ï»¿

Restart Signaling for IS-IS

Abstract

   This document describes a mechanism for a restarting router to signal
   to its neighbors that it is restarting, allowing them to reestablish
   their adjacencies without cycling through the DOWN state while still
   correctly initiating database synchronization.

   This document additionally describes a mechanism for a router to
   signal its neighbors that it is preparing to initiate a restart while
   maintaining forwarding-plane state. This allows the neighbors to
   maintain their adjacencies until the router has restarted but also
   allows the neighbors to bring the adjacencies down in the event of
   other topology changes.

   This document additionally describes a mechanism for a restarting
   router to determine when it has achieved Link State Protocol Data
   Unit (LSP) database synchronization with its neighbors and a
   mechanism to optimize LSP database synchronization while minimizing
   transient routing disruption when a router starts.

   This document obsoletes RFC 5306.

Status of This Memo

   This is an Internet Standards Track document.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Further information on
   Internet Standards is available in Section 2 of RFC 7841.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   https://www.rfc-editor.org/info/rfc8706.

Copyright Notice

Table of Contents

1. Overview

   The Intermediate System to Intermediate System (IS-IS) routing
   protocol [RFC1195] [ISO10589] is a link state intra-domain routing
   protocol. Normally, when an IS-IS router is restarted, temporary
   disruption of routing occurs due to events in both the restarting
   router and the neighbors of the restarting router.

   The router that has been restarted computes its own routes before
   achieving database synchronization with its neighbors. The results
   of this computation are likely to be non-convergent with the routes
   computed by other routers in the area/domain.

   Neighbors of the restarting router detect the restart event and cycle
   their adjacencies with the restarting router through the DOWN state.
   The cycling of the adjacency state causes the neighbors to regenerate
   their LSPs describing the adjacency concerned. This in turn causes a
   temporary disruption of routes passing through the restarting router.

   In certain scenarios, the temporary disruption of the routes is
   highly undesirable. This document describes mechanisms to avoid or
   minimize the disruption due to both of these causes.

   When an adjacency is reinitialized as a result of a neighbor
   restarting, a router does three things:

   1. It causes its own LSP(s) to be regenerated, thus triggering
       Shortest Path First (SPF) runs throughout the area (or in the
       case of Level 2, 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 complete set
       of Complete Sequence Number PDUs (CSNPs), 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
   synchronization of the LSP database.

   The third action above minimizes the number of LSPs that must be
   exchanged and, if made reliable, provides a means of determining when
   the LSP databases of the neighboring routers have been synchronized.
   This is desirable whether or not the router is being restarted (so
   that the overload bit can be cleared in the router's own LSP, for
   example).

   This document describes a mechanism for a restarting router to signal
   to its neighbors that it is restarting. The mechanism further allows

the neighbors to reestablish their adjacencies with the restarting
router without cycling through the DOWN state while still correctly
initiating database synchronization.

This document additionally describes a mechanism for a restarting
router to determine when it has achieved LSP database synchronization
with its neighbors and a mechanism to optimize LSP database
synchronization and minimize transient routing disruption when a
router starts.

It is assumed that the three-way handshake [RFC5303] is being used on
Point-to-Point circuits.

2. Conventions Used in This Document

If the control and forwarding functions in a router can be maintained
independently, it is possible for the forwarding function state to be
maintained across a resumption of control function operations. This
functionality is assumed when the terms "restart/restarting" are used
in this document.

The terms "start/starting" are used to refer to a router in which the
control function has either commenced operations for the first time
or has resumed operations, but the forwarding functions have not been
maintained in a prior state.

The terms "(re)start/(re)starting" are used when the text is
applicable to both a "starting" and a "restarting" router.

The terms "normal IIH" or "IIH normal" refer to IS-IS Hellos (IIHs)
in which the Restart TLV (defined later in this document) has no
flags set.

2.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and
"OPTIONAL" in this document are to be interpreted as described in
BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all
capitals, as shown here.

3. Approach

3.1. Timers

Three additional timers (T1, T2, and T3) are required to support the
mechanisms defined in this document. Timers T1 and T2 are used both
by a restarting router and a starting router. Timer T3 is used only
by a restarting router.

NOTE: These timers are NOT applicable to a router that is preparing
to do a planned restart.

An instance of the timer T1 is maintained per interface and indicates
the time after which an unacknowledged (re)start attempt will be
repeated. A typical value is 3 seconds.

An instance of the timer T2 is maintained for each LSP database
(LSPDB) present in the system. For example, for a Level 1/2 system,
there will be an instance of the timer T2 for Level 1 and an instance
for Level 2. This is the maximum time that the system will wait for
LSPDB synchronization. A typical value is 60 seconds.

A single instance of the timer T3 is maintained for the entire
system. It indicates the time after which the router will declare
that it has failed to achieve database synchronization (by setting
the overload bit in its own LSP). This is initialized to 65535
seconds but is set to the minimum of the remaining times of received
IIHs containing a Restart TLV with the Restart Acknowledgement (RA)
set and an indication that the neighbor has an adjacency in the UP
state to the restarting router. (See item a in Section 3.2.1.)

3.2. Restart TLV

A new TLV is defined to be included in IIH PDUs. The TLV includes
flags that are used to convey information during a (re)start. The
absence of this TLV indicates that the sender supports none of the
functionality defined in this document. Therefore, if a router
supports any of the functionality defined in this document it MUST
include this TLV in all transmitted IIHs.

Type:
   211

Length:
   Number of octets in the Value field (1 to (3 + ID Length))

Value:

```
                                     No. of octets
         +------------------------+
         |    Flags               |    1
         +------------------------+
         | Remaining Time         |    2
         +------------------------+
         | Restarting Neighbor ID |    ID Length
         +------------------------+
```

Flags (1 octet)

```
          0  1  2  3  4  5  6  7
         +--+--+--+--+--+--+--+--+
         |Reserved|PA|PR|SA|RA|RR|
         +--+--+--+--+--+--+--+--+
```

   RR - Restart Request
   RA - Restart Acknowledgement
   SA - Suppress adjacency advertisement
   PR - Restart is planned
   PA - Planned restart acknowledgement

Remaining Time (2 octets)
   Remaining Holding Time (in seconds).

   Required when the RA, PR, or PA bit is set. Otherwise, this
   field SHOULD be omitted when sent and MUST be ignored when
   received.

Restarting Neighbor System ID (ID Length octets)
   The System ID of the neighbor to which an RA/PA refers.

   Required when the RA or PA bit is set. Otherwise, this field
   SHOULD be omitted when sent and MUST be ignored when received.

   Note: Very early draft versions of the restart functionality
   did not include the Restarting Neighbor System ID in the TLV.
   RFC 5306 allowed for the possibility of interoperating with
   legacy implementations by stating that a router that is
   expecting an RA on a LAN circuit should assume that the
   acknowledgement is directed at the local system if the TLV is
   received with RA set and Restarting Neighbor System ID is not
   present. It is an implementation choice whether to continue to
   accept (on a LAN) a TLV with RA set and Restarting Neighbor
   System ID absent. Note that the omission of the Restarting
   Neighbor System ID only introduces ambiguity in the case where
   there are multiple systems on a LAN simultaneously performing
   restart.

The RR and SA flags may both be set in the TLV under the conditions
described in Section 3.3.2. All other combinations where multiple
flags are set are invalid and MUST NOT be transmitted. Received TLVs
that have invalid flag combinations set MUST be ignored.

3.2.1. Use of RR and RA Bits

The RR bit is used by a (re)starting router to signal to its
neighbors that a (re)start is in progress, that an existing adjacency
SHOULD be maintained even under circumstances when the normal
operation of the adjacency state machine would require the adjacency
to be reinitialized, to request a set of CSNPs, and to request
setting of the SRMflags.

The RA bit is sent by the neighbor of a (re)starting router to
acknowledge the receipt of a Restart TLV with the RR bit set.

When the neighbor of a (re)starting router receives an IIH with the
Restart TLV having the RR bit set, if there exists on this interface
an adjacency in the UP state 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 Three-Way Adjacency" option
(Point-to-Point circuits):

a. the state of the adjacency is not changed. If this is the first
    IIH with the RR bit set that this system has received associated
    with this adjacency, then the adjacency is marked as being in
    "Restart mode" and the adjacency Holding Time is refreshed --
    otherwise, the Holding Time is not refreshed. The Remaining Time
    transmitted according to (b) below MUST reflect the actual time
    after which the adjacency will now expire. Receipt of an IIH
    with the RR bit reset will clear the "Restart mode" state. This
    procedure allows the restarting router to cause the neighbor to
    maintain the adjacency long enough for restart to successfully
    complete while also preventing repetitive restarts from
    maintaining an adjacency indefinitely. Whether or not an
    adjacency is marked as being in "Restart mode" has no effect on
    adjacency state transitions.

b. immediately (i.e., without waiting for any currently running
    timer interval to expire but with a small random delay of a few
    tens of milliseconds on LANs to avoid "storms") transmit over the
    corresponding interface an IIH including the Restart TLV with the
    RR bit clear and the RA bit set, in the case of Point-to-Point
    adjacencies having updated the "Point-to-Point Three-Way
    Adjacency" option to reflect any new values received from the
    (re)starting router. (This allows a restarting router to quickly
    acquire the correct information to place in its hellos.) The
    Remaining Time MUST be set to the current time (in seconds)
    before the holding timer on this adjacency is due to expire. If
    the corresponding interface is a LAN interface, then the
    Restarting Neighbor System ID SHOULD be set to the System ID of
    the router from which the IIH with the RR bit set was received.
    This is required to correctly associate the acknowledgement and
    Holding Time in the case where multiple systems on a LAN restart
    at approximately the same time. This IIH SHOULD be transmitted
    before any LSPs or SNPs are transmitted as a result of the
    receipt of the original IIH.

c. if the corresponding interface is a Point-to-Point interface, or
    if the receiving router has the highest LnRouterPriority (with
    the highest source Media Access Control (MAC) address breaking
    ties) among those routers to which the receiving router has an
    adjacency in the UP state on this interface whose IIHs contain
    the Restart TLV, excluding adjacencies to all routers that are
    considered in "Restart mode" (note the actual Designated
    Intermediate System (DIS) 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 in the UP state to the
System ID in question), process the IIH as normal by reinitializing
the adjacency and setting the RA bit in the returned IIH.

3.2.2. Use of the SA Bit

The SA bit is used by a starting router to request that its neighbor
suppress advertisement of the adjacency to the starting router in the
neighbor's LSPs.

A router that is starting has no maintained forwarding function
state. This may or may not be the first time the router has started.
If this is not the first time the router has started, copies of LSPs
generated by this router in its previous incarnation may exist in the
LSP databases of other routers in the network. These copies are
likely to appear "newer" than LSPs initially generated by the
starting router due to the reinitialization of LSP fragment sequence
numbers by the starting router. This may cause temporary blackholes
to occur until the normal operation of the update process causes the
starting router to regenerate and flood copies of its own LSPs with
higher sequence numbers. The temporary blackholes can be avoided if
the starting router's neighbors suppress advertising an adjacency to
the starting router until the starting router has been able to
propagate newer versions of LSPs generated by previous incarnations.

When a router receives an IIH with the Restart TLV having the SA bit
set, if there exists on this interface an adjacency in the UP state
with the same System ID and, in the case of a LAN circuit, with the
same source LAN address, then the router MUST suppress advertisement
of the adjacency to the neighbor in its own LSPs. Until an IIH with
the SA bit clear has been received, the neighbor advertisement MUST
continue to be suppressed. If the adjacency transitions to the UP
state, the new adjacency MUST NOT be advertised until an IIH with the
SA bit clear has been received.

Note that a router that suppresses advertisement of an adjacency MUST
NOT use this adjacency when performing its SPF calculation. In
particular, if an implementation follows the example guidelines
presented in [ISO10589], Annex C.2.5, Step 0:b) "pre-load TENT with
the local adjacency database", the suppressed adjacency MUST NOT be
loaded into TENT.

3.2.3. Use of PR and PA Bits

The PR bit is used by a router that is planning to initiate a restart
to signal to its neighbors that it will be restarting. The router
sending an IIH with PR bit set SHOULD set the Remaining Time to a
value greater than the expected control-plane restart time. The PR
bit SHOULD remain set in IIHs until the restart is initiated.

The PA bit is sent by the neighbor of a router planning to restart to
acknowledge receipt of a Restart TLV with the PR bit set.

When the neighbor of a router planning a restart receives an IIH with
the Restart TLV having the PR bit set, if there exists on this
interface an adjacency in the UP state with the same System ID and,
in the case of a LAN circuit, with the same source LAN address, then:

a. if this is the first IIH with the PR bit set that this system has
    received associated with this adjacency, then the adjacency is
    marked as being in Planned Restart State and the adjacency
    Holding Time is refreshed -- otherwise, the Holding Time is not
    refreshed. The Holding Time SHOULD be set to the Remaining Time
    specified in the received IIH with PR set. The Remaining Time
    transmitted according to (b) below MUST reflect the actual time
    after which the adjacency will now expire. Receipt of an IIH
    with the PR bit reset will clear the Planned Restart State and
    cause the receiving router to set the adjacency Holding Time to
    the locally configured value. This procedure allows the router
    planning a restart to cause the neighbor to maintain the
    adjacency long enough for restart to successfully complete.
    Whether or not an adjacency is marked as being in Planned Restart
    State has no effect on adjacency state transitions.

b. immediately (i.e., without waiting for any currently running
    timer interval to expire, but with a small random delay of a few

tens of milliseconds on LANs to avoid "storms") transmit over the
corresponding interface an IIH including the Restart TLV with the
PR bit clear and the PA bit set. The Remaining Time MUST be set
to the current time (in seconds) before the holding timer on this
adjacency is due to expire. If the corresponding interface is a
LAN interface, then the Restarting Neighbor System ID SHOULD be
set to the System ID of the router from which the IIH with the PR
bit set was received. This is required to correctly associate
the acknowledgement and Holding Time in the case where multiple
systems on a LAN are planning a restart at approximately the same
time.

NOTE: Receipt of an IIH with PA bit set indicates to the router
planning a restart that the neighbor is aware of the planned restart
and -- in the absence of topology changes as described below -- will
maintain the adjacency for the Remaining Time included in the IIH
with PA set.

By definition, a restarting router maintains forwarding state across
the control-plane restart (see Section 2). But while a control-plane
restart is in progress, it is expected that the restarting router
will be unable to respond to topology changes. It is therefore
useful to signal a planned restart so that the neighbors of the
restarting router can determine whether it is safe to maintain the
adjacency if other topology changes occur prior to the completion of
the restart. Signaling a planned restart in the absence of
maintained forwarding-plane state is likely to lead to significant
traffic loss and MUST NOT be done.

Neighbors of the router that have signaled planned restart SHOULD
maintain the adjacency in a Planned Restart State until it receives
an IIH with the RR bit set, it receives an IIH with both PR and RR
bits clear, or the adjacency Holding Time expires -- whichever occurs
first. Neighbors that choose not to follow the recommended behavior
need to consider the impact on traffic delivery of not using the
restarting router for forwarding traffic during the restart period.

While the adjacency is in Planned Restart State, some or all of the
following actions MAY be taken:

a. If additional topology changes occur, the adjacency that is in
    Planned Restart State MAY be brought down even though the Holding
    Time has not yet expired. Given that the neighbor that has
    signaled a planned restart is not expected to update its
    forwarding plane in response to signaling of the topology changes
    (since it is restarting) traffic that transits that node is at
    risk of being improperly forwarded. On a LAN circuit, if the
    router in Planned Restart State is the DIS at any supported
    level, the adjacency or adjacencies SHOULD be brought down
    whenever any LSP update is either generated or received so as to
    trigger a new DIS election. Failure to do so will compromise the
    reliability of the update process on that circuit. What other
    criteria are used to determine what topology changes will trigger
    bringing the adjacency down is a local implementation decision.

b. If a Bidirectional Forwarding Detection (BFD) [RFC5880] Session
    to the neighbor that signals a planned restart is in the UP state
    and subsequently goes down, the event MAY be ignored since it is
    possible this is an expected side effect of the restart. Use of
    the Control-Plane Independent state as signaled in BFD control
    packets SHOULD be considered in the decision to ignore a BFD
    Session DOWN event.

c. On a Point-to-Point circuit, transmission of LSPs, CSNPs, and
    Partial Sequence Number PDU (PSNPs) MAY be suppressed. It is
    expected that the PDUs will not be received.

Use of the PR bit provides a means to safely support restart periods
that are significantly longer than standard Holding Times.

3.3. Adjacency (Re)Acquisition

Adjacency (re)acquisition is the first step in (re)initialization.
Restarting and starting routers will make use of the RR bit in the
Restart TLV, though each will use it at different stages of the
(re)start procedure.

3.3.1. Adjacency Reacquisition during Restart

The restarting router explicitly notifies its neighbor that the
adjacency is being reacquired and, hence, that it SHOULD NOT
reinitialize the adjacency. This is achieved by setting the RR bit
in the Restart TLV. When the neighbor of a restarting router
receives an IIH with the Restart TLV having the RR bit set, if there
exists on this interface an adjacency in the UP state with the same
System ID and, in the case of a LAN circuit, with the same source LAN
address, then the procedures described in Section 3.2.1 are followed.

A router that does not support the restart capability will ignore the
Restart TLV and reinitialize the adjacency as normal, returning an
IIH without the Restart TLV.

On restarting, a router initializes the timer T3, starts the timer T2
for each LSPDB, and for each interface (and in the case of a LAN
circuit, for each level) starts the timer T1 and transmits an IIH
containing the Restart TLV with the RR bit set.

On a Point-to-Point circuit, the restarting router SHOULD set the
"Adjacency Three-Way State" to "Init", because the receipt of the
acknowledging IIH (with RA set) MUST cause the adjacency to enter the
UP state immediately.

On a LAN circuit, the LAN-ID assigned to the circuit SHOULD be the
same as that used prior to the restart. In particular, for any
circuits for which the restarting router was previously DIS, the use
of a different LAN-ID would necessitate the generation of a new set
of pseudonode LSPs and corresponding changes in all the LSPs
referencing them from other routers on the LAN. By preserving the
LAN-ID across the restart, this churn can be prevented. To enable a
restarting router to learn the LAN-ID used prior to restart, the LAN-
ID specified in an IIH with RR set MUST be ignored.

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

When a restarting router receives an IIH a local adjacency is
established as usual, and if the IIH contains a Restart TLV with the
RA bit set (and on LAN circuits with a Restart Neighbor System ID
that matches that of the local system), the receipt of the
acknowledgement over that interface is noted. When the RA bit is set
and the state of the remote adjacency is UP, then the timer T3 is set
to the minimum of its current value and the value of the Remaining
Time field in the received IIH.

On a Point-to-Point link, receipt of an IIH not containing the
Restart TLV is also treated as an acknowledgement, since it indicates
that the neighbor is not restart capable. However, since no CSNP is
guaranteed to be received over this interface, the timer T1 is
canceled immediately without waiting for a complete set of CSNPs.
Synchronization may therefore be deemed complete even though there
are some LSPs that are held (only) by this neighbor (see
Section 3.4). In this case, we also want to be certain that the
neighbor will reinitialize the adjacency in order to guarantee that
the SRMflags have been set on its database, thus ensuring eventual
LSPDB synchronization. This is guaranteed to happen except in the
case where the Adjacency Three-Way State in the received IIH is UP
and the Neighbor Extended Local Circuit ID matches the Extended Local
Circuit ID assigned by the restarting router. In this case, the
restarting router MUST force the adjacency to reinitialize by setting
the local Adjacency Three-Way State to DOWN and sending a normal IIH.

In the case of a LAN interface, receipt of an IIH not containing the
Restart TLV is unremarkable since synchronization can still occur so
long as at least one of the non-restarting neighboring routers on the
LAN supports restart. Therefore, T1 continues to run in this case.
If none of the neighbors on the LAN are restart capable, T1 will
eventually expire after the locally defined number of retries.

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 three-way
handshake information. The presence of Neighbor Extended Local
Circuit ID information that does not match the value currently in use
by the local system is ignored (since the IIH may have been
transmitted before the neighbor had received the new value from the
restarting router), but the adjacency remains in the initializing
state until the correct information is received.

In the case of a LAN circuit, the source neighbor information (e.g.,
SNPAAddress) is recorded and used for adjacency establishment and
maintenance as normal.

When BOTH a complete set of CSNPs (for each active level, in the case
of a Point-to-Point circuit) and an acknowledgement have been
received over the interface, the timer T1 is canceled.

Once the timer T1 has been canceled, subsequent IIHs are transmitted
according to the normal algorithms but including the Restart 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 reinitialized their adjacencies.

If an interface is active but does not have any neighboring router
reachable over that interface, the timer T1 would never be canceled,
and according to Section 3.4.1.1, the SPF would never be run.
Therefore, timer T1 is canceled after some predetermined number of
expirations (which MAY be 1).

3.3.2. Adjacency Acquisition during Start

The starting router wants to ensure that in the event that a
neighboring router has an adjacency to the starting router in the UP
state (from a previous incarnation of the starting router), this
adjacency is reinitialized. The starting router also wants
neighboring routers to suppress advertisement of an adjacency to the
starting router until LSP database synchronization is achieved. This
is achieved by sending IIHs with the RR bit clear and the SA bit set
in the Restart TLV. The RR bit remains clear and the SA bit remains
set in subsequent transmissions of IIHs until the adjacency has
reached the UP state and the initial T1 timer interval (see below)
has expired.

Receipt of an IIH with the RR bit clear will result in the
neighboring router utilizing normal operation of the adjacency state
machine. This will ensure that any old adjacency on the neighboring
router will be reinitialized.

Upon receipt of an IIH with the SA bit set, the behavior described in
Section 3.2.2 is followed.

Upon starting, a router starts timer T2 for each LSPDB.

For each interface (and in the case of a LAN circuit, for each
level), when an adjacency reaches the UP state, the starting router
starts a timer T1 and transmits an IIH containing the restart TLV
with the RR bit clear and SA bit set. Upon expiry of the timer T1,
it is restarted and the IIH is retransmitted with both RR and SA bits
set (only the RR bit has changed state from earlier IIHs).

Upon receipt of an IIH with the RR bit set (regardless of whether or not the SA bit is set), the behavior described in Section 3.2.1 is followed.

When an IIH is received by the starting router and the IIH contains a Restart TLV with the RA bit set (and on LAN circuits with a Restart Neighbor System ID that matches that of the local system), the receipt of the acknowledgement over that interface is noted.

On a Point-to-Point link, receipt of an IIH not containing the Restart TLV is also treated as an acknowledgement, since it indicates that the neighbor is not restart capable. Since the neighbor will have reinitialized the adjacency, this guarantees that SRMflags have been set on its database, thus ensuring eventual LSPDB synchronization. However, since no CSNP is guaranteed to be received over this interface, the timer T1 is canceled immediately without waiting for a complete set of CSNPs. Synchronization may therefore be deemed complete even though there are some LSPs that are held (only) by this neighbor (see Section 3.4).

In the case of a LAN interface, receipt of an IIH not containing the Restart TLV is unremarkable since synchronization can still occur so long as at least one of the non-restarting neighboring routers on the LAN supports restart. Therefore, T1 continues to run in this case. If none of the neighbors on the LAN are restart capable, T1 will eventually expire after the locally defined number of retries. The usual operation of the update process will ensure that synchronization is eventually achieved.

When BOTH a complete set of CSNPs (for each active level, in the case of a Point-to-Point circuit) and an acknowledgement have been received over the interface, the timer T1 is canceled. Subsequent IIHs sent by the starting router have the RR and RA bits clear and the SA bit set in the Restart TLV.

Timer T1 is canceled after some predetermined number of expirations (which MAY be 1).

When the T2 timer(s) are canceled or expire, transmission of "normal IIHs" will begin.

3.3.3. Multiple Levels

A router that is operating as both a Level 1 and a Level 2 router on a particular interface MUST perform the above operations for each level.

On a LAN interface, it MUST send and receive both Level 1 and Level 2 IIHs and perform the CSNP synchronizations independently for each level.

On a Point-to-Point interface, only a single IIH (indicating support for both levels) is required, but it MUST perform the CSNP synchronizations independently for each level.

3.4. Database Synchronization

When a router is started or restarted, it can expect to receive a complete set of CSNPs over each interface. The arrival of the CSNP(s) is now guaranteed, since an 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.

When (re)starting, a router starts an instance of timer T2 for each LSPDB, as described in Section 3.3.1 or Section 3.3.2. In addition to normal processing of the CSNPs, the set of LSPIDs contained in the first complete set of CSNPs received over each interface is recorded, together with their remaining lifetime. In the case of a LAN

interface, a complete set of CSNPs MUST consist of CSNPs received
from neighbors that are not restarting. If there are multiple
interfaces on the (re)starting 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 an LSP arrives before the CSNP containing the
reference). When an LSPID has been held in the list for its
indicated remaining lifetime, it is removed from the list. When the
list of LSPIDs is empty and the timer T1 has been canceled for all
the interfaces that have an adjacency at this level, the timer T2 is
canceled.

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) that 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 and those with a short remaining lifetime are
deleted from the list when the lifetime expires, cancellation of the
timer T2 will not be prevented by waiting for an LSP that will never
arrive.

3.4.1. LSP Generation and Flooding and SPF Computation

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

3.4.1.1. Restarting

In order to avoid causing unnecessary routing churn in other routers,
it is highly desirable that the router's 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 reestablished and any information required
for propagation into the local LSPs is fully available. Ideally, the
information is 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 may 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 restarting router, none of the router's own LSPs are
transmitted, nor are the router's own forwarding tables updated while
the timer T3 is running.

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) MUST 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 is to 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 a router's own LSP -- and hence churn in

remote routers. Even if the local system has one or more of the
router's own LSPs (which it 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 router's own LSP will need to be
generated when the LSP databases have synchronized.

During this period, a restarting router SHOULD send CSNPs as it
normally would. Information about the router's own LSPs MAY be
included, but if it is included, it MUST be based on LSPs that have
been received, not on versions that have been generated (but not yet
transmitted). This restriction is necessary to prevent premature
removal of an LSP from the global LSP database.

When the timer T2 expires or is canceled, indicating that
synchronization for that level is complete, the SPF for that level is
run in order to derive any information that is required to be
propagated to another level, but the forwarding tables are not yet
updated.

Once the other level's SPF has run and any inter-level propagation
has been resolved, the router's own LSPs can be generated and
flooded. Any own LSPs that were previously ignored, but that are not
part of the current set of own LSPs (including pseudonodes), MUST
then be purged. Note that it is possible that a Designated Router
change may have taken place and, consequently, the router SHOULD
purge those pseudonode LSPs that it previously owned but that are now
no longer part of its set of pseudonode LSPs.

When all the T2 timers have expired or been canceled, the timer T3 is
canceled, and the local forwarding tables are updated.

If the timer T3 expires before all the T2 timers have expired or been
canceled, this indicates that the synchronization process is taking
longer than the minimum Holding Time of the neighbors. The router's
own LSP(s) for levels that have not yet completed their first SPF
computation are then flooded with the overload bit set to indicate
that the router's LSPDB is not yet synchronized (and therefore other
routers MUST NOT compute routes through this router). Normal
operation of the update process resumes, and the local forwarding
tables are updated. In order to prevent the neighbor's adjacencies
from expiring, IIHs with the normal interface value for the Holding
Time are transmitted over all interfaces with neither RR nor RA set
in the Restart TLV. This will cause the neighbors to refresh their
adjacencies. The router's own LSP(s) will continue to have the
overload bit set until timer T2 has expired or been canceled.

3.4.1.2. Starting

In the case of a starting router, as soon as each adjacency is
established, and before any CSNP exchanges, the router's own zeroth
LSP is 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 router's own zeroth LSP is still present in
the network) while any timer T2 is running.

When all the T2 timers have been canceled, the router's own LSP(s)
MAY be regenerated with the overload bit clear (assuming the router
is not in fact overloaded, and there is no other reason, such as
incomplete BGP convergence, to keep the overload bit set) and flooded
as normal.

Other LSPs owned by this router (including pseudonodes) are generated
and flooded as normal, irrespective of the timer T2. The SPF is also
run as normal and the Routing Information Base (RIB) and Forwarding
Information Base (FIB) updated as routes become available.

To avoid the possible formation of temporary blackholes, the starting
router sets the SA bit in the Restart TLV (as described in

Section 3.3.2) in all IIHs that it sends.

When all T2 timers have been canceled, the starting router MUST
transmit IIHs with the SA bit clear.

4. State Tables

This section presents state tables that summarize the behaviors
described in this document. Other behaviors, in particular adjacency
state transitions and LSP database update operations, are NOT
included in the state tables except where this document modifies the
behaviors described in [ISO10589] and [RFC5303].

The states named in the columns of the tables below are a mixture of
states that are specific to a single adjacency (ADJ suppressed, ADJ
Seen RA, ADJ Seen CSNP) and states that are indicative of the state
of the protocol instance (Running, Restarting, Starting, SPF Wait).

Three state tables are presented from the point of view of a running
router, a restarting router, and a starting router.

4.1. Running Router

Event	Running	ADJ suppressed
RX PR	Set Planned Restart State Update Holding Time Send PA	
RX PR clr and RR clr	Clear Planned Restart State Restore Holding Time to local value	
RX RR	Maintain ADJ State Send RA Set SRM, send CSNP (Note 1) Update Holding Time, set Restart Mode (Note 2)	
RX RR clr	Clr Restart mode	
RX SA	Suppress IS neighbor TLV in LSP(s) Goto ADJ Suppressed	
RX SA clr		Unsuppress IS neighbor TLV in LSP(s) Goto Running

                   Table 1: Running Router

Note 1: CSNPs are sent by routers in accordance with item c in
         Section 3.2.1

Note 2: If Restart Mode clear

4.2. Restarting Router

Event	Restarting	ADJ Seen RA	ADJ Seen CSNP	SPF Wait
Restart planned	Send PR			

Event	Restarting			
Planned restart canceled	Send PR clr			
RX PA	Proceed with planned restart			
Router restarts	Send IIH/RR ADJ Init Start T1, T2, T3			
RX RR	Send RA			
RX RA	Adjust T3 Goto ADJ Seen RA		Cancel T1 Adjust T3	
RX CSNP set	Goto ADJ Seen CSNP	Cancel T1		
RX IIH w/o Restart TLV	Cancel T1 (Point-to-point only)			
T1 expires	Send IIH/RR Restart T1	Send IIH/RR Restart T1	Send IIH/RR Restart T1	
T1 expires nth time	Send IIH/normal	Send IIH/ normal	Send IIH/ normal	
T2 expires	Trigger SPF Goto SPF Wait			
T3 expires	Set overload bit Flood local LSPs Update fwd plane			
LSP DB Sync	Cancel T2 and T3 Trigger SPF Goto SPF wait			
All SPF done				Clear overload bit Update fwd plane Flood local LSPs Goto Running

Table 2: Restarting Router

4.3. Starting Router

Event	Starting	ADJ Seen RA	ADJ Seen CSNP
Router starts	Send IIH/SA Start T1 and T2		

```
+------------+-------------------------+-----------+-----------+
| RX RR      | Send RA                 |           |           |
+------------+-------------------------+-----------+-----------+
| RX RA      | Goto ADJ Seen RA        |           | Cancel T1 |
+------------+-------------------------+-----------+-----------+
| RX CSNP Set| Goto ADJ Seen CSNP      | Cancel T1 |           |
+------------+-------------------------+-----------+-----------+
| RX IIH w no| Cancel T1 (Point-to-    |           |           |
| Restart TLV| Point only)             |           |           |
+------------+-------------------------+-----------+-----------+
| ADJ UP     | Start T1                |           |           |
|            | Send local LSPs with    |           |           |
|            | overload bit set        |           |           |
+------------+-------------------------+-----------+-----------+
| T1 expires | Send IIH/RR and SA      | Send IIH/ | Send IIH/ |
|            | Restart T1              | RR and SA | RR and SA |
|            |                         | Restart T1| Restart T1|
+------------+-------------------------+-----------+-----------+
| T1 expires | Send IIH/SA             | Send IIH/ | Send IIH/ |
| nth time   |                         | SA        | SA        |
+------------+-------------------------+-----------+-----------+
| T2 expires | Clear overload bit      |           |           |
|            | Send IIH normal         |           |           |
|            | Goto Running            |           |           |
+------------+-------------------------+-----------+-----------+
| LSP DB Sync| Cancel T2               |           |           |
|            | Clear overload bit      |           |           |
|            | Send IIH normal         |           |           |
+------------+-------------------------+-----------+-----------+
```

                    Table 3: Starting Router

5. IANA Considerations

    This document defines the following IS-IS TLV that is listed in the
    "IS-IS TLV Codepoints" registry.

```
+------+-------------+-----+-----+-----+-------+
| Type | Description | IIH | LSP | SNP | Purge |
+======+=============+=====+=====+=====+=======+
| 211  | Restart TLV | y   | n   | n   | n     |
+------+-------------+-----+-----+-----+-------+
```

                          Table 4

    IANA has updated the entry in registry to point to this document.

6. Security Considerations

    Any new security issues raised by the procedures in this document
    depend upon the ability of an attacker to inject a false but
    apparently valid IIH, the ease/difficulty of which has not been
    altered.

    If the RR bit is set in a false IIH, neighbors who receive such an
    IIH will continue to maintain an existing adjacency in the UP state
    and may (re)send a complete set of CSNPs. While the latter action is
    wasteful, neither action causes any disruption in correct protocol
    operation.

    If the RA bit is set in a false IIH, a (re)starting router that
    receives such an IIH may falsely believe that there is a neighbor on
    the corresponding interface that supports the procedures described in
    this document. In the absence of receipt of a complete set of CSNPs
    on that interface, this could delay the completion of (re)start
    procedures by requiring the timer T1 to time out the locally defined
    maximum number of retries. This behavior is the same as would occur
    on a LAN where none of the (re)starting router's neighbors support
    the procedures in this document and is covered in Sections 3.3.1 and
    3.3.2.

If the SA bit is set in a false IIH, this could cause suppression of the advertisement of an IS neighbor, which could either continue for an indefinite period or occur intermittently with the result being a possible loss of reachability to some destinations in the network and/or increased frequency of LSP flooding and SPF calculation.

If the PR bit is set in a false IIH, neighbors who receive such an IIH could modify the Holding Time of an existing adjacency inappropriately. In the event of topology changes, the neighbor might also choose to not flood the topology updates and/or bring the adjacency down in the false belief that the forwarding plane of the router identified as the source of the false IIH is not currently processing announced topology changes. This would result in unnecessary forwarding disruption.

If the PA bit is set in a false IIH, a router that receives such an IIH may falsely believe that the neighbor on the corresponding interface supports the planned restart procedures defined in this document. If such a router is planning to restart, it might then proceed to initiate a restart in the false expectation that the neighbor has updated its Holding Time as requested. This may result in the neighbor bringing down the adjacency while the receiving router is restarting, causing unnecessary disruption to forwarding.

The possibility of IS-IS PDU spoofing can be reduced by the use of authentication, as described in [RFC1195] and [ISO10589], and especially by the use of cryptographic authentication, as described in [RFC5304] and [RFC5310].

7. Manageability Considerations

These extensions that have been designed, developed, and deployed for many years do not have any new impact on management and operation of the IS-IS protocol via this standardization process.

8. Normative References

[ISO10589] International Organization for Standardization, "Information technology -- Telecommunications and information exchange between systems -- Intermediate System to Intermediate System intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service (ISO 8473)", ISO/IEC 10589:2002, Second Edition, November 2002.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, DOI 10.17487/RFC1195, December 1990, <https://www.rfc-editor.org/info/rfc1195>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC5303] Katz, D., Saluja, R., and D. Eastlake 3rd, "Three-Way Handshake for IS-IS Point-to-Point Adjacencies", RFC 5303, DOI 10.17487/RFC5303, October 2008, <https://www.rfc-editor.org/info/rfc5303>.

[RFC5304] Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", RFC 5304, DOI 10.17487/RFC5304, October 2008, <https://www.rfc-editor.org/info/rfc5304>.

[RFC5310] Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, DOI 10.17487/RFC5310, February 2009, <https://www.rfc-editor.org/info/rfc5310>.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010,

                <https://www.rfc-editor.org/info/rfc5880>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC
            2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174,
            May 2017, <https://www.rfc-editor.org/info/rfc8174>.

Appendix A. Summary of Changes from RFC 5306

   This document extends RFC 5306 by introducing support for signaling
   the neighbors of a restarting router that a planned restart is about
   to occur. This allows the neighbors to be aware of the state of the
   restarting router so that appropriate action may be taken if other
   topology changes occur while the planned restart is in progress.
   Since the forwarding plane of the restarting router is maintained
   based upon the pre-restart state of the network, additional topology
   changes introduce the possibility that traffic may be lost if paths
   via the restarting router continue to be used while the restart is in
   progress.

   In support of this new functionality, two new flags have been
   introduced:

   PR - Restart is planned

   PA - Planned restart acknowledgement

   No changes to the post-restart exchange between the restarting router
   and its neighbors have been introduced.

Acknowledgements

   For RFC 5306, the authors acknowledged contributions made by Jeff
   Parker, Radia Perlman, Mark Schaefer, Naiming Shen, Nischal Sheth,
   Russ White, and Rena Yang.

   The authors of this updated document acknowledge the contribution of
   Mike Shand, coauthor of RFC 5306.

Authors' Addresses

   Les Ginsberg
   Cisco Systems, Inc.

   Email: ginsberg@cisco.com

   Paul Wells
   Cisco Systems, Inc.

   Email: pauwells@cisco.com