Reliable IPTV Transport Network

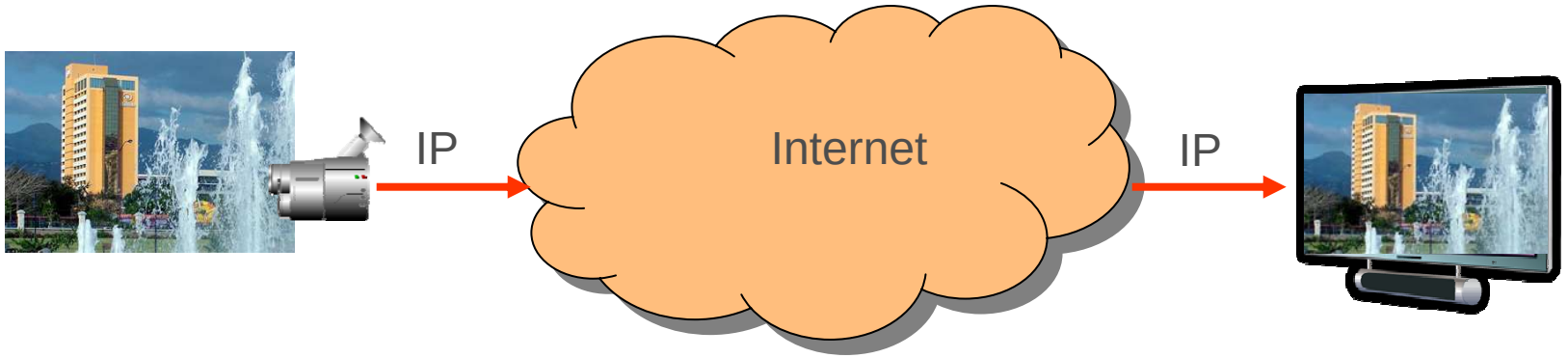
Dongmei Wang
AT&T labs-research
Florham Park, NJ

Outline

- Background on IPTV
- Motivations for IPTV
- Technical challenges
- How to design a reliable IPTV backbone network
  - Smart IGP weight setting
  - Make-before-break tree switching
- Future works

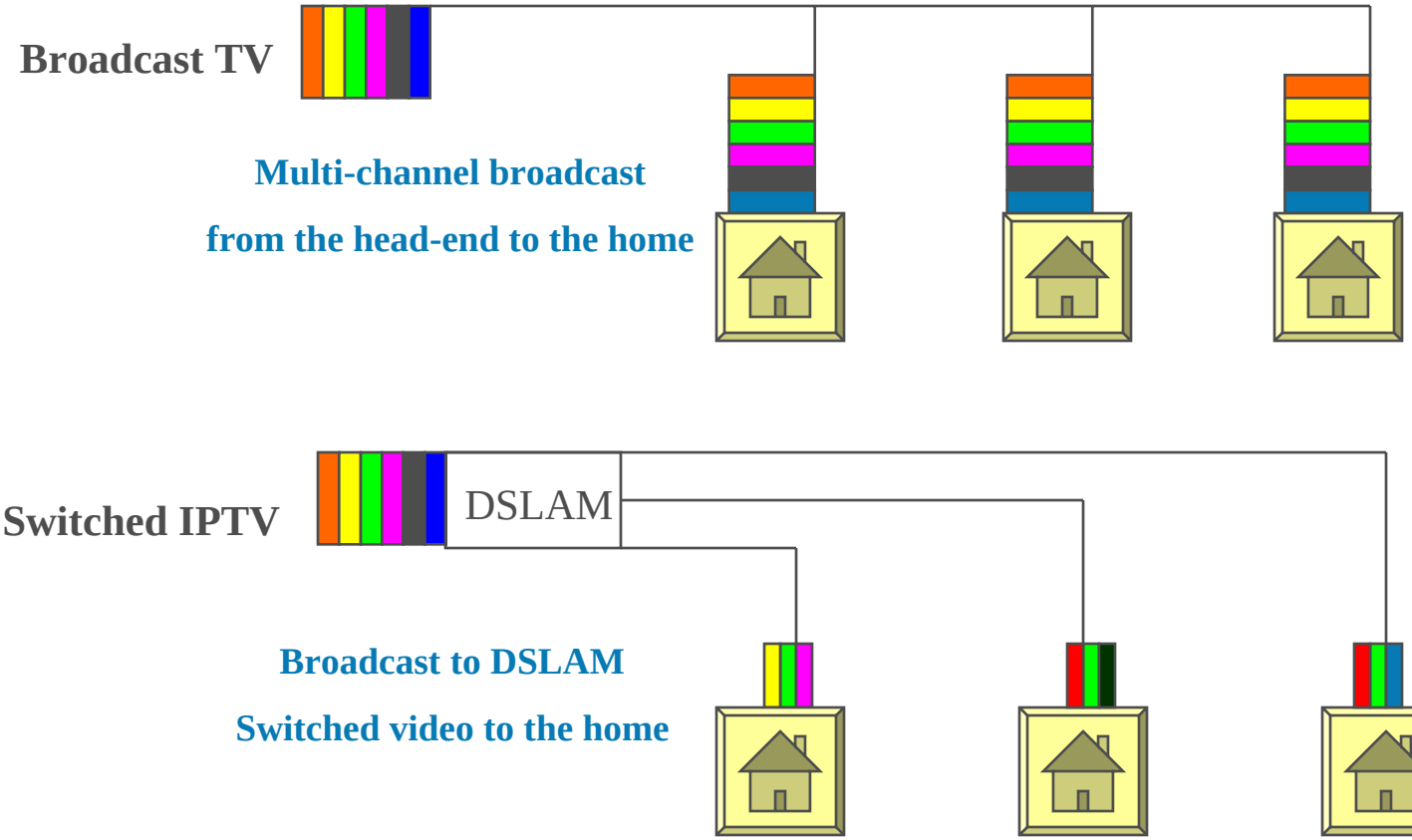
What is IPTV

IPTV: Internet Protocol Television



Further defined:
A technology that Telcos are deploying to compete with cable TV
Using internet protocol and IP multicast protocol to deliver IP packets of digital video.
IPTV packets are delivered over private networks.

IPTV vs. Cable TV



Why IPTV

- **Business**
  - Critical component to triple play bundle
  - Attracts new subscribers
  - Grow Average revenue per customer (ARPU)
- **Customer benefits**
  - Improved price
  - Enhanced services
    - Caller ID displayed on TV
    - Unified messaging
    - Picture-in-Picture
    - Search functionality

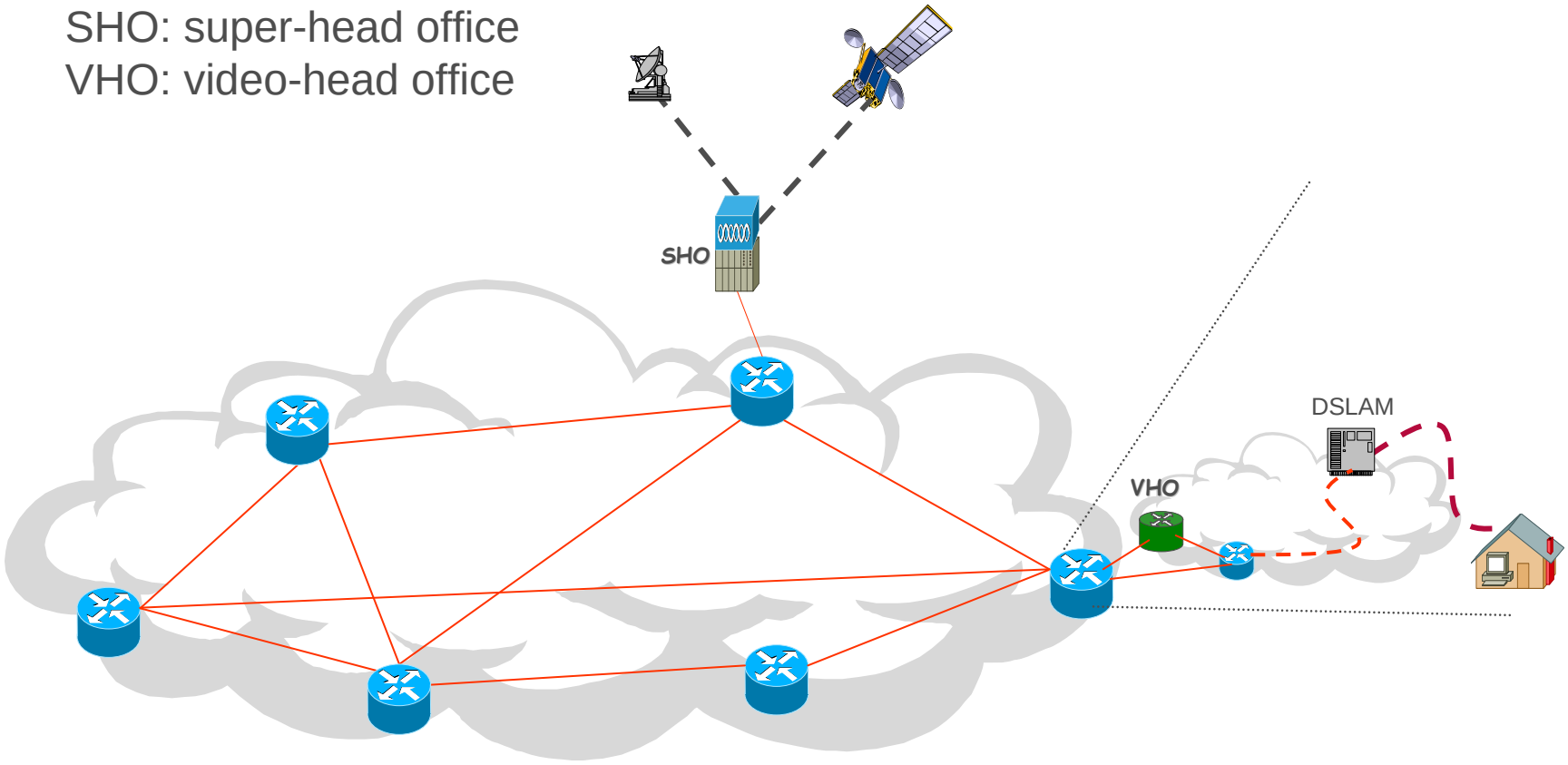
IPTV Basic Requirements

- Relatively stable high bandwidth
  - 1~4 mbps per video stream, 6~8 mbps HDTV
  - About 300~500 channels ➔ 1.5 Gbps
- High availability
  - 99.99% ~ 99.999% ->5~50 minutes downtime per year
- Tight jitter (<10ms) and loss constraints (<0.1%)

FAST RESTORATION (<50ms)?

IPTV Backbone Architecture

SHO: super-head office
VHO: video-head office



How to handle failures

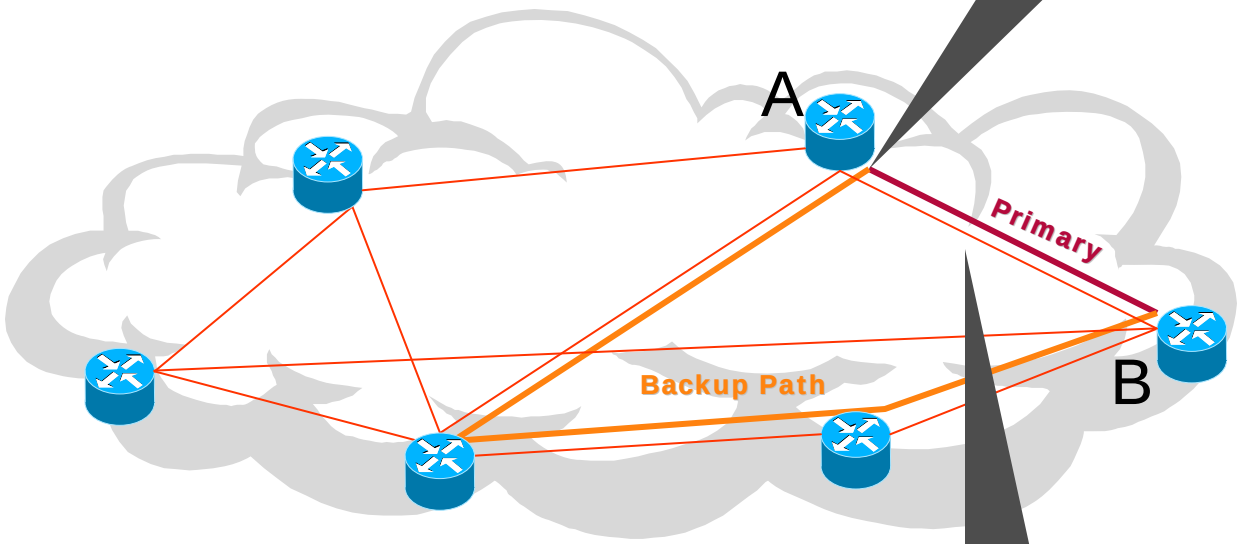
- Protocols
  - OSPF routing protocol
  - PIM-SSM: source specific multicast
- Protocol re-convergence upon failure
  - 5~30 seconds for OSPF convergence
  - 200 ms for PIM-SSM
  - Does not satisfy IPTV restoration requirement (<50ms) !!

Link-Based FRR

Normal traffic forwards on primary link

Primary link fails, MPLS FRR to backup

No OSPF/PIM-SSM convergences

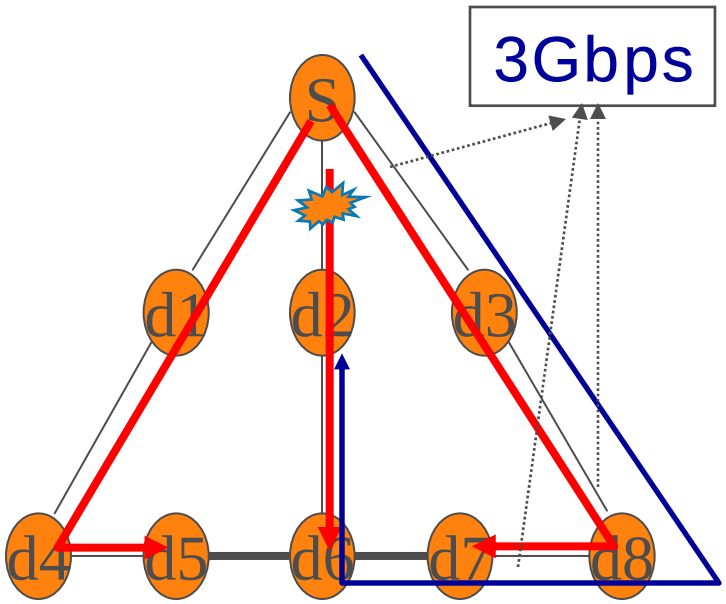


Virtual link between AB with virtual interfaces

Virtual link consists both primary/backup path

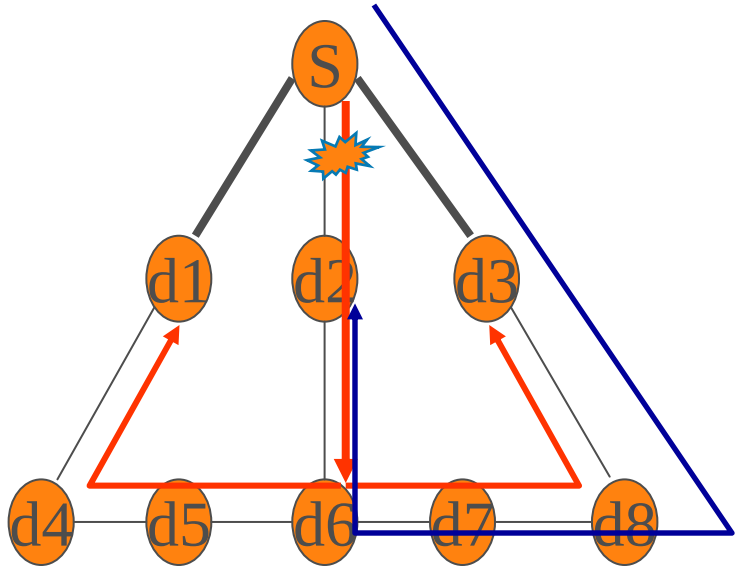
OSPF LSA on top of virtual link

Why Smart Link Weight



(a) Bad

Link d5-d6 and link d6-d7 have weight 2, other links have weight 1



(b) Good

Link S-d1 and link S-d3 have weight 2, other links have weight 1

overlap: a packet travels more than once on the same link along the same direction

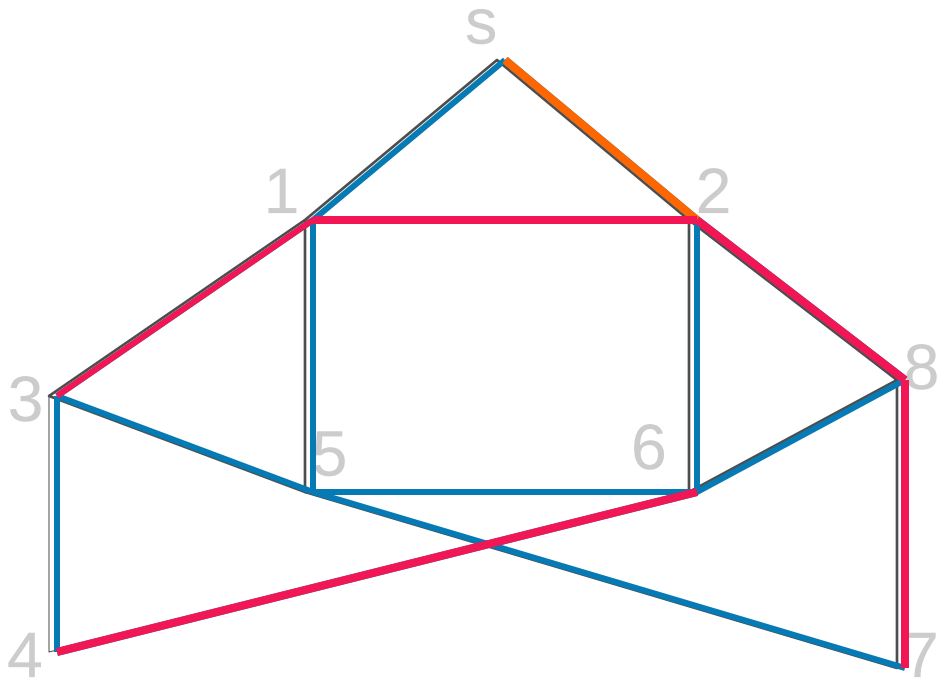
Smart Link Weights

- Assumption:
  - Given a 2-connected network topology
  - A source node
- Objective:
  - Separate links: high cost and low cost
  - Low cost links form a multicast tree
  - Each link on the multicast tree has a backup path
  - No overlap between backup traffic and multicast traffic

Algorithm

1. Find a set of links to form a ring, including source
2. Assign weights for the ring links:
   1. Set one link adjacent to source as high cost
   2. Set other links on the ring with low cost
   3. All links with weights form graph *G*
3. Find a set of links to form a line with two ends of the line staying on *G* from remaining links
4. Assign weights for the links on the new line
   1. Set one end link as high cost
   2. Set other links on the line as low cost
   3. Add the new line with weights to *G*
5. Repeating steps 3-4 until all links are in *G*

Example



Steps:

1: select ring S-1-5-6-2

2: select chain 1-3-5

3: select chain 3-4-6

4: select chain 2-8-6

5: select chain 5-7-8

6: select chain 1-2

Low link weight

High link weight

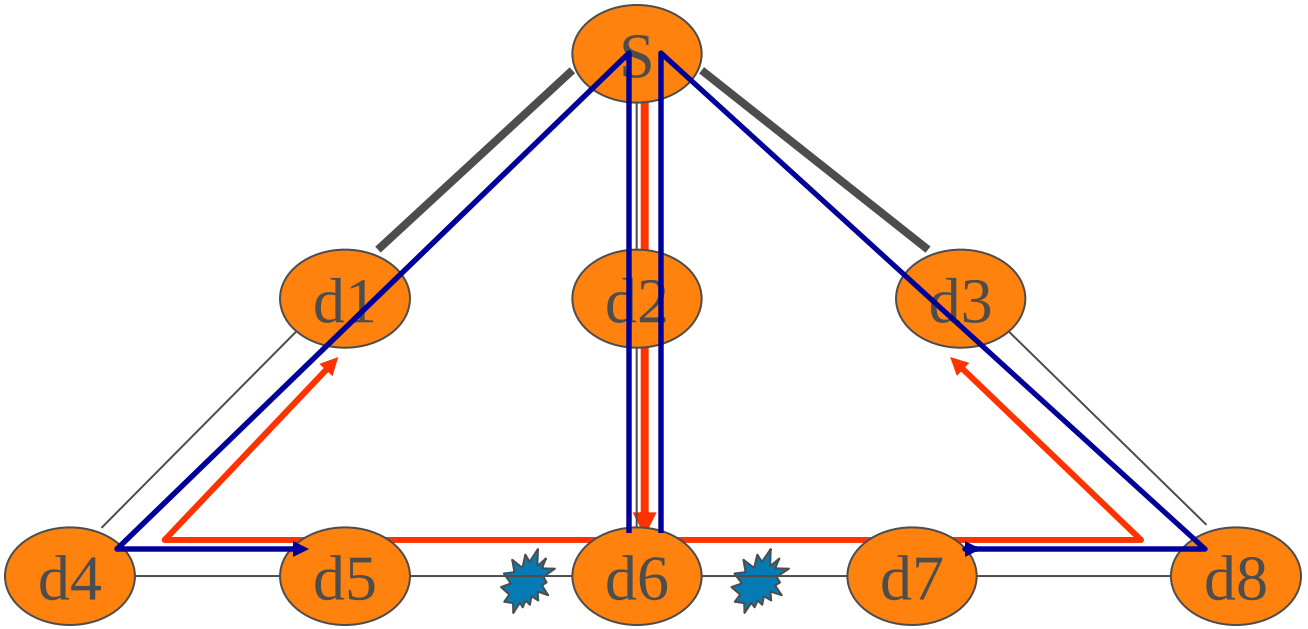
Correctness of Algorithm

- Induction proof
  - Base: ring topology
  - Assumption for k new lines are added
  - Proof after (k+1)th new line is added
    - First we need to prove the existence of such a new line. Then we pick any two nodes on graph G, we prove that there is one path from one node to another without overlapping the multicast tree traffic. Then we prove the correctness of our algorithm (see Infocom 2007)

Summary on FRR with smart weight setting

- Achieved
  - Fast Switch to the backup path (<50ms) upon link failure
  - No routing re-convergence as long as either the link or its backup path is available
  - Guaranteed fast restoration (<50ms) for single link failure
  - Upon router failure, routing protocol re-converges and PIM rebuilds the multicast tree.
- Problem:
  - No guarantee for dual/multiple link failures

Double Failure Congestion

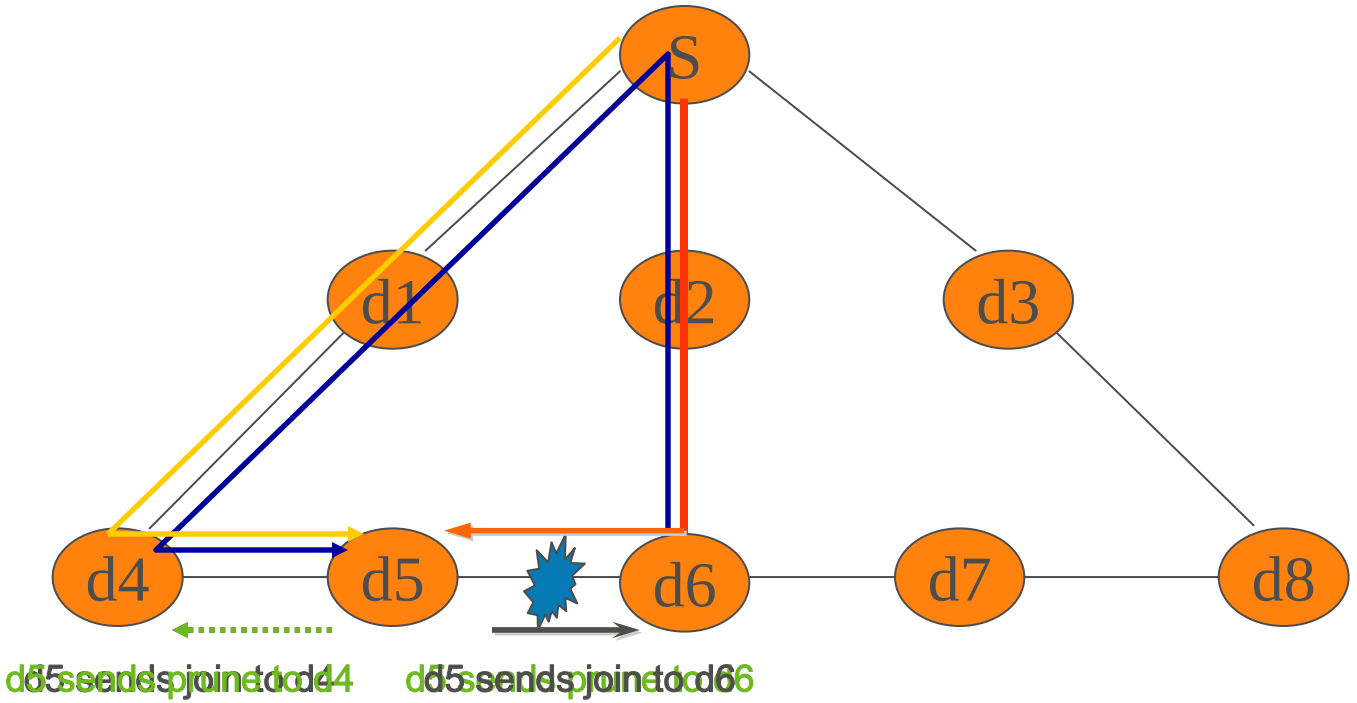


- Link d6-d5 has backup path d6-d2-S-d1-d4-d5
- Link d6-d7 has backup path d6-d2-S-d3-d8-d7
- If d6-d5 and d6-d7 fail, there are traffic overlapping on links d6-d2 and d2-S, which could cause congestion and may last a few more hours

Backup path for transit period only

- Proposed approach
  - Fast reroute traffic to backup path upon link/interface failure
  - Cost-out the backup path to trigger routing re-convergence.
  - After routing re-converges, PIM rebuilds multicast tree. The backup path is only used during protocol convergence period.
- Problem:
  - Potential double hits during single failure

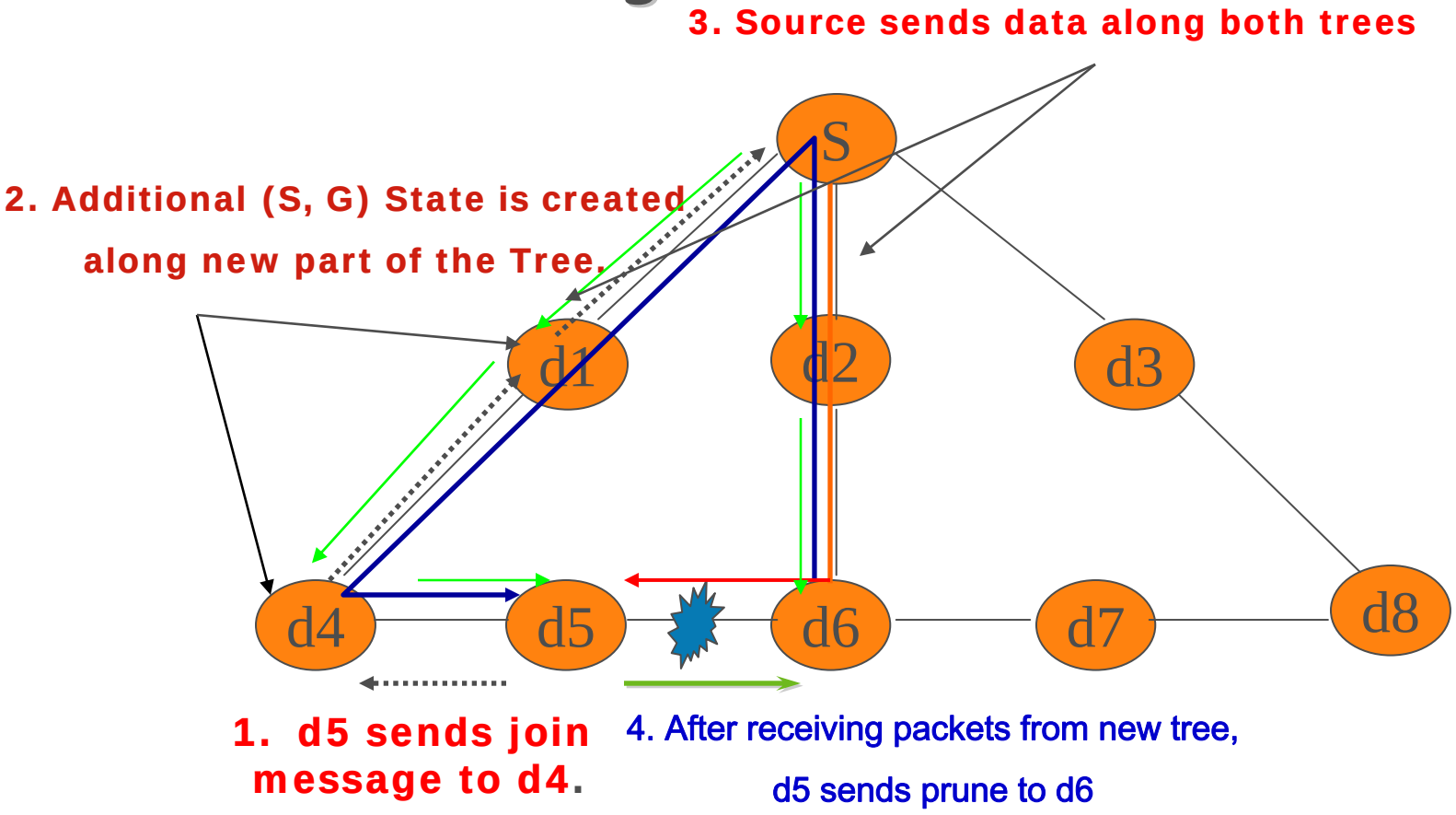
Potential double hits per single failure



First hit: d5 stops receiving packets from d6 even though routing in S has not converged

Second hit: after failure repair, d5 switches back to the original tree too quick.

Hitless tree switching



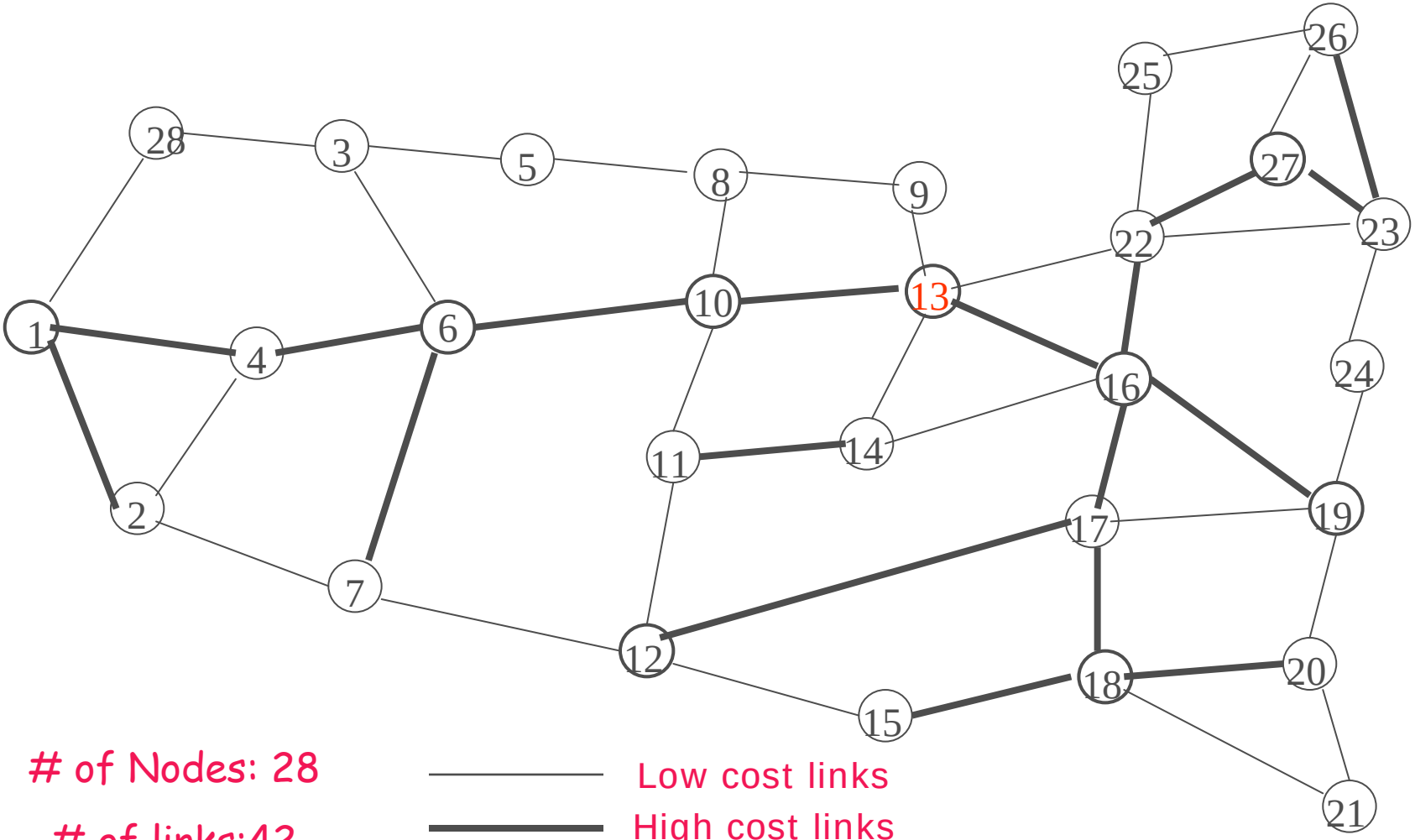
Problem Solved?

- Restoration time <50ms for single link failure
- Restoration time is bounded by protocol convergence time (10s) for multiple link failures
- Restoration time is bounded by protocl convergence time (10s) for router failure
- Is this sufficient to guarantee the required QoS??

Performance Analysis

- Assumptions:
  - Network unicast routing protocol, for example OSPF
    - Covergence time: 10s
  - Network multicast routing protocol: PIM-SSM
    - Converegnce time 200 ms
  - Link based Fast ReRoute (FRR) (50ms)
    - No service interruption
  - Hitless tree switching (50ms)
    - No service interruption
  - Optical transport layer only provides pure connectivity to IP layer.
  - All restoration process is carrying out via IP layer

Using A Hypothetical US Backbone



# of Nodes: 28

# of links:42

Low cost links

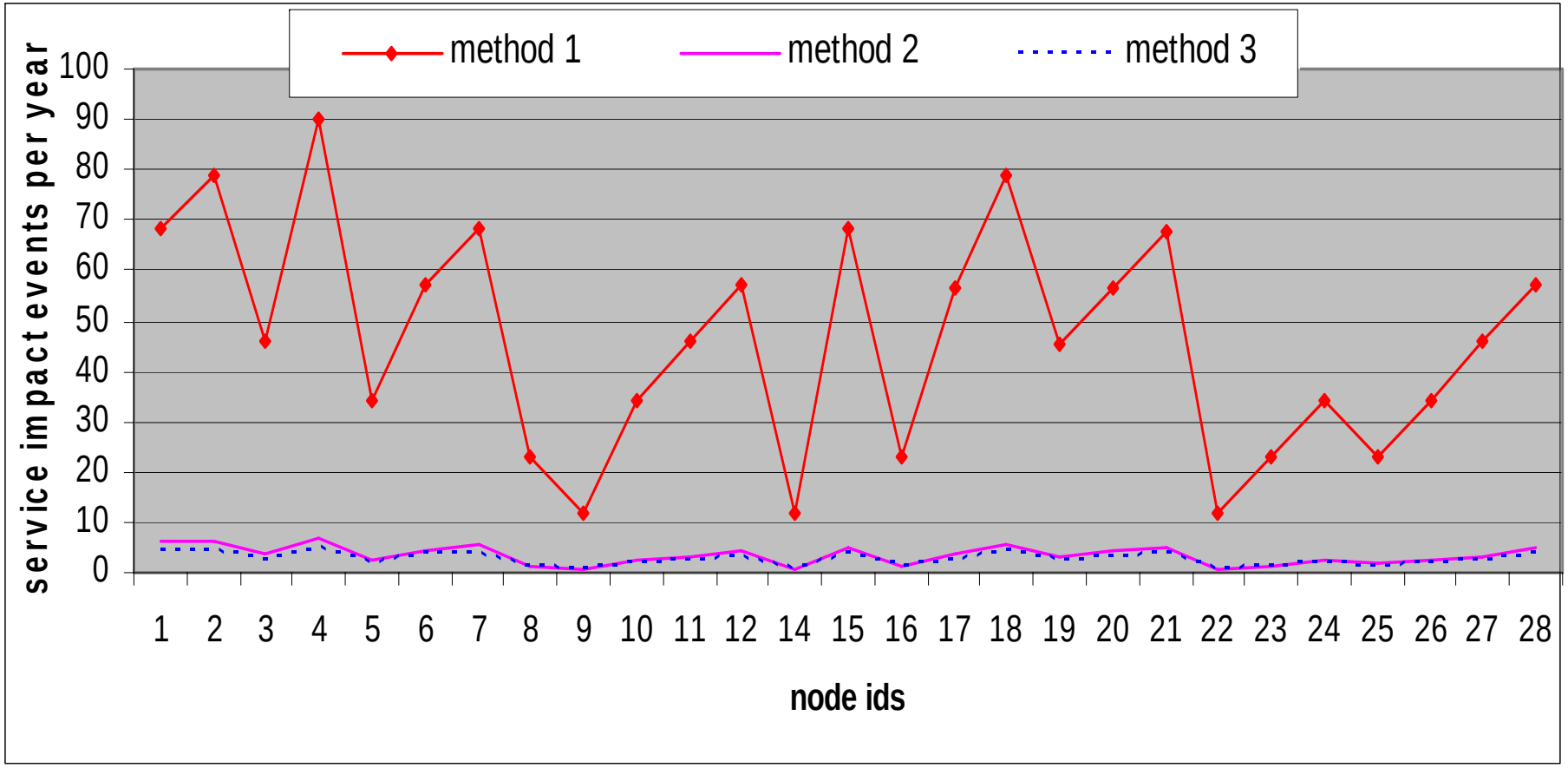
High cost links

at&t

Performance analysis (continue)

- Compare three methods
  - Method1: IGP re-convergence only
  - Method2: Link based fast reroute
  - Method3: fast reroute plus hitless multicast re-convergence
- Metrics
  - Service impact events per year
    - Events last more than 50ms
  - Total down time per year
- Event generation
  - Network performance analyzer
  - Using probability model to generate the events including single failure and multiple failures

Service Impact Events per Year



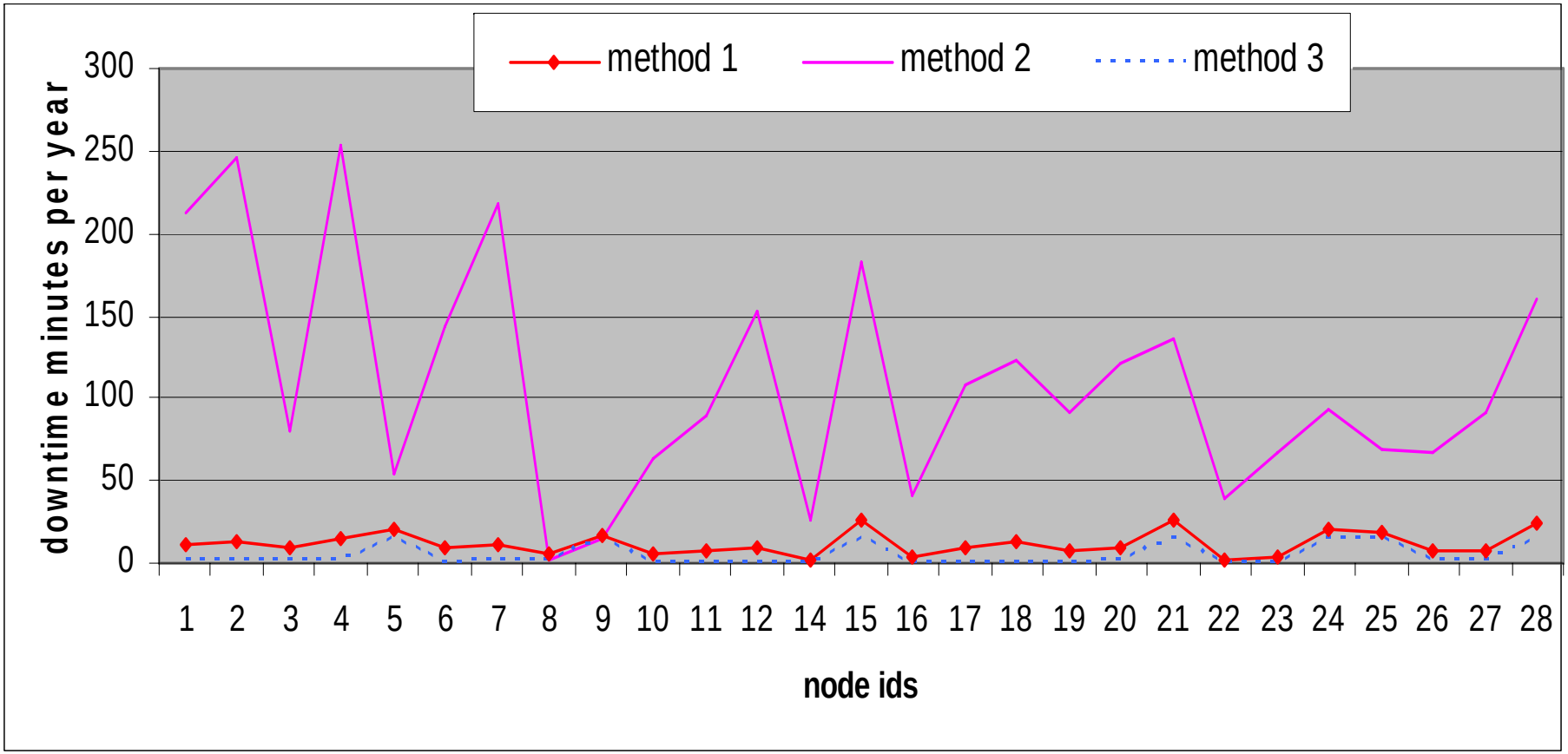
Method1: IGP re-convergence only

Method2: Link based fast reroute

Method3: fast reroute plus hitless multicast re-convergence

at&t

Down-time Minutes per Year



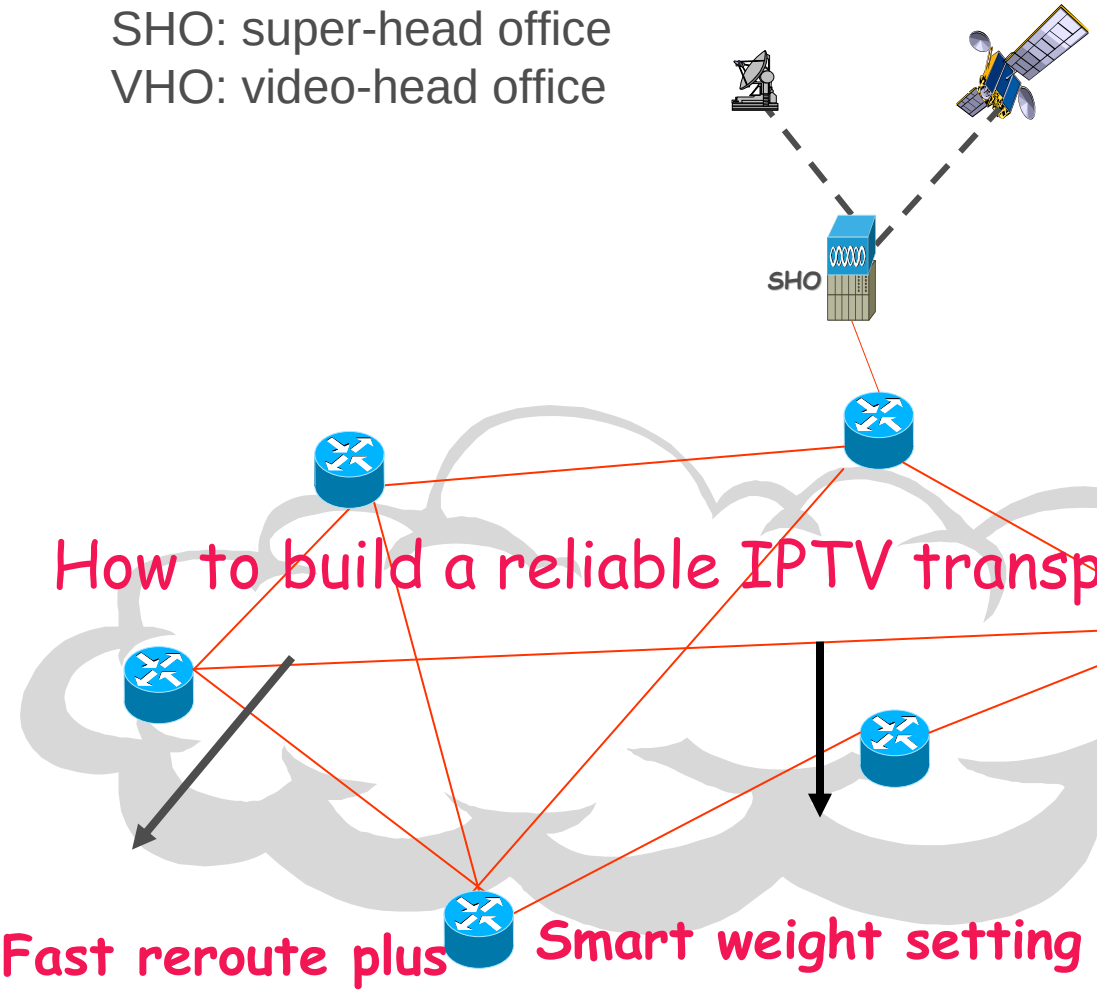
Method1: IGP re-convergence only

Method2: Link based fast reroute

Method3: fast reroute plus hitless multicast re-convergence

Conclusion

SHO: super-head office
VHO: video-head office



How to build a reliable IPTV transport network?

Fast reroute plus hitless tree switching

Smart weight setting algorithm

Performance analysis: Minimize service impact

Questions?