RTGWG
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2011

W. Lu
A. Tian
S. Kini
Ericsson

October 18, 2010

# Fast Notification Framework
# draft-lu-fast-notification-framework-00

## Abstract

This document describes an architectural work that competes with the IP Fast Re-Route (IPFRR) solution which aims to minimize the network down time in the event of equipments failure. The work provides a layered framework based upon which applications such as the domain-wide fast convergence may be achieved through the transport layer fast delivery of failure notifications.

## Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on April 21, 2011.

Table of Contents

## 1. Introduction

The ability to recover rapidly from network failures is one of the most sought network characteristics. Few solutions address this issue to the satisfactory.

IPFRR [RFC5714] is one such solution. It mimics MPLS-FRR [RFC4090] solution. The difference is that the MPLS-FRR is path based, or source routing based in other words. This implies that the re-route decision can be carried out by the PLR (point-of-local-repair) router alone, with no need of cooperation of other LSRs in the network.

Unfortunately, IP based FRR is by nature not source routing based. Its re-route decision may not be honored by other routers in the network. The consequence can be very severe, either traffic outage or even routing loops.

Many methods were proposed around IPFRR concept but none is close to be satisfactory. Some methods such as LFA described in [RFC5286] require lot of computation and have coverage issue. Some others such as Not-Via [I-D.ietf-rtgwg-ipfrr-notvia-addresses] are extremely complicated and are prohibitive to be useful.

The primary reason for such difficulties can be understood from the following passage which is quoted from [RFC5714] first paragraph of section 1:

However, there is an alternative approach, which is to compute backup routes that allow the failure to be repaired locally by the router(s) detecting the failure without the immediate need to inform other routers of the failure.

The phrase "without the immediate need to inform other routers of the failure" is against the very nature of the IP network in which the domain-wide synchronization is the key.

In this document we propose a method which directly addresses the rapid network synchronization needs. It is not IPFRR based. However it can achieve the same or better result without much complexity and compromise.

The method lays out a framework which decouples the improvement in the forwarding plane from the control plane. The design also allows and promotes future innovations based upon the framework.

### 1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

### 1.2. Acronyms

FRR - Fast Re-Route

IPFRR - IP Fast Re-Route

MPLS - Multi-Protocol Label Switch

LFA - Loop Free Alternative

TLV - Type Length Value tuple

IGP - Interior Gateway protocol

OSPF - Open Shortest Path First

IS-IS - Intermediate System to Intermediate System

PDU - Protocol Data Unit

DoS - Denial of Service

FNF - Fast Notification Framework

## 2. Event Framework

An event framework is introduced for the purpose of rapid disseminating of events to all interested receivers in a network.

The framework is application independent. Many applications can generate the events and/or register to receive the events. A TLV based framework is proposed to ensure separation between application and the delivery framework.

The event framework is also independent of the underlying delivery mechanisms. Different delivery mechanisms may be introduced, each with different properties suitable for different requirements. For example, some delivery mechanism is solely optimized for simplicity; while other may improve on reliability.

One of the use cases of this event framework is Fast Failure

Notification, which can be used to improve network convergence time. When a failure occurs in a network, routers adjacent to the failure can detect it and quickly disseminate the failure notifications to other routers throughout the area. Routing protocols on different routers can register and receive such failure notifications, then quickly react to the failure to achieve fast convergence.

The routing protocols discussed in this work are Interior Gateway Protocols (IGP) with the focus on the Link State Routing Protocols such as Open Shortest Path First [RFC2328] and Intermediate System to Intermediate System [RFC1195] [ISO.10589.1992].

The event in the scope of this architecture is specifically the link-down event or node-down event. The up events are not fast flooded for the sake of network stability.

## 3. Layered Structure

The framework can be viewed as a layered structure in which various routing functions can be rearranged. This arrangement is based on the principle of separation of functions. It will facilitate the innovation in various component building blocks and in the mean while allow them to integrate in a systematic manner.

There are two layers that make the framework. One is for routing protocol specific functionality. The other is the data transport layer. Figure 1 depicts this concept.

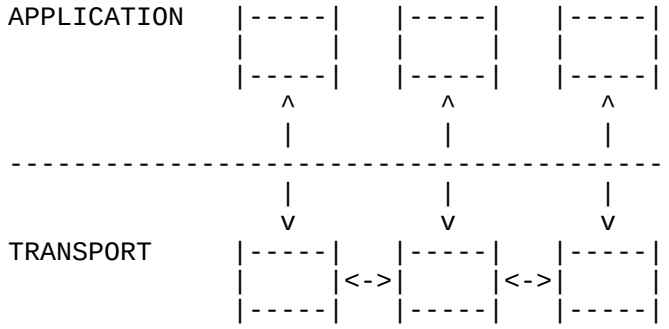


Figure 1: Fast Notification Architecture

Regular routing protocol performs the flooding in store-and-forward manner. While this is reliable (retransmission) and secure (adjacency check), it involves control plane operation and the control plane to data plane communication. It inevitably drags the

network-wide convergence.

With the fast notification architecture, the delivery function is detached from the application layer and moved onto the transport layer. More precisely, the transport layer provides domain-wide fast delivery platform. The normal flooding function is still kept in the application layer to ensure ultimate synchronization in case the fast flooding does not reach some intended routers for whatever reasons.

The speed of the fast flooding needs not to be faster than the data traffic. As long as the messenger travels at the same speed of the data traffic, it always gives the next-hop router the same amount of time for processing as it gives the previous router.

## 4. Operation

Fast failure notification operates on following steps:

1. Failure detection;

2. Notification composing and dispatching;

3. Notification flooding;

4. Notification receiving;

5. Routing/forwarding table update.

### 4.1. Failure detection

This can be made in many ways. But it has to be fast and light-weight. Layer-2 link-event monitoring and signaling is obvious an option. Bidirectional Forwarding detection (BFD) is also a good candidate. There may be more, or combinations of them.

The fast notification architecture encourages the innovation in this area which can be pursued freely and independently.

### 4.2. Notification Origination

This part involves the message format. This document does not specify or endorse a particular format. It is open to any format as long as it fulfills the fast flooding purpose. The detecting router is responsible for the initiation of the fast notification process. Its action is the starting point of the fast flooding.

There are two packet formats worth of mentioning.

#### 4.2.1. IGP PDU

The simplest approach is to use the IGP packet format directly. For example, the OSPF Router-LSA packet which reflects a broken adjacency (one fewer router link) can be fast-flooded to all routers without special modification.

The benefit is that the receivers can process the packet as usual. Moreover since the packet is no different than the one in normal flooding, it guarantees the seamless transition when the "slow" flooding catches up. Plus, there will be no duplicate effort of fast and slow convergence. Flooding stops wherever a router is updated (already fast flooded).

The drawback is that the message cannot be made uniform for multiple protocols. Other protocol such as IS-IS will have to devise a different format. In addition, since IS-IS PDU is not IP based, it may require encapsulation in some cases.

Another drawback is that the normal IGP flooding uses adjacency check to prevent DoS attack or PDU replay from un-trusted parties. The check has to be bypassed for the fast-flooded packets to be accepted. This opens door to the DoS or some other attacks. Domain-wide authentication may be adopted for protection.

#### 4.2.2. Uniform Message

This format must include essential and sufficient information about the broken link. The message will be treated on the receiver router as a local event. The uniformed messaging provides freedom for future expansion. The format thus is recommended TLV-based.

Cautions must be taken in case the message is mistakenly flooded due to bugs or some error conditions. Timeout machinery may be used to protect against such issues.

The detecting router is responsible for the initiation of the fast notification process. Its action is the starting point of the fast flooding.

### 4.3. Fast Flooding

The fast flooding does not specify the fast flooding mechanism. It is up to the routing society to figure out and single out good solutions. The requirement is that the flooding has to be

a. Reliable in that it reaches all participants even after failures occur;

b. Loop-free;

c. Simple;

d. Can be authenticated.

### 4.4. Notification Receiving and Handling

This involves upon the arrival of the notification message, how it is forwarded to the routing protocol for further processing. If the fast-flooding scheme uses specific IP destination addresses or MAC addresses, the receiving router has to recognize it.

When the message reaches the protocol process, it may have to relax its acceptance criteria.

If in the future, some algorithm is developed that the notification handling takes very few CPU cycles, this process may be performed in real-time. Therefore it is worthy of considering move the notification handling into the data plane. This will cut a large chunk of delay and may lead to hitless domain-wide convergence.

### 4.5. Routing/Forwarding Table Update

This should be the same as normal IGP decision process. It is also possible to pre-download the changes to the data plane if the complexity can be limited. This will improve the overall convergence time dramatically.

## 5. Convergence Analyses

### 5.1. Definition of Convergence Time

The convergence time is measured by dividing the number of lost packets with the traffic flow rate between any two routers in the domain. This SHOULD equal to the domain wide network convergence time if all individual routers have the same computing power and the same convergence time.

### 5.2. Domain Wide Convergence

Due to the propagation delay, all routers do not converge at the same time. The traffic loss, however, stops immediately after the first router repairs.

This is because the data traffic has to go through the same propagation delay, which exactly compensate the late starting of the convergence at remote routers.

Take a ring topology for example, as shown in Figure 2.

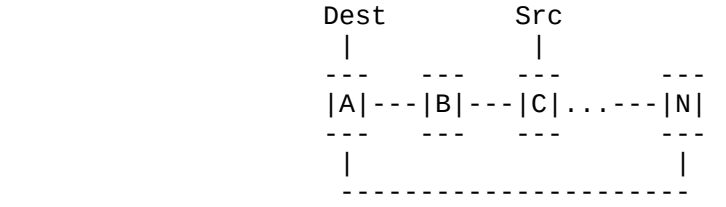


Figure 2: Ring Topology

Assume all routers have same convergence time 50 milliseconds. Assume the transmission delay over each hop is 20 milliseconds.

Upon link A-B failure, B floods its Link State Update to C. Table 1 shows the convergence timeline.

| Node | Converge Starts | converge Completes |
|---|---|---|
| B | 0 | 50ms |
| C | 20ms | 70ms |

During the first 50 milliseconds, packets from B to A are dropped. Right after 50th milliseconds, B re-routes packets toward C. Those packets, after traveling 20 milliseconds, arrive C at 70th milliseconds when C is just repaired. Since C and all downstream routers will correct themselves one by one right before those packets arrive, they will arrive at the destination via the corrected path successfully. The overall convergence time is thus same as B's.

## 5.3. Micro-looping

If routers' convergence time is different, micro looping may form, although packets will still be delivered after several loops. Still use Figure 2 for example. Assume C needs 90 milliseconds to converge. When B re-routes packets back to C at 70th milliseconds, C has not finished its updating yet. It continues to use its old forwarding table and bounces packets back to B. B in turn re-route packets again to C. This time packets arrive at C at 110th milliseconds. C has done updating and will forward packets

correctly. The packets are looped once.

The micro-looping does not form easily with Fast Flooding method. The routers have to differ in computing speed and differ significantly.

### 5.4. Packet Reordering

Due to the different convergence timeline, packets may be temporarily forwarded in wrong direction before being placed on the right track. This will not cause packet loss, but will result in packet reordering.

Packet reordering affects TCP communication adversely in that new sequence numbered packets may arrive ahead of the older ones.

This problem is common in IPFRR solutions, and remains an open issue. Not-Via for example, may have packets reordered when it switches to use the final stable routes from the temporary LFAs. On the other hand, the connectionless network by nature never promises ordered packet delivery. This type of problem deserves a separate topic and is beyond the scope of this document.

## 6. Scalability Analyses

Fast Flooding scales with networks of any size and any topology. At least it scales no inferior to the normal IGP flooding.

## 7. Traffic Analyses

Traffics that did not route through the broken link are intact. Traffics that did will be successfully re-routed as soon as the affected router converges (as opposed to all routers converge).

Upon the convergence of the affected router, Fast Flooding guarantees correct routes for all affected traffics.

## 8. Acknowledgements

TBD

## 9. IANA Considerations

This memo includes no request to IANA.

## 10. Security Considerations

TBD

## 11. References

### 11.1. Normative References

[ISO.10589.1992]
International Organization for Standardization, "Intermediate system to intermediate system intra-domain-routing routine information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)", ISO Standard 10589, 1992.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

### 11.2. Informative References

[I-D.ietf-rtgwg-ipfrr-notvia-addresses]
Shand, M., Bryant, S., and S. Previdi, "IP Fast Reroute Using Not-via Addresses", draft-ietf-rtgwg-ipfrr-notvia-addresses-05 (work in progress), March 2010.

[RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

[RFC5286] Atlas, A. and A. Zinin, "Basic Specification for IP Fast Reroute: Loop-Free Alternates", RFC 5286, September 2008.

[RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, January 2010.

Authors' Addresses

Wenhu Lu
Ericsson
300 Holger Way
San Jose, California 95134
USA

Phone: 408 750-5436
Email: wenhu.lu@ericsson.com

Albert Tian
Ericsson
300 Holger Way
San Jose, California 95134
USA

Phone: 408 750-8739
Email: albert.tian@ericsson.com

Sriganesh Kini
Ericsson
300 Holger Way
San Jose, California 95134
USA

Phone: 408 750-5210
Email: sriganesh.kini@ericsson.com