Cross-Layer Failure Restoration Techniques for a Robust IPTV Service

Murat Yuksel, K. K. Ramakrishnan, and Robert D. Doverspike
yuksem@cse.unr.edu, kkrama@research.att.com, rdd@research.att.com

***Abstract*— Broadcast TV distribution over an IP network requires stringent QoS constraints, such as low latency and loss. The main challenge to achieving these QoS objectives is how to design the network to respond to network failures. Streaming content in IPTV is typically delivered to the distribution points on the IP backbone using IP multicast, and in the case being considered, with Protocol Independent Multicast Source Specific Mode (PIM-SSM). A proven failure restoration technique at the IP layer is link-restoration using MPLS or layer-2 Fast Reroute (FRR). Link-based FRR creates a pseudo-wire or tunnel in parallel to the IP adjacencies (links) along the forwarding path used by the PIM tree. For each such tunnel both a primary and backup path are defined. The backup path is Layer-1-disjoint from the physical link and when the link fails, the pseudo-wire can be rapidly restored. Thus, single link failures are transparent to the Interior Gateway Protocol (IGP). Although one may choose the back-up path's IGP link weights to avoid traffic overlap during any single link failure, multiple failures may still cause traffic overlap with FRR. We present a cross-layer restoration approach that combines both FRR-based restoration for single link failure and "hitless" (i.e., without loss) PIM tree reconfiguration algorithms to prevent traffic overlap when multiple failures occur.**

***Index Terms*— multicast, fast-reroute, cross-layer design, IPTV, PIM-SSM reconfiguration**

I. Introduction

DISTRIBUTION of real-time multimedia over an IP backbone has been gaining momentum with content and service providers [1,2,3,4]. However, unlike traditional cable-based broadcast infrastructures that provide "broadcast" analog-based video (e.g., TV), using an IP backbone for real-time broadcast video distribution imposes stringent requirements for protection and restoration after a failure. Distribution of real-time 'linear' (also called broadcast) TV requires that the delay experienced by a user viewing the TV content be limited to less than a few seconds. This limits the size of the playout buffer at the receiving IPTV set-top box. Moreover, the mechanisms for recovering from packet loss have a limited capability to recover from burst packet losses. The combination of player loss-concealment algorithms, recovery through retransmissions and the use of packet-level redundancy mechanisms such as FEC are designed to recover from burst losses that are no more than a few tens of milliseconds. The tight QoS constraints of low latency and loss need to be met even under failures. The network challenge to meeting these constraints while providing high network availability, in turn, requires a methodology for rapid restoration [3].

Use of IP-based Protocol Independent Multicast (PIM) [5,6,7] to distribute the content from the source to the various distribution points on the IP backbone allows the infrastructure to be cost-competitive with the more mature cable broadcast infrastructure. However, PIM-SSM depends on a "join" and "prune" process to rebuild the multi-cast tree after a network failure. This process, when combined with the IGP reconfiguration process, takes too much time to restore the packet flow to the receivers after the failure. Thus, the packet loss concealment and recovery mechanisms alone are not effective in recovering from such a long burst of packet losses. A common approach to provide the needed high availability under this IP framework is link based Fast Reroute (FRR) [8,9,10]. However, a drawback of FRR is that it re-routes traffic on a local basis (instead of end-to-end). Therefore, it has a higher tendency to suffer from traffic overlap (and thereby congestion) during network failures, especially with multiple failures. Large-scale, real-time video streams are characterized by high bandwidth, often 1 Gb/s or more in aggregate across all the channels, and because of its stringent QoS constraints, traffic congestion often has the same affect on the customer's perception of service as an unprotected link failure. Therefore, to achieve the low-cost logical multicast tree design without significant over-provisioning of the distribution network (and all links having roughly the same bandwidth), the network has to be designed to avoid congestion from traffic overlap due to restoration.

To understand the motivation for our approach, we will describe in more detail the interaction of multiple failures and FRR. Consider a router node pair (k,j) in an IP network used for IPTV distribution where the routers in the network are part of a multicast tree using PIM-SSM [25]. Assuming the adjacency between these routers is restorable via FRR, this node pair actually consists of four different links that are visible to the Interior Gateway Protocol (IGP) topology: the Pseudowire from k to j (P_{kj}), the Pseudowire from j to k (P_{jk}), the physical layer "PHY" link from k to j (L_{kj}), and the "PHY" link from j to k (L_{jk}). Assume the primary path for each Pseudowire is its corresponding PHY link and the backup path (i.e., FRR path) is over other PHY links disjoint from the primary path. The IGP link costs (which we shall generically

This work has been supported in part by AT&T, Inc. and the U.S. National Science Foundation under award 0721600.

Murat Yuksel is with the University of Nevada – Reno, Reno, NV 89557.

K. K. Ramakrishnan is with AT&T Labs Research, Florham Park, NJ 07932 USA.

Robert D. Doverspike is with AT&T Labs Research, Middletown, NJ 07748 USA.

refer to as "weights") on the Pseudowire links are lower than the PHY links. This causes the IGP shortest path algorithm to route over the Pseudowire links rather than the PHY links in a non-failure state. Then, when either of the PHY links, L_k or L_{jk}, fails, both links are taken out of service and the two Pseudowires are switched from their primary paths to their backup paths, usually with a target switching time of 50ms or less. This switching time is sufficiently small that the higher layer packet loss concealment and recovery protocols can recover from this failure with little or no perceived video quality disruption. In addition, the back up Pseudowire will be up and running well before the IGP timers expire. Therefore, when the IGP Link State Advertisements (LSAs) are broadcast, although they show that the PHY links are down, the Pseudowire link states are unchanged. Because the Pseudowire links have lower weights than the PHY links comprising their primary paths, there is no change to the shortest path to the source and hence no change to the multicast tree.

Typically, the FRR backup path is kept active until the PHY links are repaired. Once the PHY links are back in service, the Pseudowires are switched back to their primary paths rapidly (again with a target switching time of 50ms or less). This methodology works well to achieve high network availability when only non-simultaneous link failures occur (by "link" we mean the pair of unidirectional PHY links between a router pair). However, if a second failure occurs during the outage interval of the first failure (which may range from a few minutes to several hours), then because IGP is unaware of active FRR backup paths, it is possible that traffic overlap may occur. Here, traffic overlap means that the packets of the same IP flows (e.g., packetized video for individual 'channels' for IPTV) travel over the same unidirectional link two or more times. Since distances can be thousands of miles and long paths over underlying Wavelength Division Multiplexing (WDM) transport systems may not be highly diverse, multiple IP-layer failures in a long distance backbone network are not rare events. This phenomenon of traffic overlap can be avoided if we modify the approach of leaving the backup path active until the PHY link is repaired. The key is to only *temporarily* route the Pseudowire over the FRR backup path and then have IGP converge to a new shortest path tree and the resulting multicast tree which does not use the affected Pseudowire. Thus, if a second failure occurs, IGP is fully aware of the routing for the tree and can avoid the failed links during IGP reconvergence. However, a requirement of this new approach is to engineer a *hitless* (i.e., without loss) multicast methodology, wherein we switch to the new tree after a FRR reroute without incurring packet loss. Otherwise, every time a failure occurs and the backup FRR path is used, we will experience a hit to converge to the new tree, thus making the new approach potentially worse than the current approach.

The approach we describe above was suggested in "Architecture 3" of [25]. The primary contributions of the present paper are to propose in more detail the algorithms to perform the switchover to the new multicast tree and then evaluate the performance via simulation. The algorithms we propose are implemented entirely locally at the nodes. For those multicast nodes hearing about the link failure as a unicast IGP routing change, we introduce two bits of state that are associated with our hitless switchover protocol. The routers (the one downstream of the failed link being the first) recompute the shortest path to the source after the failure and issue a join on the new upstream interface, while still not issuing a prune on the old interface. Once multicast data begins to flow on the new incoming interface, the router then issues a prune on the old interface. More details of the algorithms are described in Section IV.

The rest of the paper is organized as follows: In Section II, we discuss related work. In Section III we discuss adaptations to the IGP reconvergence process needed to make our approach work. Section IV describes in detail the hitless PIM-SSM process we propose. Section V analyzes its performance through discrete-event simulation, followed by a summary.

II. Related Work

Most of research in IPTV focuses on architecture design [4,12], protocol design and selection, multimedia stream coding/decoding techniques [18], as well as potential new applications of multicast [19]. There has been extensive IP multicast research and experiments, although much less on multimedia IP network design. IETF has standardized multiple IP multicast protocols, including PIM-DM [5], PIM-SM [6], and PIM-SSM [7], and has made recommendations for reliable IP multicast [20]. To improve IP multicast performance, many models and techniques have been proposed including the overlay model [21] and MPLS P2MP [22], resource reservation and admission control to avoid congestion [23], and priority queueing or fair queueing to guarantee quality of service. These solutions have to be adapted to be useable in a carrier's multimedia distribution service.

Network restoration has also been an active research area for many decades. As video and real-time services migrate to IP based environments, IP network restoration has also increased in importance. But, little work has been done on performance evaluation of different restoration schemes in multimedia backbone design. Although simulations show that large IP networks are able to achieve sub-second convergence for the routing protocol by tuning the routing protocol's timers [13], service providers have not adapted such schemes due to concerns regarding network stability. Other schemes like failure insensitive routing [24] apply to unicast routing and are not really suitable for our environment which involves routing of multicast flows.

IETF RFC-2362 [6] outlines an approach to a "make-before-break" mechanism for PIM-SM. Each multicast group G may define two trees: a shared tree (*,G) and source shortest path tree (S,G). RFC 2362 [6] proposes three steps to switch from a shared tree to the shortest path tree using the "make-before-break" method as follows:

- The last-hop router issues the join (S, G) toward the source (S) to set up the shortest path tree;
- The data packets start to flow on the shortest path tree

once the tree setup is complete;

- The last hop router issues a prune message to remove the old path after receiving the data packet from the shortest path tree.

The method we propose here is similar, but goes into more careful initiation and progress of steps so that we achieve both hitless multicast tree reconfiguration and avoidance of traffic overlap for IP video services when the IP network uses FRR.

III. The Architecture: Hitless Multicast Reconfiguration Due to Routing Changes

Our architecture, shown in Figure 1, for providing a hitless multicast reconfiguration is motivated by the need to transition from using the backup path offered by link-based FRR soon after a link failure, so as to minimize the exposure of the network to traffic overlap from any subsequent failure prior to the repair of the initial link failure. The intent is to transition to a new multicast tree (at least portions of the tree upstream from the router affected by the failure) as soon as feasible, but to do so in a manner that does not result in a significant traffic "hit", both in terms of packet loss or delivery of duplicate packets. As described earlier, IGP routes the traffic over a Pseudowire link, which is routed over a primary LSP that is set up on the PHY link. Upon a PHY link failure, the Pseudowire link is routed over the backup LSP. For the remainder of this paper, we describe our multicast tree reconfiguration methodology in general terms, and we refer to the Pseudowire as a "link" observed by IGP.

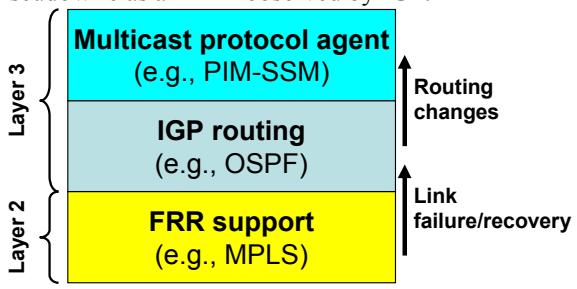

Fig. 1. Cross-layer architecture for failure restoration.

Normally, if we were to adopt an IP-only approach for dealing with link failures and recovery, the underlying changes in the topology are first propagated by IGP to the rest of the nodes, using standard IGP (e.g., OSPF) techniques. This involves first the link failure detection (through a lower layer indication such as SONET alarms or the lack of receipt of HELLOs within a *RouterDeadInterval*), then the propagation via flooding of LSAs and subsequent topology convergence, which may be a function of both computation of the SPF as well as the *spfDelay and spfHoldTime* timers [26]. All these timers are likely to be of the order of several seconds. Subsequent to that, the PIM-SSM tree has to be reconfigured, which again may take several seconds to tens of seconds (e.g., a new join is issued after 30 seconds, as part of the standard process of refreshing the soft-state for multicast). These timers are usually set to be very large so as to meet the needs of a large scale IP network and control the message and computational overhead, and cannot accommodate the real-time performance requirements of multimedia streaming applications.

In [25], the "Architecture 3" option described uses the FRR mechanism, but effectively limits its use only to the period prior to full IGP re-convergence subsequent to a link failure because the link failure is exposed to the IGP. After this initial rerouting period, the failed link (Pseudowire in this particular IPTV case) is avoided (by setting the link weight to a very high value) so that the new IGP topology and the resulting multicast tree do not use the backup path (LSP). After the IGP converges, the portion of the PIM-SSM tree from the point of the failure is rebuilt by issuing a join (after PIM rejoin timers fire) along the new shortest path to the source. This achieves the benefit of rapid restoration from single link failures, yet allows the multicast tree to dynamically adapt to multiple failures and thus avoid traffic overlap. However, this method is still vulnerable to traffic overlaps when multiple failures occur, and uses network bandwidth longer than necessary since the PIM rejoin timers are typically set to several tens of seconds.

The PIM-SSM tree is typically reconfigured after an IGP reconvergence event in a distributed manner, where each router independently computes the shortest path to the source for each multicast group. The portion of the tree to the router downstream of the failure is formally reconfigured by each router issuing a *join* message (and a *prune* on the old path) to attach to its new parent node. During this process, packet loss can occur, which we refer to here as a performance *hit*. There is already a potential *hit* when the link failure occurs. The failure has to be detected and the local link level FRR has to switch over to the backup path. The failure detection mechanisms and the switchover to the backup path are typically designed to try and keep this potential *hit* to a minimum – often within 50 milliseconds. However, if we simply reconfigure the PIM tree after a short FRR transient period subsequent to a network link failure, a second significant hit could occur at the time of reconfiguration of the PIM multicast tree. Furthermore, the design has to be careful to avoid another hit occurring after the link failure is repaired (e.g., to return to the original PIM tree). If this is the case, then the advantages of reconfiguring the tree to avoid congestion may be somewhat obviated by the performance hits from single link failures and their subsequent repair. As noted previously, the mean time between single link failures in a long distance network is short enough so that the occurrence of multiple failures should not be considered a rare event.

Therefore, a key component of the method briefly outlined in [25] is *make-before-break*, i.e., the requirement to switch traffic from the old multicast tree to the new multicast tree with minimal loss of traffic. In this paper, we propose more algorithmic detail on the tree switchover, to carefully weave together the join and prune into the new tree and out of the old tree as well as the IGP reconvergence process to achieve *hitless* multicast. The primary elements in our method are that routers coordinate the calculation of the shortest path with local decisions on choosing new parent nodes (and the corresponding branches) in the new multicast tree and then sending a join to build a path from its new parent. However,

the prune message to remove the branch to the previous parent is not sent until the router receives PIM-SSM data packets from its new parent for the corresponding (S,G) group. In this way, each router is guaranteed to continuously receive (without loss) data packets during the switchover period.

IV. PIM Reconfiguration with Hitless Switchover

"Hitless switchover" to a new multicast tree has been described as the "make-before-break" mechanism in the literature, e.g., RFC 2362 [6]. Though these make-before-break mechanisms attempt to minimize the amount of packet loss, they were designed for switching to a new tree over a stable network without failures. The key shortcoming of the approach (outlined in Section II) is that it neglects link failures and switches to the new tree only after the routing has completely stabilized across the entire topology. Further, it does not guarantee a coherent switchover if the new upstream node on the new tree is also one of the current downstream nodes of the router that is involved in the switchover to the new tree. We introduce a hitless switchover protocol which makes the switchover more proactive and reconfigures the multicast tree immediately after a routing change. We expose the link failures to the IGP routing layer here, while taking advantage of FRR support during the reconvergence process. Key characteristics of our method are:

- It guarantees reception of all data packets even after a failure, except the packets in transit which are lost on the failed link and packets buffered for that interface at the router adjacent to the link;
- It can be initiated when a failure is detected locally by the router and does not have to wait until routing has converged network-wide;
- It works even if the new upstream router is one of the current downstream routers.

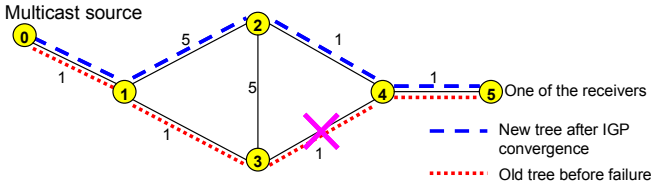

Fig. 2. A sample link failure causing multicast tree to change.

A. Make-Before-Break Protocol with Hitless Switchover

To make a hitless switchover to the new tree after failure, we introduce a small amount of additional state information at every router but do not change any of the protocol messages used by standard PIM-SSM. The routers use this additional state information when they are in a "transient state" which lasts until the IGP converges subsequent to the first failure. We believe that this switchover method is applicable to the interaction of any variant of IP multicast and link-state routing protocols, even though we primarily focus on PIM-SSM and OSPF in our description.

Following is a detailed description of our method and the sequence of events that take place while restoring the failure:

Step 1: *Failure occurs.* The failure of a link on the existing multicast tree will be first detected by the routers adjacent to the failed link. For example in Figure 2, routers 3 and 4 will detect the failure of link 3↔4 locally. This failure detection will trigger SPF calculation in the IGP at routers 3 and 4.

Step 2: *Multicast protocol agent is notified.* With standard multicast operation, the failure of the link 3↔4 is not made visible to the multicast protocol. With our approach, the multicast protocol agent (see Figure 1) is notified of the IGP routing change. More specifically, after the router performs the SPF calculation, router 4 will inform its multicast agent that the upstream routing interface to the source has changed.

Step 3: *Multicast protocol agent moves to transient state for those (S,G) groups whose upstream interface has changed.* The multicast agent, when notified of the routing change, checks if the new routing interface for the upstream router to the source of the subscribed source-specific multicast session (S,G) is different than the previous one. If so, it will mark the (S,G) group as being in the transient state, waiting to receive multicast data for that group on the interface to the new upstream router to the source (let us call this state "waiting for multicast tree branch to form"). Note that the multicast agent will perform this check and mark the state (if needed) for each (S,G).

Step 4: *Multicast protocol agent tries sending a join on the new upstream interface.* For each (S,G) the multicast agent has marked as being in the transient state, it checks to send a join message on the new upstream interface. There are two cases:

Case I: If the new upstream interface is not among the list of outgoing downstream interfaces of that (S,G), then the agent immediately sends the join upstream on the new upstream interface and then goes to Step 5. That is, in Figure 2, router 4 will send a join to router 2. That join will be then sent further up to router 1. This will make router 1 forward the (S,G) multicast traffic to both 2 and 3.

Case II: If the new upstream interface is among the list of current outgoing downstream interfaces of that (S,G), it will have to wait until it receives a prune message on that interface first. In that case, the agent marks this interface as "waiting-to-send-join" and proceeds to Step 4.a. This delay in sending the join assures that all the upstream routers on the new tree have completed their SPF calculation when they are notified of the link failure through link state advertisements. To understand the reason for the delay, consider the scenario in Figure 3 where there is a second receiver at router 2. Consider the case where router 4 detects the failure of link 3↔4 and immediately sends the join message to its new upstream router 2 (without checking if it is currently a downstream interface or not). It is certainly possible that router 2 has not yet performed its SPF calculation, Router 2 considers that router 4 is still its upstream router for the group (S,G). Then, router 2 will either reject that join or forward it back to 4 causing the join message to be lost and not serving its purpose to reconfigure that portion of the multicast tree.

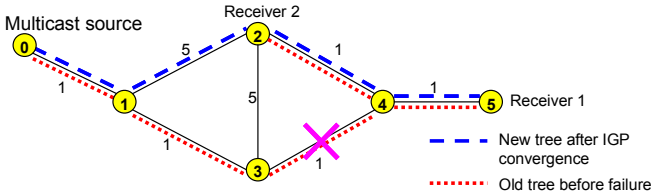

Fig. 3. A sample link failure causing a previously downstream router to be the upstream one on the new multicast tree.

Step 4.a: *Multicast protocol agent receives a prune on the new upstream interface on which it could not send a join.* After all the upstream routers on the new tree have safely passed through their transient state (if they indeed had to transition into that state), the new upstream router will also transition out of its transient state and send a prune message (see Step 5) on its old upstream interface, which happens to be the new upstream interface of the router waiting in the transient state. Then, this "waiting-to-send-join" router will send a join message upon receiving the prune message. In Figure 3, router 2 will first send the prune message to router 4 when it starts receiving data packets from 1. Receipt of the prune message triggers router 4 to send the join message to router 2 (new upstream interface) that it was waiting to send.

Step 5: *Multicast protocol agent sends a prune on the* ***old*** *upstream interface after receiving data packets on the* ***new*** *upstream interface.* After sending the join on the new upstream interface, the multicast agent at the router in the transient state "waiting for multicast tree branch to form" (either coming directly from Step 4, or first Step 4 and then through Step 4a) will wait to receive data packets on the new upstream interface without deleting the old upstream interface yet. Once data arrives on the new upstream interface, the agent will actually "install" the new upstream interface and send a prune on the old interface. This is particularly important with FRR support because data may be forwarded by router 3 to router 4 on the backup path for link 3↔4. In Figure 2, router 4 will send a prune to router 3 (along the backup path for link 4↔3), which then forwards it upstream to 1. Then, router 1 will remove the outgoing downstream interface to router 3 and only 2 will be a downstream router.

Step 6: *Switchover is completed and transient state information on (S,G)s are deleted.* Sending the prune to the old interface completes the switchover to the new tree and the router leaves its transient state. This means that all the state information that had to be stored for each (S,G) during the switchover can now be deleted, e.g., for the old upstream interface, and (if any) the state on a downstream link that it is waiting to send a join on.

V. Simulations

In order to perform an initial evaluation of our cross-layer failure restoration framework, we performed simulation experiments with ns-2 [27]. Specifically, our experiments aim to reveal how much *packet loss* occurs and how many *duplicate packets* are received at the multicast subscribers.

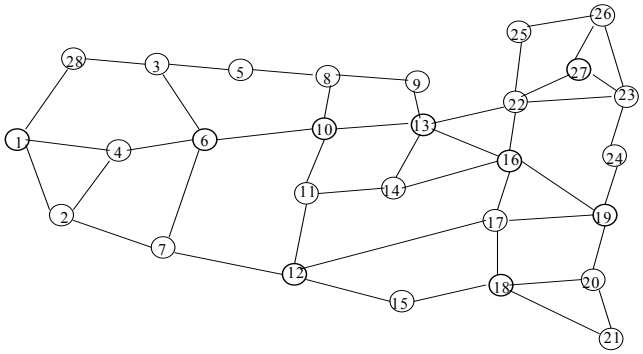

Fig. 4. U.S. Backbone Network Topology.

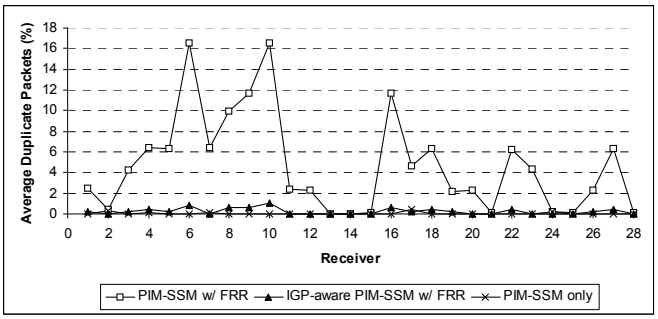

Fig. 5. Average percentage of duplicate packets received.

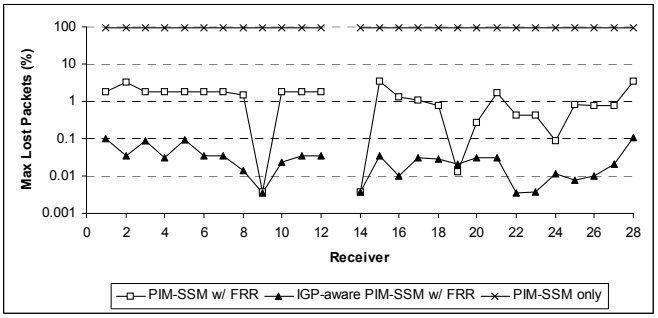

Fig. 6. Maximum percentage of packets lost.

We used MPLS [28] to provide FRR support at the routers and implemented PIM-SSM over MPLS in ns-2. For the IGP routing, we used an OSPF implementation in ns-2 [26]. Finally, we made the necessary changes in MPLS, OSPF, and PIM-SSM implementations to make sure that OSPF is informed about link failures (even though MPLS forwards the data packets on the FRR path immediately after the failure) and PIM-SSM is informed about OSPF routing changes.

A. Experimental Setup

We performed our simulations on a hypothetical US backbone topology (with 28 routers and 45 links) shown in Figure 4. The multicast source is at router 13, roughly at the center of the topology. We assigned equal capacity and 2ms propagation delay to all links. The multicast source generated UDP traffic with packet size of 1000 Bytes. The rate of the multicast traffic was 70% of the capacity of the links. We budgeted for 120ms of buffer time at each link, i.e., a link with a 100Mb/s capacity had a buffer size of 1500 packets.

We used default OSPF timer settings, e.g., 5s *spfDelayTime*, and 10s *spfHoldTime*. To make our comparison more conservative against PIM-SSM, we used a relatively short rejoin interval of 60s for PIM-SSM. Normally, this timer

is supposed to be set to several minutes, resulting in more loss and duplication of packets after a failure than observed here.

Our simulation scenarios consisted of failing a link in the topology and observing the multicast data traffic during re-convergence of the protocols. We measured packets lost or duplicated at each receiver until the restoration from the link failure is complete. We compared three scenarios: (i) PIM-SSM only (i.e., no FRR support and no IGP-aware tree reconfiguration), (ii) PIM-SSM w/ FRR (i.e., FRR support exists but no IGP-aware tree reconfiguration), and (iii) IGP-aware PIM-SSM w/ FRR (i.e., both FRR support and IGP-aware tree reconfiguration are used – our proposal). To compare these three scenarios, we measured the packet loss and duplication metrics until PIM-SSM rejoins, i.e., 60s. Note that the PIM-SSM convergence time is normally 3-4 times longer than the rejoin timer, as all unnecessary downstream branches will have to be pruned. However, we are conservative and use only the rejoin time in our comparison, since PIM-SSM will guarantee that multicast traffic will flow to all receivers after this timer expires. We assigned OSPF link weights by using the algorithm in [25] which assures that single failures do not cause any traffic overlap for the "PIM-SSM w/ FRR" case. When selecting the FRR paths, we chose the shortest non-overlapping path restoring the failure.

B. *Results*

We failed each of the 45 links in the topology one by one and measured the average number of duplicated or lost packets at each receiver (shown in Figures 5 and 6). The plots clearly show that our IGP-aware mechanism to reconfigure PIM-SSM consistently achieves little or no packet loss. Indeed the only packet loss that takes place in our scheme is due to the packets in transit on the failing link at the time of failure, and potential congestion on the FRR path. Since our scheme utilizes the FRR path less than PIM-SSM w/ FRR, the loss due to FRR path congestion is significantly less as well. Further, our scheme achieves this hitless switchover without introducing any significant packet duplication as seen in Figure 5.

VI. SUMMARY

We presented a method to make PIM-SSM reconvergence aware of the underlying network failure conditions. The method employs FRR support at the link layer, but makes the IGP layer aware of the failure and also notifies the multicast agent of the routing changes. Our method reduces the amount of packet loss and duplication for a multicast session.

As we outlined, during a switchover to a new multicast tree, a crucial problem is the overlap of the FRR path with the new tree. Future work will involve investigation of the congestion on these overlapping links, i.e., the congested common link problem. Further, we also plan to investigate the performance of our schemes under multiple simultaneous failures.

ACKNOWLEDGMENT

We thank Guangzhi Li, Kostas Oikonomou, Rakesh Sinha and Dongmei Wang of AT&T Labs Research for assistance in coming up with the cross-layer restoration architecture and Justin Burke for sharing his SSM code.

REFERENCES

[1] http://www.sbc.com/gen/press-room?pid=5838
[2] http://www22.verizon.com/FiosForHome/Channels/fios/FiosTV_comingsoon.aspx
[3] http://www.iptvnews.net
[4] http://www.sbc.com/Common/files/pdf/IPvideonetwork.pdf
[5] A. Adams, J. Nicholas, and W. Siadak, "Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol specification (revised)," *IETF RFC 3973*, 2005.
[6] D. Estrin, Ed., "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol specification", *IETF RFC 2362*, 1998.
[7] S. Bhattacharyya, Ed., "An overview of Source-Specific Multicast (SSM)," *IETF RFC 3569*, 2003.
[8] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching architecture," *IETF RFC 3031*, 2001.
[9] P. Pan, G. Swallow, and A. Atlas (Editors), "Fast reroute extensions to RSVP-TE for LSP tunnels," *IETF RFC 4090*, 2005.
[10] Cisco document, "MPLS traffic engineering fast reroute: Link protection," http://www.cisco.com.
[11] J. Moy, "*OSPF Anatomy of an Internet Routing Protocol*", Addison Wesley, 2000.
[12] M. Cha, W. A. Chaovalitwongse, Z. Ge, J. Yates, and S. Moon, "Path protection routing with SRLG constraints to support IPTV in WDM mesh networks," *Proc. of IEEE Global Internet Symposium*, 2006.
[13] P. Francois, C. Filsfils, J. Evans, and O. Bonaventure, "Achieving sub-second IGP convergence in large IP networks," *ACM SIGCOMM Computer Communication Review*, 35(3), pp. 35-44, July 2005.
[14] K. Oikonomou, R. Sinha, and R. Doverspike, "Multi-layer network performance and reliability analysis", submitted to *J. of Operations Research*.
[15] D. Velten, R. Hinden, and J. Sax, "Reliable Data Protocol (RDP)," *IETF RFC 908*, 1984.
[16] W. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Transactions on Circuits and Systems for Video Technology*, 11(3), pp. 373-386, 2001.
[17] Y. Xiong and L. Mason, "Restoration strategies and spare capacity requirements in self-healing ATM networks," *IEEE/ACM Transactions on Networking*, 7(1), pp. 98-110, 1999.
[18] K. Nahrstedt and R. Steinmetz, "*Multimedia Fundamentals, Volume 1: Media Coding and Content Processing*", 2nd Ed. Prentice Hall, 2002.
[19] B. Quinn, K. Almeroth, "IP multicast applications: Challenges and solutions," *IETF RFC 3170*, 2001.
[20] M. Handley et al, "The reliable multicast design space for bulk data transfer," *IETF RFC 2887*, 2000.
[21] M. Cha, S. Moon, C.-D. Park, and A. Shaikh, "Placing relay nodes for intra-domain path diversity," *Proc. of IEEE INFOCOM*, April 2006.
[22] G. Li, D. Wang, and R. Doverspike, "Efficient distributed MPLS P2MP fast reroute," *Proc. of IEEE INFOCOM*, April 2006.
[23] R. Braden (Ed.), L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- version 1 functional specification," *IETF RFC 2205*, 1997.
[24] S. Lee, Y. Yu, S. Nelakuditi, Z.-L. Zhang, and C.-N. Chuah., "Proactive vs. reactive approaches to failure resilient routing," *Proc. of IEEE INFOCOM*, March 2004.
[25] R. Doverspike, G. Li, K. Oikonomou, K. K. Ramakrishnan, and D. Wang, "IP backbone design for multimedia distribution: architecture and performance," *Proceedings of IEEE INFOCOM*, April 2007.
[26] M. Goyal, K. K. Ramakrishnan, and W.-C. Feng, "Achieving faster failure detection in OSPF networks", *Proc. of IEEE ICC*, May 2003.
[27] The Network Simulator, ns-2. http://www.isi.edu/nsnam/ns.
[28] G. Ahn and W. Chun, "Design and implementation of MPLS network simulator supporting LDP and CR-LDP," *Proc. of IEEE International Conference on Networks (ICON)*, 2000.