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.

Xin Wang, IRT, Columbia University June 18, 2000
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 $\text{RouterDeadInterval} = 3 \times \text{HelloInterval}$
  
  – 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
Figure 4: PIM DM load change with the variation of network redundancy factor a) and receiver percentage b)
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

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 Interval*=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

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

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

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

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

Xin Wang, IRT, Columbia University June 18, 2000
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