Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on August 18, 2008.

Copyright Notice

Copyright (C) The IETF Trust (2008).

Abstract

The goal of this document is to investigate the solution space for improving the recovery time of multicast trees built by the variants of the PIM routing protocol in the case of various topological failures (including links and nodes).
Table of Contents

1. Introduction ............................................. 3
2. Motivation .............................................. 3
3. Problem analysis ........................................... 4
   3.1. Failure types ......................................... 4
   3.2. Convergence and recovery time analysis ............... 4
      3.2.1. Failure detection ................................ 5
      3.2.2. Failure notification time ......................... 5
      3.2.3. Path re-computation / update of the unicast RIB . 5
      3.2.4. Unicast RIB update to MRIB update time .......... 5
      3.2.5. Distributed Multicast Tree update time .......... 5
      3.2.6. TIB to MFIB update time ........................... 6
4. Solution evaluation framework ............................. 6
   4.1. Scaling factors ....................................... 6
   4.2. Necessary resources to support the FRR scheme .... 6
   4.3. Coverage ............................................. 6
5. Solution space ............................................ 6
6. References .................................................. 7
Authors’ Addresses ............................................ 7
Intellectual Property and Copyright Statements ............... 9
1. Introduction

Recently, Service Providers (SP) have increased their interest in the restoration times of the services supported by multicast routing protocols. A typical deployment scenario for the support of such services in the core of a SP network uses the Protocol Independent Multicast protocol. Such deployments typically rely on unicast interior gateway protocols (IGP) such as IS-IS [1] or OSPF [2].

The goal of this document is to investigate the solution space for improving the recovery time of multicast trees built by the variants of the PIM routing protocol [3] in the case of failures.

The first section of this draft describes the motivation of the framework. Section 2 draws tracks on the failure types to be considered in the framework. Section 3 provides an evaluation framework for the solution proposals. Section 4 discusses the solution space.

2. Motivation

This first version of the draft aims at triggering community interest and collaboration. Another aim of the draft is to solicit input and views of the community on the following aspects of the problem.

i) Which "modes" of PIM should be investigated. PIM {SM, SSM, BIDIR, DM}

ii) Which are failure cases (e.g., link, node, SRLG, PIM-SM RP, source, leaves, ...) that should be addressed by new proposals

iii) What is the expected performance for the proposed solutions (coverage, recovery time, state overhead, bandwidth usage)

iv) How to evaluate the possible solutions.

Each of the investigated PIM variants considered by the framework would be evaluated for its current strengths and weaknesses w.r.t. convergence time and corrective measures proposed and evaluated.

Investigating Fast Reroute (FRR) solutions for multicast could introduce another set of coverage/performance issues compared to the techniques that are worked on for unicast [4]. For some variants of PIM, and for some levels of coverage, these may be small "pin-point" improvements, while for others more fundamental reconsideration may be required.
A basic (and then maybe advanced) FRR approach may thus be followed as for the unicast IP-FRR work.

Some solutions to improve multicast resiliency have already been introduced to solve specific aspects of the problem space. For instance, Anycast RP for PIM-SM [5] has been introduced to provide a faster convergence in the case of the failure of an RP.

The goal of this framework would be to analyze existing solutions as well as new proposals aimed at reducing the impact of the scaling factors applicable during PIM convergence.

3. Problem analysis

This section aims at describing the failure types that should be considered for the improvement of PIM resiliency, and provides an overview of the main components of involved in PIM convergence.

3.1. Failure types

All types of failures triggering a convergence of the unicast routing protocols on which PIM relies should be in scope, i.e., point-to-point, point-to-multipoint links (e.g. using P2MP MPLS LSP), and multi-access link failures (e.g. Ethernet LAN segments), node failures, and SRLG failures.

Also, failures should be analyzed with respect to the multicast-specific role of the failed entity. This includes for example the failure of a Source, a Leaf, a Rendez-Vous Point (RP), and DR for the case of PIM-SM.

3.2. Convergence and recovery time analysis

We term as recovery time the time between the interruption of a multicast stream and when all the receivers receive multicast packets of that stream, as a result of the failure affecting a multicast distribution tree.

We term as convergence time the time after which all the MFIB updates have been performed by all the routers as a result of the failure affecting a multicast distribution tree.

PIM routing recovery and convergence times depend on the variant of PIM that is used, the size and the shape of the network topology and the number of groups affected by the failure.

The main components of the convergence can be sketched as follows.
Note that the listed components do not necessarily take place in sequence during the convergence process.

3.2.1. Failure detection

We identify three types of failure detection related to multicast convergence.

1. Multicast Routing Protocol dependent failure detection: MSDP (remote), PIM Hellos (local)
2. Unicast Routing protocol dependent failure detection: IGP Hellos (local)

3.2.2. Failure notification time

This is the IGP Link-State update flooding time to routers that will have to perform an MFIB update during the convergence.

3.2.3. Path re-computation / update of the unicast RIB

This the time to update the Unicast RIB entries that will themselves trigger a change of RPF neighbours in PIM.

3.2.4. Unicast RIB update to MRIB update time

This is the time between the change of an entry in the unicast RIB and the time at which the corresponding RPF neighbour information is updated in the MRIB.

3.2.5. Distributed Multicast Tree update time

This is the time required to let the distributed collection of TIB states of the multicast routers re-form a tree.

This component can be decomposed as

- The time between a change in the MRIB and the time at which a join and a prune message is sent as a result of the state change.
- The time required to update the oif lists according to the received Join and Prune messages.
- The time required to re-elect a designated router with the exchange of Assert messages.
3.2.6. TIB to MFIB update time

This is the time to propagate the Tree Information Base changes to the MFIB entries in the forwarding plane.

4. Solution evaluation framework

Multiple aspects of the proposed solutions should be evaluated and compared with the conventional PIM recovery process.

4.1. Scaling factors

The scaling factors of the recovery time induced by the solution should be described.

4.2. Necessary resources to support the FRR scheme

The necessary resources to support the solution should be evaluated. This includes the capacity overhead to support the solution when a failure occurs and the FRR scheme is activated. This also includes the state to be maintained and the mRIB/mFIB complexity to support the solution.

4.3. Coverage

The detailed "coverage" of the proposed scheme should be analyzed.

Proposals should define the PIM variants that they targets. It is also acknowledged that multiple enhancements may be considered such as to cover e.g. Shared trees and Source trees in PIM-SM.

The ability to protect multicast distribution trees from the diverse failure types considered in the framework should be discussed.

If the ability of the solution to protect multicast distribution trees depends on the characteristics of the topology (layout and link metrics), it should be mentioned and the applicability to typical deployments should be evaluated.

5. Solution space

To initiate investigation, we preliminarily identify two main tracks of proposals.

o) Track 1: unicast FRR solutions used to protect multicast traffic.
Such solutions would basically re-use or extend an existing unicast Fast Reroute scheme to protect multicast trees.

This implies that the unicast FRR scheme incorporates a certain level of "multicast-awareness". One can see two components of this PIM routing awareness: either implicit or explicit. In the former, the re-routing scheme is extended so as to decrease the time required for PIM routing protocol messages to be exchanged after failure occurrence (such as decrease re-convergence time of the multicast state). This involves mainly tuning the unicast routing re-convergence so to decrease time for operation described in Section 3.2.3, 3.2.4, and 3.2.5. In the latter, the unicast FRR scheme is extended to decrease time required to propagate fail-over information by retro-fitting it into the unicast FRR scheme. This involves mainly tuning the unicast routing re-convergence so to decrease time for operation described in Section 3.2.2.

An open investigation point is whether or not a solution that will be recommended for multicast will be the same as its unicast counterpart.

o) Track 2: PIM built-in extensions to improve resiliency capabilities.

Basic and advanced FRR solutions can be proposed in both tracks.

Existing solutions: Anycast RP, Dual multicast topologies, Push conventional PIM convergence to the limits

These solutions tackle specific failure cases and rely on abstracting reachability and/or topology. Another approach consists in tweaking Hello timers. Such approach leads to faster failure detection but also increases processing overhead and results in PIM neighbor being declared down due to missed Hellos (if Hello packets are not prioritized). Another drawback is the dependency created between PIM Hello exchanges for maintaining link/interface liveness. Indeed, PIM Hello messages are sent on each PIM-enabled interface to learn about the neighboring PIM routers on each interface, elect a Designated Router (DR), and to negotiate additional capabilities.

The alternative suggest in this track consists in extending PIM mechanisms (potentially by the use of another for fast failure detection) to improve the convergence time. The multicast routing-specific components that can benefit from such improvement are mainly related to the time needed for sending a Join/Prune message as a result of the multicast state change. This improvement must be accompanied by a set of conditions to prevent transients loops that may be induced from the use of multiple MFIBs entries for the multicast group (resulting from PIM Join exchanges prior and after failure).
6. References


Authors’ Addresses

Dimitri Papadimitriou
Alcatel-Lucent
BE

Email: dimitri.papadimitriou@alcatel-lucent.be

Pierre Francois
Universite catholique de Louvain
Place Ste Barbe, 2
Louvain-la-Neuve 1348
BE

Email: pierre.francois@uclouvain.be