Internet Engineering Task Force (IETF)
Request for Comments: 6307
Category: Standards Track
ISSN: 2070-1721

D. Black, Ed.
EMC Corporation
L. Dunbar, Ed.
Huawei Technologies
M. Roth
Infinera
R. Solomon
Orckit-Corrigent

April 2012

Encapsulation Methods for Transport of Fibre Channel Traffic over MPLS Networks

Abstract

A Fibre Channel pseudowire (PW) is used to carry Fibre Channel traffic over an MPLS network. This enables service providers to take advantage of MPLS to offer "emulated" Fibre Channel services. This document specifies the encapsulation of Fibre Channel traffic within a pseudowire. It also specifies the common procedures for using a PW to provide a Fibre Channel service.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6307.

Table of Contents

1. Introduction

Fibre Channel (FC) is a high-speed communications technology, used primarily for Storage Area Networks (SANs). Within a single site (e.g., data center), an FC-based SAN connects servers to storage systems, and FC can be extended across sites. When FC is extended across multiple sites, the most common usage is storage replication in support of recovery from disasters (e.g., flood or fire that takes a site out of operation). This is particularly the case over longer distances where network latency results in unacceptable performance for a server whose storage is not at the same site. Fibre Channel is standardized by the INternational Committee for Information Technology Standards (INCITS) Technical Committee T11 [T11], and multiple methods for encapsulating and transporting FC traffic over other networks have been developed [FC-BB-6].

Fibre Channel Over TCP/IP (FCIP), as described in [RFC3821] and [FC-BB-6], interconnects otherwise isolated FC SANs over IP Networks. FCIP uses FC Frame Encapsulation [RFC3643] to encapsulate FC frames for tunneling over an IP-based network. Since IP networks may drop or reorder packets, FCIP relies on TCP to retransmit dropped frames and restore the delivery order of reordered frames. Due to possible delay variation and TCP timeouts, special timing mechanisms are required to ensure correct Fibre Channel operation over FCIP [FC-BB-6].

MPLS networks can be provisioned and operated with very low loss rates and very low probability of reordering, making it possible to directly interconnect Fibre Channel ports over MPLS. A Fibre Channel pseudowire (FC PW) is a method to transparently transport FC traffic over an MPLS network resulting in behavior similar to a pair of FC ports that are directly connected by a physical FC link. The result is simpler control processing in comparison to FCIP.

This document specifies the encapsulation of FC traffic into an MPLS pseudowire and related PW procedures to transport FC traffic over MPLS PWs. The complete FC pseudowire specification consists of this document and the FC PW portion of the T11 [FC-BB-5/AM1] standard. The following subsections describe some of the requirements for transporting FC traffic over an MPLS network.

1.1. Transparency

Transparent extension of an FC link is a key requirement for transporting FC traffic over a PW. This requires the FC PW to emulate an FC link between two FC ports, similar to the approach defined for FC over GFPT in [FC-BB-6]. GFPT is an Asynchronous Transparent Generic Framing Procedure specified by ITU-T; see

[FC-BB-6] for details and reference to the ITU-T specifications. This results in transparent forwarding of FC traffic over the MPLS network from both the FC fabric and the network operator points of view.

Transparency distinguishes the FC PW approach from FCIP. An FC PW logically connects the FC port on the FC link attached to one end of the PW directly with the FC port on the far end of the FC link attached to the other end of the PW, whereas FCIP introduces FC B_Ports at both ends of the extended FC link; each FC B_Port is connected to an FC E_Port in an FC switch on the same side of the link extension.

1.2. Bandwidth Efficiency

The bandwidth allocated to a PW may be less than the rate of the attached FC port. When there is no data exchange on a native FC link, Idle Primitive Signals are continuously exchanged between the two FC ports. In order to improve the bandwidth efficiency across the MPLS network, it is necessary for the FC PW Provider Edge (PE) to suppress (or drop) the Idle Primitive Signals generated by its adjacent FC ports. The far-end FC PW PE regenerates Idle Primitive Signals to send to its adjacent FC port as required; see [FC-BB-5/AM1].

FC link control protocols require an FC port to continuously send the same FC Primitive Sequence [FC-FS-2] until a reply is received or some other event occurs. To improve bandwidth efficiency, the FC PW PE encapsulates a subset of repeated FC Primitive Sequences to send across the WAN [FC-BB-5/AM1]. For example, in a sequence of identical received primitives, only every fourth primitive may be sent across the MPLS network. Alternatively, a time-based approach may be used to send a copy of the repeated FC Primitive Sequence once every few milliseconds. The far-end FC PW PE regenerates the FC link behavior by continuously sending the Primitive Sequence most recently received from the WAN until a new primitive signal, primitive sequence, or data frame is received from the WAN.

The sending FC PW PE may unilaterally choose any convenient subset for sending the same FC Primitive Sequence. This is acceptable because the receiving FC PW PE generates a continuous stream of the most recently received FC Primitive Sequence on the outgoing native FC link, independent of the arrival rate of that FC Primitive Sequence from the WAN. In practice, a 10:1 reduction in FC Primitive Sequence transmission rate achieves 90% of the bandwidth benefits without loss of FC functionality, and sending a copy every few milliseconds does not pose a serious risk of exceeding the timeouts specified in Section 5 below.

These bandwidth-efficiency techniques may cause changes in the FC traffic that traverses an FC PW (e.g., number of Idle Primitive Signals or number of identical Primitive Sequences), but the far-end FC PW PE's regeneration of FC link behavior on the attached FC port is transparent to the FC ports connected to each PW PE.

1.3. Reliability

Fibre Channel does not employ a native frame retransmission protocol and treats most frame delivery failures as errors. FC SAN traffic requires a very low frame loss rate because the typical result of a failure to deliver a frame is an I/O operation failure. Recovery from such I/O failures involves I/O operation retries after what may be a significant delay (30-second and 60-second timeouts are common). In addition, such retries are likely to be logged as errors indicating possible problems with FC equipment or cables. Hence, drops, errors, and discards of FC frames must be very rare for an FC PW.

FC SAN implementations have limited tolerance for frame reordering. Any reordering affecting more than a few frames within a single higher-level operation (e.g., a read or write I/O) is usually treated as an error by the destination FC port, resulting in discards of the frames involved; some deployed FC implementations treat all such within-operation frame reordering as errors that result in frame discards. As a result, FC frame reordering must be minimized for an FC PW.

The FC PW does not compensate for frame drops, discards, or reordering. The MPLS network that hosts the FC PW is expected to be designed and operated in a fashion that makes such events very rare.

In contrast to the Time to Live (TTL) field in an IP packet, FC uses a constant delivery timeout value (R_A_TOV) for which 10 seconds is the default. Each FC frame must be delivered or discarded within that timeout period after it is sent; see Section 5.

1.4. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Reference Model

An FC PW extends a native FC link over an MPLS network. This document specifies the PW encapsulation for FC. Figure 1 describes the reference models (derived from [RFC3985]) that support the FC PW. FC traffic is received by PE1's FC attachment channel, encapsulated at PE1, transported across MPLS network, decapsulated at PE2, and transmitted onward via the PE2's FC attachment channel. This document assumes that a pseudowire can be provisioned statically or via a signaling protocol as defined in [RFC4447].

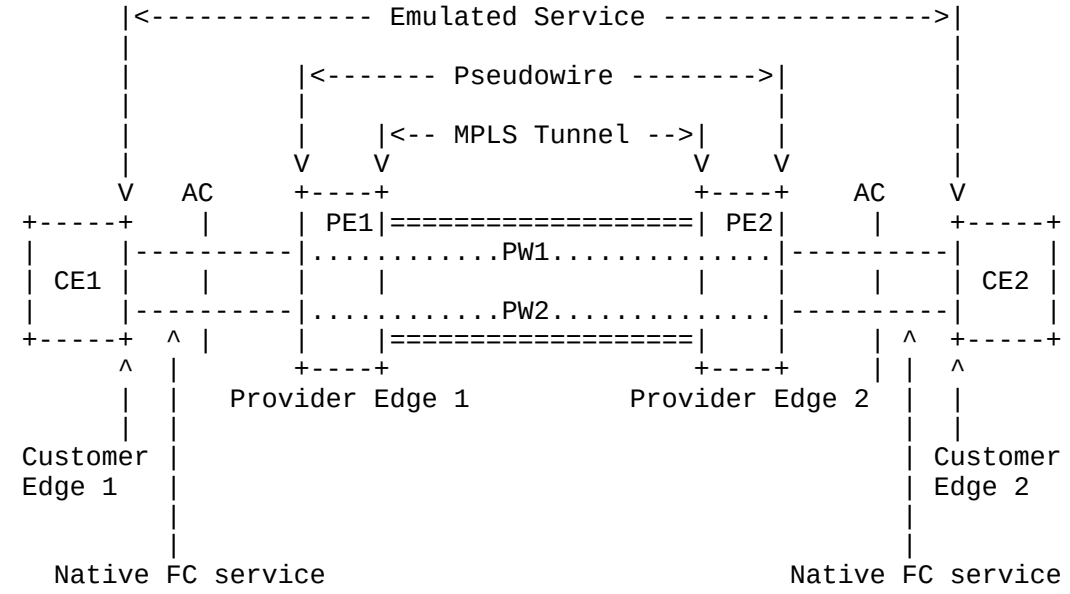

Figure 1 - PWE3 FC Interface Reference Configuration

The following reference model describes the termination point of each end of the PW within the PE:

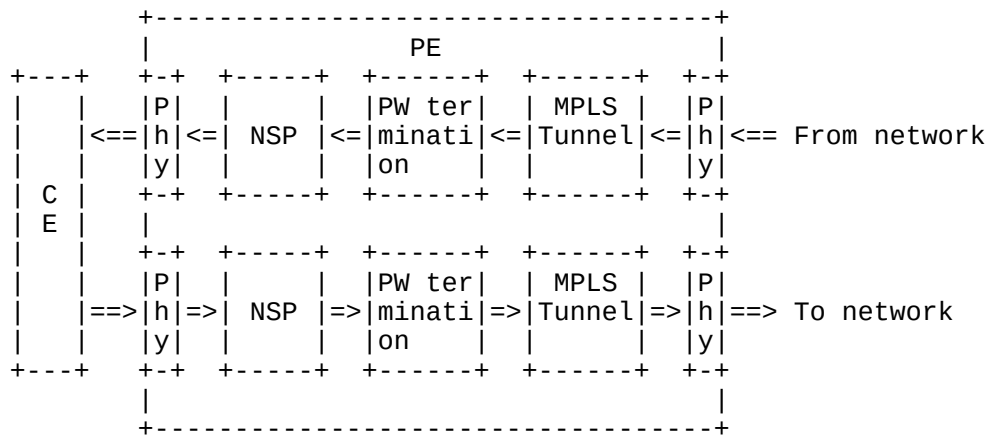

Figure 2 - PW Reference Diagram

The Native Service Processing (NSP) function includes the following functionality:

- Idle Suppression: any FC Idle Primitive Signals received from the source PE's attached FC port are suppressed and regenerated at the destination PE to send on its attached FC port when there is no other FC traffic to send;

- FC Primitive Sequence Reduction: a subset of repetitive FC Primitive Sequences received from the attached FC port at the source PE is selected for WAN transmission, with the destination PE sending the FC Primitive Sequence most recently received from the WAN on the destination PE's attached FC port continuously until a new packet is received from the WAN; and

- Flow Control: the Alternate Simple Flow Control (ASFC) protocol is used for buffer management in concert with the peer PW PE's NSP function so that FC traffic is not dropped. ASFC is a simple pause/resume protocol that allows operation repetition; the receiver responds to the first pause or resume operation in an identical sequence of operations and ignores the rest of the sequence.

The NSP flow control functionality is required to extend FC's credit-based flow control to address situations where the number of buffer credits available to an FC link is insufficient to utilize the available bandwidth over the additional distance and latency

represented by the FC pseudowire. The NSPs avoid this problem by inserting ASFC into FC's link flow control used on the attached FC ports; see [FC-BB-5/AM1].

In contrast, Idle Suppression and FC Primitive Sequence Reduction are bandwidth optimizations that are included in the NSP for clarity in this document. Analogous optimizations are not treated as part of the NSP by other pseudowires (e.g., Asynchronous Transfer Mode (ATM) idle frame suppression is not considered to be an NSP function by [RFC4717]).

The NSP function is specified in detail by [FC-BB-5/AM1].

3. Encapsulation

This specification provides port-to-port transport of FC-encapsulated traffic. There are a number of port types defined by Fibre Channel, including:

- N_port: a port on the node (e.g., host or storage device) used with both FC-P2P (Point to Point) or FC-SW (switched fabric) topologies. Also known as a Node port.
- NL_port: a port on the node used with an FC-AL (Arbitrated Loop) topology. Also known as a Node Loop port.
- F_port: a port on the switch that connects to a node point- to-point (i.e., connects to an N_port). Also known as a Fabric port. An F_port is not loop capable.
- FL_port: a port on the switch that connects to an FC-AL loop (i.e., to NL_ports). Also known as a Fabric Loop port.
- E_port: a port used to connect two Fibre Channel switches. Also known as an Expansion port. When E_ports between two switches are connected to form a link, that link is referred to as an inter-switch link (ISL).

Among the port types listed above, only the following FC connections (as specified in [FC-BB-5/AM1]) are supported by an FC PW over MPLS:

- N_Port to N_Port, established by an FC PLOGI (Port Login) operation
- N_Port to F_Port, established by an FC FLOGI (Fabric Login) operation

- o E_Port to E_Port, established by an FC ELP (Exchange Link Parameters) operation

FC traffic flowing over an FC PW is subdivided into four payload types (PTs) that are encoded in the PW Control Word (see Section 3.1):

1. FC login traffic (PT = 1): FC login operations and responses that establish connections between FC ports. The three FC login operations are PLOGI, FLOGI, and ELP. These operations and their responses may require the NSP to allocate buffer resources. See the specification of Login Exchange Monitors in [FC-BB-5/AM1].

2. FC data traffic (PT = 0): All FC frames other than those involved in an FC login operation.

3. FC Primitive Sequences and Signals (PT = 2): Native FC link control operations; 4-character primitive sequences and signals that are not encapsulated in FC frames. See [FC-BB-5/AM1] and [FC-FS-2].

4. FC PW Control (PT = 6): FC PW control operations exchanged only between the endpoints of the PW. FC PW control operations are used for ASFC flow control, ping (e.g., for round-trip latency measurement), and reporting native FC link errors. See [FC-BB-5/AM1].

This FC PW specification is limited to use with FC service classes 2, 3, and F; see [FC-FS-2]. Other FC service classes (e.g., 1, 4, and 6) MUST NOT be used with an FC PW. Numbered FC service classes are used for end-to-end FC traffic, whereas service class F is used for inter-switch traffic in an FC switched fabric.

This FC PW specification is limited to native FC attachment links that employ an 8b/10b transmission code (see [FC-FS-2]). The protocol specified in this document converts a received 10b code to its 8b counterpart for PW encapsulation and hence does not support attached FC links that use a 64b/66b transmission code (e.g., 10GFC and 16GFC); such links MUST NOT be attached to an FC PW PE unless their link speed can be negotiated to one that uses 8b/10b encoding. If an invalid 10b code that cannot be converted to an 8b code is received from an FC link, the PE sends an FC PW control frame to report the error (see [FC-BB-5/AM1]).

3.1. The Control Word

The Generic PW Control Word, as defined in "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN" [RFC4385], MUST be used for FC PW to facilitate the transport of short packets (by setting the Length field as detailed below) and convey the flag bits defined below. The structure of the Control Word for the FC PW is as follows:

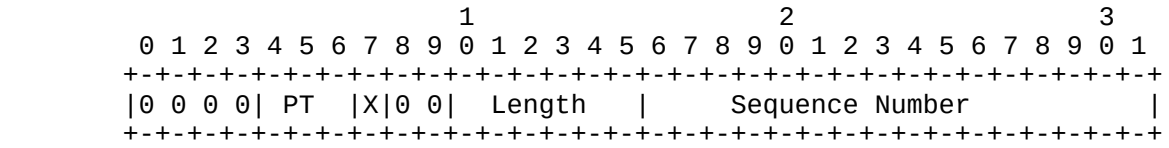

Figure 3 - Control Word Structure

The first four bits of the PW Control Word MUST be set to 0 by the ingress PE to indicate PW data.

Three of the four flag bits are used to convey the Payload Type (PT) indication. The 3-bit binary value in this field identifies the payload type carried by a PW packet. The following types are defined:

PT = 0: FC data frame.

PT = 1: FC login frame.

PT = 2: FC Primitive Sequence(s) and/or Primitive Signal(s).

PT = 6: FC PW control frame (refer to [FC-BB-5/AM1] for usage).

Packets with other values in the PT field are not valid for the FC PW and MUST be discarded by the receiving FC PW PE.

The X flag bit is not used by this version of the protocol. It SHOULD be set to zero by the sender and MUST be ignored by the receiver.

The fragmentation bits (bits 8-9) are not used by the FC PW protocol. These bits may be used in the future for FC-specific indications as defined in [RFC4385]. The fragmentation bits SHOULD be set to zero by the ingress PE and MUST be ignored by the egress PE.

The Length field enables recovery of the original pseudowire packet when a short packet is padded to the minimum 64-octet packet size required for Ethernet; see [RFC4385]. The Length field MUST be used for packets shorter than 64 octets, MUST be set to zero for longer packets, and MUST be processed according to the rules specified in [RFC4385].

The sequence number is not used for the FC PW; it MUST be set to 0 by the ingress PE and MUST be ignored by the egress PE.

3.2. MTU Requirements

The MPLS network MUST be able to transport the largest Fibre Channel frame after encapsulation, including the overhead associated with the encapsulation. The maximum FC frame size is 2164 octets without PW and MPLS labels (refer to Figure 4); this maximum size is a constant value that is required for all FC implementations [FC-FS-2]. The MPLS network SHOULD accommodate frames of up to 2500 octets in order to support possible future increases in the maximum FC frame size.

Fragmentation, as described in [RFC4623], SHALL NOT be used for an FC PW; therefore, the network MUST be configured with a minimum MTU that is sufficient to transport the largest encapsulated FC frame.

3.3. Mapping of FC Traffic to PW Packets

FC frames, Primitive Sequences, and Primitive Signals are transported over the PW. All packet types are carried over a single PW. In addition to the PW Control Word, an FC Encapsulation Header is included in the PW packet. This FC Encapsulation Header is not used in this version of the protocol; it SHOULD be set to zero by the sender and MUST be ignored by the receiver.

3.3.1. FC Data Frames (PT=0) and FC Login Frames (PT=1)

FC data frames and FC login frames share a common encapsulation format, except that the PT field in the FC PW Control Word is set to 0 for data frames and is set to 1 for login frames. An FC login frame contains an FC PLOGI, FLOGI, or ELP operation or response that requires special processing by the NSP in support of flow control; see [FC-BB-5/AM1].

Each FC data frame or login frame is mapped to a PW packet, including the Start Of Frame (SOF) delimiter, frame header, Cyclic Redundancy Check (CRC) field, and the End Of Frame (EOF) delimiter, as shown in Figure 4.

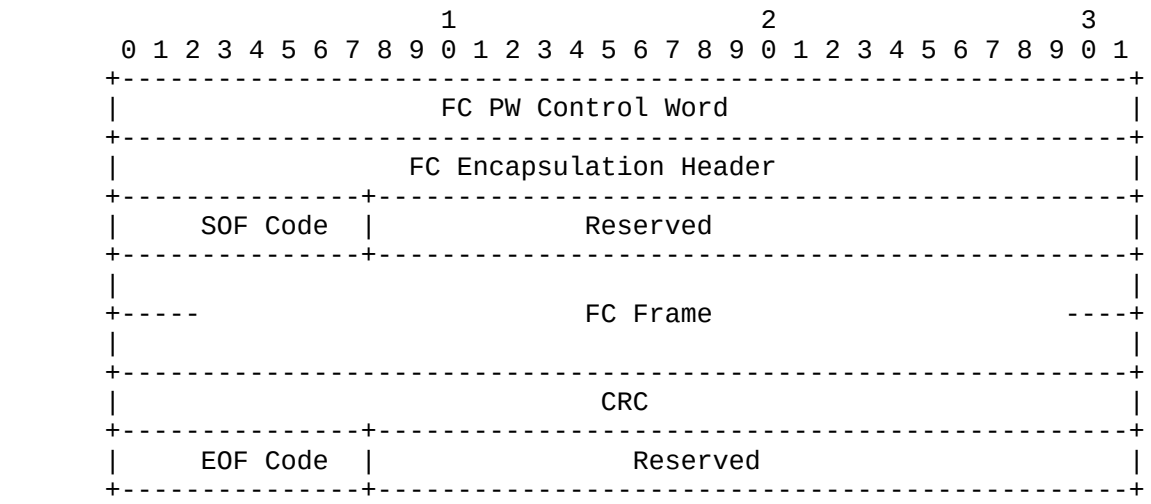

Figure 4 - FC Frame (SOF/Data/CRC/EOF) Encapsulation in PW Packet

The SOF and EOF frame delimiters are each encoded into a single octet as specified in [RFC3643], except that the codes for delimiters that apply only to FC service class 4 (SOFi4, SOFc4, SOFn4, EOFdt, EOFdti, EOFrt, and EOFrti -- see [FC-FS-2]) MUST NOT be used.

The CRC in the frame is obtained directly from the FC attachment channel, so that the PW PE is not required to recalculate the CRC or to check the CRC in the received frame. The CRC will be checked by the FC port that receives the frame, ensuring that coverage is provided for data errors that occur between the PW endpoints. This CRC behavior differs from the Frame Check Sequence (FCS) retention technique for PWs defined in [RFC4720], which states that "as usual, the FCS MUST be examined at the ingress PE, and errored frames MUST be discarded".

3.3.2. FC Primitive Sequences and Primitive Signals (PT=2)

FC Primitive Sequences and Primitive Signals are FC Ordered Sets. On an 8b/10b-coded FC link, an Ordered Set consists of four 10b characters, starting with the K28.5 character, followed by three Dxx.y data characters. All FC Ordered Sets start with a K28.5 control character, but the three following Dxx.y data characters differ depending on the Ordered Set. A Kxx.y control character has a different 10b code from the corresponding Dxx.y data character but uses the same 8b code (e.g., K28.5 and D28.5 both use the 8b code 0xBC). Here are two examples of Ordered Sets:

- Idle (IDLE) is K28.5 - D21.4 - D21.5 - D21.5. This FC Primitive Signal is sent when the FC link is idle; it is suppressed by the FC PW NSP and not sent over the WAN.

- Link Reset Response (LRR) is K28.5 - D21.1 - D31.5 - D9.2. This FC Primitive Sequence is used as part of FC link initialization and recovery.

Each Ordered Set is encapsulated in a PW packet containing the encoded K28.5 control character [FC-BB-5/AM1], followed by three encoded data characters, as shown in Figure 5.

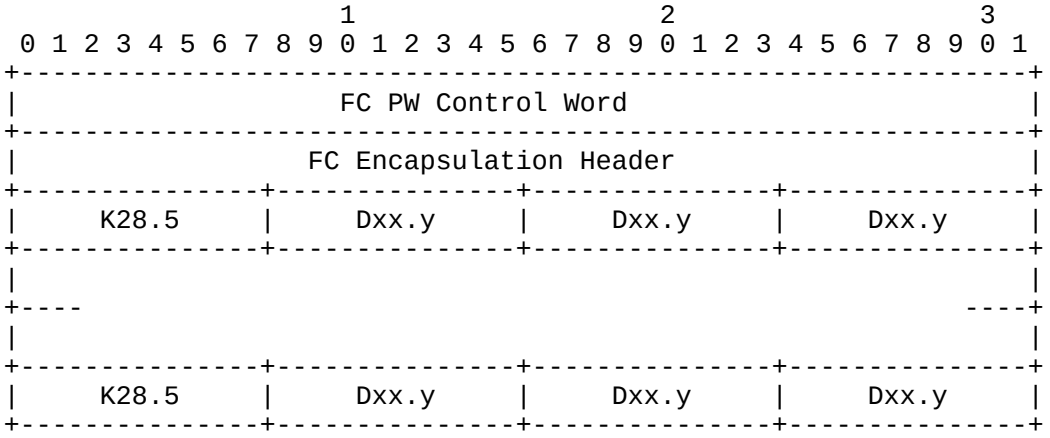

Figure 5 - FC Ordered Sets Encapsulation in PW Packet

The K28.5 10b control character received from the PE's attached FC link is encoded for the FC PW as its 8b counterpart (0xBC). Because the same 8b value (0xBC) is used to encode a D28.5 data word, the receiving FC PW PE:

- MUST check for presence of an 8b K28.5 value (0xBC) at the start of each Ordered Set (see Figure 5) and MUST send that value as a 10b K28.5 character on the attached FC link.

- MUST send the following three Dxx.y 8b values as Dxx.y 10b characters on the attached FC link and MUST NOT send any of these Dxx.y 8b values as 10b Kxx.y characters on the attached FC link.

A PW packet may contain one or more encoded FC Ordered Sets [FC-BB-5/AM1]. The Length field in the FC PW Control Word is used to indicate the packet length when the PW packet contains multiple Ordered Sets. For this reason, FC PW packets that contain FC Ordered

Sets MUST NOT be larger than 60 octets (8 octets of header words plus at most 13 Ordered Sets), in order to ensure that the Length field contains a non-zero value (see [RFC4385]).

Idle Primitive Signals could be carried over the PW in the same manner as Primitive Sequences. However, [FC-BB-5/AM1] requires that Idle Primitive Signals be dropped by the Ingress PE and regenerated by the egress PE in order to reduce bandwidth consumption (see [FC-BB-5/AM1] for further details).

The egress PE extracts the Primitive Sequence or Primitive Signal from the received PW packet. For a Primitive Sequence, the PE continues transmitting the same FC Ordered Set to its attached FC port until an FC frame or another Ordered Set is received over the PW; see Section 1.2 above for discussion of ingress PE transmission behavior for Primitive Sequences. A Primitive Signal is sent once, except that Idle Primitive Signals are sent continuously when there is nothing else to send.

3.3.3. FC PW Control Frames (PT=6)

FC PW control frames are transported over the PW by encapsulating each frame in a PW packet with PT=6 in the Control Word. FC PW control frame payloads are generated and terminated by the corresponding FC entity. FC PW control frames are used for FC PW flow control (ASFC), ping, and transmission of error indications. [FC-BB-5/AM1] specifies the generation and processing of FC PW control frames. FC PW control frames are always shorter than 64 octets, and hence the Length field in the FC Control Word indicates their length.

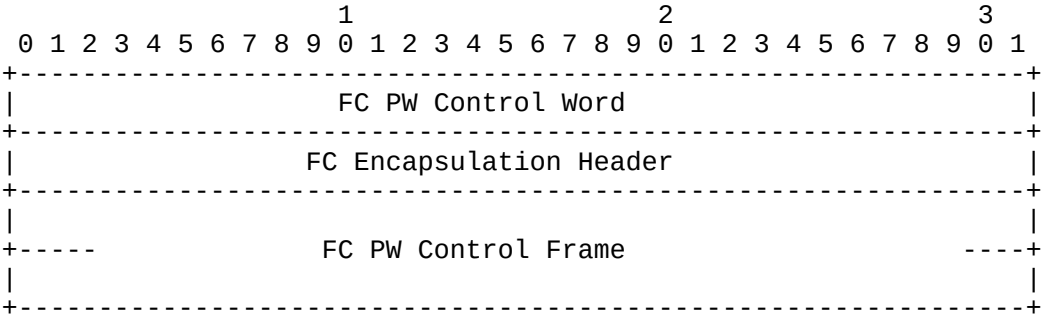

Figure 6 - FC PW Control Frame Encapsulation in PW Packet

3.4. PW Failure Mapping

PW failures are detected through PW signaling failure, PW status notifications as defined in [RFC4447], or PW Operations, Administration, and Maintenance (OAM) mechanisms and MUST be mapped to emulated signal failure indications. Sending the FC link failure indication to its attached FC link is performed by the NSP, as defined by [FC-BB-5/AM1].

4. Signaling of FC Pseudowires

[RFC4447] specifies the use of the MPLS Label Distribution Protocol (LDP) as a protocol for setting up and maintaining pseudowires. This section describes the use of specific fields and error codes used to control FC PW.

The PW Type field in the PWid Forwarding Equivalence Class (FEC) element and PW generalized ID FEC elements MUST be set to the "FC Port Mode" value in Section 8.

The Control Word is REQUIRED for FC pseudowires. Therefore, the C-Bit in the PWid FEC element and PW generalized ID FEC elements MUST be set. If the C-Bit is not set, the pseudowire MUST NOT be established, and a Label Release MUST be sent with an "Illegal C-Bit" status code [RFC4447].

The Fragmentation Indicator (Parameter ID = 0x09) is specified in [RFC4446], and its usage is defined in [RFC4623]. Since fragmentation is not used in FC PW, the fragmentation indicator parameter MUST be omitted from the Interface Parameter Sub-TLV.

The Interface MTU Parameter (Parameter ID = 0x01) is specified in [RFC4447]. Since all FC interfaces have the same MTU, this parameter MUST be omitted from the Interface Parameter Sub-TLV.

The FCS Retention Indicator (Parameter ID = 0x0A) is specified in [RFC4720]. Since the CRC treatment defined in this document differs from one that is specified in [RFC4720], this parameter MUST be omitted from the Interface Parameter Sub-TLV.

5. Timing Considerations

Correct Fibre Channel link operation requires that the FC link latency between CE1 and CE2 (refer to Figure 1) be:

- no more than one-half of the R_T_TOV (Receiver Transmitter Timeout Value, default value: 100 milliseconds) of the attached devices for Primitive Sequences;

- no more than one-half of the E_D_TOV (Error Detect Timeout Value, default value: 2 seconds) of the attached devices for frames; and

- within the R_A_TOV (Resource Allocation Timeout Value, default value: 10 seconds) of the attached fabric(s), if any. The FC standards require that the E_D_TOV value for each FC link be set so that the R_A_TOV value for the fabric is respected when the worst-case latency occurs for each link (see [FC-FS-2]).

An FC PW MUST adhere to these three timing requirements and MUST NOT be used in environments where high or variable latency may cause these requirements to be violated.

These three timeout values are ordered (R_T_TOV < E_D_TOV < R_A_TOV), so adherence to one-half of R_T_TOV for all FC PW traffic is sufficient. See [FC-FS-2] for definitions of the FC timeout values.

The R_T_TOV is used by the FC link initialization protocol. If an FC PW's latency exceeds one-half R_T_TOV, initialization of the FC link that is encapsulated by the FC PW may fail, leaving that FC link in a non-operational state.

The E_D_TOV is used to detect failures of operational FC links. If an FC PW's latency exceeds the one-half E_D_TOV requirement, the FC link that is encapsulated by the FC PW may fail. The usual FC response to such a link failure is to attempt to recover the FC link by initializing it. That initialization will also fail if the FC PW latency exceeds one-half R_T_TOV (a tighter requirement).

The R_A_TOV is used to determine when FC communication resources (e.g., values that identify FC frames) may be reused. If an FC PW's violation of the one-half E_D_TOV requirement is sufficient to also cause the FC fabric to violate the R_A_TOV requirement, then FC reuse of frame identification values after an R_A_TOV timeout may result in multiple FC frames with the same identification values, causing incorrect Fibre Channel operation. For example, if two such frames are swapped between I/O operations, the result may corrupt data in the I/O operations.

The PING and PING_ACK FC PW control frames defined in Section 6.4.7 of [FC-BB-5/AM1] SHOULD be used to measure the current FC pseudowire latency between the Customer Edge (CE) devices. If the measured latency violates any of the timing requirements, then the FC PW PE MUST generate a WAN Down event as specified in [FC-BB-5/AM1].

The WAN Down event causes the PE to continuously send NOS (an FC Primitive Sequence) on the native FC link to the FC port at the other end of that link (typically an E_Port on a switch in this case).

This immediately causes the FC link that is carried by the PW to become non-operational, halting transmission of FC traffic. However, it is not necessary to tear down the pseudowire itself in this situation (e.g., destroy the MPLS path set up by LDP).

The Transparent FC-BB initialization state machine in [FC-BB-5/AM1] specifies the protocol used to attempt to recover from a WAN Down event (i.e., bring the WAN back up). If that protocol brings the WAN back up, FC traffic will resume and the standard FC link recovery protocol will bring the encapsulated FC link back up. If the previous pseudowire was destroyed, attempts will be made to re-establish the path via LDP as part of recovering from the WAN Down event. If the PW round-trip latency remains above R_T_TOV, the initialization protocol for the FC PW will repeatedly time out in attempting to recover from the WAN Down event, preventing recovery of the FC link carried by the PW; see [FC-BB-5/AM1].

6. Security Considerations

The FC PW is an MPLS pseudowire; for MPLS pseudowire security considerations, see the security considerations sections of [RFC3985] and [RFC4385].

The protocols used to implement security in a Fibre Channel fabric are defined in [FC-SP]. These protocols operate at higher layers of the FC hierarchy and are transparent to the FC PW.

The FC timing requirements (see Section 5) create an exposure of the FC PW to inserted latency. Injection of latency sufficient to cause the round-trip time for an FC PW to exceed R_T_TOV (default: 100 ms) may cause the FC PW to fail in an active fashion because the FC link initialization protocol repeatedly times out. OAM functionality for deployed FC PWs SHOULD monitor for persistence of this situation and respond accordingly (e.g., shut down the FC PW in order to avoid wasting WAN bandwidth on an FC PW whose FC link cannot be successfully initialized due to excessive latency).

7. Applicability Statement

FC PW allows the transparent transport of FC traffic between Fibre Channel ports while saving network bandwidth by removing FC Idle Primitive Signals and reducing the number of FC Primitive Sequences.

- The pair of CE devices operates as if they were directly connected by an FC link. In particular, they react to Primitive Sequences on their local FC links as specified by the FC standards.

- The FC PW carries only FC data frames, FC Primitive Signals, and a subset of the copies of an FC Primitive Sequence. Idle Primitive Signals are suppressed, and long streams of the same Primitive Sequence are reduced over the PW, thus saving bandwidth.

- The PW PE MUST generate Idle Primitive Signals to the attached FC link when there is no other traffic to transmit on the attached FC link [FC-FS-2].

- The PW PE MUST send Primitive Sequences continuously to the attached FC port, as required by the FC standards [FC-FS-2].

FC PW traffic should only traverse MPLS networks that are provisioned based on traffic engineering to provide dedicated bandwidth for FC PW traffic. The MPLS network should enforce ingress traffic policing so that delivery of FC PW traffic can be assured. To extend FC across a network that does not satisfy these requirements, FCIP SHOULD be used instead of an FC PW (see [RFC3821] and [FC-BB-6]).

This document does not provide any mechanisms for protecting an FC PW against network outages. As a consequence, resilience of the emulated FC service to such outages is dependent upon the underlying MPLS network, which should be protected against failures. When a network outage is detected, the PE SHOULD use a WAN Down event (as specified in [FC-BB-5/AM1]) to convey the PW status to the CE and enable faster outage handling.

8. IANA Considerations

IANA has assigned a new MPLS Pseudowire (PW) type as follows:

PW type	Description	Reference
0x001F	FC Port Mode	RFC 6307

IANA has reserved the following Pseudowire Interface Parameters Sub-TLV Types. These Sub-TLV types were used for the FC PW Selective Retransmission protocol, which the PWE3 working group has decided to eliminate. This action prevents future use of these values for other purposes, as there is at least one implementation of the Selective Retransmission protocol that has been deployed.

Parameter ID	Length	Description	Reference
0x12		Reserved	RFC 6307
0x13		Reserved	RFC 6307
0x14		Reserved	RFC 6307
0x15		Reserved	RFC 6307

9. Acknowledgments

Previous versions of this document were authored by Moran Roth, Ronen Solomon, and Munefumi Tsurusawa; their efforts and contributions are gratefully acknowledged. The authors would like to thank Stewart Bryant, Elwyn Davies, Steve Hanna, Dave Peterson, Yaakov Stein, Alexander Vainshtein, and the members of the IESG for helpful comments on this document.

The protocol specified in this document is intended to be used in conjunction with the Fibre Channel pseudowire portion of the FC-BB-5 Amendment 1 specification developed by INCITS Technical Committee T11. The authors would like to thank the members of both the IETF and T11 organizations who have supported and contributed to this work.

10. Normative References

[FC-BB-5/AM1] "Fibre Channel - Backbone-5 / Amendment 1", INCITS 462-2010/AM 1-2012, June 2012.

[FC-FS-2] "Fibre Channel - Framing and Signaling-2 (FC-FS-2)", ANSI INCITS 424:2007, August 2007.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3643] Weber, R., Rajagopal, M., Travostino, F., O'Donnell, M., Monia, C., and M. Merhar, "Fibre Channel (FC) Frame Encapsulation", RFC 3643, December 2003.

[RFC3985] Bryant, S., Ed., and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, March 2005.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, February 2006.

[RFC4446] Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", BCP 116, RFC 4446, April 2006.

[RFC4447] Martini, L., Ed., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.

[RFC4623] Malis, A. and M. Townsley, "Pseudowire Emulation Edge-to-Edge (PWE3) Fragmentation and Reassembly", RFC 4623, August 2006.

[RFC4720] Malis, A., Allan, D., and N. Del Regno, "Pseudowire Emulation Edge-to-Edge (PWE3) Frame Check Sequence Retention", RFC 4720, November 2006.

11. Informative References

[FC-BB-6] "Fibre Channel Backbone-6" (FC-BB-6), T11 Project 2159-D, Rev 1.04, Work in Progress, January 2012.

[FC-SP] "Fibre Channel - Security Protocols" (FC-SP), ANSI INCITS 426:2007, February 2007.

[RFC3821] Rajagopal, M., Rodriguez, E., and R. Weber, "Fibre Channel Over TCP/IP (FCIP)", RFC 3821, July 2004.

[RFC4717] Martini, L., Jayakumar, J., Bocci, M., El-Aawar, N., Brayley, J., and G. Koleyni, "Encapsulation Methods for Transport of Asynchronous Transfer Mode (ATM) over MPLS Networks", RFC 4717, December 2006.

[T11] INCITS Technical Committee T11, http://www.t11.org, January 2011.

Authors' Addresses

David L. Black (editor)
EMC Corporation
176 South Street
Hopkinton, MA 01748
USA
Phone: +1 (508) 293-7953
EMail: david.black@emc.com

Linda Dunbar (editor)
Huawei Technologies
1700 Alma Drive, Suite 500
Plano, TX 75075
USA
Phone: +1 (972) 543-5849
EMail: ldunbar@huawei.com

Moran Roth
Infinera Corporation
169 Java Drive
Sunnyvale, CA 94089
USA
Phone: (408) 572-5200
EMail: MRoth@infinera.com

Ronen Solomon
Orckit-Corrigent Systems
126, Yigal Alon st.
Tel Aviv
Israel
Phone: +972-3-6945316
EMail: ronens@corrigent.com