17 Distributed Switch Environment

17.1 Overview

A Distributed Switch is a set of FCDFs associated with at least one Controlling Switch, that controls the operations of the set of FCDFs. Figure 42 shows an example of Distributed Switch composed of a Controlling Switch and two FCDFs.

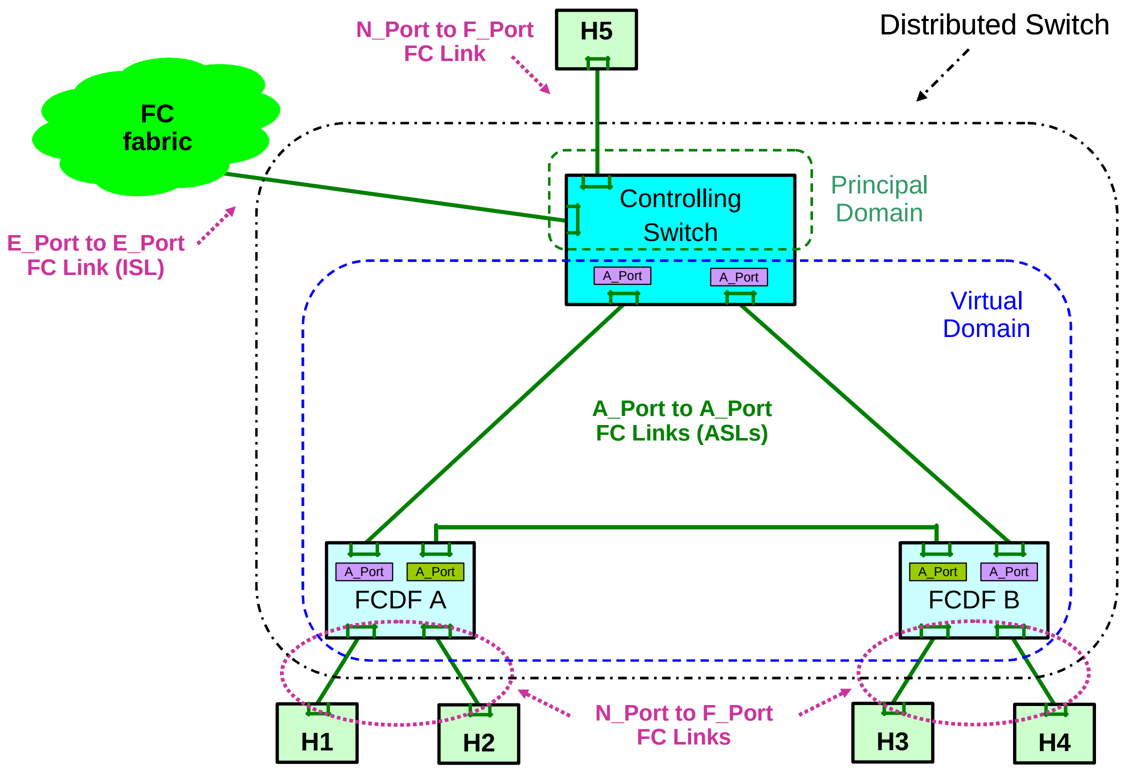

Figure 42 – Example of Distributed Switch

From an external point of view (i.e., outside the dotted and dashed black line in figure 42), a Distributed Switch behaves as a Fibre Channel Switch. In particular, a Distributed Switch supports the instantiation of N_Port to F_Port links and of E_Port to E_Port links (ISLs). N_Port to F_Port links are supported by both FCDFs and Controlling Switches, while ISLs are supported only by Controlling Switches. This means that it is possible to connect a Distributed Switch to another Switch only through a Controlling Switch, not through an FCDF.

From an internal point of view (i.e., inside the dotted and dashed black line in figure 42), A_Port to A_Port links (ASLs) enable FC frames forwarding between Controlling Switch and FCDFs, as well as between FCDFs. ASLs are also used to exchange control information between Controlling Switch and FCDFs.

The Controlling Switch uses one or more Virtual Domain_IDs to perform N_Port_ID allocations for N_Ports connected to the FCDF Set of the Distributed Switch (i.e., a Virtual Domain_ID is used as the most significant byte in the N_Port_IDs allocated to N_Ports that are attached to the FCDF Set). The Controlling Switch uses also another Domain_ID, called Principal Domain, for its normal functions as a Fibre Channel Switch. As a result, a Distributed Switch such as the one shown in figure 42 uses at least two Domain_IDs: one for the Principal Domain and one or more for the Virtual Domain.

To properly support the operations of a Virtual Domain, a Controlling Switch shall have at least one Switch_Name to associate with the Virtual Domain, in addition to its own Switch_Name.

FCDFs are not able to operate properly without a Controlling Switch, therefore the Controlling Switch is a single point of failure in a Distributed Switch configuration with only one Controlling Switch, as the one shown in figure 42. To avoid this issue, Distributed Switches may support a redundant configuration of two Controlling Switches, a Primary one and a Secondary one. The Secondary Controlling Switch keeps its state synchronized with the Primary and is able to take its place in case of failure according to the Controlling Switch Redundancy Protocol.

Figure 43 shows an example of Distributed Switch including a redundant pair of Controlling Switches.

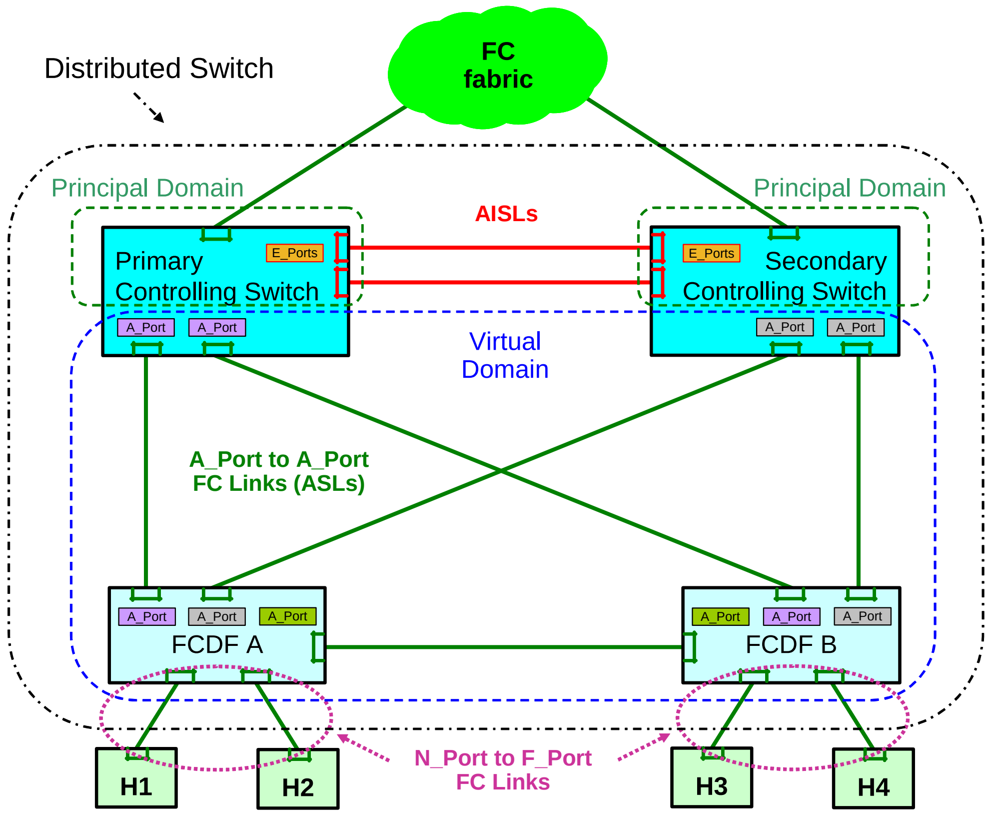

Figure 43 – Example of Redundant Distributed Switch

The two Controlling Switches in a redundant Distributed Switch instantiate at least two Augmented ISLs (AISLs) between themselves, where the term ‘augmented’ indicates that link is used also for the Redundancy protocol, in addition to normal E_Port operation.

The Controlling Switches use one or more Virtual Domain_IDs to perform N_Port_ID allocations for N_Ports connected to the FCDF Set of the Distributed Switch (i.e., a Virtual Domain_ID is used as the most significant byte in the N_Port_IDs allocated to N_Ports that are attached to the FCDF Set). Using Virtual Domain_IDs to assign N_Port_IDs enables seamless operation in case of failures of one of the two redundant Controlling Switches. Each Controlling Switch uses also another Domain_ID, called Principal Domain, for its normal functions as a Fibre Channel Switch. As a result, a redundant Distributed Switch typically uses three or more Domain_IDs: one for each Principal Domain and one or more for the Virtual Domain. To properly support the operations of a Virtual Domain, a Controlling Switch shall have at least a Switch_Name to associate with the Virtual Domain, in addition to its own Switch_Name.

The two redundant Controlling Switches instantiate ASLs to enable the forwarding of FC frames and the communication of control information between Controlling Switches and FCDFs. In a redundant configuration, FCDFs instantiate ASLs to each of the Controlling Switches and between themselves.

A Distributed Switch may have a cascaded FCDF configuration. Figure 44 shows an example of such a configuration.

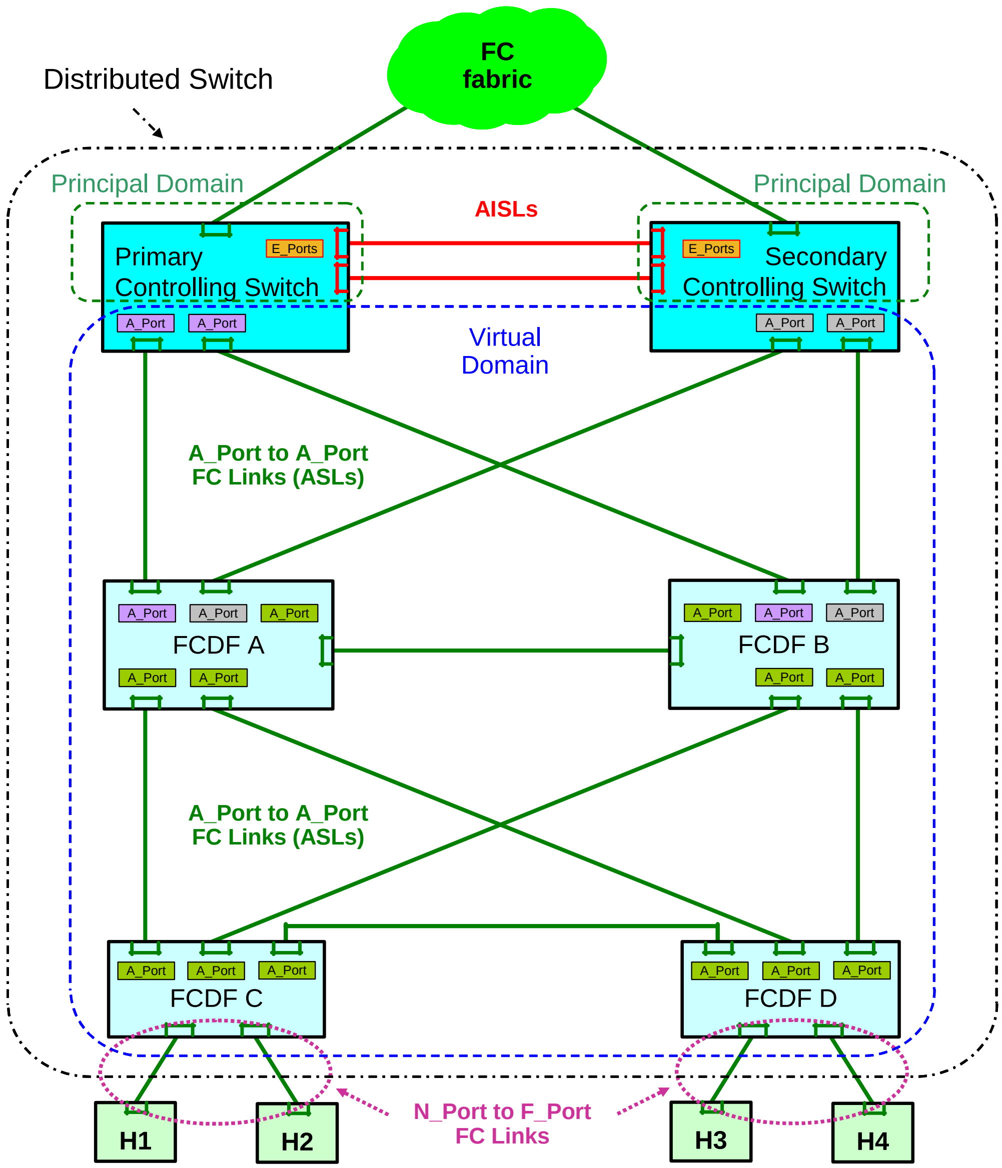

Figure 44 – Example of Distributed Switch with Cascaded FCDFs

A Controlling Switch is uniquely identified by its Switch_Name Name_Identifier, as an FC Switch. An FCDF is uniquely identified by its Switch_Name Name_Identifier. A Distributed Switch is defined by an administrative configuration on the Controlling Switches, listing:

a) the Switch_Names of the two Controlling Switches that act as the Primary/Secondary pair for that Distributed Switch (i.e., the Controlling Switch Set); and

b) the Switch_Names of the FCDFs that are part of that Distributed Switch (i.e., the FCDF Set).

For proper operation of a redundant Distributed Switch it is recommended for the FCDFs directly connected to at least one Controlling Switch to be directly connected to all Controlling Switches.

17.2 Controlling Switch Functional Model

Figure 45 shows the functional model of a Controlling Switch.

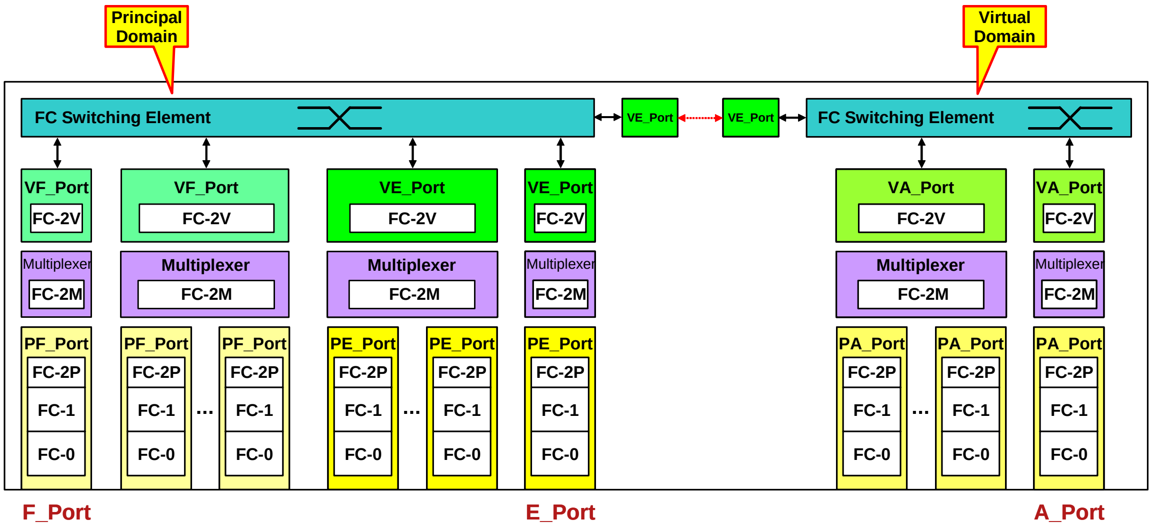

Figure 45 – Controlling Switch Functional Model

A Controlling Switch is an FC Switch that supports the instantiation of VA_Ports, in addition to VF_Ports and VE_Ports. As any FC Switch, a Controlling Switch is able to aggregate its physical ports in sets that behave as virtual ports, providing higher bandwidth than the one available to a single physical port.

For a Controlling Switch, a physical port is an LCF (see FC-FS-3), that may behave as a Physical F_Port (PF_Port), as a Physical E_Port (PE_Port), or as a Physical A_Port (PA_Port). A virtual port is an instance of the FC-2V sublevel of Fibre Channel (see FC-FS-3), that may behave as a Virtual F_Port (VF_Port), as a Virtual E_Port (VE_Port), or as a Virtual A_Port (VA_Port).

As shown in figure 45, a Controlling Switch is functionally modeled as having two FC Switching Elements, one for the Principal Domain and one for the Virtual Domain, connected by an internal VE_Port to VE_Port link. The Switching Element associated with the Principal Domain supports the intantiation of VF_Ports and VE_Ports, the Switching Element associated with the Virtual Domain supports the intantiation of VA_Ports.

17.3 FCDF Functional Model

Figure 46 shows the functional model of an FCDF.

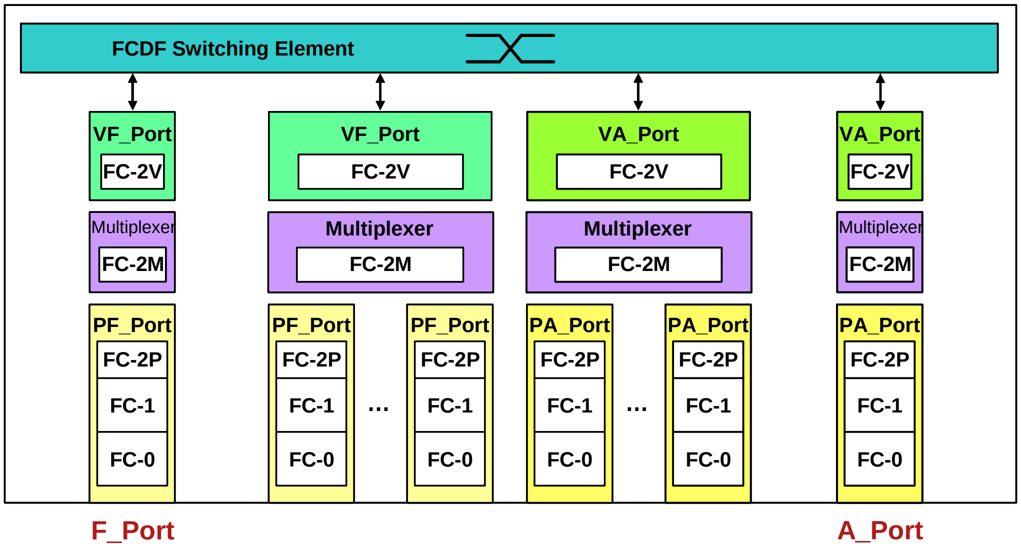

Figure 46 – FCDF Functional Model

An FCDF is a simplified FC switching entity that forwards FC frames among VA_Ports and VF_Ports through a FCDF Switching Element. As any FC Switch, an FCDF is able to aggregate its physical ports in sets that behave as virtual ports, providing higher bandwidth than the one available to a single physical port. An FCDF shall support at least one VA_Port operating together with a PA_Port (i.e., an A_Port) and may support one or more F_Ports.

For an FCDF, a physical port is an LCF (see FC-FS-3), that may behave as a Physical F_Port (PF_Port) or as a Physical A_Port (PA_Port). A virtual port is an instance of the FC-2V sublevel of Fibre Channel (see FC-FS-3), that may behave as a Virtual F_Port (VF_Port) or as a Virtual A_Port (VA_Port).

Figure 47 shows the model of the FCDF Switching Element, composed by a Switch Construct, a Routing Table Update function, and a FCDF Controller function.

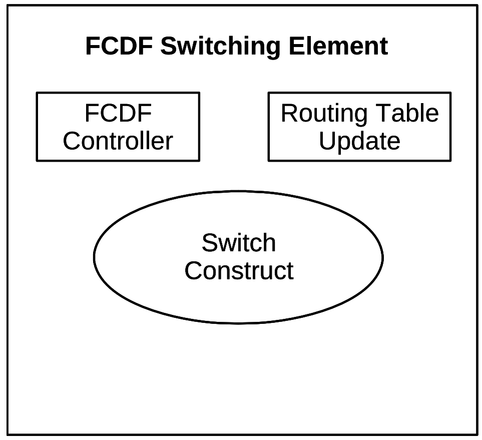

Figure 47 – FCDF Switching Element

The Switch Construct is the entity performing FC frames forwarding based on the FC frame's D_ID field according to a routing table. The structure of the Switch Construct is undefined and beyond the scope of this document.

The Routing Table Update is a logical entity that updates the Switch Construct's routing table through the VA_Port protocols.

The FCDF Controller is a logical entity that performs the management of the FCDF through the VA_Port protocols. The FCDF Controller has the characteristics of a VN_Port.

17.4 FCDF Handling of Well Known Addresses

N_Ports use Well Known Addresses (WKAs) and Domain Controller address identifiers to exchange information with the Fabric, either through ELSs or through the Common Transport protocol.

An FCDF supports VF_Ports, therefore it shall terminate FC frames destined to the F_Port Controller WKA. This implies local processing by the FCDF of the FLOGI, FDISC, LOGO, and RLS ELSs.

The handling of other WKAs and Domain Controllers address identifiers is performed by the Primary Controlling Switch, therefore an FCDF shall forward all FC frames having as D_ID the address identifiers listed in table 208 to the Primary Controlling Switch through a VA_Port. The NPRD SW_ILS provides to FCDFs the routing information needed to reach the Primary Controlling Switch.

Table 208 – Forwarded Domain Controller and Well Known Address Identifiers

Address Value	Description
FFFC01h .. FFFCFEh	Domain Controller Address Identifiers
FFFFF4h	Event Service WKA
FFFFF6h	Clock Synchronization Service WKA
FFFFF7h	Security Key Distribution Service WKA
FFFFFAh	Management Service WKA
FFFFFBh	Time Service WKA
FFFFFCh	Directory Service WKA
FFFFFDh	Fabric Controller WKA

The AISLs used for the redundancy protocol between the Primary and Secondary Controlling Switch are used as paths to reach the Primary Controlling Switch when an FCDF is connected to the Secondary Controlling Switch but not anymore to the Primary one. In order to do so, the Secondary Controlling Switch shall forward to the Primary Controlling Switch over the AISLs:

a) any FC frame destined to the address identifier FFFFF9h (i.e., the VA_Port Controller); and

b) any FC frame destined to the address identifiers shown in table 208 when they are received from a VA_Port.

17.5 A_Port Operation

An A_Port is the combination of one PA_Port and one VA_Port operating together. A PA_Port is the LCF within the Fabric that attaches to another PA_Port through a link. A VA_Port is an instance of the FC-2V sublevel of Fibre Channel that connects to another VA_Port. A VA_Port is uniquely identified by an A_Port_Name Name_Identifier and is addressable by the VA_Port connected to it through the A_Port Controller address identifier (i.e., FFFFF9h).

An A_Port is the point at which a Controlling Switch is connected to an FCDF to create a Distributed Switch. Also, an A_Port is the point at which an FCDF is connected to another FCDF. It normally functions as a conduit among FCDFs and between FCDFs and Controlling Switches for frames destined for remote N_Ports and NL_Ports. An A_Port is also used to carry frames between Controlling Switch and FCDFs for purposes of configuring and maintaining the Distributed Switch.

An A_Port shall support the Class F service. An A_Port shall also be capable of forwarding one or more of the following classes of service: Class 2 service, Class 3 service. An A_Port shall not admit to its FCDF or Controlling Switch any Primitive Sequences, or any Primitive Signals other than Idle, that the A_Port receives on its inbound fibre.

The model of an A_Port on an FC-FS-3 Transport is shown in figure 48.

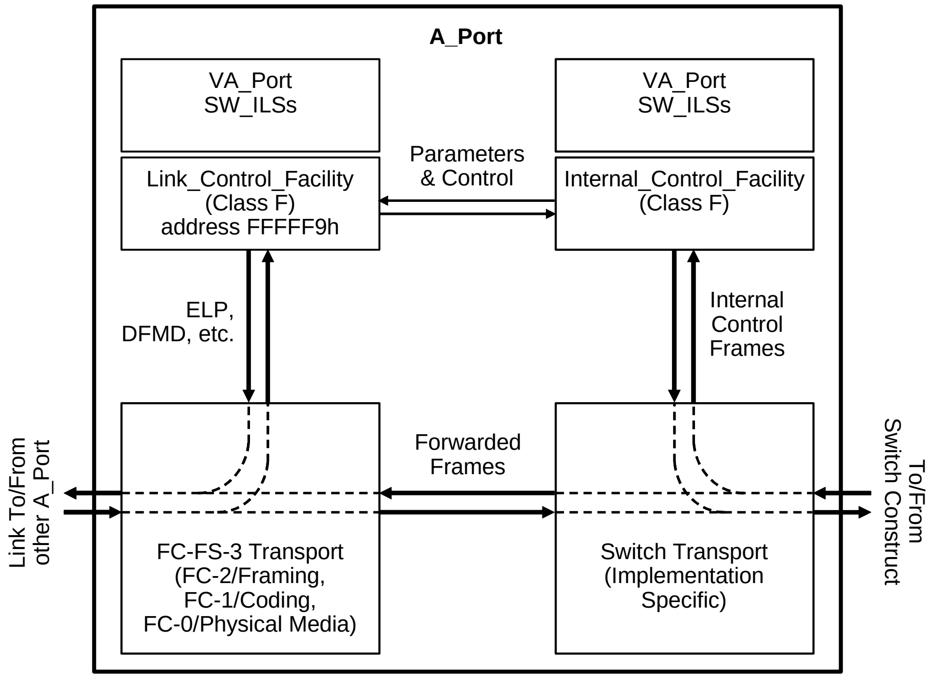

Figure 48 – A_Port Model

An A_Port contains an FC-FS-3 Transport element through which all frames are passed, and Primitives are transferred across the Link to and from the other A_Port. Frames received from the other A_Port are either directed to the Switch Construct via the Switch Transport element, or directed to the Link_Control_Facility. The Link_Control_Facility receives frames related to the VA_Port SW_ILSs, and transmits responses to those frames.

Frames received from the FC-FS-3 Transport element that are destined for other ports are directed by the Switch Transport to the Switch Construct for further forwarding. Frames received from the Switch Construct by the Switch Transport are directed either to the FC-FS-3 Transport for transmission to the other A_Port, or to the Internal_Control_Facility. The Internal_Control_Facility receives frames related to VA_Port SW_ILSs, and transmits responses to those frames.

Information is passed between the Internal_Control_Facility and the Link_Control_Facility to effect the control and configuration of the Transport elements.

17.6 A_Port to A_Port Links (ASLs)

An ASL becomes operational on successful completion of an ELP Exchange between a Controlling Switch and a FCDF or between two FCDFs. Two additional bits in the flags field of the ELP payload indicate if the originator of the ELP Request or SW_ACC is a Controlling Switch or an FCDF.

Bits 13 and 12 in the flags field of the ELP payload indicate if the originator of the ELP Request or SW_ACC is a Controlling Switch or an FCDF.Bits 13 and 12 in the flags field of the ELP payload indicate if the originator of the ELP Request or SW_ACC is a Controlling Switch or an FCDF (see 6.1.4).

A received ELP Request having both these bits set to one is invalid, shall be rejected, and the link shall be Isolated. A received SW_ACC having both these bits set to one is invalid and the link shall be Isolated. Table 209 shows the meaning of the values of these bits.

Table 209 – VA_Port ELP Flags

Bit 13 value	Bit 12 value	Description
0b	0b	The originator of the ELP Request or SW_ACC is a normal FC Switch or FCF
0b	1b	The originator of the ELP Request or SW_ACC is an FCDF or an FDF
1b	0b	The originator of the ELP Request or SW_ACC is a Controlling Switch or a Controlling FCF
1b	1b	Invalid combination

A port of a Controlling Switch shall transmit an ELP Request after completing Link Initialization. This ELP Request has the Controlling FCF/Switch flag set to one and the FDF/FCDF flag set to zero.

If the ELP is accepted by the neighbor and

a) the received ELP SW_ACC has both the Controlling FCF/Switch flag and the FDF/FCDF flag set to zero; or

b) the received ELP SW_ACC has the Controlling FCF/Switch flag set to one and the neighbor Switch is not the peer Controlling Switch of this Distributed Switch

then the Controlling Switch port behaves as an E_Port (i.e., an ISL is instantiated).

If the ELP is accepted and the received ELP SW_ACC has the Controlling FCF/Switch flag set to one and the neighbor Switch is the peer Controlling Switch of this Distributed Switch then the Controlling Switch port behaves as an Augmented E_Port for this Distributed Switch (i.e., an AISL is instantiated), used for the redundancy protocol of the Distributed Switch (see 17.8).

If the ELP is accepted and the received ELP SW_ACC has the FDF/FCDF flag set to one and the neighbor FCDF is part of this Distributed Switch FCDF Set, then the Controlling Switch port behaves as an A_Port (i.e., an ASL is instantiated) when the Controlling Switch is operational (i.e., when in state P2 or S2 of of the Controlling Switch Redundancy Protocol, see 17.8), otherwise (i.e., when the Controlling Switch is not yet operational) the Controlling Switch port shall transition to the Isolated state.

If the ELP is accepted and the received ELP SW_ACC has the FDF/FCDF flag set to one and the neighbor FCDF is not part of this Distributed Switch FCDF Set, then the Controlling Switch port shall go in Isolated state.

A port of a Controlling Switch shall reject a received ELP Request with the FDF/FCDF flag set to one with Reason Code ‘Protocol Error’ and Reason Code Explanation ‘Invalid Request’.

A port of a Controlling Switch shall reply to a received ELP Request with the FDF/FCDF flag set to zero according to the normal ELP rules (i.e., acceptance or rejection includes considering the involved Switch_Names). If the ELP Request is accepted and

a) the received ELP Request has both the Controlling FCF/Switch flag and the FDF/FCDF flag set to zero; or

b) the received ELP Request has the Controlling FCF/Switch flag set to one and the neighbor Switch is not the peer Controlling Switch of this Distributed Switch

then the Controlling Switch port behaves as an E_Port (i.e., an ISL is instantiated).

If the ELP is accepted and the received ELP Request has the Controlling FCF/Switch flag set to one and the neighbor Switch is the peer Controlling Switch of this Distributed Switch, then the Controlling Switch port behaves as an Augmented E_Port for this Distributed Switch (i.e., an AISL is instantiated), used for the Redundancy protocol of the Distributed Switch (see 17.8).

The ports of an FCDF that has not yet received from the Primary Controlling Switch the Distributed Switch’s FCDF Set through the DFMD SW_ILS shall wait to receive an ELP Request after completing Link Initialization.

After having received from the Primary Controlling Switch the Distributed Switch’s FCDF Set through the DFMD SW_ILS and routing information through the NPRD SW_ILS, the ports of an FCDF that have completed Link Initialization, except the one from which the DFMD Request has been received, shall transmit an ELP Request with the FDF/FCDF flag set to one.

On Receiving an ELP Request with the Controlling FCF/Switch flag set to one or the FDF/FCDF flag set to one, the FCDF port shall process it irrespective of the value of the Switch_Name field in the ELP Request payload (i.e., acceptance or rejection shall be based on the other ELP parameters, not on the involved Switch_Names). If the ELP is accepted then the FCDF Port behaves as an A_Port (i.e., an ASL is instantiated).

NOTE 48 – These rules enable an ordered establishments of ASLs from the Controlling Switch(es) to the peripheral FCDFs in a Distributed Switch with cascaded FCDFs.

An FCDF does not support E_Ports, therefore a port of an FCDF shall reject a received ELP Request with both Controlling FCF/Switch flag and FDF/FCDF flag set to zero (i.e., a ELP Request coming from a Switch that is not a Controlling Switch) with Reason Code ‘Protocol Error’ and Reason Code Explanation ‘Invalid Request’. After an FCDF has received the Distributed Switch's FCDF Set from the Primary Controlling Switch, that FCDF shall reject received ELP Requests coming from a Controlling Switch other than the Controlling Switches in the Controlling Switch set, with Reason Code ‘Logical Error’ and Reason Code Explanation ‘Not Authorized’.

Upon instantiating an ASL with another FCDF, an FCDF may receive a DFMD Request over that ASL. If the received DFMD Request contains a Virtual Domain Switch_Name different than its Virtual Domain Switch_Name, then the DFMD Request shall be rejected and the ASL Isolated.

17.7 VA_Port SW_ILSs

17.7.1 Overview

The VA_Port SW_ILSs are used to exchange information between Controlling Switches and FCDFs (i.e., they are not used to exchange information between FCDFs). If a Distributed Switch includes cascaded FCDFs, the intermediate FCDFs relay the SW_ILSs through a chain of Exchanges, as shown in figure 49. If one Exchange of this chain of Exchanges is abnormally terminated, then the other Exchanges in the chain shall be terminated as well.

.

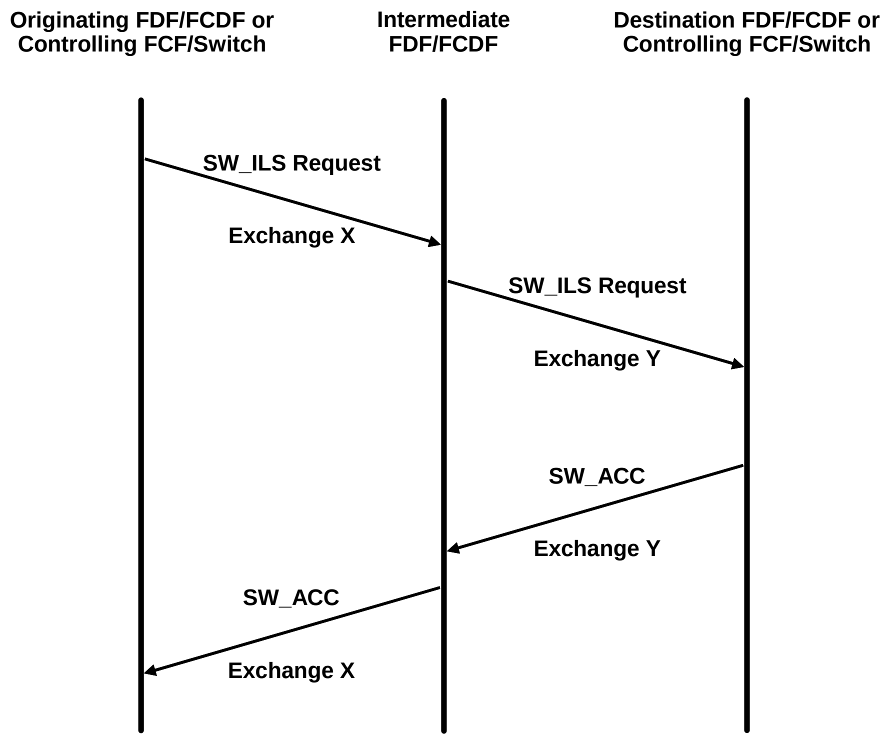

Figure 49 – VA_Port SW_ILS Relay

To enable this relay, all VA_Port SW_ILSs include the originating and destination FCDF or Controlling Switch Switch_Names in the first two fields of their payload. The subsequent part of a VA_Port SW_ILS is a list of self-identifying descriptors, as defined in 17.7.2. The descriptor list may be null.

The need for the originating and destination FCDF or Controlling Switch Switch_Names in the first two fields of the payload requires the definition of an updated SW_RJT, called here VA_RJT (see 17.7.3.1).

The VA_Port SW_ILSs have the same high-order byte in their command code, denoted here as XXh. Table 210 shows the VA_Port SW_ILSs command codes.

Table 210 – VA_Port SW_ILSs Command Codes

Encoded Value	Description	Abbreviation
A000 0001h	VN_Port Reachability Notification	VNRN
A000 0002h	VN_Port Unreachability Notification	VNUN
A000 0003h	FCDF Reachability Notification	FDRN
A000 0004h	FCDF Unreachability Notification	FDUN
A000 0005h	N_Port_ID Route Distribution	NPRD
A000 0006h	N_Port_ID and Zoning ACL Distribution	NPZD
A000 0007h	Active Zoning ACL Distribution	AZAD
A000 0008h	Distributed Switch Membership Distribution	DFMD

17.7.2 VA_Port SW_ILS Descriptors

17.7.2.1 Descriptor Format

Each VA_Port SW_ILS descriptor has the format shown in table 211. This format applies also to the descriptors for the Redundancy Protocol SW_ILSs (see 17.8).

Table 211 – Descriptor Format

Item	Size (Bytes)
Descriptor Tag	4
Descriptor Length	4
Descriptor Value	variable

Descriptor Tag: the two most significant bytes of this field are reserved, the two least significant bytes contain the tag value. The defined tag values are shown in table 212.

Table 212 – Descriptor Tags

Tag Value	Descriptor	Reference
0001h	VN_Port Reachability	17.7.2.2
0002h	FLOGI/NPIV FDISC Parameters	17.7.2.3
0003h	VN_Port Unreachability	17.7.2.4
0004h	FCDF Reachability	17.7.2.5
0005h	Sequence Number	17.7.2.6
0006h	Controlling Switch Reachability	17.7.2.7
0007h	N_Port_IDs Reachability	17.7.2.8
0008h	Domain_IDs Reachability	17.7.2.9
0009h	Allocation Status	17.7.2.10
000Ah	Peering Status	17.7.2.11
000Bh	Membership Set	17.7.2.12
000Ch	Integrity	17.7.2.13
000Dh	FCDF Identification	17.7.2.14
000Eh	Reject	17.7.2.15
0011h	Controlling Switch State	17.8.2.2
0012h	FCDF Topology	17.8.2.3
0013h	FCDF N_Port_IDs	17.8.2.4
0014h	RHello Interval	17.8.2.5
all others	Reserved	

Descriptor Length: contains the length in bytes of the Descriptor Value.

Descriptor Value: contains the specific information carried in the descriptor.

17.7.2.2 VN_Port Reachability Descriptor

The format of the VN_Port Reachability descriptor is shown in table 213.

Table 213 – VN_Port Reachability Descriptor Format

Item	Size (Bytes)
Tag Value = 0001h	4
Length = 12	4
F_Port_Name	8
Physical Port Number	4

F_Port_Name: contains the F_Port_Name of the VF_Port to which the newly reachable VN_Port is being associated.

Physical Port Number: contains the physical port number where an FLOGI or NPIV FDISC Request has been received.

17.7.2.3 FLOGI/NPIV FDISC Parameters Descriptor

The format of the FLOGI/NPIV FDISC Parameters descriptor is shown in table 214.

Table 214 – FLOGI/NPIV FDISC Parameters Descriptor Format

Item	Size (Bytes)
Tag Value = 0002h	4
Length = 116	4
FLOGI/NPIV FDISC Parameters	116

FLOGI/NPIV FDISC Parameters Descriptor: contains the payload of an FLOGI or NPIV FDISC (see FC-LS-2).

17.7.2.4 VN_Port Unreachability Descriptor

The format of the VN_Port Unreachability descriptor is shown in table 215.

Table 215 – VN_Port Unreachability Descriptor Format

Item	Size (Bytes)
Tag Value = 0003h	4
Length = 20	4
Flags	1
Unreachable N_Port_ID	3
Unreachable N_Port_Name	8
F_Port_Name	8

Flags: 8 flag bits. The following flag bits are defined:

Bit 8 .. 1: reserved.

Bit 0: indicates if only one VN_Port is unreachable or if all the VN_Ports associated to a VF_Port are unreachable. This flag is set to zero to indicate that only one VN_Port is unreachable and to one to indicate that all the VN_Ports associated to a VF_Port are unreachable.

Unreachable N_Port_ID: when bit 0 of the flag field is set to zero contains the N_Port_ID of the unreachable VN_Port. When bit 0 of the flag field is set to one contains 000000h.

Unreachable N_Port_Name: when bit 0 of the flag field is set to zero contains the N_Port_Name of the unreachable VN_Port. When bit 0 of the flag field is set to one contains 0000 0000 0000 0000h.

F_Port_Name: contains the F_Port_Name of the involved VF_Port.

17.7.2.5 FCDF Reachability Descriptor

The format of the FCDF Reachability descriptor is shown in table 216.

Table 216 – FCDF Reachability Descriptor Format

Item	Size (Bytes)
Tag Value = 0004h	4
Length = 28	4
FCDF or Controlling Switch Switch_Name	8
Local A_Port_Name	8
Adjacent A_Port_Name	8
Reserved	2
Link Cost	2

FCDF or Controlling Switch Switch_Name: contains the Switch_Name of the adjacent entity with which an ASL has been instantiated or deinstantiated.

Local A_Port_Name: contains the local A_Port_Name of the instantiated or deinstantiated ASL.

Adjacent A_Port_Name: contains the adjacent A_Port_Name of the instantiated or deinstantiated ASL.

Link Cost: contains the cost of the instantiated or deinstantiated ASL.

17.7.2.6 Sequence Number Descriptor

The format of the Sequence Number descriptor is shown in table 217.

Table 217 – Sequence Number Descriptor Format

Item	Size (Bytes)
Tag Value = 0005h	4
Length = 8	4
Sequence Number	8

Sequence Number: contains a monotonically increasing sequence number. When the sequence number reaches the value FFFFFFFF FFFFFFFFh it wraps to 00000000 00000000h.

17.7.2.7 Controlling Switch Reachability Descriptor

The format of the Controlling Switch Reachability descriptor is shown in table 218.

Table 218 – Controlling Switch Reachability Descriptor Format

Item	Size (Bytes)
Tag Value = 0006h	4
Length = variable	4
Controlling Switch Switch_Name	8
Number of Paths to the Controlling Switch (j)	4
Next-hop Switch_Name #1	8
Local A_Port_Name #1	8
Path #1 cost	4
Next-hop Switch_Name #2	8
Local A_Port_Name #2	8
Path #2 cost	4
...	
Next-hop Switch_Name #j	8
Local A_Port_Name #j	8
Path #j cost	4

Controlling Switch Switch_Name: contains the Switch_Name of the Controlling Switch.

Number of Paths to the Controlling Switch: contains the number of paths toward the Controlling Switch. Each path that follows is expressed by the Switch_Name of the next-hop FCDF or Controlling Switch followed by the local A_Port_Name of the involved ASL and by the path cost.

17.7.2.8 N_Port_IDs Reachability Descriptor

The format of the N_Port_IDs Reachability descriptor is shown in table 219.

Table 219 – N_Port_IDs Reachability Descriptor Format

Item	Size (Bytes)
Tag Value = 0007h	4
Length = variable	4
Number of N_Port_ID Reachability Entries (p)	4
N_Port_ID Reachability Entry #1	see table 220
N_Port_ID Reachability Entry #2	see table 220
…	
N_Port_ID Reachability Entry #p	see table 220

Number of N_Port_ID Reachability Entries: contains the number of N_Port_ID Reachability Entries that follow. There shall be an N_Port_ID Reachability Entry for each FCDF currently belonging the Distributed Switch. The N_Port_ID Reachability Entry format is shown in table 220.

Table 220 – N_Port_ID Reachability Entry Format

Item	Size (bytes)
Reachable FCDF Switch_Name	8
Number of Equal Cost Paths to the Reachable FCDF (w)	4
Next-hop Switch_Name #1	8
Local A_Port_Name #1	8
Next-hop Switch_Name #2	8
Local A_Port_Name #2	8
...	
Next-hop Switch_Name #w	8
Local A_Port_Name #w	8
Number of N_Port_ID Ranges (q)	4
N_Port_ID Range #1	4
N_Port_ID Range #2	4
…	
N_Port_ID Range #q	4

Reachable FCDF Switch_Name: contains the Switch_Name of the FCDF to which the subsequent next-hops and N_Port_ID Ranges refer.

Number of Equal Cost Paths to the Reachable FCDF: contains the number of equal cost paths having the lowest cost toward the destination FCDF. Each path that follows is expressed as the Switch_Name of the next-hop FCDF or Controlling Switch followed by the local A_Port_Name of the involved ASL. Only the first path listed shall be used to relay VA_Port SW_ILSs to the Reachable FCDF.

Number of N_Port_ID Ranges: contains the number of N_Port_ID Range Entries that follow. The N_Port_ID Range is defined by an N_Port_ID in the least significant three bytes, and by the number of bits defining the range in the most significant byte (e.g., the range 020200h .. 02027Fh is expressed as ‘7 || 020200h’). The set of N_Port_ID Range Entries encodes in a compact form all the N_Port_IDs currently allocated to VN_Ports logged into the reachable FCDF.

17.7.2.9 Domain_IDs Reachability Descriptor

The format of the Domain_IDs Reachability descriptor is shown in table 221.

Table 221 – Domain_IDs Reachability Descriptor Format

Item	Size (Bytes)
Tag Value = 0008h	4
Length = variable	4
Number of Reachable Domain_ID Entries (r)	4
Reachable Domain_ID Entry #1	see table 222
Reachable Domain_ID Entry #2	see table 222
…	
Reachable Domain_ID Entry #r	see table 222

Number of Reachable Domain_ID Entries: contains the number of Reachable Domain_ID Entries that follow. The Reachable Domain_ID Entry format is shown in table 222.

Table 222 – Reachable Domain_ID Entry Format

Item	Size (bytes)
Reachable Domain_ID	4
Number of Equal Cost Paths to the Reachable Domain_ID (y)	4
Next-hop Switch_Name #1	8
Local A_Port_Name #1	8
Next-hop Switch_Name #2	8
Local A_Port_Name #2	8
...	
Next-hop Switch_Name #y	8
Local A_Port_Name #y	8

Reachable Domain_ID: contains the reachable Domain_ID. The three most significant bytes of this field are reserved.

Number of Equal Cost Paths to the Reachable Domain_ID: contains the number of equal cost paths having the lowest cost toward the destination Domain_ID. Each path that follows is expressed as the Switch_Name of the next-hop FCDF or Controlling Switch followed by the local A_Port_Name of the involved ASL.

17.7.2.10 Allocation Status Descriptor

The format of the Allocation Status descriptor is shown in table 223.

Table 223 – Allocation Status Descriptor Format

Item	Size (Bytes)
Tag Value = 0009h	4
Length = variable	4
Number of Allocation / Deallocation Entries (z)	4
Allocation / Deallocation Entry #1	see table 224
Deallocation Entry #2	see table 224
...	
Deallocation Entry #z	see table 224

Number of Allocation / Deallocation Entries: contains the number of Allocation / Deallocation Entries that follow. Only one Allocation Entry may be present, multiple Deallocation Entries may be present. The Allocation / Deallocation Entry format is shown in table 224.

Table 224 – Allocation / Deallocation Entry Format

Item	Size (bytes)
Flags	4
Allocated / Deallocated N_Port_ID	4
N_Port_Name associated with the Allocated/Deallocated N_Port_ID	8
Switch_Name of the FCDF associated with the Allocated/Deallocated N_Port_ID	8
FLOGI / NPIV FDISC LS_ACC Parameters	116

Flags: 32 flag bits. The following flag bits are defined:

Bit 32 .. 2: reserved.

Bit 1: indicates if the FLOGI / NPIV FDISC LS_ACC Parameters field is present in the payload. The field is present when this flag is set to one and not present when this flag is set to zero. This flag shall not be set to one when bit 0 indicates deallocation (i.e., the FLOGI / NPIV FDISC LS_ACC Parameters field may be present only when an N_Port_ID allocation is performed).

Bit 0: indicates if the operation is an allocation or a deallocation. This flag is set to zero to indicate allocation and to one to indicate deallocation.

Allocated / Deallocated N_Port_ID: contains the N_Port_ID that the Primary Controlling Switch allocated or deallocated in the least significant three bytes. The most significant byte is reserved.

N_Port_Name associated with the Allocated/Deallocated N_Port_ID: contains the N_Port_Name of the VN_Port for which an N_Port_ID is allocated or deallocated.

Switch_Name of the FCDF associated with the Allocated/Deallocated N_Port_ID: contains the Switch_Name of the FCDF associated with the VN_Port for which an N_Port_ID is allocated or deallocated.

FLOGI / NPIV FDISC LS_ACC Parameters: this field is present when bit 1 of the flags field is set to one. It contains the payload of the LS_ACC generated by the Primary Controlling Switch in response to the FLOGI or NPIV FDISC payload provided in the VNRN Request Sequence.

17.7.2.11 Peering Status Descriptor

The format of the Peering Status descriptor is shown in table 225.

Table 225 – Peering Status Descriptor Format

Item	Size (Bytes)
Tag Value = 000Ah	4
Length = variable	4
Number of Peering Entries (h)	4
Peering Entry #1	see table 226
Peering Entry #2	see table 226
…	
Peering Entry #h	see table 226

Number of Peering Entries: contains the number of Peering Entries that follow. Each Peering Entry contains a complete list of the Peer N_Port_IDs with which the Principal N_Port_ID is allowed to communicate according to the current fabric zoning configuration. The Peering Entry format is shown in table 226.

Table 226 – Peering Entry Format

Item	Size (bytes)
Principal N_Port_ID	4
Number of Allowed Peers (k)	4
Peer N_Port_ID #1	4
Peer N_Port_ID #2	4
…	
Peer N_Port_ID #q	4

Principal N_Port_ID: contains the N_Port_ID to which the subsequent Peer N_Port_IDs refer.

Number of Allowed Peers: contains the number of N_Port_IDs to which the Principal N_Port_ID is allowed to communicate.

Peer N_Port_ID: contains an N_Port_ID in the least significant three bytes and the most significant byte is reserved.

17.7.2.12 Membership Set Descriptor

The format of the Membership Set descriptor is shown in table 227.

Table 227 – Membership Set Descriptor Format

Item	Size (Bytes)
Tag Value = 000Bh	4
Length = variable	4
Fabric_Name	8
Virtual Domain Switch_Name	8
Primary Controlling Switch Switch_Name	8
Secondary Controlling Switch Switch_Name	8
Number of FCDFs (n)	4
FCDF Switch_Name #1	8
FCDF Switch_Name #2	8
…	
FCDF Switch_Name #n	8

Fabric_Name: contains the Fabric_Name of the Distributed Switch's associated Fabric.

Virtual Domain Switch_Name: contains the Switch_Name of the Virtual Domain.

Primary Controlling Switch Switch_Name: contains the Switch_Name of the Primary Controlling Switch.

Secondary Controlling Switch Switch_Name: contains the Switch_Name of the Secondary Controlling Switch. This field shall be set to 00000000 00000000h when there is no Secondary Controlling Switch.

Number of FCDFs: contains the number of FCDF Switch_Names that follow. This list of FCDF Switch_Names is the FCDF Set of the Distributed Switch. If the number of FCDF Switch_Names is zero, then any FCDF is allowed in the Distributed Switch.

17.7.2.13 Integrity Descriptor

The format of the Integrity descriptor is shown in table 228.

Table 228 – Integrity Descriptor Format

Item	Size (Bytes)
Tag Value = 000Ch	4
Length = variable	4
Integrity Type	4
Integrity Check Value Length	4
Integrity Check Value	variable

Integrity Type: indicates, in the least significant byte, the type of cryptographic integrity that protects the payload. The defined values are:

00h: No integrity

01h: HMAC-SHA-256-128 integrity

02h .. FFh: Reserved

Integrity Check Value Length: contains the length expressed in bytes of the Integrity Check Value.

Integrity Check Value: contains the cryptographic hash of the payload computed using the shared key according to the specified Integrity Type.

17.7.2.14 FCDF Identification Descriptor

The format of the FCDF Identification descriptor is shown in table 229.

Table 229 – FCDF Identification Descriptor Format

Item	Size (Bytes)
Tag Value = 000Dh	4
Length = 36	4
Number of Physical Ports	4
RNID Specific Node-Identification Data	32

Number of Physical Ports: contains the number of physical ports that the FCDF has.

RNID Specific Node-Identification Data: see FC-SB-4.

17.7.2.15 Reject Descriptor

The format of the Reject descriptor is shown in table 230.

Table 230 – Reject Descriptor Format

Item	Size (Bytes)
Tag Value = 000Eh	4
Length = 4	4
Reserved	1
Reason Code	1
Reason Code Explanation	1
Vendor Specific	1

17.7.3 VA_Port SW_ILSs Definition

17.7.3.1 VA_RJT

The VA_RJT SW_ILS is used in place of an SW_RJT as a reply Sequence to a VA_Port SW_ILS Request to reject that request.

Addressing: The S_ID field shall be set to the value of the D_ID field in the SW_ILS request. The D_ID field shall be set to the value of the S_ID field in the SW_ILS request.

Payload: the format of the VA_RJT Payload is shown in table 231.

Table 231 – VA_RJT Payload

Item	Size (bytes)
SW_ILS Code = 0300 0000h	4
Destination Switch_Name	8
Originating Switch_Name	8
Descriptor List Length	4
Reject Descriptor	see 17.7.2.15

Destination Switch_Name: contains the Switch_Name of the destination entity.

Originating Switch_Name: contains the Switch_Name of the originating entity.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

Reject Descriptor: see 17.7.2.15

17.7.3.2 VN_Port Reachability Notification (VNRN)

The VN_Port Reachability Notification SW_ILS is used by an FCDF to communicate to the Primary Controlling Switch that a VN_Port is attempting Fabric login through an FLOGI Request or a NPIV FDISC Request. If the FCDF does not have an ASL with the Primary Controlling Switch, the VNRN SW_ILS is relayed to the Primary Controlling Switch by the intermediate FCDFs.

VNRN Request Sequence

Addressing: the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port.

Payload: the format of the VNRN Request Sequence Payload is shown in table 232.

Table 232 – VNRN Request Payload

Item	Size (bytes)
SW_ILS Code = A000 0001h	4
Destination Controlling Switch Switch_Name	8
Originating FCDF Switch_Name	8
Descriptor List Length	4
VN_Port Reachability Descriptor	see 17.7.2.2
FLOGI/NPIV FDISC Parameters Descriptor	see 17.7.2.3

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating FCDF Switch_Name: contains the Switch_Name of the originating FCDF.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

VN_Port Reachability Descriptor: see 17.7.2.2.

FLOGI/NPIV FDISC Parameters Descriptor: contains the payload of the received FLOGI or NPIV FDISC Request (see FC-LS-2).

VNRN Reply Sequence

VA_RJT: indicates the rejection of the VNRN Request Sequence. As a result, a FLOGI LS_RJT or a NPIV FDISC LS_RJT is sent as response to the FLOGI Request or NPIV FDISC Request that caused the issuance of the VNRN Request.

SW_ACC: indicates the acceptance of the VNRN Request Sequence. The format of the VNRN SW_ACC Payload is shown in table 233.

Table 233 – VNRN SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination FCDF Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
FLOGI / NPIV FDISC Parameters Descriptor	see 17.7.2.3

FLOGI / NPIV FDISC Parameters Descriptor: this descriptor contains the payload of the LS_ACC generated by the Primary Controlling Switch in response to the FLOGI or NPIV FDISC payload provided in the VNRN Request Sequence.

17.7.3.3 VN_Port Unreachability Notification (VNUN)

The VN_Port Unreachability Notification SW_ILS is used by an FCDF to communicate to the Primary Controlling Switch that one or more of its VN_Ports have been logged out. If the FCDF does not have an ASL with the Primary Controlling Switch, the VNUN SW_ILS is relayed to the Primary Controlling Switch by the intermediate FCDFs.

VNUN Request Sequence

Addressing: the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port.

Payload: the format of the VNUN Request Sequence Payload is shown in table 234.

Table 234 – VNUN Request Payload

Item	Size (bytes)
SW_ILS Code = A000 0002h	4
Destination Controlling Switch Switch_Name	8
Originating FCDF Switch_Name	8
Decriptor List Length	4
VN_Port Unreachability Descriptor	see 17.7.2.4

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating FCDF Switch_Name: contains the Switch_Name of the requesting FCDF.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

VN_Port Unreachability Descriptor: see 17.7.2.4**.**

VNUN Reply Sequence

VA_RJT: indicates the rejection of the VNUN Request Sequence.

SW_ACC: indicates the acceptance of the VNUN Request Sequence. The format of the VNUN SW_ACC Payload is shown in table 235.

Table 235 – VNUN SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination FCDF Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length = 0000 0000h	4

17.7.3.4 FCDF Reachability Notification (FDRN)

The FCDF Reachability Notification SW_ILS is used by an FCDF to communicate to the Primary Controlling Switch that it has instantiated an ASL with another FCDF or with the Secondary Controlling Switch. If the FCDF does not have an ASL with the Primary Controlling Switch, the FDRN SW_ILS is relayed to the Primary Controlling Switch by the intermediate FCDFs.

The FDRN SW_ILS is also used between Primary and Secondary Controlling Switch to keep their state synchronized.

FDRN Request Sequence

Addressing: when used between a FCDF and the Primary Controlling Switch the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port. When used between the two Controlling Switches the S_ID field shall be set to FFFFFDh, indicating the originating VE_Port, and the D_ID field shall be set to FFFFFDh, indicating the destination VE_Port.

Payload: the format of the FDRN Request Sequence Payload is shown in table 236.

Table 236 – FDRN Request Payload

Item	Size (bytes)
SW_ILS Code = A000 0003h	4
Destination Controlling Switch Switch_Name	8
Originating FCDF Switch_Name	8
Descriptor List Length	4
FCDF Reachability Descriptor	see 17.7.2.5

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating FCDF Switch_Name: contains the Switch_Name of the requesting FCDF.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

FCDF Reachability Descriptor: describes the instantiated ASL **(**see 17.7.2.5).

FDRN Reply Sequence

VA_RJT: indicates the rejection of the FDRN Request Sequence.

SW_ACC: indicates the acceptance of the FDRN Request Sequence. The format of the FDRN SW_ACC Payload is shown in table 237.

Table 237 – FDRN SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination FCDF Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length = 0000 0000h	4

17.7.3.5 FCDF Unreachability Notification (FDUN)

The FCDF Unreachability Notification SW_ILS is used by an FCDF to communicate to the Primary Controlling Switch that it has deinstantiated an ASL with another FCDF or with the Secondary Controlling Switch. If the FCDF does not have an ASL with the Primary Controlling Switch, the FDUN SW_ILS is relayed to the Primary Controlling Switch by the intermediate FCDFs.

The FDUN SW_ILS is also used between Primary and Secondary Controlling Switch to keep their state synchronized.

FDUN Request Sequence

Addressing: when used between a FCDF and the Primary Controlling Switch the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port. When used between the two Controlling Switches the S_ID field shall be set to FFFFFDh, indicating the originating VE_Port, and the D_ID field shall be set to FFFFFDh, indicating the destination VE_Port.

Payload: the format of the FDUN Request Sequence Payload is shown in table 238.

Table 238 – FDUN Request Payload

Item	Size (bytes)
SW_ILS Code = A000 0004h	4
Destination Controlling Switch Switch_Name	8
Originating FCDF Switch_Name	8
Descriptor List Length	4
FCDF Reachability Descriptor	see 17.7.2.5

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating FCDF Switch_Name: contains the Switch_Name of the requesting FCDF.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

FCDF Reachability Descriptor: describes the deinstantiated ASL **(**see 17.7.2.5)..

FDUN Reply Sequence

VA_RJT: indicates the rejection of the FDUN Request Sequence.

SW_ACC: indicates the acceptance of the FDUN Request Sequence. The format of the FDUN SW_ACC Payload is shown in table 239.

Table 239 – FDUN SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination FCDF Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length = 0000 0000h	4

17.7.3.6 N_Port_ID Route Distribution (NPRD)

The N_Port_ID Route Distribution SW_ILS is used by the Primary Controlling Switch to communicate to an FCDF the N_Port_ID routing information for the Distributed Switch. If the Primary Controlling Switch does not have an ASL with the destination FCDF, the NPRD SW_ILS is relayed to the destination FCDF by the intermediate FCDFs.

NPRD Request Sequence

Addressing: the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port.

Payload: the format of the NPRD Request Sequence Payload is shown in table 240.

Table 240 – NPRD Request Payload

Item	Size (bytes)
SW_ILS Code = A000 0005h	4
Destination FCDF Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
Sequence Number Descriptor	see 17.7.2.6
Primary Controlling Switch Reachability Descriptor	see 17.7.2.7
Secondary Controlling Switch Reachability Descriptor	see 17.7.2.7
N_Port_IDs Reachability Descriptor	see 17.7.2.8
Domain_IDs Reachability Descriptor	see 17.7.2.9

Destination FCDF Switch_Name: contains the Switch_Name of the destination FCDF.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the requesting Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

Sequence Number Descriptor: see 17.7.2.6.

Primary Controlling Switch Reachability Descriptor: contains the reachability information toward the Primary Controlling Switch.

NOTE 49 – Paths toward the Primary Controlling Switch are fundamental for the operation of an FCDF. Specifying higher cost paths enables more redundancy, because if the lowest cost path toward the Primary Controlling Switch fails, a higher cost path may be used.

Secondary Controlling Switch Reachability Descriptor: contains the reachability information toward the Secondary Controlling Switch.

N_Port_IDs Reachability Descriptor: see 17.7.2.8.

Domain_IDs Reachability Descriptor: see 17.7.2.9.

NPRD Reply Sequence

VA_RJT: indicates the rejection of the NPRD Request Sequence.

SW_ACC: indicates the acceptance of the NPRD Request Sequence. The format of the NPRD SW_ACC Payload is shown in table 241.

Table 241 – NPRD SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination Controlling Switch Switch_Name	8
Originating FCDF Switch_Name	8
Descriptor List Length = 0000 0000h	4

17.7.3.7 N_Port_ID and Zoning ACL Distribution (NPZD)

The N_Port_ID and Zoning ACL Distribution SW_ILS is used by the Primary Controlling Switch to communicate to the Secondary Controlling Switch the allocation of an N_Port_ID and/or the deallocation of one or more N_Port_IDs and to communicate to an FCDF the allocation of an N_Port_ID and its associated Zoning ACL information and/or the deallocation of one or more N_Port_IDs and their associated Zoning ACL information. Upon receiving an NPZD Request, an FCDF shall update its Zoning enforcement according to the received Zoning ACLs only for the listed Principal N_Port_IDs. If the Primary Controlling Switch does not have an ASL with the destination FCDF, the NPZD SW_ILS is relayed to the destination FCDF by the intermediate FCDFs.

NPZD Request Sequence

Addressing: when used between a FCDF and the Primary Controlling Switch the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port. When used between the two Controlling Switches the S_ID field shall be set to FFFFFDh, indicating the originating VE_Port, and the D_ID field shall be set to FFFFFDh, indicating the destination VE_Port.

Payload: the format of the NPZD Request Sequence Payload is shown in table 242.

Table 242 – NPZD Request Payload

Item	Size (bytes)
SW_ILS Code = A000 0006h	4
Destination FCDF or Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
Sequence Number Descriptor	see 17.7.2.6
Allocation Status Descriptor	see 17.7.2.10
Peering Status Descriptor	see 17.7.2.11

Destination FCDF or Controlling Switch Switch_Name: contains the Switch_Name of the destination FCDF or Controlling Switch.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the requesting Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

Sequence Number: see 17.7.2.6.

Allocation Status Descriptor: see 17.7.2.10.

When an N_Port_ID is deallocated and allocated at the same time (e.g., as a result of a re-login), that N_Port_ID is listed only in the allocation entry (i.e., there is no deallocation entry for that N_Port_ID in the Allocation Status descriptor).

Peering Status Descriptor: see 17.7.2.11.

When present, the Peering Status descriptor contains Peering entries per each VN_Port currently logged into the destination FCDF and with which the allocated N_Port_ID is allowed to communicate or with which the deallocated N_Port_IDs were allowed to communicate, according to the current fabric zoning configuration. In case of allocation, the Peering Status descriptor for the FCDF that receives the allocated N_Port_ID also contains a Peering Entry with a Principal N_Port_ID equal to the allocated N_Port_ID. In case of deallocation, the Zoning ACLs for the deallocated N_Port_IDs are implicitly removed and the Peering Status descriptor for the FCDF that had the deallocated N_Port_IDs does not contain Peering Entries with a Principal N_Port_ID equal to any the deallocated N_Port_IDs.

NPZD Reply Sequence

VA_RJT: indicates the rejection of the NPZD Request Sequence.

SW_ACC: indicates the acceptance of the NPZD Request Sequence. The format of the NPZD SW_ACC Payload is shown in table 243.

Table 243 – NPZD SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination Controlling Switch Switch_Name	8
Originating FCDF Switch_Name	8
Descriptor List Length = 0000 0000h	4

17.7.3.8 Active Zoning ACL Distribution (AZAD)

The Active Zoning ACL Distribution SW_ILS is used by the Primary Controlling Switch to communicate to an FCDF new Zoning ACL information when a new Zone Set is activated in the fabric. Upon receiving an AZAD Request, an FCDF shall completely replace its Zoning enforcement according to the received Zoning ACLs. If the Primary Controlling Switch does not have an ASL with the destination FCDF, the AZAD SW_ILS is relayed to the destination FCDF by the intermediate FCDFs.

AZAD Request Sequence

Addressing: the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port.

Payload: the format of the AZAD Request Sequence Payload is shown in table 244.

Table 244 – AZAD Request Payload

Item	Size (bytes)
SW_ILS Code = A000 0007h	4
Destination FCDF Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
Sequence Number Descriptor	see 17.7.2.6
Peering Status Descriptor	see 17.7.2.11

Destination FCDF Switch_Name: contains the Switch_Name of the destination FCDF.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the requesting Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

Sequence Number: see 17.7.2.6.

Peering Status Descriptor: see 17.7.2.11.

AZAD Reply Sequence

VA_RJT: indicates the rejection of the AZAD Request Sequence.

SW_ACC: indicates the acceptance of the AZAD Request Sequence. The format of the AZAD SW_ACC Payload is shown in table 245.

Table 245 – AZAD SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination Controlling Switch Switch_Name	8
Originating FCDF Switch_Name	8
Descriptor List Length = 0000 0000h	4

17.7.3.9 Distributed Switch Membership Distribution (DFMD)

The Distributed Switch Membership Distribution SW_ILS is used by the Primary Controlling Switch to communicate to an FCDF the identities of the Primary and Secondary Controlling Switches and of all the FCDFs that compose the Distributed Switch. The DFMD payload may be integrity protected by a cryptographic hash; in this case the involved entities shall be provided with a shared key. If the Primary Controlling Switch does not have an ASL with the destination FCDF, the DFMD SW_ILS is relayed to the destination FCDF by the intermediate FCDFs.

DFMD Request Sequence

Addressing: the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port.

Payload: the format of the DFMD Request Sequence Payload is shown in table 246.

Table 246 – DFMD Request Payload

Item	Size (bytes)
SW_ILS Code = A000 0008h	4
Destination FCDF Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
Membership Set Descriptor	see 17.7.2.12
Integrity Descriptor	see 17.7.2.13

Destination FCDF Switch_Name: contains the Switch_Name of the destination FCDF.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the originating Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

Membership Set Descriptor: see 17.7.2.12.

Integrity Descriptor: see 17.7.2.13.

DFMD Reply Sequence

VA_RJT: indicates the rejection of the DFMD Request Sequence.

SW_ACC: indicates the acceptance of the DFMD Request Sequence. The format of the DFMD SW_ACC Payload is shown in table 247.

Table 247 – DFMD SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination Controlling Switch Switch_Name	8
Originating FCDF Switch_Name	8
Descriptor List Length	4
FCDF Identification Descriptor	see 17.7.2.14

17.7.4 VA_Port SW_ILS Timeouts

Table 248 shows the timeouts associated to each VA_Port SW_ILS.

Table 248 – VA_Port SW_ILSs Timeouts

Description	Abbreviation	Timeout
VN_Port Reachability Notification	VNRN	2000 ms
VN_Port Unreachability Notification	VNUN	500 ms
FCDF Reachability Notification	FDRN	500 ms
FCDF Unreachability Notification	FDUN	500 ms
N_Port_ID Route Distribution	NPRD	1000 ms
N_Port_ID and Zoning ACL Distribution	NPZD	1000 ms
Active Zoning ACL Distribution	AZAD	1000 ms
Distributed Switch Membership Distribution	DFMD	1000 ms

17.8 Redundancy Protocol SW_ILSs

17.8.1 Overview

The Redundancy Protocol SW_ILSs are used to exchange redundancy information between Controlling Switches. Redundancy Protocol SW_ILSs include the originating and destination Controlling Switch Switch_Names in the first two fields of their payload. The subsequent part of a Redundancy Protocol SW_ILS is a list of self-identifying descriptors, as defined in 17.8.2. The descriptor list may be null.

The Redundancy Protocol SW_ILSs have the same high-order byte in their command code, denoted here as YYh. Table 249 shows the Redundancy Protocol SW_ILSs command codes.

Table 249 – Redundancy Protocol SW_ILSs Command Codes

Encoded Value	Description	Abbreviation
A100 0001h	Exchange Redundancy Parameters	ERP
A100 0002h	Get FCDF Topology State	GFTS
A100 0003h	Get FDCF N_Port_IDs State	GFNS
A100 0004h	Secondary Synchronization Achieved	SSA
A100 0005h	Redundancy Hello	RHello

17.8.2 Redundancy Protocol Descriptors

17.8.2.1 Descriptor Format

The Redundancy Protocol descriptors have the same format of the VA_Port SW_ILS descriptors (see 17.7.2.1). Descriptor tags are shown in table 212.

17.8.2.2 Controlling Switch State Descriptor

The format of the Controlling Switch State descriptor is shown in table 250.

Table 250 – Controlling Switch State Descriptor Format

Item	Size (Bytes)
Tag Value = 0011h	4
Length = variable	4
Originating Controlling Switch Priority	4
Virtual Domain Switch_Name	8
Number of Allocated N_Port_ID Ranges (q)	4
Allocated N_Port_ID Range #1	4
Allocated N_Port_ID Range #2	4
…	
Allocated N_Port_ID Range #q	4

Originating Controlling Switch Priority: contains the operational Priority of the originating Controlling Switch in the least significant byte and three reserved bytes in the three most significant bytes.

Virtual Domain Switch_Name: contains the Switch_Name of the Virtual Domain. This field shall be set to 00000000 00000000h if the Switch_Name of the Virtual Domain has not yet been assigned.

Number of Allocated N_Port_ID Ranges: contains the number of Allocated N_Port_ID Range Entries that follow. This list of Allocated N_Port_ID Ranges identifies the N_Port_IDs allocated by the originating Controlling Switch. The N_Port_ID Range is defined by an N_Port_ID in the least significant three bytes, and by the number of bits defining the range in the most significant byte (e.g., the range 020200h .. 02027Fh is expressed as '7 || 020200h').

17.8.2.3 FCDF Topology Descriptor

The format of the FCDF Topology descriptor is shown in table 251.

Table 251 – FCDF Topology Descriptor Format

Item	Size (Bytes)
Tag Value = 0012h	4
Length = variable	4
Number of FCDF Connectivity Records (n)	4
FCDF Connectivity Record #1	see table 252
FCDF Connectivity Record #2	see table 252
...	
FCDF Connectivity Record #n	see table 252

Number of FCDF Connectivity Records: contains the number of FCDF Connectivity Records that follow. The format of the FCDF Connectivity Record is shown in table 252.

Table 252 – FCDF Connectivity Record Format

Item	Size (bytes)
FCDF Switch_Name	8
Number of ASL Records (m)	4
ASL Record #1	28
ASL Record #2	28
...	
ASL Record #m	8

FCDF Switch_Name: contains the Switch_Name of the FCDF whose ASLs are being described.

Number of ASL Records: contains the number of ASL Records that follow. The format of the ASL Record is shown in table 252

Table 253 – FCDF Connectivity Record Format

Item	Size (bytes)
Switch_Name of Neighbor	8
Local A_Port_Name	8
Adjacent A_Port_Name	8
Link Cost	4

Switch_Name of Neighbor: contains the Switch_Name of the FCDF or Controlling Switch at the other end of the described ASL.

Local A_Port_Name: contains the local A_Port_Name of the described ASL.

Adjacent A_Port_Name: contains the adjacent A_Port_Name of the described ASL.

Link Cost: contains the link cost of the described ASL in the two least significant bytes and two reserved bytes in the two most significant bytes.

17.8.2.4 FCDF N_Port_IDs Descriptor

The format of the FCDF N_Port_IDs descriptor is shown in table 217.

Table 254 – FCDF N_Port_IDs Descriptor Format

Item	Size (Bytes)
Tag Value = 0013h	4
Length = variable	4
Number of Virtual Domain_ID Records (z)	4
Virtual Domain_ID Record #1	see table 255
Virtual Domain_ID Record #2	see table 255
...	
Virtual Domain_ID Record #z	see table 255
Number of FCDF Allocation Records (n)	4
FCDF Allocation Record #1	see table 256
FCDF Allocation Record #2	see table 256
...	
FCDF Allocation Record #n	see table 256

Number of Virtual Domain_ID Records: contains the number of Virtual Domain_ID Records that follow. The format of the Virtual Domain_ID Record is shown in table 256.

Table 255 – Virtual Domain_ID Record Format

Item	Size (bytes)
Virtual Domain_ID Value	4
Distributed Switch Switch_Name	8

Virtual Domain_ID Value: contains a Virtual Domain_ID for the Distributed Switch in the least significant byte and three reserved bytes in the three most significant bytes.

Distributed Switch Switch_Name: contains a Switch_Name for the Distributed Switch, Switch_Name associated with the Virtual Domain_ID value.

Number of FCDF Allocation Records: contains the number of FCDF Allocation Records that follow. The format of the FCDF Allocation Record is shown in table 256.

Table 256 – FCDF Allocation Record Format

Item	Size (bytes)
FCDF Switch_Name	8
Number of Allocated N_Port_ID Ranges (s)	4
Allocated N_Port_ID Range #1	4
Allocated N_Port_ID Range #2	4
...	
Allocated N_Port_ID Range #s	4

FCDF Switch_Name: contains the Switch_Name of the FCDF whose N_Port_IDs allocation is provided.

Number of Allocated N_Port_ID Ranges: contains the number of Allocated N_Port_ID Range Entries that follow. This list of Allocated N_Port_ID Ranges identifies the N_Port_IDs allocated to the described FCDF. The N_Port_ID Range is defined by an N_Port_ID in the least significant three bytes, and by the number of bits defining the range in the most significant byte (e.g., the range 020200h .. 02027Fh is expressed as ‘7 || 020200h’).

17.8.2.5 RHello Interval Descriptor

The format of the RHello Interval descriptor is shown in table 257.

Table 257 – RHello Interval Descriptor Format

Item	Size (Bytes)
Tag Value = 0014h	4
Length = 4	4
RHello_Interval	4

RHello_Interval: contains the RHello_Interval value expressed in ms.

17.8.3 Redundancy Protocol SW_ILSs

17.8.3.1 Exchange Redundancy Parameters (ERP)

The Exchange Redundancy Parameter (ERP) SW_ILS is used by the redundancy protocol to determine which Controlling Switch behaves as Primary and which one behaves as Secondary.

ERP Request Sequence

Addressing: the S_ID field shall be set to FFFFFDh, indicating the originating VE_Port, and the D_ID field shall be set to FFFFFDh, indicating the destination VE_Port.

Payload: The format of the ERP Request Sequence Payload is shown in table 258.

Table 258 – ERP Request Payload

Item	Size (bytes)
SW_ILS Code = A100 0001h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
Controlling Switch State Descriptor	see 17.8.2.2

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the requesting Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

Controlling Switch State Descriptor: see 17.8.2.2.

ERP Reply Sequence

SW_RJT: indicates the rejection of the ERP Request Sequence.

SW_ACC: indicates the acceptance of the ERP Request Sequence. The format of the ERP SW_ACC Payload is shown in table 259.

Table 259 – ERP SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
Controlling Switch State Descriptor	see 17.8.2.2

17.8.3.2 Get FCDF Topology State (GFTS)

The Get FCDF Topology State (GFTS) SW_ILS is used by the Secondary Controlling Switch to request to the Primary the Virtual Domain_ID value(s) and the current FCDF topology, in order to synchronize its state with the one of the Primary.

GFTS Request Sequence

Addressing: the S_ID field shall be set to FFFFFDh, indicating the originating VE_Port, and the D_ID field shall be set to FFFFFDh, indicating the destination VE_Port.

Payload: The format of the GFTS Request Sequence Payload is shown in table 260.

Table 260 – GFTS Request Payload

Item	Size (bytes)
SW_ILS Code = A100 0002h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length = 0000 0000h	4

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the originating Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

GFTS Reply Sequence

SW_RJT: indicates the rejection of the GFTS Request Sequence.

SW_ACC: indicates the acceptance of the GFTS Request Sequence. The format of the GFTS SW_ACC Payload is shown in table 261.

Table 261 – GFTS SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
FCDF Topology Descriptor	see 17.8.2.3

FCDF Topology Descriptor: see 17.8.2.3.

17.8.3.3 Get FCDF N_Port_IDs State (GFNS)

The Get FDCF N_Port_IDs State (GFNS) SW_ILS is used by the Secondary Controlling Switch to request to the Primary the current allocation of N_Port_IDs to each FCDF of the Distributed Switch, in order to synchronize its state with the one of the Primary.

GFNS Request Sequence

Addressing: the S_ID field shall be set to FFFFFDh, indicating the originating VE_Port, and the D_ID field shall be set to FFFFFDh, indicating the destination VE_Port.

Payload: The format of the GFNS Request Sequence Payload is shown in table 262.

Table 262 – GFNS Request Payload

Item	Size (bytes)
SW_ILS Code = A100 0003h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length = 0000 0000h	4

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the originating Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

GFNS Reply Sequence

SW_RJT: indicates the rejection of the GFNS Request Sequence.

SW_ACC: indicates the acceptance of the GFNS Request Sequence. The format of the GFNS SW_ACC Payload is shown in table 263.

Table 263 – GFNS SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
FCDF N_Port_IDs Descriptor	see 17.8.2.4

FCDF N_Port_IDs Descriptor: see 17.8.2.4.

17.8.3.4 Secondary Synchronization Achieved (SSA)

The Secondary Synchronization Achieved (SSA) SW_ILS is used by the Secondary Controlling Switch to communicate to the Primary that it achieved state synchronization.

SSA Request Sequence

Addressing: the S_ID field shall be set to FFFFFDh, indicating the originating VE_Port, and the D_ID field shall be set to FFFFFDh, indicating the destination VE_Port.

Payload: The format of the SSA Request Sequence Payload is shown in table 264.

Table 264 – SSA Request Payload

Item	Size (bytes)
SW_ILS Code = A100 0004h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length = 0000 0000h	4

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the originating Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

SSA Reply Sequence

SW_RJT: indicates the rejection of the SSA Request Sequence.

SW_ACC: indicates the acceptance of the SSA Request Sequence. The format of the SSA SW_ACC Payload is shown in table 265.

Table 265 – SSA SW_ACC Payload

Item	Size (bytes)
SW_ILS Code = 0200 0000h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length = 0000 0000h	4

17.8.3.5 Redundancy Hello (RHello)

The Redundancy Hello (RHello) SW_ILS is used by the redundancy protocol. The Rhello SW_ILS is transmitted in a unidirectional Exchange (i.e., it does not have a Reply Sequence).

RHello Request Sequence

Addressing: when used over an AISL, the S_ID field shall be set to FFFFFDh, indicating the originating VE_Port, and the D_ID field shall be set to FFFFFDh, indicating the destination VE_Port. When used over an ASL, the S_ID field shall be set to FFFFF9h, indicating the originating VA_Port, and the D_ID field shall be set to FFFFF9h, indicating the destination VA_Port.

Payload: The format of the RHello Request Sequence Payload is shown in table 266.

Table 266 – RHello Request Payload

Item	Size (bytes)
SW_ILS Code = A100 0005h	4
Destination Controlling Switch Switch_Name	8
Originating Controlling Switch Switch_Name	8
Descriptor List Length	4
RHello Interval Descriptor	see 17.8.2.5

Destination Controlling Switch Switch_Name: contains the Switch_Name of the destination Controlling Switch.

Originating Controlling Switch Switch_Name: contains the Switch_Name of the originating Controlling Switch.

Descriptor List Length: contains the length in bytes of the subsequent list of descriptors.

RHello_Interval: see 17.8.2.5.

17.8.4 Redundancy Protocol Timeouts

Table 267 shows the timeouts associated to each Redundancy Protocol SW_ILS.

Table 267 – Redundancy Protocol SW_ILSs Timeouts

Description	Abbreviation	Timeout
Exchange Redundancy Parameter	ERP	1000 ms
Get FCDF Topology State	GFTS	1000 ms
Get FDCF N_Port_IDs State	GFNS	1000 ms
Secondary Synchronization Achieved	SSA	1000 ms

17.9 Distributed Switch Operations

17.9.1 Overview

In a Distributed Switch, the Primary Controlling Switch defines the routes for the FCDF topology and performs N_Port_ID allocations and deallocations for all its controlled FCDFs. When two Controlling Switches are present in a Distributed Switch, the two Controlling Switches keep their state synchronized.

17.9.2 FCDF Routing

A Controlling Switch establishes its role when in state P2 or S2 of of the Controlling Switch Redundancy Protocol (see 17.8). When a Controlling Switch established its role, it instantiates ASLs with the FCDFs that are directly connected and are part of its FCDF Set.

Upon instantiating an ASL with an FCDF, the Primary Controlling Switch shall initiate an FDRN Exchange describing that link with the Secondary Controlling Switch, if available, to keep the state synchronized. Upon completion of this FDRN Exchange, the Primary Controlling Switch shall provide to that FCDF the Distributed Switch Membership information through a DFMD Exchange. The Primary Controlling Switch shall recompute the N_Port_ID routes and distribute them to each FCDF belonging to the Distributed Switch through NPRD Exchanges. At this point the instantiated ASL becomes part of the Distributes Switch Internal Topology (i.e., the set of ASLs internal to the Distributed Switch).

Upon de-instantiating an ASL with an FCDF, the Primary Controlling Switch shall initiate an FDUN Exchange describing that disappeared link with the Secondary Controlling Switch, if available, to keep the state synchronized. Upon completion of this FDUN Exchange, the Primary Controlling Switch shall recompute the N_Port_ID routes and distribute them to each FCDF belonging to the Distributed Switch through NPRD Exchanges.

When becoming operational, an FCDF waits for a Controlling Switch or another FCDF to initiate an ELP Exchange with it, in order to set up a ASL. Upon completing the DFMD and NPRD Exchanges with the Primary Controlling Switch, the FCDF becomes able to initiate ELP Requests to instantiate other ASLs with other FCDFs. At this point the FCDF is part of the Distributed Switch internal topology and enables its ports for logins from Nodes; any FLOGI received on a FCDF port before this point is responded by the FCDF with a LS_RJT having reason code ‘Logical busy’ and reason code explanation ‘No additional explanation’.

Upon instantiating a ASL with another FCDF or with the Secondary Controlling Switch, an FCDF shall perform a FDRN Exchange with the Primary Controlling Switch to inform it of the new link. Upon completing a FDRN Exchange with an FCDF, the Primary Controlling Switch shall initiate another FDRN Exchange with the same parameters with the Secondary Controlling Switch, if available, to keep the state synchronized. After completing this FDRN Exchange the primary Controlling Switch shall provide to the newly reported FCDF the Distributed Switch Membership information through a DFMD Exchange, if that FCDF did not not already receive a DFMD Exchange in a previous step. At this point the instantiated ASL becomes part of the Distributed Switch internal topology (i.e., the set of ASLs internal to the Distributed Switch). Upon completion of this DFMD Exchange, the Primary Controlling Switch shall recompute the N_Port_ID routes and distribute them to each FCDF belonging to the Distributed Switch through NPRD Exchanges.

NOTE 50 – An ASL with the Secondary Controlling Switch may be instantiated before the ASL with the Primary Controlling Switch. The FCDF recognizes the Primary Controlling Switch because it is the one from which it receives the DFMD Request. In this case, the FCDF initiates with the Primary Controlling Switch the FDRN Exchange describing the link with the Secondary Controlling Switch upon completing the DFMD Exchange.

Upon deinstantiating an ASL with another FCDF or with the Secondary Controlling Switch, an FCDF shall perform a FDUN Exchange with the Primary Controlling Switch to inform it of the disappeared link. Upon completing a FDUN Exchange with an FCDF, the Primary Controlling Switch shall initiate another FDUN Exchange with the same parameters with the Secondary Controlling Switch, if available, to keep the state synchronized. Upon completion of this FDUN Exchange, the Primary Controlling Switch shall recompute the N_Port_ID routes and distribute them to each FCDF belonging to the Distributed Switch through NPRD Exchanges.

The distribution of NPRD Requests shall take precedence over the distribution of AZAD and NPZD Requests.

17.9.3 N_Port_ID Handling

Upon receiving on a port a FLOGI Request or a NPIV FDISC Request from a Node, an FCDF shall send a VNRN Request to the Primary Controlling Switch to inform it of the newly reachable VN_Port. If the Primary Controlling Switch rejects the VNRN Request, the FCDF shall also reject the FLOGI Request or NPIV FDISC Request. Upon receiving the VNRN Request, the Primary Controlling Switch performs the following processing:

a) if the VNRN Request carried a FLOGI Request and that VN_Port was not already logged in or if the VNRN Request carried a NPIV FDISC Request, then the Primary Controlling Switch shall allocate to the newly reachable VN_Port an N_Port_ID from a Virtual Domain_ID; or

b) if the VNRN Request carried a FLOGI Request and that VN_Port was already logged in, then the Primary Controlling Switch shall implicitly log out that VN_Port and all the VN_Ports associated to the VF_Port that VN_Port was associated with and then allocate to that VN_Port an N_Port_ID from a Virtual Domain_ID.

The Primary Controlling Switch shall also recompute the Zoning ACLs for the affected N_Port_IDs, generate appropriate RSCN(s), and update the Fibre Channel Name Server. The Primary Controlling Switch shall distribute the Zoning ACLs and N_Port_ID allocation/deallocation information to the Secondary Controlling Switch, if available, and to each FCDF belonging to the Distributed Switch through an appropriate NPZD Exchange. The NPZD Request sent to the Secondary Controlling Switch shall carry no Peering Entries. The NPZD Requests sent to the Secondary Controlling Switch shall include the FLOGI / NPIV FDISC LS_ACC Parameters; the NPZD Requests sent to the FCDFs shall not include them. The Primary Controlling Switch shall wait to receive the NPZD SW_ACC from the Secondary Controlling Switch and the NPZD SW_ACC from the FCDF that sent the VNRN Request before sending the VNRN SW_ACC containing the FLOGI/NPIV FDISC LS_ACC parameters to the FCDF that sent the VNRN Request. Upon receiving the VNRN SW_ACC, containing the FLOGI / NPIV FDISC LS_ACC Parameters, the FCDF that sent the VNRN Request shall accept the FLOGI Request or NPIV FDISC Request and complete the N_Port Fabric login. If the FLOGI or NPIV FDISC Exchange that triggered the VNRN Request has been terminated, then an FLOGI LS_ACC or an NPIV FDISC LS_ACC shall not be generated upon receiving the VNRN SW_ACC. In this case, if upon receiving the VNRN SW_ACC or upon termination of the VNRN Exchange there is no Fabric Login in progress or established for the VN_Port that began the terminated FLOGI or NPIV FDISC Exchange, then the FCDF shall perform a VNUN Exchange with the Primary Controlling Switch to inform the Primary Controlling Switch that the VN_Port is now unreachable.

If while performing the processing of a VNRN Request the Primary Controlling Switch receives a second VNRN Request for the same VN_Port (i.e., the second VNRN Request is received while the first VNRN Exchange is still open) and the processing of the second VNRN Request results in NPZD Requests having the same payloads as the ones generated for the first VNRN Request, then the Controlling Switch may skip sending the second set of NPZD Requests and reply to the second VNRN Request once NPZD processing for the first VNRN Request is considered completed.

When a VN_Port is logged out or when a VF_Port is de-instantiated, an FCDF shall perform a VNUN Exchange with the Primary Controlling Switch to inform it that the VN_Port is now unreachable or that all the VN_Ports associated with that VF_Port are unreachable. Upon completing a VNUN Exchange, the Primary Controlling Switch shall deallocate the N_Port_ID(s) assigned to the affected VN_Port(s), recompute the Zoning ACLs for the affected N_Port_IDs, generate appropriate RSCN(s), and update the Fibre Channel Name Server. The Primary Controlling Switch shall then distribute this information to the Secondary Controlling Switch, if available, and to each FCDF belonging to the Distributed Switch through NPZD Requests indicating N_Port_ID(s) deallocation.

The Primary Controlling Switch maintains a sequence number for each FCDF in the FCDF Set. The sequence number is incremented by one and included in the NPZD, NPRD, or AZAD sequence number descriptor each time an NPZD, NPRD, or AZAD Request is sent.

Upon receipt of an NPZD, NPRD, or AZAD Request, an FCDF compares the sequence number in the received sequence number descriptor to that of the last processed NPZD, NPRD, or AZAD Request, or to 00000000 00000001h if none of these commands (NPZD, NPRD, AZAD) has previously been processed. If the received sequence number is lower, except in the case where a sequence number wrap condition has been detected, the received NPZD, NPRD, or AZAD request shall be discarded and a VA_RJT shall be sent with Reason Code of ‘Logical Error’ and Reason Code Explanation of ‘Out of Order’. If the received sequence number is higher, or a wrap condition has been detected, then the NPZD, NPRD, or AZAD is processed.

An FCDF considers an N_Port_ID to be allocated when it has successfully received the N_Port_ID in an Allocation Entry of the current or previous NPZD Request. If an NPZD Request contains a peering entry with a Principal N_Port_ID that has not been allocated, that entire peering entry shall be ignored. If an NPZD Request contains a peering entry with a Principal N_Port_ID that is currently allocated, but that peering entry contains Peer N_Port_ID(s) that have not been allocated, then those Peer N_Port_ID(s) shall be ignored.

Whenever an NPZD Request is retransmitted for any reason (e.g., timeout) the Zoning ACLs for the affected N_Port_IDs shall be recomputed and a new NPZD Request including a new sequence number and the newly computed peering entries shall be sent.

If a Primary Controlling Switch receives a VA_RJT with a Reason Code of ‘Logical Error’ and Reason Code Explanation of ‘Out of Order’ in response to an NPZD Request, the Primary Controlling Switch shall retransmit the NPZD Request.

When a new Zone Set is activated in the Fabric, the Primary Controlling Switch shall recompute the Zoning ACLs for all N_Port_IDs allocated in the Virtual Domain and distribute them to the FCDFs of the Distributed Switch through AZAD Exchanges.

If the Primary Controlling Switch has to send an AZAD request to an FCDF, any NPZD or NPRD requests outstanding to that FCDF shall first be completed. Any AZAD requests outstanding shall also be completed prior to initiating any subsequent NPZD or NPRD requests with that FCDF.

If the Primary Controlling Switch has to send an NPRD request to an FCDF, any NPZD or AZAD requests outstanding to that FCDF shall first be completed. Any NPRD requests outstanding shall also be completed prior to initiating any subsequent NPZD or AZAD requests with that FCDF.

Upon receiving on a port a FLOGI Request or a NPIV FDISC Request from a Node, a Controlling Switch shall allocate to the newly reachable VN_Port an N_Port_ID from the Principal Domain_ID if it accepts the received FLOGI or NPIV FDISC Request.

17.9.4 Distribution Tree

The Primary Controlling Switch shall compute a distribution tree to distribute VA_Port SW_ILSs to the FCDFs. The distribution tree information is encoded in the NPRD Request (i.e., only the first path listed in a N_Port_ID Reachability Entry is used to relay VA_Port SW_ILSs).

17.10 Distributed Switch Redundancy Protocol

17.10.1 Redundancy Protocol Overview

The purpose of the Controlling Switch Redundancy protocol is to avoid any single point of failure in a Distributed Switch. Figure 50 shows an example of redundant Distributed Switch, including the two Principal Domains and the Virtual Domain.

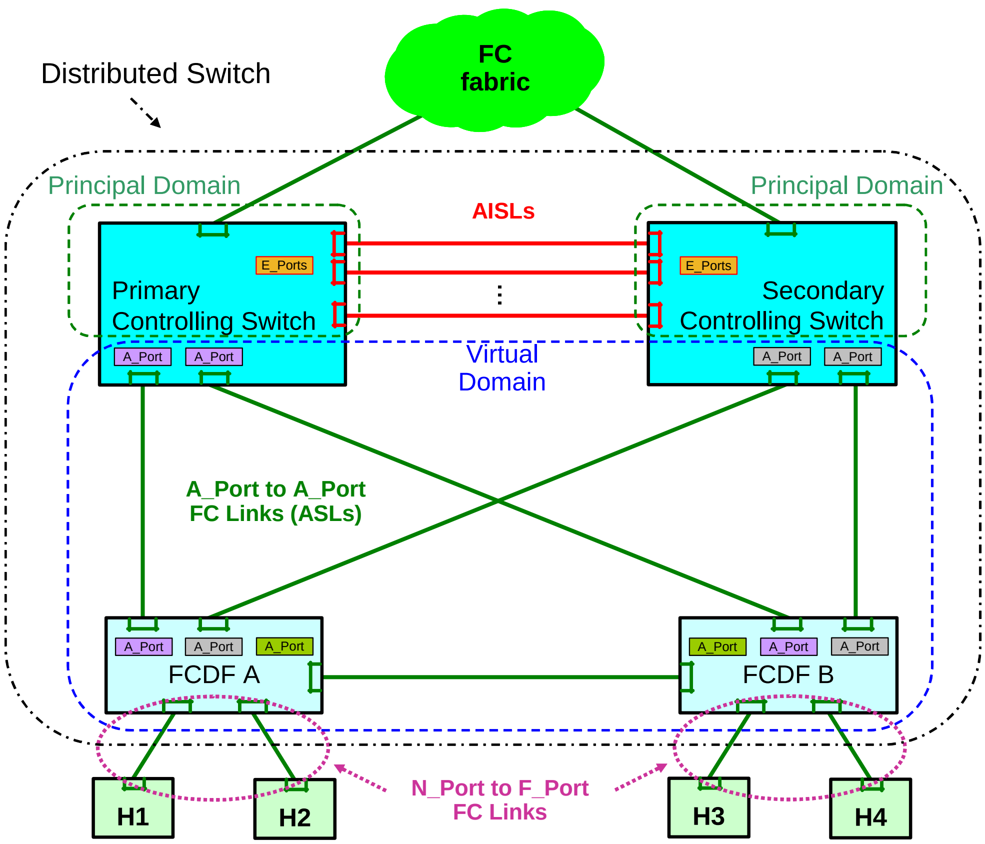

Figure 50 – Example of Redundant Distributed Switch

The Controlling Switch Redundancy protocol uses a set of Augmented E_Port to E_Port links (AISLs) between the Primary and Secondary Controlling Switches. This set is referred to as the AISL Set. It is strongly recommended to deploy at least two AISLs in the AISL Set, in order to distinguish the case of an AISL failure from the case of a Controlling Switch failure. Additional AISLs provide additional resiliency.

In a Redundant Distributed Switch the Primary Controlling Switch generates the LSR(s) describing the Virtual Domain in the Distributed Switch. In addition, both Primary and Secondary Controlling Switch list the Virtual Domain as a directly attached Domain in their LSR. The resulting FSPF topology is depicted in figure 51, where Z1 .. Zn are the Domain_IDs belonging to the Virtual Domain and X and Y are the Domain_IDs of the Principal Domains of the two Controlling Switches. X and Y are also connected between themselves by virtue of the AISLs.

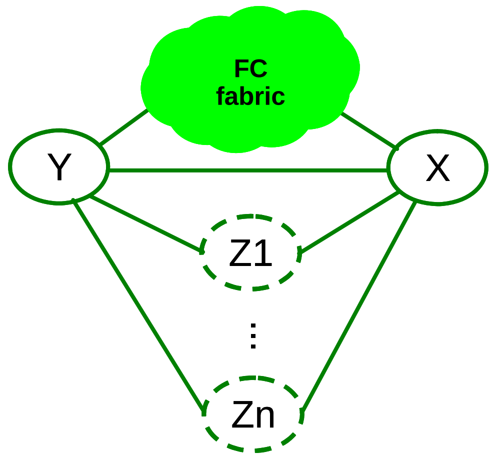

Figure 51 – Distributed Switch FSPF Topology

17.10.2 Redundancy Protocol State Machine

The redundancy protocol state machine reacts to AISLs failures in a timed fashion. To this end, the redundancy protocol state machine uses indications from the physical layer to determine if a link failed together with periodic Redundancy Hello messages (RHello) to verify the health of the Controlling Switches. The redundancy protocol state machine uses the following time intervals and timers:

RHello_Interval: Time interval between RHellos, expressed in milliseconds. The default value is 200 ms.

Down_Interval: Time interval for a Controlling Switch to declare the other one down. Calculated as 2.5 * RHello_Interval.

To determine which Controlling Switch behaves as Primary and which one as Secondary, the redundancy protocol uses a Priority value associated to each Controlling Switch. Priority values are shown in table 268.

Table 268 – Controlling Switch Priority Values

Value	Description
00h	Reserved
01h	Highest Priority value. This value is administratively configured to force the election of a Controlling Switch to Primary.
02h[a]	Primary Controlling Switch priority. This value is used by the Redundancy protocol to identify a Controlling Switch as Primary.
03 .. FEh	Higher to lower Priority values. The default value is 128.
FFh[a]	This value indicates that a Controlling Switch is not willing to operate as Primary. This is used by the Primary Controlling Switch to trigger a transition of the Secondary Controlling Switch to Primary without having to wait for the current Primary to timeout, if appropriate.
[a] These values are used by the Redundancy protocol and not available to an administrator.	

Figure 52 shows the redundancy protocol state machine.

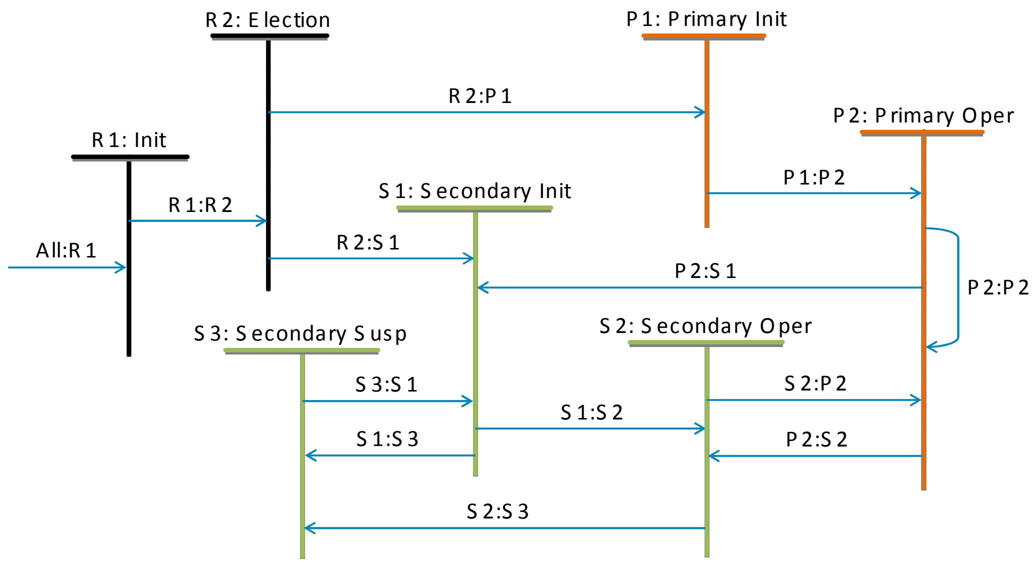

Figure 52 – Redundancy Protocol State Machine

State R1:Init. In this state a Controlling Switch clears its state and waits to begin the processing for the redundancy protocol.

Transition R1:R2. Occurs when processing for the redundancy protocol begins. The redundancy protocol processing begins when:

a) the redundancy protocol is enabled;

b) the Controlling Switch Set and the FCDF Set are configured; and

c) Fabric configuration is completed.

Transition All:R1. Occurs when the redundancy protocol is re-initialized.

State R2:Election. In this state a Controlling Switch determines if it operates as Primary or Secondary. If the AISL Set is NULL, then the Controlling Switch exits this state. If the AISL Set is not NULL, then an ERP Exchange is performed.

NOTE 51 – In this state the ERP payload does not contain N_Port_ID Ranges, nor a non-zero Virtual Domain Switch_Name given that the Controlling Switch cleared its state in state R1.

If the ERP Exchange shows that the local Controlling Switch Priority is 01h and the remote Controlling Switch Priority is 01h (i.e., both Controlling Switches are manually configured to be Primary) then the Redundancy protocol is disabled and an error is logged.

Transition R2:P1. Occurs when:

a) the AISL Set is NULL;

b) the AISL Set is not NULL and the ERP Exchange showed that the local Controlling Switch Priority is lower than the remote Controlling Switch Priority; or

c) the AISL Set is not NULL, the ERP Exchange showed that the local Controlling Switch Priority is equal to the remote Controlling Switch Priority, and the local Switch_Name is lower than the remote Switch_Name.

Transition R2:S1. Occurs when:

a) the AISL Set is not NULL and the ERP Exchange showed that the local Controlling Switch Priority is higher than the remote Controlling Switch Priority; or

b) the AISL Set is not NULL, the ERP Exchange showed that the local Controlling Switch Priority is equal to the remote Controlling Switch Priority, and the local Switch_Name is greater than the remote Switch_Name.

State P1:Primary Initialization. In this state a Controlling Switch performs the operations to become the Primary Controlling Switch of the Distributed Switch. The Controlling Switch sets its Priority to 02h, selects the Virtual Domain Switch_Name (i.e., a Switch_Name for the Virtual Domain), and obtains an additional Domain_ID value (i.e., a Virtual Domain_ID) from the Principal Switch of the fabric by generating an RDI Request on behalf of the Virtual Domain Switch_Name.

Transition P1:P2. Occurs when the Virtual Domain_ID is available.

State P2:Primary Operational. In this state the Controlling Switch is operational as Primary. On entering this state the Controlling Switch:

a) sets its Priority to 02h;

b) initiates an ERP Exchange with the Secondary Controlling Switch, if available;

c) sends a DFMD SW_ILS to all reachable FCDF of the FCDF Set declaring itself as Primary Controlling Switch;

d) on native Fibre Channel links that were Isolated because connected to FCDFs, if any, it performs an ELP; and

e) on FCoE interfaces, it establishes VA_Port to VA_Port Virtual Links with neighbor FDFs belonging to the FDF Set to which no VA_Port to VA_Port Virtual Links has been established, if any.

While in this state, the Controlling Switch:

a) performs the Distributed Switch operations (see 17.9);

b) generates the FSPF LSR(s) describing the Virtual Domain(s) in the Distributed Switch and lists the Virtual Domain(s) as a directly attached Domain(s) in its FSPF LSR;

c) on receiving an SSA SW_ILS (i.e., when the Secondary Controlling Switch completed its state synchronization) sends a DFMD SW_ILS to all reachable FCDFs of the FCDF Set declaring itself as Primary and the Secondary as Secondary;

d) if the Secondary Controlling Switch is available sends RHello Requests every RHello_Interval over each of its AISLs and over each ASL through which the Secondary is reachable;

e) resets the Down_Timer to Down_Interval everytime an RHello Request from the Secondary Controlling Switch is received over at least one AISL or ASL;

f) when the Secondary Controlling Switch is not anymore available (i.e., when Down_Timer expires) sends a DFMD SW_ILS to all reachable FCDFs of the FCDF Set declaring itself as Primary; and

g) if the AISL Set goes from NULL to not-NULL (i.e., a Controlling Switch becomes available), it performs an ERP Exchange with the other Controlling Switch.

Transition P2:P2: Occurs following the ERP Exchange performed when the AISL Set went from NULL to not-NULL if:

a) the received Controlling Switch State Descriptor contains a zero Virtual Domain Switch_Name and no N_Port_ID Ranges;

b) the received Controlling Switch State Descriptor contains the Virtual Domain Switch_Name of the Distributed Switch, there is no allocated N_Port_ID conflict between the two Controlling Switches, and this Switch is the one selected to remain Primary;

c) the received Controlling Switch State Descriptor contains the Virtual Domain Switch_Name of the Distributed Switch and there is an allocated N_Port_ID conflict between the two Controlling Switches. In this case the AISL shall be Isolated; or

d) the received Controlling Switch State Descriptor contains a Virtual Domain Switch_Name different than the one of the Distributed Switch. In this case the AISL shall be isolated.

Transition P2:S1: Occurs following the ERP Exchange performed when the AISL Set went from NULL to not-NULL if the received Controlling Switch State Descriptor contains the Virtual Domain Switch_Name of the Distributed Switch, there is no allocated N_Port_IDs conflict between the two Controlling Switches, and this Switch is the one selected to become Secondary.

State S1:Secondary Initialization. In this state a Controlling Switch performs the operations to become the Secondary Controlling Switch of the Distributed Switch. The Controlling Switch has to synchronize its state with the one of the Primary Controlling Switch. To this end the Controlling Switch:

1) Requests to the Primary the FCDF topology through the GTFS (Get FCDF Topology State) SW_ILS;

2) Requests to the Primary the Virtual Domain_IDs and N_Port_IDs Allocation state in the Distributed Switch through the GFNS (Get FDCF N_Port_IDs State) SW_ILS;

3) Obtains the information associated with each N_Port_ID in the Name Server through the GE_ID CT Request; and

4) Communicates the achieved state synchronization to the Primary through the SSA (Secondary Synchronization Achieved) SW_ILS.

While in this state the Controlling Switch:

a) processes FDUN, FDRN, and NPZD Requests from the Primary Controlling Switch;

b) sends RHello Requests every RHello_Interval over each of its AISLs and over each ASL through which the Primary is reachable; and

c) resets the Down_Timer to Down_Interval everytime an RHello Request is received over at least one AISL or ASL.

Transition S1:S2. Occurs when the Secondary Controlling Switch has synchronized its state with the Primary.

State S2:Secondary Operational. In this state the Controlling Switch is operational as Secondary. On entering this state the Controlling Switch:

a) sets its Priority to its configured value;

b) initiates an ERP Exchange with the Primary Controlling Switch;

c) on native Fibre Channel links that were Isolated because connected to FCDFs, if any, it performs an ELP; and

d) on FCoE interfaces, it establishes VA_Port to VA_Port Virtual Links with neighbor FDFs belonging to the FDF Set to which no VA_Port to VA_Port Virtual Links has been established, if any.

While in this state, the Secondary Controlling Switch:

a) participates in the Distributed Switch operations (see 17.9);

b) lists the Virtual Domain(s) as a directly attached Domain(s) in its FSPF LSR;

c) sends RHello Requests every RHello_Interval over each of its AISLs and over each ASL through which the Primary is reachable; and

d) resets the Down_Timer to Down_Interval everytime an RHello Request is received over at least one AISL or ASL.

Transition S2:P2. Occurs when the Secondary Controlling Switch becomes Primary. This occurs when:

a) the Secondary Controlling Switch Down_Timer expires and the Primary Controlling Switch is no longer part of the fabric FSPF topology (i.e., the Primary Controlling Switch is no longer available); or

b) the Priority field in a received ERP Request has a value of FFh. This is an indication that the Primary Controlling Switch determined to become Secondary.

Transition S2:S3. Occurs when the Secondary Controlling Switch Down_Timer expires and the Primary Controlling Switch is still part of the fabric FSPF topology (i.e., the Secondary Controlling Switch is unable to maintain synchronization with the Primary but the Primary is still part of the fabric).

State S3. Secondary Suspended. In this state the controlling switch suspends its operations as Controlling witch to avoid split-brain scenarios. On entering this state the Controlling Switch shall Isolate all of its ASLs.

Transition S3:S1. Occurs when an AISL with the Primary Controlling Switch is instantiated.

Transition S1:S3. Occurs when the Secondary Controlling Switch Down_Timer expires and the Primary Controlling Switch is still part of the fabric FSPF topology (i.e., the Secondary Controlling Switch is unable to maintain synchronization with the Primary but the Primary is still part of the fabric).

Transition P2:S2. Occurs when the Primary Controlling Switch determines to become Secondary by setting its Priority to FFh. This may happen as result of an administrative action.