



Figure 5: Example: Switching from shared tree to shortest path tree. Actions are numbered in the order they occur

the (S_n, G) entry.

A PIM-Join message will be sent upstream to the best next hop towards the new source, S_n , with S_n in the join list: Multicast-Address= G , PIM-join= S_n , PIM-prune=NULL. The best next hop is determined by the unicast routing protocol.

When a router which has a (S_n, G) entry with the SPT bit cleared, starts to receive packets from the new source S_n on the interface used to reach S_n , it sets the SPT-bit. The router will send a PIM-prune toward the RP if its shared tree incoming interface differs from its shortest path tree incoming interface; indicating that it no longer wants to receive packets from S_n via the RP tree. In the PIM message toward the RP, it includes S_n in the prune list, with the WC-bit set indicating that a negative cache should be set up on the way to the RP. A negative cache entry is a (S, G) entry with null outgoing interface list. Data packets matching the negative cache are discarded silently.

When the S_n, G entry is created, the outgoing interface list is copied from $*, G$, i.e. all local shared tree branches are replicated in the new shortest path tree. In this way when a data packet from S_n arrives and matches on this entry, all receivers will continue to receive source packets along this path unless and until the receivers choose to prune themselves.

Note that a DR may adopt a policy of not setting up a (S, G) entry (and therefore not sending a PIM-Join message toward the source) until it has received m data packets from the source within some interval of n seconds. This would eliminate the overhead of (S, G) state upstream when small numbers of packets are sent sporadically (at the expense of data packet delivery over the suboptimal paths of the shared RP tree). The DR may also choose to remain on the RP-distribution tree indefinitely instead of moving to the shortest path tree. Note that if the DR does join the SPT, the path changes for all directly connected and downstream receivers. As a result, we do not "guarantee" that a receiver will remain on the RP tree; if receiver A's

RP tree overlaps with another receiver B's SPT, receiver A may receive its packets over the SPT. A multicast distribution tree is a resource shared by all members of the group; to satisfy individual receiver-specific requirements or policies the multicast tree might degenerate into a set of receiver-specific unicast paths.

3.4 Steady state maintenance of router state

In the steady state each router sends periodic refreshes of PIM messages upstream to each of the next hop routers that is en route to each source, $(S, *)$ for which it has a multicast forwarding entry (S, G) ; as well as for the RP listed in the $(*, G)$ entry. These messages are sent periodically to capture state, topology, and membership changes. A PIM message is also sent on an event-triggered basis each time a new forwarding entry is established for some new (S_n, G) (note that some damping function may be applied, e.g., a merge time). Optionally the PIM message could contain only the incremental information about the new source. The delivery of PIM messages does not depend on positive acknowledgement; lost packets will be recovered from at the next periodic refresh time.

3.5 Multicast Data Packet Processing

Data packets are processed in a manner similar to existing multicast schemes. An incoming interface check is performed and if it fails the packet is dropped, otherwise the packet is forwarded to all the interfaces listed in the outgoing interface list (whose timers have not expired). There are two exception actions that are introduced if packets are to be delivered continuously, even during the transition from a shared to shortest path tree.

1. When a data packet matches on an (S, G) entry with a cleared SPT bit, if the packet does not match the incoming interface for that entry, then the packet is forwarded according to the $*, G$ entry; i.e., it is sent