IP Multicast Fault Recovery in PIM over OSPF

Xin Wang, Chienming Yu, Henning Schulzrinne, Paul Stirpe, and Wei Wu
Dept. of Electrical Engineering
Columbia University
New York, New York
xwang@ctr.columbia.edu

Outline

- Motivation: many IP multicast applications require high availability
- We study failure recovery in a complete architecture: IGMP + OSPF (unicast) + PIM-SM (multicast); consider single link and router faults
- develop sequence of events and interactions under different failures
- provide some analytical results under different failures (not shown here)
- simulate failures in OPNET; measure control overheads and recovery times
- study failure recovery and implementation issues on small test-bed

Network Failure and Recovery Scenarios

- Failure Recovery in WAN
  - OSPF:
    - Detect link failure within “carrier delay” or *RouterDeadInterval*.
    - Send updated router-SLA to neighbors
    - Neighbors recalculate their shortest paths through Dijkstra’s algorithm.
  - PIM:
    - Learn failure through notify message or polling of unicast routing table
    - Determine new Reverse Path Forwarding (RPF) router
    - Send Join/Graft on the new RPF interface, re-build multicast tree.

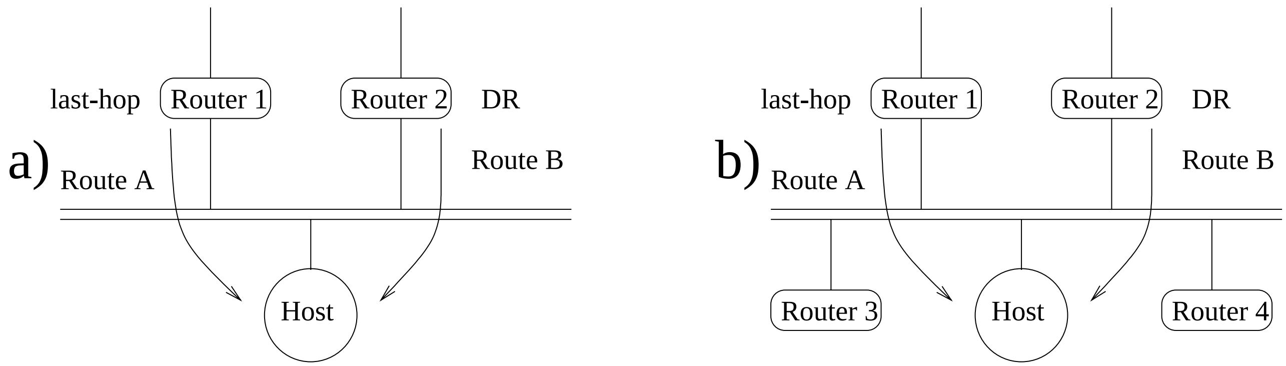


Figure 1: LAN failure scenario, DR and last-hop router are different routers

- Failure Recovery in LAN – PIM-SM
  - DR and Last-hop router are separate:
    - ∗ Upstream link of DR fails: wait for IGMP report to reactivate the pruned interface.
    - ∗ Link between DR and LAN fails: new DR election and multicast entry re-build.
    - ∗ Upstream link of last-hop router fails: send join right away to new RPF.
    - ∗ Link between last-hop router and LAN fails: wait for new IGMP report and recover through DR; Or recover by downstream router join.

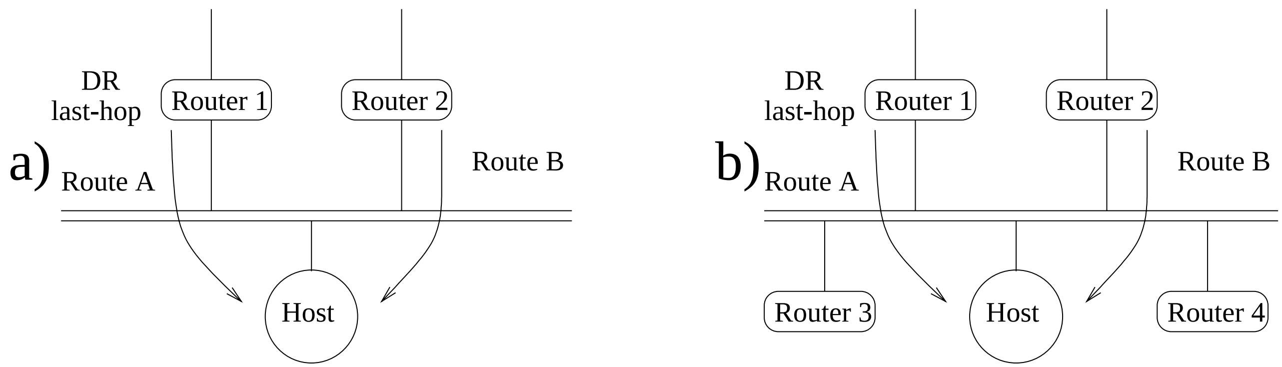


Figure 2: LAN failure scenario, DR and last-hop router are the same router

- Protocol Interaction in LAN
  - DR and last-hop router are same:
    - Upstream link of DR fails: DR will recover the multicast channel immediately.
    - Link between DR and LAN fails: new DR election and multicast entry re-build; Or recover by downstream router join.

- Protocol Interaction in LAN – PIM-DM
  - Upstream link of router-Other fails: graft immediately if with active entry; otherwise, wait for new IGMP report.
  - Upstream link of last-hop router fails: send graft immediately to new RPF.
  - Link between last-hop router: wait for new IGMP report and recover through DR; Or recover by downstream router join.

- Simulation Model
  - Network topology: 36 nodes random topology, default redundancy factor = 4, percentage of receivers set to 80% for the (single) group.
  - OSPF parameters: the *RouterDeadInterval* = 3x*HelloInterval*
  - PIM parameters: unicast table polling interval = 0.2 s.
  - Application layer parameters: data rate set to a low value, end to end recovery time is measured.

OSPF Control Load versus OSPF Hello Interval

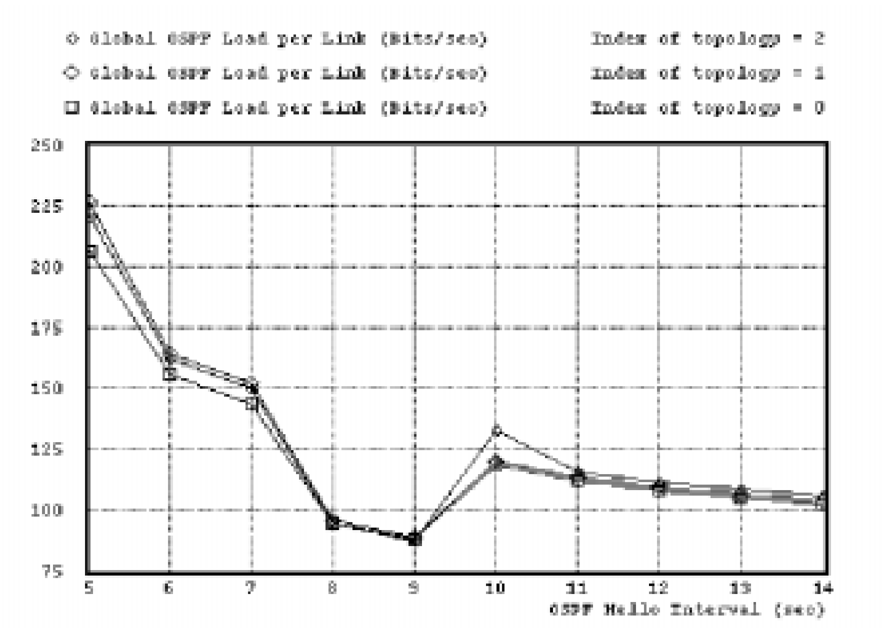


Figure 3: OSPF load change with the variation of Hello interval

PIM DM Control Load versus Network Topology

a) b)

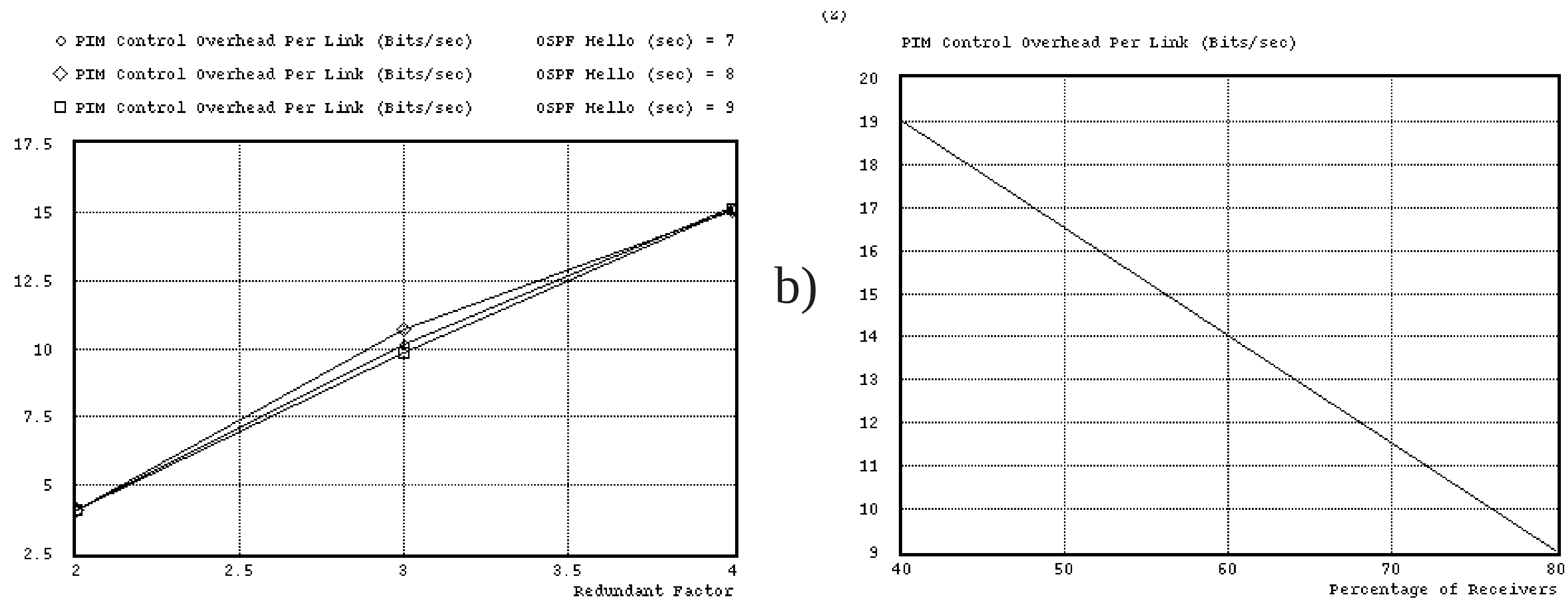


Figure 4: PIM DM load change with the variation of network redundancy factor a) and receiver percentage b)

Single Multicast Channel Recovery Time

a)

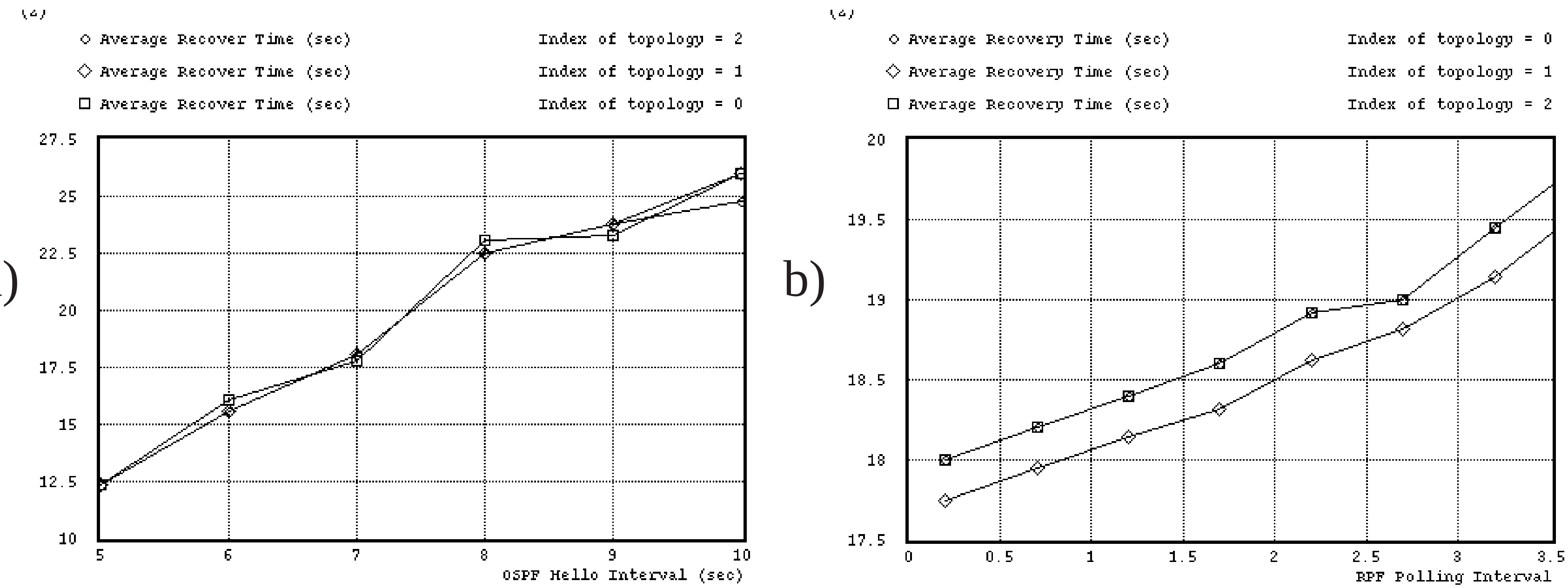


Figure 5: a) Variation of multicast channel recovery time with the OSPF Hello interval (PIM polling interval set to 0.2 s) b) Variation of multicast channel recovery time with the PIM polling interval

Network Load Change during Failure Recovery

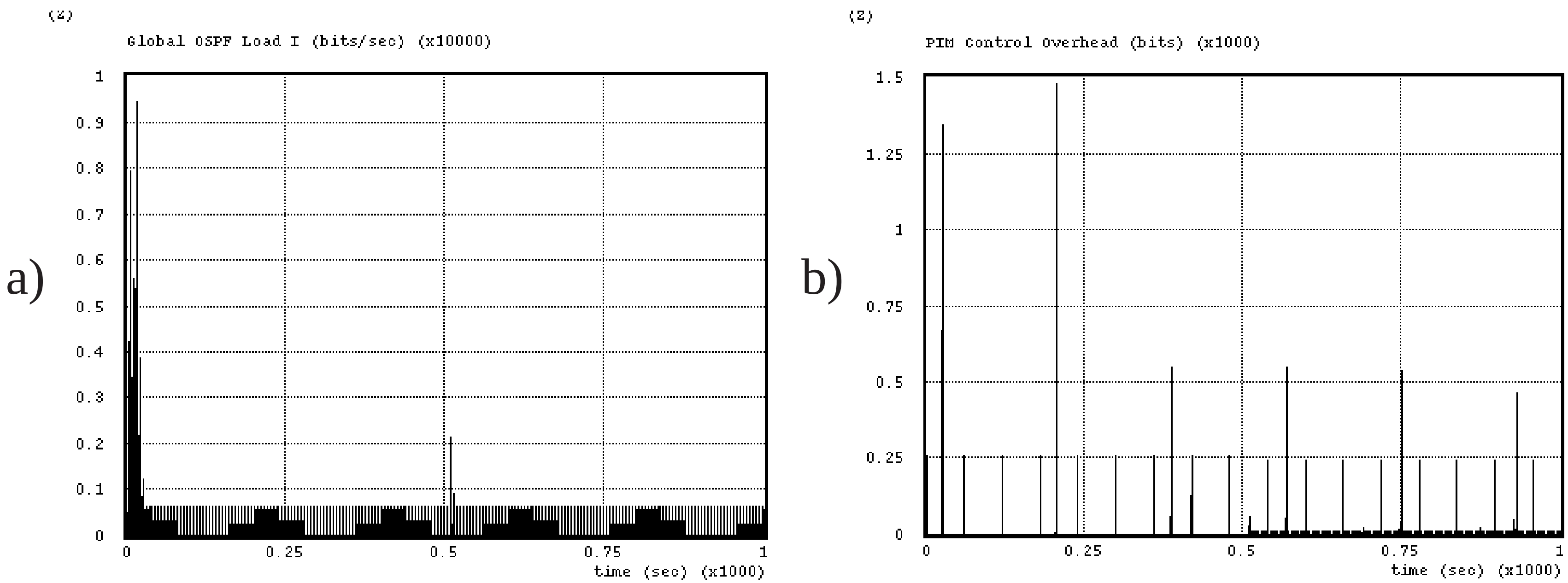


Figure 6: OSPF load change (a) and PIM DM load change (b) during failure recovery, beginning at t=500 seconds

Testbed Setup and Parameters

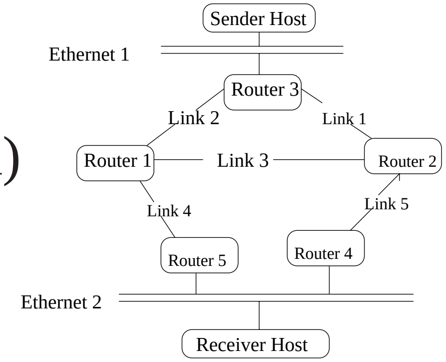
a) 

Figure 7: Testbed topology

- IGMP parameters:
  - *Query Interval*=125 s; *Query Response Interval*=10 s; *Other Querier Present Interval*=255 s.
- OSPF parameters:
  - *HelloInterval*=1 s; *RouterDeadInterval*=3s.
- PIM parameters:
  - *Hello Intercal*=2 s; therefore DR failure detection=6 s; Join/Prune interval=60 s; unicast table polling interval=5 s.

Example test-bed data in a LDAP directory

Failure Event	OSPF Recovery	PIM Recovery	Join Latency	Total Recovery	Router Perspective	Initial Route before failure
link 1	2.11853	2.87677	0.05926	5.05456	R2	R3 ⟶ R2 ⟶ R4
link 5	2.02733	3.38755	0.05251	5.46739	R4	R3 ⟶ R2 ⟶ R4
Router 2	2.06035	4.60794	0.06246	6.73075	R4	R3 ⟶ R2 ⟶ R4
Router 4 (FWD&DR)	3.012	4.176	0.006	7.194	R5	R3 ⟶ R2 ⟶ R4
Router 5 (FWD) SM	2.470	64.027	0.128	66.625	R4	R3 ⟶ R1 ⟶ R5
Router 5 (FWD) DM	2.470	95.025	0.128	97.623	R4	R3 ⟶ R1 ⟶ R5

Table 1: Fail-over time (in seconds) with OSPF totally stubby area

Failure Event	OSPF Recovery	PIM Recovery	Join Latency	Total Recovery	Router Perspective	Initial Route before failure
link 1 (step1)	2.1431	4.32362	0.01918	6.4859	R2	R3 ⟶ R2 ⟶ R4
(step2)	0	3.28387	0.01574	3.29961	R4	R3 ⟶ R2 ⟶ R4
link 5	2.65603	3.40131	0.08288	6.14022	R4	R3 ⟶ R2 ⟶ R4
Router 2	2.12218	4.16531	0.04512	6.33261	R4	R3 ⟶ R2 ⟶ R4
Router 4 (FWD&DR)	2.563	4.001	0.007	6.971	R5	R3 ⟶ R2 ⟶ R4
Router 5 (FWD) SM	2.638	60.024	0.023	62.685	R4	R3 ⟶ R1 ⟶ R5
Router 5 (FWD) DM	2.638	92.012	0.023	94.673	R4	R3 ⟶ R1 ⟶ R5

Table 2: Fail-over time (in seconds) with OSPF non-stubby area

Conclusion

- General observations
  - Channel recovery time: dominated by unicast table re-construction time.
  - Protocol control loads: PIM DM control load increases proportionally with the redundancy factor and decreases inversely with the percentage of receivers; OSPF load increases proportionally as OSPF *Hello* interval decreases.
  - Neither PIM nor OSPF has high control traffic during failure recovery.
- PIM Enhancement for Fault Recovery
  - Fast recovery from DR failure: reduce *Hello-Holdtime* to detecting neighbor failure faster; Backup DR; IGMP group information caching in all LAN routers.

- Fast recovery from last-hop router failure: DR could record the last-hop router address, would not need to wait for an IGMP report to reactivate its *oif* to the LAN; Backup router can be used in PIM DM acting as DR for rapid detection of the last-hop router failure.
- Reduce extra delay due to polling by using interrupts