

Restart signaling for ISIS
draft-shand-isis-restart-00.txt

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#) [1].

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>.

1. Abstract

The IS-IS routing protocol ([RFC 1142](#) [2], ISO/IEC 10589 [3]) is a link state intra-domain routing protocol. Normally, when an IS-IS router is re-started, the neighboring routers detect the restart event and cycle their adjacencies with the restarting router through the down state. This is necessary in order to invoke the protocol mechanisms to ensure correct re-synchronization of the LSP database. However, the cycling of the adjacency state causes the neighbors to regenerate their LSPs describing the adjacency concerned. This in turn causes temporary disruption of routes passing through the restarting router.

In certain scenarios such temporary disruption of the routes is highly undesirable.

This draft describes a mechanism for a restarting router to signal that it is restarting to its neighbors, and allow them to re-

establish their adjacencies without cycling through the down state, while still correctly initiating database synchronization.

When such a router is restarted, it is highly desirable that it does not re-compute its own routes until it has achieved database synchronization with its neighbors. Re-computing its routes before synchronization is achieved will result in its own routes being temporarily incorrect.

This draft additionally describes a mechanism for a restarting router to determine when it has achieved synchronization with its neighbors.

[2.](#) 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](#) [4].

[3.](#) Overview

There are two related problems with the existing specification of IS-IS with regard to re-synchronization of LSP databases when a router is re-started.

Firstly, when a routing process restarts, and an adjacency to a neighboring router is re-initialized the neighboring routing process does three things

1. It re-initializes the adjacency and causes its own LSP(s) to be regenerated, thus triggering SPF runs throughout the domain.
2. It sets SRMflags on its own LSP database on the adjacency concerned.
3. In the case of a Point-to-Point link it transmits a (set of) CSNP(s) over the adjacency.

In the case of a restarting router process, the first of these is highly undesirable, but the second is essential in order to ensure re-synchronization of the LSP database.

Secondly, whether or not the router is being re-started, it is desirable to be able to determine when the LSP databases of the neighboring routers have been synchronized (so that the overload bit can be cleared in the router's own LSP, for example). This document describes modifications to achieve this.

It is assumed that the three-way handshake [5] is being used on Point-to-Point circuits.

Shand

Expires Aug 2001

[Page 2]

INTERNET DRAFT

IS-IS restart

Feb 2001

[4. Approach](#)

[4.1 Adjacency re-acquisition](#)

Adjacency re-acquisition is the first step in re-initialization. The restarting router explicitly notifies its neighbor that the adjacency is being re-acquired, and hence that it should not re-initialize the adjacency. This is achieved by the inclusion of a new "re-start" option (TLV) in the IIH PDU. The presence of this TLV indicates that the sender supports the new restart capability and it carries flags that are used to convey information during a restart. All IIHs transmitted by a router that supports this capability **MUST** include this TLV.

Type [TBD]

Length 1

Value (1 octet)

Bit 1 - Restart Request (RR)

Bit 2 - Restart Acknowledgment (RA)

Bits 3-8 - Reserved

On receipt of an IIH with the "re-start" TLV having the RR bit set, if there exists on this interface an adjacency in state "Up" with the same System ID, and in the case of a LAN circuit, with the same source LAN address, then, irrespective of the other contents of the "Intermediate System Neighbors" option (LAN circuits), or the "Point-to-Point Adjacency State" option (Point-to-Point circuits):-

- a) Refresh the timer on the adjacency and leave the adjacency in state "Up",
- b) immediately (i.e. without waiting for any currently running timer

interval to expire, but with a small random delay of a few 10s of milliseconds on LANs to avoid "storms"), transmit over the corresponding interface an IIH including the "re-start" TLV with the RR bit clear and the RA bit set, having updated the "Point-to-Point Adjacency State" option to reflect any new values received from the re-starting router. (This allows the restarting router to quickly acquire the correct information to place in its hellos.),

- c) if the corresponding interface is a Point-to-Point interface, or
 - i) if the receiving router is the LAN level n designated router (where n is the level of the IIH), or
 - ii) if the transmitting router is currently (according to the receiving router) the LAN level n designated router (where n is the level of the IIH) and the receiving router would be elected the LAN level n designated router if the transmitting router were ignored (note the actual DR is NOT changed by this process),

initiate the transmission over the corresponding interface of a complete set of CSNPs, and set SRMflags on the corresponding interface for all LSPs in the local LSP database.

Otherwise (i.e. if there was no adjacency to the system ID in question), process the IIH as normal by re-initializing the adjacency, and setting the RA bit in the returned IIH.

A router that does not support the re-start capability will ignore the "re-start" TLV and re-initialize the adjacency as normal, returning an IIH without the "re-start" TLV.

On starting, a router starts a timer T1 and transmits an IIH containing the "re-start" TLV with the RR bit set.

1. On a LAN circuit the IIH contains an empty "Intermediate Systems Neighbors" TLV.
2. On a Point-to-Point circuit the IIH contains a "Point-to-Point Adjacency State" option with state "Initializing", and with empty "Neighbor System ID" and "Neighbor Extended Local Circuit ID" options. The values of the "LocalCircuitID" and the "Extended Local CircuitID" may, but need not be, the same

as those used previously for this circuit.

Transmission of "normal" IIHs is inhibited until the conditions described below are met (in order to avoid causing an unnecessary adjacency re-initialization). On expiry of the timer T1, it is restarted and the IIH is retransmitted as above.

On receipt of an IIH by the restarting router, a local adjacency is established as usual, and if the IIH contains a "re-start" TLV with the RA bit set, the receipt of the acknowledgement over that interface is noted.

Receipt of an IIH not containing the "re-start" option is also treated as an acknowledgement, since it indicates that the neighbor is not re-start capable. In this case the neighbor will have re-initialized the adjacency as normal, which in the case of a Point-to-Point link will guarantee that SRMflags have been set on its database. In the case of LAN interface, the usual operation of the update process will ensure that synchronization is eventually achieved.

In the case of a Point-to-Point circuit, the "LocalCircuitID" and "Extended Local Circuit ID" information contained in the IIH can be used immediately to generate an IIH containing the correct 3-way handshake information. The presence of "Neighbor System ID" or "Neighbor Extended Local Circuit ID" information which does not match the values currently in use by the local system is ignored (since the IIH may have been transmitted before the neighbor had received the new values from the re-starting router), but the

adjacency remains in the initializing state until the correct information is received.

In the case of a LAN circuit the information in the Intermediate Systems Neighbors option is recorded and used for the generation of subsequent IIHs as normal.

When BOTH a complete set of CSNP(s) and an acknowledgement have been received over the interface, the timer T1 is cancelled.

Once T1 has been cancelled, subsequent IIHs are transmitted according to the normal algorithms, but including the "re-start" TLV with both RR and RA clear.

If a LAN contains a mixture of systems, only some of which support the new algorithm, database synchronization is still guaranteed, but the "old" systems will have re-initialized their adjacencies.

If an interface is active, but does not have any neighboring router reachable over that interface the timer T1 would never be cancelled, and according to clause 4.2.1.2 the SPF would never be run. Therefore timer T1 is cancelled after some pre-determined expirations. (By this time any existing adjacency on a remote system would probably have expired anyway.)

4.1.1 State Table

The above operations can be summarized by the following state table.

Event	Running	Restarting	Seen RA	Seen CSNP
RX RR	Set SRM	Set SRM	Set SRM	Set SRM
	Send RA	Send RA	Send RA	Send RA
	Send CSNP	Send CSNP	Send CSNP	Send CSNP

RX RA		Goto Seen RA		Cancel T1 Goto Running
RX CSNP		Goto Seen CSNP	Cancel T1 Goto Running	
RX IIH with no Reset TLV		Cancel T1 Goto Running	Cancel T1 Goto Running	Cancel T1 Goto Running
T1 Expires		Send RR Send CSNP Start T1	Send RR Send CSNP Start T1	Send RR Send CSNP Start T1
T1 Expires n times		Cancel T1 Goto Running	Cancel T1 Goto Running	Cancel T1 Goto Running
Router Restarted	Set SRM Send RR Send CSNP Start T1 Goto Restarting	Set SRM Send RR Send CSNP Start T1 Goto Restarting	Set SRM Send RR Send CSNP Start T1 Goto Restarting	Set SRM Send RR Send CSNP Start T1 Goto Restarting

4.2 Database synchronization

When a router is started or re-started it can expect to receive a (set of) CSNP(s) over each interface. The arrival of the CSNP(s) is now guaranteed, since the "re-start" IIH with the RR bit set will be retransmitted until the CSNP(s) are correctly received.

The CSNPs describe the set of LSPs that are currently held by each neighbor. Synchronization will be complete when all these LSPs have been received.

On starting, a router starts a timer T2 for each active level. In addition to normal processing of the CSNPs, the set of LSPIDs contained in the first complete set of CSNP(s) received over each interface is recorded. If there are multiple interfaces on the restarting router, the recorded set of LSPIDs is the union of those received over each interface. LSPs with a remaining lifetime of zero are NOT so recorded.

As LSPs are received (by the normal operation of the update process) over any interface, the corresponding LSPID entry is removed (it is also removed if the LSP had arrived before the CSNP containing the reference). When the list of LSPIDs becomes empty, the timer T2 is cancelled.

At this point the local database is guaranteed to contain all the LSP(s) (either the same sequence number, or a more recent sequence number) which were present in the neighbors' databases at the time of re-starting. LSPs that arrived in a neighbor's database after the time of re-starting may, or may not, be present, but the normal operation of the update process will guarantee that they will eventually be received. At this point the local database is deemed to be "synchronized".

Since LSPs mentioned in the CSNP(s) with a zero remaining lifetime are not recorded, it is unlikely that cancellation of the timer T2 will be prevented by waiting for an LSP which will never arrive. However it is possible that an LSP with a small remaining lifetime was placed in the list. If the re-synchronization process takes longer than 'ZeroAgeLifetime' (default 60 seconds), the corresponding LSP will have been removed from the neighbors' LSP databases and will never arrive. Under these circumstances, the timer T2 may expire, and the databases are deemed to be synchronized.

[4.2.1](#) LSP generation and flooding and SPF computation

The operation of a router starting, as opposed to re-starting is somewhat different. These two cases are dealt with separately below.

[4.2.1.1](#). Starting for the first time

In the case of a starting router, the router's own zeroth LSP is first transmitted with the overload bit set. This prevents other routers from computing routes through the router until it has reliably acquired the complete set of LSPs. The overload bit remains set in subsequent transmissions of the zeroth LSP (such as will occur if a previous copy of the routers LSP is still present in the network) while the timer T2 is running.

When the timer T2 expires, or is cancelled, the own LSP is regenerated with the overload bit clear (assuming the router isn't in fact overloaded), and flooded as normal.

Other 'own' LSPs (including pseudonodes) are generated and flooded as normal, irrespective of the timer T2. The SPF is also run as

normal and the RIB and FIB updated as routes become available.

[4.2.1.2](#). Re-starting

In order to avoid causing unnecessary routing churn in other routers, it is highly desirable that the own LSPs generated by the restarting system are the same as those previously present in the network (assuming no other changes have taken place). It is important therefore not to regenerate and flood the LSPs until all the adjacencies have been re-established and any information required for propagation into the local LSPs is fully available. Ideally, the information should be loaded into the LSPs in a deterministic way, such that the same information occurs in the same place in the same LSP (and hence the LSPs are identical to their previous versions). If this can be achieved, the new versions will not even cause SPF to be run in other systems. However, provided the same information is included in the set of LSPs (albeit in a different order, and possibly different LSPs), the result of running the SPF will be the same and will not cause churn to the forwarding tables.

In the case of a re-starting router, none of the router's own non-pseudonode LSPs are transmitted, nor is the SPF run to update the forwarding tables while the timer T2 is running, or while the timer T1 is still running on any of the interfaces.

Redistribution of inter-level information must be regenerated before this router's LSP is flooded to other nodes. Therefore the level-n non-pseudonode LSP(s) should not be flooded until the other level's T2 timer has expired and its SPF has been run. This ensures that any inter-level information that should be propagated can be included in the level-n LSP(s).

During this period, if one of the router's own (including pseudonodes) LSPs is received, which the local router does not currently have in its own database, it is NOT purged. Under normal operation, such an LSP would be purged, since the LSP clearly should not be present in the global LSP database. However, in the present circumstances, this would be highly undesirable, because it could cause premature removal of an own LSP -- and hence churn in remote

routers. Even if the local system has one or more own LSPs (which is has generated, but not yet transmitted) it is still not valid to compare the received LSP against this set, since it may be that as a result of propagation between level 1 and level 2 (or vice versa) a further own LSP will need to be generated when the LSP databases have synchronized.

When the timer T2 expires, or is cancelled, the SPF is run to update the RIB and FIB.

Once the other level's SPF has run and any inter-level propagation has been resolved, the 'own' LSPs can be generated and flooded. Any 'own' LSPs which were previously ignored, but which are not part of the current set of 'own' LSPs (including pseudonodes) should then be

Shand

Expires Aug 2001

[Page 8]

INTERNET DRAFT

IS-IS restart

Feb 2001

purged. Note that it is possible that a Designated Router change may have taken place, and consequently the router should purge those pseudonode LSPs which it previously owned, but which are now no longer part of its set of pseudonode LSPs.

5. Security Considerations

This memo does not create any new security issues for the IS-IS protocol. Security considerations for the base IS-IS protocol are covered in [2] and [3].

6. References

- 1 Bradner, S., "The Internet Standards Process -- Revision 3", [BCP 9](#), [RFC 2026](#), October 1996.
- 2 Callon, R., "OSI IS-IS for IP and Dual Environment," [RFC 1195](#), December 1990.
- 3 ISO, "Intermediate system to Intermediate system routing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473)," ISO/IEC 10589:1992.
- 4 Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997
- 5 Katz, D., "Three-Way Handshake for IS-IS Point-to-Point

Adjacencies", [draft-ietf-isis-3way-03.txt](#), July 2000

7. Acknowledgments

The author would like to acknowledge contributions made by Radia Perlman, Mark Schaefer, Russ White and Rena Yang.

8. Author's Addresses

Mike Shand
Cisco Systems
4, The Square,
Stockley Park,
UXBRIDGE,
Middlesex
UB11 1BN, UK

Phone: +44 20 8756 8690
Email: mshand@cisco.com