

Network Working Group
Internet-Draft
Intended status: Informational
Expires: May 3, 2009

M. Shand
S. Bryant
Cisco Systems
P. Francois
Universite catholique de Louvain
October 30, 2008

Mechanisms for safely abandoning loop-free convergence (AAH)
draft-bryant-francois-shand-ipfrr-aah-01

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/lid-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 May 3, 2009.

Abstract

IPFRR and loop-free convergence techniques can deal with single topology change events, multiple correlated change events, and in some cases even certain uncorrelated events. However, in all cases there are events which cannot be dealt with and the mechanism needs to quickly revert to normal convergence. This is known as "Abandoning All Hope" (AAH). This document describes the nature of the problem, and various proposed mechanisms to deal with it.

Internet-Draft

Abandon All Hope (AAH)

October 2008

Table of Contents

1.	Conventions used in this document	3
2.	Introduction	3
3.	Possible Solutions	4
3.1.	Hold-down timer only	4
3.2.	Basic per event AAH messages	4
3.3.	AAH messages	5
3.3.1.	Per Router State Machine	6
3.3.2.	Per Neighbor State Machine	8
4.	Management Considerations	9
5.	Scope and applicability	9
6.	IANA Considerations	9
7.	Security Considerations	9
8.	Acknowledgements	9
9.	References	10
9.1.	Normative References	10
9.2.	Informative References	10
	Authors' Addresses	10
	Intellectual Property and Copyright Statements	12

Internet-Draft

Abandon All Hope (AAH)

October 2008

1. Conventions used in this document

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](#) [1].

2. Introduction

IPFRR[2] and loop-free convergence techniques[3] can deal with single topology change events, multiple correlated change events, and in some cases even certain uncorrelated events. However, in all cases there are events which cannot be dealt with and the mechanism needs to quickly revert to normal convergence. This is known as "Abandoning All Hope" (AAH).

A good example is the case of the ordered FIB loop-free convergence technique (oFIB)[4], however the problem and the mechanisms described here for its resolution are equally applicable to any loop free convergence mechanism, such as PLSN[5]. All the routers performing the calculation must have an identical view of the set of topology changes under consideration. One technique to ensure this is to start a hold-down timer on reception of the first event in the hope that all subsequent events related to the same root cause will arrive before the timer expires. If this is the case, then all routers in the network will have acquired an identical set of changes and processing can continue correctly. However, in some cases the timer value will be too short to ensure that all the related events have arrived at all routers (perhaps because there was some unexpected propagation delay, or one or more of the events are slow in being detected). In other cases, a completely unrelated event may occur after the timer has expired, but before the processing is complete. In either case it is necessary to "Abandon all Hope" and revert to traditional convergence.

There are a number of problems with this naive approach. Firstly,

since the timer is started at each router on reception of the first LSP announcing a topology change, the actual starting time is dependant upon the propagation time of the first LSP. So, for a subsequent event occurring around the time of the timer expiry, because of variations in propagation delay it may reach some routers before the timer expires and others after it has expired. In the former case this LSP will be included in the set of changes to be considered, while in the latter it will be excluded and would invoke an AAH in the routers receiving it. Clearly this would be a dangerous condition, and it is therefore necessary to arrange that an AAH invoked anywhere in the network causes ALL routers to AAH. This can be achieved by reliably propagating an AAH message throughout the

network. However, this raises a second problem, the need to synchronize the exit from AAH state throughout the network.

While in AAH state any topology changes previously received, or which are subsequently received, should be processed immediately using the traditional convergence algorithms i.e. without invoking controlled convergence. If the exit from the AAH state is not correctly synchronized, a new event may be processed by some routers immediately (as AAH), while those which have already left AAH state will treat it as the first of a new batch of changes and attempt controlled convergence.

[3.](#) Possible Solutions

A number of approaches to this problem have been proposed, in increasing order of complexity:

1. Hold-down timer only. This is the solution proposed in PLSN.
2. Basic per event AAH messages
3. Synchronization of AAH state using AAH messages.

These are described below. The purpose of this draft is to trigger discussion on the trade-offs between complexity and robustness in the AAH solution-space.

[3.1.](#) Hold-down timer only

This method uses a hold-down to acquire a set of LSPs which should be processed together. On expiry of the local hold-down timer, the router begins processing the batch of LSPs according to the loop free prevention algorithm.

[3.2.](#) Basic per event AAH messages

This method uses signaling between neighbors to announce the abandoning of controlled convergence.

A router individually decides when it should abandon controlled convergence for a given (set of) LSP(s). It bases this decision on the LSP reception timings and the hold down timers defined for the controlled convergence mechanism used.

When a router makes a decision to abandon controlled convergence for an LSP, it sends an AAH message to a selected subset of its neighbors. The message identifies the LSPs for which controlled

convergence was abandoned.

The reception of such a message MUST trigger the decision to abandon controlled convergence for this LSP by the receiver. The receiver SHOULD also abandon controlled convergence for the other pending LSPs.

A router is only allowed to send AAH messages for a given event once. This can be achieved for example with a one bit flag in the LSP of the LSDB, stating whether convergence has been abandoned and signaled for this LSP. This can also be achieved by storing the identification of the LSPs for which convergence was abandoned for a time that is an order of magnitude longer than a typical IGP convergence (i.e., 10 seconds). The subset of neighbors to which an AAH message must be sent by a router R depends on the controlled convergence mechanism. It can be equal to all the neighbors of R, but not necessarily.

For any controlled convergence mechanism, the selection of this subset MUST be such that if a router R abandons controlled convergence, all the routers who could create a forwarding loop with R by not abandoning controlled convergence will eventually abandon

controlled convergence.

For the case of controlled convergence using ordered-FIB :

- o In the case of a link up / node up / metric decrease event, the set MUST include the neighbors of R that are on the shortest paths between R and the originator of the LSP for which controlled convergence is abandoned.
- o In the case of a link down / node down / metric increase event, the set MUST include the neighbors of R that are upstream of R on the paths towards the originator of the LSP for which controlled convergence is abandoned.

[3.3.](#) AAH messages

Like the others, this method uses a hold-down to acquire a set of LSPs which should be processed together. On expiry of the local hold-down timer, the router begins processing the batch of LSPs according to the loop free prevention algorithm. This is the same behaviour as the hold-down timer only method. However, if any router, having started the loop-free convergence process receives an LSP which would trigger a topology change, it locally abandons the controlled convergence process, and sends an AAH message to all its neighbors. This eventually triggers all routers to abandon the controlled convergence. The routers remain in AAH state (i.e.

processing topology changes using normal "fast" convergence), until a period of quiescence has elapsed. The exit from AAH state is synchronized by using a two step process.

To achieve the required synchronization, two additional messages are required, AAH and AAH ACK. The AAH message is reliably exchanged between neighbours using the AAH ACK message. These could be implemented as a new message within the routing protocol or carried in existing routing hello messages.

Two types of state machines are needed. A per-router AAH state machine and a per neighbour AAH state machine(PNSM). These are described below.

[3.3.1.](#) Per Router State Machine

Per Router State Table

EVENT	Q	Hold	CC	AAH	AAH-hold
RX LSP triggering change	Start hold-down timer [Hold]	-	TX-AAH Start AAH timer. [AAH]	Re-start AAH timer. [AAH]	TX-AAH Start AAH timer. [AAH]
RX AAH (Neighbor's PNSM processes RX AAH.)	TX-AAH Start AAH timer. [AAH]	TX-AAH Start AAH timer. [AAH]	TX-AAH Start AAH timer. [AAH]	[AAH]	TX-AAH Start AAH timer. [AAH]
Timer expiry	-	Trigger CC. [CC]	-	Start AAH-hold timer. [AAH-hold]	[Q]
Controlled convergence completed	-	-	[Q]	-	-

TX-AAH = Send "goto TX-AAH" to all other PNSMs.

Operation of the per-router state machine is as follows:

Operation of this state machine under normal topology change involves only states: Quiescent (Q), Hold-down (Hold) and Controlled Convergence (CC). The remaining states are associated with an AAH

event.

The resting state is Quiescent. When the router in the Quiescent state receives an LSP indicating a topology change, which would normally trigger an SPF, it starts the Hold-down timer and changes state to Hold-down. It normally remains in this state, collecting additional LSPs until the Hold-down timer expires. Note that all routers MUST use a common value for the Hold-down timer. When the

Hold-down timer expires the router then enters Controlled Convergence (CC) state and executes the CC mechanism to re-converge the topology. When the CC process has completed on the router, the router re-enters the Quiescent state.

If this router receives a topology changing LSP whilst it is in the CC state, it enters AAH state, and sends a "goto TX-AAH" command to all per neighbour state machines which causes each per-neighbour state machine to signal this state change to its neighbour. Alternatively, if this router receives an AAH message from any of its neighbors whilst in any state except AAH, it starts the AAH timer and enters the AAH state. The per neighbor state machine corresponding to the neighbor from which the AAH was received executes the RX AAH action (which causes it to send an AAH ACK), while the remainder are sent the "goto TX-AAH" command. The result is that the AAH is acknowledged to the neighbor from which it was received and propagated to all other neighbors. On entering AAH state, all CC timers are expired and normal convergence takes place.

Whilst in the AAH state, LSPs are processed in the traditional manner. Each time an LSP is received, the AAH timer is restarted. In an unstable network ALL routers will remain in this state for some time and the network will behave in the traditional uncontrolled convergence manner.

When the AAH timer expires, the router enters AAH-hold state and starts the AAH hold timer. The purpose of the AAH-hold state is to synchronize the transition of the network from AAH to Quiescent. The additional state ensures that the network cannot contain a mixture of routers in both AAH and Quiescent states. If, whilst in AAH-Hold state the router receives a topology changing LSP, it re-enters AAH state and commands all per neighbour state machines to "goto TX-AAH". If, whilst in AAH-Hold state the router receives an AAH message from one of its neighbours, it re-enters the AAH state and commands all other per neighbour state machines to "goto TX-AAH". Note that the per-neighbor state machine receiving the AAH message will autonomously acknowledge receipt of the AAH message. Commanding the per-neighbour state machine to "goto TX-AAH" is necessary, because routers may be in a mixture of Quiescent, Hold-down and AAH-hold state, and it is necessary to rendezvous the entire network back to

When the AAH Hold timer expires the router changes to state Quiescent and is ready for loop free convergence.

3.3.2. Per Neighbor State Machine

Per Neighbor State Table

EVENT	Idle	TX-AAH
RX AAH	Send ACK. [IDLE]	Send ACK. Cancel timer. [IDLE]
RX ACK	ignore	Cancel timer. [IDLE]
RX "goto TX-AAH" from Router State Machine	Send AAH [TX-AAH]	ignore
Timer expires	impossible	Send AAH Restart timer. [TX-AAH]

There is one instance of the per-neighbour (PN) state machine for each neighbour within the convergence control domain.

The normal state is IDLE.

On command ("goto TX-AAH") from the router state machine, the state machine enters TX-AAH state, transmits an AAH message to its neighbour and starts a timer.

On receipt of an AAH ACK in state TX-AAH the state machine cancels the timer and enters IDLE state.

In states IDLE, any AAH ACK message received is ignored.

On expiry of the timer in state TX-AAH the state machine transmits an AAH message to the neighbour and restarts the timer. (The timer cannot expire in any other state.)

In any state, receipt of an AAH causes the state machine to transmit an AAH ACK and enter the IDLE state.

Note that for correct operation the state machine MUST remain in

state TX-AAH, until an AAH ACK or an AAH is received, or the state machine is deleted. Deletion of the per neighbor state machine occurs when routing determines that the neighbour has gone away, or when the interface goes away.

When routing detects a new neighbour it creates a new instance of the per-neighbour state machine in state Idle. The consequent generation of the router's own LSP will then cause the router state machine to execute the LSP receipt actions, which will if necessary result in the new per-neighbour state machine receiving a "goto TX-AAH" command and transitioning to TX-AAH state.

[4.](#) Management Considerations

The management requirements will depend upon the solution adopted, but at the very least there needs to be reporting of the current state.

[5.](#) Scope and applicability

The initial scope of this work is in the context of link state IGPs.

[6.](#) IANA Considerations

There are no IANA considerations that arise from this document.

[7.](#) Security Considerations

This document does not itself introduce any security issues, but attention must be paid to the security implications of any proposed solutions to the problem.

[8.](#) Acknowledgements

The authors would like to acknowledge contributions made by Les Ginsberg.

[9.](#) References

Internet-Draft

Abandon All Hope (AAH)

October 2008

[9.1.](#) Normative References

- [1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

[9.2.](#) Informative References

- [2] Shand, M. and S. Bryant, "IP Fast Reroute Framework", [draft-ietf-rtgwg-ipfrr-framework-09](#) (work in progress), October 2008.
- [3] Shand, M. and S. Bryant, "A Framework for Loop-free Convergence", [draft-ietf-rtgwg-lf-conv-frmwk-02](#) (work in progress), February 2008.
- [4] Francois, P., "Loop-free convergence using oFIB", [draft-ietf-rtgwg-ordered-fib-02](#) (work in progress), February 2008.
- [5] Zinin, A., "Analysis and Minimization of Microloops in Link-state Routing Protocols", [draft-ietf-rtgwg-microloop-analysis-01](#) (work in progress), October 2005.

Authors' Addresses

Mike Shand
Cisco Systems
250, Longwater Avenue.
Reading, Berks RG2 6GB
UK

Email: mshand@cisco.com

Stewart Bryant
Cisco Systems
250, Longwater Avenue.
Reading, Berks RG2 6GB

UK

Email: stbryant@cisco.com

Shand, et al.

Expires May 3, 2009

[Page 10]

Internet-Draft

Abandon All Hope (AAH)

October 2008

Pierre Francois
Universite catholique de Louvain

Email: pierre.francois@uclouvain.be
URI: <http://inl.info.ucl.ac.be/pfr>

Full Copyright Statement

Copyright (C) The IETF Trust (2008).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information

on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.