Inter-Chassis Communication Protocol for L2VPN PE Redundancy

draft-martini-pwe3-iccp-01.txt

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on August 17, 2009

Abstract

This document specifies an inter-chassis communication protocol (ICCP) that enables PE redundancy for Virtual Private Wire Service (VPWS) and Virtual Private LAN Service (VPLS) applications. The protocol runs within a set of two or more PEs, forming a redundancy group, for the purpose of synchronizing data amongst the systems. It accommodates multi-chassis attachment circuit as well as pseudowire redundancy mechanisms.
Table of Contents

1 Specification of Requirements .................................. 3
2 Acknowledgments ................................................ 4
3 Introduction ....................................................... 4
4 ICCP Overview .................................................... 4
4.1 Redundancy Model & Topology .................................. 4
4.2 ICCP Interconnect Scenarios ................................... 6
4.2.1 Co-located Dedicated Interconnect .......................... 6
4.2.2 Co-located Shared Interconnect .............................. 7
4.2.3 Geo-redundant Dedicated Interconnect ....................... 7
4.2.4 Geo-redundant Shared Interconnect .......................... 8
4.3 ICCP Requirements ............................................. 9
5 ICC LDP Protocol Extension Specification ....................... 10
5.1 LDP ICCP Capability Advertisement ............................ 11
5.2 RG Membership Management ..................................... 11
5.3 Application Connection Management ............................ 12
5.4 Application Versioning ......................................... 13
5.5 Application Data Transfer ....................................... 14
5.6 Dedicated Redundancy Group LDP session ........................ 14
6 ICCP PE Node Failure Detection Mechanism ....................... 14
7 ICCP Message Formats ........................................... 15
7.1 Encoding ICC into LDP Messages ............................... 15
7.1.1 ICC Header .................................................. 16
7.1.2 Message Encoding .......................................... 17
7.2 RG Connect Message .......................................... 19
7.2.1 Sender Name TLV ........................................... 19
7.3 RG Disconnect Message ........................................ 20
7.4 RG Notification Message ....................................... 23
7.4.1 Notification Message TLVs ................................ 23
7.5 RG Application Data Message .................................. 26
8 Application TLVs ................................................. 27
8.1 Pseudowire Redundancy (PW-RED) Application TLVs .......... 27
8.1.1 PW-RED Connect TLV ........................................ 27
8.1.2 PW-RED Disconnect TLV ..................................... 28
8.1.3 PW-RED Config TLV ......................................... 29
8.1.4 Service Name TLV ........................................... 29
8.1.5 PW ID TLV .................................................. 29
8.1.6 Generalized PW ID TLV ..................................... 30
8.2 Multi-chassis LACP (mLACP) Application TLVs ............... 31
8.2.1 mLACP Connect TLV ......................................... 32
8.2.2 mLACP Disconnect TLV ....................................... 33
8.2.3 mLACP System Config TLV .................................. 33
8.2.4 mLACP Port Config TLV ..................................... 35
8.2.5 mLACP Change Port Priority TLV ............................ 36
8.2.6 mLACP Port State TLV ....................................... 37
1. Specification of Requirements

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.
2. Acknowledgments

The authors wish to acknowledge the important contributions of Dennis Cai, Neil McGill and Amir Maleki.

3. Introduction

Network availability is a critical metric for service providers as it has a direct bearing on their profitability. Outages translate not only to lost revenue but also to potential penalties mandated by contractual agreements with customers running mission-critical applications that require tight SLAs. This is true for any carrier network, and networks employing Layer2 Virtual Private Network (L2VPN) technology are no exception. Network high-availability can be achieved by employing intra and inter-chassis redundancy mechanisms. The focus of this document is on the latter. The document defines an Inter-Chassis Communication Protocol (ICCP) that allows synchronization of state and configuration data between a set of two or more PEs forming a Redundancy Group (RG). The protocol supports multi-chassis redundancy mechanisms that can be employed on either the attachment circuit or pseudowire front.

4. ICCP Overview

4.1. Redundancy Model & Topology

The focus of this document is on PE node redundancy. It is assumed that a set of two or more PE nodes are designated by the operator to form a Redundancy Group (RG). Members of a Redundancy Group fall under a single administration (e.g. service provider) and employ a common redundancy mechanism towards the access (attachment circuits or access pseudowires) and/or towards the core (pseudowires) for any given service instance. It is possible, however, for members of a RG to make use of disparate redundancy mechanisms for disjoint services. The PE devices may be offering any type of L2VPN service, i.e. VPWS or VPLS. As a matter of fact, the use of ICCP may even be applicable for Layer 3 service redundancy, but this is considered to be outside the scope of this document.

The PEs in a RG offer multi-homed connectivity to either individual devices (e.g. CE, DSLAM, etc...) or entire networks (e.g. access network). Figure 1 below depicts the model.
In the topology of Figure 1, the redundancy mechanism employed towards the access node/network can be one of a multitude of technologies, e.g. it could be IEEE 802.3ad Link Aggregation Groups with Link Aggregation Control Protocol (LACP), or SONET APS. The specifics of the mechanism are out of the scope of this document. However, it is assumed that the PEs in the RG are required to communicate amongst each other in order for the access redundancy mechanism to operate correctly. As such, it is required to run an inter-chassis communication protocol among the PEs in the RG in order to synchronize configuration and/or running state data.

Furthermore, the presence of the inter-chassis communication channel allows simplification of the pseudowire redundancy mechanism. This is primarily because it allows the PEs within a RG to run some arbitration algorithm to elect which pseudowire(s) should be in active or standby mode for a given service instance. The PEs can then advertise the outcome of the arbitration to the remote-end PE(s), as opposed to having to embed a hand-shake procedure into the pseudowire redundancy status communication mechanism, and every other possible Layer 2 status communication mechanism.
4.2. ICCP Interconnect Scenarios

When referring to ‘interconnect’ in this section, we are concerned with the links or networks over which Inter-Chassis Communication Protocol messages are transported, and not normal data traffic between PEs. The PEs which are members of a RG may be either physically co-located or geo-redundant. Furthermore, the physical interconnect between the PEs over which ICCP is to run may comprise of either dedicated back-to-back links or a shared connection through the PSN network (e.g., core). This gives rise to a matrix of four interconnect scenarios, described next.

4.2.1. Co-located Dedicated Interconnect

In this scenario, the PEs within a RG are co-located in the same physical location (POP, CO). Furthermore, dedicated links provide the interconnect for ICCP among the PEs.

![Figure 2: ICCP Co-located PEs Dedicated Interconnect Scenario](image)

Given that the PEs are connected back-to-back in this case, it is possible to rely on Layer 2 redundancy mechanisms to guarantee the robustness of the links carrying the ICCP. For example, if the interconnect comprises of IEEE 802.3 Ethernet links, it is possible to provide redundant interconnect by means of IEEE 802.3ad Link Aggregation Groups.
4.2.2. Co-located Shared Interconnect

In this scenario, the PEs within a RG are co-located in the same physical location (POP, CO). However, unlike the previous scenario, there are no dedicated links between the PEs. The interconnect for ICCP is provided through the core network to which the PEs are connected. Figure 3 depicts this model.

![Diagram of co-located shared interconnect scenario]

Figure 3: ICCP Co-located PEs Shared Interconnect Scenario

Given that the PEs in the RG are connected over the Packet Switched Network (PSN), then PSN Layer mechanisms can be leveraged to ensure the resiliency of the interconnect against connectivity failures. For example, it is possible to employ RSVP LSPs with Fast ReRoute (FRR) and/or end-to-end backup LSPs.

4.2.3. Geo-redundant Dedicated Interconnect

In this variation, the PEs within a Redundancy Group are located in different physical locations to provide geographic redundancy. This may be desirable, for example, to protect against natural disasters or the like. A dedicated interconnect is provided to link the PEs, which is a costly option, especially when considering the possibility of providing multiple such links for interconnect robustness. The resiliency mechanisms for the interconnect are similar to those highlighted in the co-located interconnect counterpart.
4.2.4. Geo-redundant Shared Interconnect

In this scenario, the PEs of a RG are located in different physical locations and the interconnect for ICCP is provided over the PSN network to which the PEs are connected. This interconnect option is more likely to be the one used for geo-redundancy as it is more economically appealing compared to the geo-redundant dedicated interconnect. The resiliency mechanisms that can be employed to guarantee the robustness of the ICCP transport are PSN Layer mechanisms as has been described in a previous section.
4.3. ICCP Requirements

The Inter-chassis Communication Protocol should satisfy the following requirements:

- **i.** Provide a control channel for communication between PEs in Redundancy Group (RG). Nodes maybe co-located or remote (refer to Interconnect Scenarios section above). It is expected that client applications which make use of ICCP services will only use this channel to communicate control information and not data-traffic. As such the protocol should cater for low-bandwidth, low-delay and highly reliable message transfer.

- **ii.** Accommodate multiple client applications (e.g. multi-chassis LACP, PW redundancy, SONET APS, etc...). This implies that the messages should be extensible (e.g. TLV-based) and the protocol should provide a robust application registration and versioning scheme.

- **iii.** Provide reliable message transport and in-order delivery between nodes in a RG with secure authentication mechanisms built into the protocol. The redundancy applications that are clients of ICCP expect reliable message transfer, and as such will assume that the protocol takes care of flow-control and retransmissions. Furthermore, given that the applications will rely on ICCP to communicate data used to synchronize state-machines on disparate nodes, it is critical that ICCP guarantees in-order message delivery. Loss of messages or out-of-sequence messages would have adverse side-effects to the operation of the client applications.

- **iv.** Provide a common mechanism to actively monitor the health of PEs in a RG. This mechanism will be used to detect PE node failure and inform the client applications. The applications require this to trigger failover according to the procedures of the employed redundancy protocol on the AC and PW. It is desired to achieve sub-second detection of loss of remote node (~ 50 - 150 msec) in order to give the client applications (redundancy mechanisms) enough reaction time to achieve sub-second service restoration time.

- **v.** Provide asynchronous event-driven state update, independent of periodic messages, for immediate notification of client applications’ state changes. In other words, the transmission of messages carrying application data should be on-demand rather than timer-based to minimize inter-chassis
state synchronization delay.

-vi. Accommodate multi-link and multi-hop interconnect between nodes. When the devices within a RG are located in different physical locations, the physical interconnect between them will comprise of a network rather than a link. As such, ICCP should accommodate the case where the interconnect involves multiple hops. Furthermore, it is possible to have multiple (redundant) paths or interconnects between a given pair of devices. This is true for both the co-located and geo-redundant scenarios. ICCP should handle this as well.

-vii. Ensure transport security between devices in a RG. This is especially important in the scenario where the members of a RG are located in different physical locations and connected over a shared (e.g. PSN) network.

-viii. Must allow operator to statically configure members of RG. Auto-discovery may be considered in the future.

-ix. Allow for flexible RG membership. It is expected that only two nodes per an RG will cover most of the redundancy applications for common deployments. However, ICCP should not preclude supporting more than two nodes in a RG by virtue of design. Furthermore, it is required to allow a single node to be member of multiple RGs simultaneously.

5. ICC LDP Protocol Extension Specification

To address the requirements identified in the previous section, ICCP is modeled to comprise of three layers:

-i. Application Layer: This provides the interface to the various redundancy applications that make use of the services of ICCP. ICCP is concerned with defining common connection management procedures and the formats of the messages exchanged at this layer; however, beyond that, it does not impose any restrictions on the procedures or state-machines of the clients, as these are deemed application-specific and lie outside the scope of ICCP. This guarantees implementation inter-operability without placing any unnecessary constraints on internal design specifics.
-ii. Inter Chassis Communication (ICC) Layer: This layer implements the common set of services which ICCP offers to the client applications. It handles protocol versioning, RG membership, PE node identification and ICCP connection management.

-iii. Transport Layer: This layer provides the actual ICCP message transport. It is responsible for addressing, route resolution, flow-control, reliable and in-order message delivery, connectivity resiliency/redundancy and finally PE node failure detection. The Transport layer may differ depending on the Physical Layer of the inter-connect.

5.1. LDP ICCP Capability Advertisement

When a RG is enabled on a particular PE, the capability of supporting ICCP must be advertised to all LDP peers in that RG. This is achieved by using the methods in [LDP-CAP] and advertising the ICCP LDP capability TLV. If an LDP peer supports the dynamic capability advertisement, this can be done by sending a new capability message with the S bit set for the ICCP capability TLV when the first RG is enabled on the PE. If the peer does not support dynamic capability advertisement, then the ICCP TLV MUST be included in the LDP initialization procedures in the capability parameter [LDP-CAP].

5.2. RG Membership Management

ICCP defines a mechanism that enables PE nodes to manage their RG membership. When a PE is configured to be a member of a RG, it will first advertise the ICCP capability to its peers. Subsequently, the PE sends a RG Connect message to the peers that have also advertised ICCP capability. The PE then waits for the peers to send their own RG Connect messages, if they haven’t already. For a given RG, the ICCP connection between two devices is considered to be operational only when both have sent and received ICCP RG Connect messages for that RG.

If a PE that has sent a particular RG Connect message doesn’t receive a corresponding RG Connect (or a Notification message with NAK) from a destination, it will remain in a state expecting the corresponding RG Connect message (or Notification message). The RG will not become operational until the corresponding RG Connect Message has been received. If a PE that has sent an RG Connect message receives a Notification message with a NAK, it will stop attempting to bring up the ICCP connection immediately. The PE MUST resume bringing up the connection after it receives a RG Connect message from the peer PE.
for the RG in question. This is achieved by responding to the incoming RG Connect message with an appropriate RG Connect.

A device MUST send a NAK for a RG Connect message if at least one of the following conditions is satisfied:

- i. the PE is not a member of the RG;
- ii. the maximum number of simultaneous ICCP connections that the PE can handle is exceeded.

A PE sends a RG Disconnect message to tear down the ICCP connection for a given RG. This is a unilateral operation and doesn’t require any acknowledgement from the other PEs. Note that the ICCP connection for a RG MUST be operational before any client application can make use of ICCP services in that RG.

5.3. Application Connection Management

ICCP provides a common set of procedures by which applications on one PE can connect to their counterparts on another PE, for purpose of inter-chassis communication in the context of a given RG. The prerequisite for establishing an application connection is to have an operational ICCP RG connection between the two endpoints. It is assumed that the association of applications with RGs is known apriori, e.g. by means of device configuration. ICCP then sends an Application-specific Connect TLV (carried in RG Connect message), on behalf of each client application, to each remote PE within the RG. The client may piggyback application-specific information in that Connect TLV, which for example can be used to negotiate parameters or attributes prior to bringing up the actual application connection. The procedures for bringing up the application connection are similar to those of the ICCP connection: An application connection between two nodes is up only when both nodes have sent and received RG Connect Messages with the proper Application-specific Connect TLVs. A PE MUST send a Notification Message to NAK an application connection request if one of the following conditions is encountered:

- i. the application doesn’t exist or is not configured for that RG;
- ii. the application connection count exceeds the PE’s capabilities.

When a PE receives such a NAK notification, it should stop attempting to bring up the application connection until it receives a new application connection request from the remote PE. This is done by
responding to the incoming RG Connect message (carrying an Application-specific Connect TLV) with an appropriate RG Connect message (carrying a corresponding Application-specific Connect TLV).

When an application is stopped on a device or it is no longer associated with a RG, it should signal ICCP to trigger sending an Application-specific Disconnect TLV (in RG Disconnect message). This is a unilateral notification to the other PEs within a RG, and as such doesn’t trigger any response.

5.4. Application Versioning

During application connection setup time, a given application on one PE can negotiate with its counterpart on a peer PE the proper application version to use for communication. If no common version is agreed upon, then the application connection is not brought up. This is achieved through the following simple rules:

- If an application receives an Application-specific Connect TLV with a version that is higher than its own, it MUST send a Notification message with a NAK TLV indicating status code "Incompatible Protocol Version" and supplying the version that is locally supported by the PE.

- If an application receives an Application-specific Connect TLV with a version that is lower than its own, it MAY respond with a RG Connect that has an Application-specific Connect TLV using the same version that was received. Alternatively, the application MAY respond with a Notification message to NAK the request using the "Incompatible Protocol Version" code, and supplying the version that is supported. The above allows an application to operate in either backwards compatible or incompatible modes.

- If an application receives an Application-specific Connect TLV with a version that is equal to its own, then the application MUST honor or reject the request based on whether the application is configured for the RG in question, and whether or not the application connection count has been exceeded.
5.5. Application Data Transfer

When an application has information to transfer over ICCP it triggers the transmission of an Application Data message. ICCP guarantees in-order and loss-less delivery of data. An application may NAK a message or a set of one or more TLVs within a message by using the Notification Message with NAK TLV. Furthermore, an application may implement its own ACK mechanism, if deemed required, by defining an application-specific TLV to be transported in an Application Data message.

5.6. Dedicated Redundancy Group LDP session

In some ICCP applications there is a requirement to exchange a fairly large amount of RG information in a very short period of time. In order to better distribute the load in a multiple processor system, and to avoid head of line blocking to other LDP applications, it may be required to initiate a separate TCP/IP session between the two LDP speakers.

This procedure is OPTIONAL, and does not change the operation of LDP or ICCP.

An LSR that requires a separate LDP session will advertise a separate LDP adjacency with a non-zero label space identifier. This will cause the remote peer to open a separate LDP session for this label space. No labels need be advertised in in this label space, as it is only used for a particular, or particular set of ICCP RGs. All relevant LDP, and ICCP procedures still apply as described in the relevant documents.

6. ICCP PE Node Failure Detection Mechanism

ICCP provides its client applications a notification when a remote PE that is member of the RG fails. This is used by the client applications to trigger failover according to the procedures of the employed redundancy protocol on the AC and PW. To that end, ICCP does not define its own KeepAlive mechanism for purpose of monitoring the health of remote PE nodes, but rather reuses existing fault detection mechanisms. The following mechanisms may be used by ICCP to detect PE node failure:

- BFD

Run a BFD session [BFD] between the PEs that are members of a given RG, and use that to detect PE node failure. This assumes...
that resiliency mechanisms are in place to protect connectivity to the remote PE nodes, and hence loss of BFD periodic messages from a given PE node can only mean that the node itself has failed.

- IP Reachability Monitoring

It is possible for a PE to monitor IP layer connectivity to other members of a RG that are participating in IGP/BGP. When connectivity to a given PE is lost, the local PE interprets that to mean loss of the remote PE node. This assumes that resiliency mechanisms are in place to protect the route to the remote PE nodes, and hence loss of IP reachability to a given node can only mean that the node itself has failed.

It is worth noting here that loss of the LDP session with a PE in a RG is not a reliable indicator that the remote PE itself is down. It is possible, for e.g. that the remote PE encounters a local event that leads to resetting the LDP session, while the PE node remains operational for purpose of traffic forwarding.

7. ICCP Message Formats

This section defines the messages exchanged at the Application and ICC layers.

7.1. Encoding ICC into LDP Messages

ICCP requires reliable, in order, state-full message delivery, as well as capability negotiation between PEs. The LDP protocol offers all these features, and is already in wide use in the applications that would also require the ICCP protocol extensions. For these reasons, ICCP takes advantage of the already defined LDP protocol infrastructure. [RFC5036] Section 3.5 defines a generic LDP message structure. A new set of LDP message types is defined to communicate the ICCP information. LDP message types in the range of 0x700 to 0x7ff will be used for ICCP.

Message types are allocated by IANA, and requested in the IANA section below.
### 7.1.1. ICC Header

Every ICCP message comprises of an ICC specific LDP Header followed by an ICCP message. The format of the ICC Header is as follows:

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|U|   Message Type              |      Message Length           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Message ID                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type=0x0005 (ICC RG ID)     |           Length=4            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          ICC RG ID                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Mandatory Parameters                      |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Optional Parameters                       |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- **U-bit**
  
  Unknown message bit. Upon receipt of an unknown message, if U is clear (=0), a notification is returned to the message originator; if U is set (=1), the unknown message is silently ignored. The following sections which define messages specify a value for the U-bit.

- **Message Type**

  Identifies the type of the ICCP message, must be in the range of 0x0700 to 0x07ff.
- Message Length

Two octet integer specifying the total length of this message in octets, excluding the U-bit, Message Type and Length fields.

- Message ID

Four octet value used to identify this message. Used by the sending PE to facilitate identifying RG Notification messages that may apply to this message. A PE sending a RG Notification message in response to this message SHOULD include this Message ID in the "NAK TLV" of the RG Notification message; see Section "RG Notification Message".

- ICC RG ID

A TLV of type 0x0005, length 4, containing 4 octects unsigned integer designating the Redundancy Group which the sending device is member of. RG ID value 0x00000000 is reserved by the protocol.

- Mandatory Parameters

Variable length set of required message parameters. Some messages have no required parameters.

For messages that have required parameters, the required parameters MUST appear in the order specified by the individual message specifications in the sections that follow.

- Optional Parameters

Variable length set of optional message parameters. Many messages have no optional parameters.

For messages that have optional parameters, the optional parameters may appear in any order.

7.1.2. Message Encoding

The generic format of an ICC parameter is:
- **U-bit**

  Unknown TLV bit. Upon receipt of an unknown TLV, if U is clear (=0), a notification MUST be returned to the message originator and the entire message MUST be ignored; if U is set (=1), the unknown TLV MUST be silently ignored and the rest of the message processed as if the unknown TLV did not exist. The sections following that define TLVs specify a value for the U-bit.

- **F-bit**

  Forward unknown TLV bit. This bit applies only when the U-bit is set and the LDP message containing the unknown TLV is to be forwarded. If F is clear (=0), the unknown TLV is not forwarded with the containing message; if F is set (=1), the unknown TLV is forwarded with the containing message. The sections following that define TLVs specify a value for the F-bit. By setting both the U- and F-bits, a TLV can be propagated as opaque data through nodes that do not recognize the TLV.

- **Type**

  Fourteen bits indicating the parameter type.

- **Length**

  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- **TLV(s)**: A set of 0 or more TLVs, that vary according to the message type.
7.2. RG Connect Message

The RG Connect Message is used to establish ICCP connection in addition to individual Application connections between PEs in a RG: a RG Connect message with no "Application-specific connect TLVs" signals establishment of the base ICCP connection. RG Connect messages with appropriate "Application-specific connect TLVs" signal the establishment of Application connections, in addition to the base ICCP connection (if not already established). A PE sends a RG Connect Message to declare its membership in a Redundancy Group. One such message should be sent to each PE that is member of the same RG. The set of PEs to which RG Connect Messages should be transmitted is known via configuration or an auto-discovery mechanism that is outside the scope of this specification. If a device is member of multiple RGs, it must send separate RG Connect Messages for each RG even if the receiving device(s) happen to be the same.

The format of the RG Connect Message is as follows:

- i. ICC header with Message type = "RG Connect Message" (0x0700)
- ii. "Sender Name TLV"
- iii. "Application specific connect TLV"

The currently defined Application-specific connect TLVs are:

- PW Redundancy Connect TLV
- mLACP Connect TLV

The details of these TLVs are discussed in the "Application TLVs" section.

The RG Connect message can contain zero, one or more Application-specific connect TLVs. Multiple application connect TLVs can be sent in a single message, or multiple messages can be sent containing different application connect TLVs, but no application connect TLV can be sent more than once.

7.2.1. Sender Name TLV

A TLV that carries the hostname of the sender encoded in UTF-8. This is used primarily for purpose of management of the RG and easing network operations. The specific format is shown below:
7.3. RG Disconnect Message

The RG Disconnect Message serves dual-purpose: to signal that a particular Application connection is being closed within a RG, or that the ICCP connection itself is being disconnected because the PE wishes to leave the RG. The format of this message is:

\[
\begin{array}{c|c|c}
0 & 1 & 2 & 3 \\
+--------------------------------+----------------+
|U|F|                   Type|     Length                        |
+--------------------------------+----------------+
|                     Sender Name |
+----------------+
|      ...                                    |
\end{array}
\]

- U=F=0
- Type set to "ICC sender name" Parameter type (from ICC parameter name space).
- Length
  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.
- Sender Name
  Hostname of sending device encoded in UTF-8, and SHOULD NOT exceed 80 characters.
- **U-bit**
  
  U=0

- **Message Type**

  The message type for RG Disconnect Message is set to (0x0701)

- **Length**

  Length of the TLV in octets excluding the U-bit, Message Type, and Message Length fields.

- **Message ID**

  Defined in the "ICC Header" section above.

- **ICC RG ID**

  Defined in the "ICC Header" section above.
- Disconnect Code TLV

The format of this TLV is as follows:

```
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|U|F|         Type=0x0004       |    Length                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      ICCP Status Code                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- U,F Bits
both U and F are set to 0.

- Type
set to "Disconnect Code TLV" (0x0004)

- Length
Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- ICCP Status Code
A status code that reflects the reason for the disconnect message. Allowed values are "ICCP RG Removed" and "ICCP Application Removed from RG".

- Optional Application-specific Disconnect TLVs
Zero, one or more Application-specific Disconnect TLVs which are defined later in the document. If the RG Disconnect message has a status code of "RG Removed", then it should not contain any Application-specific Disconnect TLVs, as the sending PE is signaling that it has left the RG and, thus, is disconnecting the entire ICCP connection, with all associated client application connections. If the message has a status code of "Application Removed from RG", then it should contain one or more Application-specific Disconnect TLVs, as the sending PE is only tearing down the connection for the specified applications. Other applications, and the base ICCP connection are not to be affected.
7.4. RG Notification Message

A PE sends a RG Notification Message to indicate one of the following: to reject an ICCP connection, to reject an application connection, to NAK an entire message or to NAK one or more TLV(s) within a message. The Notification message can only be sent to a PE that is already part of a RG.

The format of the Notification Message is:

- i. ICC header with Message type = "RG Notification Message" (0x0702)
- ii. Notification Message TLVs.

The currently defined Notification message TLVs are:

- i. Sender Name TLV
- ii. NAK TLV.

7.4.1. Notification Message TLVs

The Sender Name TLV uses the same format as in the RG Connect message, and was described above.

The NAK TLV is defined as follows:

```
+-----------------+-----------------+-----------------+-----------------+
 | U | F |             |             |
 +-----------------+-----------------+-----------------+-----------------+  
 | Type=0x0002    | Length          |                |
 +-----------------+-----------------+-----------------+-----------------+  
 | ICCP Status Code|
 +-----------------+-----------------+-----------------+-----------------+  
 | Rejected Message ID|
 +-----------------+-----------------+-----------------+-----------------+  
 | Optional TLV(s) |
 +-----------------+-----------------+-----------------+-----------------+  

Martini, et al. [Page 23]
```
- **U,F Bits**
  
  both \( U \) and \( F \) are set to 0.

- **Type**
  
  set to "NAK TLV" (0x0002)

- **Length**
  
  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- **ICCP Status Code**
  
  A status code that reflects the reason for the NAK TLV. Allowed values are:

  - **i. Unknown RG (0x00010001)**
    
    This code is used to reject a new incoming ICCP connection for a RG that is not configured on the local PE. When this code is used, the Rejected Message ID field must contain the message ID of the rejected "RG Connect" message.

  - **ii. ICCP Connection Count Exceeded (0x00010002)**
    
    This is used to reject a new incoming ICCP connection that would cause the local PE’s ICCP connection count to exceed its capabilities. When this code is used, the Rejected Message ID field must contain the message ID of the rejected "RG Connect" message.

  - **iii. Application Connection Count Exceeded (0x00010003)**
    
    This is used to reject a new incoming application connection that would cause the local PE’s ICCP connection count to exceed its capabilities. When this code is used, the Rejected Message ID field must contain the message ID of the rejected "RG Connect" message and the corresponding Application Connect TLV must be included in the "Optional TLV".

  - **iv. Application not in RG (0x00010004)**
    
    This is used to reject a new incoming application connection when the local PE doesn’t support the application, or the application is not configured in the
RG. When this code is used, the Rejected Message ID field must contain the message ID of the rejected "RG Connect" message and the corresponding Application Connect TLV must be included in the "Optional TLV".

-v. Incompatible Protocol Version (0x00010005)

This is used to reject a new incoming application connection when the local PE has an incompatible version of the application. When this code is used, the Rejected Message ID field must contain the message ID of the rejected "RG Connect" message and the corresponding Application Connect TLV must be included in the "Optional TLV".

-vi. Rejected Message (0x00010006)

This is used to reject a RG Application Data message, or one or more TLV(s) within the message. When this code is used, the Rejected Message ID field must contain the message ID of the rejected "RG Application Data" message.

-vii. ICCP Administratively Disabled (0x00010007)

This is used to reject any ICCP messages from a peer from which the PE is not allowed to exchange ICCP messages due to local administrative policy.

- Rejected Message ID

If non-zero, 32-bit value that identifies the peer message to which the NAK TLV refers. If zero, no specific peer message is being identified.

- Optional TLV(s)

A set of one or more optional TLVs. If the status code is "Rejected Message" then this field contains the TLV(s) that were rejected. If the entire message is rejected, all its TLVs MUST be present in this field; otherwise, the subset of TLVs that were rejected MUST be echoed in this field.

If the status code is "Incompatible Protocol Version" then this field contains the "Requested Protocol Version TLV" defined as follows:
- **U and F Bits**
  
  Both are set to 0.

- **Type**
  
  set to 0x0003 for "Requested Protocol Version TLV"

- **Length**
  
  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- **Connection Reference**
  
  This field is set to the Type field of the Application specific Connect TLV that was rejected because of incompatible version.

- **Requested Version**
  
  The version of the application supported by the transmitting device. For this version of the protocol it is set to 0x0001.

### 7.5. RG Application Data Message

The RG Application Data Message is used to transport application data between PEs within a RG. A single message can be used to carry data from multiple applications, as long as all these applications are part of the same RG. Such multiplexing is possible because the transported TLVs are application specific which allows for identifying the target application for each TLV at the receiving side. The format of the Application Data Message is:

- i. ICC header with Message type = "RG Application Data Message" (0x703)
ii. "Application specific TLVs"

The details of these TLVs are discussed in the "Application TLVs" section.

8. Application TLVs

8.1. Pseudowire Redundancy (PW-RED) Application TLVs

This section discusses the ICCP TLVs for the Pseudowire Redundancy application.

8.1.1. PW-RED Connect TLV

This TLV is included in the RG Connect message to signal the establishment of PW-RED application connection.

```
| U | F | Type=0x0010 | Length |
+---+---+------------+--------+
|   |   |            |        |
+---+---+------------+--------+
| Protocol Version | Optional Sub-TLVs |
|                 |                   |
+-----------------+------------------+
|                 |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
|                  |                  |
+-----------------+------------------+
```

- **U and F Bits**
  - Both are set to 0.
- **Type**
  - set to 0x0010 for "PW-RED Connect TLV"
- **Length**
  - Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.
- Protocol Version

The version of this particular protocol for the purposes of ICCP. This is set to 0x0001.

- Optional Sub-TLVs

There are no optional Sub-TLVs defined for this version of the protocol.

8.1.2. PW-RED Disconnect TLV

This TLV is used in a RG Disconnect Message to indicate that the connection for the PW-RED application is to be terminated.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|U|F|   Type=0x0011             |    Length                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Optional Sub-TLVs                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- U and F Bits

Both are set to 0.

- Type

set to 0x0011 for "PW-RED Disconnect TLV"

- Length

Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- Optional Sub-TLVs

There are no optional Sub-TLVs defined for this version of the protocol.
8.1.3. PW-RED Config TLV

The PW-RED Config TLV is used in RG Application Data message and is composed of the following TLVs in the following order:

- i. Service Name TLV
- ii. PW ID TLV or Generalized PW ID TLV

In the PW-RED Config TLV the U and F Bits are both set to 0, and the TLV type is set to 0x0012.

8.1.4. Service Name TLV

```
+---------- 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
|U|F|   Type                    |    Length                     |
+-----------------------------------------------+
|                        Service Name                           |
+-----------------------------------------------+
```

- U and F Bits
  Both are set to 0.
- Type
  set to 0x0013 for "Service Name TLV"
- Length
  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.
- Service Name
  The name of the L2VPN service instance encoded in UTF-8 format and up to 80 character in length.

8.1.5. PW ID TLV

This TLV is used to communicate the configuration of PWs for VPWS.
- U and F Bits
  Both are set to 0.

- Type
  set to 0x0014 for "PW ID TLV"

- Length
  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- Peer ID
  Four octet LDP Router ID of the peer at the far end of the PW.

- Group ID
  Same as Group ID in [RFC4447] section 5.2.

- PW ID
  Same as PW ID in [RFC4447] section 5.2.

8.1.6. Generalized PW ID TLV

This TLV is used to communicate the configuration of PWs for VPLS.
8.2. Multi-chassis LACP (mLACP) Application TLVs

This section discusses the ICCP TLVs for Ethernet attachment circuit redundancy using the multi-chassis LACP (mLACP) application.
8.2.1. mLACP Connect TLV

This TLV is included in the RG Connect message to signal the establishment of mLACP application connection.

```
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|U|F|   Type=0x0030             |    Length                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Protocol Version         |   Optional Sub-TLVs           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
~                                                               ~
~                                                               ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             ...                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- **U and F Bits**
  
  Both are set to 0.

- **Type**
  
  set to 0x0030 for "mLACP Connect TLV"

- **Length**
  
  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- **Protocol Version**
  
  The version of this particular protocol for the purposes of ICCP. This is set to 0x0001.

- **Optional Sub-TLVs**
  
  There are no optional Sub-TLVs defined for this version of the protocol.
8.2.2. mLACP Disconnect TLV

This TLV is used in a RG Disconnect Message to indicate that the connection for the mLACP application is to be terminated.

```
          0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+----------------------------------+-+-+-+-+-+-+-+-+-
|U|F|   Type=0x0031             |    Length                     |
+----------------------------------+-+-+-+-+-+-+-+-+-
|                       Optional Sub-TLVs                       |
+----------------------------------+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- **U and F Bits**
  - Both are set to 0.
- **Type**
  - set to 0x0031 for "mLACP Disconnect TLV"
- **Length**
  - Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.
- **Optional Sub-TLVs**
  - There are no optional Sub-TLVs defined for this version of the protocol.

8.2.3. mLACP System Config TLV

The mLACP System Config TLV is sent in the RG Application Data message. This TLV announces the local node’s LACP System Parameters to the RG peers.
Both are set to 0.

- **Type**
  
  Set to 0x0032 for "mLACP System Config TLV"

- **Length**
  
  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- **System ID**
  
  6 octets field encoding the System ID used by LACP as specified in [IEEE-802.3] section 43.3.2.

- **System Priority**
  
  2 octets encoding the LACP System Priority as defined in [IEEE-802.3] section 43.3.2.

- **Node ID**
  
  One octet, LACP node ID. Used to ensure that the LACP Port IDs are unique across all devices in a RG. Valid values are in the range 0 - 7.
8.2.4. mLACP Port Config TLV

The mLACP Port Config TLV is sent in the RG Application Data message. This TLV is used to notify RG peers about the local configuration state of a port.

```
+-----+-+-+-+-+-+-+-+
| U | F |   Type=0x0033 |    Length                     |
+-----+-+-+-+-+-+-+-+
|       Port Number |    MAC Address                |
+-------------------+                               |
+---------------------------------------------------------------+
|       Admin Key |     Port Priority             |
+-----------------+---------------------------------+
|                |                                |
+-------------------+                               |
| Flags | IF Name Len |    Interface Name             |
| ~ |  ~ |                                      |
+-------------------+                               |
```

- **U and F Bits**
  
  Both are set to 0.

- **Type**

  set to 0x0033 for "mLACP Port Config TLV"

- **Length**

  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- **Port Number**

  Two octets, LACP Port Number for the corresponding interface as specified in [IEEE-802.3] section 43.3.4. When the value of this field is 0, it denotes the Aggregator whose key is specified in the "Admin Key" field.

- **MAC Address**

  Six octets encoding the port MAC address.
- Admin Key

  Two octets, LACP Admin key for the corresponding interface, as specified in [IEEE-802.3] section 43.3.5.

- Port Priority

  Two octets, LACP port priority for the corresponding interface, as specified in [IEEE-802.3] section 43.3.4. This field is valid only when the "Flags" field has "Priority Set" asserted.

- Flags

  Valid values are:

  -i. Synchronized (0x01)

    Indicates that the sender has concluded transmitting all member link port configurations for a given Aggregator. Also, indicates to the receiving device that its local port priorities will not be overridden.

  -ii. Purge Configuration (0x02)

    Indicates that the port is no longer configured for mLACP operation.

  -iii. Priority Set (0x04)

    Indicates that the "Port Priority" field is valid.

- IF Name Len

  One octet, length of the "Interface Name" field in octets.

- Interface Name

  Interface name encoded in UTF-8 format, up to a maximum of 20 characters.

8.2.5. mLACP Change Port Priority TLV

  The mLACP Port State TLV is used in RG Application Data message. This TLV is used by a device to authoritatively request that a particular member of a RG change its port priority.
8.2.6. mLACP Port State TLV

The mLACP Port State TLV is used in RG Application Data message. This TLV is used by a device to report its LACP port status to other members in the RG.
Both are set to 0.

- Type
  set to 0x0035 for "mLACP Port State TLV"

- Length
  Length of the TLV in octets excluding the U-bit, F-bit, Type, and Length fields.

- LACP Partner System ID
  6 octets, the LACP Partner System ID for the corresponding interface, encoded as a MAC address as specified in [IEEE-802.3] section 43.4.2.2 item r.

- LACP Partner System Priority
  2 octets field specifying the LACP Partner System Priority as specified in [IEEE-802.3] section 43.4.2.2 item q.

- LACP Partner Port Number
  2 octets encoding the LACP Partner Port Number as specified in [IEEE-802.3] section 43.4.2.2 item u.
- LACP Partner Port Priority

  2 octets field encoding the LACP Partner Port Priority as specified in [IEEE-802.3] section 43.4.2.2 item t.

- LACP Partner Oper Key

  2 octets field representing the LACP Partner Key as defined in [IEEE-802.3] section 43.4.2.2 item s.

- LACP Partner Oper State

  1 octet field encoding the LACP Partner State Variable as defined in [IEEE-802.3] section 43.4.2.2 item v.

- LACP Actor Oper State

  1 octet encoding the LACP Actor’s State Variable for the port as specified in [IEEE-802.3] section 43.4.2.2 item m.

- LACP Actor Port Number

  2 octets field representing the LACP Actor Port Number as specified in [IEEE-802.3] section 43.3.4. When the value of this field is 0, it denotes the Aggregator whose key is specified in the "Actor Operational Key" field.

- LACP Actor Oper Key

  2 octet field encoding the LACP Actor Operational Key as specified in [IEEE-802.3] section 43.3.5.

- Port State

  1 octet encoding the operational state of the port as follows:
  - 0x00 Up
  - 0x01 Down
  - 0x02 Administrative Down
  - 0x03 Test (e.g. IEEE 802.3ah OAM Intrusive Loopback mode)
9. LDP Capability Negotiation

As required in [LDP-CAP] the following TLV is defined to indicate the ICCP capability:

```
+--------------------------------+  +--------------------------------+
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 |
+--------------------------------+  +--------------------------------+
|U|F|  TLV Code Point=0x405         |
|S|  Reserved | Reserved               |
|  VER/Maj |  Ver/Min                |
+--------------------------------+  +--------------------------------+
```

where:

- **U-bit**
  - SHOULD be 1 (ignore if not understood).

- **F-bit**
  - SHOULD be 0 (don’t forward if not understood).

- **TLV Code Point**
  - The TLV type, which identifies a specific capability. For the ICCP code point is requested in the IANA allocation section below.

- **S-bit** The State Bit indicates whether the sender is advertising or withdrawing the ICCP capability. The State bit is used as follows:
  - 1 - The TLV is advertising the capability specified by the TLV Code Point.
  - 0 - The TLV is withdrawing the capability specified by the TLV Code Point.

- **Ver/Maj**
  - The major version revision of the ICCP protocol, this document specifies 1.0. This field is then set to 1

- **Ver/Min**
  - The minor version revision of the ICCP protocol, this document specifies 1.0. This field is then set to 0
ICCP capability is advertised to a LDP peer if there is at least one RG enabled on the local PE.

10. Client Applications

10.1. Pseudowire Redundancy Application Procedures

This section defines the procedures for the Pseudowire Redundancy (PW-RED) Application.

10.1.1. Initial Setup

When a RG is configured on a system and multi-chassis pseudowire redundancy is enabled in that RG, the PW-RED application should send an "RG Connect" message with "PW-RED Connect TLV" to each PE that is member of the same RG. When the system receives similar "RG Connect" messages from a PE, the two devices can start exchanging "RG Application Data" messages for the PW-RED application.

If a system receives an "RG Connect" message with "PW-RED Connect TLV" that has a differing Protocol Version, it must follow the procedures outlined in the "Application Versioning" section above.

When the PW-RED application is disabled on the device, or is unconfigured for the RG in question, the system should send an "RG Disconnect" message with "PW-RED Disconnect TLV".

10.1.2. Pseudowire Configuration

A system should advertise its local PW configuration to other PEs that are members of the same RG. This allows the PEs to build a view of the redundant nodes and pseudowires that are protecting the same service instances. The advertisement should be initiated when the PW-RED application connection first comes up, as well as upon any subsequent PW configuration change. To that end, the system should send "RG Application Data" messages with "PW-RED Config TLV". It is possible to send configuration information for multiple PWs in a single "RG Application Data" message.

The "Service Name TLV" is used on the receiving system for the purpose of associating PW information advertised by some PE with the corresponding AC information received over ICCP from that PE’s AC redundancy application. The Service Name has a global context in a RG, so redundant PWs for the same service on disparate member PEs should share the same Service Name, in order to be correlated.
10.1.3. Pseudowire Status Synchronization

On a given PE, the forwarding status of the PW (Active or Standby) is derived from the state of the associated AC(s). This simplifies the operation of the multi-chassis redundancy solution (Figure 1) and eliminates the possibility of deadlock conditions between the AC and PW redundancy mechanisms. The rules by which the PW state is derived from the AC state are as follows:

- **VPWS**
  
  For VPWS, there’s a single AC per service instance. If the AC is Active, then the PW status should be Active. If the AC is Standby, then the PW status should be Standby.

- **VPLS**
  
  For VPLS, there could be multiple ACs per service instance (i.e. VFI). If AT LEAST ONE AC is Active, then the PW status should be Active. If ALL ACs are Standby, then the PW status should be Standby.

The PW-RED application does not synchronize PW status across chassis, per se. Rather, the AC Redundancy application should synchronize AC status between chassis, in order to determine which AC (and subsequently which PE) is Active or Standby for a given service. When that is determined, each PE will then adjust its local PWs state according to the rules described above.

10.1.4. PE Node Failure

When a PE node detects that a remote PE, that is member of the same RG, has gone down, the local PE examines if it has redundant PWs for the affected services. If the local PE has the highest priority (after the failed PE) then it becomes the active node for the services in question, and subsequently activates its associated PWs.

10.2. Attachment Circuit Redundancy Application Procedures

10.2.1. Common AC Procedures

This section describes generic procedures for AC Redundancy applications, independent of the type of the AC (ATM, FR or Ethernet).
10.2.2. AC Failure

When the AC Redundancy mechanism on the Active PE detects a failure of the AC, it should send an ICCP Application Data message to inform the redundant PEs of the need to take over. The AC failures can be categorized into the following scenarios:

- Failure of CE interface connecting to PE
- Failure of CE uplink to PE
- Failure of PE interface connecting to CE

10.2.3. PE Node Failure

When a PE node detects that a remote PE, that is member of the same RG, has gone down, the local PE examines if it has redundant ACs for the affected services. If the local PE has the highest priority (after the failed PE) then it becomes the active node for the services in question, and subsequently activates its associated ACs.

10.2.4. PE Isolation

When a PE node detects that it has been isolated from the core network (i.e. all core facing interfaces/links are not operational), then it should instruct its AC Redundancy mechanism to change the status of any active ACs to Standby. The AC Redundancy application should then send ICCP Application Data messages in order to trigger failover to a standby PE.

10.2.5. ATM AC Procedures

10.2.6. Frame Relay AC Procedures

10.2.7. Ethernet AC Procedures

10.2.8. Multi-chassis LACP (mLACP) Application Procedures

This section defines the procedures that are specific to the multi-chassis LACP (mLACP) application.
10.2.8.1. Initial Setup

When a RG is configured on a system and mLACP is enabled in that RG, the mLACP application should send an "RG Connect" message with "mLACP Connect TLV" to each PE that is member of the same RG. When the system receives similar "RG Connect" message from a PE, the two devices can start exchanging "RG Application Data" messages for the mLACP application.

If a system receives an "RG Connect" message with "mLACP Connect TLV" that has a differing Protocol Version, it must follow the procedures outlined in the "Application Versioning" section above.

After the mLACP application connection has been established, every PE must communicate its system level configuration to its peers via the use of "mLACP System Config TLV". This allows every PE to discover the Node ID and the locally configured System ID and System Priority values of its peers. It is necessary for all PEs in a RG to agree upon the System ID and System Priority values to be used ubiquitously. To achieve this, every PE MUST use the numerically lowest value (among RG members) for each of the two parameters. This guarantees that the PEs always agree on uniform values, which yield the highest System Priority.

When the mLACP application is disabled on the device, or is unconfigured for the RG in question, the system should send an "RG Disconnect" message with "mLACP Disconnect TLV".

10.2.8.2. mLACP Port Configuration

A system must synchronize the configuration of its mLACP operating ports with other RG members. To that end, a system must use the "Port Config TLVs". An implementation must advertise the configuration of Aggregators prior to advertising the configuration of any of their associated member links. Aggregators are identified by using the Port Number 0 (which is not a valid LACP port number) and the associated Key. If the "Priority Set" flag is asserted in such TLV, it indicates that the same Port Priority applies to all member links that are attached to the Aggregator in question. When the configuration of all ports for member links associated with a given Aggregator has been sent by a device, it asserts that fact by setting the "Synchronized" flag in the last port’s "Port Config TLV". This also serves as a cue for the receiving system that its local port priorities will not be remotely overridden by the sending PE.

Furthermore, for a given port, an implementation must advertise the port’s configuration prior to advertising its state (via the "mLACP
Port State TLV").

When mLACP is unconfigured on a port, a PE must send a "Port Config TLV" with the "Purge Configuration" flag asserted. This allows receiving PEs to purge any state maintained for the decommissioned port.

10.2.8.3. mLACP Port Status Synchronization

PEs within a RG need to synchronize their state-machines for proper mLACP operation with a multi-homed device. This is achieved by having each system advertise its ports’ running state in "mLACP Port State TLVs". Whenever any port parameter, whether on the Partner (i.e. multi-homed device) or the Actor (i.e. PE) side, is changed a system MUST transmit an updated "mLACP Port State TLV" for the affected port.

10.2.8.4. Triggering Failover

A PE MAY trigger a failover to a redundant PE within the RG by sending an "mLACP Change Port Priority TLV" specifying the affected Aggregator and a priority value that causes the remote PE to have a higher Port Priority thereby moving to active forwarding state.

A PE MAY assume active role within the RG by sending an "mLACP Change Port Priority TLV" to the currently active PE, specifying the affected Aggregator and a port priority value that is less than its local port priority for the links associated with that Aggregator.

11. Security Considerations

The security considerations described in [RFC5036] and [RFC4447] that apply to the base LDP specification, and to the PW LDP control protocol extensions apply to the capability mechanism described in this document.

The ICCP protocol is not intended to be applicable when the redundancy group spans PE in different administrative domains. Furthermore, implementations MUST provide a mechanism to select to which LDP peers the ICCP capability will be advertised, and from which LDP peers the ICCP messages will be accepted.
12. IANA Considerations

12.1. MESSAGE TYPE NAME SPACE

This document uses several new LDP message types, IANA already maintains a registry of name "MESSAGE TYPE NAME SPACE" defined by [RFC5036]. The following values are suggested for assignment:

<table>
<thead>
<tr>
<th>Message type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0700</td>
<td>RG Connect Message</td>
</tr>
<tr>
<td>0x0701</td>
<td>RG Disconnect Message</td>
</tr>
<tr>
<td>0x0702</td>
<td>RG Notification Message</td>
</tr>
<tr>
<td>0x0703</td>
<td>RG Application Data Message</td>
</tr>
</tbody>
</table>

12.2. TLV TYPE NAME SPACE

This document uses a new LDP TLV type, IANA already maintains a registry of name "TLV TYPE NAME SPACE" defined by [RFC5036]. The following value is suggested for assignment:

<table>
<thead>
<tr>
<th>TLV Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x700</td>
<td>ICCP capability TLV.</td>
</tr>
<tr>
<td>0x701</td>
<td>LDP TCP/IP Port TLV.</td>
</tr>
</tbody>
</table>

12.3. ICC RG Parameter Type Space

IANA needs to set up a registry of "ICC RG parameter type". These are 14-bit values. Parameter Type values 1 through 0x000F are specified in this document, Parameter Type values 0x0010 through 0x1FFF are to be assigned by IANA, using the "Expert Review" policy defined in [RFC5226]. Parameter Type values 0x2000 through 0x2FFF, 0x3FFF, and 0 are to be allocated using the IETF consensus policy defined in [RFC5226]. Parameter Type values 0x3000 through 0x3FFE are reserved for vendor proprietary extensions and are to be assigned by IANA, using the "First Come First Served" policy defined in [RFC5226]. A Parameter Type description is required for any assignment from this registry. Additionally, for the vendor proprietary extensions range a citation of a person or company name is also required. A document reference should also be provided.

Initial ICC RG parameter type space value allocations are specified below:

<table>
<thead>
<tr>
<th>Parameter Type</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0001</td>
<td>ICC sender name</td>
<td>[RFCxxxx]</td>
</tr>
<tr>
<td>0x0002</td>
<td>NAK TLV</td>
<td>[RFCxxxx]</td>
</tr>
</tbody>
</table>
12.4. STATUS CODE NAME SPACE

This document uses several new Status codes, IANA already maintains a registry of name "STATUS CODE NAME SPACE" defined by [RFC5036]. The following values is suggested for assignment: The "E" column is the required setting of the Status Code E-bit.

<table>
<thead>
<tr>
<th>Range/Value</th>
<th>E</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x00010001</td>
<td>0</td>
<td>Unknown ICCP RG</td>
<td></td>
</tr>
<tr>
<td>0x00010002</td>
<td>0</td>
<td>ICCP Connection Count Exceeded</td>
<td></td>
</tr>
<tr>
<td>0x00010003</td>
<td>0</td>
<td>ICCP Application Connection Count Exceeded</td>
<td></td>
</tr>
<tr>
<td>0x00010004</td>
<td>0</td>
<td>ICCP Application not in RG</td>
<td></td>
</tr>
<tr>
<td>0x00010005</td>
<td>0</td>
<td>Incompatible ICCP Protocol Version</td>
<td></td>
</tr>
<tr>
<td>0x00010006</td>
<td>0</td>
<td>ICCP Rejected Message</td>
<td></td>
</tr>
<tr>
<td>0x00010007</td>
<td>0</td>
<td>ICCP Administratively Disabled</td>
<td></td>
</tr>
<tr>
<td>0x00010010</td>
<td>0</td>
<td>ICCP RG Removed</td>
<td></td>
</tr>
<tr>
<td>0x00010011</td>
<td>0</td>
<td>ICCP Application Removed from RG</td>
<td></td>
</tr>
</tbody>
</table>
13. Normative References


14. Informative References


15. Author’s Addresses

Luca Martini
Cisco Systems, Inc.
9155 East Nichols Avenue, Suite 400
Englewood, CO, 80112
e-mail: lmartini@cisco.com

Samer Salam
Cisco Systems, Inc.
595 Burrard Street, Suite 2123
Vancouver, BC V7X 1J1
Canada
e-mail: ssalam@cisco.com
Ali Sajassi  
Cisco Systems, Inc.  
170 West Tasman Drive  
San Jose, CA 95134  
e-mail: sajassi@cisco.com

Satoru Matsushima  
Softbank Telecom  
1-9-1, Higashi-Shinbashi, Minato-ku  
Tokyo 105-7313, JAPAN  
e-mail: satoru.matsushima@tm.softbank.co.jp

Thomas D. Nadeau  
BT  
BT Centre  
81 Newgate Street  
London, EC1A 7AJ  
United Kingdom  
e-mail: tom.nadeau@bt.com

Full Copyright Statement

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents in effect on the date of publication of this document (http://trustee.ietf.org/licenseinfo). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.
Acknowledgments

The authors wish to acknowledge the contributions of Satoru Matsushima, Wei Luo, Neil Mcgill, Skip Booth, Neil Hart, Michael Hua, and Tiberiu Grigoriu.

Expiration Date: August 2009