

Multicast Routing Blackhole Avoidance

draft-asati-pim-multicast-routing-blackhole-avoid-00

Rajiv Asati

Mike McBride

Agenda

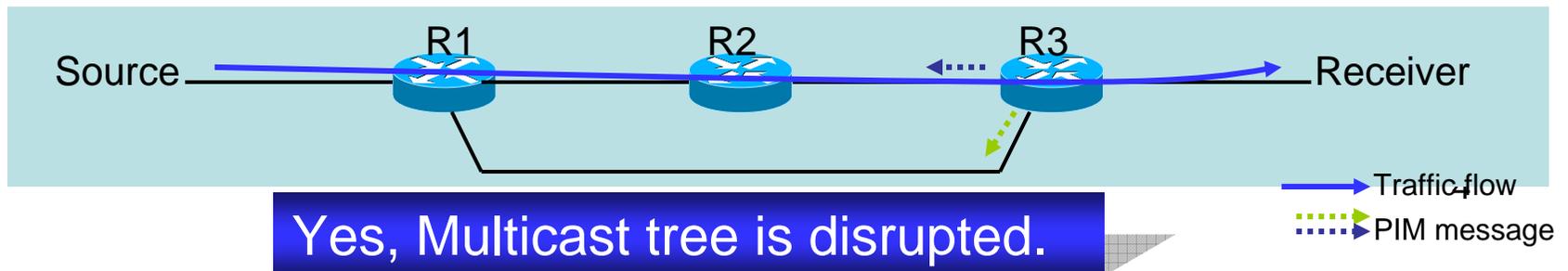
- Background
- Problem Statement
 - Root Cause
- Solution
- Advantage
- Next Steps

Background

- In a network, interface DOWN and UP events may happen.
- There are scenarios in which multicast traffic may get dropped after the local interface UP event or metric change etc.
- This draft proposes a simple and straightforward solution for such scenarios.

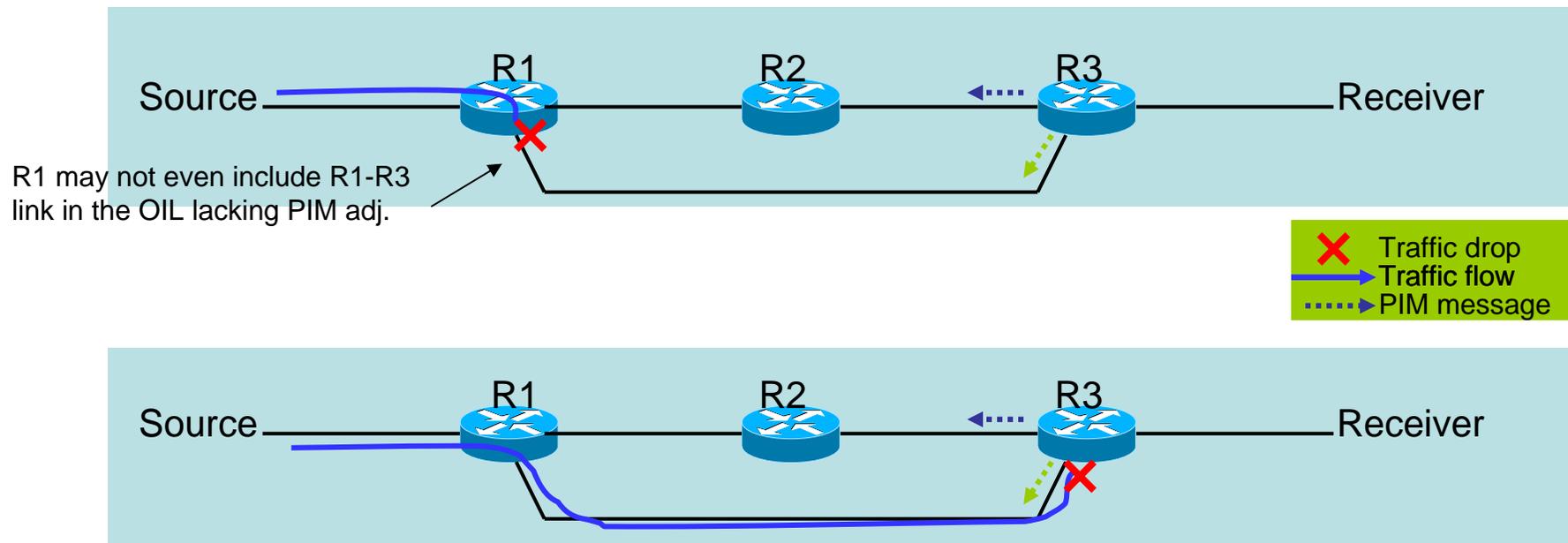
Problem

- The multicast traffic is dropped wherever the related multicast tree is disrupted.
- Is there a mcast tree disruption in this sequence ??
 1. Multicast tree is set up to deliver the traffic as R1 -> R2 -> R3
 2. Interface between R1 and R3 comes UP
 3. R1-R3 interface becomes usable per unicast routing IGP/BGP
 4. R3 calculates the new RPF interface for source to be R1-R3 int.
 5. R3 changes the multicast tree to be R1 -> R3
 6. R3 attempts to send **PIM join** to R1 and **PIM Prune** to R3
 7. What if the R1-R3 int doesn't yet have the PIM adjacency !!! ☹



Problem

- The multicast tree gets disrupted either on R1 or R3, and traffic gets blackholed.



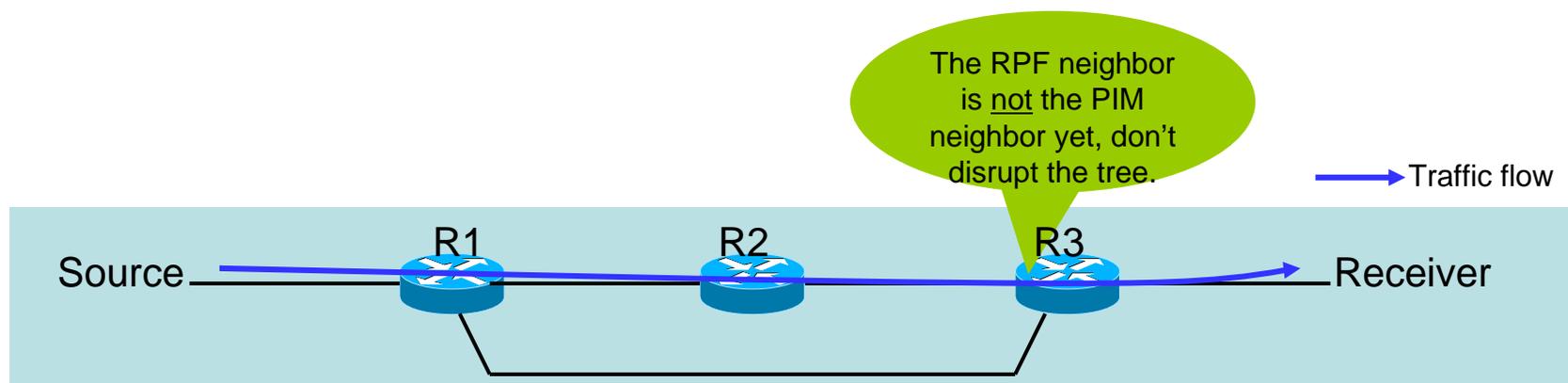
- The blackhole continues until the PIM adjacency is established on R1-R3.

Root Cause

- Firstly, the PIM neighbor establishment on an interface may take longer
 - than the total time taken to establish unicast routing neighbor on that interface, determine the new RPF neighbor, if any, and update the corresponding multicast route entry(s).
- Secondly, a multicast routing entry may be updated with the new RPF neighbor information immediately after the unicast routing convergence.
 - There is no check for whether the RPF neighbor is also the PIM neighbor prior to this update.
- The latter is really what that causes the disruption in multicast tree.

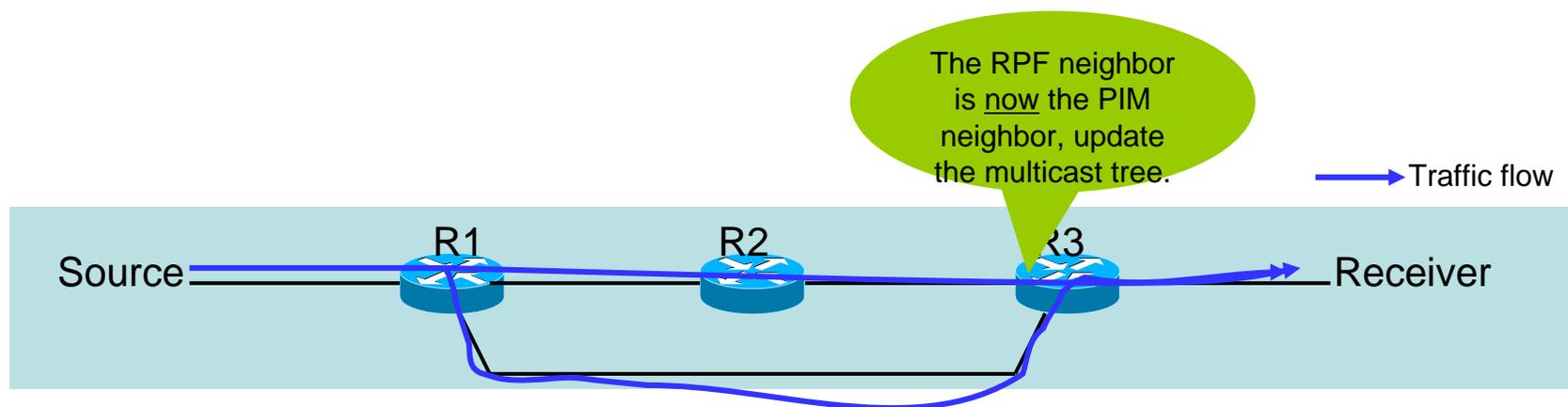
Solution

- The solution is to replace the current RPF neighbor with the new RPF neighbor for a multicast routing entry ONLY if the new RPF neighbor is determined to be the PIM neighbor.



Solution

- The solution is to replace the current RPF neighbor with the new RPF neighbor for a multicast routing entry ONLY if the new RPF neighbor is determined to be the PIM neighbor.



Solution Details (Section 3.1)

The multicast routing sequence becomes as following –

1. New RPF neighbor is determined after the unicast routing convergence.
2. Check whether the new RPF neighbor is also PIM neighbor on the RPF interface
3. If it is not, then initiate the PIM neighbor establishment procedure by sending PIM Hellos etc. on the new RPF interface, and wait.
4. If it is, then update the multicast forwarding entry by replacing the old RPF neighbor (and interface) information with the new RPF neighbor (and interface) information
5. Send the PIM Join message for the related multicast routes (S,G or *,G).
6. Send the PIM Prune message to the old RPF neighbor for the related multicast routes (S,G or *,G)

Advantage

- This solution ensures that the multicast distribution tree is not disrupted unnecessarily and the multicast traffic is not blackholed just because a link comes UP.
 - Of course, the multicast distribution tree may not be able to make use of the changed topology i.e. new link for brief time period.
 - Paves the way for make-before-break (to be updated).
- Simple and straight-forward PIM based solution.
- Changes are local to the router. No interoperability desired.

Next Steps

- Incorporate the feedback received so far.
 - Will update the section 4 in the next version.
- **Request WG Adoption** (as an informational draft)!!!

Comparison with other solutions

- This solution nicely avoids the tree disruption due to the local link Up event.
- This draft does not attempt to address the tree disruption due to remote link Up event.
 - Well, it is out of scope for any hop-by-hop protocol (since it is not a link-state protocol) .
 - LDP based MPLS LSP, for example.
- For the remote link Up event, a solution based on having incongruent multicast topology and dedicated SPF for that topology may be desired. However, it comes with the baggage that may not be suited for every network -
 - Should **incongruence be mandatory** or optional?
 - Should the **multi-topology be really mandated** ?
 - Should the router be forced to have multiple RIBs with common information?
 - Can the **operations/troubleshooting network mgt applications** deal with multi-topology etc. right away?
 - Will there be any **negative impact** on ucast or mcast convergence ?