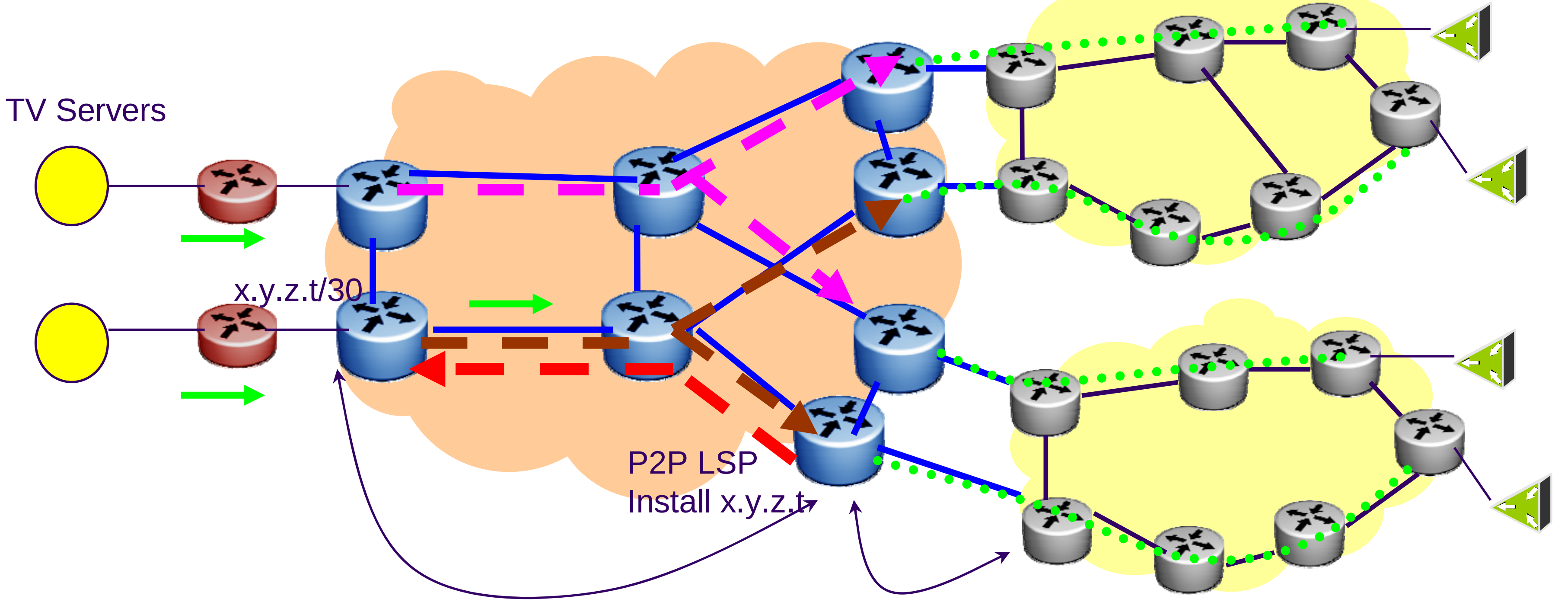
- One deployment soon in one domestic network, for DSL TV
  - Full Juniper core network

- P2MP MPLS-TE in the core and PIM in the aggregation

- Recovery Option 1, but with some specificities: Option 1bis

- Successful lab testing and field trial, with always sub-second convergence

Resiliency: Option 1 bis



TV Servers

x.y.z.t/30

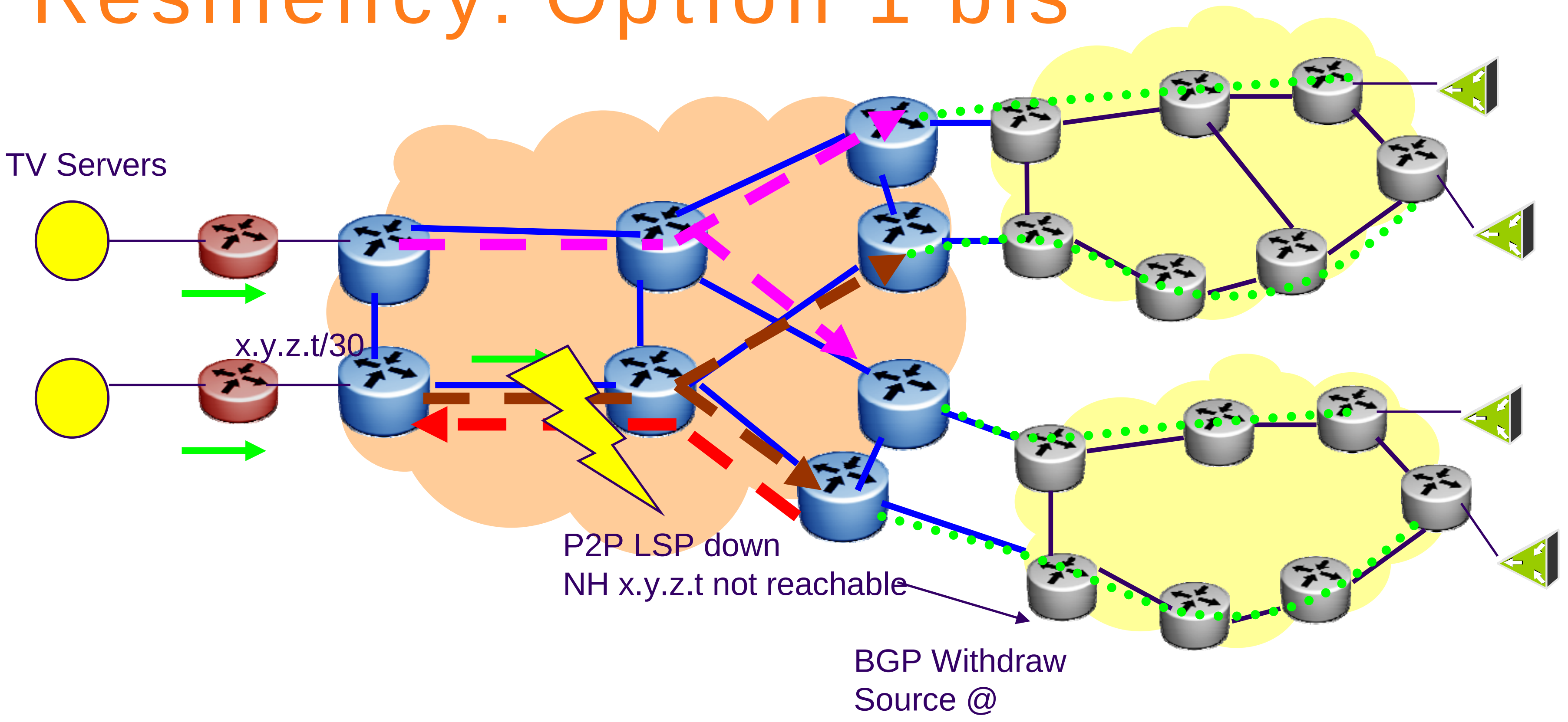
P2P LSP
Install x.y.z.t

BGP session
Advertise Source @
NH x.y.z.t

BGP session
Advertise Source @

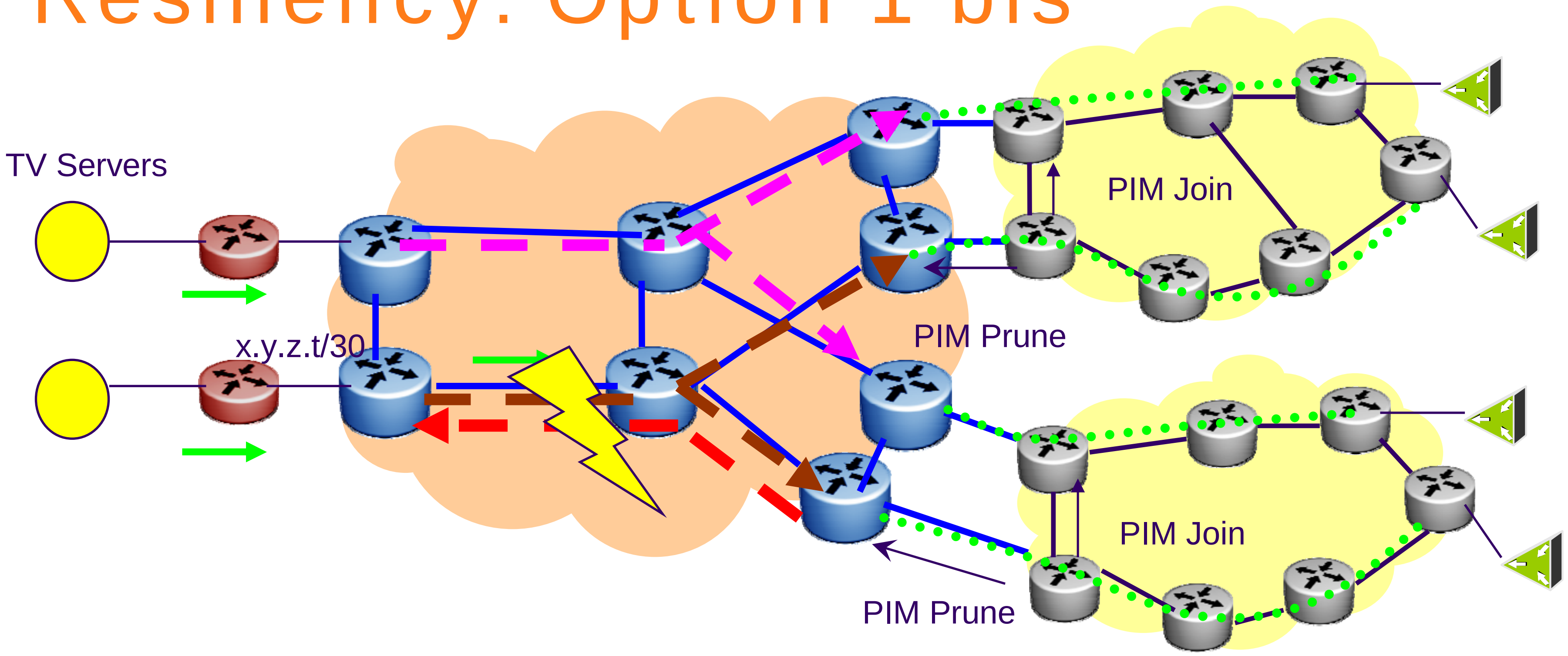
- Not always possible to play with metrics so that a failure in the core triggers rerouting in the backhaul
- Solution = a static P2P LSP from a Leaf PE to the Root PE, that follows the tree path in the reverse direction
- iBGP session root PE - Leaf PE with a next-hop reachable via the P2P LSP only
- Failure => the P2P LSP is down => BGP next hop no longer reacheable => Source withdrawn => PIM convergence

Resiliency: Option 1 bis



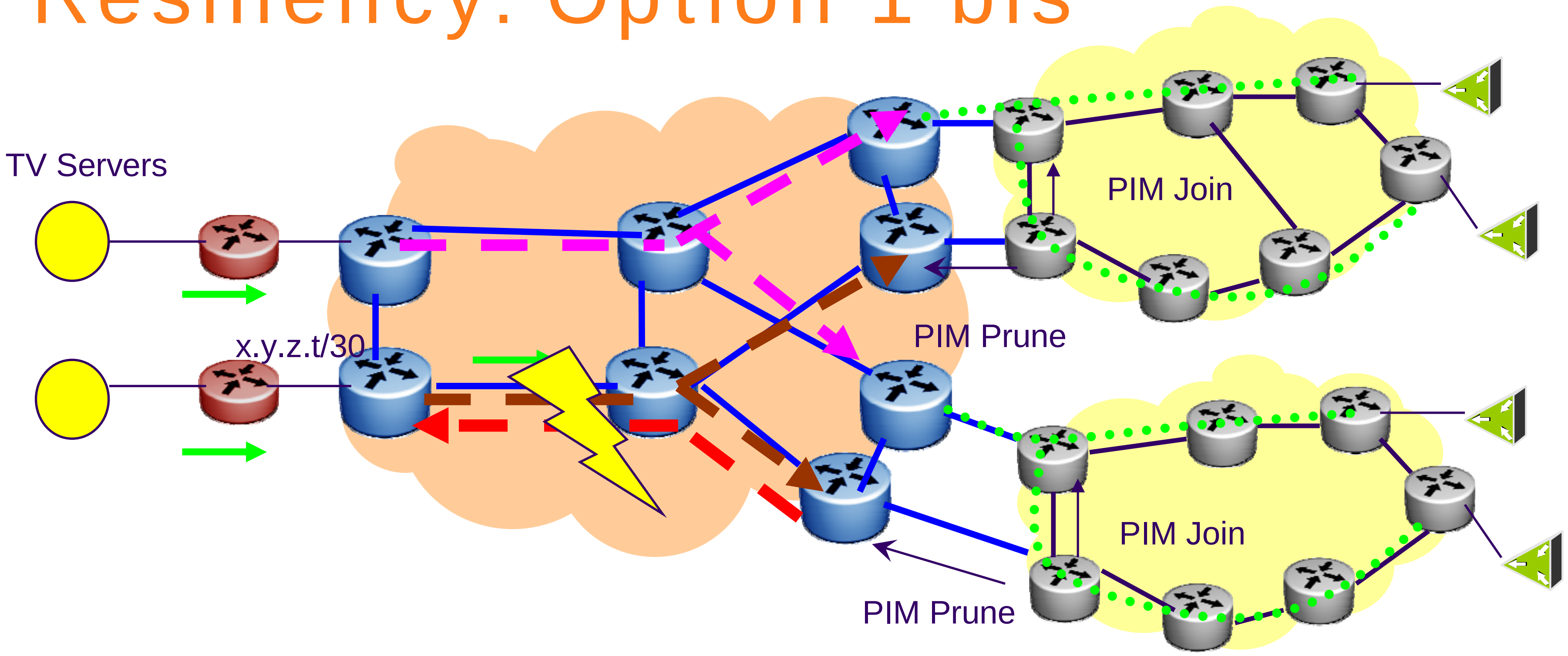
- Not always possible to play with metrics so that a failure in the core triggers rerouting in the backhaul
- Solution = a static P2P LSP from a Leaf PE to the Root PE, that follows the tree path in the reverse direction
- iBGP session root PE - Leaf PE with a next-hop reachable via the P2P LSP only
- Failure => the P2P LSP is down => BGP next hop no longer reacheable => Source withdrawn => PIM convergence

Resiliency: Option 1 bis



- Not always possible to play with metrics so that a failure in the core triggers rerouting in the backhaul
- Solution = a static P2P LSP from a Leaf PE to the Root PE, that follows the tree path in the reverse direction
- iBGP session root PE - Leaf PE with a next-hop reachable via the P2P LSP only
- Failure => the P2P LSP is down => BGP next hop no longer reacheable => Source withdrawn => PIM convergence

Resiliency: Option 1 bis



- Not always possible to play with metrics so that a failure in the core triggers rerouting in the backhaul
- Solution = a static P2P LSP from a Leaf PE to the Root PE, that follows the tree path in the reverse direction
- iBGP session root PE - Leaf PE with a next-hop reachable via the P2P LSP only
- Failure => the P2P LSP is down => BGP next hop no longer reacheable => Source withdrawn => PIM convergence

Ongoing studies

- Fast Recovery upon source failures
  - Various options under study including measurement based
    - Measure multicast traffic load and trigger routing events
  - Could rely on Junos Event Scripts

- Fast Reroute with P2MP Bypass Tunnels
  - Allows for link and node protection with significant bandwidth savings

- Combination of Fast Reroute and MPEG Error Correction (FEC)
  - MPEG COP#3 FEC allows correction upon missing N successive frames
  - Combined with FRR this could allow completely avoiding the impact of packet loss upon failure
    - This requires FRR perf allowing less than N loss upon failure
    - E.g with a 300pps flow and N=10 we need FRR < 30ms => Achievable

Closing Remarks

- IPTV services rapidly growing
- P2MP MPLS-TE well suited to IPTV
- Dynamicity not yet supported => restrict the usage today to the core only
- Integration with NGEN Multicast VPN will bring useful features
  - Dynamic leaf addition/removal, Admission Control, Root resiliency…
  - Allows extending the scope of P2MP MPLS-TE to aggregation networks
- To be deployed in one Orange Group network soon
  - Juniper P2MP MPLS in the core & IP Multicast PIM in the metro
  - Successful lab testing and field trial

References

- Yasukawa, S. et al. "Signaling Requirements for P2MP TE-LSPs", <RFC 4461>, April 2006
- Aggarwal, R., Papadimitriou, D., Yasukawa, S.,."Extensions to RSVP-TE for Point-to-Multipoint TE LSPs" <RFC 4875>, May 2007
- Swallow, G., Nadeau, T., Aggarwal, R., "Connectivity Verification for Multicast Label Switched Paths", <draft-ietf-mpls-mcast-cv>
- Rosen, E., Aggarwal, R., et al. "Multicast in MPLS/BGP IP VPNs" <draft-ietf-l3vpn-2547bis-mcast>
- R. Aggarwal, E. Rosen, T. Morin, Y. Rekhter, C. Codeboniya, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs" <draft-ietf-l3vpn-2547bis-mcast-bgp>
- Le Roux, J.L., Aggarwal, R, et al. "P2MP MPLS-TE Fast Reroute with P2MP Bypass Tunnels", <draft-ietf-mpls-p2mp-te-bypass>

THANKS !