Multicast "primer"

Per Jensen, per@cisco.com

December 2006

Agenda

- IP multicast – why SSM
- Resiliency
- Channel changing
- Admission control / CAC

IP multicast primer (SSM)

… as required for IPTV…

The key IP multicast services

- **Network level Modes/Services:**
  - **ASM – The well known IP Multicast**
    - Since 1995, lots of innovations, PIM-SM -> Bidir-PIM
  - **SSM – Source Specific Multicast**
    - Easier to deploy/manage, provides implicit protection.
    - Cisco has set of transition solutions for this!
    - Requires server/receiver applications support

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!)
    - Simple point to multipoint tree building == (S,G) SPTs only
- Cisco, (IETF ?)
  - Source redundancy (option – other options too)
    - Anycast/Prioritycast source addresses with RIPv2 signaling

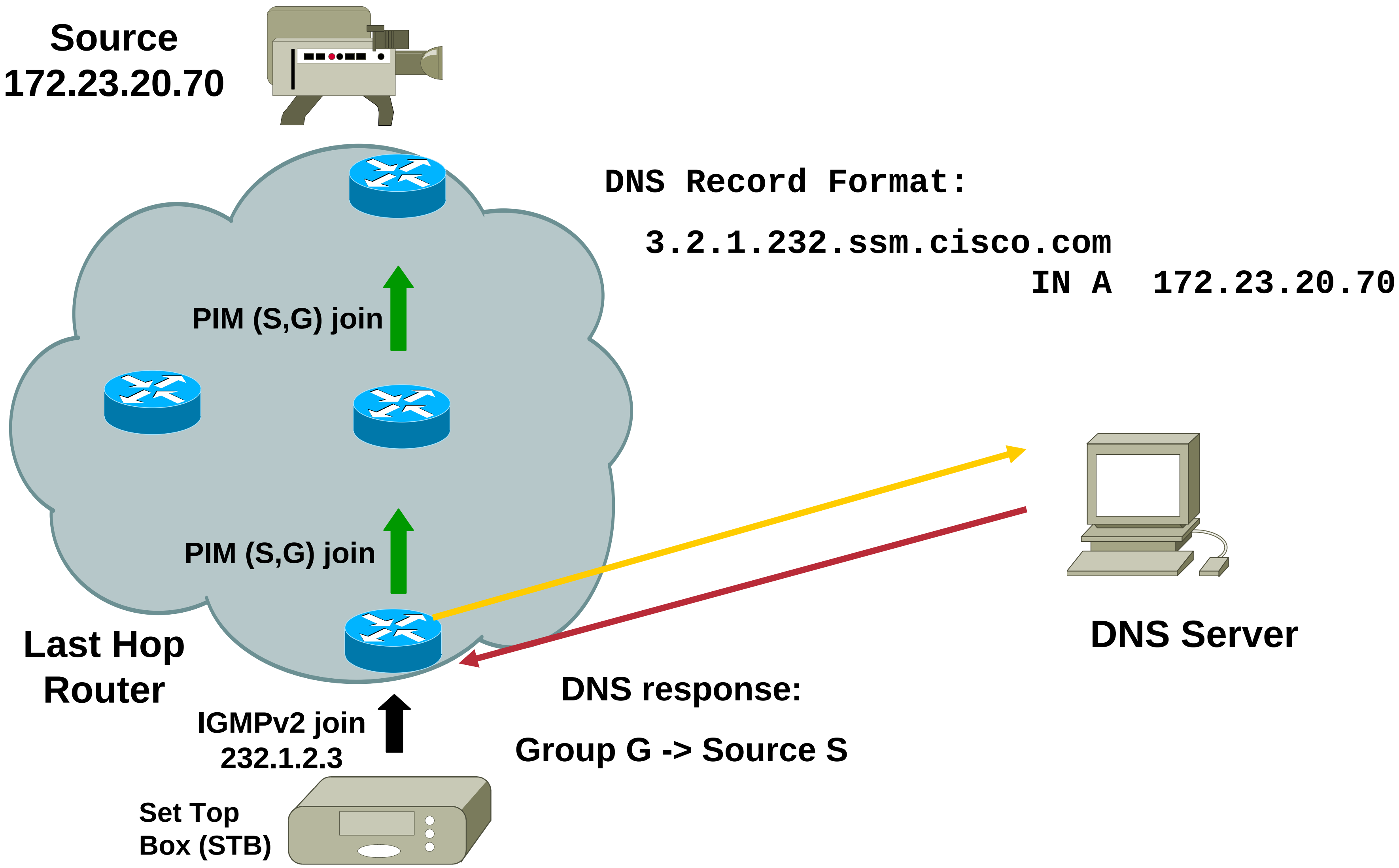
Host-to-router signalling
The issue with IGMPv3 / SSM ?

- **While SSM was “invented”in 1999/2000 and has been available in routers since end of 2000, support on receiver hosts has been lagging**
  - **Windows XP and Linux started to support it in 2001 – 2003**
  - **Set Top Box Vendors – need to be pushed by the SP customers, and those had in the past issues recognizing the importance of SSM.**
  - **L2 IGMP snooping devices similar story**

SSM Mapping

- **One of three Cisco SSM transition solutions**
  - **URD: If application is started from browser**
  - **IGMPv3lite: Application can support SSM, but OS not**
  - **SSM mapping: Host can not do SSM at all**
- **Last Hop Router maps group to source(s)**
  - More than one sources when redundancy is needed
  - Static mapping for quick deployment / testing
  - Mapping via DNS for operational separation
    - Network operator of last hop router only needs to configure the feature (3 CLI lines)
    - Application operator (TV Broadcast etc..) operates DNS server that provides the mapping
    - Delivers simplification benefit to network operator

SSM Mapping – DNS Example



Configuration

Enabling SSM mapping on the router

```
ip igmp ssm-map enable
```

For static mapping:

```
ip igmp ssm-map static <acl-1> <source-1 IP address>
ip igmp ssm-map static <acl-2> <source-2 IP address>
```

For DNS mapping (existing commands):

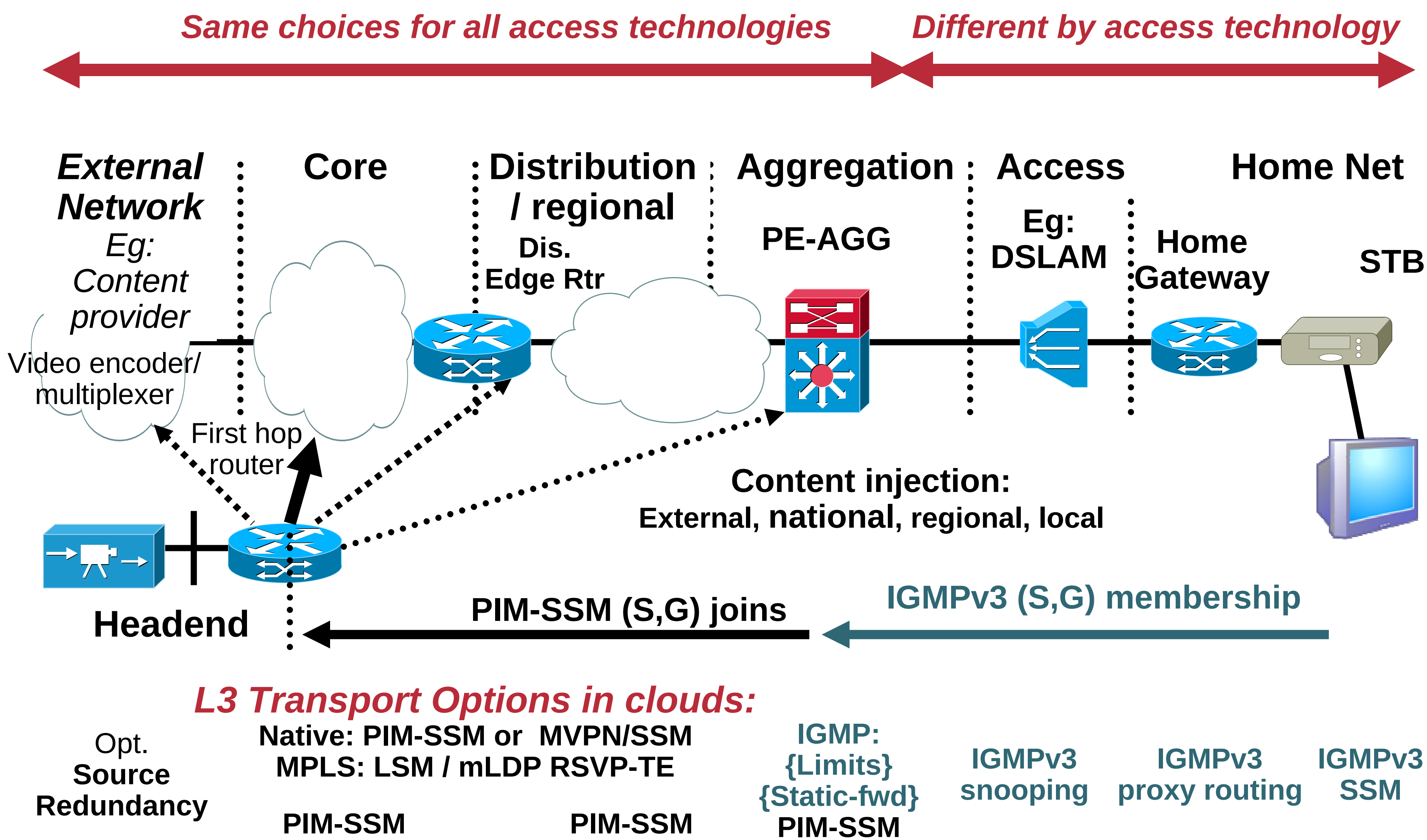
```
ip domain-server <ip address>
ip domain-name <domain.com>
```

To disable DNS mapping

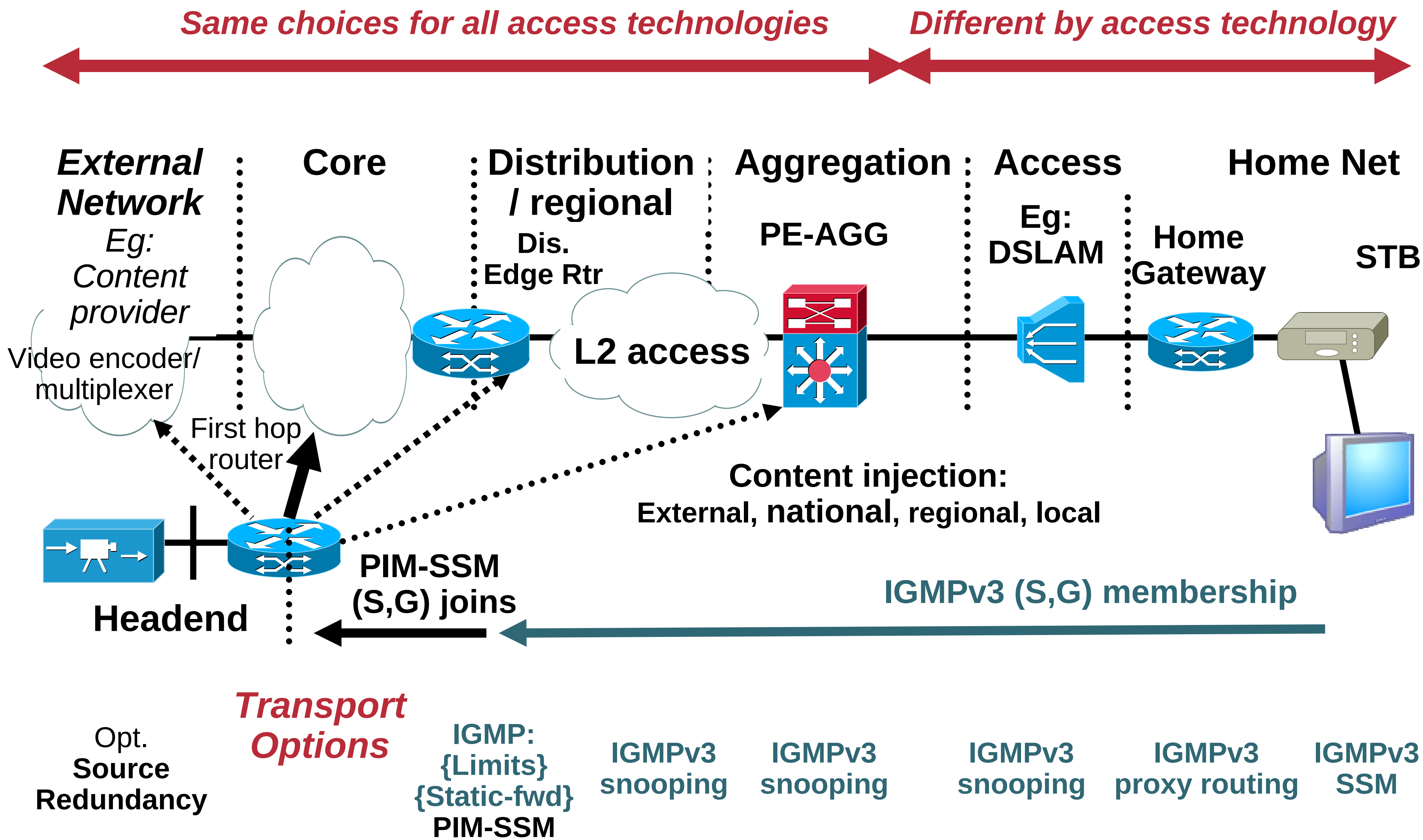
```
no ip igmp ssm-map query dns
```

```
DNS Record Format:   3.2.1.232       IN A   172.23.20.70
```

End-to-end protocol view



End-to-end protocol view with L2 distribution



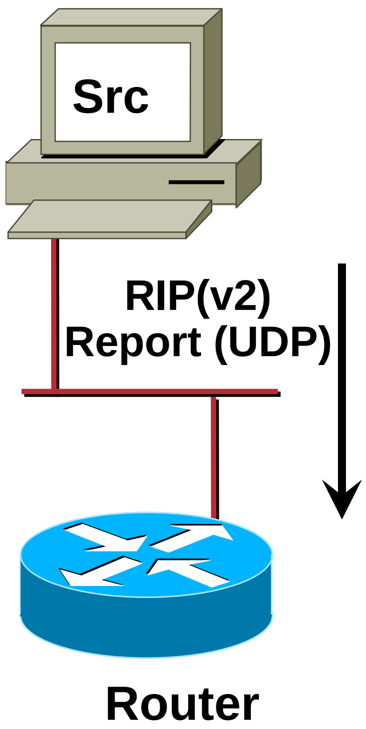
Resiliency

Redundancy models

- **Dual streams (1+1 RTP sessions)**
  - **Let the receiver decide which one to take**
- **Heartbeat**
  - **Active sends periodic hello to standby (muted) source**
- **Receiver driven**
  - **Same group with two sources. STB decides which one to join using IGMPv3**
- **Anycast-Source**
  - **Two (or more) sources actively sending with same origin IP address**
  - **Network decides which one to use using its metrics**
  - **Disaster-recovery and redundant headend applications**

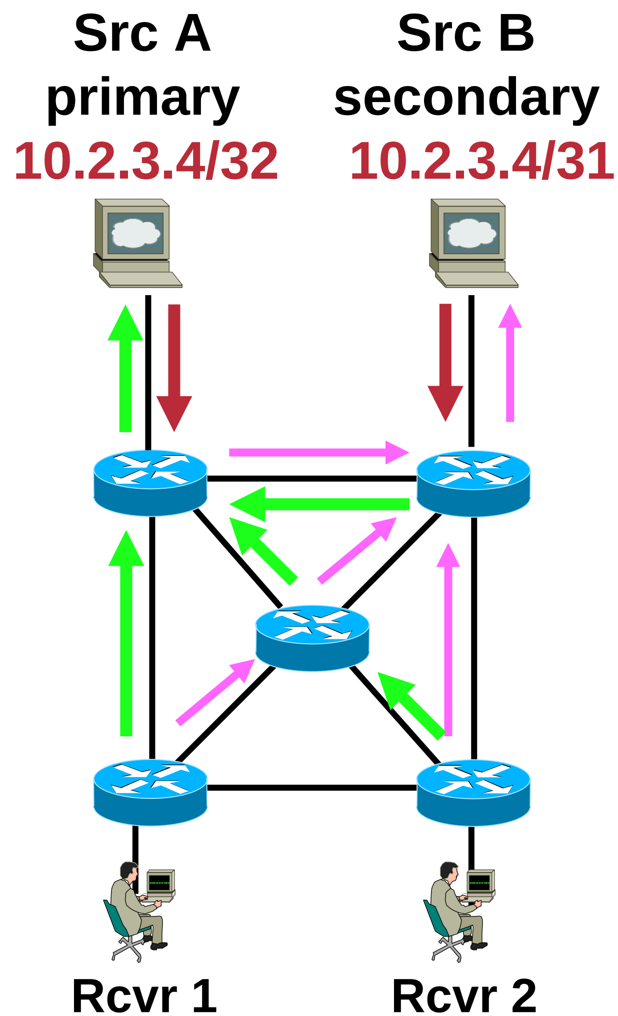
Source Redundancy
Anycast/Prioritycast signaling

- **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 periodicy for fast failure detection**
  - **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, DNS**
- **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**
- **Move source instances “freely” around the network**
  - **Runs within IGP area**
  - **Not good for regional to national encoder failover (BGP)**
- **No vendor proprietary source sync protocol required**

Multicast Fast Convergence

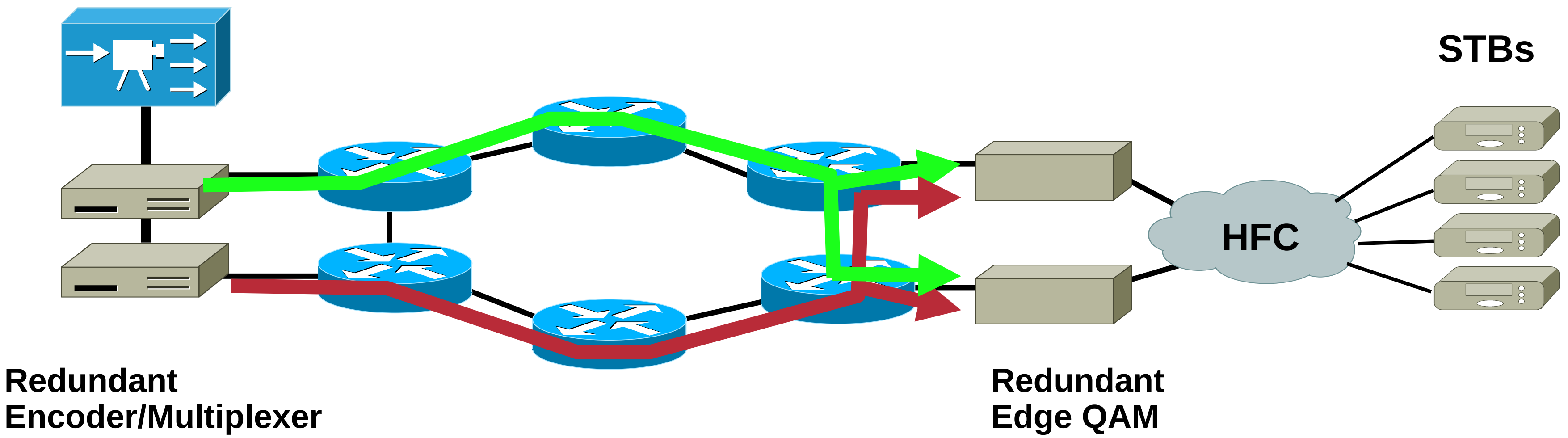
- IP multicast
  - Failures / topology changes are compensated for by reconverging the trees
  - Reconvergence time is sum of:
    - Failure detection time (only for failure cases)
    - Unicast reconvergence time
    - ~ No. Multicast-trees (PIM reconvergence time)
  - Possible
    - ~ 300 msec … 400 msec for few hundred trees
    - >= 1 sec for >= 1000 trees ?

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**

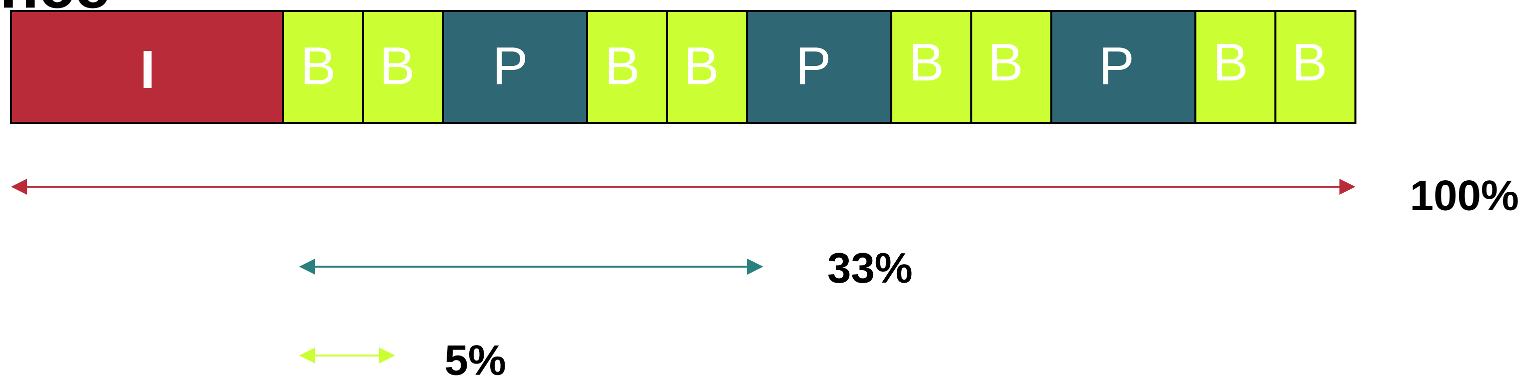
Stream redundancy with path separation
Candidate example from broadcast-TV in cable



- **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: MPEG timestamp).**
  - **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.**

GOP and network failures

- **Group Of Pictures (GOP) and frame relevance**



- **MPEG-2 GOP of size 12 => 480 msec (25 fps)**
  - **Hitting an I-frame (> 20% probability) affects the whole GOP**
  - **< 300 msec should cause a single-GOP loss, i.e. minor glitch**
  - **STB vendor dependencies**

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.**

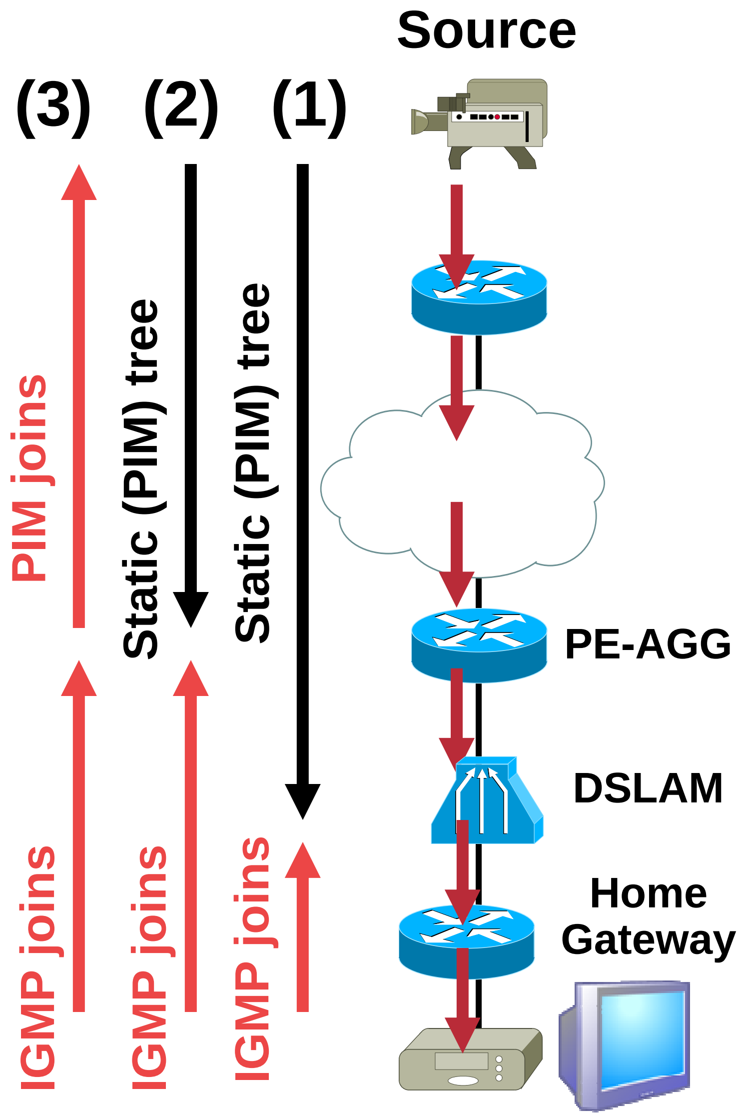
Channel changing

The (should be) obsolete problem

- **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 Resolved with IGMPv3/MLDv2**
  - **Ability for explicit tracking (vendor specific)**
  - **Can immediately stop forwarding upon leaves**

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



Fast Join/Leave for Faster Channel Change

Problem Description:

In networks where bandwidth is constrained between multicast routers and hosts (like in xDSL deployments), fast channel changes can easily lead to *bandwidth oversubscription*, resulting in a temporary degradation of traffic flow for all users.

Solution:

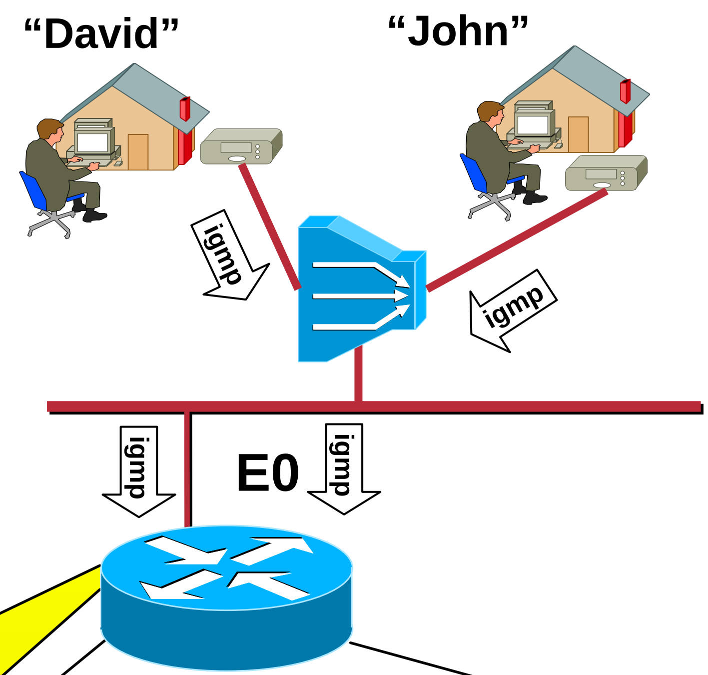
Reduce the leave latency during a channel change by extending the IGMPv3 protocol.

Benefits:

- **Faster channel changing without BW oversubscription**
- **Improved diagnostics capabilities**

Multicast Fast Join/Leave for Faster Channel Change

- **Relies on IGMPv3**
- **Router tracks both User and Channel(s) being watched**
- **When user leaves channel no one else is watching, router immediately prunes the channel off the interface compared to IGMPv2 (up to 3 seconds) and IGMPv1 (up to 180 seconds)!**



Configuration:

```
interface Ethernet 0
 ip pim sparse-mode
  ip igmp version 3
  ip igmp explicit-tracking
```

Int	Channel	User
~~E0~~	~~10.0.0.1, 239.1.1.1~~	~~"David"~~
E0	10.0.0.1, 239.2.2.2	"John"
E0	10.0.0.1, 239.3.3.3	"David"

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
  - RSVP for multicast admission control (Unicast/VoD)

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

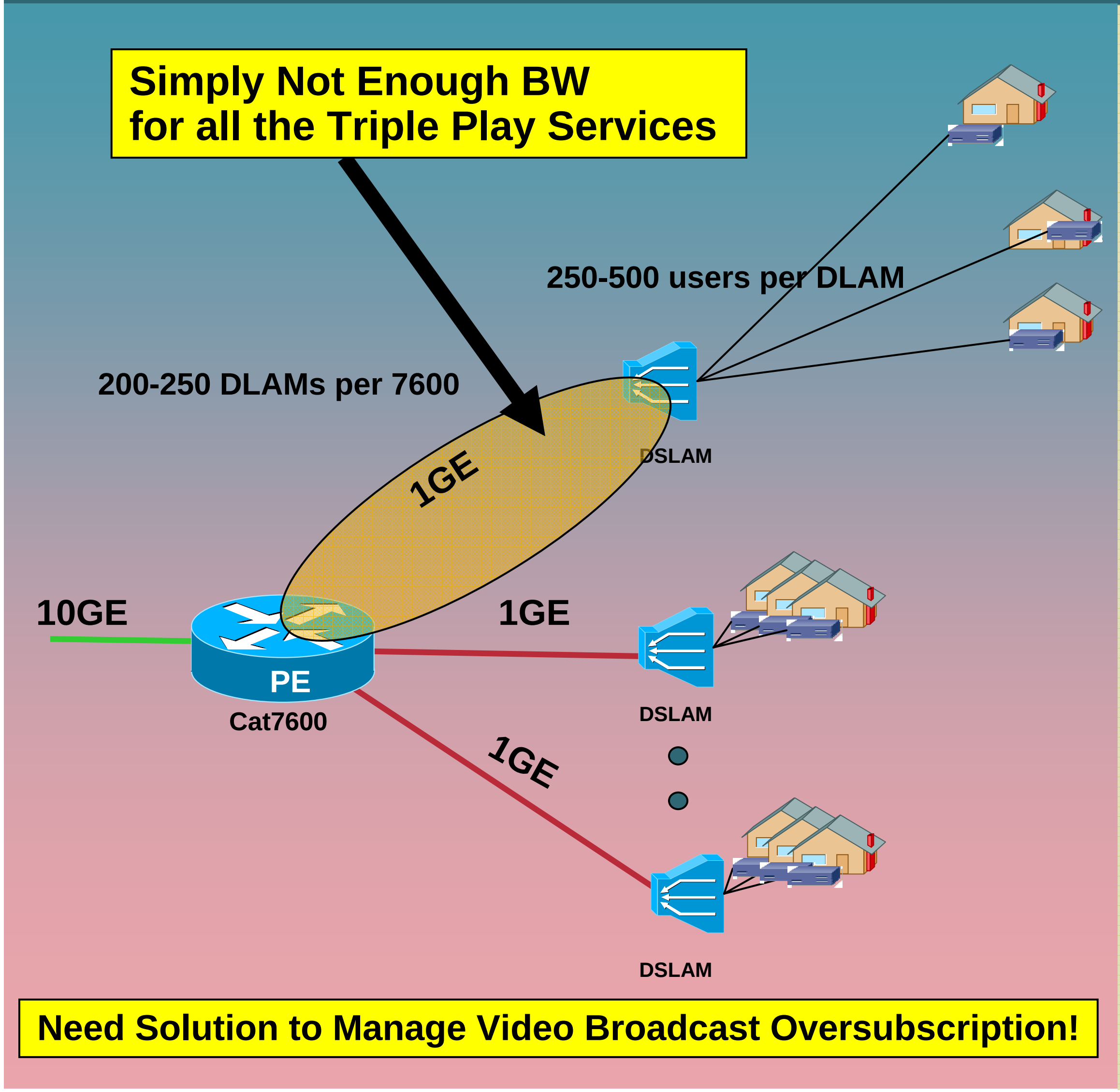
- GOP size of N seconds causes channel change latency >= N seconds
  - Can not start decoding before next I-frame
- Need/should-have channel change acceleration for GOP sizes > 0.5 sec ?
- Unclear
  - How much bandwidth is saved in different codecs by raising GOP size (same quality)
  - Eg: WM9/AV ~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

Multicast CAC

Local policy CAC in IOS

- Branch (“interface”) access control for flows
  - Ip igmp access-group <flows-acl>
  - Ip multicast boundary <flows-acl> [in | out]
- Branch (“interface”) limits for #flows
  - Ip igmp limit <n> [except <flows-acl>]
    - Output side only, IGMP only
  - Ip multicast limit <flows-acl> <n> [in | out]
  - Cost factor to multicast limit
    - Driven by customer requirements to provide fair share of bandwidth for flows from multiple content providers (See example)

Oversubscription on aggregation link to DSLAM



Problem statement

1. 250 – 500 end users need to be supported on a 1 Gbps DSLAM uplink.
2. Triple Play Services need to support 250 – 500 users
   - Voice : 2 IP phone connections per home
   - Video : 200 – 500 cable channels
   - Data : Internet
   - VOD : 10 % users using VOD
3. If 500 homes on a single DSLAM are all watching a different channel, the total BW required for video alone would be 500 x 4Mbps = 2Gbps!!!
4. This is NOT enough BW for a good user experience in the worst case scenario.
5. And don't forget, we still need BW to accomodate Voice, VoD & Data!
6. CONCLUSION…

Multicast Call Admission Control (CAC) phases

	Phases	Description	12.4T/12.2S Availability
1	Single per interface mroute state limit	Limits mroute state per interface. Introduced by Cisco® IOS IGMP State Limit feature.	*CSCdt86093* 12.2(15)T, 12.2(14)S
2	multiple per interface mroute state limits	Limits mroute state for different ACL-classified sets multicast traffic individually on an interface. Introduced by Cisco® IOS "Per Interface Mroute State Limit" feature.	*CSCdz51630* 12.3(14)T Future: Cisco-7600
3	Cost-factor addition to per interface mroute state limits	Allows for bandwidth based admission control on a per-interface bassis (ingress or egress).	*CSCej51837* Future: Cisco-7600

Phase 1 –Feature: (Single) Per Interface Mroute State Limit

Example CAC use:

1. Say the total number of SDTV channels offered by a Service Provider is 300.

2. Each SDTV channel is approximately 4Mbps.

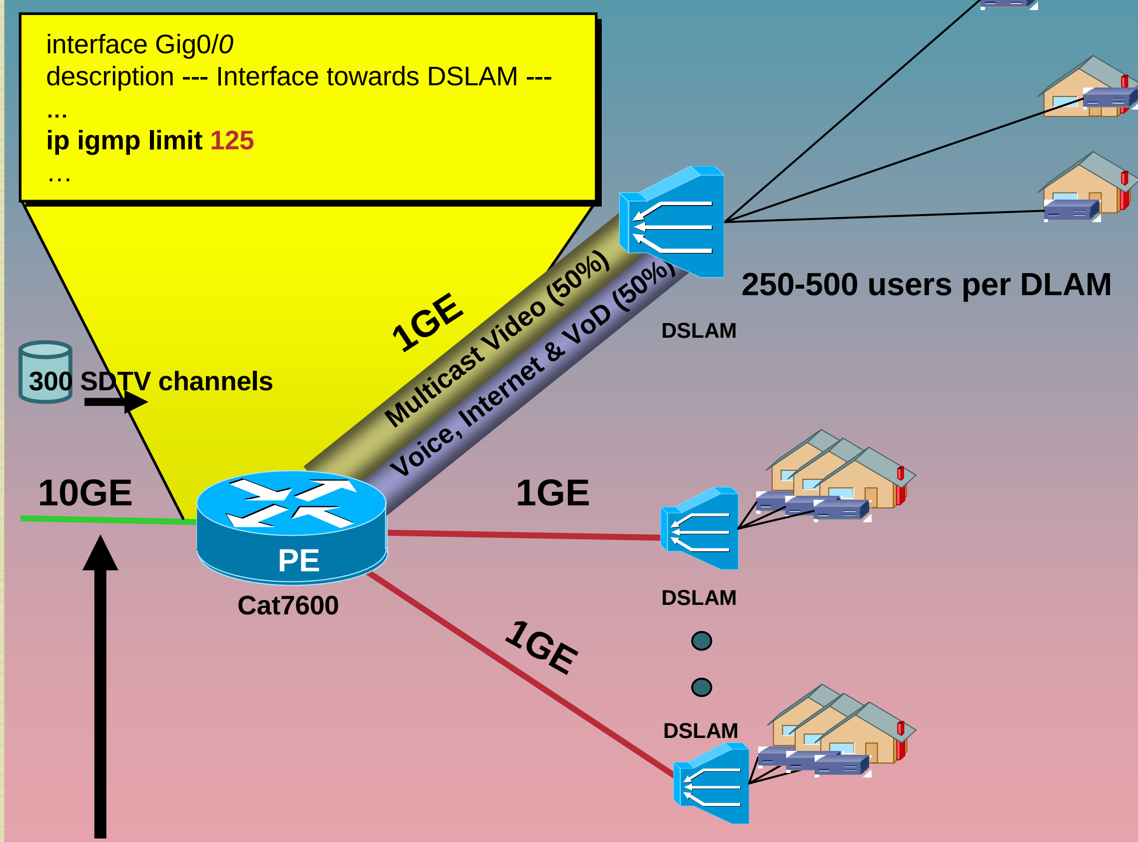
3. 50% of each outgoing 1Gbps link (500Mbps) needs to be provisioned for multicast video leaving the remaining 50% for Voice, Internet & VoD.

4. The required CAC needed per interface comes out to:

   500Mbps/4Mbps = 125 mroutes

Supports ASM and SSM IGMPv2/v3, but not PIM. Receiver side only.

Cisco® IOS IGMP State Limit Feature



300 channels x 4Mbps = 1.2Gbps > 1GE

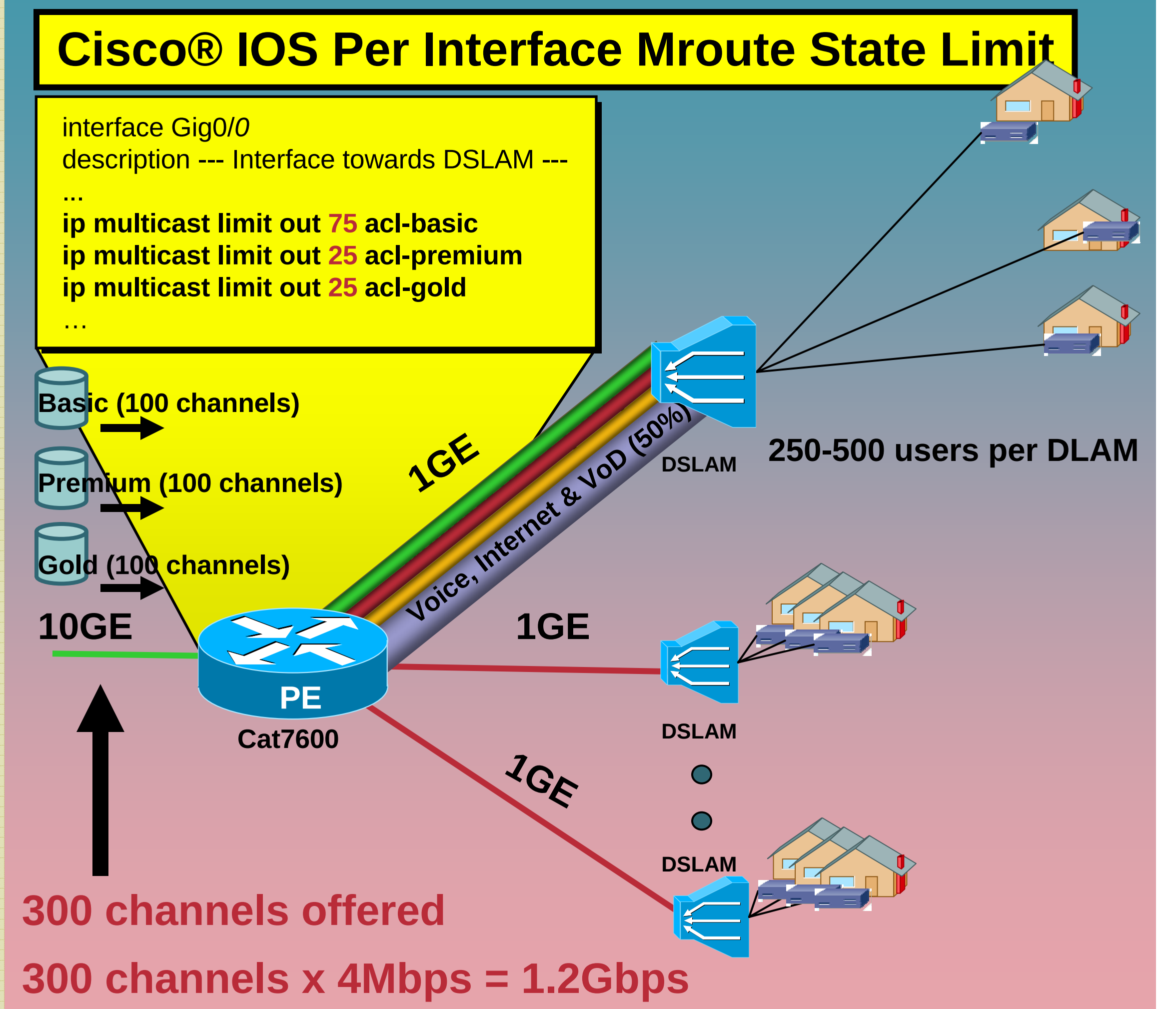
Phase 2 – Feature: (multiple) Per Interface Mroute State Limit

Generic interface multicast route limit feature with support for Ingress, egress, PIM/IGMP, ASM/SSM.

Example CAC use:

1. Say the total number of SDTV channels offered by a Service Provider is 300.
2. Each SDTV channel is ~ 4Mbps.
3. Service Provider will offer three TV bundles (Basic, Premium, Gold). Each bundle will have 100 channels.
4. 50% of 1Gbps link for mcast video Rest for Voice, Internet & VoD.
5. Within this provisioned 50%:
   60% will be for Basic (300Mbps)
   20% will be for Premium (100Mbps)
   20% will be for Gold (100Mbps)
6. The required CAC needed per interface comes out to:

   Basic mroute limit = 300/4 = 75
   Premium mroute limit = 100/4 = 25
   Gold mroute limit = 100/4 = 25



Phase 3 – Cost factor for per-interface Mroute State Limits

Example CAC use:

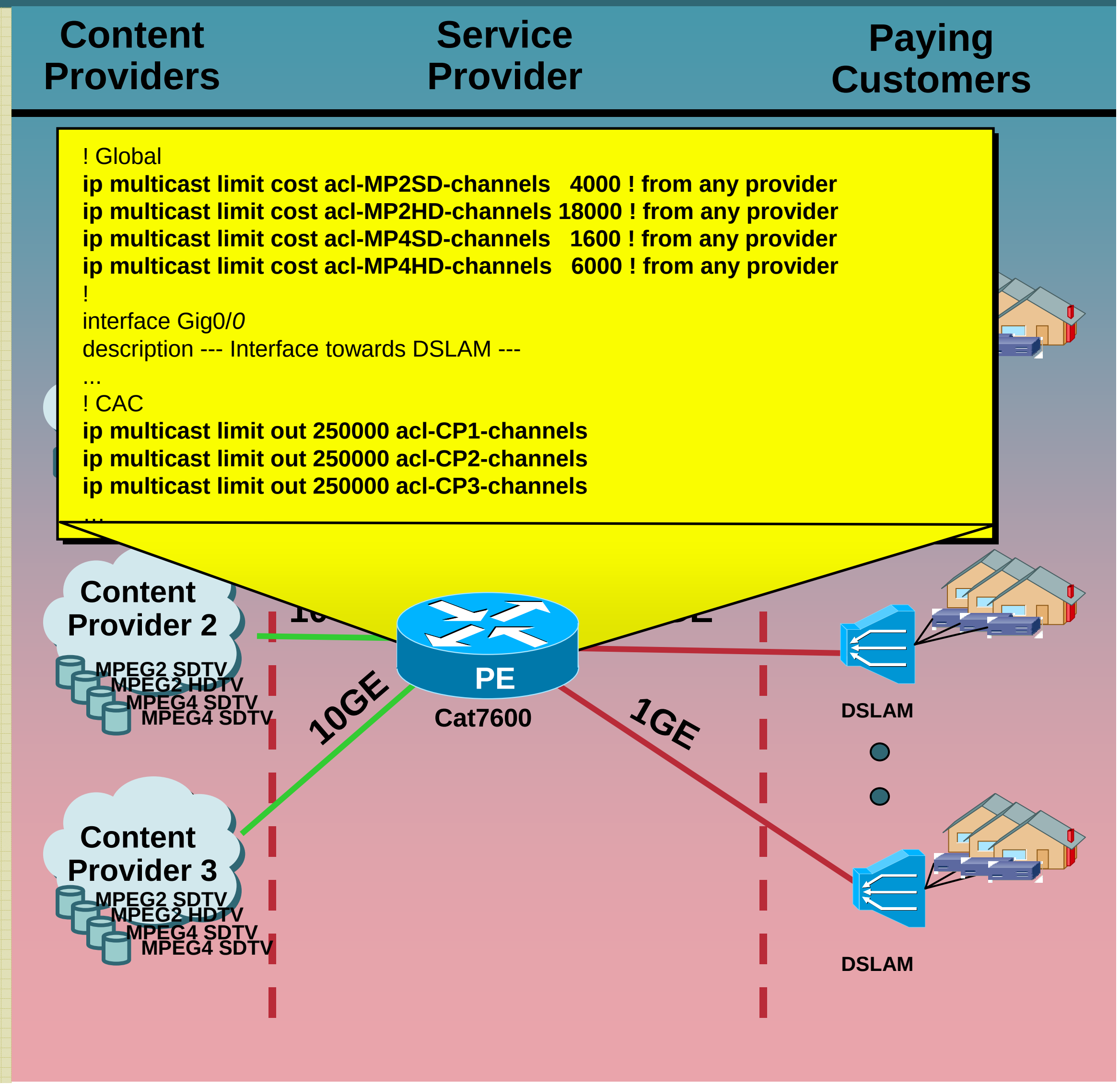
1. Consider the following. Three Content Providers (CPs) are providing multicast content.
2. Multiple CP will have TV programs w/ different BW:

   - MPEG2 SDTV: 4 Mbps
   - MPEG2 HDTV: 18 Mbps
   - MPEG4 SDTV: 1.6 Mbps
   - MPEG4 HDTV: 6 Mbps

3. Service Provider (SP) would like to provision fair sharing of bandwidth between these three content providers to its consumers across 1Gbps links.

4. 250Mbps for each CP, 250 Mbps for Voice/Internet/VoD.

5. Simple extension of multicast limits: global cost factor config.



Cisco Systems