Network Working Group
Request for Comments: 4172
Category: Standards Track

C. Monia
Consultant
R. Mullendore
McDATA
F. Travostino
Nortel
W. Jeong
Troika Networks
M. Edwards
Adaptec (UK) Ltd.
September 2005

iFCP - A Protocol for Internet Fibre Channel Storage Networking

Status of This Memo

Copyright Notice

Abstract

This document specifies an architecture and a gateway-to-gateway protocol for the implementation of fibre channel fabric functionality over an IP network. This functionality is provided through TCP protocols for fibre channel frame transport and the distributed fabric services specified by the fibre channel standards. The architecture enables internetworking of fibre channel devices through gateway-accessed regions with the fault isolation properties of autonomous systems and the scalability of the IP network.

Table of Contents

1. Introduction

1.1. 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 BCP 14, RFC 2119 [RFC2119].

Unless specified otherwise, numeric quantities are given as decimal values.

All diagrams that portray bit and byte ordering, including the depiction of structures defined by fibre channel standards, adhere to the IETF conventions whereby bit 0 is the most significant bit and the first addressable byte is in the upper left corner. This IETF convention differs from that used for INCITS T11 fibre channel standards, in which bit 0 is the least significant bit.

1.1.1. Data Structures Internal to an Implementation

To facilitate the specification of required behavior, this document may define and refer to internal data structures within an iFCP implementation. Such structures are intended for explanatory purposes only and need not be instantiated within an implementation as described in this specification.

1.2. Purpose of This Document

This is a standards-track document that specifies a protocol for the implementation of fibre channel transport services on a TCP/IP network. Some portions of this document contain material from standards controlled by INCITS T10 and T11. This material is included here for informational purposes only. The authoritative information is given in the appropriate NCITS standards document.

The authoritative portions of this document specify the mapping of standards-compliant fibre channel protocol implementations to TCP/IP. This mapping includes sections of this document that describe the "iFCP Protocol" (see Section 5).

2. iFCP Introduction

iFCP is a gateway-to-gateway protocol that provides fibre channel fabric services to fibre channel devices over a TCP/IP network. iFCP uses TCP to provide congestion control, error detection, and recovery. iFCP's primary objective is to allow interconnection and

networking of existing fibre channel devices at wire speeds over an IP network.

The protocol and method of frame address translation described in this document permit the attachment of fibre channel storage devices to an IP-based fabric by means of transparent gateways.

The protocol achieves this transparency by allowing normal fibre channel frame traffic to pass through the gateway directly, with provisions, where necessary, for intercepting and emulating the fabric services required by a fibre channel device.

2.1. Definitions

Terms needed to describe the concepts presented in this document are presented here.

Address-translation mode -- A mode of gateway operation in which the scope of N_PORT fabric addresses, for locally attached devices, are local to the iFCP gateway region in which the devices reside.

Address-transparent mode -- A mode of gateway operation in which the scope of N_PORT fabric addresses, for all fibre channel devices, are unique to the bounded iFCP fabric to which the gateway belongs.

Bounded iFCP Fabric -- The union of two or more gateway regions configured to interoperate in address-transparent mode.

DOMAIN_ID -- The value contained in the high-order byte of a 24-bit N_PORT fibre channel address.

F_PORT -- The interface used by an N_PORT to access fibre channel switched-fabric functionality.

Fabric -- From [FC-FS]: "The entity that interconnects N_PORTs attached to it and is capable of routing frames by using only the address information in the fibre channel frame."

Fabric Port -- The interface through which an N_PORT accesses a fibre channel fabric. The type of fabric port depends on the fibre channel fabric topology. In this specification, all fabric port interfaces are considered functionally equivalent.

FC-2 -- The fibre channel transport services layer, described in [FC-FS].

FC-4 -- The fibre channel mapping of an upper-layer protocol, such as [FCP-2], the fibre channel to SCSI mapping.

Fibre Channel Device -- An entity implementing the functionality accessed through an FC-4 application protocol.

Fibre Channel Network -- A native fibre channel fabric and all attached fibre channel nodes.

Fibre Channel Node -- A collection of one or more N_PORTs controlled by a level above the FC-2 layer. A node is attached to a fibre channel fabric by means of the N_PORT interface, described in [FC-FS].

Gateway Region -- The portion of an iFCP fabric accessed through an iFCP gateway by a remotely attached N_PORT. Fibre channel devices in the region consist of all those locally attached to the gateway.

iFCP -- The protocol discussed in this document.

iFCP Frame -- A fibre channel frame encapsulated in accordance with the FC Frame Encapsulation Specification [ENCAP] and this specification.

iFCP Portal -- An entity representing the point at which a logical or physical iFCP device is attached to the IP network. The network address of the iFCP portal consists of the IP address and TCP port number to which a request is sent when the TCP connection is created for an iFCP session (see Section 5.2.1).

iFCP Session -- An association comprised of a pair of N_PORTs and a TCP connection that carries traffic between them. An iFCP session may be created as the result of a PLOGI fibre channel login operation.

iSNS -- The server functionality and IP protocol that provide storage name services in an iFCP network. Fibre channel name services are implemented by an iSNS name server, as described in [ISNS].

Locally Attached Device -- With respect to a gateway, a fibre channel device accessed through the fibre channel fabric to which the gateway is attached.

Logical iFCP Device -- The abstraction representing a single fibre channel device as it appears on an iFCP network.

N_PORT -- An iFCP or fibre channel entity representing the interface to fibre channel device functionality. This interface implements the fibre channel N_PORT semantics, specified in [FC-FS]. Fibre channel defines several variants of this interface that depend on the fibre channel fabric topology. As used in this document, the term applies equally to all variants.

N_PORT Alias -- The N_PORT address assigned by a gateway to represent a remote N_PORT accessed via the iFCP protocol.

N_PORT fabric address -- The address of an N_PORT within the fibre channel fabric.

N_PORT ID -- The address of a locally attached N_PORT within a gateway region. N_PORT IDs are assigned in accordance with the fibre channel rules for address assignment, specified in [FC-FS].

N_PORT Network Address -- The address of an N_PORT in the iFCP fabric. This address consists of the IP address and TCP port number of the iFCP Portal and the N_PORT ID of the locally attached fibre channel device.

Port Login (PLOGI) -- The fibre channel Extended Link Service (ELS) that establishes an iFCP session through the exchange of identification and operation parameters between an originating N_PORT and a responding N_PORT.

Remotely Attached Device -- With respect to a gateway, a fibre channel device accessed from the gateway by means of the iFCP protocol.

Unbounded iFCP Fabric -- The union of two or more gateway regions configured to interoperate in address-translation mode.

3. Fibre Channel Communication Concepts

Fibre channel is a frame-based, serial technology designed for peer-to-peer communication between devices at gigabit speeds and with low overhead and latency.

This section contains a discussion of the fibre channel concepts that form the basis for the iFCP network architecture and protocol described in this document. Readers familiar with this material may skip to Section 4.

Material presented in this section is drawn from the following T11 specifications:

-- The Fibre Channel Framing and Signaling Interface, [FC-FS]

-- Fibre Channel Switch Fabric -2, [FC-SW2]

-- Fibre Channel Generic Services, [FC-GS3]

-- Fibre Channel Fabric Loop Attachment, [FC-FLA]

The reader will find an in-depth treatment of the technology in [KEMCMP] and [KEMALP].

3.1. The Fibre Channel Network

The fundamental entity in fibre channel is the fibre channel network. Unlike a layered network architecture, a fibre channel network is largely specified by functional elements and the interfaces between them. As shown in Figure 1, these consist, in part, of the following:

a) N_PORTs -- The end points for fibre channel traffic. In the FC standards, N_PORT interfaces have several variants, depending on the topology of the fabric to which they are attached. As used in this specification, the term applies to any one of the variants.

b) FC Devices -- The fibre channel devices to which the N_PORTs provide access.

c) Fabric Ports -- The interfaces within a fibre channel network that provide attachment for an N_PORT. The types of fabric port depend on the fabric topology and are discussed in Section 3.2.

d) The network infrastructure for carrying frame traffic between N_PORTs.

e) Within a switched or mixed fabric (see Section 3.2), a set of auxiliary servers, including a name server for device discovery and network address resolution. The types of service depend on the network topology.

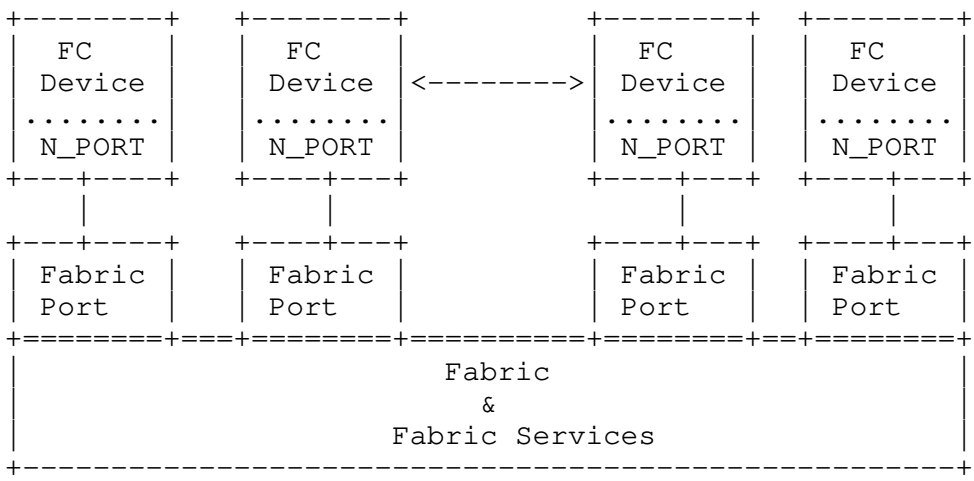


Figure 1. A Fibre Channel Network

The following sections describe fibre channel network topologies and give an overview of the fibre channel communications model.

3.2. Fibre Channel Network Topologies

The principal fibre channel network topologies consist of the following:

a) Arbitrated Loop -- A series of N_PORTs connected together in daisy-chain fashion. In [FC-FS], loop-connected N_PORTs are referred to as NL_PORTs. Data transmission between NL_PORTs requires arbitration for control of the loop in a manner similar to that of a token ring network.

b) Switched Fabric -- A network consisting of switching elements, as described in Section 3.2.1.

c) Mixed Fabric -- A network consisting of switches and "fabric-attached" loops. A description can be found in [FC-FLA]. A loop-attached N_PORT (NL_PORT) is connected to the loop through an L_PORT and accesses the fabric by way of an FL_PORT.

Depending on the topology, the N_PORT and its means of network attachment may be one of the following:

FC Network Topology	Network Interface	N_PORT Variant
Loop	L_PORT	NL_PORT
Switched	F_PORT	N_PORT
Mixed	FL_PORT via L_PORT	NL_PORT
	F_PORT	N_PORT

The differences in each N_PORT variant and its corresponding fabric port are confined to the interactions between them. To an external N_PORT, all fabric ports are transparent, and all remote N_PORTs are functionally identical.

3.2.1. Switched Fibre Channel Fabrics

An example of a multi-switch fibre channel fabric is shown in Figure 2.

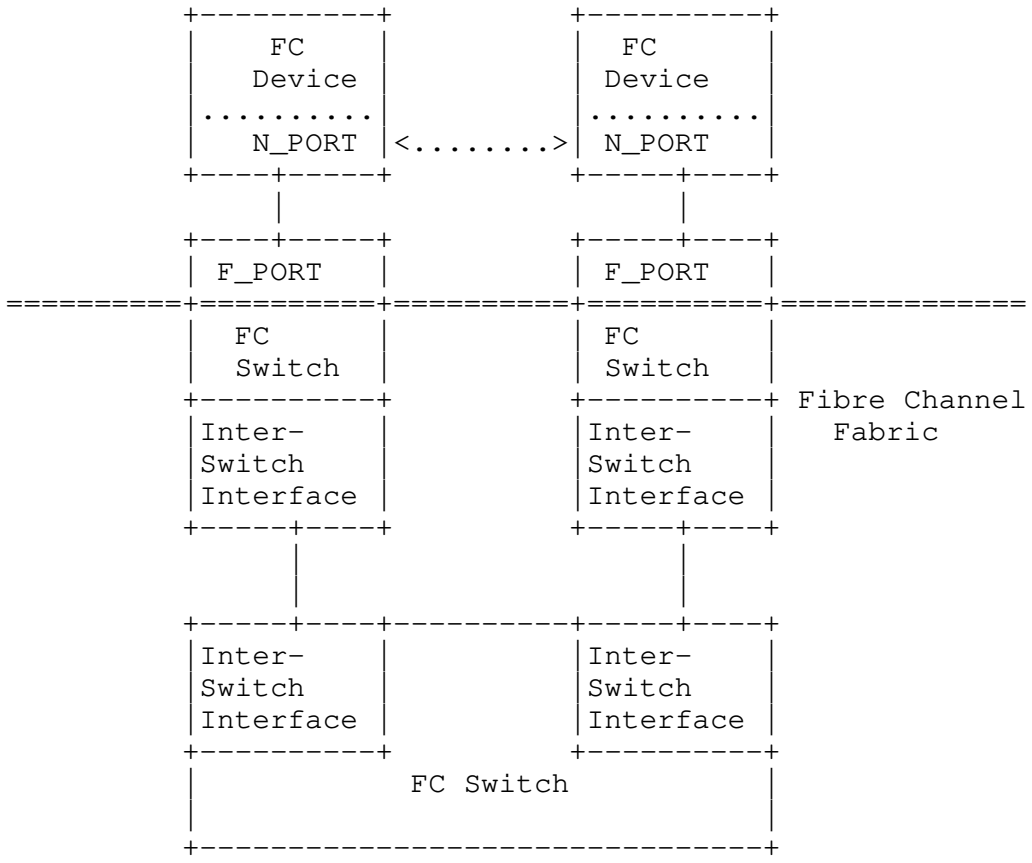


Figure 2. Multi-Switch Fibre Channel Fabric

The interface between switch elements is either a proprietary interface or the standards-compliant E_PORT interface, which is described by the FC-SW2 specification, [FC-SW2].

3.2.2. Mixed Fibre Channel Fabric

A mixed fabric contains one or more arbitrated loops connected to a switched fabric as shown in Figure 3.

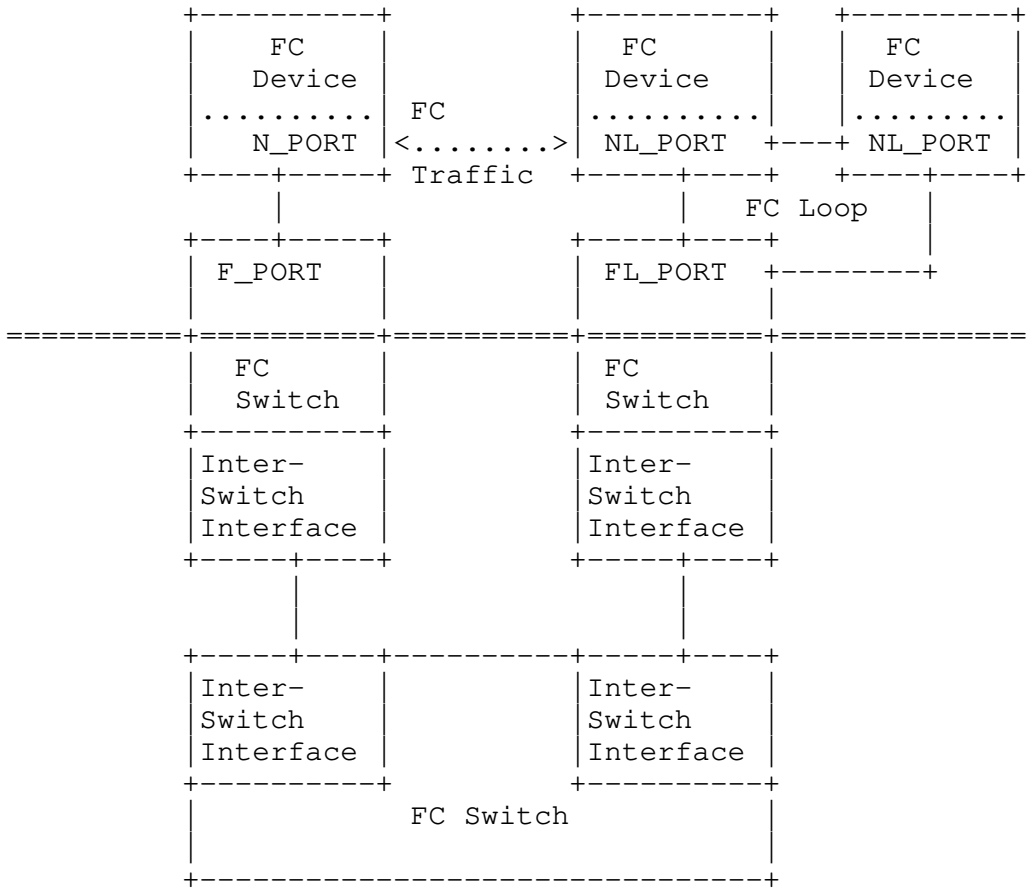


Figure 3. Mixed Fibre Channel Fabric

As noted previously, the protocol for communications between peer N_PORTs is independent of the fabric topology, N_PORT variant, and type of fabric port to which an N_PORT is attached.

3.3. Fibre Channel Layers and Link Services

A fibre channel consists of the following layers:

FC-0 -- The interface to the physical media.

FC-1 -- The encoding and decoding of data and out-of-band physical link control information for transmission over the physical media.

FC-2 -- The transfer of frames, sequences, and Exchanges comprising protocol information units.

FC-3 -- Common Services.

FC-4 -- Application protocols such as the fibre channel protocol for SCSI (FCP).

In addition to the layers defined above, a fibre channel defines a set of auxiliary operations, some of which are implemented within the transport layer fabric, called link services. These are required in order to manage the fibre channel environment, establish communications with other devices, retrieve error information, perform error recovery, and provide other similar services. Some link services are executed by the N_PORT. Others are implemented internally within the fabric. These internal services are described in the next section.

3.3.1. Fabric-Supplied Link Services

Servers that are internal to a switched fabric handle certain classes of Link Service requests and service-specific commands. The servers appear as N_PORTs located at the 'well-known' N_PORT fabric addresses specified in [FC-FS]. Service requests use the standard fibre channel mechanisms for N_PORT-to-N_PORT communications.

All switched fabrics must provide the following services:

Fabric F_PORT server -- Services N_PORT requests to access the fabric for communications.

Fabric Controller -- Provides state change information to inform other FC devices when an N_PORT exits or enters the fabric (see Section 3.5).

Directory/Name Server - Allows N_PORTs to register information in a database, retrieve information about other N_PORTs, and to discover other devices as described in Section 3.5.

A switched fabric may also implement the following optional services:

Broadcast Address/Server -- Transmits single-frame, class 3 sequences to all N_PORTs.

Time Server -- Intended for the management of fabric-wide expiration timers or elapsed time values; not intended for precise time synchronization.

Management Server - Collects and reports management information, such as link usage, error statistics, link quality, and similar items.

Quality of Service Facilitator - Performs fabric-wide bandwidth and latency management.

3.4. Fibre Channel Nodes

A fibre channel node has one or more fabric-attached N_PORTs. The node and its N_PORTs have the following associated identifiers:

a) A worldwide-unique identifier for the node.

b) A worldwide-unique identifier for each N_PORT associated with the node.

c) For each N_PORT attached to a fabric, a 24-bit fabric-unique address with the properties defined in Section 3.7.1. The fabric address is the address to which frames are sent.

Each worldwide-unique identifier is a 64-bit binary quantity with the format defined in [FC-FS].

3.5. Fibre Channel Device Discovery

In a switched or mixed fabric, fibre channel devices and changes in the device configuration may be discovered by means of services provided by the fibre channel Name Server and Fabric Controller.

The Name Server provides registration and query services that allow a fibre channel device to register its presence on the fabric and to discover the existence of other devices. For example, one type of query obtains the fabric address of an N_PORT from its 64-bit worldwide-unique name. The full set of supported fibre channel name server queries is specified in [FC-GS3].

The Fabric Controller complements the static discovery capabilities provided by the Name Server through a service that dynamically alerts a fibre channel device whenever an N_PORT is added or removed from the configuration. A fibre channel device receives these notifications by subscribing to the service as specified in [FC-FS].

3.6. Fibre Channel Information Elements

The fundamental element of information in fibre channel is the frame. A frame consists of a fixed header and up to 2112 bytes of payload with the structure described in Section 3.7. The maximum frame size that may be transmitted between a pair of fibre channel devices is negotiable up to the payload limit, based on the size of the frame buffers in each fibre channel device and the path maximum transmission unit (MTU) supported by the fabric.

Operations involving the transfer of information between N_PORT pairs are performed through 'Exchanges'. In an Exchange, information is transferred in one or more ordered series of frames, referred to as Sequences.

Within this framework, an upper layer protocol is defined in terms of transactions carried by Exchanges. In turn, each transaction consists of protocol information units, each of which is carried by an individual Sequence within an Exchange.

3.7. Fibre Channel Frame Format

A fibre channel frame consists of a header, payload and 32-bit CRC bracketed by SOF and EOF delimiters. The header contains the control information necessary to route frames between N_PORTs and manage Exchanges and Sequences. The following diagram gives a schematic view of the frame.

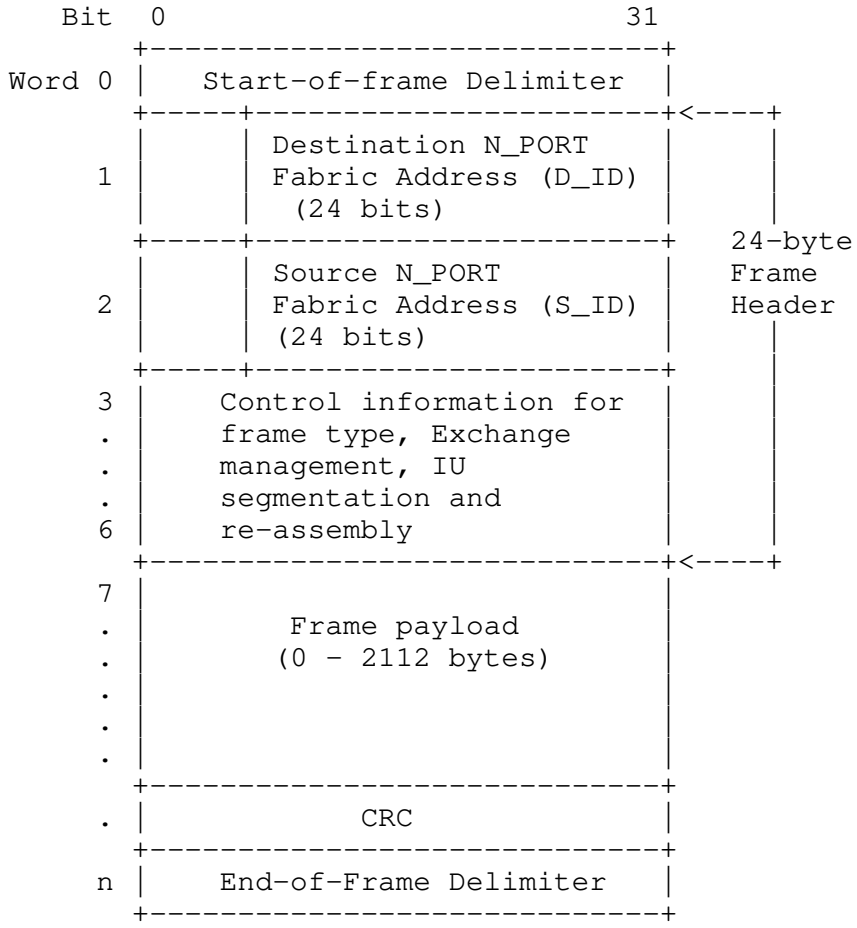


Figure 4. Fibre Channel Frame Format

The source and destination N_PORT fabric addresses embedded in the S_ID and D_ID fields represent the physical addresses of originating and receiving N_PORTs, respectively.

3.7.1. N_PORT Address Model

N_PORT fabric addresses are 24-bit values with the following format, defined by the fibre channel specification [FC-FS]:

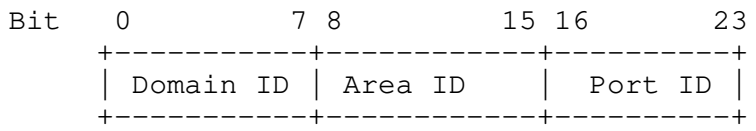


Figure 5. Fibre Channel Address Format

A fibre channel device acquires an address when it logs into the fabric. Such addresses are volatile and subject to change based on modifications in the fabric configuration.

In a fibre channel fabric, each switch element has a unique Domain ID assigned by the principal switch. The value of the Domain ID ranges from 1 to 239 (0xEF). Each switch element, in turn, administers a block of addresses divided into area and port IDs. An N_PORT connected to an F_PORT receives a unique fabric address, consisting of the switch's Domain ID concatenated with switch-assigned area and port IDs.

A loop-attached NL_PORT (see Figure 3) obtains the Port ID component of its address during the loop initialization process described in [FC-AL2]. The area and domain IDs are supplied by the fabric when the fabric login (FLOGI) is executed.

3.8. Fibre Channel Transport Services

N_PORTs communicate by means of the following classes of service, which are specified in the fibre channel standard ([FC-FS]):

Class 1 - A dedicated physical circuit connecting two N_PORTs.

Class 2 - A frame-multiplexed connection with end-to-end flow control and delivery confirmation.

Class 3 - A frame-multiplexed connection with no provisions for end-to-end flow control or delivery confirmation.

Class 4 -- A connection-oriented service, based on a virtual circuit model, providing confirmed delivery with bandwidth and latency guarantees.

Class 6 -- A reliable multicast service derived from class 1.

Classes 2 and 3 are the predominant services supported by deployed fibre channel storage and clustering systems.

Class 3 service is similar to UDP or IP datagram service. Fibre channel storage devices using this class of service rely on the ULP implementation to detect and recover from transient device and transport errors.

For class 2 and class 3 service, the fibre channel fabric is not required to provide in-order delivery of frames unless it is explicitly requested by the frame originator (and supported by the fabric). If ordered delivery is not in effect, it is the

responsibility of the frame recipient to reconstruct the order in which frames were sent, based on information in the frame header.

3.9. Login Processes

The Login processes are FC-2 operations that allow an N_PORT to establish the operating environment necessary to communicate with the fabric, other N_PORTs, and ULP implementations accessed via the N_PORT. Three login operations are supported:

a) Fabric Login (FLOGI) -- An operation whereby the N_PORT registers its presence on the fabric, obtains fabric parameters, such as classes of service supported, and receives its N_PORT address,

b) Port Login (PLOGI) -- An operation by which an N_PORT establishes communication with another N_PORT.

c) Process Login (PRLOGI) -- An operation that establishes the process-to-process communications associated with a specific FC-4 ULP, such as FCP-2, the fibre channel SCSI mapping.

Since N_PORT addresses are volatile, an N_PORT originating a login (PLOGI) operation executes a Name Server query to discover the fibre channel address of the remote device. A common query type involves use of the worldwide-unique name of an N_PORT to obtain the 24-bit N_PORT fibre channel address to which the PLOGI request is sent.

4. The iFCP Network Model

The iFCP protocol enables the implementation of fibre channel fabric functionality on an IP network in which IP components and technology replace the fibre channel switching and routing infrastructure described in Section 3.2.

The example of Figure 6 shows a fibre channel network with attached devices. Each device accesses the network through an N_PORT connected to an interface whose behavior is specified in [FC-FS] or [FC-AL2]. In this case, the N_PORT represents any of the variants described in Section 3.2. The interface to the fabric may be an L_PORT, F_PORT, or FL_PORT.

Within the fibre channel device domain, addressable entities consist of other N_PORTs and fibre channel devices internal to the network that perform the fabric services defined in [FC-GS3].

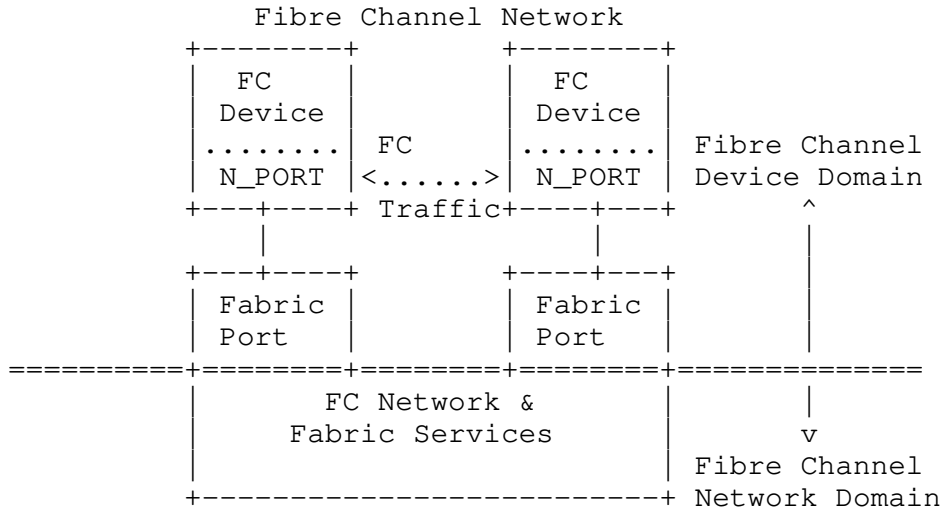


Figure 6. A Fibre Channel Network

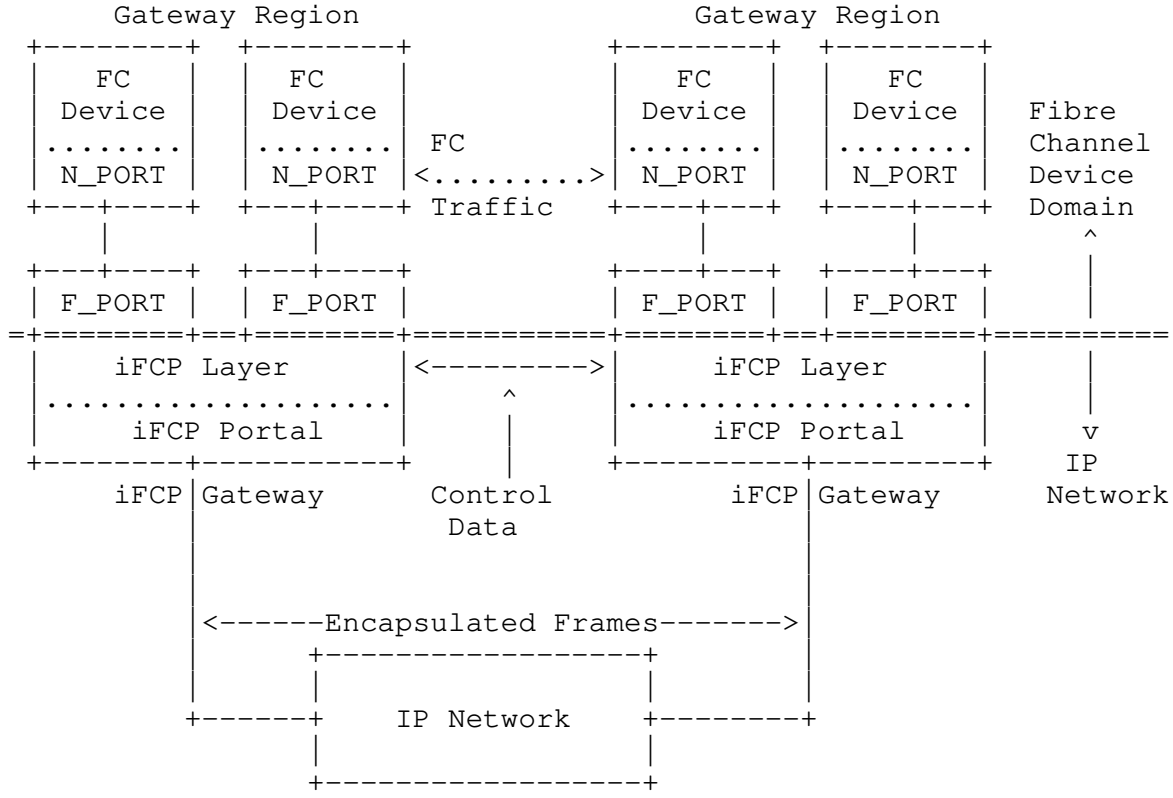


Figure 7. An iFCP Fabric Example

One example of an equivalent iFCP fabric is shown in Figure 7. The fabric consists of two gateway regions, each accessed by a single iFCP gateway.

Each gateway contains two standards-compliant F_PORTs and an iFCP Portal for attachment to the IP network. Fibre channel devices in the region are those locally connected to the iFCP fabric through the gateway fabric ports.

Looking into the fabric port, the gateway appears as a fibre channel switch element. At this interface, remote N_PORTs are presented as fabric-attached devices. Conversely, on the IP network side, the gateway presents each locally connected N_PORT as a logical fibre channel device.

Extrapolating to the general case, each gateway region behaves like an autonomous system whose configuration is invisible to the IP network and other gateway regions. Consequently, in addition to the F_PORT shown in the example, a gateway implementation may transparently support the following fibre channel interfaces:

Inter-Switch Link -- A fibre channel switch-to-switch interface used to access a region containing fibre channel switch elements. An implementation may support the E_PORT defined by [FC-SW2] or one of the proprietary interfaces provided by various fibre channel switch vendors. In this case, the gateway acts as a border switch connecting the gateway region to the IP network.

FL_PORT -- An interface that provides fabric access for loop-attached fibre channel devices, as specified in [FC-FLA].

L_PORT -- An interface through which a gateway may emulate the fibre channel loop environment specified in [FC-AL2]. As discussed in appendix B, the gateway presents remotely accessed N_PORTS as loop-attached devices.

The manner in which these interfaces are provided by a gateway is implementation specific and therefore beyond the scope of this document.

Although each region is connected to the IP network through one gateway, a region may incorporate multiple gateways for added performance and fault tolerance if the following conditions are met:

a) The gateways MUST coordinate the assignment of N_PORT IDs and aliases so that each N_PORT has one and only one address.

b) All iFCP traffic between a given remote and local N_PORT pair MUST flow through the same iFCP session (see Section 5.2.1). However, iFCP sessions to a given remotely attached N_PORT need not traverse the same gateway.

Coordinating address assignments and managing the flow of traffic is implementation specific and outside the scope of this specification.

4.1. iFCP Transport Services

N_PORT to N_PORT communications that traverse a TCP/IP network require the intervention of the iFCP layer within the gateway. This consists of the following operations:

a) Execution of the frame-addressing and -mapping functions described in Section 4.4.

b) Encapsulation of fibre channel frames for injection into the TCP/IP network and de-encapsulation of fibre channel frames received from the TCP/IP network.

c) Establishment of an iFCP session in response to a PLOGI directed to a remote device.

Section 4.4 discusses the iFCP frame-addressing mechanism and the way that it is used to achieve communications transparency between N_PORTs.

4.1.1. Fibre Channel Transport Services Supported by iFCP

An iFCP fabric supports Class 2 and Class 3 fibre channel transport services, as specified in [FC-FS]. An iFCP fabric does not support Class 4, Class 6, or Class 1 (dedicated connection) service. An N_PORT discovers the classes of transport services supported by the fabric during fabric login.

4.2. iFCP Device Discovery and Configuration Management

An iFCP implementation performs device discovery and iFCP fabric management through the Internet Storage Name Service defined in [ISNS]. Access to an iSNS server is required to perform the following functions:

a) Emulate the services provided by the fibre channel name server described in Section 3.3.1, including a mechanism for asynchronously notifying an N_PORT of changes in the iFCP fabric configuration.

b) Aggregate gateways into iFCP fabrics for interoperation.

c) Segment an iFCP fabric into fibre channel zones through the definition and management of device discovery scopes, referred to as 'discovery domains'.

d) Store and distribute security policies, as described in Section 10.2.9.

e) Implementation of the fibre channel broadcast mechanism.

4.3. iFCP Fabric Properties

A collection of iFCP gateways may be configured for interoperation as either a bounded or an unbounded iFCP fabric.

Gateways in a bounded iFCP fabric operate in address transparent mode, as described in Section 4.5. In this mode, the scope of a fibre channel N_PORT address is fabric-wide and is derived from domain IDs issued by the iSNS server from a common pool. As discussed in Section 4.3.2, the maximum number of domain IDs allowed by the fibre channel limits the configuration of a bounded iFCP fabric.

Gateways in an unbounded iFCP fabric operate in address translation mode as described in Section 4.6. In this mode, the scope of an N_PORT address is local to a gateway region. For fibre channel traffic between regions, the translation of frame-embedded N_PORT addresses is performed by the gateway. As discussed below, the number of switch elements and gateways in an unbounded iFCP fabric may exceed the limits of a conventional fibre channel fabric.

All iFCP gateways MUST support unbounded iFCP fabrics. Support for bounded iFCP fabrics is OPTIONAL.

The decision to support bounded iFCP fabrics in a gateway implementation depends on the address transparency, configuration scalability, and fault tolerance considerations given in the following sections.

4.3.1. Address Transparency

Although iFCP gateways in an unbounded fabric will convert N_PORT addresses in the frame header and payload of standard link service messages, a gateway cannot convert such addresses in the payload of vendor- or user-specific fibre channel frame traffic.

Consequently, although both bounded and unbounded iFCP fabrics support standards-compliant FC-4 protocol implementations and link services used by mainstream fibre channel applications, a bounded iFCP fabric may also support vendor- or user-specific protocol and link service implementations that carry N_PORT IDs in the frame payload.

4.3.2. Configuration Scalability

The scalability limits of a bounded fabric configuration are a consequence of the fibre channel address allocation policy discussed in Section 3.7.1. As noted, a bounded iFCP fabric using this address allocation scheme is limited to a combined total of 239 gateways and fibre channel switch elements. As the system expands, the network may grow to include many switch elements and gateways, each of which controls a small number of devices. In this case, the limitation in switch and gateway count may become a barrier to extending and fully integrating the storage network.

Since N_PORT fibre channel addresses in an unbounded iFCP fabric are not fabric-wide, the limits imposed by fibre channel address allocation only apply within the gateway region. Across regions, the number of iFCP gateways, fibre channel devices, and switch elements that may be internetworked are not constrained by these limits. In exchange for improved scalability, however, implementations must consider the incremental overhead of address conversion, as well as the address transparency issues discussed in Section 4.3.1.

4.3.3. Fault Tolerance

In a bounded iFCP fabric, address reassignment caused by a fault or reconfiguration, such as the addition of a new gateway region, may cascade to other regions, causing fabric-wide disruption as new N_PORT addresses are assigned. Furthermore, before a new gateway can be merged into the fabric, its iSNS server must be slaved to the iSNS server in the bounded fabric to centralize the issuance of domain IDs. In an unbounded iFCP fabric, coordinating the iSNS databases requires only that the iSNS servers exchange client attributes with one another.

A bounded iFCP fabric also has an increased dependency on the availability of the iSNS server, which must act as the central address assignment authority. If connectivity with the server is lost, new DOMAIN_ID values cannot be automatically allocated as gateways and fibre channel switch elements are added.

4.4. The iFCP N_PORT Address Model

This section discusses iFCP extensions to the fibre channel addressing model of Section 3.7.1, which are required for the transparent routing of frames between locally and remotely attached N_PORTs.

In the iFCP protocol, an N_PORT is represented by the following addresses:

a) A 24-bit N_PORT ID. The fibre channel N_PORT address of a locally attached device. Depending on the gateway addressing mode, the scope is local either to a region or to a bounded iFCP fabric. In either mode, communications between N_PORTs in the same gateway region use the N_PORT ID.

b) A 24-bit N_PORT alias. The fibre channel N_PORT address assigned by each gateway operating in address translation mode to identify a remotely attached N_PORT. Frame traffic is intercepted by an iFCP gateway and directed to a remotely attached N_PORT by means of the N_PORT alias. The address assigned by each gateway is unique within the scope of the gateway region.

c) An N_PORT network address. A tuple consisting of the gateway IP address, TCP port number, and N_PORT ID. The N_PORT network address identifies the source and destination N_PORTs for fibre channel traffic on the IP network.

To provide transparent communications between a remote and local N_PORT, a gateway MUST maintain an iFCP session descriptor (see Section 5.2.2.2) reflecting the association between the fibre channel address representing the remote N_PORT and the remote device's N_PORT network address. To establish this association, the iFCP gateway assigns and manages fibre channel N_PORT fabric addresses as described in the following paragraphs.

In an iFCP fabric, the iFCP gateway performs the address assignment and frame routing functions of an FC switch element. Unlike an FC switch, however, an iFCP gateway must also direct frames to external devices attached to remote gateways on the IP network.

In order to be transparent to FC devices, the gateway must deliver such frames using only the 24-bit destination address in the frame header. By exploiting its control of address allocation and access to frame traffic entering or leaving the gateway region, the gateway is able to achieve the necessary transparency.

N_PORT addresses within a gateway region may be allocated in one of two ways:

a) Address Translation Mode - A mode of N_PORT address assignment in which the scope of an N_PORT fibre channel address is unique to the gateway region. The address of a remote device is represented in that gateway region by its gateway-assigned N_PORT alias.

b) Address Transparent Mode - A mode of N_PORT address assignment in which the scope of an N_PORT fibre channel address is unique across the set of gateway regions comprising a bounded iFCP fabric.

In address transparent mode, gateways within a bounded fabric cooperate in the assignment of addresses to locally attached N_PORTs. Each gateway in control of a region is responsible for obtaining and distributing unique domain IDs from the address assignment authority, as described in Section 4.5.1. Consequently, within the scope of a bounded fabric, the address of each N_PORT is unique. For that reason, gateway-assigned aliases are not required for representing remote N_PORTs.

All iFCP implementations MUST support operations in address translation mode. Implementation of address transparent mode is OPTIONAL but, of course, must be provided if bounded iFCP fabric configurations are to be supported.

The mode of gateway operation is settable in an implementation-specific manner. The implementation MUST NOT:

a) allow the mode to be changed after the gateway begins processing fibre channel frame traffic,

b) permit operation in more than one mode at a time, or

c) establish an iFCP session with a gateway that is not in the same mode.

4.5. Operation in Address Transparent Mode

The following considerations and requirements apply to this mode of operation:

a) iFCP gateways in address transparent mode will not interoperate with iFCP gateways that are not in address transparent mode.

b) When interoperating with locally attached fibre channel switch elements, each iFCP gateway MUST assume control of DOMAIN_ID assignments in accordance with the appropriate fibre channel standard or vendor-specific protocol specification. As described in Section 4.5.1, DOMAIN_ID values that are assigned to FC switches internal to the gateway region must be issued by the iSNS server.

c) When operating in address transparent Mode, fibre channel address translation SHALL NOT take place.

When operating in address transparent mode, however, the gateway MUST establish and maintain the context of each iFCP session in accordance with Section 5.2.2.

4.5.1. Transparent Mode Domain ID Management

As described in Section 4.5, each gateway and fibre channel switch in a bounded iFCP fabric has a unique domain ID. In a gateway region containing fibre channel switch elements, each element obtains a domain ID by querying the principal switch as described in [FC-SW2] -- in this case, the iFCP gateway itself. The gateway, in turn, obtains domain IDs on demand from the iSNS name server acting as the central address allocation authority. In effect, the iSNS server assumes the role of principal switch for the bounded fabric. In that case, the iSNS database contains:

a) The definition for one or more bounded iFCP fabrics, and

b) For each bounded fabric, a worldwide-unique name identifying each gateway in the fabric. A gateway in address transparent mode MUST reside in one, and only one, bounded fabric.

As the Principal Switch within the gateway region, an iFCP gateway in address transparent mode SHALL obtain domain IDs for use in the gateway region by issuing the appropriate iSNS query, using its worldwide name.

4.5.2. Incompatibility with Address Translation Mode

Except for the session control frames specified in Section 6, iFCP gateways in address transparent mode SHALL NOT originate or accept frames that do not have the TRP bit set to one in the iFCP flags field of the encapsulation header (see Section 5.3.1). The iFCP gateway SHALL immediately terminate all iFCP sessions with the iFCP gateway from which it receives such frames.

4.6. Operation in Address Translation Mode

This section describes the process for managing the assignment of addresses within a gateway region that is part of an unbounded iFCP fabric, including the modification of FC frame addresses embedded in the frame header for frames sent and received from remotely attached N_PORTs.

As described in Section 4.4, the scope of N_PORT addresses in this mode is local to the gateway region. A principal switch within the gateway region, possibly the iFCP gateway itself, oversees the assignment of such addresses, in accordance with the rules specified in [FC-FS] and [FC-FLA].

The assignment of N_PORT addresses to locally attached devices is controlled by the switch element to which the device is connected.

The assignment of N_PORT addresses for remotely attached devices is controlled by the gateway by which the remote device is accessed. In this case, the gateway MUST assign a locally significant N_PORT alias to be used in place of the N_PORT ID assigned by the remote gateway. The N_PORT alias is assigned during device discovery, as described in Section 5.2.2.1.

To perform address conversion and to enable the appropriate routing, the gateway MUST establish an iFCP session and generate the information required to map each N_PORT alias to the appropriate TCP/IP connection context and N_PORT ID of the remotely accessed N_PORT. These mappings are created and updated by means specified in Section 5.2.2.2. As described in that section, the required mapping information is represented by the iFCP session descriptor reproduced in Figure 8.

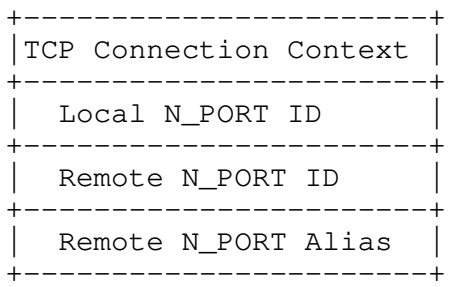


Figure 8. iFCP Session Descriptor (from Section 5.2.2.2)

Except for frames comprising special link service messages (see Section 7.2), outbound frames are encapsulated and sent without modification. Address translation is deferred until receipt from the IP network, as specified in Section 4.6.1.

4.6.1. Inbound Frame Address Translation

For inbound frames received from the IP network, the receiving gateway SHALL reference the session descriptor to fill in the D_ID field with the destination N_PORT ID and the S_ID field with the N_PORT alias it assigned. The translation process for inbound frames is shown in Figure 9.

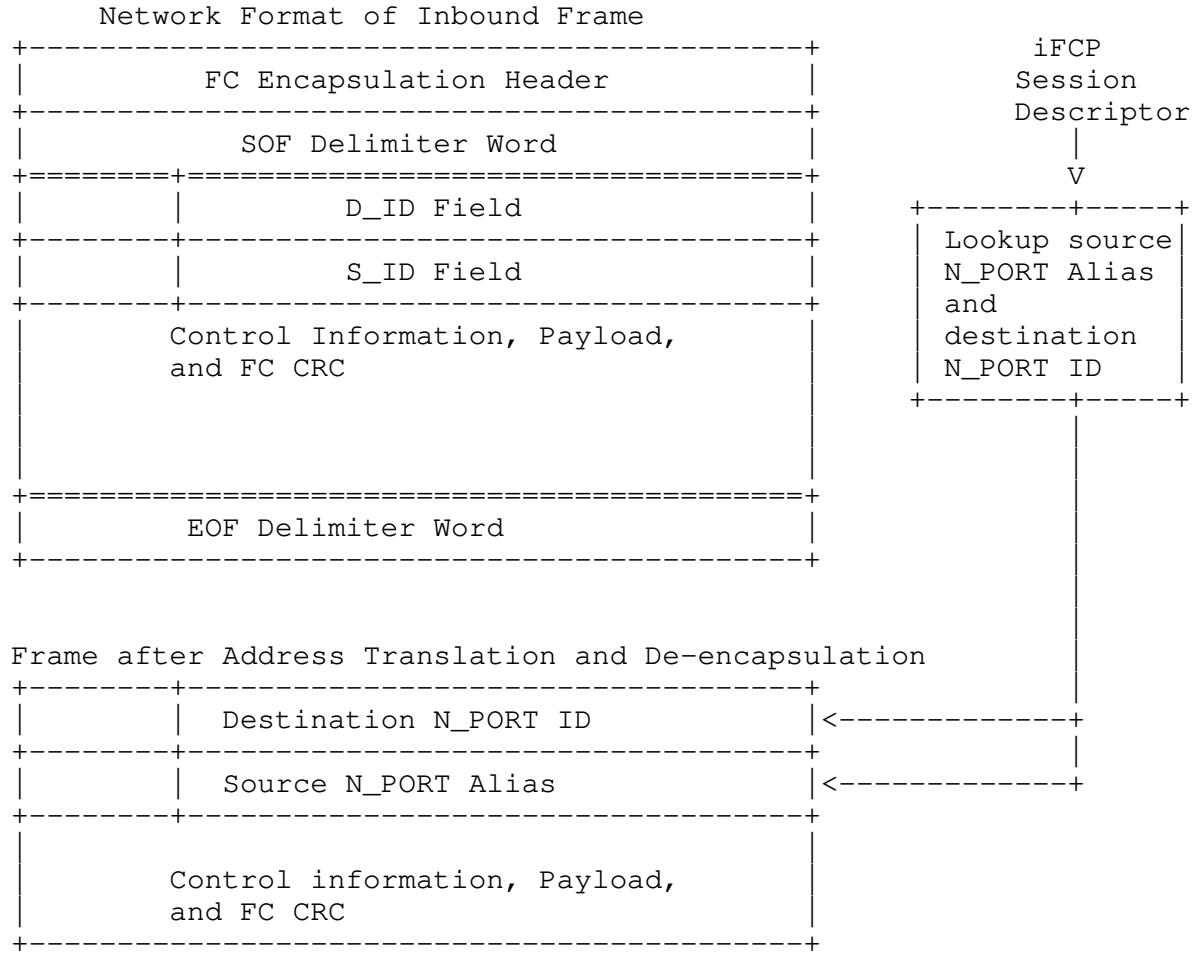


Figure 9. Inbound Frame Address Translation

The receiving gateway SHALL consider the contents of the S_ID and D_ID fields to be undefined when received. After replacing these fields, the gateway MUST recalculate the FC CRC.

4.6.2. Incompatibility with Address Transparent Mode

iFCP gateways in address translation mode SHALL NOT originate or accept frames that have the TRP bit set to one in the iFCP flags field of the encapsulation header. The iFCP gateway SHALL immediately abort all iFCP sessions with the iFCP gateway from which it receives frames such as those described in Section 5.2.3.

5. iFCP Protocol

5.1. Overview

5.1.1. iFCP Transport Services

The main function of the iFCP protocol layer is to transport fibre channel frame images between locally and remotely attached N_PORTs.

When transporting frames to a remote N_PORT, the iFCP layer encapsulates and routes the fibre channel frames comprising each fibre channel Information Unit via a predetermined TCP connection for transport across the IP network.

When receiving fibre channel frame images from the IP network, the iFCP layer de-encapsulates and delivers each frame to the appropriate N_PORT.

The iFCP layer processes the following types of traffic:

a) FC-4 frame images associated with a fibre channel application protocol.

b) FC-2 frames comprising fibre channel link service requests and responses.

c) Fibre channel broadcast frames.

d) iFCP control messages required to set up, manage, or terminate an iFCP session.

For FC-4 N_PORT traffic and most FC-2 messages, the iFCP layer never interprets the contents of the frame payload.

iFCP does interpret and process iFCP control messages and certain link service messages, as described in Section 5.1.2.

5.1.2. iFCP Support for Link Services

iFCP must intervene in the processing of those fibre channel link service messages that contain N_PORT addresses in the message payload or that require other special handling, such as an N_PORT login request (PLOGI).

In the former case, an iFCP gateway operating in address translation mode MUST supplement the payload with additional information that enables the receiving gateway to convert such embedded N_PORT addresses to its frame of reference.

For out bound fibre channel frames comprising such a link service, the iFCP layer creates the supplemental information based on frame content, modifies the frame payload, and then transmits the resulting fibre channel frame with supplemental data through the appropriate TCP connection.

For incoming iFCP frames containing supplemented fibre channel link service frames, iFCP must interpret the frame, including any supplemental information, modify the frame content, and forward the resulting frame to the destination N_PORT for further processing.

Section 7.1 describes the processing of these link service messages in detail.

5.2. TCP Stream Transport of iFCP Frames

5.2.1. iFCP Session Model

An iFCP session consists of the pair of N_PORTs comprising the session endpoints joined by a single TCP/IP connection. No more than one iFCP session SHALL exist between a given pair of N_PORTs.

An N_PORT is identified by its network address, consisting of:

a) the N_PORT ID assigned by the gateway to which the N_PORT is locally attached, and

b) the iFCP Portal address, consisting of its IP address and TCP port number.

Because only one iFCP session may exist between a pair of N_PORTs, the iFCP session is uniquely identified by the network addresses of the session end points.

TCP connections that may be used for iFCP sessions between pairs of iFCP portals are either "bound" or "unbound". An unbound connection

is a TCP connection that is not actively supporting an iFCP session. A gateway implementation MAY establish a pool of unbound connections to reduce the session setup time. Such pre-existing TCP connections between iFCP Portals remain unbound and uncommitted until allocated to an iFCP session through a CBIND message (see Section 6.1).

When the iFCP layer creates an iFCP session, it may select an existing unbound TCP connection or establish a new TCP connection and send the CBIND message down that TCP connection. This allocates the TCP connection to that iFCP session.

5.2.2. iFCP Session Management

This section describes the protocols and data structures required to establish and terminate an iFCP session.

5.2.2.1. The Remote N_PORT Descriptor

In order to establish an iFCP session, an iFCP gateway MUST maintain information allowing it to locate a remotely attached N_PORT. For explanatory purposes, such information is assumed to reside in a descriptor with the format shown in Figure 10.

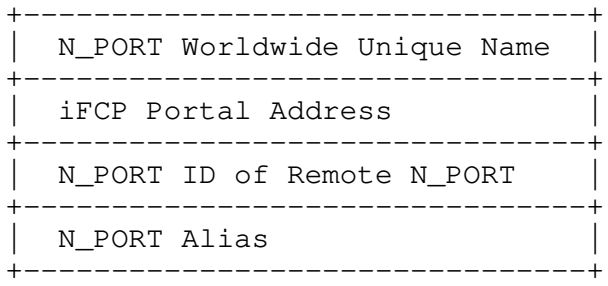


Figure 10. Remote N_PORT Descriptor

Each descriptor aggregates the following information about a remotely attached N_PORT:

N_PORT Worldwide Unique Name -- 64-bit N_PORT worldwide name as specified in [FC-FS]. A Remote N_PORT descriptor is uniquely identified by this parameter.

iFCP Portal Address -- The IP address and TCP port number referenced when creation of the TCP connection associated with an iFCP session is requested.

N_PORT ID -- N_PORT fibre channel address assigned to the remote device by the remote iFCP gateway.

N_PORT Alias -- N_PORT fibre channel address assigned to the remote device by the 'local' iFCP gateway when it operates in address translation mode.

An iFCP gateway SHALL have one and only one descriptor for each remote N_PORT it accesses. If a descriptor does not exist, one SHALL be created using the information returned by an iSNS name server query. Such queries may result from:

a) a fibre channel Name Server request originated by a locally attached N_PORT (see Sections 3.5 and 9.3), or

b) a CBIND request received from a remote fibre channel device (see Section 5.2.2.2).

When creating a descriptor in response to an incoming CBIND request, the iFCP gateway SHALL perform an iSNS name server query using the worldwide port name of the remote N_PORT in the SOURCE N_PORT NAME field within the CBIND payload. The descriptor SHALL be filled in using the query results.

After creating the descriptor, a gateway operating in address translation mode SHALL create and add the 24-bit N_PORT alias.

5.2.2.1.1. Updating a Remote N_PORT Descriptor

A Remote N_PORT descriptor SHALL only be updated as the result of an iSNS query to obtain information for the specified worldwide port name or from information returned by an iSNS state change notification. Following such an update, a new N_PORT alias SHALL NOT be assigned.

Before such an update, the contents of a descriptor may have become stale because of an event that invalidated or triggered a change in the N_PORT network address of the remote device, such as a fabric reconfiguration or the device's removal or replacement.

A collateral effect of such an event is that a fibre channel device that has been added or whose N_PORT ID has changed will have no active N_PORT logins. Consequently, FC-4 traffic directed to such an N_PORT, because of a stale descriptor, will be rejected or discarded.

Once the originating N_PORT learns of the reconfiguration, usually through the name server state change notification mechanism, information returned in the notification or the subsequent name server lookup needed to reestablish the iFCP session will automatically purge such stale data from the gateway.

5.2.2.1.2. Deleting a Remote N_PORT Descriptor

Deleting a remote N_PORT descriptor is equivalent to freeing up the corresponding N_PORT alias for reuse. Consequently, the descriptor MUST NOT be deleted while there are any iFCP sessions that reference the remote N_PORT.

Descriptors eligible for deletion should be removed based on a last in, first out policy.

5.2.2.2. Creating an iFCP Session

An iFCP session may be in one of the following states:

OPEN -- The session state in which fibre channel frame images may be sent and received.

OPEN PENDING -- The session state after a gateway has issued a CBIND request but no response has yet been received. No fibre channel frames may be sent.

The session may be initiated in response to a PLOGI ELS (see Section 7.3.1.7) or for any other implementation-specific reason.

The gateway SHALL create the iFCP session as follows:

a) Locate the remote N_PORT descriptor corresponding to the session end point. If the session is created in order to forward a fibre channel frame, then the session endpoint may be obtained by referencing the remote N_PORT alias contained in the frame header D_ID field. If no descriptor exists, an iFCP session SHALL NOT be created.

b) Allocate a TCP connection to the gateway to which the remote N_PORT is locally attached. An implementation may use an existing connection in the Unbound state, or a new connection may be created and placed in the Unbound state.

   When a connection is created, the IP address and TCP Port number SHALL be obtained by referencing the remote N_PORT descriptor as specified in Section 5.2.2.1.

c) If the TCP connection cannot be allocated or cannot be created due to limited resources, the gateway SHALL terminate session creation.

d) If the TCP connection is aborted for any reason before the iFCP session enters the OPEN state, the gateway SHALL respond in accordance with Section 5.2.3 and MAY terminate the attempt to create a session or MAY try to establish the TCP connection again.

e) The gateway SHALL then issue a CBIND session control message (see Section 6.1) and place the session in the OPEN PENDING state.

f) If a CBIND response is returned with a status other than "Success" or "iFCP session already exists", the session SHALL be terminated, and the TCP connection returned to the Unbound state.

g) A CBIND STATUS of "iFCP session already exists" indicates that the remote gateway has concurrently initiated a CBIND request to create an iFCP session between the same pair of N_PORTs. A gateway receiving such a response SHALL terminate this attempt and process the incoming CBIND request in accordance with Section 5.2.2.3.

h) In response to a CBIND STATUS of "Success", the gateway SHALL place the session in the OPEN state.

Once the session is placed in the OPEN state, an iFCP session descriptor SHALL be created, containing the information shown in Figure 11:

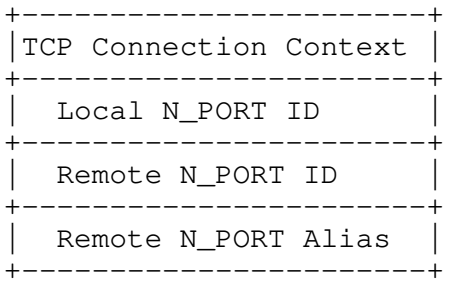


Figure 11. iFCP Session Descriptor

TCP Connection Context -- Information required to identify the TCP connection associated with the iFCP session.

Local N_PORT ID -- N_PORT ID of the locally attached fibre channel device.

Remote N_PORT ID -- N_PORT ID assigned to the remote device by the remote gateway.

Remote N_PORT Alias -- Alias assigned to the remote N_PORT by the local gateway when it operates in address translation mode. If in this mode, the gateway SHALL copy this parameter from the Remote N_PORT descriptor. Otherwise, it is not filled in.

5.2.2.3. Responding to a CBIND Request

The gateway receiving a CBIND request SHALL respond as follows:

a) If the receiver has a duplicate iFCP session in the OPEN PENDING state, then the receiving gateway SHALL compare the Source N_PORT Name in the incoming CBIND payload with the Destination N_PORT Name.

b) If the Source N_PORT Name is greater, the receiver SHALL issue a CBIND response of "Success" and SHALL place the session in the OPEN state.

c) If the Source N_PORT Name is less, the receiver shall issue a CBIND RESPONSE of Failed - N_PORT session already exists. The state of the receiver-initiated iFCP session SHALL BE unchanged.

d) If there is no duplicate iFCP session in the OPEN PENDING state, the receiving gateway SHALL issue a CBIND response. If a status of Success is returned, the receiving gateway SHALL create the iFCP session and place it in the OPEN state. An iFCP session descriptor SHALL be created as described in Section 5.2.2.2.

e) If a remote N_PORT descriptor does not exist, one SHALL be created and filled in as described in Section 5.2.2.1.

5.2.2.4. Monitoring iFCP Connectivity

During extended periods of inactivity, an iFCP session may be terminated due to a hardware failure within the gateway or through loss of TCP/IP connectivity. The latter may occur when the session traverses a stateful intermediate device, such as a NA(P)T box or firewall, that detects and purges connections it believes are unused.

To test session liveness, expedite the detection of connectivity failures, and avoid spontaneous connection termination, an iFCP gateway may maintain a low level of session activity and monitor the session by requesting that the remote gateway periodically transmit the LTEST message described in Section 6.3. All iFCP gateways SHALL support liveness testing as described in this specification.

A gateway requests the LTEST heartbeat by specifying a non-zero value for the LIVENESS TEST INTERVAL in the CBIND request or response message as described in Section 6.1. If both gateways seek to monitor liveness, each must set the LIVENESS TEST INTERVAL in the CBIND request or response.

Upon receiving such a request, the gateway providing the heartbeat SHALL transmit LTEST messages at the specified interval. The first message SHALL be sent as soon as the iFCP session enters the OPEN state. LTEST messages SHALL NOT be sent when the iFCP session is not in the OPEN state.

An iFCP session SHALL be terminated as described in Section 5.2.3 if:

a) the contents of the LTEST message are incorrect, or

b) an LTEST message is not received within twice the specified interval or the iFCP session has been quiescent for longer than twice the specified interval.

The gateway to receive the LTEST message SHALL measure the interval for the first expected LTEST message from when the session is placed in the OPEN state. Thereafter, the interval SHALL be measured relative to the last LTEST message received.

To maximize liveness test coverage, LTEST messages SHOULD flow through all the gateway components used to enter and retrieve fibre channel frames from the IP network, including the mechanisms for encapsulating and de-encapsulating fibre channel frames.

In addition to monitoring a session, information in the LTEST message encapsulation header may also be used to compute an estimate of network propagation delay, as described in Section 8.2.1. However, the propagation delay limit SHALL NOT be enforced for LTEST traffic.

5.2.2.5. Use of TCP Features and Settings

This section describes ground rules for the use of TCP features in an iFCP session. The core TCP protocol is defined in [RFC793]. TCP implementation requirements and guidelines are specified in [RFC1122].

Feature	Applicable RFCs	RFC Status	Peer-Wise Agreement Required?	Requirement Level
Keep Alive	[RFC1122] (discussion)	None	No	Should not use
Tiny Segment Avoidance (Nagle)	[RFC896]	Standard	No	Should not use
Window Scale	[RFC1323]	Proposed Standard	No	Should use
Wrapped Sequence Protection (PAWS)	[RFC1323]	Proposed Standard	No	SHOULD use

Table 1. Usage of Optional TCP Features

The following sections describe these options in greater detail.

5.2.2.5.1. Keep Alive

Keep Alive speeds the detection and cleanup of dysfunctional TCP connections by sending traffic when a connection would otherwise be idle. The issues are discussed in [RFC1122].

In order to test the device more comprehensively, fibre channel applications, such as storage, may implement an equivalent keep alive function at the FC-4 level. Alternatively, periodic liveness test messages may be issued as described in Section 5.2.2.4. Because of these more comprehensive end-to-end mechanisms and the considerations described in [RFC1122], keep alive at the transport layer should not be implemented.

5.2.2.5.2. 'Tiny' Segment Avoidance (Nagle)

The Nagle algorithm described in [RFC896] is designed to avoid the overhead of small segments by delaying transmission in order to agglomerate transfer requests into a large segment. In iFCP, such small transfers often contain I/O requests. The transmission delay of the Nagle algorithm may decrease I/O throughput. Therefore, the Nagle algorithm should not be used.

5.2.2.5.3. Window Scale

Window scaling, as specified in [RFC1323], allows full use of links with large bandwidth - delay products and should be supported by an iFCP implementation.

5.2.2.5.4. Wrapped Sequence Protection (PAWS)

TCP segments are identified with 32-bit sequence numbers. In networks with large bandwidth - delay products, it is possible for more than one TCP segment with the same sequence number to be in flight. In iFCP, receipt of such a sequence out of order may cause out-of-order frame delivery or data corruption. Consequently, this feature SHOULD be supported as described in [RFC1323].

5.2.3. Terminating iFCP Sessions

iFCP sessions SHALL be terminated in response to one of the events in Table 2:

Event	iFCP Sessions to Terminate
PLOGI terminated with LS_RJT response	Peer N_PORT
State change notification indicating N_PORT removal or reconfiguration.	All iFCP Sessions from the reconfigured N_PORT
LOGO ACC response from peer N_PORT	Peer N_PORT
ACC response to LOGO ELS sent to F_PORT server (D_ID = 0xFF-FF-FE) (fabric logout)	All iFCP sessions from the originating N_PORT
Implicit N_PORT LOGO as defined in [FC-FS]	All iFCP sessions from the N_PORT logged out
LTEST Message Error (see Section 5.2.2.4)	Peer N_PORT
Non fatal encapsulation error as specified in Section 5.3.3	Peer N_PORT
Failure of the TCP connection associated with the iFCP session	Peer N_PORT
Receipt of an UNBIND session control message	Peer N_PORT
Gateway enters the Unsynchronized state (see Section 8.2.1)	All iFCP sessions
Gateway detects incorrect address mode to peer gateway(see Section 4.6.2)	All iFCP sessions with peer gateway

Table 2. Session Termination Events

If a session is being terminated due to an incorrect address mode with the peer gateway, the TCP connection SHALL be aborted by means of a connection reset (RST) without performing an UNBIND. Otherwise, if the TCP connection is still open following the event, the gateway SHALL shut down the connection as follows:

a) Stop sending fibre channel frames over the TCP connection.

b) Discard all incoming traffic, except for an UNBIND session control message.

c) If an UNBIND message is received at any time, return a response in accordance with Section 6.2.

d) If session termination was not triggered by an UNBIND message, issue the UNBIND session control message, as described in Section 6.2.

e) If the UNBIND message completes with a status of Success, the TCP connection MAY remain open at the discretion of either gateway and may be kept in a pool of unbound connections in order to speed up the creation of a new iFCP session.

   If the UNBIND fails for any reason, the TCP connection MUST be terminated. In this case, the connection SHOULD be aborted with a connection reset (RST).

For each terminated session, the session descriptor SHALL be deleted. If a session was terminated by an event other than an implicit LOGO or a LOGO ACC response, the gateway shall issue a LOGO to the locally attached N_PORT on behalf of the remote N_PORT.

To recover resources, either gateway may spontaneously close an unbound TCP connection at any time. If a gateway terminates a connection with a TCP close operation, the peer gateway MUST respond by executing a TCP close.

5.3. Fibre Channel Frame Encapsulation

This section describes the iFCP encapsulation of fibre channel frames. The encapsulation complies with the common encapsulation format defined in [ENCAP], portions of which are included here for convenience.

The format of an encapsulated frame is shown below:

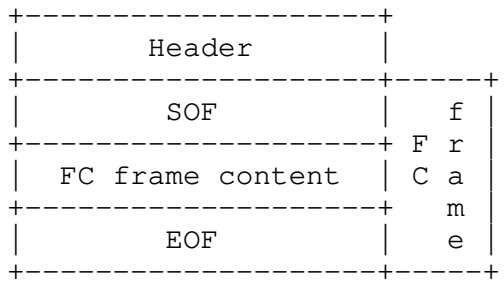


Figure 12. Encapsulation Format

The encapsulation consists of a 7-word header, an SOF delimiter word, the FC frame (including the fibre channel CRC), and an EOF delimiter word. The header and delimiter formats are described in the following sections.

5.3.1. Encapsulation Header Format

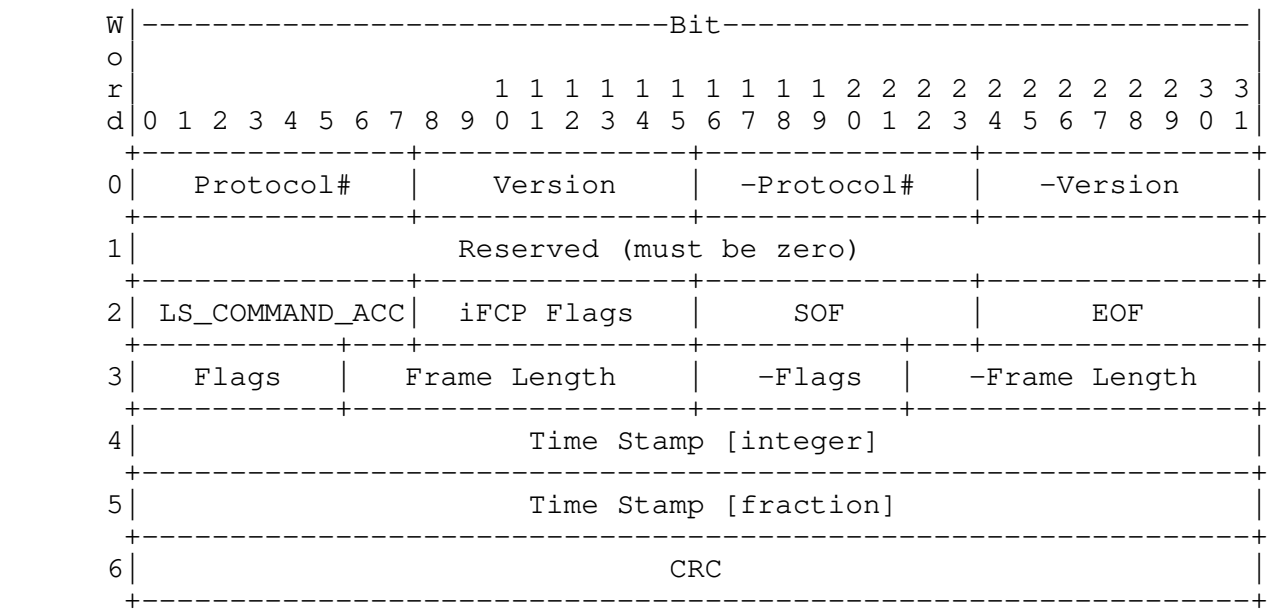


Figure 13. Encapsulation Header Format

Common Encapsulation Fields:

Protocol# IANA-assigned protocol number identifying the protocol using the encapsulation. For iFCP, the value assigned by [ENCAP] is 2.

Version	Encapsulation version, as specified in [ENCAP].
-Protocol#	Ones complement of the Protocol#.
-Version	Ones complement of the version.
Flags	Encapsulation flags (see 5.3.1.1).
Frame Length	Contains the length of the entire FC Encapsulated frame, including the FC Encapsulation Header and the FC frame (including SOF and EOF words) in units of 32-bit words.
-Flags	Ones complement of the Flags field.
-Frame Length	Ones complement of the Frame Length field.
Time Stamp [integer]	Integer component of the frame time stamp, as specified in [ENCAP].
Time Stamp [fraction]	Fractional component of the time stamp, as specified in [ENCAP].
CRC	Header CRC. MUST be valid for iFCP.

The time stamp fields are used to enforce the limit on the lifetime of a fibre channel frame as described in Section 8.2.1.

iFCP-Specific Fields:

LS_COMMAND_ACC	For a special link service ACC response to be processed by iFCP, the LS_COMMAND_ACC field SHALL contain a copy of bits 0 through 7 of the LS_COMMAND to which the ACC applies. Otherwise, the LS_COMMAND_ACC field SHALL be set to zero.
iFCP Flags	iFCP-specific flags (see below).
SOF	Copy of the SOF delimiter encoding (see Section 5.3.2).
EOF	Copy of the EOF delimiter encoding (see Section 5.3.2).

The iFCP flags word has the following format:

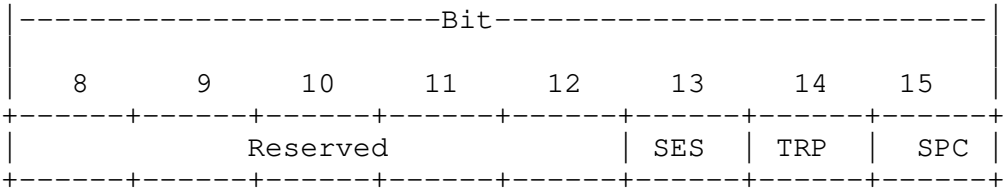


Figure 14. iFCP Flags Word

iFCP Flags:

SES 1 = Session control frame (TRP and SPC MUST be 0)

TRP 1 = Address transparent mode enabled

0 = Address translation mode enabled

SPC 1 = Frame is part of a link service message requiring special processing by iFCP prior to forwarding to the destination N_PORT.

5.3.1.1. Common Encapsulation Flags

The iFCP usage of the common encapsulation flags defined in [ENCAP] is shown in Figure 15:

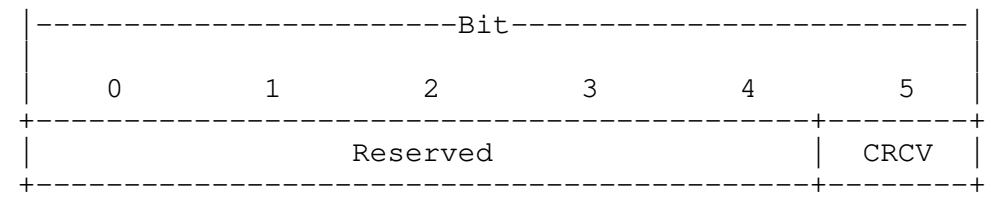


Figure 15. iFCP Common Encapsulation Flags

For iFCP, the CRC field MUST be valid, and CRCV MUST be set to one.

5.3.2. SOF and EOF Delimiter Fields

The format of the delimiter fields is shown below.

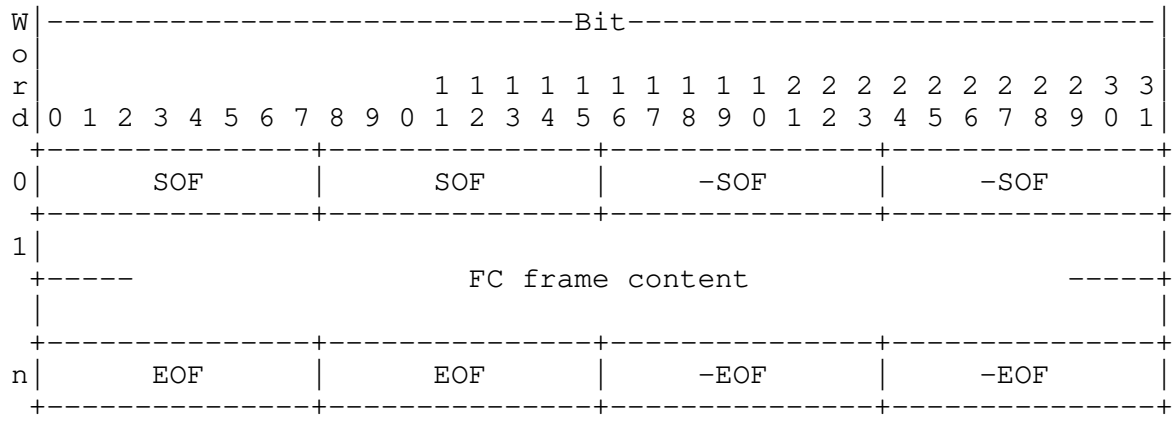


Figure 16. FC Frame Encapsulation Format

SOF (bits 0-7 and bits 8-15 in word 0): iFCP uses the following subset of the SOF fields specified in [ENCAP]. For convenience, these are reproduced in Table 3. The authoritative encodings should be obtained from [ENCAP].

FC SOF	SOF Code
SOFi2	0x2D
SOFn2	0x35
SOFi3	0x2E
SOFn3	0x36

Table 3. Translation of FC SOF Values to SOF Field Contents

-SOF (bits 16-23 and 24-31 in word 0): The -SOF fields contain the ones complement the value in the SOF fields.

EOF (bits 0-7 and 8-15 in word n): iFCP uses the following subset of EOF fields specified in [ENCAP]. For convenience, these are reproduced in Table 4. The authoritative encodings should be obtained from [ENCAP].

FC EOF	EOF Code
EOFn	0x41
EOFt	0x42

Table 4. Translation of FC EOF Values to EOF Field Contents

-EOF (bits 16-23 and 24-31 in word n): The -EOF fields contain the ones complement the value in the EOF fields.

iFCP implementations SHALL place a copy of the SOF and EOF delimiter codes in the appropriate header fields.

5.3.3. Frame Encapsulation

A fibre channel Frame to be encapsulated MUST first be validated as described in [FC-FS]. Any frames received from a locally attached fibre channel device that do not pass the validity tests in [FC-FS] SHALL be discarded by the gateway.

If the frame is a PLOGI ELS, the creation of an iFCP session, as described in Section 7.3.1.7, may precede encapsulation. Once the session has been created, frame encapsulation SHALL proceed as follows.

The S_ID and D_ID fields in the frame header SHALL be referenced to look up the iFCP session descriptor (see Section 5.2.2.2). If no iFCP session descriptor exists, the frame SHALL be discarded.

Frame types submitted for encapsulation and forwarding on the IP network SHALL have one of the SOF delimiters in Table 3 and an EOF delimiter from Table 4. Other valid frame types MUST be processed internally by the gateway as specified in the appropriate fibre channel specification.

If operating in address translation mode and processing a special link service message requiring the inclusion of supplemental data, the gateway SHALL format the frame payload and add the supplemental information specified in Section 7.1. The gateway SHALL then calculate a new FC CRC on the reformatted frame.

Otherwise, the frame contents SHALL NOT be modified and the gateway MAY encapsulate and transmit the frame image without recalculating the FC CRC.

The frame originator MUST then create and fill in the header and the SOF and EOF delimiter words, as specified in Sections 5.3.1 and 5.3.2.

5.3.4. Frame De-encapsulation

The receiving gateway SHALL perform de-encapsulation as follows:

Upon receiving the encapsulated frame, the gateway SHALL check the header CRC. If the header CRC is valid, the receiving gateway SHALL check the iFCP flags field. If one of the error conditions in Table 5 is detected, the gateway SHALL handle the error as specified in Section 5.2.3.

Condition	Error Type
Header CRC Invalid	Encapsulation error
SES = 1, TRP or SPC not 0	Encapsulation error
SES = 0, TRP set incorrectly	Incorrect address mode

Table 5. Encapsulation Header Errors

The receiving gateway SHALL then verify the frame propagation delay as described in Section 8.2.1. If the propagation delay is too long, the frame SHALL be discarded. Otherwise, the gateway SHALL check the SOF and EOF in the encapsulation header. A frame SHALL be discarded if it has an SOF code that is not in Table 3 or an EOF code that is not in Table 4.

The gateway SHALL then de-encapsulate the frame as follows:

a) Check the FC CRC and discard the frame if the CRC is invalid.

b) If operating in address translation mode, replace the S_ID field with the N_PORT alias of the frame originator, and the D_ID with the N_PORT ID, of the frame recipient. Both parameters SHALL be obtained from the iFCP session descriptor.

c) If processing a special link service message, replace the frame with a copy whose payload has been modified as specified in Section 7.1.

The de-encapsulated frame SHALL then be forwarded to the N_PORT specified in the D_ID field. If the frame contents have been modified by the receiving gateway, a new FC CRC SHALL be calculated.

6. TCP Session Control Messages

TCP session control messages are used to create and manage an iFCP session as described in Section 5.2.2. They are passed between peer iFCP Portals and are only processed within the iFCP layer.

The message format is based on the fibre channel extended link service message template shown below.

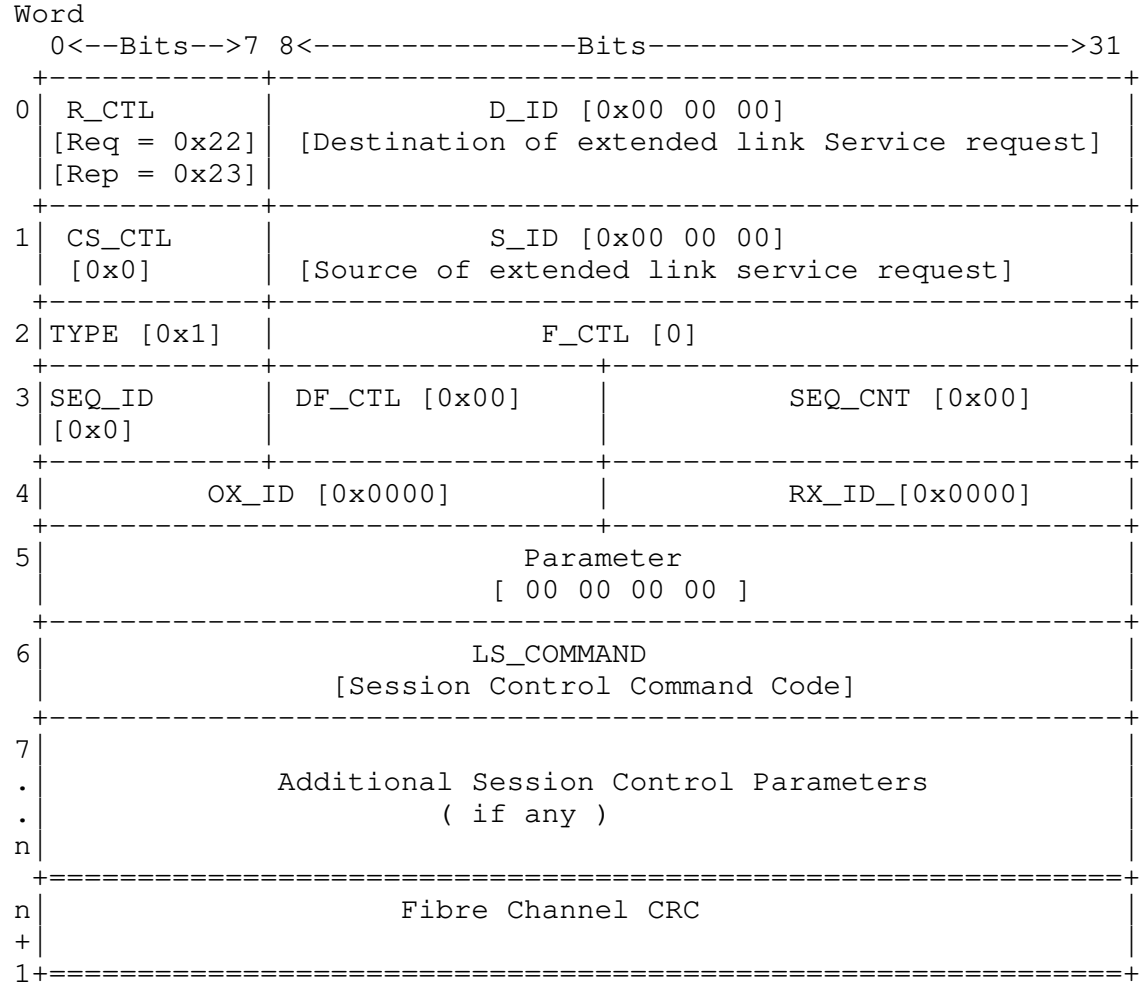


Figure 17. Format of Session Control Message

The LS_COMMAND value for the response remains the same as that used for the request.

The session control frame is terminated with a fibre channel CRC. The frame SHALL be encapsulated and de-encapsulated according to the rules specified in Section 5.3.

The encapsulation header for the link Service frame carrying a session control message SHALL be set as follows:

Encapsulation Header Fields:

LS_COMMAND_ACC	0
iFCP Flags	SES = 1
	TRP = 0
	INT = 0
SOF code	SOFi3 encoding (0x2E)
EOF code	EOFt encoding (0x42)

The encapsulation time stamp words SHALL be set as described for each message type.

The SOF and EOF delimiter words SHALL be set based on the SOF and EOF codes specified above.

Table 6 lists the values assigned to byte 0 of the LS_COMMAND field for iFCP session control messages.

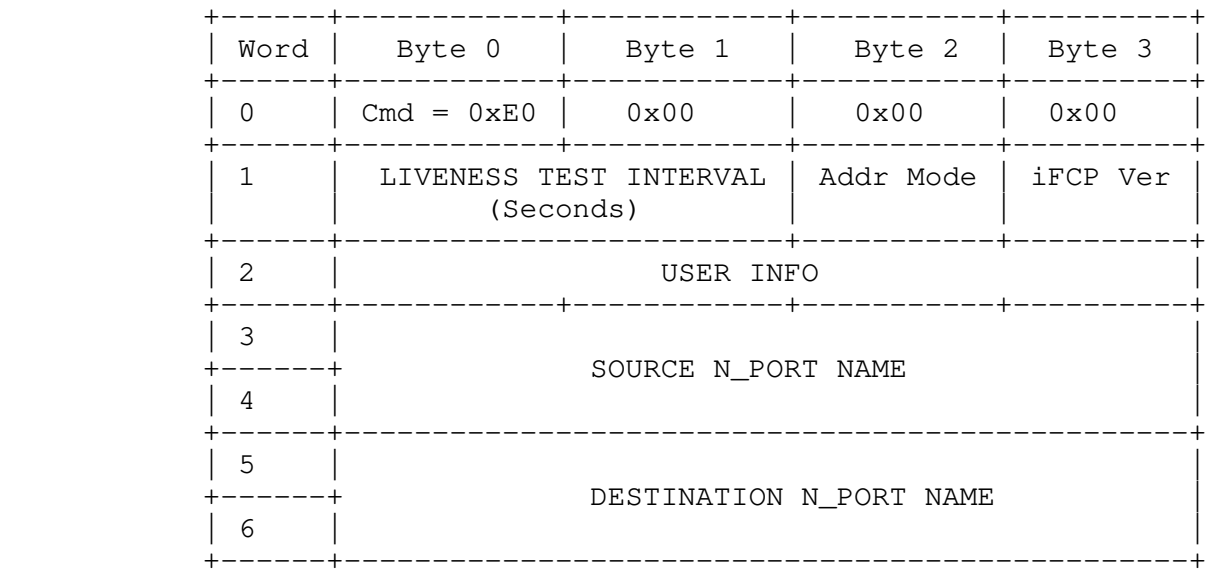
LS_COMMAND field, byte 0	Function	Mnemonic	iFCP Support
0xE0	Connection Bind	CBIND	REQUIRED
0xE4	Unbind Connection	UNBIND	REQUIRED
0xE5	Test Connection Liveness	LTEST	REQUIRED
0x01-0x7F	Vendor-Specific		
0x00	Reserved -- Unassignable		
All other values	Reserved		

Table 6. Session Control LS_COMMAND Field, Byte 0 Values

6.1. Connection Bind (CBIND)

As described in Section 5.2.2.2, the CBIND message and response are used to bind an N_PORT login to a specific TCP connection and establish an iFCP session. In the CBIND request message, the source and destination N_PORTs are identified by their worldwide port names. The time stamp words in the encapsulation header SHALL be set to zero in the request and response message frames.

The following shows the format of the CBIND request.



Addr Mode: The addressing mode of the originating gateway. 0 = Address Translation mode; 1 = Address Transparent mode.

iFCP Ver: iFCP version number. SHALL be set to 1.

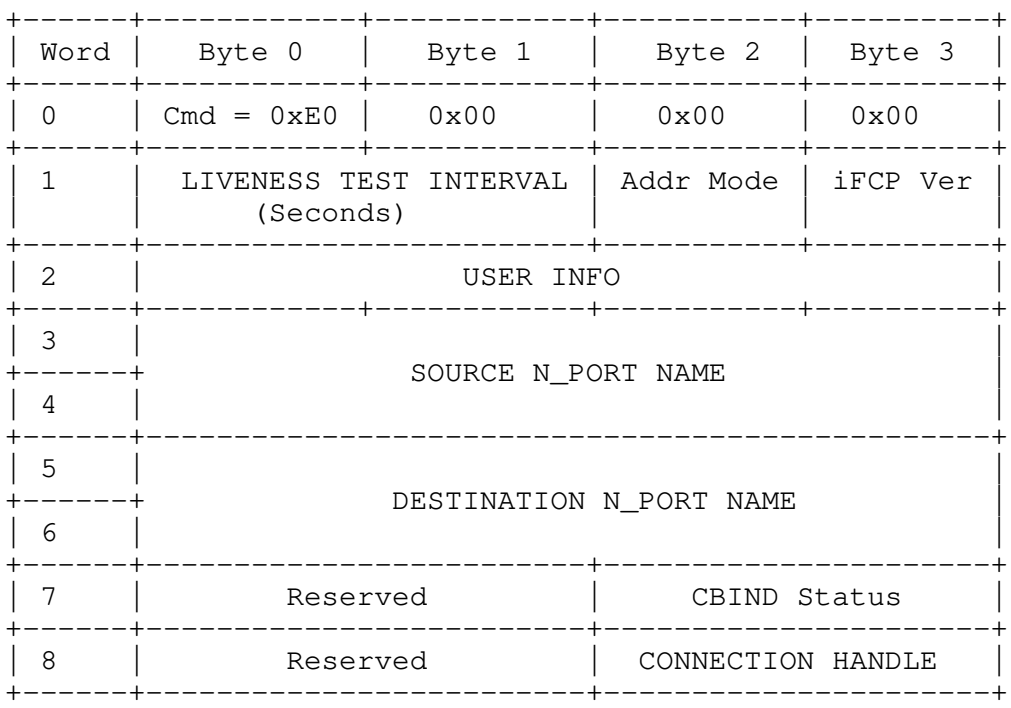
LIVENESS TEST INTERVAL: If non-zero, requests that the receiving gateway transmit an LTEST message at the specified interval in seconds. If set to zero, LTEST messages SHALL NOT be sent.

USER INFO: Contains any data desired by the requestor. This information MUST be echoed by the recipient in the CBIND response message.

SOURCE N_PORT NAME: The Worldwide Port Name (WWPN) of the N_PORT locally attached to the gateway originating the CBIND request.

DESTINATION N_PORT NAME: The Worldwide Port Name (WWPN) of the N_PORT locally attached to the gateway receiving the CBIND request.

The following shows the format of the CBIND response.



Total Length = 36

Addr Mode: The address translation mode of the responding gateway. 0 = Address Translation mode, 1 = Address Transparent mode.

iFCP Ver: iFCP version number. Shall be set to 1.

LIVENESS TEST INTERVAL: If non-zero, requests that the gateway receiving the CBIND RESPONSE transmit an LTEST message at the specified interval in seconds. If zero, LTEST messages SHALL NOT be sent.

USER INFO: Echoes the value received in the USER INFO field of the CBIND request message.

SOURCE N_PORT NAME:	Contains the Worldwide Port Name (WWPN) of the N_PORT locally attached to the gateway issuing the CBIND request.
DESTINATION N_PORT NAME:	Contains the Worldwide Port Name (WWPN) of the N_PORT locally attached to the gateway issuing the CBIND response.
CBIND STATUS:	Indicates success or failure of the CBIND request. CBIND values are shown below.
CONNECTION HANDLE:	Contains a value assigned by the gateway to identify the connection. The connection handle is required when the UNBIND request is issued.

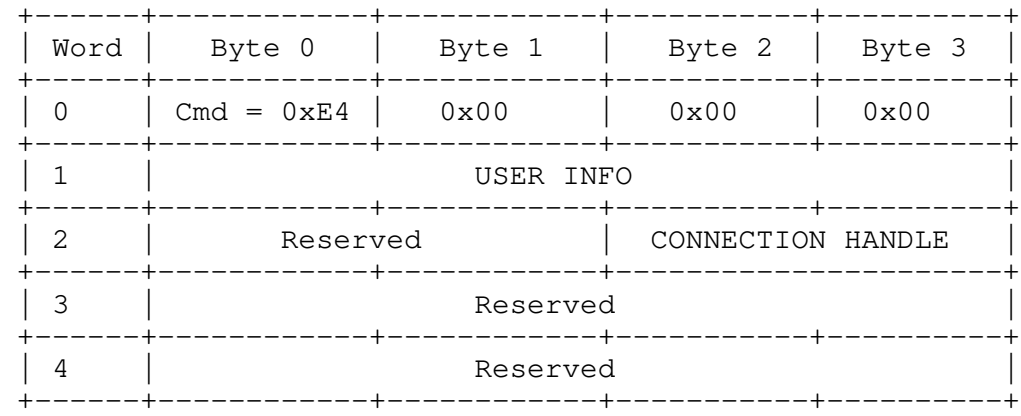
CBIND Status	Description
0	Success
1 - 15	Reserved
16	Failed - Unspecified Reason
17	Failed - No such device
18	Failed - iFCP session already exists
19	Failed - Lack of resources
20	Failed - Incompatible address translation mode
21	Failed - Incorrect protocol version number
22	Failed - Gateway not Synchronized (see Section 8.2)
Others	Reserved

6.2. Unbind Connection (UNBIND)

UNBIND is used to terminate an iFCP session and disassociate the TCP connection as described in Section 5.2.3.

The UNBIND message is transmitted over the connection that is to be unbound. The time stamp words in the encapsulation header shall be set to zero in the request and response message frames.

The following is the format of the UNBIND request message.

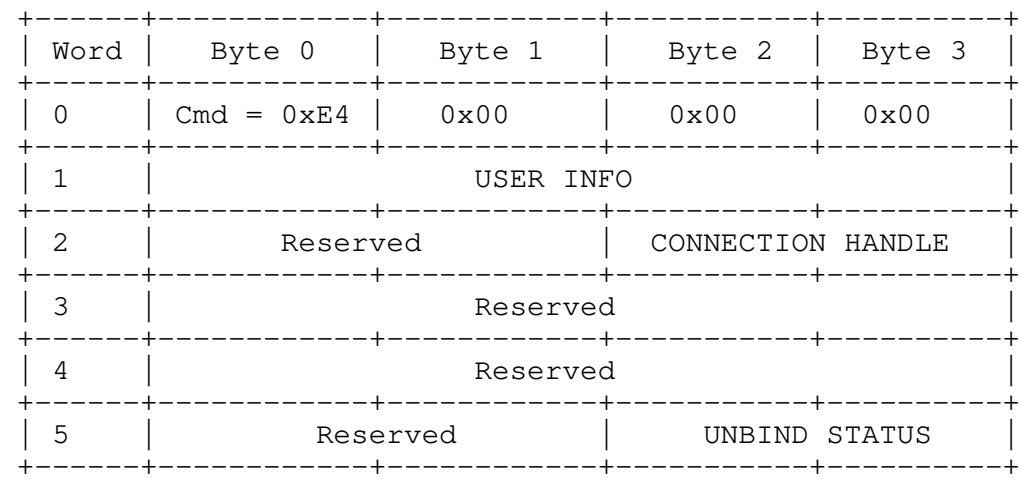


Word	Byte 0	Byte 1	Byte 2	Byte 3
0	Cmd = 0xE4	0x00	0x00	0x00
1	USER INFO			
2	Reserved		CONNECTION HANDLE	
3	Reserved			
4	Reserved			

USER INFO — Contains any data desired by the requestor. This information MUST be echoed by the recipient in the UNBIND response message.

CONNECTION HANDLE: — Contains the gateway-assigned value from the CBIND request.

The following shows the format of the UNBIND response message.



Word	Byte 0	Byte 1	Byte 2	Byte 3
0	Cmd = 0xE4	0x00	0x00	0x00
1	USER INFO			
2	Reserved		CONNECTION HANDLE	
3	Reserved			
4	Reserved			
5	Reserved		UNBIND STATUS	

USER INFO — Echoes the value received in the USER INFO field of the UNBIND request message.

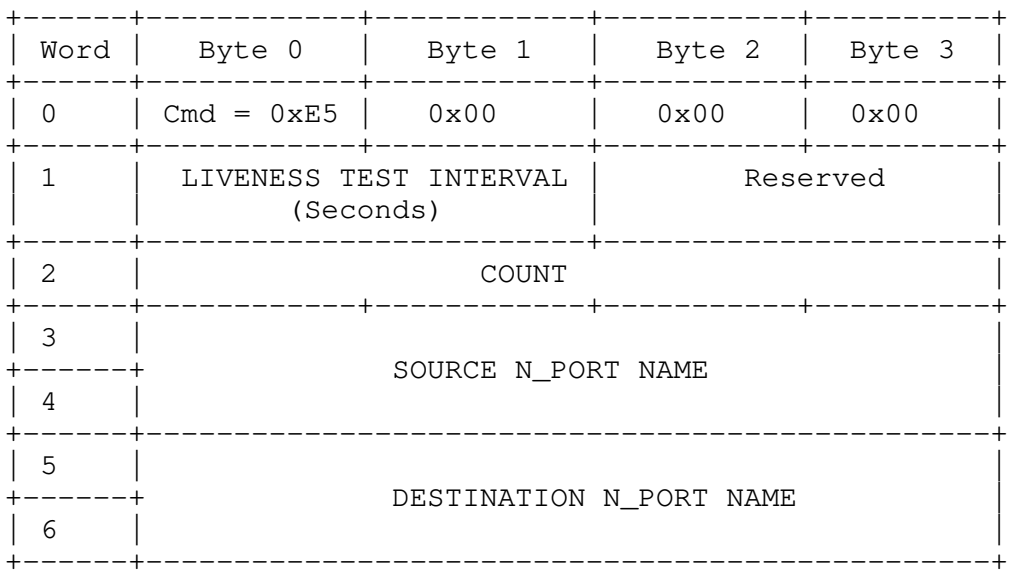
CONNECTION HANDLE: — Echoes the CONNECTION HANDLE specified in the UNBIND request message.

UNBIND STATUS: Indicates the success or failure of the UNBIND request as follows:

Unbind Status	Description
0	Successful - No other status
1 - 15	Reserved
16	Failed - Unspecified Reason
18	Failed - Connection ID Invalid
Others	Reserved

6.3. LTEST -- Test Connection Liveness

The LTEST message is sent at the interval specified in the CBIND request or response payload. The LTEST encapsulation time stamp SHALL be set as described in Section 8.2.1 and may be used by the receiver to compute an estimate of propagation delay. However, the propagation delay limit SHALL NOT be enforced.



LIVENESS TEST INTERVAL: Copy of the LIVENESS TEST INTERVAL specified in the CBIND request or reply message.

COUNT: Monotonically increasing value, initialized to 0 and incremented by one for each successive LTEST message.

SOURCE N_PORT NAME:	Contains a copy of the SOURCE N_PORT NAME specified in the CBIND request.
DESTINATION N_PORT NAME:	Contains a copy of the DESTINATION N_PORT NAME specified in the CBIND request.

7. Fibre Channel Link Services

Link services provide a set of fibre channel functions that allow a port to send control information or request another port to perform a specific control function.

There are three types of link services:

a) Basic

b) Extended

c) ULP-specific (FC-4)

Each link service message (request and reply) is carried by a fibre channel sequence and can be segmented into multiple frames.

The iFCP layer is responsible for transporting link service messages across the IP network. This includes mapping link service messages appropriately from the domain of the fibre channel transport to that of the IP network. This process may require special processing and the inclusion of supplemental data by the iFCP layer.

Each link service MUST be processed according to one of the following rules:

a) Pass-through - The link service message and reply MUST be delivered to the receiving N_PORT by the iFCP protocol layer without altering the message payload. The link service message and reply are not processed by the iFCP protocol layer.

b) Special - Applies to a link service reply or request requiring the intervention of the iFCP layer before forwarding to the destination N_PORT. Such messages may contain fibre channel addresses in the payload or may require other special processing.

c) Rejected - When issued by a locally attached N_PORT, the specified link service request MUST be rejected by the iFCP gateway. The gateway SHALL return an LS_RJT response with a Reason Code of 0x0B (Command Not Supported), and a Reason Code Explanation of 0x0 (No Additional Explanation).

This section describes the processing for special link services, including the manner in which supplemental data is added to the message payload.

Appendix A enumerates all link services and the iFCP processing policy that applies to each.

7.1. Special Link Service Messages

Special link service messages require the intervention of the iFCP layer before forwarding to the destination N_PORT. Such intervention is required in order to:

a) service any link service message that requires special handling, such as a PLOGI, and

b) service any link service message that has an N_PORT address in the payload in address translation mode only .

Unless the link service description specifies otherwise, support for each special link service is MANDATORY.

Such messages SHALL be transmitted in a fibre channel frame with the format shown in Figure 18 for extended link services or Figure 19 for FC-4 link services.

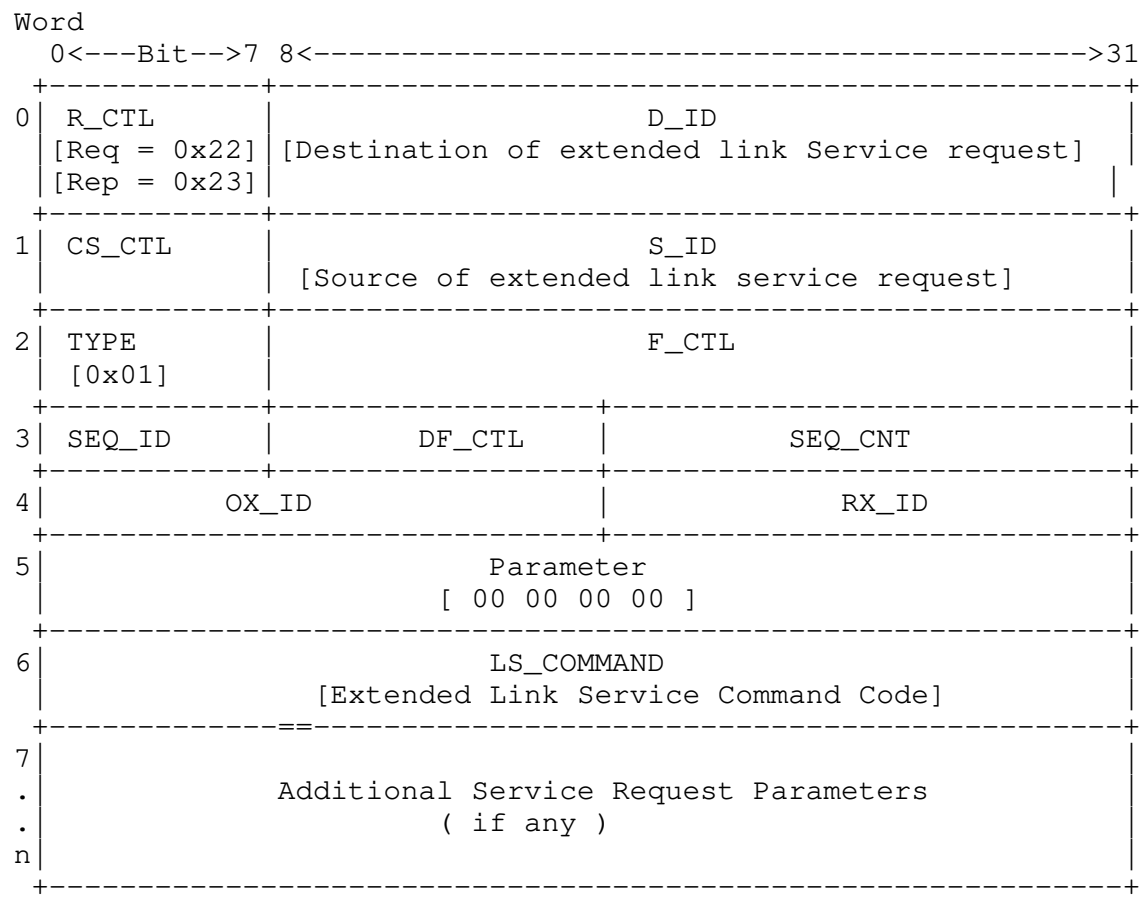


Figure 18. Format of an Extended Link Service Frame

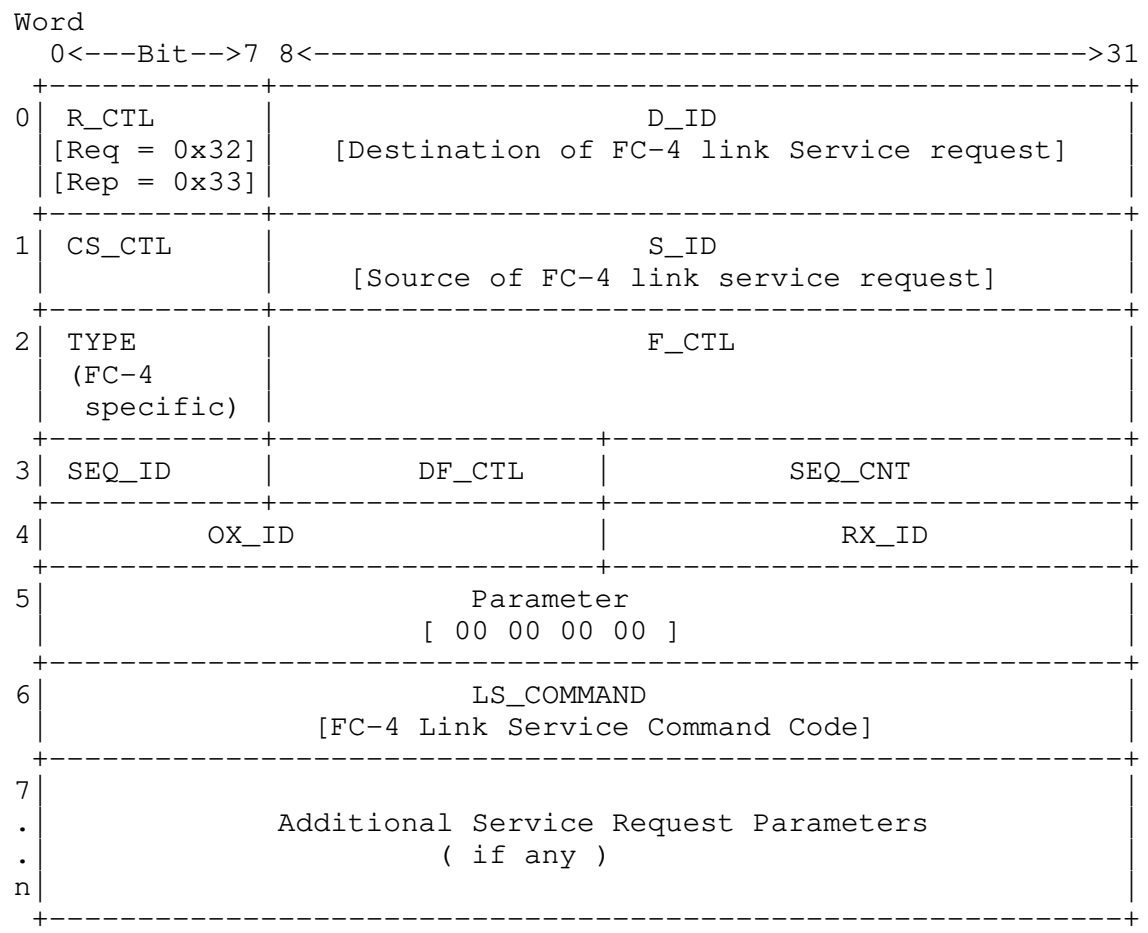


Figure 19. Format of an FC-4 Link Service Frame

7.2. Link Services Requiring Payload Address Translation

This section describes the handling for link service frames containing N_PORT addresses in the frame payload. Such addresses SHALL only be translated when the gateway is operating in address translation mode. When operating in address transparent mode, these addresses SHALL NOT be translated, and such link service messages SHALL NOT be sent as special frames unless other processing by the iFCP layer is required.

Supplemental data includes information required by the receiving gateway to convert an N_PORT address in the payload to an N_PORT address in the receiving gateway's address space. The following rules define the manner in which such supplemental data shall be packaged and referenced.

For an N_PORT address field, the gateway originating the frame MUST set the value in the payload to identify the address translation type as follows:

0x00 00 01 - The gateway receiving the frame from the IP network MUST replace the contents of the field with the N_PORT alias of the frame originator. This translation type MUST be used when the address to be converted is that of the source N_PORT.

0x00 00 02 - The gateway receiving the frame from the IP network MUST replace the contents of the field with the N_PORT ID of the destination N_PORT. This translation type MUST be used when the address to be converted is that of the destination N_PORT

0x00 00 03 - The gateway receiving the frame from the IP network MUST reference the specified supplemental data to set the field contents. The supplemental information is the 64-bit worldwide identifier of the N_PORT, as set forth in the fibre channel specification [FC-FS]. If not otherwise part of the link service payload, this information MUST be appended in accordance with the applicable link service description. Unless specified otherwise, this translation type SHALL NOT be used if the address to be converted corresponds to that of the frame originator or recipient.

Since fibre channel addressing rules prohibit the assignment of fabric addresses with a domain ID of 0, the above codes will never correspond to valid N_PORT fabric IDs.

If the sending gateway cannot obtain the worldwide identifier of an N_PORT, the gateway SHALL terminate the request with an LS_RJT message as described in [FC-FS]. The Reason Code SHALL be set to 0x07 (protocol error), and the Reason Explanation SHALL be set to 0x1F (Invalid N_PORT identifier).

Supplemental data is sent with the link service request or ACC frames in one of the following ways:

a) By appending the necessary data to the end of the link service frame.

b) By extending the sequence with additional frames.

In the first case, a new frame SHALL be created whose length includes the supplemental data. The procedure for extending the link service sequence with additional frames is dependent on the link service type.

For each field requiring address translation, the receiving gateway SHALL reference the translation type encoded in the field and replace it with the N_PORT address as shown in Table 7.

Translation Type Code	N_PORT Translation
0x00 00 01	Replace field contents with N_PORT alias of frame originator.
0x00 00 02	Replace field contents with N_PORT ID of frame recipient.
0x00 00 03	Lookup N_PORT via iSNS query. If locally attached, replace with N_PORT ID. If remotely attached, replace with N_PORT alias from remote N_PORT. descriptor (see Section 5.2.2.1).

Table 7. Link Service Address Translation

For translation type 3, the receiving gateway SHALL obtain the information needed to fill in the field in the link service frame payload by converting the specified N_PORT worldwide identifier to a gateway IP address and N_PORT ID. This information MUST be obtained through an iSNS name server query. If the query is unsuccessful, the gateway SHALL terminate the request with an LS_RJT response message as described in [FC-FS]. The Reason Code SHALL be set to 0x07 (protocol error), and the Reason Explanation SHALL be set to 0x1F (Invalid N_PORT identifier).

After applying the supplemental data, the receiving gateway SHALL forward the resulting link service frames to the destination N_PORT with the supplemental information removed.

7.3. Fibre Channel Link Services Processed by iFCP

The following Extended and FC-4 Link Service Messages must receive special processing.

Extended Link Service Messages	LS_COMMAND	Mnemonic
Abort Exchange	0x06 00 00 00	ABTX
Discover Address	0x52 00 00 00	ADISC
Discover Address Accept	0x02 00 00 00	ADISC ACC
FC Address Resolution Protocol Reply	0x55 00 00 00	FARP-REPLY
FC Address Resolution Protocol Request	0x54 00 00 00	FARP-REQ
Logout	0x05 00 00 00	LOGO
Port Login	0x30 00 00 00	PLOGI
Read Exchange Concise	0x13 00 00 00	REC
Read Exchange Concise Accept	0x02 00 00 00	REC ACC
Read Exchange Status Block	0x08 00 00 00	RES
Read Exchange Status Block Accept	0x02 00 00 00	RES ACC
Read Link Error Status Block	0x0F 00 00 00	RLS
Read Sequence Status Block	0x09 00 00 00	RSS
Reinstate Recovery Qualifier	0x12 00 00 00	RRQ
Request Sequence Initiative	0x0A 00 00 00	RSI
Scan Remote Loop	0x7B 00 00 00	SRL
Third Party Process Logout	0x24 00 00 00	TPRLO
Third Party Process Logout Accept	0x02 00 00 00	TPRLO ACC

FC-4 Link Service Messages	LS_COMMAND	Mnemonic
FCP Read Exchange Concise	0x13 00 00 00	FCP REC
FCP Read Exchange Concise Accept	0x02 00 00 00	FCP REC ACC

Each encapsulated fibre channel frame that is part of a special link service MUST have the SPC bit set to one in the iFCP FLAGS field of the encapsulation header, as specified in Section 5.3.1. If an ACC link service response requires special processing, the responding gateway SHALL place a copy of LS_COMMAND bits 0 through 7, from the

link service request frame, in the LS_COMMAND_ACC field of the ACC encapsulation header. Supplemental data (if any) MUST be appended as described in the following section.

The format of each special link service message, including supplemental data, where applicable, is shown in the following sections. Each description shows the basic format, as specified in the applicable FC standard, followed by supplemental data as shown in the example below.

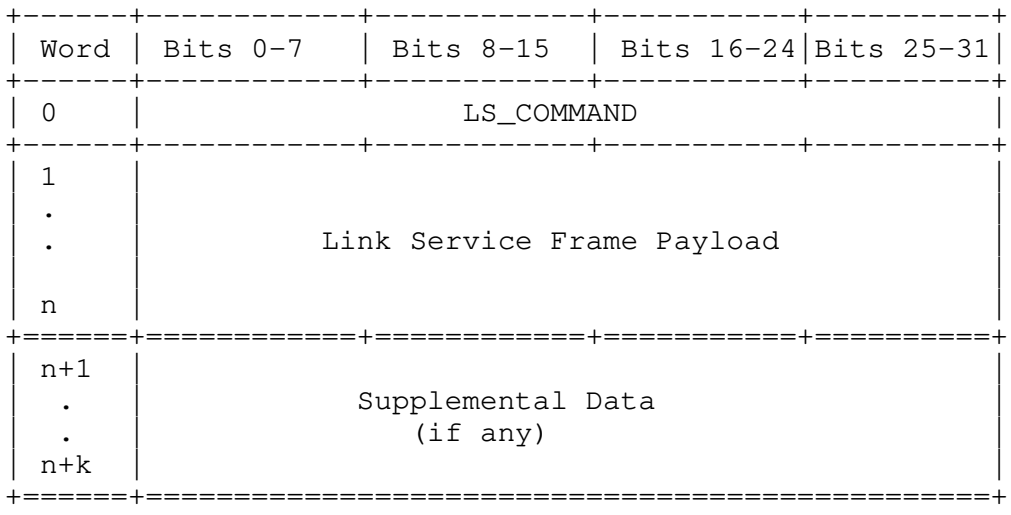


Figure 20. Special Link Service Frame Payload

7.3.1. Special Extended Link Services

The following sections define extended link services for which special processing is required.

7.3.1.1. Abort Exchange (ABTX)

ELS Format:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x6	0x00	0x00	0x00
1	RRQ Status	Exchange Originator S_ID		
2	OX_ID of Tgt exchange		RX_ID of tgt exchange	
3-10	Optional association header (32 bytes			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1, 2	N/A

Other Special Processing:

None.

7.3.1.2. Discover Address (ADISC)

Format of ADISC ELS:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x52	0x00	0x00	0x00
1	Reserved	Hard address of ELS Originator		
2-3	Port Name of Originator			
4-5	Node Name of originator			
6	Rsvd	N_PORT ID of ELS Originator		

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
N_PORT ID of ELS Originator	1	N/A

Other Special Processing:

The Hard Address of the ELS originator SHALL be set to 0.

7.3.1.1.3. Discover Address Accept (ADISC ACC)

Format of ADISC ACC ELS:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x20	0x00	0x00	0x00
1	Reserved	Hard address of ELS Originator		
2-3	Port Name of Originator			
4-5	Node Name of originator			
6	Rsvd	N_PORT ID of ELS Originator		

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
N_PORT ID of ELS Originator	1	N/A

Other Special Processing:

The Hard Address of the ELS originator SHALL be set to 0.

7.3.1.4. FC Address Resolution Protocol Reply (FARP-REPLY)

The FARP-REPLY ELS is used in conjunction with the FARP-REQ ELS (see Section 7.3.1.5) to perform the address resolution services required by the FC-VI protocol [FC-VI] and the fibre channel mapping of IP and ARP specified in RFC 2625 [RFC2625].

Format of FARP-REPLY ELS:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x55	0x00	0x00	0x00
1	Match Addr Code Points	Requesting N_PORT Identifier		
2	Responder Action	Responding N_PORT Identifier		
3-4	Requesting N_PORT Port_Name			
5-6	Requesting N_PORT Node_Name			
7-8	Responding N_PORT Port_Name			
9-10	Responding N_PORT Node_Name			
11-14	Requesting N_PORT IP Address			
15-18	Responding N_PORT IP Address			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Requesting N_PORT Identifier	2	N/A
Responding N_PORT Identifier	1	N/A

Other Special Processing:

None.

7.3.1.5. FC Address Resolution Protocol Request (FARP-REQ)

The FARP-REQ ELS is used in conjunction with the FC-VI protocol [FC-VI] and IP-to-FC mapping of RFC 2625 [RFC2625] to perform IP and FC address resolution in an FC fabric. The FARP-REQ ELS is usually directed to the fabric broadcast server at well-known address 0xFF-FF-FF for retransmission to all attached N_PORTs.

Section 9.4 describes the iFCP implementation of FC broadcast server functionality in an iFCP fabric.

Format of FARP_REQ ELS:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x54	0x00	0x00	0x00
1	Match Addr Code Points	Requesting N_PORT Identifier		
2	Responder Action	Responding N_PORT Identifier		
3-4	Requesting N_PORT Port_Name			
5-6	Requesting N_PORT Node_Name			
7-8	Responding N_PORT Port_Name			
9-10	Responding N_PORT Node_Name			
11-14	Requesting N_PORT IP Address			
15-18	Responding N_PORT IP Address			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Requesting N_PORT Identifier	3	Requesting N_PORT Port Name
Responding N_PORT Identifier	3	Responding N_PORT Port Name

Other Special Processing:

None.

7.3.1.6. Logout (LOGO) and LOGO ACC

ELS Format:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x5	0x00	0x00	0x00
1	Rsvd	N_PORT ID being logged out		
2-3	Port name of the LOGO originator (8 bytes)			

This ELS SHALL always be sent as a special ELS regardless of the translation mode in effect.

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
N_PORT ID Being Logged Out	1	N/A

Other Special Processing:

See Section 5.2.3.

7.3.1.7. Port Login (PLOGI) and PLOGI ACC

A PLOGI ELS establishes fibre channel communications between two N_PORTs and triggers the creation of an iFCP session if one does not exist.

The PLOGI request and ACC response carry information identifying the originating N_PORT, including a specification of its capabilities. If the destination N_PORT accepts the login request, it sends an Accept response (an ACC frame with PLOGI payload) specifying its capabilities. This exchange establishes the operating environment for the two N_PORTs.

The following figure is duplicated from [FC-FS], and shows the PLOGI message format for both the request and Accept (ACC) response. An N_PORT will reject a PLOGI request by transmitting an LS_RJT message containing no payload.

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x3 Acc = 0x2	0x00	0x00	0x00
1-4	Common Service Parameters			
5-6	N_PORT Name			
7-8	Node Name			
9-12	Class 1 Service Parameters			
13-17	Class 2 Service Parameters			
18-21	Class 3 Service Parameters			
22-25	Class 4 Service Parameters			
26-29	Vendor Version Level			

Figure 21. Format of PLOGI Request and ACC Payloads

Details of the above fields, including common and class-based service parameters, can be found in [FC-FS].

Special Processing

As specified in Section 5.2.2.2, a PLOGI request addressed to a remotely attached N_PORT MUST cause the creation of an iFCP session if one does not exist. Otherwise, the PLOGI and PLOGI ACC payloads MUST be passed through without modification to the destination N_PORT using the existing iFCP session. In either case, the SPC bit must be set in the frame encapsulation header as specified in 5.3.3.

If the CBIND to create the iFCP session fails, the issuing gateway SHALL terminate the PLOGI with an LS_RJT response. The Reason Code and Reason Code Explanation SHALL be selected from Table 8 based on the CBIND failure status.

CBIND Failure Status	LS_RJT Reason Code	LS_RJT Reason Code Explanation
Unspecified Reason (16)	Unable to Perform Command Request (0x09)	No Additional Explanation (0x00)
No Such Device (17)	Unable to Perform Command Request (0x09)	Invalid N_PORT Name (0x0D)
Lack of Resources (19)	Unable to Perform Command Request (0x09)	Insufficient Resources to Support Login (0x29)
Incompatible Address Translation Mode (20)	Unable to Perform Command Request (0x09)	No Additional Explanation (0x00)
Incorrect iFCP Protocol version Number (21)	Unable to Perform Command Request (0x09)	No Additional Explanation (0x00)
Gateway Not Synchronized (22)	Unable to Perform Command Request (0x09)	No Additional Explanation (0x00)

Table 8. PLOGI LS_RJT Status for CBIND Failures

7.3.1.8. Read Exchange Concise (REC)

Link Service Request Format:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x13	0x00	0x00	0x00
1	Rsvd	Exchange Originator S_ID		
2	OX_ID		RX_ID	
3-4	Port Name of the Exchange Originator (8 bytes) (present only for translation type 3)			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1, 2, or 3	Port Name of the Exchange Originator

Other Special Processing:

None.

7.3.1.9. Read Exchange Concise Accept (REC ACC)

Format of REC ACC Response:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Acc = 0x02	0x00	0x00	0x00
1	OX_ID		RX_ID	
2	Rsvd	Originator Address Identifier		
3	Rsvd	Responder Address Identifier		
4	FC4VALUE (FC-4-Dependent Value)			
5	E_STAT (Exchange Status)			
6-7	Port Name of the Exchange Originator (8 bytes)			
8-9	Port Name of the Exchange Responder (8 bytes)			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Originator Address Identifier	1, 2, or 3	Port Name of the Exchange Originator
Responder Address Identifier	1, 2, or 3	Port Name of the Exchange Responder

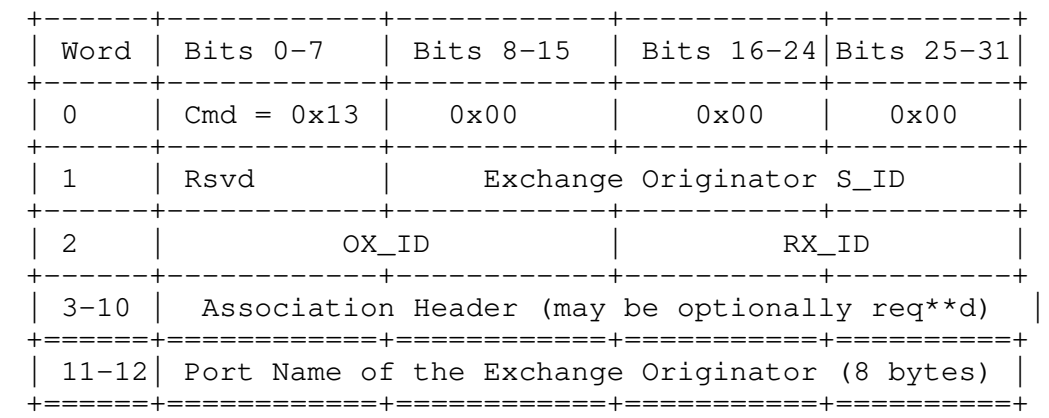
When supplemental data is required, the frame SHALL always be extended by 4 words as shown above. If the translation type for the Originator Address Identifier or the Responder Address Identifier is 1 or 2, the corresponding 8-byte port name SHALL be set to all zeros.

Other Special Processing:

None.

7.3.1.10. Read Exchange Status Block (RES)

ELS Format:



Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x13	0x00	0x00	0x00
1	Rsvd	Exchange Originator S_ID		
2	OX_ID		RX_ID	
3-10	Association Header (may be optionally req**d)			
11-12	Port Name of the Exchange Originator (8 bytes)			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1, 2, or 3	Port Name of the Exchange Originator

Other Special Processing:

None.

7.3.1.11. Read Exchange Status Block Accept (RES ACC)

Format of ELS Accept Response:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Acc = 0x02	0x00	0x00	0x00
1	OX_ID		RX_ID	
2	Rsvd	Exchange Originator N_PORT ID		
3	Rsvd	Exchange Responder N_PORT ID		
4	Exchange Status Bits			
5	Reserved			
6-n	Service Parameters and Sequence Statuses as described in [FC-FS]			
n+1-n+2	Port Name of the Exchange Originator (8 bytes)			
n+3-n+4	Port Name of the Exchange Responder (8 bytes)			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Exchange Originator N_PORT ID	1, 2, or 3	Port Name of the Exchange Originator
Exchange Responder N_PORT ID	1, 2, or 3	Port Name of the Exchange Responder

When supplemental data is required, the ELS SHALL be extended by 4 words as shown above. If the translation type for the Exchange Originator N_PORT ID or the Exchange Responder N_PORT ID is 1 or 2, the corresponding 8-byte port name SHALL be set to all zeros.

Other Special Processing:

None.

7.3.1.12. Read Link Error Status (RLS)

ELS Format:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x0F	0x00	0x00	0x00
1	Rsvd	N_PORT Identifier		
2-3	Port Name of the N_PORT (8 bytes)			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
N_PORT Identifier	1, 2, or 3	Port Name of the N_PORT

Other Special Processing:

None.

7.3.1.13. Read Sequence Status Block (RSS)

ELS Format:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x09	0x00	0x00	0x00
1	SEQ_ID	Exchange Originator S_ID		
2	OX_ID		RX_ID	
3-4	Port Name of the Exchange Originator (8 bytes)			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1, 2, or 3	Port Name of the Exchange Originator

Other Special Processing:

None.

7.3.1.14. Reinstate Recovery Qualifier (RRQ)

ELS Format:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x12	0x00	0x00	0x00
1	Rsvd	Exchange Originator S_ID		
2	OX_ID		RX_ID	
3-10	Association Header (may be optionally req**d)			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1 or 2	N/A

Other Special Processing:

None.

7.3.1.15. Request Sequence Initiative (RSI)

ELS Format:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x0A	0x00	0x00	0x00
1	Rsvd	Exchange Originator S_ID		
2	OX_ID		RX_ID	
3-10	Association Header (may be optionally req**d)			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1 or 2	N/A

Other Special Processing:

None.

7.3.1.16. Scan Remote Loop (SRL)

SRL allows a remote loop to be scanned to detect changes in the device configuration. Any changes will trigger a fibre channel state change notification and subsequent update of the iSNS database.

ELS Format:

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x7B	Reserved		
1	Flag	Address Identifier of the FL_PORT (see B.1)		
2-3	Worldwide Name of the Remote FL_PORT			

Fields Requiring Address Translation	Translation Type (see Section 7.2)	Supplemental Data (type 3 only)
Address Identifier of the FL_PORT	3	Worldwide Name of the Remote FL_PORT

Other Special Processing:

The D_ID field is the address of the Domain Controller associated with the remote loop. The format of the Domain Controller address is the hex 'FF FC' || Domain_ID, where Domain_ID is the gateway-assigned alias representing the remote gateway or switch element being queried. After translation by the remote gateway, the D_ID identifies the gateway or switch element to be scanned within the remote gateway region.

The FLAG field defines the scope of the SRL. If set to 0, all loop port interfaces on the given switch element or gateway are scanned. If set to one, the loop port interface on the gateway or switch element to be scanned MUST be specified in bits 8 through 31.

If the Flag field is zero, the SRL request SHALL NOT be sent as a special ELS.

If the Domain_ID represents a remote switch or gateway and an iFCP session to the remote Domain Controller does not exist, the requesting gateway SHALL create the iFCP session.

7.3.1.17. Third Party Process Logout (TPRLO)

TPRLO provides a mechanism for an N_PORT (third party) to remove one or more process login sessions that exist between the destination N_PORT and other N_PORTs specified in the command. This command includes one or more TPRLO LOGOUT PARAMETER PAGEs, each of which, when combined with the destination N_PORT, identifies a process login to be terminated by the command.

Word	Bits 0-7	Bits 8-15	Bits 16 - 31
0	Cmd = 0x24	Page Length (0x10)	Payload Length
1	TPRLO Logout Parameter Page 0		
5	TPRLO Logout Parameter Page 1		
	....		
(4*n)+1	TPRLO Logout Parameter Page n		

Figure 22. Format of TPRLO ELS

Each TPRLO parameter page contains parameters identifying one or more image pairs and may be associated with a single FC-4 protocol type that is common to all FC-4 protocol types between the specified image pair or global to all specified image pairs. The format of a TPRLO page requiring address translation is shown in Figure 23. Additional information on TPRLO can be found in [FC-FS].

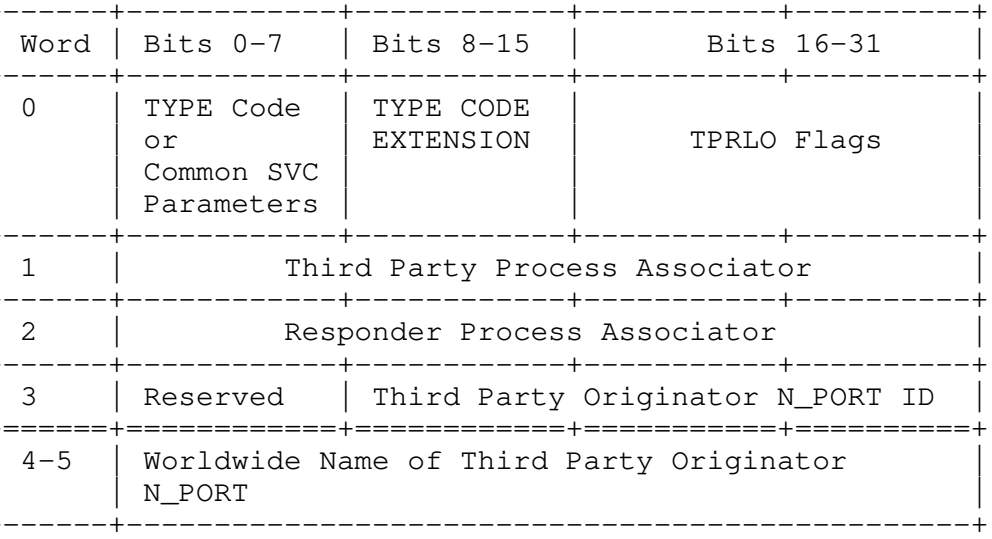


Figure 23. Format of an Augmented TPRLO Parameter Page

The TPRLO flags that affect supplemented ELS processing are as follows:

Bit 18: Third party Originator N_PORT Validity. When set to one, this bit indicates that word 3, bits 8-31 (Third Party Originator N_PORT ID), are meaningful.

Bit 19: Global Process logout. When set to one, this bit indicates that all image pairs for all N_PORTs of the specified FC-4 protocol shall be invalidated. When the value of this bit is one, only one logout parameter page is permitted in the TPRLO payload.

If bit 18 has a value of zero and bit 19 has a value of one in the TPRLO flags field, then the ELS SHALL NOT be sent as a special ELS.

Otherwise, the originating gateway SHALL process the ELS as follows:

a) The first word of the TPRLO payload SHALL NOT be modified.

b) Each TPRLO parameter page shall be extended by two words as shown in Figure 23.

c) If word 0, bit 18 (Third Party Originator N_PORT ID validity), in the TPRLO flags field has a value of one, then the sender shall place the worldwide port name of the fibre channel device's N_PORT in the extension words. The N_PORT ID SHALL be set to 3. Otherwise, the contents of the extension words and the Third Party Originator N_PORT ID SHALL be set to zero.

d) The ELS originator SHALL set the SPC bit in the encapsulation header of each augmented frame comprising the ELS (see Section 5.3.1).

e) If the ELS contains a single TPRLO parameter page, the originator SHALL increase the frame length as necessary to include the extended parameter page.

f) If the ELS to be augmented contains multiple TPRLO parameter pages, the FC frames created to contain the augmented ELS payload SHALL NOT exceed the maximum frame size that can be accepted by the destination N_PORT.

   Each fibre channel frame SHALL contain an integer number of extended TPRLO parameter pages. The maximum number of extended TPRLO parameter pages in a frame SHALL be limited to the number that can be held without exceeding the above upper limit. New frames resulting from the extension of the TPRLO pages to include the supplemental data SHALL be created by extending the SEQ_CNT in the fibre channel frame header. The SEQ_ID SHALL NOT be modified.

The gateway receiving the augmented TPRLO ELS SHALL generate ELS frames to be sent to the destination N_PORT by copying word 0 of the ELS payload and processing each augmented parameter page as follows:

a) If word 0, bit 18, has a value of one, create a parameter page by copying words 0 through 2 of the augmented parameter page. The Third Party Originator N_PORT ID in word 3 shall be generated by referencing the supplemental data as described in Section 7.2.

b) If word 0, bit 18, has a value of zero, create a parameter page by copying words 0 through 3 of the augmented parameter page.

The size of each frame to be sent to the destination N_PORT MUST NOT exceed the maximum frame size that the destination N_PORT can accept. The sequence identifier in each frame header SHALL be copied from the augmented ELS, and the sequence count SHALL be monotonically increasing.

7.3.1.18. Third Party Logout Accept (TPRLO ACC)

The format of the TPRLO ACC frame is shown in Figure 24.

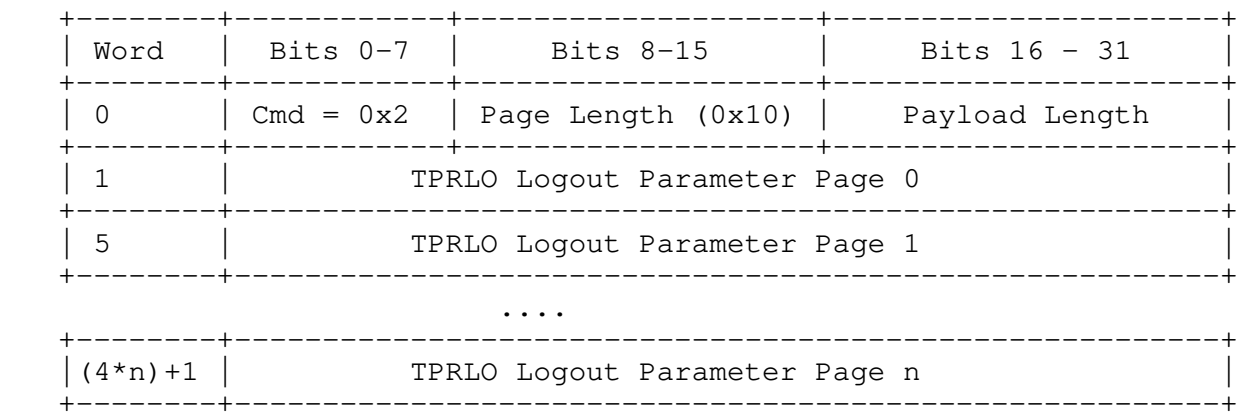


Figure 24. Format of TPRLO ACC ELS

The format of the parameter page and rules for parameter page augmentation are as specified in Section 7.3.1.17.

7.3.2. Special FC-4 Link Services

The following sections define FC-4 link services for which special processing is required.

7.3.2.1. FC-4 Link Services Defined by FCP

The format of FC-4 link service frames defined by FCP can be found in [FCP-2].

7.3.2.1.1. FCP Read Exchange Concise (FCP REC)

The payload format for this link service is identical to the REC extended link service specified in Section 7.3.1.8 and SHALL be processed as described in that section. The FC-4 version will become obsolete in [FCP-2]. However, in order to support devices implemented against early revisions of FCP-2, an iFCP gateway MUST support both versions.

7.3.2.1.2. FCP Read Exchange Concise Accept (FCP REC ACC)

The payload format for this link service is identical to the REC ACC extended link service specified in Section 7.3.1.9 and SHALL be processed as described in that section. The FC-4 version will become obsolete in [FCP-2]. However, in order to support devices implemented against earlier revisions of FCP-2, an iFCP gateway MUST support both versions.

7.4. FLOGI Service Parameters Supported by an iFCP Gateway

The FLOGI ELS is issued by an N_PORT that wishes to access the fabric transport services.

The format of the FLOGI request and FLOGI ACC payloads are identical to the PLOGI request and ACC payloads described in Section 7.3.1.7.

Word	Bits 0-7	Bits 8-15	Bits 16-24	Bits 25-31
0	Cmd = 0x4 Acc = 0x2	0x00	0x00	0x00
1-4	Common Service Parameters			
5-6	N_PORT Name			
7-8	Node Name			
9-12	Class 1 Service Parameters			
13-17	Class 2 Service Parameters			
18-21	Class 3 Service Parameters			
22-25	Class 4 Service Parameters			
26-29	Vendor Version Level			

Figure 25. FLOGI Request and ACC Payload Format

A full description of each parameter is given in [FC-FS].

This section tabulates the protocol-dependent service parameters supported by a fabric port attached to an iFCP gateway.

The service parameters carried in the payload of an FLOGI extended link service request MUST be set in accordance with Table 9.

Service Parameter	Fabric Login Class			
	1	2	3	4
Class Validity	n	M	M	n
Service Options				
Intermix Mode	n	n	n	n
Stacked Connect-Requests	n	n	n	n
Sequential Delivery	n	M	M	n
Dedicated Simplex	n	n	n	n
Camp On	n	n	n	n
Buffered Class 1	n	n	n	n
Priority	n	n	n	n
Initiator/Recipient Control				
Clock Synchronization ELS Capable	n	n	n	n

Table 9. FLOGI Service Parameter Settings

Notes:

1) "n" indicates a parameter or capability that is not supported by the iFCP protocol.

2) "M" indicates an applicable parameter that MUST be supported by an iFCP gateway.

8. iFCP Error Detection

8.1. Overview

This section specifies provisions for error detection and recovery in addition to those in [FC-FS], which continue to be available in the iFCP network environment.

8.2. Stale Frame Prevention

Recovery from fibre channel protocol error conditions requires that frames associated with a failed or aborted exchange drain from the fabric before exchange resources can be safely reused.

Since a fibre channel fabric may not preserve frame order, there is no deterministic way to purge such frames. Instead, the fabric guarantees that frame the lifetime will not exceed a specific limit (R_A_TOV).

R_A_TOV is defined in [FC-FS] as "the maximum transit time within a fabric to guarantee that a lost frame will never emerge from the fabric". For example, a value of 2 x R_A_TOV is the minimum time that the originator of an ELS request or FC-4 link service request must wait for the response to that request. The fibre channel default value for R_A_TOV is 10 seconds.

An iFCP gateway SHALL actively enforce limits on R_A_TOV as described in Section 8.2.1.

8.2.1. Enforcing R_A_TOV Limits

The R_A_TOV limit on frame lifetimes SHALL be enforced by means of the time stamp in the encapsulation header (see Section 5.3.1) as described in this section.

The budget for R_A_TOV SHOULD include allowances for the propagation delay through the gateway regions of the sending and receiving N_PORTs, plus the propagation delay through the IP network. This latter component is referred to in this specification as IP_TOV.

IP_TOV should be set well below the value of R_A_TOV specified for the iFCP fabric and should be stored in the iSNS server. IP_TOV should be set to 50 percent of R_A_TOV.

The following paragraphs describe the requirements for synchronizing gateway time bases and the rules for measuring and enforcing propagation delay limits.

The protocol for synchronizing a gateway time base is SNTP [RFC2030]. In order to ensure that all gateways are time aligned, a gateway SHOULD obtain the address of an SNTP-compatible time server via an iSNS query. If multiple time server addresses are returned by the query, the servers must be synchronized and the gateway may use any server in the list. Alternatively, the server may return a multicast group address in support of operation in Anycast mode. Implementation of Anycast mode is as specified in [RFC2030], including the precautions defined in that document. Multicast mode SHOULD NOT be used.

An SNTP server may use any one of the time reference sources listed in [RFC2030]. The resolution of the time reference MUST be 125 milliseconds or better.

Stability of the SNTP server and gateway time bases should be 100 ppm or better.

With regard to its time base, the gateway is in either the Synchronized or Unsynchronized state.

When in the synchronized state, the gateway SHALL

a) set the time stamp field for each outgoing frame in accordance with the gateway's internal time base;

b) check the time stamp field of each incoming frame, following validation of the encapsulation header CRC, as described in Section 5.3.4;

c) if the incoming frame has a time stamp of 0,0 and is not one of the session control frames that require a 0,0 time stamp (see Section 6), the frame SHALL be discarded;

d) if the incoming frame has a non-zero time stamp, the receiving gateway SHALL compute the absolute value of the time in flight and SHALL compare it against the value of IP_TOV specified for the IP fabric;

e) if the result in step (d) exceeds IP_TOV, the encapsulated frame shall be discarded. Otherwise, the frame shall be de-encapsulated as described in Section 5.3.4.

A gateway SHALL enter the Synchronized state upon receiving a successful response to an SNTP query.

A gateway shall enter the Unsynchronized state:

a) upon power-up and before successful completion of an SNTP query, and

b) whenever the gateway looses contact with the SNTP server, such that the gateway's time base may no longer be in alignment with that of the SNTP server. The criterion for determining loss of contact is implementation specific.

Following loss of contact, it is recommended that the gateway enter the Unsynchronized state when the estimated time base drift relative to the SNTP reference is greater than ten percent of the IP_TOV limit. (Assuming that all timers have an accuracy of 100 ppm and IP_TOV equals 5 seconds, the maximum allowable loss of contact duration would be about 42 minutes.)

As the result of a transition from the Synchronized to the Unsynchronized state, a gateway MUST abort all iFCP sessions as described in Section 5.2.3. While in the Unsynchronized state, a gateway SHALL NOT permit the creation of new iFCP sessions.

9. Fabric Services Supported by an iFCP Implementation

An iFCP gateway implementation MUST support the following fabric services:

N_PORT ID Value	Description	Section
0xFF-FF-FE	F_PORT Server	9.1
0xFF-FF-FD	Fabric Controller	9.2
0xFF-FF-FC	Directory/Name Server	9.3

In addition, an iFCP gateway MAY support the FC broadcast server functionality described in Section 9.4.

9.1. F_PORT Server

The F_PORT server SHALL support the FLOGI ELS, as described in Section 7.4, as well as the following ELSs specified in [FC-FS]:

a) Request for fabric service parameters (FDISC).

b) Request for the link error status (RLS).

c) Read Fabric Timeout Values (RTV).

9.2. Fabric Controller

The Fabric Controller SHALL support the following ELSs as specified in [FC-FS]:

a) State Change Notification (SCN).

b) Registered State Change Notification (RSCN).

c) State Change Registration (SCR).

9.3. Directory/Name Server

The Directory/Name server provides a registration service allowing an N_PORT to record or query the database for information about other N_PORTs. The services are defined in [FC-GS3]. The queries are issued as FC-4 transactions using the FC-CT command transport protocol specified in [FC-GS3].

In iFCP, each name server request MUST be translated to the appropriate iSNS query defined in [ISNS]. The definitions of name server objects are specified in [FC-GS3].

The name server SHALL support record and query operations for directory subtype 0x02 (Name Server) and 0x03 (IP Address Server) and MAY support the FC-4 specific services as defined in [FC-GS3].

9.4. Broadcast Server

Fibre channel frames are broadcast throughout the fabric by addressing them to the fibre channel broadcast server at the well-known fibre channel address 0xFF-FF-FF. The broadcast server then replicates and delivers the frame to each attached N_PORT in all zones to which the originating device belongs. Only class 3 (datagram) service is supported.

In an iFCP system, the fibre channel broadcast function is emulated by means of a two-tier architecture comprising the following elements:

a) A local broadcast server residing in each iFCP gateway. The local server distributes broadcast traffic within the gateway region and forwards outgoing broadcast traffic to a global server for distribution throughout the iFCP fabric.

b) A global broadcast server that re-distributes broadcast traffic to the local server in each participating gateway.

c) An iSNS discovery domain defining the scope over which broadcast traffic is propagated. The discovery domain is populated with a global broadcast server and the set of local servers it supports.

The local and global broadcast servers are logical iFCP devices that communicate using the iFCP protocol. The servers have an N_PORT Network Address consisting of an iFCP portal address and an N_PORT ID set to the well-known fibre channel address of the FC broadcast server (0xFF-FF-FF).

As noted above, an N_PORT originates a broadcast by directing frame traffic to the fibre channel broadcast server. The gateway-resident local server distributes a copy of the frame locally and forwards a copy to the global server for redistribution to the local servers on other gateways. The global server MUST NOT echo a broadcast frame to the originating local server.

9.4.1. Establishing the Broadcast Configuration

The broadcast configuration is managed with facilities provided by the iSNS server by the following means:

a) An iSNS discovery domain is created and seeded with the network address of the global broadcast server N_PORT. The global server is identified as such by setting the appropriate N_PORT entity attribute.

b) Using the management interface, each broadcast server is preset with the identity of the broadcast domain.

During power up, each gateway SHALL invoke the iSNS service to register its local broadcast server in the broadcast discovery domain. After registration, the local server SHALL wait for the global broadcast server to establish an iFCP session.

The global server SHALL register with the iSNS server as follows:

a) The server SHALL query the iSNS name server by attribute to obtain the worldwide port name of the N_PORT pre-configured to provide global broadcast services.

b) If the worldwide port name obtained above does not correspond to that of the server issuing the query, the N_PORT SHALL NOT perform global broadcast functions for N_PORTs in that discovery domain.

c) Otherwise, the global server N_PORT SHALL register with the discovery domain and query the iSNS server to identify all currently registered local servers.

d) The global broadcast server SHALL initiate an iFCP session with each local broadcast server in the domain. When a new local server registers, the global server SHALL receive a state change notification and respond by initiating an iFCP session with the newly added server. The gateway SHALL obtain these notifications using the iSNS provisions for lossless delivery.

Upon receiving the CBIND request to initiate the iFCP session, the local server SHALL record the worldwide port name and N_PORT network address of the global server.

9.4.2. Broadcast Session Management

After the initial broadcast session is established, the local or global broadcast server MAY choose to manage the session in one of the following ways, depending on resource requirements and the anticipated level of broadcast traffic:

a) A server MAY keep the session open continuously. Since broadcast sessions are often quiescent for long periods of time, the server SHOULD monitor session connectivity as described in Section 5.2.2.4.

b) A server MAY open the broadcast session on demand only when broadcast traffic is to be sent. If the session is reopened by the global server, the local server SHALL replace the previously recorded network address of the global broadcast server.

9.4.3. Standby Global Broadcast Server

An implementation may designate a local server to assume the duties of the global broadcast server in the event of a failure. The local server may use the LTEST message to determine whether the global server is functioning and may assume control if it is not.

When assuming control, the standby server must register with the iSNS server as the global broadcast server in place of the failed server and must install itself in the broadcast discovery domain as specified in steps c) and d) of Section 9.4.1.

10. iFCP Security

10.1. Overview

iFCP relies upon the IPSec protocol suite to provide data confidentiality and authentication services, and it relies upon IKE as the key management protocol. Section 10.2 describes the security requirements arising from iFCP's operating environment, and Section

10.3 describes the resulting design choices, their requirement levels, and how they apply to the iFCP protocol.

Detailed considerations for use of IPsec and IKE with the iFCP protocol can be found in [SECIPS].

10.2. iFCP Security Threats and Scope

10.2.1. Context

iFCP is a protocol designed for use by gateway devices deployed in enterprise data centers. Such environments typically have security gateways designed to provide network security through isolation from public networks. Furthermore, iFCP data may have to traverse security gateways in order to support SAN-to-SAN connectivity across public networks.

10.2.2. Security Threats

Communicating iFCP gateways may be subjected to attacks, including attempts by an adversary to:

a) acquire confidential data and identities by snooping data packets,

b) modify packets containing iFCP data and control messages,

c) inject new packets into the iFCP session,

d) hijack the TCP connection carrying the iFCP session,

e) launch denial-of-service attacks against the iFCP gateway,

f) disrupt the security negotiation process,

g) impersonate a legitimate security gateway, or

h) compromise communication with the iSNS server.

It is imperative to thwart these attacks, given that an iFCP gateway is the last line of defense for a whole fibre channel island, which may include several hosts and fibre channel switches. To do so, the iFCP gateway must implement and may use confidentiality, data origin authentication, integrity, and replay protection on a per-datagram basis. The iFCP gateway must implement and may use bi-directional authentication of the communication endpoints. Finally, it must implement and may use a scalable approach to key management.

10.2.3. Interoperability with Security Gateways

Enterprise data center networks are considered mission-critical facilities that must be isolated and protected from all possible security threats. Such networks are usually protected by security gateways, which, at a minimum, provide a shield against denial-of-service attacks. The iFCP security architecture is capable of leveraging the protective services of the existing security infrastructure, including firewall protection, NAT and NAPT services, and IPSec VPN services available on existing security gateways. Considerations regarding intervening NAT and NAPT boxes along the iFCP-iSNS path can be found in [ISNS].

10.2.4. Authentication

iFCP is a peer-to-peer protocol. iFCP sessions may be initiated by either peer gateway or both. Consequently, bi-directional authentication of peer gateways must be provided in accordance with the requirement levels specified in Section 10.3.1.

N_PORT identities used in the Port Login (PLOGI) process shall be considered authenticated if the PLOGI request is received from the remote gateway over a secure, IPSec-protected connection.

There is no requirement that the identities used in authentication be kept confidential.

10.2.5. Confidentiality

iFCP traffic may traverse insecure public networks, and therefore implementations must have per-packet encryption capabilities to provide confidentiality in accordance with the requirements specified in Section 10.3.1.

10.2.6. Rekeying

Due to the high data transfer rates and the amount of data involved, an iFCP implementation must support the capability to rekey each phase 2 security association in the time intervals dictated by sequence number space exhaustion at a given link rate. In the rekeying scenario described in [SECIPS], for example, rekeying events happen as often as every 27.5 seconds at a 10 Gbps rate.

The iFCP gateway must provide the capability for forward secrecy in the rekeying process.

10.2.7. Authorization

Basic access control properties stem from the requirement that two communicating iFCP gateways be known to one or more iSNS servers before they can engage in iFCP exchanges. The optional use of discovery domains [ISNS], Identity Payloads (e.g., ID_FQDNs), and certificate-based authentication (e.g., with X509v3 certificates) enables authorization schemas of increasing complexity. The definition of such schemas (e.g., role-based access control) is outside of the scope of this specification.

10.2.8. Policy Control

This specification allows any and all security mechanisms in an iFCP gateway to be administratively disabled. Security policies MUST have, at most, iFCP Portal resolution. Administrators may gain control over security policies through an adequately secured interaction with a management interface or with iSNS.

10.2.9. iSNS Role

iSNS [ISNS] is an invariant in all iFCP deployments. iFCP gateways MUST use iSNS for discovery services and MAY use security policies configured in the iSNS database as the basis for algorithm negotiation in IKE. The iSNS specification defines mechanisms for securing communication between an iFCP gateway and iSNS server(s). Additionally, the specification indicates how elements of security policy concerning individual iFCP sessions can be retrieved from iSNS server(s).

10.3. iFCP Security Design

10.3.1. Enabling Technologies

Applicable technology from IPsec and IKE is defined in the following suite of specifications:

[RFC2401] Security Architecture for the Internet Protocol

[RFC2402] IP Authentication Header

[RFC2404] The Use of HMAC-SHA-1-96 within ESP and AH

[RFC2405] The ESP DES-CBC Cipher Algorithm with Explicit IV

[RFC2406] IP Encapsulating Security Payload

[RFC2407] The Internet IP Security Domain of Interpretation for ISAKMP

[RFC2408] Internet Security Association and Key Management Protocol (ISAKMP)

[RFC2409] The Internet Key Exchange (IKE)

[RFC2410] The NULL Encryption Algorithm and Its Use With IPSEC

[RFC2451] The ESP CBC-Mode Cipher Algorithms

[RFC2709] Security Model with Tunnel-mode IPsec for NAT Domains

The implementation of IPsec and IKE is required according to the following guidelines.

Support for the IP Encapsulating Security Payload (ESP) [RFC2406] is MANDATORY to implement. When ESP is used, per-packet data origin authentication, integrity, and replay protection MUST be used.

For data origin authentication and integrity with ESP, HMAC with SHA1 [RFC2404] MUST be implemented, and the Advanced Encryption Standard [AES] in CBC MAC mode with Extended Cipher Block Chaining SHOULD be implemented in accordance with [AESCBC].

For confidentiality with ESP, 3DES in CBC mode [RFC2451] MUST be implemented, and AES counter mode encryption [AESCTR] SHOULD be implemented. NULL encryption MUST be supported as well, as defined in [RFC2410]. DES in CBC mode SHOULD NOT be used due to its inherent weakness. Since it is known to be crackable with modest computation resources, it is inappropriate for use in any iFCP deployment scenario.

A conforming iFCP protocol implementation MUST implement IPsec ESP [RFC2406] in tunnel mode [RFC2401] and MAY implement IPsec ESP in transport mode.

Regarding key management, iFCP implementations MUST support IKE [RFC2409] for bi-directional peer authentication, negotiation of security associations, and key management, using the IPsec DOI. There is no requirement that the identities used in authentication be kept confidential. Manual keying MUST NOT be used since it does not provide the necessary keying support. According to [RFC2409], pre-shared secret key authentication is MANDATORY to implement, whereas certificate-based peer authentication using digital signatures MAY be implemented (see Section 10.3.3 regarding the use of certificates). [RFC2409] defines the following requirement levels for IKE Modes:

Phase-1 Main Mode MUST be implemented.

Phase-1 Aggressive Mode SHOULD be implemented.

Phase-2 Quick Mode MUST be implemented.

Phase-2 Quick Mode with key exchange payload MUST be implemented.

With iFCP, Phase-1 Main Mode SHOULD NOT be used in conjunction with pre-shared keys, due to Main Mode's vulnerability to man-in-the-middle-attackers when group pre-shared keys are used. In this scenario, Aggressive Mode SHOULD be used instead. Peer authentication using the public key encryption methods outlined in [RFC2409] SHOULD NOT be used.

The DOI [RFC2407] provides for several types of Identification Payloads.

When used for iFCP, IKE Phase 1 exchanges MUST explicitly carry the Identification Payload fields (IDii and IDir). Conforming iFCP implementations MUST use ID_IPV4_ADDR, ID_IPV6_ADDR (if the protocol stack supports IPv6), or ID_FQDN Identification Type values. The ID_USER_FQDN, IP Subnet, IP Address Range, ID_DER_ASN1_DN, ID_DER_ASN1_GN Identification Type values SHOULD NOT be used. The ID_KEY_ID Identification Type values MUST NOT be used. As described in [RFC2407], the port and protocol fields in the Identification Payload MUST be set to zero or UDP port 500.

When used for iFCP, IKE Phase 2 exchanges MUST explicitly carry the Identification Payload fields (IDci and IDcr). Conforming iFCP implementations MUST use either ID_IPV4_ADDR or ID_IPV6_ADDR Identification Type values (according to the version of IP supported). Other Identification Type values MUST NOT be used. As described in Section 5.2.2, the gateway creating the iFCP session must query the iSNS server to determine the appropriate port on which to initiate the associated TCP connection. Upon a successful IKE Phase 2 exchange, the IKE responder enforces the negotiated selectors on the IPsec SAs. Any subsequent iFCP session creation requires the iFCP peer to query its iSNS server for access control (in accordance with the session creation requirements specified in Section 5.2.2.1).

10.3.2. Use of IKE and IPsec

A conforming iFCP Portal is capable of establishing one or more IKE Phase-1 Security Associations (SAs) to a peer iFCP Portal. A Phase-1 SA may be established when an iFCP Portal is initialized or may be deferred until the first TCP connection with security requirements is established.

An IKE Phase-2 SA protects one or more TCP connections within the same iFCP Portal. More specifically, the successful establishment of an IKE Phase-2 SA results in the creation of two uni-directional IPsec SAs fully qualified by the tuple <SPI, destination IP address, ESP>.

These SAs protect the setup process of the underlying TCP connections and all their subsequent TCP traffic. The number of TCP connections in an IPsec SA, as well as the number of SAs, is practically driven by security policy considerations (i.e., security services are defined at the granularity of an IPsec SA only), QoS considerations (e.g., multiple QoS classes within the same IPsec SA increase odds of packet reordering, possibly falling outside the replay window), and failure compartmentalization considerations. Each of the TCP connections protected by an IPsec SA is either in the unbound state, or bound to a specific iFCP session.

In summary, at any point in time:

-- there exist 0..M IKE Phase-1 SAs between peer iFCP portals,

-- each IKE Phase-1 SA has 0..N IKE Phase-2 SAs, and

-- each IKE Phase-2 SA protects 0..Z TCP connections.

The creation of an IKE Phase-2 SA may be triggered by a policy rule supplied through a management interface or by iFCP Portal properties registered with the iSNS server. Similarly, the use of a Key Exchange payload in Quick Mode for perfect forward secrecy may be dictated through a management interface or by an iFCP Portal policy rule registered with the iSNS server.

If an iFCP implementation makes use of unbound TCP connections, and such connections belong to an iFCP Portal with security requirements, then the unbound connections MUST be protected by an SA at all times just like bound connections.

Upon receipt of an IKE Phase-2 delete message, there is no requirement to terminate the protected TCP connections or delete the associated IKE Phase-1 SA. Since an IKE Phase-2 SA may be associated with multiple TCP connections, terminating these connections might in fact be inappropriate and untimely.

To minimize the number of active Phase-2 SAs, IKE Phase-2 delete messages may be sent for Phase-2 SAs whose TCP connections have not handled data traffic for a while. To minimize the use of SA

resources while the associated TCP connections are idle, creation of a new SA should be deferred until new data are to be sent over the connections.

10.3.3. Signatures and Certificate-Based Authentication

Conforming iFCP implementations MAY support peer authentication via digital signatures and certificates. When certificate authentication is chosen within IKE, each iFCP gateway needs the certificate credentials of each peer iFCP gateway in order to establish a security association with that peer.

Certificate credentials used by iFCP gateways MUST be those of the machine. Certificate credentials MAY be bound to the interface (IP Address or FQDN) of the iFCP gateway used for the iFCP session, or to the fabric WWN of the iFCP gateway itself. Since the value of a machine certificate is inversely proportional to the ease with which an attacker can obtain one under false pretenses, it is advisable that the machine certificate enrollment process be strictly controlled. For example, only administrators may have the ability to enroll a machine with a machine certificate. User certificates SHOULD NOT be used by iFCP gateways for establishment of SAs protecting iFCP sessions.

If the gateway does not have the peer iFCP gateway's certificate credentials, then it can obtain them:

a) by using the iSNS protocol to query for the peer gateway's certificate(s) stored in a trusted iSNS server, or

b) through use of the ISAKMP Certificate Request Payload (CRP) [RFC2408] to request the certificate(s) directly from the peer iFCP gateway.

When certificate chains are long enough, IKE exchanges using UDP as the underlying transport may yield IP fragments, which are known to work poorly across some intervening routers, firewalls, and NA(P)T boxes. As a result, the endpoints may be unable to establish an IPsec security association.

Due to these fragmentation shortcomings, IKE is most appropriate for intra-domain usage. Known solutions to the fragmentation problem include sending the end-entry machine certificate rather than the chain, reducing the size of the certificate chain, using IKE implementations over a reliable transport protocol (e.g., TCP) assisted by Path MTU discovery and code against black-holing as per [RFC2923], or installing network components that can properly handle fragments.

IKE negotiators SHOULD check the pertinent Certificate Revocation List (CRL) [RFC2408] before accepting a certificate for use in IKE's authentication procedures.

10.4. iSNS and iFCP Security

iFCP implementations MUST use iSNS for discovery and management services. Consequently, the security of the iSNS protocol has an impact on the security of iFCP gateways. For a discussion of potential threats to iFCP gateways through use of iSNS, see [ISNS].

To provide security for iFCP gateways using the iSNS protocol for discovery and management services, the IPSec ESP protocol in tunnel mode MUST be supported for iFCP gateways. Further discussion of iSNS security implementation requirements is found in [ISNS]. Note that iSNS security requirements match those for iFCP described in Section 10.3.

10.5. Use of iSNS to Distribute Security Policy

Once communication between iFCP gateways and the iSNS server has been secured through use of IPSec, the iFCP gateways have the capability to discover the security settings that they need to use (or not use) to protect iFCP traffic. This provides a potential scaling advantage over device-by-device configuration of individual security policies for each iFCP gateway. It also provides an efficient means for each iFCP gateway to discover the use or non-use of specific security capabilities by peer gateways.

Further discussion on use of iSNS to distribute security policies is found in [ISNS].

10.6. Minimal Security Policy for an iFCP Gateway

An iFCP implementation may be able to disable security mechanisms for an iFCP Portal administratively through a management interface or through security policy elements set in the iSNS server. As a consequence, IKE or IPsec security associations will not be established for any iFCP sessions that traverse the portal.

For most IP networks, it is inappropriate to assume physical security, administrative security, and correct configuration of the network and all attached nodes (a physically isolated network in a test lab may be an exception). Therefore, authentication SHOULD be used in order to provide minimal assurance that connections have initially been opened with the intended counterpart. The minimal iFCP security policy only states that an iFCP gateway SHOULD authenticate its iSNS server(s) as described in [ISNS].

11. Quality of Service Considerations

11.1. Minimal Requirements

Conforming iFCP protocol implementations SHALL correctly communicate gateway-to-gateway, even across one or more intervening best-effort IP regions. The timings with which such gateway-to gateway communication is performed, however, will greatly depend upon BER, packet losses, latency, and jitter experienced throughout the best-effort IP regions. The higher these parameters, the higher the gap measured between iFCP observed behaviors and baseline iFCP behaviors (i.e., as produced by two iFCP gateways directly connected to one another).

11.2. High Assurance

It is expected that many iFCP deployments will benefit from a high degree of assurance regarding the behavior of intervening IP regions, with resulting high assurance on the overall end-to-end path, as directly experienced by fibre channel applications. Such assurance on the IP behaviors stems from the intervening IP regions supporting standard Quality-of-Service (QoS) techniques that are fully complementary to iFCP, such as:

a) congestion avoidance by over-provisioning of the network,

b) integrated Services [RFC1633] QoS,

c) differentiated Services [RFC2475] QoS, and

d) Multi-Protocol Label Switching [RFC3031].

One may load an MPLS forwarding equivalence class (FEC) with QoS class significance, in addition to other considerations such as protection and diversity for the given path. The complementarity and compatibility of MPLS with Differentiated Services is explored in [MPSLDS], wherein the PHB bits are copied to the EXP bits of the MPLS shim header.

In the most general definition, two iFCP gateways are separated by one or more independently managed IP regions that implement some of the QoS solutions mentioned above. A QoS-capable IP region supports the negotiation and establishment of a service contract specifying the forwarding service through the region. Such contract and negotiation rules are outside the scope of this document. In the case of IP regions with DiffServ QoS, the reader should refer to Service Level Specifications (SLS) and Traffic Conditioning Specifications (TCS) (as defined in [DIFTERM]). Other aspects of a

service contract are expected to be non-technical and thus are outside of the IETF scope.

Because fibre channel Class 2 and Class 3 do not currently support fractional bandwidth guarantees, and because iFCP is committed to supporting fibre channel semantics, it is impossible for an iFCP gateway to infer bandwidth requirements autonomously from streaming fibre channel traffic. Rather, the requirements on bandwidth or other network parameters need to be administratively set into an iFCP gateway, or into the entity that will actually negotiate the forwarding service on the gateway's behalf. Depending on the QoS techniques available, the stipulation of a forwarding service may require interaction with network ancillary functions, such as admission control and bandwidth brokers (via RSVP or other signaling protocols that an IP region may accept).

The administrator of a iFCP gateway may negotiate a forwarding service with IP region(s) for one, several, or all of an iFCP gateway's TCP sessions used by an iFCP gateway. Alternately, this responsibility may be delegated to a node downstream. Since one TCP connection is dedicated to each iFCP session, the traffic in an individual N_PORT to N_PORT session can be singled out by iFCP-unaware network equipment as well.

For rendering the best emulation of fibre channel possible over IP, it is anticipated that typical forwarding services will specify a fixed amount of bandwidth, null losses, and, to a lesser degree of relevance, low latency and low jitter. For example, an IP region using DiffServ QoS may support SLSes of this nature by applying EF DSCPs to the iFCP traffic.

12. IANA Considerations

The IANA-assigned port for iFCP traffic is port number 3420.

An iFCP Portal may initiate a connection using any TCP port number consistent with its implementation of the TCP/IP stack, provided each port number is unique. To prevent the receipt of stale data associated with a previous connection using a given port number, the provisions of [RFC1323], Appendix B, SHOULD be observed.

13. Normative References

[AESCBC] Frankel, S. and H. Herbert, "The AES-XCBC-MAC-96 Algorithm and Its Use With IPsec", RFC 3566, September 2003.

[AESCTR] Housley, R., "Using Advanced Encryption Standard (AES) Counter Mode With IPsec Encapsulating Security Payload (ESP)", RFC 3686, January 2004.

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

[FC-FS] dpANS INCITS.XXX-200X, "Fibre Channel Framing and Signaling (FC-FS), Rev 1.70, INCITS Project 1331D, February 2002

[FC-GS3] dpANS X3.XXX-200X, "Fibre Channel Generic Services -3 (FC-GS3)", revision 7.01, INCITS Project 1356-D, November 2000

[FC-SW2] dpANS X3.XXX-2000X, "Fibre Channel Switch Fabric -2 (FC-SW2)", revision 5.2, INCITS Project 1305-D, May 2001

[FCP-2] dpANS T10, "Fibre Channel Protocol for SCSI, Second Version", revision 8, INCITS Project 1144D, September 2002

[ISNS] Tseng, J., Gibbons, K., Travostino, F., Du Laney, C., and J. Souza, "Internet Storage Name Service (iSNS)", RFC 4171, September 2005.

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

[RFC2401] Kent, S. and R. Atkinson, "Security Architecture for the Internet Protocol", RFC 2401, November 1998.

[RFC2402] Kent, S. and R. Atkinson, "IP Authentication Header", RFC 2402, November 1998.

[RFC2404] Madson, C. and R. Glenn, "The Use of HMAC-SHA-1-96 within ESP and AH", RFC 2404, November 1998.

[RFC2406] Kent, S. and R. Atkinson, "IP Encapsulating Security Payload (ESP)", RFC 2406, November 1998.

[RFC2407] Piper, D., "The Internet IP Security Domain of Interpretation for ISAKMP", RFC 2407, N.

[RFC2408] Maughan, D., Schertler, M., Schneider, M., and J. Turner, "Internet Security Association and Key Management Protocol (ISAKMP)", RFC 2408, November 1998.

[RFC2409] Harkins, D. and D. Carrel, "The Internet Key Exchange (IKE)", RFC 2409, November 1998.

[RFC2410] Glenn, R. and S. Kent, "The NULL Encryption Algorithm and Its Use With IPsec", RFC 2410, November 1998.

[RFC2451] Pereira, R. and R. Adams, "The ESP CBC-Mode Cipher Algorithms", RFC 2451, November 1998.

[RFC793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, September 1981.

[SECIPS] Aboba, B., Tseng, J., Walker, J., Rangan, V., and F. Travostino, "Securing Block Storage Protocols Over IP", RFC 3723, April 2004.

14. Informative References

[AES] FIPS Publication XXX, "Advanced Encryption Standard (AES)", Draft, 2001, Available from http://csrc.nist.gov/publications/drafts/dfips-AES.pdf

[DIFTERM] Grossman, D., "New Terminology and Clarifications for Diffserv", RFC 3260, April 2002.

[FC-AL2] dpANS X3.XXX-199X, "Fibre Channel Arbitrated Loop (FC-AL-2)", revision 7.0, NCITS Project 1133D, April 1999

[FC-FLA] TR-20-199X, "Fibre Channel Fabric Loop Attachment (FC-FLA)", revision 2.7, NCITS Project 1235-D, August 1997

[FC-VI] ANSI/INCITS 357:2002, "Fibre Channel Virtual Interface Architecture Mapping Protocol (FC-VI)", NCITS Project 1332-D, July 2000.

[KEMALP] Kembel, R., "The Fibre Channel Consultant, Arbitrated Loop", Robert W. Kembel, Northwest Learning Associates, 2000, ISBN 0-931836-84-0

[KEMCMP] Kembel, R., "Fibre Channel, A Comprehensive Introduction", Northwest Learning Associates Inc., 2000, ISBN 0-931836-84-0

[MPSLDS] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, May 2002.

[RFC1122] Braden, R., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, October 1989.

[RFC1323] Jacobson, V., Braden, R., and D. Borman, "TCP Extensions for High Performance", RFC 1323, May 1992.

[RFC1633] Braden, R., Clark, D., and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.

[RFC2030] Mills, D., "Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI", RFC 2030, October 1996.

[RFC2405] Madson, C. and N. Doraswamy, "The ESP DES-CBC Cipher Algorithm With Explicit IV", RFC 2405, November 1998.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Service", RFC 2475, December 1998.

[RFC2625] Rajagopal, M., Bhagwat, R., and W. Rickard, "IP and ARP over Fibre Channel", RFC 2625, June 1999.

[RFC2709] Srisuresh, P., "Security Model with Tunnel-mode IPsec for NAT Domains", RFC 2709, October 1999.

[RFC2923] Lahey, K., "TCP Problems with Path MTU Discovery", RFC 2923, September 2000.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC896] Nagle, J., "Congestion control in IP/TCP internetworks", RFC 896, January 1984.

Appendix A. iFCP Support for Fibre Channel Link Services

For reference purposes, this appendix enumerates all the fibre channel link services and the manner in which each shall be processed by an iFCP implementation. The iFCP processing policies are defined in Section 7.

In the following sections, the name of a link service specific to a particular FC-4 protocol is prefaced by a mnemonic identifying the protocol.

A.1. Basic Link Services

The basic link services are shown in the following table:

Basic Link Services

Name	Description	iFCP Policy
ABTS	Abort Sequence	Transparent
BA_ACC	Basic Accept	Transparent
BA_RJT	Basic Reject	Transparent
NOP	No Operation	Transparent
PRMT	Preempted	Rejected (Applies to Class 1 only)
RMC	Remove Connection	Rejected (Applies to Class 1 only)

A.2. Pass-Through Link Services

As specified in Section 7, the link service requests of Table 10 and the associated ACC response frames MUST be passed to the receiving N_PORT without altering the payload.

Name	Description
ADVC	Advise Credit
CSR	Clock Synchronization Request
CSU	Clock Synchronization Update
ECHO	Echo
ESTC	Estimate Credit
ESTS	Establish Streaming
FACT	Fabric Activate Alias_ID
FAN	Fabric Address Notification

FCP_RJT	FCP FC-4 Link Service Reject
FCP SRR	FCP Sequence Retransmission Request
FDACT	Fabric Deactivate Alias_ID
FDISC	Discover F_Port Service Parameters
FLOGI	F_Port Login
GAID	Get Alias_ID
LCLM	Login Control List Management
LINIT	Loop Initialize
LIRR	Link Incident Record Registration
LPC	Loop Port Control
LS_RJT	Link Service Reject
LSTS	Loop Status
NACT	N_Port Activate Alias_ID
NDACT	N_Port Deactivate Alias_ID
PDISC	Discover N_Port Service Parameters
PRLI	Process Login
PRLO	Process Logout
QoSR	Quality of Service Request
RCS	Read Connection Status
RLIR	Registered Link Incident Report
RNC	Report Node Capability
RNFT	Report Node FC-4 Types
RNID	Request Node Identification Data
RPL	Read Port List
RPS	Read Port Status Block
RPSC	Report Port Speed Capabilities
RSCN	Registered State Change Notification
RTV	Read Timeout Value
RVCS	Read Virtual Circuit Status
SBRP	Set Bit-Error Reporting Parameters
SCN	State Change Notification
SCR	State Change Registration
TEST	Test
TPLS	Test Process Login State

Table 10. Pass-Through Link Services

A.3. Special Link Services

The extended and FC-4 link services of Table 11 are processed by an iFCP implementation as described in the sections referenced in the table.

Name	Description	Section
ABTX	Abort Exchange	7.3.1.1
ADISC	Discover Address	7.3.1.2
ADISC ACC	Discover Address Accept	7.3.1.3
FARP-REPLY	Fibre Channel Address Resolution Protocol Reply	7.3.1.4
FARP-REQ	Fibre Channel Address Resolution Protocol Request	7.3.1.5
LOGO	N_PORT Logout	7.3.1.6
PLOGI	Port Login	7.3.1.7
REC	Read Exchange Concise	7.3.1.8
REC ACC	Read Exchange Concise Accept	7.3.1.9
FCP REC	FCP Read Exchange Concise (see [FCP-2])	7.3.2.1.1
FCP REC ACC	FCP Read Exchange Concise Accept (see [FCP-2])	7.3.2.1.2
RES	Read Exchange Status Block	7.3.1.10
RES ACC	Read Exchange Status Block Accept	7.3.1.11
RLS	Read Link Error Status Block	7.3.1.12
RRQ	Reinstate Recovery Qualifier	7.3.1.14
RSI	Request Sequence Initiative	7.3.1.15
RSS	Read Sequence Status Block	7.3.1.13
SRL	Scan Remote Loop	7.3.1.16
TPRLO	Third Party Process Logout	7.3.1.17
TPRLO ACC	Third Party Process Logout Accept	7.3.1.18

Table 11. Special Link Services

Appendix B. Supporting the Fibre Channel Loop Topology

A loop topology may be optionally supported by a gateway implementation in one of the following ways:

a) By implementing the FL_PORT public loop interface specified in [FC-FLA].

b) By emulating the private loop environment specified in [FC-AL2].

Private loop emulation allows the attachment of fibre channel devices that do not support fabrics or public loops. The gateway presents such devices to the fabric as though they were fabric-attached. Conversely, the gateway presents devices on the fabric, whether they are locally or remotely attached, as though they were connected to the private loop.

Private loop support requires gateway emulation of the loop primitives and control frames specified in [FC-AL2]. These frames and primitives MUST be locally emulated by the gateway. Loop control frames MUST NOT be sent over an iFCP session.

B.1. Remote Control of a Public Loop

A gateway MAY disclose that a remotely attached device is connected to a public loop. If it does, it MUST also provide aliases representing the corresponding Loop Fabric Address (LFA), DOMAIN_ID, and FL_PORT Address Identifier through which the public loop may be remotely controlled.

The LFA and FL_PORT address identifier both represent an N_PORT that services remote loop management requests contained in the LINIT and SRL extended link service messages. To support these messages, the gateway MUST allocate an NL_PORT alias so that the corresponding alias for the LFA or FL_PORT address identifier can be derived by setting the Port ID component of the NL_PORT alias to zero.

Acknowledgements

The authors are indebted to those who contributed material and who took the time to carefully review and critique this specification including David Black (EMC), Rory Bolt (Quantum/ATL), Victor Firoiu (Nortel), Robert Peglar (XIOtech), David Robinson (Sun), Elizabeth Rodriguez, Joshua Tseng (Nishan), Naoke Watanabe (HDS) and members of the IPS working group. For review of the iFCP security policy, the authors are further indebted to the authors of the IPS security document [SECIPS], which include Bernard Aboba (Microsoft), Ofer Biran (IBM), Uri Elzer (Broadcom), Charles Kunziger (IBM), Venkat Rangan (Rhapsody Networks), Julian Satran (IBM), Joseph Tardo (Broadcom), and Jesse Walker (Intel).

Author's Addresses

Comments should be sent to the ips mailing list (ips@ece.cmu.edu) or to the authors.

Charles Monia
7553 Morevern Circle
San Jose, CA 95135

EMail: charles_monia@yahoo.com

Rod Mullendore
McDATA
4555 Great America Pkwy
Suite 301
Santa Clara, CA 95054

Phone: 408-519-3986
EMail: Rod.Mullendore@MCDATA.com

Franco Travostino
Nortel
600 Technology Park Drive
Billerica, MA 01821 USA

Phone: 978-288-7708
EMail: travos@nortel.com

Wayland Jeong
TROIKA Networks, Inc.
2555 Townsgate Road, Suite 105
Westlake Village, CA 91361

Phone: 805-371-1377
EMail: wayland@TroikaNetworks.com

Mark Edwards
Adaptec (UK) Ltd.
4th Floor, Howard House
Queens Ave, UK. BS8 1SD

Phone: +44 (0)117 930 9600
EMail: mark_edwards@adaptec.com

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.