MPLS WG
Internet Draft
Document: draft-farrel-mpls-ldp-ft-00.txt
Expiration Date: August 2000

A. Farrel
P. Brittain
Data Connection Ltd
February 2000

Fault Tolerance for LDP and CR-LDP

Status of this Memo

NOTE: The new TLV type numbers, bit values for flags specified in this draft, and new LDP status code values are preliminary suggested values and have yet to be approved by IANA or the MPLS WG. See the section "IANA Considerations" for further details.

Abstract

MPLS systems will be used in core networks where system downtime must be kept to an absolute minimum. Many MPLS LSRs may, therefore, exploit Fault Tolerant (FT) hardware or software to provide high-availability of the core networks.

The details of how FT is achieved for the various components of an FT LSR, including LDP, CR-LDP, the switching hardware and TCP, are implementation specific. This document identifies issues in the CR-LDP specification [2] and the LDP specification [4] that make it difficult to implement an FT LSR using the current LDP and CR-LDP protocols, and proposes enhancements to the LDP specification to ease such FT LSR implementations.

The extensions described here are equally applicable to CR-LDP.

Contents

1. Conventions and Terminology used in this document

Definitions of key words and terms applicable to LDP and CR-LDP are inherited from [2] and [4].

The term "FT label" is introduced in this document to indicated a label for which fault-tolerant operation is used. A "non-FT label" is not fault-tolerant and is handled as specified in [2] and [4].

The extensions to LDP specified in this document are collectively referred to as the "LDP FT enhancements".

In the examples quoted, the following notation is used.

Ln : An LSP. For example L1.
Pn : An LDP peer. For example P1.

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 [3].

2. Introduction

High Availability (HA) is typically claimed by equipment vendors when their hardware achieves availability levels of at least 99.999% (five 9s). To implement this, the equipment must be capable of recovering from local hardware and software failures through a process called Fault Tolerance (FT).

The usual approach to FT involves provisioning backup copies of hardware and software. When a primary copy fails, processing is switched to the backup copy. This process, called failover, should result in minimal disruption in both the Data and the Control Planes.

In an FT system, backup resources are sometimes provisioned on a one-to-one basis (1:1), sometimes as many-to-one (1:n), and occasionally as many-to-many (m:n). Whatever backup provisioning is made, the system must switch to the backup automatically on failure of the primary, and the software and hardware state in the backup must be set up to replicate the state in the primary at the point of failure.

FT systems are well placed to facilitate hot-swaps of hardware. A card can be simply removed from the system and replaced with a new one. The removal of a card is treated as hardware failure.

Similarly, an FT system can implement online software upgrades by

swapping from primary to backup under management control.

2.1 Fault-tolerance for MPLS

MPLS systems will be used in core networks where system downtime must be kept to an absolute minimum. Many MPLS LSRs may, therefore, exploit Fault Tolerant (FT) hardware or software to provide high-availability of core networks.

An FT MPLS system should be capable of failing over
- with minimal disruption to the data flow using established labels
- without loss of control information for established labels
- in a way that allows recovery of control information for labels that are being established or torn down.

It may be acceptable for some data to be lost, especially if failover involves swapping between two sets of switching hardware.

It is not acceptable for there to be a significant loss of service on any established label or LSP during failover. The target of at most 50ms disruption applied during discussions of LSP protection, should apply here too, for established LSPs.

It is, therefore, clearly unacceptable for established labels or LSPs transitting an LSR or a card within an LSR to be torn down during failover or upgrade processing.

Finally, it is not acceptable that resources (such as bandwidth allocation) should be lost during failover. This might arise if MPLS CR-LSP setup or tear-down are not completed correctly.

2.2 Issues with LDP and CR-LDP

LDP and CR-LDP use TCP to provide reliable connections between LSRs over which to exchange protocol messages to distribute labels and to set up LSPs. A pair of LSRs which have such a connection are referred to as LDP peers.

TCP enables LDP and CR-LDP to assume reliable transfer of protocol messages. This means that some of the messages do not need to be acknowledged (for example, Label Release).

LDP and CR-LDP are defined such that if the TCP connection fails, the LSR should immediately tear down the LSPs associated with the session between the LDP peers, and release any labels and resources assigned to those LSPs.

It is notoriously hard to provide a fault tolerant implementation of TCP. To do so might involve making copies of all data sent and received. This is an issue familiar to implementers of other TCP applications such as BGP.

During failover affecting the TCP or LDP stacks, therefore, the TCP connection may be lost. Recovery from this position is made worse by the fact that LDP or CR-LDP control messages may have been lost during the connection failure. Since these messages are unconfirmed, it is possible that LSP or label state information will be lost.

This draft describes a solution which involves
- negotiation between LDP peers of the intent to support extensions to LDP that facilitate recovery from failover without loss of LSPs
- selection of FT survival on a per LSP/label basis
- acknowledgement messages to ensure that a full handshake is performed on label distribution and LSP setup/teardown of FT labels
- sending duplicate messages after failover to ensure that LSP/label state is correctly reflected at the peer for FT labels.

Other objectives of this draft are to
- offer back-compatability with LSRs that do not implement these proposals
- preserve existing protocol rules described in [2] and [4] for handling unexpected duplicate messages and for processing unexpected messages referring to unknown LSPs/labels
- integrate with the LSP modification function described in [5]
- avoid full state refresh solutions (such as those present in RSVP: see [6], [7] and [8]) whether they be full-time, or limited to post-failover recovery.

This draft does not attempt to describe how to modify the routing of an LSP or the resources allocated to a label or LSP, which is covered by [5].

This draft also does not address how to provide automatic layer 2/3 protection switching for a label or LSP, which is a separate area for study.

2.3 Data Forwarding During TCP Connection Failure

An LSR that implements the LDP FT enhancements SHOULD preserve the programming of the switching hardware across a failover. This ensures that data forwarding is unaffected by the state of the TCP connection between LSRs.

It is an integral part of FT failover processing in some hardware configurations that some data packets might be lost. If data loss is not acceptable to the applications using the MPLS network, the LDP FT enhancements described in this draft SHOULD NOT be used.

2.4 FT Recovery Support on Non-FT LSRs

In order to take full advantage of the FT capabilities of LSRs in the network, it may be that an LSR that does not itself contain the ability to recover from local hardware or software faults still needs to support the LDP FT enhancements described in this draft.

Consider an LSR, P1, that is an LDP peer of a fully fault tolerant LSR, P2. If P2 experiences a fault in the hardware or software that serves an LDP session between P1 and P2, it may fail the TCP connection between the peers. When the connection is recovered, the LSPs/labels between P1 and P2 can only be recovered if both LSRs were applying the FT recovery procedures to the LDP session.

3. Overview of LDP FT Enhancements

The LDP FT enhancements consist of the following main elements, which are described in more detail in the sections that follow.

- An FT Session Flag on the LDP Initialization message that indicates whether an LSR supports the LDP FT enhancements on this session.

- An FT State Flag on the LDP Initialization message that indicates whether an LSR has preserved FT label state across a failure of the TCP connection.

- An FT Reconnection Timeout, exchanged on the LDP Initialization message, that indicates the maximum time peer LSRs will preserve FT label state after a failure of the TCP connection.

- An FT Default Flag for each TCP connection, exchanged on the LDP initialization message, plus an optional FT Label Flag for an individual label, to indicate whether labels should be treated as FT labels or non-FT labels.

- Enhanced handshaking on label operations affecting FT labels to enable peer LSRs to correctly complete any operations on FT labels that are interrupted by a failure of the TCP connection.

3.1 Establishing an FT LDP Session

In order that the extensions to LDP [4] and CR-LDP [2] described in this draft can be used successfully on an LDP session between a pair of LDP peers, they MUST negotiate that the LDP FT enhancements are to be used on the LDP session.

This is done on the LDP Initialization message exchange using a new FT Session Flag, that indicates whether the peer wants to support the LDP FT enhancements on this LDP session. This flag is carried in a new FT Session TLV (see section "FT Session TLV").

If the FT Session Flag is not set by the active LSR in the LDP initialization message, the LDP FT enhancements MUST NOT be used on this LDP session.

If the active LSR sets the FT Session Flag, but the passive LSR does not set the FT Session Flag in the LDP Initialization message, the LDP FT enhancements MUST NOT be used on this LDP session.

If both LDP peers on an LDP session indicate support for the LDP FT enhancements in the LDP Initialization messages, each peer MUST use the FT label operation procedures indicated in this draft for FT labels.

3.1.1 Interoperation with Non-FT LSRs

If the active LSR does not include the FT Session TLV in its LDP Initialization message, the passive LSR MUST NOT include the FT Session TLV in its LDP Initialization message.

If the passive LSR does not support the LDP FT enhancements, for example because it implements the base LSP specification in [4], it MUST reject the LDP Initialization message sent by the active LSR using a Notification message indicating an unknown TLV. The Notification message MUST contain the Unknown TLV status code, as specified in [4]. In such cases, the active LSR SHOULD retry LDP initialization omitting the FT Session TLV, as specified in [4].

An LSR MAY present different FT/non-FT behavior through different values for the FT Session Flag on different LDP sessions, even if those sessions are successive instantiations of the LDP session between the same LDP peers.

3.2 LDP Session Failure

If the LDP FT session enhancements are not in use on an LDP session, the action of the LDP peers on failure of the LDP session is as specified in [2] and [4].

All state information and resources associated with non-FT labels MUST be released on the failure of the LDP session, including deprogramming the non-FT label from the switching hardware. This is equivalent to the behaviour specified in [4].

If the LDP FT enhancements are in use on an LDP session, both LDP peers SHOULD preserve state information and resources associated with FT labels exchanged on the LDP session. Both LDP peers SHOULD use a timer to release the preserved state information and resources associated with FT-labels if the LDP session is not reconnected within a reasonable period. The behavior when this timer expires is equivalent to the LDP session failure behavior described in [4].

The FT Reconnection Timeout each LDP peer intends to apply to the LDP

session is carried in the FT Session TLV on the LDP Initialization messages. It is RECOMMENDED that both LDP peers use the lower timeout value from the LDP Initialization exchange when setting their reconnection timer after a TCP connection failure.

3.3 LDP Session Re-initialization

When a TCP connection is recovered after FT failure, the LDP peers MUST re-exchange LDP Initialization messages.

If an LDP peer sets the FT Session Flag in the LDP Initialization message for the new instantiation of the LDP session, it MUST also set the FT State Flag if the peer has preserved state for all FT labels exchanged on previous instantiations of the TCP connection. The FT State Flag is carried in the FT Session TLV (see below).

If an LDP peer has been unable to preserve state for all FT labels exchanged on previous instantiations of the LDP session, it MUST NOT set the FT State Flag on the LDP Initialization message.

If either LDP peer does not set the FT State Flag in the LDP Initialization message, both LDP peers MUST release any state information and resources associated with all FT labels preserved from previous instantiations of the LDP session between the same LDP peers. This ensures that network resources are not permanently lost if one of the LDP peers is forced to undergo a cold start.

If both LDP peers set the FT State Flag, both LDP peers MUST use the FT label operation procedures indicated in this draft to complete any label operations on FT labels that were interrupted by the LDP session failure.

3.4 Operations on FT Labels

Label operations on FT labels are made fault-tolerant by enhancing the handshaking information available to each LDP peer in order that the LDP peers can recover from an interruption to the LDP session while one or more label operations are in progress. This is achieved by a combination of adding acknowledgements to label operations that are not acknowledged in [2] or [4], and procedures for reissuing unacknowledged label operations after re-connection of the LDP session between two LDP peers that are using the LDP FT enhancements.

Using these acknowledgements and procedures, it is not necessary for LDP peers to perform a complete re-synchronization of state for all FT labels, either on re-connection of the LDP session between the LDP peers or on a timed basis.

The message exchanges used to achieve acknowledgement of label operations and the procedures used to complete interrupted label operations are detailed in the section "Use of FT Labels".

3.5 Notes on an Alternate Solution

An alternate solution to this issue does not require retransmission of unacknowledged state changes. Instead it says that any LSP/label with unacknowledged state on TCP connection recovery MUST be torn down.

This appears to be a relatively small saving in processing and some loss of function. It also changes the recovery model in Downstream Unsolicited label distribution mode.

4. Use of FT Labels

Once an LDP session has been established as supporting the LFP FT enhancements FT recovery using the procedures described in section "Establishing an FT LDP Session", both LDP peers MUST apply the procedures described in this section for FT labels.

If the LDP session has been negotiated to not use the LDP FT enhancements, these procedures MUST NOT be used.

4.1 Identifying FT Labels

A label is identified as being an FT label if the initial Label Request or Label Mapping message relating to that label carries the FT Label Flag. This flag is carried in a new FT Protection TLV (see section "FT Protection TLV) optionally present on Label Request and Label Mapping messages.

If the FT Protection TLV is present on the initial Label Request or Label Mapping message for a label, but the FT Label Flag is not set, the label MUST be treated as a non-FT label.

If the FT Protection TLV is present on the initial Label Request or Label Mapping message for a label, and the FT Label Flag is set, the label MUST be treated as an FT label.

The setting of the FT Label Flag in subsequent message exchanges between the LDP peers MUST be ignored once the label has been identified as an FT/non-FT label.

4.1.1 Defaulting FT Label Status

If the FT Protection TLV is not present on the initial Label Request or Label Mapping message for a label, the FT/non-FT status of the label MUST be determined from the FT Default Flag, which is carried on the FT Session TLV.

The FT Default Flag value for an FT LDP session is determined by both LDP peers from the setting of the FT Default Flag in the LDP initialization message sent by the active LSR.

If the active LSR includes an FT Session TLV in its LDP Initialization message with the FT Default Flag set, all labels exchanged between the LDP peers MUST be treated as FT labels unless specifically overridden by the FT Label Flag for an individual label.

If the active LSR does not set the FT Default Flag in its LDP Initialization message, or if it omits the FT Session TLV altogether, all labels exchanged between the LDP peers MUST be treated as non-FT labels.

The setting of the FT Default Flag on re-connection of an LDP session MUST NOT change the FT/non-FT status of any labels for which state information and resources have been preserved since previous instantiations of the LDP session between the same LDP peers.

4.1.2 Scope of FT Labels

The scope of the FT/non-FT status of a label is limited to the LDP message exchanges between a pair of LDP peers. In Ordered Control, when the message is forwarded downstream or upstream, the TLV may be present or absent according to the requirements of the LSR sending the message.

4.2 Label Operation Handshaking

Once a label is identified as an FT label, both LSRs MUST apply the following handshaking procedure. This handshaking procedure MUST also be applied to the Label Request or Label Mapping message that identified the label as an FT label.

The message exchanges described in [2] and [4] do not provide for full handshaking of the allocation or deallocation of a label exchanged between LDP peers as the current LDP specification assumes that all labels and associated resources are torn down if the TCP connection fails.

Additional handshaking is required to ensure that both peers can synchronize on the result of a label operation even if that operation is interrupted by a TCP connection failure.

The LDP FT enhancements achieve handshaking of label operations by use of additional LDP messages to explicitly acknowledge the completion of a label operation between the LDP peers.

The type of acknowledgement used for each message that relates to label exchange or LSP setup is given below.

- A Label Request message MUST be acknowledged using a Label

Mapping message, to successfully return a label, or a Notification message giving the reason for the failure of the Label Request, as specified in [4].

- A Label Mapping message that is sent in response to a Label Request in Downstream On Demand Label Advertisement mode SHOULD NOT be acknowledged as it forms the response to the Label Request.

- A Label Mapping message that is sent in Downstream Unsolicited Label Advertisement mode MUST be acknowledged using a Notification message containing an OK status code or giving the reason for the failure of the Label Mapping. The peer LSR is not required to make use of use the label specified in a downstream unsolicited Label Mapping message, but it MUST acknowledge the Label Mapping message if it relates to an FT label.

- A Label Withdraw message MUST be acknowledged using a Label Release message, as specified in [4].

- A Label Release message that is sent in response to a Label Withdraw SHOULD NOT be acknowledged as it forms the response to the Label Withdraw.

- A Label Release message that is sent other than in response to a Label Withdraw message MUST be acknowledged using a Notification message containing an OK response code or giving the reason for the failure of the Label Release.

- A Label Abort message MUST NOT be explicitly acknowledged if it crosses with the Label Mapping message sent in response to the aborted Label Request, as specified in [4], or a Label Withdraw message for the same FEC. In such cases, the upstream LDP peer MUST issue a Label Release message on receipt of the Label Mapping or Label Withdraw message. See section "Procedure after TCP Re-connection" for further details of how a Label Abort may cross with a Label Withdraw message.

- A Label Abort message that does not cross with the associated Label Mapping message MUST be acknowledged using a Notification message specifying the Label Request Aborted status code, as specified in [4].

- Notification messages relating to a label operation MUST NOT be acknowledged, as they form the response handshake to a previous label operation.

4.3 Preservation of Label State

If the LDP FT enhancements are in use on an LDP session, each LDP peer MUST NOT release the state information and resources associated with FT labels exchanged on that LDP session when the TCP connection fails. This is contrary to [2] and [4], but allows label operations on FT labels to be completed after re-connection of the TCP

connection.

Both LDP peers on a LDP session that is using the LDP FT enhancements MUST preserve the state information and resources it holds for an FT label exchanged between the LDP peers until one of the following occurs:

- An upstream LDP peer SHOULD release the resources (in particular bandwidth) associated with an FT label when it initiates a Label Release or Label Abort message for the label. The upstream LDP peer MUST preserve state information for the label, even if it releases the resources associated with the label, as it may have to reissue the label operation if the TCP connection is interrupted.

- An upstream LDP peer MUST release the state information and resources associated with an FT label when it receives an explicit acknowledgement to a Label Release or Label Abort message that it sent for the label, or when it sends a Label Release message in response to a Label Withdraw message received from the downstream LDP peer.

- A downstream LDP peer SHOULD NOT release the resources associated with an FT label when it sends a Label Withdraw message for the label as it has not yet received confirmation that the upstream LDP peer has ceased to send data using the label. The downstream LDP peer MUST NOT release the state information it holds for the label as it may yet have to reissue the label operation if the TCP connection is interrupted.

- A downstream LDP peer MUST release the resources and state information associated with an FT label when receives an acknowledgement to a Label Withdraw message for the label.

- When the FT Reconnection Timeout expires, an LSR SHOULD release all state information and resources preserved for FT labels from previous instantiations of the (permanently) failed LDP session. If an LDP peer does release state information and resources in this situation, it MUST NOT set the FT State Flag (see section "LDP Session Re-initialization") on the next instantiation of the TCP connection between the same LDP peers. Otherwise it MUST set the FT State Flag on the next instantiation of the TCP connection between the same LDP peers.

- When an LSR receives a Status TLV with the E-bit set in the status code, which causes it to close the TCP connection, the LSR SHOULD release all state information and resources preserved for FT labels if it is not immediately going to try to re-establish the TCP connection. See section "Error Cases" for further discussion of the handling of the E-bit in Status TLVs.

The release of state information and resources associated with non-FT labels is as described in [2] and [4].

4.4 Procedure After TCP Failure

TCP connection failure may be notified to an LDP or CR-LDP implementation in an implementation-specific way. It may also be discovered by failure to receive any LDP message (including a KeepAlive message) on the connection within the life of the KeepAlive Timer.

When an LSR discovers or is notified of a TCP connection failure it SHOULD start an FT Reconnection Timer to allow a period for re-connection of the TCP connection between the LDP peers.

Once the TCP connection between LDP peers has failed, the active LSR SHOULD attempt to re-establish the TCP connection. The mechanisms, timers and retry counts to re-establish the TCP connection are an implementation choice. It is RECOMMENDED that any attempt to re-establish the connection take account of the failover processing necessary on the peer LSR, the nature of the network between the LDP peers, and the FT Reconnection Timeout chosen on the previous instantiation of the TCP connection (if any).

If the TCP connection cannot be re-established within the FT Reconnection Timeout period, the LDP session between the peers SHOULD be deemed to have failed permanently. The LDP implementation SHOULD fail all FT labels exchanged between the LDP peers, as described in the previous section.

If the TCP connection is successfully re-established within the FT Reconnection Timeout, both peers MUST re-synchronize the state of the label operations that were interrupted by the TCP connection failure. This procedure is described below.

4.4.1 Label Operations During TCP Failure

If an LSR determines that it needs to issue a new operation on an existing FT Label to an LDP peer to which the TCP connection has currently failed, it MUST complete that operation with the LDP peer when the TCP connection is restored, unless the label operation is overridden by a subsequent additional label operation during the TCP connection failure. For example, an LSR SHOULD NOT issue a queued Label Mapping message for a new downstream unsolicited FT label on re-establishment of the TCP connection to the upstream LDP peer if it also queues a Label Withdraw operation for the same FT label before the TCP connection is re-established.

FT label operations that cannot be correctly forwarded because of a TCP connection failure MAY be processed immediately (provided sufficient state is kept to forward the label operation) or queued for processing when the onward TCP connection is restored.

Consider the case when an upstream LSR has sent a Label Release for a label in ordered distribution mode. An LSR receiving the Label Release that also needs to forward it downstream MAY queue the Label Release and process it when the TCP connection is restored. Alternatively, the LSR MAY action the Label Release immediately, freeing up the label resources, provided that it retains sufficient state information to forward the Label Release downstream when the TCP connection is restored.

It is RECOMMENDED that Label Request operations for new FT labels are not queued awaiting the re-establishment of TCP connection that is awaiting recovery at the time the LSR determines that it needs to issue the Label Request message. Instead, such Label Request operations SHOULD be failed and, if necessary, a notification message containing the No LDP Connection status code sent upstream.

Label Requests for new non-FT labels MUST be rejected during TCP connection failure, as specified in [2] and [4].

4.5 Procedure After TCP Re-connection

The label operation handshaking described above means that all state changes for FT labels are confirmed or reproducible at each LSR.

If the TCP connection between LDP peers fails but is re-connected within the FT Reconnection Timeout, both LDP peers on the connection MUST complete any label operations for FT labels that were interrupted by the failure and re-connection of the TCP connection. Label operation are completed using the procedure described below.

4.5.1 Issuing FT Duplicate Messages

On restoration of the TCP connection between LDP peers, any label operations on FT labels that were interrupted by the TCP connection failure are completed by each LDP peer re-issuing any messages that were unacknowledged at the time of the TCP failure. The LDP peer that receives a re-issued message either processes the message from scratch (if the previous copy of the message was lost), or returns the same result for the label operation as it had previously returned (if the acknowledgement was lost).

Re-issued messages MUST carry the FT Duplicate Flag to indicate that they contain a message that has been re-issued to complete an outstanding label operation on an FT label.

The messages that may or may not need to be re-issued on re-establishment of the TCP connection between LDP peers are as follows:

- A Label Request message MUST be re-issued if an acknowledgement had not previously been received, unless a Label Abort will be re-issued for the same Label Request.

- A Label Mapping message MUST NOT be re-issued if it was originally sent in response to a Label Request message.

- A Label Mapping message that was previously sent other than in response to a Label Request message MUST be re-issued if the acknowledgement to the Label Mapping had not previously been received, unless a Label Withdraw message will be issued for the same FT label.

- A Label Withdraw message MUST be re-issued if an acknowledgement had not previously been received.

- A Label Release message MUST NOT be re-issued if it was originally sent in response to a Label Withdraw message.

- A Label Release message that was previously sent other than in response to a Label Withdraw message MUST be re-issued if the acknowledgement to the Label Release had not previously been received.

- A Label Abort message MUST be re-issued if an acknowledgement or crossing Label Mapping message had not previously been received.

Re-issued messages SHOULD NOT contain the same Message ID as the orignal message.

Any FT label operations that were queued (see section "Label Operations During TCP Failure") during the TCP connection failure MUST be issued on re-establishment of the LDP session. Queued Label Request or Label Mapping messages SHOULD NOT be issued if a Label Abort or Label Withdraw message is also queued for the same FT label.

4.5.2 Receiving FT Duplicate Messages

If an LSR receives a re-issued message marked with the FT Duplicate Flag on an LDP session for which both the FT Session Flag and FT State Flag were set during LDP Initialization, it MUST process the message according to the following procedure:

- All re-issued messages received by an LSR that it can match to a previously received message for the same FT Label MUST be acknowledged using the same message/status code as the LSR previously sent in response to the original copy of the message.

- If an LSR receives a re-issued Label Withdraw of which it has no record (because it has already released the state information when it received the original request), it MUST return a Label Release message for the same label.

- If an LSR receives a re-issued Label Release of which it has no record (because it has already released the state information when it received the original request), it MUST return a Notification message containing an OK status code.

- If an LSR receives a re-issued Label Abort message for a label of which it has no record, it MUST return a Notification message containing a Label Request Aborted status code.

- If an LSR receives a re-issued Label Mapping message for an FT label of which it has no record, it MUST process the Label Mapping from scratch. (This is most likely to happen in Downstream Unsolicited Label Advertisement mode, especially if the upstream LSR does not retain all such labels).

- If an LSR receives a re-issued Label Request message for a FT label of which it has no record, it MUST process the Label Request from scratch.

4.5.3 Forwarding FT Duplicate Messages

When forwarding a re-issued message upstream or downstream, an LSR must follows the procedure described below:

- If an LDP peer operating in ordered distribution mode receives a re-issued message for an FT label, it MUST NOT forward the message upstream or downstream (as appropriate to the message type) if it had previously forwarded the original message upstream or downstream.

- If an LDP peer receives a re-issued message, but it had not received the original message, the LDP peer MUST process the re-issued message from scratch and forward the message upstream or downstream as required.

- If an LDP peer forwards a re-issued message upstream or downstream it MUST NOT set the FT Duplicate Flag in the forwarded messages.

4.5.4 Error Cases

If an LSR receives an apparent duplicate message that is not marked with the FT Duplicate Flag, it MUST reject the message using a Notification message containing the Unexpected Duplicate/Duplicate Not Marked status code.

If an LSR that supports the LDP FT enhancements receives a message marked with the LDP Duplicate Flag, but the LDP Initialization for the LDP session did not set both the FT Session Flag or the

FT State Flag, the LSR MUST reject the message using a Notification Message containing the Unexpected Duplicate/Session Not FT status code.

If an LSR is using the LDP FT enhancements on an LDP session and receives a message on that session that relates to a non-FT label but that is marked with the FT Duplicate Flag, it SHOULD reject the message using a Notification message containing the Unexpected Duplicate/Non-FT Label status code. The LSR MAY ignore the FT Duplicate Flag if it is set on Label Release, Label Abort or Label Withdraw message for a non-FT label provided that it releases the state information and resources associated with that label.

If an LSR receives a Status TLV message containing the Unexpected Duplicate/Non-FT Label status code, but for a label that it believes to be an FT label, it SHOULD Release/Withdraw the label (without setting the FT Duplicate Flag) and issue a new Label Request/Label Mapping message to request an FT label.

If an LSR receives a Status TLV with the E-bit set, it SHOULD NOT, on re-establishment of the TCP connection, immediately re-issue the message (if any) indicated in the Status TLV as being in error. In particular, if an LSR receives the Unexpected Duplicate/Session Not FT status code, it should release all state and resources held for FT labels associated with this session before attempting to re-establish the TCP connection.

Receipt of duplicate messages on an LDP session that does not support the LDP FT enhancements is outside the scope of this draft and MUST be handled as described in [5], [2] and [4], regardless of the setting of the FT Duplicate flag on the message.

4.5.5 Interaction with CR-LDP LSP Modification

Re-issuing LDP messages for FT operation is orthogonal to the use of duplicate messages marked with the Modify ActFlg, as specified in [5]. Each time an LSR uses the modification procedure for an LSP to issue a new Label Request message, the FT label operation procedures MUST be separately applied to the new Label Request message.

5. Changes to Existing Messages

5.1 LDP Initialization Message

The LDP FT enhancements add the following optional parameter to a LDP Initialization message

Optional Parameter	Length	Value
FT Session TLV	4	See below

The encoding for FT Session TLV is found in Section "FT Session TLV".

FT Session
If present, specifies the FT behavior of the LDP session.

5.2 Label Request Message

The LDP FT enhancements add the following optional parameter to a Label Request message

Optional Parameter	Length	Value
FT Protection TLV	4	See below

The encoding for FT Protection TLV is found in Section "FT Protection TLV".

FT Protection
If present, specifies the FT/non-FT characteristics for the label and whether the message has been re-issued for FT recovery.

5.3 Label Mapping Message

The LDP FT enhancements add the following optional parameter to a Label Mapping message

Optional Parameter	Length	Value
FT Protection TLV	4	See below

The encoding for FT Protection TLV is found in Section "FT Protection TLV".

FT Protection
If present, specifies the FT/non-FT characteristics for the label and whether the message has been re-issued for FT recovery.

5.4 Label Release Message

The LDP FT enhancements add the following optional parameter to a Label Release message

Optional Parameter	Length	Value
FT Protection TLV	4	See below

The encoding for FT Protection TLV is found in Section "FT Protection TLV".

FT Protection
If present, specifies whether the Label Release message has

been re-issued for FT recovery.

5.5 Label Withdraw Message

The LDP FT enhancements add the following optional parameter to a Label Withdraw message

Optional Parameter	Length	Value
FT Protection TLV	4	See below

The encoding for FT Protection TLV is found in Section "FT Protection TLV".

FT Protection
If present, specifies whether the Label Withdraw message has been re-issued for FT recovery.

5.6 Label Abort Message

The LDP FT enhancements add the following optional parameter to a Label Abort message

Optional Parameter	Length	Value
FT Protection TLV	4	See below

The encoding for FT Protection TLV is found in Section "FT Protection TLV".

FT Protection
If present, specifies whether the Label Abort message has been re-issued for FT recovery.

5.7 Notification Request Message

The LDP FT enhancements add the following optional parameter to a Notification message

Optional Parameter	Length	Value
FT Protection TLV	4	See below

The encoding for FT Protection TLV is found in Section "FT Protection TLV".

FT Protection
If present, specifies whether the Label Release message has been issued in response to an FT Duplicate LDP message.
for FT

6. New Fields and Values

6.1 Status Codes

The following new status codes are defined to indicate various error conditions specific to the LDP FT enhancements. These status codes are carried in the Status TLV of a Notification message.

The "E" column is the required setting of the Status Code E-bit; the "Status Data" column is the value of the 30-bit Status Data field in the Status Code TLV.

Note that the setting of the Status Code F-bit is at the discretion of the LSR originating the Status TLV. However, it is RECOMMENDED that the F-bit is not set on Notification messages containing status codes 0x00000017 - 0x00000019 because the duplication of messages SHOULD be restricted to being a per-hop behavior.

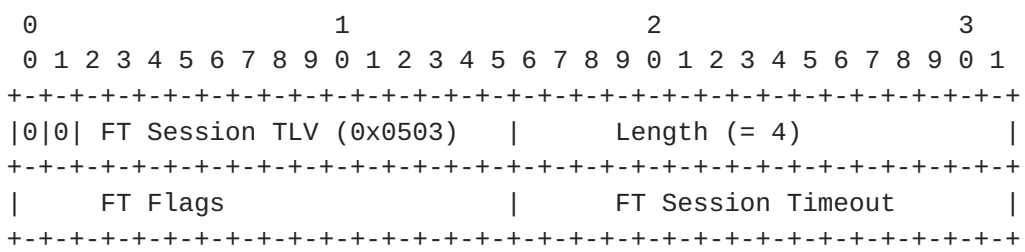
Status Code	E	Status Data
No LDP Session	0	0x00000016
Unexpected Duplicate / Duplicate Not Marked	1	0x00000017
Unexpected Duplicate / Session Not FT	1	0x00000018
Unexpected Duplicate / Non-FT Label	0	0x00000019
Temporary Shutdown	0	0x0000001A

The Temporary Shutdown status code SHOULD be used in place of the Shutdown status code (which carries the E-bit) if the LSR that is shutting down wishes to inform its LDP peer that it expects to be able to preserve FT label state and to return to service before the FT Reconnection Timer expires.

6.2 FT Session TLV

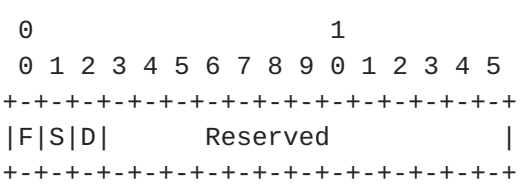
LDP peers can negotiate whether the LDP session between them supports FT extensions by using a new OPTIONAL parameter, the FT Session TLV, on LDP Initialization Messages.

The FT Session TLV is encoded as follows.



FT Flags

FT Flags: A 16 bit field that indicates various attributes the FT support on this LDP session. This fields is formatted as follows:



F: FT Session Flag.
Set to 1 if this LDP session is to use the LDP FT Enhancements defined in this draft, and set to 0 otherwise. See the section "Establishing an FT LDP Session" for details on how this flag is used.

S: FT State Flag.
Set to 1 if the sending LSR has preserved state and resources for all FT-labels since the previous LDP session between the same LDP peers, and set to 0 otherwise. See the section "LDP Session Re-initialization" for details of how this flag is used.

D: FT Default Flag.
Set to 1 if all labels exchanged on this LDP session are to be treated as FT labels, or set to 0 if labels are to default to non-FT labels. See the section "Identifying FT Labels" for details of how this flag is used.

All other bits in this field are currently reserved and SHOULD be set to zero on transmission and ignored on receipt.

FT Session Timeout

The period of time the sending LSR will preserve state and resources for FT labels exchanged on the previous instantiation of an FT LDP session that has currently failed. The timeout is encoded as a 16-bit unsigned integer number of seconds.

The value of 0 for this field is reserved and MUST NOT be used.

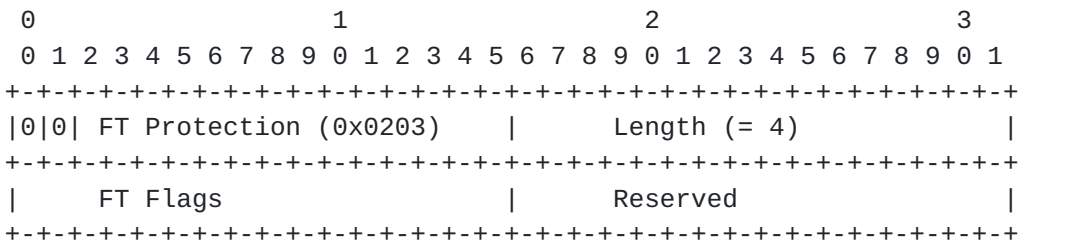
See the section "LDP Session Failure" for details of how this field is used.

6.3 FT Protection TLV

LDP peers use the FT Protection TLV to override the FT Default Flag (see section "FT Session TLV") for a specific label and to identify LDP messages that have been reissued as part of the FT recovery procedures.

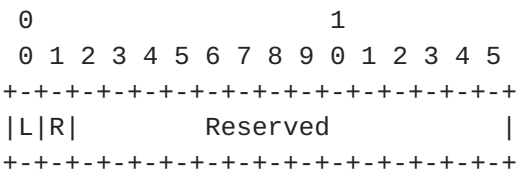
The FT Protection TLV MUST NOT be used in messages flowing on an LDP session that does not support the LDP FT enhancements.

The FT Protection TLV is encoded as follows.



FT Protection

FT Protection: A 16 bit field that indicates various attributes of the label and the LDP message. This field is formatted as follows:



L: FT Label Flag.
Set to 1 if this label is an FT label, and set to 0 otherwise. See the section "Identifying FT Labels" for details on how this flag is used. This field MUST be ignored on all messages other than the initial Label Request or Label Mapping message for a new label.

R: FT Duplicate Flag.
Set to 1 if the LDP message has been re-issued for FT recovery, and set to 0 otherwise. See the section "Procedure after TCP Re-connection" for details of how this flag is used.

All other bits in this field are currently reserved and SHOULD be set to zero on transmission and ignored on receipt.

Reserved
A 16-bit field that is currently reserved. This field SHOULD be set to zero on transmission and ignored on receipt.

7. Example Use

Consider two LDP peers, P1 and P2, implementing CR-LDP over a TCP connection that connects them.

Let there be eleven LSPs in the following states when the TCP connection fails and is recovered.

L1: P1 has received a Label Request message from upstream and has sent a Label Request message to P2. This request has not been received by P2.

L2: P1 has received a Label Request message from upstream and has sent a Label Request message to P2. This request has been received at P2 which has sent a Label Request message on downstream. No Label Mapping message has been received by P2.

L3: P1 has received a Label Request from upstream and has sent a Label Request to P2. This has been rejected by P2 which has sent a Notification message to P1. This Notification has not been received by P1.

L4: P1 has received a Label Request message from upstream and has sent a Label Request message to P2. This request has been received at P2 which has sent a Label Request message on downstream.
P2 has received a Label Mapping message from down stream and has forwarded it to P1. The Label Mapping message has not been received by P1.

L5: P1 has received a Label Request message from upstream and has sent a Label Request message to P2. This request has been received at P2 which has sent a Label Request message on downstream.
P2 has received a Label Mapping message from down stream and has forwarded it to P1. P1 has received the Label Mapping message and has forwarded it upstream.

L6: This LSP was previously successfully established. P1 has received a Label Release message from upstream and has sent a Label Release message to P2. This request has not been received by P2.

L7: This LSP was previously successfully established. P1 has received a Label Release message from upstream and has sent a Label Release message to P2. This request has been received by P2 which forwarded it downstream and sent a Notification acknowledgement (with OK status code) upstream to P1.
The Notification has not been received by P1.

L8: P2 sent a unsolicited Label Mapping to P1. P1 has
sent a Notification acknowledgement, but this has not been
received by P2.

L9: P1 has received a Label Request message from upstream and has sent a Label Request message to P2. This request has been received at P2 which has sent a Label Request message on downstream.
P1 has received a Label Abort from Upstream and has sent the Label Abort to P2, but this has not been received by P2.

L10: As L9, but P2 has received a Label Mapping from downstream and sent it on to P1, though this has not been received by P1.

L11: As 9, but P2 also receives a Label Withdraw from downstream while the TCP connection is down.

On recovery of the TCP connection, the processing of each LSP at each LSR is as follows.

L1: P1: Resend Label Request message to P2.
P2: Treat duplicate Label Request message as new.
Forward Label Request downstream.

L1: P1: Resend Label Request message to P2.
P2: Ignore duplicate Label Request message.

L3: P1: Resend Label Request message to P2.
P2: Treat duplicate Label Request message as new.
If failure reason persists, send new Notification to P1
Otherwise, forward Label Request downstream.

L4: P1: Resend Label Request message to P2.
P2: Respond to duplicate Label Request message with duplicate Label Mapping message.

L5: P1: No work required. LSP is preserved.
P2: No work required. LSP is preserved.

L6: P1: Resend Label Release message to P2.
P2: Treat duplicate Label Release message as new.
Forward Label Release message downstream.
Send Notification (OK) to P1.

L7: P1: Resend Label Release message to P2.
P2: Fail to match duplicate Label Release message to LSP.
Send Notification (OK) to P1.

L8: P2: Resend Label Mapping
P1: Respond to duplicate Label Mapping with duplicate Notification (OK) to P2.

L9: P1: Send duplicate Label Abort
P2: Treat Label Abort message as new.

Forward Label Abort downstream.

```
L10: P1: Resend Label Abort.
     P2: Resend Label Mapping
     P1: Receive duplicate Label Mapping and treat it as new.
         Send Label Release to P2 to tear down label.
     P2: Ignore duplicate Label Abort.
         Process Label Release as new when it is received.

L11: P1: Resend Label Abort.
     P2: Send Label Withdraw.
         Ignore duplicate Label Abort.
     P1: Receive Label Withdraw and process it as new.
```

8. Security Considerations

The LDP FT enhancements inherit similar security considerations to those discussed in [2] and [4].

The LDP FT enhancements allow the re-establishment of a TCP connection between LDP peers without a full re-exchange of the attributes of established labels, which renders LSRs that implement the extensions specified in this draft vulnerable to additional denial-of-service attacks as follows:

- An intruder may impersonate an LDP peer in order to force a failure and reconnection of the TCP connection, but where the intruder does not set the FT State Flag on re-connection. This forces all FT labels to be released.

- Similarly, an intruder could set the FT State Flag on re-establishment of the TCP session without preserving the state and resources for FT labels.

- An intruder could intercept the traffic between LDP peers and override the setting of the FT Label Flag to be set to 0 for all labels.

All of these attacks may be countered by use of an authentication scheme between LDP peers, such as the scheme outlined in [4].

Alternative authentication schemes for LDP peers are outside the scope of this draft, but could be deployed to provide enhanced security to implementations of LDP, CR-LDP and the LDP FT enhancements.

The RECOMMENDED use of new message IDs for re-issued messages (see section "Issuing FT Duplicate Messages") is intended to help counter replay attacks, when used in conjunction with encryption or signing of the LDP session traffic. See [4] for details of how to

apply MD5 signatures to LDP session traffic.

9. Acknowledgments

The work in this draft is based on the LDP and CR-LSP ideas expressed by the authors of [2] and [4].

The authors would also like to acknowledge the careful review and comments of Nick Weeds, Piers Finlayson, Tim Harrison and Duncan Archer at Data Connection Ltd.

10. Intellectual Property Consideration

Data Connection may seek patent or other intellectual property protection for some of the technologies disclosed in this document. If any standards arising from this document are or become protected by one or more patents assigned to Data Connection, Data Connection intends to make a license available to any qualified applicant under reasonable and non-discriminatory terms.

11. References

1 Bradner, S., "The Internet Standards Process -- Revision 3", BCP 9, RFC 2026, October 1996.

2 Jamoussi, B., et. al., Constraint-Based LSP Setup using LDP, draft-ietf-mpls-cr-ldp-03.txt, September 1999,(work in progress).

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

4 Andersson, L., et. al., LDP Specification, draft-ietf-mpls-ldp-06.txt, October 1999 (work in progress).

5 Ash, G., et al., LSP Modification Using CR-LDP, draft-ietf-mpls-crlsp-modify-00.txt, December 1999 (work in progress).

6 Braden, R., et al., Resource ReSerVation Protocol (RSVP) -- Version 1, Functional Specification, RFC 2205, September 1997.

7 Berger, L., et al., RSVP Refresh Reduction Extensions, draft-ietf-rsvp-refresh-reduct-02.txt, February 2000 (work in progress).

8 Swallow, G., et al,. Extensions to RSVP for LSP Tunnels, draft-ietf-mpls-rsvp-lsp-tunnel-04.txt, September 1999 (work in progress).

12. Authors' Addresses

Adrian Farrel
Data Connection Ltd.
Windsor House
Pepper Street
Chester
Cheshire
CH1 1DF
UK
Phone: +44-(0)-1244-313440
Fax: +44-(0)-1244-312422
Email: af@datcon.co.uk

Paul Brittain
Data Connection Ltd.
Windsor House
Pepper Street
Chester
Cheshire
CH1 1DF
UK
Phone: +44-(0)-1244-313440
Fax: +44-(0)-1244-312422
Email: pjb@datcon.co.uk

13. Full Copyright Statement

14. IANA Considerations

This draft requires the use of a number of new TLVs and status codes from the number spaces within the LDP protocol. This section explains the logic used by the authors to choose the most appropriate number space for each new entity, and is intended to assist in the

determination of any final values assigned by IANA or the MPLS WG in the event that the MPLS WG chooses to advance this draft on the standards track.

14.1 FT Session TLV

The FT Session TLV carries attributes that affect the entire LDP session between LDP peers. It is suggested that the type for this TLV should be chosen from the 0x05xx range for TLVs that is used in [4] by other TLVs carrying session-wide attributes. At the time of this writing, the next available number in this range is 0x0503.

14.2 FT Protection TLV

The FT Protection TLV carries attributes that affect a single label exchanged between LDP peers. It is suggested that the type for this TLV should be chosen from the 0x02xx range for TLVs that is used in [4] by other TLVs carrying label attributes. At the time of this writing, the next available number in this range is 0x0203.

Consideration was given to carrying the FT Label Flag and FT Duplicate Flag in the ActFlg field within the LSPID TLV [2]. The authors felt that this would be inappropriate as the LSPID TLV is not used on a "pure LDP" (as opposed to CR-LDP) session.

Consideration was also given to modifying the existing Label TLVs to carry these flags. The authors felt this would be too great a change to the use of the existing Label TLVs to introduce at this stage in the development of [4].

14.3 Status Codes

The authors' current understanding is that MPLS status codes are not sub-divided into specific ranges for different types of error. Hence, the numeric status code values suggested in this draft are simply the next available values at the time of writing and may be substituted for other numeric values.

See section "Status Codes" for details of the status codes defined in this draft.