IP Storage Working Group
INTERNET DRAFT
Expires July 2002
<draft-ietf-ips-ifcp-09.txt>

Charles Monia
Rod Mullendore
Josh Tseng
Nishan Systems

Franco Travostino
Nortel Networks

David Robinson
Sun Microsystems

Wayland Jeong
Troika Networks

Rory Bolt
Quantum/ATL

Mark Edwards
Eurologic

January 2002

iFCP - A Protocol for Internet Fibre Channel Storage Networking

Status of this Memo

Comments

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

1. Abstract

This document specifies an architecture and gateway-to-gateway protocol for the implementation of Fibre Channel fabric functionality on a network in which TCP/IP switching and routing elements replace Fibre Channel components. The protocol enables the attachment of Fibre Channel devices to an IP network by supporting the fabric services required by such devices.

2. About This Document

2.1 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 RFC-2119 [RFC2119].

All frame formats are in big endian network byte order.

~~1.2~~2.2 Purpose of this document

This is a standards-track document, which 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 NCITS 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 Channelprotocol implementations to TCP/IP. This mapping includes sections of this document which describe the "iFCP Protocol" (see section 6).

3. iFCP Introduction

iFCP is a gateway-to-gateway protocol, which 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.

~~1.1~~3.1 Definitions

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

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

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

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.

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~~logical~~ fabric to which the gateway belongs.

Gateway Region – The portion of the iFCP storage network accessed through an iFCP gateway. Fibre Channel devices in the region consist of all Fibre Channel devices locally attached to the gateway.

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

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

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

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].

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

Fabric - The components of a network that provide the transport services defined in [FC-FS]. A fabric may be implemented in the IP framework by means of the architecture and protocols discussed in this document.

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 to be functionally equivalent.

FC-2 - The Fibre Channel transport services layer described in [FC-FS].

FC-4 - The Fibre Channel application layer. This layer is functionally equivalent to the TCP/IP application layer.

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.

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 ~~are ~~depend~~ant~~ on the Fibre Channel fabric topology. As used in this document, the term applies equally to all variants.

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 I/Ds are assigned in accordance with the Fibre Channel rules for address assignment specified in [FC-FS].

N_PORT Alias -- The N_PORT address assigned by a gateway to represent a remote N_PORT accessed via the iFCP protocol. When routing frame traffic in address translation mode, the gateway automatically converts N_PORT aliases to N_PORT network addresses and vice versa.

N_PORT Network Address - The address of an N_PORT in the iFCP~~IP~~ 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.

F_PORT - The interface used by an N_PORT to access Fibre Channel switched fabric functionality.

iFCP - The protocol discussed in this document.

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

iSNS - The server functionality and IP protocol which provides storage name services in an iFCP network. Fibre Channel Name services are implemented by an iSNS name server as described in [ISNS].

N_PORT Session - An association created when two N_PORTS have executed a PLOGI operation. It is comprised of the N_PORTs and TCP connection that carries traffic between them.

iFCP Frame - A Fibre Channel frame encapsulated in accordance with the Common Encapsulation Specification [ENCAP] and this specification.

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

DOMAIN_ID – The value contained in the high-order byte of a 24-bit N_PORT Fibre Channel address.

4. 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 5.

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].

~~1.1~~4.1 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~~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 interface within a fabric that provides Fibre Channel attachment for an N_PORT. The types of fabric port depend on the fabric topology and are discussed in section 4.2.

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

e) Within a switched or mixed fabric (see section 4.2), 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.

```
  +--------+   +--------+          +--------+  +--------+
  |  FC    |   |  FC    |          |  FC    |  |  FC    |
  | Device |   | Device |<-------->| Device |  | Device |
  |........|   |........|          |........|  |........|
  | N_PORT |   | N_PORT |          | N_PORT |  | N_PORT |
  +---+----+   +----+---+          +----+---+  +----+---+
      |             |                   |           |
  +---+----+   +----+---+          +----+---+  +----+---+
  | Fabric |   | Fabric |          | Fabric |  | Fabric |
  | Port   |   | Port   |          | Port   |  | Port   |
  +========+===+========+==========+========+==+========+
  |                        Fabric                       |
  |                          &                          |
  |                     Fabric Services                 |
  +-----------------------------------------------------+
```

Figure 1 -- A Fibre Channel Network

The following sections describe Fibre Channel fabric topologies and give an overview of the Fibre Channel communications model.

~~1.2~~4.2 Fabric Topologies

The principal Fibre Channel fabric topologies consist of the following:

a) Arbitrated Loop -- A series of N_PORTs connected together in daisy-chain fashion. Data transmission between N_PORTs requires arbitration for control of the loop in a manner similar to a token ring network.

b) Switched Fabric -- A fabric consisting of switching elements, as described in section 4.2.1.

c) Mixed Fabric -- A fabric consisting of switches and "fabric-attached" loops. A description can be found in [FC-FLA].

Depending on the topology, the N_PORT and fabric port variants through which a Fibre Channel device is attached to the network may be one of the following:

Fabric Topology	Fabric Port Type	N_PORT Variant
Loop	L_PORT	NL_PORT
Switched	F_PORT	N_PORT
Mixed	FL_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.

~~1.1.1~~4.2.1 Switched Fibre Channel Fabrics

An example of a multi-switch Fibre Channel fabric is shown below.

```
            +----------+          +----------+
            |    FC    |          |  FC      |
            |   Device |          | Device   |
            |..........|          |..........|
            |   N_PORT |<........>| N_PORT   |
            +----+-----+          +-----+----+
                 |                      |
            +----+-----+          +-----+----+
            | F_PORT   |          | F_PORT   |
  ==========+==========+==========+==========+==============
            |  FC      |          | FC       |
            |  Switch  |          | Switch   |
            +----------+          +----------+ Fibre Channel
            |Inter-    |          |Inter-    |   Fabric
            |Switch    |          |Switch    |
            |Interface |          |Interface |
            +-----+----+          +-----+----+
                  |                     |
                  |                     |
            +-----+----+----------+-----+----+
            |Inter-    |          |Inter-    |
            |Switch    |          |Switch    |
            |Interface |          |Interface |
            +----------+          +----------+
            |            FC Switch           |
            |                                |
            +--------------------------------+
```

Figure 3~~2~~ -- Multi-Switch Fibre Channel Fabric

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

~~1.1.2~~4.2.2 Mixed Fibre Channel Fabric

A mixed fabric contains one or more arbitrated loops connected to a switched fabric as shown in Figure 4~~Figure 3~~.

```
            +----------+          +----------+   +---------+
            |    FC    |          |  FC      |   |  FC     |
            |   Device |          | Device   |   | Device  |
            |..........|          |..........|   |.........|
            |   N_PORT |<........>| NL_PORT  +---+ NL_PORT |
            +----+-----+          +-----+----+   +----+----+
                 |                      |   FC Loop   |
            +----+-----+          +-----+----+        |
            | F_PORT   |          | FL_PORT  +--------+
            |          |          |          |
  ==========+==========+==========+==========+==============
            |  FC      |          | FC       |
            |  Switch  |          | Switch   |
            +----------+          +----------+
            |Inter-    |          |Inter-    |
            |Switch    |          |Switch    |
            |Interface |          |Interface |
            +-----+----+          +-----+----+
                  |                     |
                  |                     |
            +-----+----+----------+-----+----+
            |Inter-    |          |Inter-    |
            |Switch    |          |Switch    |
            |Interface |          |Interface |
            +----------+          +----------+
            |            FC Switch           |
            |                                |
            +--------------------------------+
```

Figure 4~~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.

~~1.3~~4.3 Fibre Channel Layers and Link Services

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 FCP, the Fibre Channel SCSI protocol.

In addition to the layers defined above, Fibre Channel defines a set of auxiliary operations, some of which are implemented within the transport layer fabric, called link services. These are required to manage the Fibre Channel environment, establish communications with other devices, retrieve error information, perform error recovery and 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.

~~1.1.1~~4.3.1 Fabric-Supplied Link Services

Servers 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 an N_PORT request 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 4.5).

Directory/Name Server – Allows N_PORTs to register information in a database, retrieve information about other N_PORTs and discover other devices as described in section 4.5.

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 and is 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.

~~1.4~~4.4 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 world-wide unique identifier for the node,

b) A world-wide unique identifier for each N_PORT associated with the nodee,

c) For each N_PORT attached to a fabric, a 24-bit fabric-unique address having the properties defined in section 4.7.1. The fabric address is the address to which frames are sent.

Each world-wide unique identifier is a 64-bit binary quantity having the format defined in [FC-FS].

~~1.5~~4.5 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 discover the existence of other devices. For example, one type of query obtains the fabric address of an N_PORT from its 64-bit world-wide 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].

~~1.6~~4.6 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 having the structure described in section 4.7. 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 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. Each transaction, in turn, consists of protocol information units, each of which is carried by an individual Sequence within an Exchange.

~~1.7~~4.7 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 highly simplified view of the frame.

```
+-----------------------------+
|   Start-of-frame Delimiter  |
+-----+-----------------------+<----+
|     | Destination N_PORT    |     |
|     | Fabric Address (D_ID) |     |
|     |  (24-bits)            |     |
+-----+-----------------------+   24-byte
|     | Source N_PORT         |   Frame
|     | Fabric Address (S_ID) |   Header
|     | (24 bits)             |     |
+-----+-----------------------+     |
|    Control information for  |     |
|    frame type, Exchange     |     |
|    management, IU           |     |
|    segmentation and         |     |
|    re-assembly              |     |
+-----------------------------+<----+
|                             |
|        Frame payload        |
|       (0 – 2112 bytes)      |
|                             |
|                             |
|                             |
+-----------------------------+
|             CRC             |
+-----------------------------+
|    End-of-Frame Delimiter   |
+-----------------------------+
```

Figure 6~~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 MAC addresses of originating and receiving N_PORTs.

~~1.1.1~~4.7.1 N_PORT Address Model

N_PORT fabric addresses are 24-bit values having the following format defined by the Fibre Channel specification [FC-FS]:

```
       Bit   23        16 15          8 7         0
            +-----------+------------+----------+
            | Domain ID | Area ID    |  Port ID |
            +-----------+------------+----------+
```

Figure 7~~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 I/D assigned by the principal switch. The value of the Domain I/D 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 a F_PORT receives a unique fabric address consisting of the switch’s Domain I/D concatenated with switch-assigned area and port I/Ds.

A loop-attached NL_PORT (see Figure 4~~Figure 3~~) obtains the Port ID component of its address during the loop initialization process described in [FC-AL2]. The area and domain I/Ds are supplied by the fabric when the FLOGI is executed.

~~1.8~~4.8 Fibre Channel Transport Services

N_PORTs communicate by means of the following classes of service 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.

Class 2 and class 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 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.

~~1.9~~4.9 Login Processes

The Login processes are the means whereby an N_PORT establishes 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 which 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 world-wide unique name of an N_PORT to obtain the 24-bit N_PORT Fibre Channel address to which the PLOGI request is sent.

5. The iFCP Network Model

The iFCP protocol enables the implementation of Fibre Channel mixed or switched fabric functionality on an IP network in which IP components and technology replace the Fibre Channel switching and routing infrastructure described in section 4.2.

The example of Figure 8~~Figure 6~~ shows a Fibre Channel fabric with attached devices. These access the fabric through an N_PORT interface connected to a Fabric Port whose behavior is specified in [FC-FS]. In this case, the N_PORT and Fabric Port represent any of the variants described in section 4.2.

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

```
                Fibre Channel Network
            +--------+        +--------+
            |  FC    |        |  FC    |
            | Device |        | Device |
            |........|        |........| Fibre Channel
            | N_PORT |<......>| N_PORT | Device Domain
            +---+----+        +----+---+       ^
                |                  |           |
            +---+----+        +----+---+       |
            | Fabric |        | Fabric |       |
            | Port   |        | Port   |       |
  ==========+========+========+========+==============
            |       Fabric &           |       |
            |    Fabric Services       |       v
            |                          | Fibre Channel
            +--------------------------+ Fabric Domain
```

Figure 8~~6~~ -- A Fibre Channel Fabric

```
      Gateway Region                   Gateway Region
 +--------+  +--------+           +--------+  +--------+
 |   FC   |  |  FC    |           |   FC   |  |   FC   |
 | Device |  | Device | Fibre     | Device |  | Device |   Fibre
 |........|  |........| Channel   |........|  |........|   Channel
 | N_PORT |  | N_PORT |<.........>| N_PORT |  | N_PORT |   Device
 +---+----+  +---+----+ Traffic   +----+---+  +----+---+   Domain
     |           |                     |           |         ^
 +---+----+  +---+----+           +----+---+  +----+---+     |
 | Fabric |  | Fabric |           | Fabric |  | Fabric |     |
 | Port   |  | Port   |           | Port   |  | Port   |     |
=+========+==+========+===========+========+==+========+==========
 |    iFCP Layer      |<--------->|     iFCP Layer     |     |
 |....................|     ^     |....................|     |
 |     iFCP Portal    |     |     |     iFCP Portal    |     v
 +--------+-----------+     |     +----------+---------+    IP
      iFCP|Gateway     Control          iFCP|Gateway     Network
          |              Data               |
          |                                  |
          |                                  |
          |                                  |
          |<------Encapsulated Frames------->|
          |      +------------------+        |
          |      |                  |        |
          +------+    IP Network    +--------+
                 |                  |
                 +------------------+
```

Figure 10~~7~~ -- An iFCP Fabric

Figure 10~~Figure 7~~ shows an implementation of an equivalent iFCP fabric consisting of two gateways, each in control of a single gateway region.

Each iFCP gateway contains two standards-compliant fibre channel 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.

~~1.1~~5.1 Fibre Channel Fabric Topologies Supported by iFCP

A property of this architecture, not shown in the examples, is that the Fibre Channel fabric configuration and topology within the gateway region are invisible to the IP network and other gateway regions. That is, the topology in the gateway region, whether it is loop- or switch-based, is hidden from the IP network and from other gateway regions. As a result, support for specific FC fabric topologies becomes a gateway implementation issue. In such cases, the gateway may implement any standards-compliant Fibre Channel interface by incorporating the functionality required to present locally attached N_PORTs as logical iFCP devices.

~~1.2~~5.2 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 5.5.

b) Execution of fabric-supplied link services addressed to one of the well-known Fibre Channel N_PORT addresses.

c) 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.

d) Establishment of an N_PORT login session in response to a PLOGI directed to a remote device.

The following sections discuss the frame addressing mechanism and the way in which it is used to achieve communications transparency between N_PORTs.

~~1.1.1~~5.2.1 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 the Class 1 (dedicated connection) service. An N_PORT discovers the classes of transport services supported by the fabric during login.

~~1.3~~5.3 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) Emulation of the services provided by the Fibre Channel name server described in section 4.3.1, including a mechanism for asynchronously notifying an N_PORT of changes in the iFCP fabric configuration,

b) Aggregation of gateways into iFCP fabrics for interoperation,

c) Segmentation of an iFCP fabric into Fibre Channel zones through the definition and management of device discovery scopes, referred to as 'discovery domains',

d) Storage and distribution of security policies as described in section 11.2.9.

e) Implementation of the Fibre Channel broadcast mechanism.

5.4 iFCP Fabric Properties

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

Gateways in a bounded iFCP fabric operate in address transparent mode as described in section 5.5.1. In this mode, the scope of a Fibre Channel N_PORT address is fabric-wide and is derived from domain I/Ds issued by the iSNS server from a common pool. As discussed below, the maximum number of domain I/Ds allowed by 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 5.5.2. 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, an unbounded iFCP fabric may have any number of switch elements and gateways.

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 discussed below.

~~1.1.1~~5.4.1 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, while both bounded and unbounded iFCP fabrics support the standards-compliant FC-4 protocols 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 I/Ds in the frame payload.

~~1.1.2~~5.4.2 Configuration Scalability

The scalability limits of a bounded fabric configuration are a consequence of the Fibre Channel address allocation policy previously discussed. As noted, a bounded iFCP fabric using this address allocation scheme is limited to a combined total of 238 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, there are no architectural limits on the number of iFCP gateways, Fibre Channel devices and switch elements that may be internetworked. 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 5.4.1.

~~1.1.3~~5.4.3 Fault Tolerance

In an unbounded iFCP fabric, limiting the scope of an N_PORT address to a gateway region reduces the likelihood that reassignment of domain I/Ds caused by a disruption in one gateway region will cascade to others.

In addition, a bounded iFCP fabric has an increased dependency on 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 to the ~~logical~~ fabric.

Finally, adding a gateway to a bounded fabric is more likely to disrupt the operation of all devices in the gateway region along with those already in the fabric as new, fabric-wide N_PORT addresses are assigned. Furthermore, before the new gateway can be merged, its iSNS server must be slaved to the iSNS server in the bounded fabric to centralize the issuance of domain I/Ds.

In contrast, adding a new gateway to an unbounded iFCP fabric can be done non-disruptively and requires only that new gateway's iSNS server import client attributes from the other iSNS servers.

~~1.5~~5.5 The iFCP N_PORT Address Model

This section discusses iFCP extensions to the Fibre Channel addressing model of section 4.7.1, 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 I/D. The Fibre Channel N_PORT address of a locally attached device. Depending on the gateway addressing mode, the scope is either local to a region or fabric-wide. In either mode, communications between N_PORTs in the same gateway region use the N_PORT I/D.

b) A 24-bit N_PORT alias. An address assigned by a gateway operating in address translation mode to identify a remotely attached N_PORT. Frame traffic is directed to a remotely attached N_PORT by means of the N_PORT alias.

c) An N_PORT network address. A tuple consisting of the gateway IP address, TCP port number and N_PORT I/D. 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 remote and local N_PORTs, a gateway in address translation mode maintains an association between the remote N_PORT alias 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, it 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 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 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 I/Ds from the address assignment authority as described in section 5.5.1.1. 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 to represent remote N_PORTs.

All iFCP implementations MUST support operation in address translation mode. Implementation of address transparent mode is OPTIONAL but 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 allow the mode to be changed after the gateway begins processing fibre channel frame traffic.

~~1.1.1~~5.5.1 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 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 5.5.1.1, DOMAIN_ID values assigned to FC switches in attached fabrics must be issued by the iSNS server.

c) When operating in address transparent Mode, no Fibre Channel address translation SHALL take place.

The process for establishing the TCP/IP context associated with an N_PORT login session in this mode is similar to that specified for address translation mode (section 5.5.2).

~~1.1.1.1~~5.5.1.1 Transparent Mode Domain I/D Management

As described above, each gateway and Fibre Channel switch in a bounded iFCP fabric MUST have a unique domain I/D. In a gateway region containing Fibre Channel switch elements, each element obtains a domain I/D by querying the principal switch as described in [FC-SW2]-- in this case the iFCP gateway itself. The gateway in turn may obtain domain I/Ds on demand from the iSNS name server acting as the central address allocation authority . In effect, the iSNS server assumes the role of master switch for the bounded fabric. In that case, the iSNS database contains:

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

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

In its role as principle switch, an iFCP gateway in address transparent mode SHALL obtain domain I/Ds for use in the gateway region by issuing the appropriate iSNS query using its world-wide name.

~~1.1.1.2~~5.5.1.2 Incompatibility with Address Translation Mode

iFCP gateways in address transparent mode SHALL NOT originate or accept frames that do not have the TRN bit set to one in the iFCP flags field of the encapsulation header (see section 6.4.1). The iFCP gateway SHALL immediately terminate all N_PORT login sessions with the iFCP gateway from which it receives such frames.

~~1.1.2~~5.5.2 Operation in Address Translation Mode

This section describes the process for managing the assignment of addresses within a gateway region, 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 5.5, 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.

When a remotely attached N_PORT is accessed, the gateway assigns a locally significant N_PORT alias. This alias is used in place of the N_PORT I/D assigned by the remote gateway. To perform address conversion and enable the appropriate routing, the gateway maintains a table mapping N_PORT aliases to the appropriate TCP/IP connection context and N_PORT ID of all remotely accessed N_PORTs. The means by which translation table entries are created and updated are described in section 5.5.3.

~~1.1.3~~5.5.3 Address Translation

This section describes how address translation SHALL be performed by a gateway operating in address translation mode. For descriptive purposes, the gateway is assumed to maintain a table containing one entry for each remotely attached N_PORT as shown in Figure 11~~Figure 8~~.

```
+---------------------------------+
|  Network Address of Remote      |
|  Gateway                        |
+---------------------------------+
|  N_PORT I/D of Remote N_PORT    |
+---------------------------------+
|  N_PORT Alias                   |
+---------------------------------+
|  N_PORT World-wide Unique Name  |
+---------------------------------+
```

Figure 11~~8~~ -- Address Translation Table Entry for Remote N_PORT

Each entry contains the following information:

Network Address of Remote Gateway -- IP address and TCP port number of the gateway to which the remote device is attached.

N_PORT I/D -- N_PORT address assigned to the remote device by the remote iFCP gateway.

N_PORT Alias -- N_PORT address assigned to the remote device by the 'local' iFCP gateway.

N_PORT World-wide Unique Name -- 64-bit N_PORT world wide name as specified in [FC-FS].

An iFCP gateway SHALL have one and only one entry for each remotely attached N_PORT it accesses. If an entry does not exist, one SHALL be built in response to one of the following transactions:

a) A Fibre Channel Name Server request issued by a locally-attached N_PORTs as part of Fibre Channel device discovery (see section 4.5) or,

b) An N_PORT PLOGI request received from the remote Fibre Channel device (see section 8.3.1.7).

An iFCP gateway SHALL convert each Fibre Channel Name Server request to an iSNS server query. Information returned in response to the query includes the IP address, TCP port number, N_PORT ID and N_PORT world wide unique name for each remote device included in the query response. After building the table entry containing this information for a specific N_PORT, the iFCP layer SHALL create and add the 24-bit N_PORT alias. This alias SHALL then be returned to the local N_PORT as the Fibre Channel address of the remotely attached device.

If a PLOGI is received from a remotely attached device and no translation table entry exists for that device, an entry SHALL be created using the following information:

a) The world-wide unique name of the N_PORT contained in the PLOGI payload,

b) The IP address and TCP port number of the remote device obtained from the TCP connection context,

c) The N_PORT I/D obtained from the S_ID field in the PLOGI frame header.

The N_PORT alias SHALL then be assigned and used in address translation as specified in section 5.5.2.

~~1.1.1.1.1~~5.5.3.1.1 Updating an Address Translation

An address translation may become stale as the result of any event that invalidates or triggers a change in the fabric-assigned 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 I/D has changed will have no N_PORT login sessions. Consequently, frames directed to an N_PORT as the result of a stale translation table entry will be rejected or discarded by the receiving Fibre Channel device.

Once the originating N_PORT learns of the reconfiguration, usually through the name server state change notification mechanism, the normal name server lookup and PLOGI mechanisms needed to reestablish the N_PORT login session will automatically purge such stale translations from the gateway.

~~1.1.1.2~~5.5.3.2 Frame Address Translation

For outbound frames, the gateway-resident address translation SHALL be referenced to map the Destination N_PORT alias to the TCP connection context and N_PORT ID assigned by the remote gateway. The translation process for outbound frames is shown below.

```
         Raw Fibre Channel Frame
+--------+-----------------------------------+    +--------------+
|        |  Destination N_PORT Alias         |--->| Lookup TCP   |
+--------+-----------------------------------+    | connection   |
|        |  Source N_PORT ID                 |    | context      |
+--------------------------------------------+    | and N_PORT ID|
|                                            |    +------+-------+
|        Control information,                |           | TCP
|        Payload and FC CRC                  |           | conn
|                                            |           | context
+--------------------------------------------+           | &
                                                         | N_PORT
                                                         | ID
                                                         |
After Address Translation and Encapsulation              |
+--------------------------------------------+           |
|          FC Encapsulation Header           |           |
+--------------------------------------------+           |
|            SOF Delimiter Word              |           |
+============================================+           |
|        |  Destination N_PORT ID            |<----------+
+--------+-----------------------------------+
|        |  Source N_PORT ID                 |
+--------+-----------------------------------+
|                                            |
|        Control information, Payload        |
|        and FC CRC                          |
+============================================+
|         EOF Delimiter Word                 |
+--------------------------------------------+
```

Figure 13~~9~~ -- Outbound Frame Address Translation

For inbound frames, a translation SHALL be performed to regenerate the N_PORT alias from the TCP connection context and N_PORT ID contained in Source N_PORT I/D field of theencapsulated FC frame. The translation process for inbound frames is shown below.

```
     Network Format of Inbound Frame
+--------------------------------------------+     TCP
|          FC Encapsulation Header           |     Connection
+--------------------------------------------+     Context
|            SOF Delimiter Word              |         |
+============================================+         V
|        |  Destination N_PORT ID            |     +---+----+
+--------+-----------------------------------+     | Lookup |
|        |  Source N_PORT ID                 |---->| Source |
+--------+-----------------------------------+     | N_PORT |
|                                            |     | Alias  |
|        Control information, Payload        |     +----+---+
|        and FC CRC                          |          | Source
+============================================+          | N_PORT
|         EOF Delimiter Word                 |          | Alias
+--------------------------------------------+          |
                                                        |
                                                        |
Frame after Address Translation and De-encapsulation    |
+--------+-----------------------------------+          |
|        |  Destination N_PORT ID            |          |
+--------+-----------------------------------+          |
|        |  Source N_PORT Alias              |<---------+
+--------+-----------------------------------+
|                                            |
|        Control information, Payload,       |
|        and FC CRC                          |
+--------------------------------------------+
```

Figure 14~~10~~ -- Inbound Frame Address Translation

In both cases, the gateway MUST recalculate the FC CRC after altering the frame contents.

~~1.1.1.3~~5.5.3.3 Incompatibility with Address Transparent Mode

iFCP gateways in address translation mode SHALL NOT originate or accept frames that have the TRN 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 such frames as described in section 6.2.3.2.

6. iFCP Protocol

6.1 Overview

6.1.1 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 setup, 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 6.1.2

~~1.1.2~~6.1.2 iFCP Support for Link Services

iFCP must intervene in the processing of those Fibre Channel link service messages which contain N_PORT addresses in the message payload or 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, 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 interprets the frame, including any supplemental information, modifies the frame content, and forwards the resulting frame to the destination N_PORT for further processing.

Section 8.1 describes the processing of these link service messages in detail.

~~1.2~~6.2 TCP Stream Transport of iFCP Frames

6.2.1 iFCP Session Model

An iFCP session consists of the pair of N_PORTs comprising the session endpoints joined by a single TCP/IP connection.

An N_PORT is identified by its network address consisting of:

a) The N_PORT I/D 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.

Since 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 7.1).

When the iFCP layer detects a Port Login (PLOGI) message creating an iFCP session between a pair of N_PORTs, 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 PLOGI login session.

~~1.1.2~~6.2.2 iFCP Session Management

This section describes the protocols for establishing and terminating an N_PORT login session.

~~1.1.1.1~~6.2.2.1 Creating an iFCP Session

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

a) OPEN -- The session state in which Fibre Channel frame images may be sent and received.

b) 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 gateway SHALL initiate the creation of an iFCP session in response to a PLOGI ELS directed to a remote N_PORT from a locally attached N_PORT as described in the following steps.

a) If no iFCP session exists, 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. The network address of the remote gateway is obtained from the address translation table created as described in section 5.5.3

b) If a connection cannot be allocated or created due to limited resources, the gateway SHALL terminate the PLOGI with an LS_RJT response. The Reason Code field in the LS_RJT message shall be set to 0x09 (Unable to Perform Command Request) and the Reason Explanation SHALL be set to 0x29 (Insufficient Resources to Support Login).

c) If an iFCP session in the OPEN state already exists to the remote N_PORT, the gateway SHALL forward the PLOGI ELS using the existing session.

d) If the iFCP session does not exist, the gateway SHALL issue a CBIND session control message (see section 7.1) and place the session in the OPEN PENDING state.

e) If a CBIND response is returned with one of the following statuses, the PLOGI shall be terminated with an LS_RJT message. Depending on the CBIND failure status, the Reason Code and Reason Explanation SHALL be set to the following values specified in [FC-FS].

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)

f) A CBIND response with a CBIND STATUS of "N_PORT 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. The receiving gateway SHALL terminate this attempt, return the connection to the Unbound state and prepare to respond to an incoming CBIND request as described below.

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 Port Name in the incoming CBIND payload with the Destination Port Name.

b) If the Source 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 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, 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.

6.2.2.2 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 NAPT box or firewall, that detects and purges connections it believes to be idle.

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 7.3. 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 7.1. If both gateways wish 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 connectivity probe 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 aborted as described in section 6.2.3.2 if:

a) The contents of the LTEST message are incorrect

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.

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 9.2.1. The propagation delay limit SHALL NOT be enforced however.

~~1.1.1.3~~6.2.2.3 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.

~~1.1.1.1.1~~6.2.2.3.1 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. For that reason and the considerations described in [RFC1122], keep alive at the transport layer should not be implemented.

~~1.1.1.1.2~~6.2.2.3.2 '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. Hence, the transmission delay of the Nagle algorithm may decrease I/O throughput. The Nagle algorithm should therefore not be used.

~~1.1.1.1.3~~6.2.2.3.3 Window Scale

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

~~1.1.1.1.4~~6.2.2.3.4 Wrapped Sequence Protection (PAWS)

TCP segments are identified with 32-bit sequence numbers. In networks with large bandwidth - delay products, it is ~~therefore~~ 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].

~~1.1.3~~6.2.3 Terminating an N_PORT Login Session

An N_PORT login session SHALL be terminated or aborted in response to one of the following events:

a) An LS_RJT response is returned to the gateway that issued the PLOGI ELS. The gateway SHALL forward the LS_RJT to the local N_PORT and complete the session as described in section 6.2.3.1.

b) An ACC received from a remote device in response to a LOGO. The gateway SHALL forward the ACC to the local N_PORT and complete the session as described in section 6.2.3.1.

c) For an FC frame received from the IP network, a gateway detects a CRC error in the encapsulation header. The gateway shall abort the session as described in section 6.2.3.2.

d) The TCP connection associated with the login session fails for any reason. The gateway detecting the failed connection shall abort the session as described in section 6.2.3.2.

The disposition of the associated TCP connection is described in sections 6.2.3.1 and 6.2.3.2

~~1.1.1.1~~6.2.3.1 N_PORT Login Session Completion

An N_PORT login session is completed in response to a rejected PLOGI request as described in section 6.2.3 or a successful LOGO ELS.

The gateway receiving one of the above responses shall issue an Unbind session control ELS as described in section 7.2.

In response to the Unbind message, either gateway may choose to close the TCP connection or return it to a pool of unbound connections.

~~1.1.1.2~~6.2.3.2 Aborting an N_PORT Login Session

An N_PORT login session SHALL be aborted if the TCP connection is spontaneously terminated or whenever one of the following occurs:

a) An encapsulation error is detected as described in section 6.4.3.

b) The gateway receives an encapsulated frame from a gateway operating in an incompatible address translation mode as specified in section 5.5.3.3 or 5.5.1.2.

In any event, the TCP connection SHOULD be terminated with a connection reset (RST). If the local N_PORT has logged in to the remote N_PORT, the gateway SHALL send a LOGO to the local N_PORT.

~~1.3~~6.3 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.

~~1.4~~6.4 Encapsulation of Fibre Channel Frames

This section describes the iFCP encapsulation of Fibre Channel frames. The encapsulation is based on the common encapsulation format defined in [ENCAP].

The format of an encapsulated frame is shown below:

```
              +--------------------+
              |       Header       |
              +--------------------+-----+
              |        SOF         |   f |
              +--------------------+ F r |
              |  FC frame content  | C a |
              +--------------------+   m |
              |        EOF         |   e |
              +--------------------+-----+
```

Figure 15~~11~~ -- 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. When operating in Address Translation mode, (see section 5.5.2) the iFCP gateway must recalculate the Fibre Channel CRC.

~~1.1.1~~6.4.1 Encapsulation Header Format

```
W|------------------------------Bit------------------------------|
o|                                                               |
r|3 3 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1                    |
d|1 0 9 8 7 6 5 4 3 2 1 0 9 8 7 6 5 4 3 2 1 0 9 8 7 6 5 4 3 2 1 0|
 +---------------+---------------+---------------+---------------+
0|   Protocol#   |    Version    |  -Protocol#   |   -Version    |
 +---------------+---------------+---------------+---------------+
1|                  Reserved (must be zero)                      |
 +---------------+---------------+---------------+---------------+
2| LS_COMMAND    |  iFCP Flags   |     SOF       |      EOF      |
 +-----------+---+---------------+-----------+---+---------------+
3|   Flags   |   Frame Length    |   -Flags  |   -Frame Length   |
 +-----------+-------------------+-----------+-------------------+
4|                      Time Stamp [integer]                     |
 +---------------------------------------------------------------+
5|                      Time Stamp [fraction]                    |
 +---------------------------------------------------------------+
6|                              CRC                              |
 +---------------------------------------------------------------+
```

Common Encapsulation Fields:

Protocol#	IANA-assigned protocol number identifying the protocol using the encapsulation. For iFCP the value is (/TBD/).
Version	Encapsulation version
-Protocol#	Ones complement of the protocol#
-Version	Ones complement of the version
Flags	Encapsulation flags (see 6.4.1.1)
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 in SNTP format [RFC2030].
Time Stamp [fraction]	Fractional component of the time stamp in SNTP format [RFC2030].
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 9.2.1.

iFCP-specific fields:

LS_COMMAND For a special link service ACC response to be processed by iFCP, the LS_COMMAND field SHALL contain bits 31 through 24 of the LS_COMMAND to which the ACC applies. Otherwise the LS_COMMAND field shall be set to zero.

iFCP Flags iFCP-specific flags (see below)

SOF Copy of the SOF delimiter encoding (see section 6.4.2)

EOF Copy of the EOF delimiter encoding (see section 6.4.2)

The iFCP flags word has the following format:

```
|------------------------Bit----------------------------|
|                                                       |
|  23     22     21     20     19     18     17    16   |
+------+------+------+------+------+------+------+------+
|             Reserved             | SES  | TRN  |  SPC |
+------+------+------+------+------+------+------+------+
```

Figure 16~~12~~ -- iFCP Flags Word

iFCP Flags:

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

TRN 1 = Address transparent mode enabled

0 = Address translation mode enabled

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

~~1.1.1.1~~6.4.1.1 Common Encapsulation Flags

The iFCP usage of the common encapsulation flags is shown below:

```
      |-----------------------Bit--------------------------|
      |                                                    |
      |   31       30       29       28       27       26  |
      +--------------------------------------------+--------+
      |                  Reserved                  |  CRCV  |
      +--------------------------------------------+--------+
```

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

~~1.1.2~~6.4.2 SOF and EOF Delimiter Fields

The format of the delimiter fields is shown below.

```
W|------------------------------Bit------------------------------|
o|                                                               |
r|3 3 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1                    |
d|1 0 9 8 7 6 5 4 3 2 1 0 9 8 7 6 5 4 3 2 1 0 9 8 7 6 5 4 3 2 1 0|
 +---------------+---------------+-------------------------------+
0|      SOF      |      SOF      |     -SOF      |     -SOF      |
 +---------------+---------------+-------------------------------+
1|                                                               |
 +-----                   FC frame content                  -----+
 |                                                               |
 +---------------+---------------+-------------------------------+
n|      EOF      |      EOF      |     -EOF      |     -EOF      |
 +---------------+---------------+-------------------------------+
```

Figure 17~~13~~ -- FC Frame Encapsulation Format

SOF (bits 31-24 and bits 23-16 in word 0): iFCP uses the following subset of the SOF fields described in [ENCAP].

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

Table 2-- Translation of FC SOF Values to SOF Field Contents

-SOF (bits 15-8 and 7-0 in word 0): The -SOF fields contain the ones complement of the value in the SOF fields.

EOF (bits 31-24 and 23-16 in word n): iFCP uses the following subset of EOF fields specified in [ENCAP].

FC EOF	EOF Code
EOFn	0x41
EOFt	0x42

Table 3 -- Translation of FC EOF Values to EOF Field Contents

-EOF (bits 15-8 and 7-0 in word n): The -EOF fields contain the one's complement of the value in the EOF fields.

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

~~1.1.3~~6.4.3 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.

Frames types submitted for encapsulation and forwarding on the IP network SHALL have one of the SOF delimiters in Table 2~~Table 2~~ and an EOF delimiter from Table 3~~Table 3~~. Other valid frame types MUST be processed internally by the gateway as specified in the appropriate Fibre Channel specification.

Prior to submitting a frame for encapsulation, a gateway in address translation mode SHALL replace the D_ID address, and, if processing a special link service message~~ELS~~ requiring the inclusion of supplemental data, SHALL format the frame payload and add the supplemental information as specified in section 8.1. The gateway SHALL then calculate a new FC CRC on the reformatted frame.

A gateway in address transparent mode 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 above.

~~1.1.4~~6.4.4 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 invalid, the gateway SHALL terminate the N_PORT login session as described in section 6.2.3.2.

After validating the header CRC, the receiving gateway SHALL verify the frame propagation delay as described in section 9.2.1. 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 2~~Table 2~~ or an EOF code that is not in Table 3~~Table 3~~.

The gateway shall then de-encapsulate the frame. If operating in address translation mode, the gateway SHALL:

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

b) Replace the S_ID with the N_PORT alias of the frame originator

c) If processing a special link service message~~ELS~~, replace the ~~ELS~~ frame with a copy whose payload has been modified as specified in section 8.1.

The de-encapsulated frame SHALL then be delivered 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.

7. TCP Session Control Messages

TCP session control messages are used to create and manage an iFCP session as described in section 6.2.2. 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.

```
Word
  31<Bits>24 23<---------------Bits------------------------->0
 +------------+------------------------------------------------+
0| R_CTL      |            D_ID [0x00 00 00]                   |
 |[Req = 0x22]| [Destination of extended link Service request] |
 |[Rep = 0x23]|                                                |
 +------------+------------------------------------------------+
1| CS_CTL     |            S_ID [0x00 00 00]                   |
 | [0x0]      | [Source of extended link service request]      |
 +------------+------------------------------------------------+
2|TYPE [0x1]  |               F_CTL [0]                        |
 +------------+------------------+-----------------------------+
3|SEQ_ID      | DF_CTL [0x00]    |          SEQ_CNT [0x00]     |
 |[0x0]       |                  |                             |
 +------------+------------------+-----------------------------+
4|         OX_ID [0x0000]        |          RX_ID_[0x0000]     |
 +-------------------------------+-----------------------------+
5|                           Parameter                         |
 |                         [ 00 00 00 00 ]                     |
 +-------------------------------------------------------------+
6|                        LS_COMMAND                           |
 |                [Session Control Command Code]               |
 +-------------------------------------------------------------+
7|                                                             |
.|             Additional Session Control Parameters           |
.|                      ( if any )                             |
n|                                                             |
 +=============================================================+
n|                    Fibre Channel CRC                        |
+|                                                             |
1+=============================================================+
```

Figure 18~~14~~ -- Format of Session Control Message

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

The session control ELS frame is terminated with a Fibre Channel CRC. The frame SHALL be encapsulated and de-encapsulated according to the rules specified in section 6.4.

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

Encapsulation Header Fields:

LS_COMMAND	0
iFCP Flags	SES = 1
	TRN = 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.

The following lists the session control messages and their corresponding LS_COMMAND values.

Request	LS_COMMAND	Short Name	iFCP Support
Connection Bind	0xE0	CBIND	REQUIRED
Unbind Connection	0xE4	UNBIND	REQUIRED
Test Connection Liveness	0xE5	LTEST	Required

~~1.1~~7.1 Connection Bind (CBIND)

As described in section 6.2.2.1, the CBIND message and response are used to bind an N_PORT login session to a specific TCP connection and establish an iFCP session. In the CBIND request message, the source and destination N_Ports are identified by the N_PORT network address (iFCP portal address and N_PORT ID). 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.

```
+------+------------+------------+-----------+----------+
| Word |   Byte 0   |   Byte 1   |   Byte 2  |  Byte 3  |
+------+------------+------------+-----------+----------+
| 0    | Cmd = 0xE0 |   0x00     |   0x00    |  0x00    |
+------+------------+------------+-----------+----------+
| 1    |  LIVENESS TEST INTERVAL | Addr Mode | iFCP Ver |
|      |        (Seconds)        |           |          |
+------+-------------------------+-----------+----------+
| 2    |                  USER INFO                      |
+------+------------+------------+-----------+----------+
| 3    |                                                 |
+------+             SOURCE N_PORT NAME                  |
| 4    |                                                 |
+------+-------------------------------------------------+
| 5    |                                                 |
+------+             DESTINATION N_PORT NAME             |
| 6    |                                                 |
+------+-------------------------------------------------+
```

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.

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 World Wide Port Name (WWPN) of the N_PORT locally attached to the gateway originating the CBIND request.

DESTINATION N_PORT NAME: The World Wide 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.

```
    +------+------------+------------+-----------+----------+
    | Word |   Byte 0   |   Byte 1   |   Byte 2  |  Byte 3  |
    +------+------------+------------+-----------+----------+
    | 0    | Cmd = 0xE0 |   0x00     |   0x00    |  0x00    |
    +------+------------+------------+-----------+----------+
    | 1    |  LIVENESS TEST INTERVAL | Addr Mode | iFCP Ver |
    |      |      (Seconds)          |           |          |
    +------+-------------------------+-----------+----------+
    | 2    |                  USER INFO                      |
    +------+------------+------------+-----------+----------+
    | 3    |                                                 |
    +------+               SOURCE N_PORT NAME                |
    | 4    |                                                 |
    +------+-------------------------------------------------+
    | 5    |                                                 |
    +------+              DESTINATION N_PORT NAME            |
    | 6    |                                                 |
    +------+-------------------------+-----------------------+
    | 7    |        Reserved         |     CBIND Status      |
    +------+-------------------------+-----------------------+
    | 8    |        Reserved         |  CONNECTION HANDLE    |
    +------+-------------------------+-----------------------+
                              Total Length = 32
```

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.

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

SOURCE N_PORT NAME: Contains the World Wide Port Name (WWPN) of the N_PORT locally attached to the gateway issuing the CBIND request.

DESTINATION N_PORT NAME: Contains the World Wide 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 issuing the UNBIND request.

CBIND Status	Description
0	Successful – No other status
1 – 15	Reserved
16	Failed – Unspecified Reason
17	Failed – No such device
18	Failed – N_PORT session already exists
19	Failed – Lack of resources
20	Failed - Incompatible address translation mode
21	Failed - Incorrect protocol version number
Others	Reserved

~~1.2~~7.2 Unbind Connection (UNBIND)

UNBIND is used to release a bound TCP connection and return it to the pool of unbound TCP connections. This message is transmitted in 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

~~1.3~~7.3 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 9.2.1 and may be used by the receiver to compute an estimate of propagation delay. However, the propagation delay limit SHALL NOT be enforced.

```
+------+-----------+-----------+----------+----------+
| Word |  Byte 0   |  Byte 1   |  Byte 2  |  Byte 3  |
+------+-----------+-----------+----------+----------+
| 0    | Cmd = 0xE5 |  0x00     |  0x00    | 0x00     |
+------+-----------+-----------+----------+----------+
| 1    | LIVENESS TEST INTERVAL |       Reserved       |
|      |      (Seconds)        |                      |
+------+-----------------------+----------------------+
| 2    |                  COUNT                       |
+------+-----------+-----------+----------+----------+
| 3    |                                              |
+------+           SOURCE N_PORT NAME                 |
| 4    |                                              |
+------+----------------------------------------------+
| 5    |                                              |
+------+           DESTINATION N_PORT NAME            |
| 6    |                                              |
+------+----------------------------------------------+
```

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.

8. 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 fabric. 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 is processed according to one of the following rules:

a) Transparent – The link service message and reply MUST be transported to the receiving N_PORT by the iFCP gateway without altering the message payload. The link service message and reply are not processed by the iFCP implementation.

b) Special - Applies to a link service reply or request requiring iFCP intervention 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 implementation. The gateway SHALL respond to a rejected link service message by returning 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.

~~1.1~~8.1 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 which requires special handling, such as a PLOGI.

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

Such messages are transmitted in a Fibre Channel frame having the format shown in Figure 19~~Figure 15~~ for extended link services or Figure 21~~Figure 16~~ for FC-4 link services.:

```
    Word
      31        24 23                                             0
     +------------+------------------------------------------------+
    0| R_CTL      |                     D_ID                       |
     |[Req = 0x22]|[Destination of extended link Service request]  |
     |[Rep = 0x23]|                                               |
     +------------+------------------------------------------------+
    1| CS_CTL     |                     S_ID                       |
     |            | [Source of extended link service request]      |
     +------------+------------------------------------------------+
    2| TYPE       |                     F_CTL                      |
     | [0x01]     |                                                |
     +------------+------------------+-----------------------------+
    3| SEQ_ID     |        DF_CTL    |          SEQ_CNT            |
     +------------+------------------+-----------------------------+
    4|         OX_ID                 |             RX_ID           |
     +-------------------------------+-----------------------------+
    5|                         Parameter                           |
     |                      [ 00 00 00 00 ]                        |
     +-------------------------------------------------------------+
    6|                         LS_COMMAND                          |
     |               [Extended Link Service Command Code]          |
     +-------------==----------------------------------------------+
    7|                                                             |
    .|             Additional Service Request Parameters           |
    .|                      ( if any )                             |
    n|                                                             |
     +-------------------------------------------------------------+
```

Figure 19~~15~~ -- Format of an Extended Link Service Frame

```
    Word
      31        24 23                                              0
     +------------+------------------------------------------------+
    0| R_CTL      |                    D_ID                        |
     |[Req = 0x32]|   [Destination of FC-4 link Service request]   |
     |[Rep = 0x33]|                                                |
     +------------+------------------------------------------------+
    1| CS_CTL     |                    S_ID                        |
     |            |    [Source of FC-4 link service request]       |
     +------------+------------------------------------------------+
    2| TYPE       |                    F_CTL                       |
     | (FC-4      |                                                |
     |  specific) |                                                |
     +------------+------------------+-----------------------------+
    3| SEQ_ID     |        DF_CTL    |          SEQ_CNT            |
     +------------+------------------+-----------------------------+
    4|         OX_ID                 |             RX_ID           |
     +-------------------------------+-----------------------------+
    5|                        Parameter                            |
     |                     [ 00 00 00 00 ]                         |
     +-------------------------------------------------------------+
    6|                        LS_COMMAND                           |
     |               [FC-4 Link Service Command Code]              |
     +-------------------------------------------------------------+
    7|                                                             |
    .|             Additional Service Request Parameters           |
    .|                      ( if any )                             |
    n|                                                             |
     +-------------------------------------------------------------+
```

Figure 21~~16~~ -- Format of an FC-4 Link Service Frame

~~1.2~~8.2 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. In addition, 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 is 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 I/D 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 world wide 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 I/D of 0, the above codes will never correspond to valid N_PORT fabric IDs.

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 world-wide identifier to a gateway IP address and N_PORT ID. This information MUST be obtained through a name server query. If the N_PORT is locally attached, the gateway MUST fill in the field with the N_PORT ID. If the N_PORT is remotely attached, the gateway MUST assign and fill in the field with an N_PORT alias. If an N_PORT alias has already been assigned, it MUST be reused.

In the event that the sending gateway cannot obtain the world wide identifier of an N_PORT, or a receiving gateway cannot obtain the IP address and N_PORT ID, the gateway detecting the error 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.

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.

When the ACC response requires iFCP intervention, the receiving gateway MUST act as a proxy for the originator, retaining the state needed to process the response from the N_PORT to which the request was directed.

~~1.3~~8.3 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 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
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	REC
FCP Read Exchange Concise Accept	0x02 00 00 00	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 6.4.1.

Supplemental data (if any) MUST be appended as described in the following section.

The formats of each special link service message, including supplemental data where applicable, are 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.

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
0	LS_COMMAND			
1 . . n	Link Service Frame Payload			
n+1 . . n+k	Supplemental Data (if any)			

Figure 23~~17~~ -- Special Link Service Frame Payload

~~1.1.1~~8.3.1 Special Extended Link Services

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

~~1.1.1.1~~8.3.1.1 Abort Exchange (ABTX)

ELS Format:

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
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	Supplemental Data (type 3 only)

	section 8.2)	
Exchange Originator S_ID	1, 2	N/A

Other Special Processing:

None

~~1.1.1.2~~8.3.1.2 Discover Address (ADISC)

Format of ADISC ELS:

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
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 I/D of ELS Originator		

Fields Requiring Address Translation	Translation Type (see section 8.2)	Supplemental Data (type 3 only)
N_PORT I/D of ELS Originator	1	N/A

Other Special Processing:

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

~~1.1.1.3~~8.3.1.3 Discover Address Accept (ADISC ACC)

Format of ADISC ACC ELS:

Word	Bits 31–24	Bits 23–16	Bits 15–8	Bits 7-0
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 I/D of ELS Originator		

Fields Requiring Address Translation	Translation Type (see section 8.2)	Supplemental Data (type 3 only)
N_PORT I/D of ELS Originator	1	N/A

Other Special Processing:

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

~~1.1.1.4~~8.3.1.4 FC Address Resolution Protocol Reply (FARP-REPLY)

The FARP-REPLY ELS is used in conjunction with the FARP-REQ ELS (see section 8.3.1.5) 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 31–24	Bits 23–16	Bits 15–8	Bits 7-0
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 8.2)	Supplemental Data (type 3 only)
Requesting N_PORT Identifier	2	N/A
Responding N_PORT identifier	1	N/A

Other Special Processing:

None.

~~1.1.1.5~~8.3.1.5 FC Address Resolution Protocol Request (FARP-REQ)

The FARP-REQ ELS is used to 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 10.4 describes the iFCP implementation of FC broadcast server functionality in an iFCP fabric.

Format of FARP_REQ ELS:

Word	Bits 31–24	Bits 23–16	Bits 15–8	Bits 7-0
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 8.2)	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.

~~1.1.1.6~~8.3.1.6 Logout (LOGO)

ELS Format:

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
0	Cmd = 0x5	0x00	0x00	0x00
1	Rsvd	N_PORT I/D being logged out		
2-3	Port name of the LOGO originator (8 bytes)			

This ELS shall always be sent as an augmented ELS regardless of the translation mode in effect.

Fields Requiring Address Translation	Translation Type(see section 8.2)	Supplemental Data (type 3 only)
N_PORT I/D Being Logged Out	1	N/A

Other Special Processing:

See section 6.2.3.1.

~~1.1.1.7~~8.3.1.7 Port Login (PLOGI) and PLOGI ACC

PLOGI provides the mechanism for establishing a login session between two N_PORTs. In iFCP, a PLOGI request addressed to a remotely attached N_PORT may trigger the creation of an iFCP session, if one does not already exist. Otherwise, the PLOGI and PLOGI ACC payloads MUST be passed transparently to the destination N_PORT.

The PLOGI request and ACC response carry information identifying the originating N_PORT, including specification of its capabilities and limitations. If the destination N_PORT accepts the login request, it sends an accept (an ACC frame with PLOGI payload), specifying its capabilities and limitations. 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 request and accept (ACC) response. A port will reject a PLOGI request by transmitting an LS_RJT message, which contains no payload.

```
Byte
Offset
       +-----------------------------------+
   0   |             LS_COMMAND            |       4 Bytes
       +-----------------------------------+
   4   |     COMMON SERVICE PARAMETERS     |      16 Bytes
       +-----------------------------------+
  20   |             PORT NAME             |       8 Bytes
       +-----------------------------------+
  28   |             NODE NAME             |       8 Bytes
       +-----------------------------------+
  36   |     CLASS 1 SERVICE PARAMETERS    |      16 Bytes
       +-----------------------------------+
  52   |     CLASS 2 SERVICE PARAMETERS    |      16 Bytes
       +-----------------------------------+
  68   |     CLASS 3 SERVICE PARAMETERS    |      16 Bytes
       +-----------------------------------+
  86   |     CLASS 4 SERVICE PARAMETERS    |      16 Bytes
       +-----------------------------------+
 102   |        VENDOR VERSION LEVEL       |      16 Bytes
       +-----------------------------------+
                Total Length = 116 bytes
```

Figure 24~~18~~ -- Format of PLOGI Request and ACC Payloads

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

~~1.1.1.8~~8.3.1.8 Read Exchange Status Block (RES)

ELS Format:

```
+------+------------+------------+-----------+----------+
| Word | Bits 31-24 | Bits 23-16 | Bits 15-8 | Bits 7-0 |
+------+------------+------------+-----------+----------+
| 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 8.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1, 2 or 3	Port Name of the Exchange Originator

Other Special Processing:

None.

~~1.1.1.9~~8.3.1.9 Read Exchange Status Block Accept

Format of ELS Accept Response:

Word	Bits 31–24	Bits 23–16	Bits 15–8	Bits 7-0
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 [FCS]			
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 8.2)	Supplemental Data (type 3 only)
Exchange Originator N_PORT I/D	1, 2 or 3	Port Name of the Exchange Originator
Exchange Responder N_PORT I/D	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 I/D or the Exchange Responder N_PORT I/D is 1 or 2, the corresponding 8-byte port name SHALL be set to all zeros.

Other Special Processing:

None.

~~1.1.1.10~~8.3.1.10 Read Link Error Status (RLS)

ELS Format:

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
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 8.2)	Supplemental Data (type 3 only)
N_PORT Identifier	1, 2 or 3	Port Name of the N_PORT

Other Special Processing:

None.

~~1.1.1.11~~8.3.1.11 Read Sequence Status Block (RSS)

ELS Format:

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
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 8.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1, 2 or 3	Port Name of the Exchange Originator

Other Special Processing:

None.

~~1.1.1.12~~8.3.1.12 Reinstate Recovery Qualifier (RRQ)

ELS Format:

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
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 8.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1 or 2	N/A

Other Special Processing:

None.

~~1.1.1.13~~8.3.1.13 Request Sequence Initiative (RSI)

ELS Format:

Word	Bits 31–24	Bits 23–16	Bits 15–8	Bits 7-0
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 8.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1 or 2	N/A

Other Special Processing:

None.

~~1.1.1.14~~8.3.1.14 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 31–24	Bits 23–16	Bits 15 - 0
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 25~~19~~ -- 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, common to all FC-4 protocol types between the specified image pair, or global to all specified image pairs. The format of aTPRLO page requiring address translation is shown in Figure 26~~Figure 20~~. Additional information on TPRLO can be found in [FC-FS].

Word	Bits 31–24	Bits 23–16	Bits 15-8	Bits 7-0
0	TYPE Code or Common SVC Parameters	TYPE CODE EXTENSION	TPRLO Flags	
1	Third Party Process Associator			
2	Responder Process Associator			
3	Reserved	Third Party Originator N_PORT ID		
4-5	World Wide Name of Third Party Originator N_PORT			

Figure 26~~20~~ -- Format of an Augmented TPRLO Parameter Page

The TPRLO flags that affect the processing of the supplementedELS are as follows:

Bit 12: Global Process logout. When set to one, this bit indicates that all image pairs for all N_PORTs of the specified ~~FC4~~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.

Bit 13: Third party Originator N_PORT Validity. When set to one, this bit indicates that word 3, bits 23-00 (Third

Party Originator N_PORT ID) are meaningful.

If bit 13 has a value of zero and bit 12 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 26~~Figure 20~~.

c) If word 0, bit 13 (Third Party Originator N_PORT I/D validity) in the TPRLO flags field has a value of one, then the sender shall place the world-wide port name of the fibre channel device's N_PORT in the extension words. The N_PORT I/D 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 6.4.1).

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 13 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 I/D in word 3 shall be generated by referencing the supplemental data as described in section 8.2.

b) If word 0, bit 13 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.

~~1.1.1.15~~8.3.1.15 Third Party Logout Accept (TPRLO ACC)

The format of the TPRLO ACC frame is shown in Figure 28~~Figure 21~~.

Word	Bits 31–24	Bits 23–16	Bits 15 - 0
0	Cmd = 0x2	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 28~~21~~ -- Format of TPRLO ACC ELS

The format of the parameter page and rules for parameter page augmentation are as specified in section 8.3.1.14.

~~1.1.2~~8.3.2 Special FC-4 Link Services

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

~~1.1.1.1~~8.3.2.1 FC-4 Link Services defined by FCP

8.3.2.1.1 Read Exchange Concise (REC)

Link Service Request Format:

Word	Bits 31–24	Bits 23–16	Bits 15–8	Bits 7-0
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 8.2)	Supplemental Data (type 3 only)
Exchange Originator S_ID	1, 2 or 3	Port Name of the Exchange Originator

Other Special Processing:

None.

~~1.1.1.1.2~~8.3.2.1.2 Read Exchange Concise Accept (REC ACC)

Format of REC ACC Response:

Word	Bits 31–24	Bits 23–16	Bits 15–8	Bits 7-0
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	Data Transfer Count			
5	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 8.2)	Supplemental Data (type 3 only)
Exchange Originator N_PORT I/D	1, 2 or 3	Port Name of the Exchange Originator
Exchange Responder N_PORT I/D	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 Exchange Originator N_PORT I/D or the Exchange Responder N_PORT I/D is 1 or 2, the corresponding 8-byte port name SHALL be set to all zeros.

Other Special Processing:

None.

~~1.4~~8.4 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 8.3.1.7. The figure in that section is duplicated below for convenience.

```
    Byte
    Offset
          +-----------------------------------+
      0   |            LS_COMMAND             |      4 Bytes
          +-----------------------------------+
      4   |     COMMON SERVICE PARAMETERS     |     16 Bytes
          +-----------------------------------+
     20   |            PORT NAME              |      8 Bytes
          +-----------------------------------+
     28   |            NODE NAME              |      8 Bytes
          +-----------------------------------+
     36   |     CLASS 1 SERVICE PARAMETERS    |     16 Bytes
          +-----------------------------------+
     52   |     CLASS 2 SERVICE PARAMETERS    |     16 Bytes
          +-----------------------------------+
     68   |     CLASS 3 SERVICE PARAMETERS    |     16 Bytes
          +-----------------------------------+
     86   |     CLASS 4 SERVICE PARAMETERS    |     16 Bytes
          +-----------------------------------+
    102   |        VENDOR VERSION LEVEL       |     16 Bytes
          +-----------------------------------+
```

Figure 30~~22~~ -- FLOGI Request and ACC Payload Format

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

This section tabulates the protocol-dependant 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 4~~Table 4~~.

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 4 -- 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.

9. iFCP Error Detection

9.1 Overview

[FC-FS] defines error detection and recovery procedures. These Fibre Channel-defined mechanisms continue to be available in the iFCP environment.

~~1.2~~9.2 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 9.2.1.

~~1.1.1~~9.2.1 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 6.4.1) 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 insure 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 Unsynchronized state, the gateway SHALL:

a) Set the time stamp field to 0,0 for all outgoing frames

b) Ignore the time stamp field for all incoming frames.

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 6.4.4.

c) If the incoming frame has a time stamp of 0,0, the receiving gateway SHALL NOT test the frame to determine if it is stale.

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 6.4.4.

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

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 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.)

In response to loss of synchronization, a gateway enforcing R_A_TOV limits as described in this section should abort all N_PORT login sessions as described in section 6.2.3.2.

10. 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	10.1
0xFF-FF-FD	Fabric Controller	10.2
0xFF-FF-FC	Directory/Name Server	10.3

In addition, an iFCP gateway MAY support the FC broadcast server functionality described in section 10.4.

~~1.1~~10.1 F_PORT Server

The F_PORT server SHALL support the FLOGI ELS as described in section 8.4 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).

10.2 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).

10.3 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, name server requests are translated to the iSNS queries 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].

~~1.4~~10.4 Broadcast Server

Fibre Channel frames are broadcast throughout the fabric by addressing them to the Fibre Channel broadcast server at 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 comprised of 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 network.

b) A global broadcast server which 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 I/D 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.

~~1.1.1~~10.4.1 Establishing the Broadcast Configuration

The broadcast configuration is managed using facilities provided by the iSNS server. Specifically:

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 world-wide port name of the N_PORT pre-configured to provide global broadcast services.

b) If the world-wide 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 world-wide port name and N_PORT network address of the global server.

~~1.1.2~~10.4.2 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 6.2.2.2.

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.

11. iFCP Security

11.1 Overview

iFCP relies upon the IPSec protocol suite to provide data confidentiality and authentication services and IKE as the key management protocol. Section 11.2 describes the security requirements arising from iFCP's operating environment while Section 11.3 describes the resulting design choices, their requirement levels, and how they apply to the iFCP protocol.

~~1.2~~11.2 iFCP Security Operating Requirements

11.2.1 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 need to traverse security gateways in order to support SAN-to-SAN connectivity across public networks.

~~1.1.2~~11.2.2 Security Threats

Communicating iFCP gateways are vulnerable to attacks. Examples of attacks include 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 security negotiation process.

g) Impersonate a legitimate security gateway.

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 switches. To do so, the iFCP protocol MUST define confidentiality, authentication, integrity, and replay protection on a per-datagram basis. It also MUST define a scalable approach to key management. Conformant implementations of the iFCP protocol MAY use such definitions.

~~1.1.3~~11.2.3 Interoperability Requirements 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 must be able to leverage the protective services of the existing security infrastructure, including firewall protection, NAT and NAPT services, and IPSec VPN services available on existing security gateways.

~~1.1.4~~11.2.4 Statically and Dynamically Assigned IP Addresses

As iFCP gateways and switches are deployed within enterprise networks, it is expected that, like most routers and switches, gateway IP addresses will be statically assigned. Consequently, IKE and IPSec features focused on supporting DHCP and other dynamic IP address assignment capabilities for mobile hosts are not strictly required. Since the iFCP protocol cannot rule out the use of dynamically assigned IP addresses however, the security definitions for the iFCP protocol shall not exhibit any vulnerability in the case of dynamically assigned IP addresses (e.g., via DHCP [RFC2131]).

~~1.1.5~~11.2.5 Authentication Requirements

iFCP is a peer-to-peer protocol. iFCP sessions may be initiated by either or both peer gateways. Consequently, bi-directional authentication of peer gateways MUST be provided.

Fibre Channel, operating system and user identities are transparent to the iFCP protocol. IKE and IPSec authentication used to protect iFCP traffic shall be based upon the IP addresses of the communicating peer gateways.

iFCP gateways shall use Discovery Domain information obtained from the iSNS server [ISNS] to determine whether the initiating Fibre Channel N_PORT should be allowed access to the target N_PORT. N_PORT identities used in the Port Login (PLOGI) process shall be considered authenticated provided 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.

~~1.1.6~~11.2.6 Confidentiality Requirements

iFCP traffic may traverse insecure public networks, and therefore implementations MUST have per-packet encryption capabilities to provide confidentiality.

~~1.1.7~~11.2.7 Rekeying Requirements

Due to the high data transfer rates and the amount of data involved, an iFCP gateway implementation MUST support the capability to rekey each phase 2 security association in time intervals as often as every 25 seconds. The iFCP gateway MUST provide the capability for forward secrecy in the rekeying process.

~~1.1.8~~11.2.8 Usage Requirements

It must be possible for compliant iFCP implementations to administratively disable any and all security mechanisms. It must also be possible to apply different security requirements to individual N_PORT login session. Implementations may elect to expose such fine level of control through a management interface or through interaction with the iSNS.

~~1.1.9~~11.2.9 iSNS Role

iSNS [ISNS] is an invariant in all iFCP deployments. iFCP gateways 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 to secure communication between an iFCP gateway and iSNS server(s). Additionally, such specification indicates how elements of security policy concerning individual iFCP sessions can be retrieved from iSNS server(s).

~~1.3~~11.3 iFCP Security Design

11.3.1 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 the following guidelines.

Support for the IP Encapsulating Security Payload (ESP) [RFC2406] is MANDATORY to implement. As stated in [RFC2406], the following authentication algorithms MUST be implemented:

a) HMAC with SHA1 [RFC2404]

b) NULL authentication

The Advanced Encryption Standard [AES] in CBC MAC mode with Extended Cipher Block Chaining [XCBC] SHOULD be implemented.

The following encryption algorithms MUST be implemented:

a) NULL encryption [RFC2410]

b) 3DES in CBC mode [RFC2451]

AES counter mode encryption [AESCTR] SHOULD be implemented.

Implementation of DES in CBC mode [RFC2405] is OPTIONAL. It is recommended that DES in CBC mode SHOULD NOT be used due to its inherent weakness. It is in fact well known that DES is crackable with modest computation resources, and so is inappropriate for use in any iFCP deployment scenario requiring levels of security.

A conformant iFCP protocol implementation MUST implement IPsec ESP [RFC2406] in tunnel mode [RFC2401]. If minimizing the size of IPsec headers is a concern, transport mode should be supported. It shall be noted that transport mode continues to have a MUST implement requirement in those host scenarios where [RFC2401] makes it a MUST (see Sections 3.3 and 4.1 of [RFC2401]).

Regarding key management, iFCP implementations MUST support IKE [RFC2409] for peer authentication, negotiation of security associations, and key management, using the IPsec DOI. 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 11.3.3 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.

Phase-1 Main Mode SHOULD NOT be used in conjunction with pre-shared keys, due to Main Mode’s vulnerability to men-in-the-middle-attackers when group pre-shared keys are used. iFCP therefore requires that Aggressive Mode MUST be implemented as a valid alternative to Main Mode.

Peer authentication using the public key encryption methods outlined in sections 5.2 and 5.3 of [RFC2409] SHOULD NOT be used.

In all Phase 1 Modes, iFCP MUST use IP addresses as identities.

The Phase 2 Quick Mode exchanges used to negotiate protection for the TCP connections used by iFCP MUST explicitly carry the Identity Payload fields (IDci and IDcr). The DOI [RFC2407] provides for several types of identification data. However, when used in conformant iFCP security implementations, each ID Payload MUST carry a single IP address and a single non-zero TCP port number, and MUST NOT use the IP Subnet or IP Address Range formats. This allows the Phase 2 security association to correspond to specific TCP and iFCP connections.

~~1.1.2~~11.3.2 Use of IKE and IPsec

Each IP address supporting iFCP communication shall be capable of establishing one or more Phase-1 IKE Security Associations (SA) to other IP addresses configured as peer iFCP gateways, using the IP address as the identity. Such a security association may be established at a gateway’s initialization time, or may be deferred until the first TCP connection with security requirements is established.

Unlike Phase-1 SAs, a Phase-2 SA maps to an individual TCP connection. It protects the setup process of the underlying TCP connection and all its subsequent TCP traffic. TCP connections protected by the phase 2 SA are either in the unbound state, or are bound to a specific N_PORT login session. The creation of an IKE Phase-2 SA may be triggered by a policy rule supplied through a management interface, or by N_PORT properties registered with the iSNS server. Similarly, the use of Key Exchange payload in Quick Mode for perfect forward secrecy may be dictated through a management interface or by N_PORT properties registered with the iSNS server. This specification allows multiple implementation strategies, in which the establishment of an IKE Phase-2 SA occurs at different times. Examples of implementation strategies include:

a) The definition of a unique security policy for all TCP connections regardless of their bound or unbound state. Thus, an unbound TCP connection can be bound to an N_PORT login session without the need to incur a new IKE Phase-2 SA.

b) Multiple security policies for unbound TCP connections and active N_PORT login sessions. In this case, an unbound TCP connection becomes bound to an N_PORT login session after establishing a new IKE Phase-2 SA matching the new security policy for that N_PORT session.

c) The implementation does not support unbound connections. In this case, a new IKE Phase-2 SA and TCP connection must be started from scratch anytime a new N_PORT login session is created.

If the implementation does use unbound TCP connections, then an IKE Phase-2 SA MUST protect each of such unbound connections.

As expected, the successful establishment of a IKE Phase-2 SA results in the creation of two uni-directional IPsec SAs fully qualified by the tuple <SPI, destination address, ESP>.

Should a TCP connection be torn down (as opposed to joining a pool of unbound connections), the associated Phase-2 SA SHALL be terminated upon expiration of the TIME WAIT timeout value (according to [RFC793]).

Upon receiving a Phase 1 delete message, an iFCP implementation SHALL tear down all the Phase 2 SAs spawned from that Phase 1 SA, followed by the Phase 1 SA itself. Upon receiving a Phase 2 delete message, iFCP implementations will behave according to the state of the TCP connection protected by the SA in question. If the TCP session was terminated (either via FINs or RSTs), then a Phase 2 delete message SHALL terminate the IPsec SAs and any state formerly associated with that Phase 2 SA. If, however, the TCP session is maintained, then a Phase 2 delete message shall trigger a new Quick Mode exchange. To minimize the use of SA resources while the TCP connection is idle, the creation of the security association may be deferred until data is sent over the connection.

~~1.1.3~~11.3.3 Signatures and Certificate-based authentication

Conformant iFCP implementations MAY support peer authentication via digital signatures and X.509 certificates. When X.509 certificate authentication is chosen within IKE, each iFCP gateway needs the certificate credentials of each peering 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) of the iFCP gateway used for the iFCP session, or 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 SA's protecting iFCP sessions.

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

a) 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, then 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. The solutions to this problem are to send the end-entry machine certificate rather than the chain, to reduce the size of the certificate chain, to use IKE implementations over a reliable transport protocol (e.g., TCP) assisted by Path MTU discovery and code against black-holing as in [RFC2923], or to install 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.

~~1.4~~11.4 iSNS and iFCP Security

iFCP is required to use iSNS for discovery and management services. Consequently, the security of the iSNS protocol has an impact on the security of iFCP gateways. In particular, the following threats exist:

a) An attacker could alter iSNS protocol messages, so as to direct iFCP gateways to establish connections with rogue peer devices, or to weaken/eliminate IPSec protection for iFCP traffic.

b) An attacker could masquerade as the real iSNS server using false iSNS heartbeat messages. This could cause iFCP gateways to use rogue iSNS servers.

c) An attacker could gain knowledge about iFCP gateways by snooping iSNS protocol messages. Such information could aid an attacker in mounting a direct attack on iFCP gateways, such as a denial-of-service attack or outright physical theft.

To address these threats, the following capabilities are required:

a) Unicast iSNS protocol messages need to have both confidentiality and authentication support.

b) Multicast iSNS protocol messages such as the iSNS heartbeat message need to have authentication support.

There is no requirement that the communicating identities in iSNS protocol messages be kept confidential. Specifically, the identity and location of the iSNS server shall not be considered confidential.

However, in order to protect against an attacker masquerading as the real iSNS server, the iSNS server MUST have the capability to allow client gateways to authenticate broadcast or multicast messages such as the iSNS heartbeat. The iSNS authentication block (which is identical in format to the SLP authentication block) may be used for this purpose. Note that the authentication block is used only for iSNS broadcast or multicast messages, and SHOULD NOT be used in unicast iSNS messages.

For protecting unicast iSNS protocol messages, iSNS servers MUST support the ESP protocol in tunnel mode for iFCP client gateways.

~~1.5~~11.5 Use of iSNS to Distribute Security Policy

Once communication between iFCP gateways and the iSNS server have been secured through use of IPSec, the iFCP gateways have the capability to discover the security settings that they need to use to protect iFCP traffic. This provides a potential scaling advantage over device-by-device configuration of individual security policies for each iFCP gateway.

The iSNS server stores security settings for each iFCP gateway. These security settings include use or non-use of IPSec, IKE, Main Mode, Aggressive Mode, PFS, Pre-shared Key, and certificates. These settings can be retrieved by peer iFCP gateways, who can then take the appropriate action. For example, IKE may not be enabled for a particular iFCP gateway. If a peer gateway can learn of this in advance by consulting the iSNS server, it will not need to waste time and resources attempting to initiate an IKE session with that iFCP gateway.

Additionally, the iSNS server can store policies that are used for ISAKMP phase 1 and phase 2 negotiations between iFCP gateways. The ISAKMP payload format includes a series of one or more proposals that the iFCP gateway will use when negotiating the appropriate IPSec policy to use to protect iFCP traffic.

~~1.6~~11.6 Minimal Security Policy for an iFCP gateway

An iFCP implementation MAY be able to administratively disable security mechanisms for individual N_PORT login sessions. This implies that IKE and IPsec security associations may not be established for one or more of such sessions. A configuration of this type may be accomplished through a management interface or through attributes set in the iSNS server.

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 a minimal assurance that connections have initially been opened with the intended counterpart. The minimal iFCP security policy thus only states that an iFCP gateway SHOULD authenticate its iSNS server(s) as described in [ISNS].

12. Quality of Service Considerations

12.1 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 will be the gap measured between iFCP observed behaviors and baseline iFCP behaviors (i.e., as produced by two iFCP gateways directly connected to one another).

~~1.2~~12.2 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, 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

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, some of which 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 its 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 outside of the IETF scope.

Due to the fact that Fibre Channel Class 2 and Class 3 do not currently support fractional bandwidth guarantees, and that iFCP is committed to supporting Fibre Channel semantics, it is impossible for an iFCP gateway to autonomously infer bandwidth requirements 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 admission control and bandwidth brokers (via RSVP or other signalling 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 N_PORT login session , the traffic in an individual N_PORT to N_PORT session can be singled out by iFCP-unaware network equipment as well.

To render 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 SLSs of this nature by applying EF DSCPs to the iFCP traffic.

13. Author's Addresses

Charles Monia
Rod Mullendore
Josh Tseng

Nishan Systems
3850 North First Street
San Jose, CA 95134
Phone: 408-519-3986
Email:
cmonia@nishansystems.com

Franco Travostino
Director, Content
Internetworking Lab,
Nortel Networks
3 Federal Street
Billerica, MA 01821
Phone: 978-288-7708
Email:
travos@nortelnetworks.com

David Robinson
Sun Microsystems
Senior Staff Engineer
M/S UNWK16-301
901 San Antonio Road
Palo Alto, CA 94303-4900
Phone: 510-936-2337
Email:
David.Robinson@sun.com

Wayland Jeong
Troika Networks
Vice President, Hardware
Engineering
2829 Townsgate Road Suite
200
Westlake Village, CA 91361
Phone: 805-370-2614
Email:
wayland@troikanetworks.com

Rory Bolt
Quantum/ATL
Director, System Design
101 Innovation Drive
Irvine, CA 92612
Phone: 949-856-7760
Email: rbolt@atlp.com

Mark Edwards
Senior Systems Architect
Eurologic Development, Ltd.
4th Floor, Howard House
Queens Ave, UK. BS8 1SD
Phone: +44 (0)117 930 9600
Email:
medwards@eurologic.com

14. References

14.1 Normative

[RFC2026] Bradner, S., "The Internet Standards Process -- Revision 3", BCP 9, RFC 2026, October 1996.

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

[FC-FS] dpANS X3.XXX-200X, "Fibre Channel Framing and Signaling Interface", Revision 1.5, NCITS Project 1331-D, February 2001

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

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

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

[ENCAP] Weber, et-al., "FC Frame Encapsulation", draft-ietf-ips-fcencapsulation-01.txt, May 2001

[ISNS] Tseng, J., et-al., "iSNS Internet Storage Name Service", draft-ietf-ips-04.txt, July 2001

[RFC791] Postel, J., RFC 791, "The Internet Protocol", September 1981

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

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

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

[RFC2406] Kent, S., Atkinson, R., RFC 2406, "Encapsulating Security Protocol", November 1998

[RFC2407] Piper, D., RFC 2407, " The Internet IP Security Domain of Interpretation for ISAKMP", November 1998

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

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

[RFC2410] Glenn, R., Kent, S., "The NULL Encryption Algorithm and Its use with IPSEC", RFC 2410, November 1998

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

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

[RFC2410] Glenn, R., Kent, S., "The NULL Encryption Algorithm and Its use with IPSEC", RFC 2410, November 1998

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

~~1.2~~14.2 Non-Normative

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

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

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

[RFC896] Nagel, J., "Congestion Control in IP/TCP Networks", RFC 896, January 1984

[RFC2625] Rajagopal, M., et-al., RFC 2625, "IP and ARP over Fibre Channel", June 1999

[RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, March 1997

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

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

[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

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

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

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

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

[RFC1323] Jacobsen, V., et-al., "TCP Extensions for High Performance", RFC 1323, May, 1992

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

[XCBC] Black, J., Rogaway, P., "A Suggestion for Handling Arbitrary Length Messages with the CBC MAC". Available from http://csrc.nist.gov/encryption/modes/proposedmodes/xcbc-mac/xcbc-mac-spec.pdf

[AESCTR] Lipmaa, H., Rogaway, P., Wagner, D., "CTR-Mode Encryption", 2001. Available from http://csrc.nist.gov/encryption/modes/proposedmodes/ctr/ctr-spec.pdf

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

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

[MPSLDS] F. Faucheur, L. Wu, B. Davie, S. Davari, P. Vaananen, R. Krishnan, P. Cheval, J. Heinanen, "MPLS Support of Differentiated Services", draft-ietf-mpls-diff-ext-09.txt, April 2001.

[DIFTERM] Grossman, D., "New Terminology and Clarifications for Diffserv", draft-ietf-diffserv-new-terms-07.txt, December 2001

Appendix A

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 8.

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 Link Services Processed Transparently

The following link service requests and responses MUST be processed transparently as defined in section 8.

Link Services Processed Transparently

Name	Description
ACC	Accept
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
SCL	Scan Remote Loop
SCN	State Change Notification
SCR	State Change Registration
TEST	Test
TPLS	Test Process Login State

A.3 iFCP-Processed Link Services

The following extended and FC-4 link services are processed by the iFCP implementation as described in the referenced section listed in the table.

Special Link Services

Name	Description	Section
ABTX	Abort Exchange	8.3.1.1
ADISC	Discover Address	8.3.1.2
ADISC ACC	Discover Address Accept	8.3.1.3

FARP-REPLY	Fibre Channel Address Resolution Protocol Reply	8.3.1.4
FARP-REQ	Fibre Channel Address Resolution Protocol Request	8.3.1.5
LOGO	N_PORT Logout	8.3.1.6
PLOGI	Port Login	8.3.1.7
FCP REC	FCP Read Exchange Concise	8.3.2.1.1
FCP REC ACC	FCP Read Exchange Concise Accept	8.3.2.1.2
RES	Read Exchange Status Block	8.3.1.8
RES ACC	Read Exchange Status Block Accept	8.3.1.9
RLS	Read Link Error Status Block	8.3.1.10
RRQ	Reinstate Recovery Qualifier	8.3.1.12
RSI	Request Sequence Initiative	8.3.1.13
RSS	Read Sequence Status Block	8.3.1.11
TPRLO	Third Party Process Logout	8.3.1.14
TPRLO ACC	Third Party Process Logout Accept	8.3.1.15

Full Copyright Statement