Claudio DeSanti (Cisco), Erik Smith (EMC),
Craig Carlson (QLogic), Ed McGlaughlin (QLogic)
Pat Thaler (Broadcom)

1 Distributed Switching

1.1 Definitions

1.1.1 A_Port (Adjacent Port): The combination of one PA_Port and one VA_Port operating together.

1.1.2 AISL (Augmented ISL): an E_Port to E_Port link used by the redundancy protocol.

1.1.3 AISL Set: The set of AISLs that connect the two Controlling Switches that are part of a Distributed Switch.

1.1.4 ASL (A_Port Switch Link): An A_Port to A_Port link.

1.1.5 Controlling Switch: A Switch able to control a set of FCDFs in order to create a Distributed Switch.

1.1.6 Controlling Switch Set: The Switch_Names of the up to two Controlling Switches that are part of a Distributed Switch.

1.1.7 Distributed Switch: A set of FCDFs associated with at least one Controlling Switch, that controls the operations of the set of FCDFs. A Distributed Switch is defined by the administrative configuration of the Controlling Switch Set and of the FCDF Set.

1.1.8 FCDF (FC Data-Plane Forwarder): A simplified FC switching entity that forwards FC frames among VA_Ports and VF_Ports through a FCDF Switching Element (see 1.4). 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.

1.1.9 FCDF Set: The Switch_Names of the FCDFs that are part of a Distributed Switch.

1.1.10 PA_Port (Physical A_Port): The LCF within the Fabric that attaches to another PA_Port through a link.

1.1.11 VA_Port (Virtual A_Port): 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).

1.2 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 1 shows an example of Distributed Switch composed of a Controlling Switch and two FCDFs.

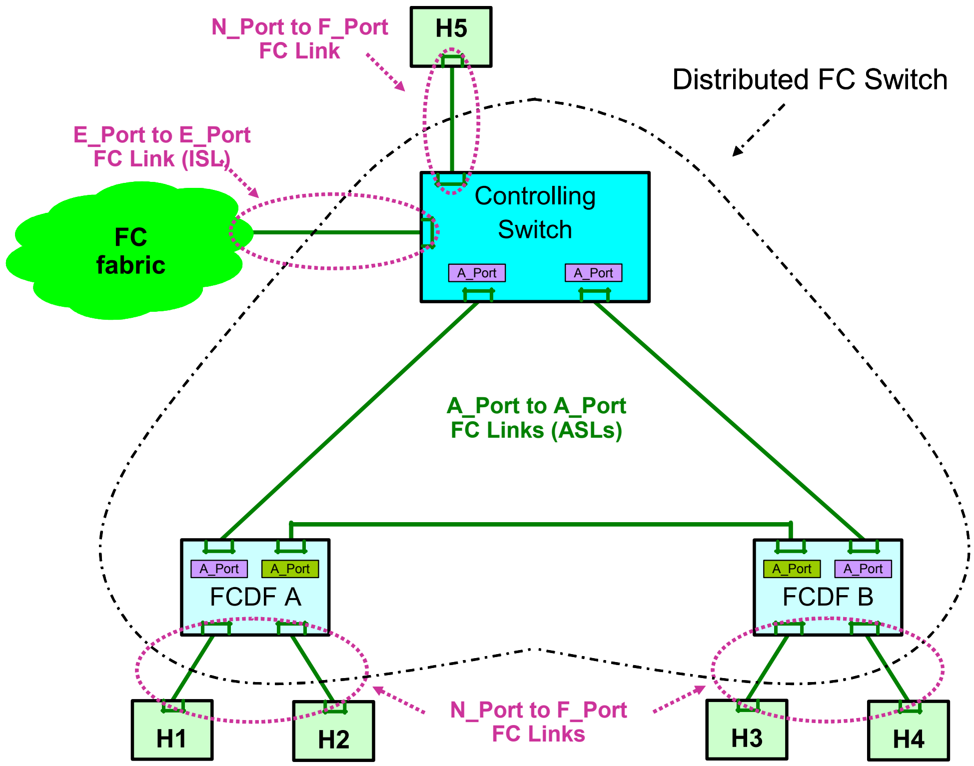

Figure 1 – Example of Distributed Switch

From an external point of view (i.e., outside the dotted and dashed black line in figure 1), 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 1), A_Port to A_Port links (ASLs) enable FC frames forwarding between Controlling Switch and FCDFs, as well as between FCDFs. In addition, ASLs are used to exchange control information between Controlling Switch and FCDFs (see T11/11-225v0).

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 1. 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 (see T11/11-224v0).

Figure 2 shows an example of Distributed Switch including a redundant pair of Controlling Switches.

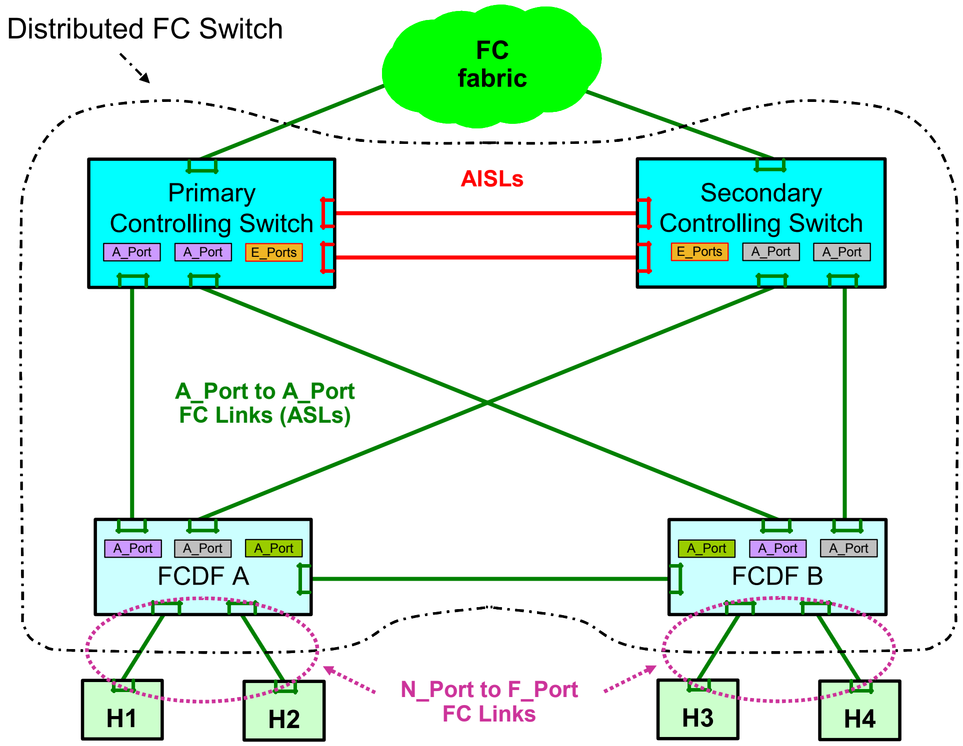

Figure 2 – Example of Redundant Distributed Switch

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 (see T11/11-224v0).

The Controlling Switches use a Virtual Domain_ID to perform N_Port_ID allocations for N_Ports connected to the FCDF Set of the Distributed Switch (i.e., the Virtual Domain_ID is used as the most significant byte in all N_Port_IDs allocated to N_Ports that are attached to the FCDF Set). Using a Virtual Domain_ID to assign N_Port_IDs enables seamless operation in case of failures of one of the two redundant Controlling Switches (see T11/11-224v0). As a result, a redundant Distributed Switch typically uses three Domain_IDs: one for each Controlling Switch and one for the Virtual Domain_ID. To properly support the operations of a Virtual Domain_ID, a Controlling Switch shall have a Switch_Name to associate with the Virtual Domain_ID, 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 3 shows an example of such a configuration.

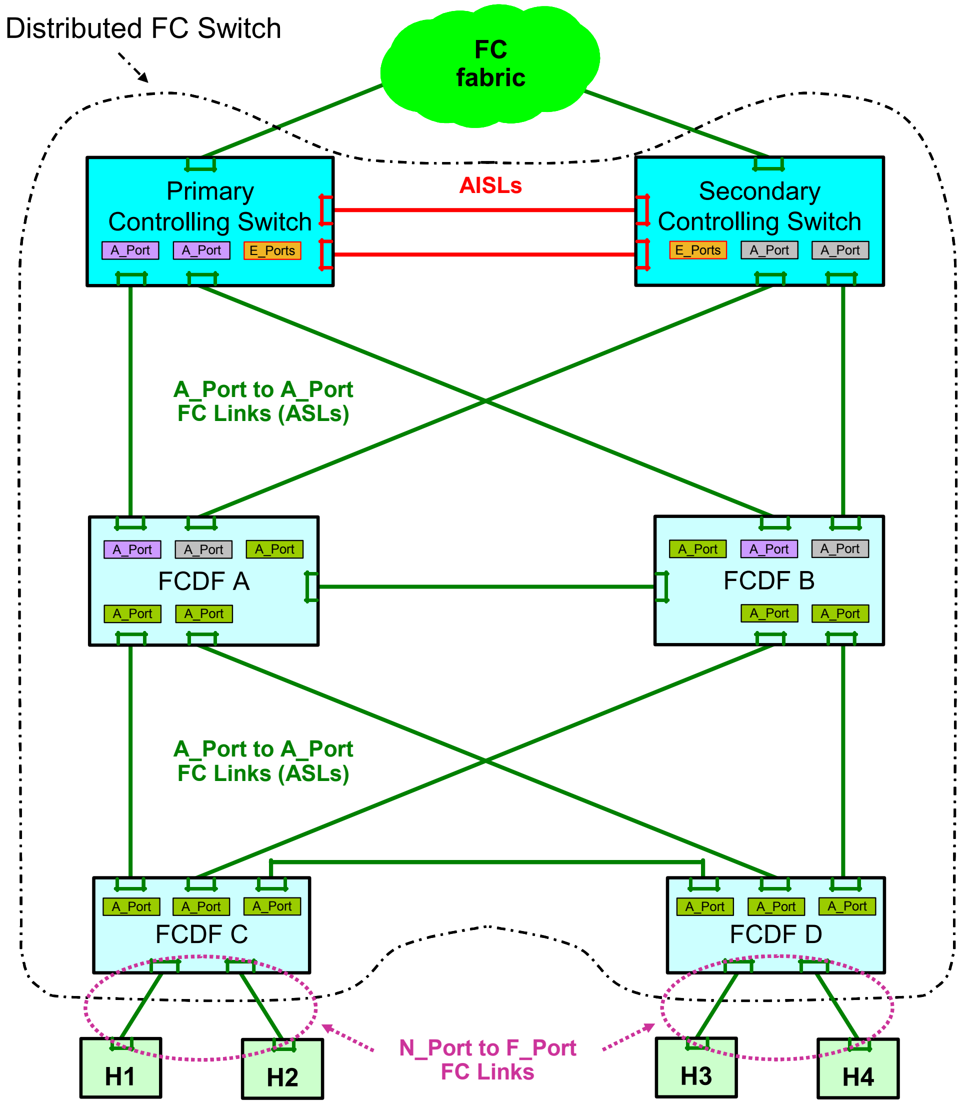

Figure 3 – Example of Distributed Switch with Cascaded FCDFs

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).

1.3 Controlling Switch Functional Model

Figure 4 shows the functional model of a Controlling Switch.

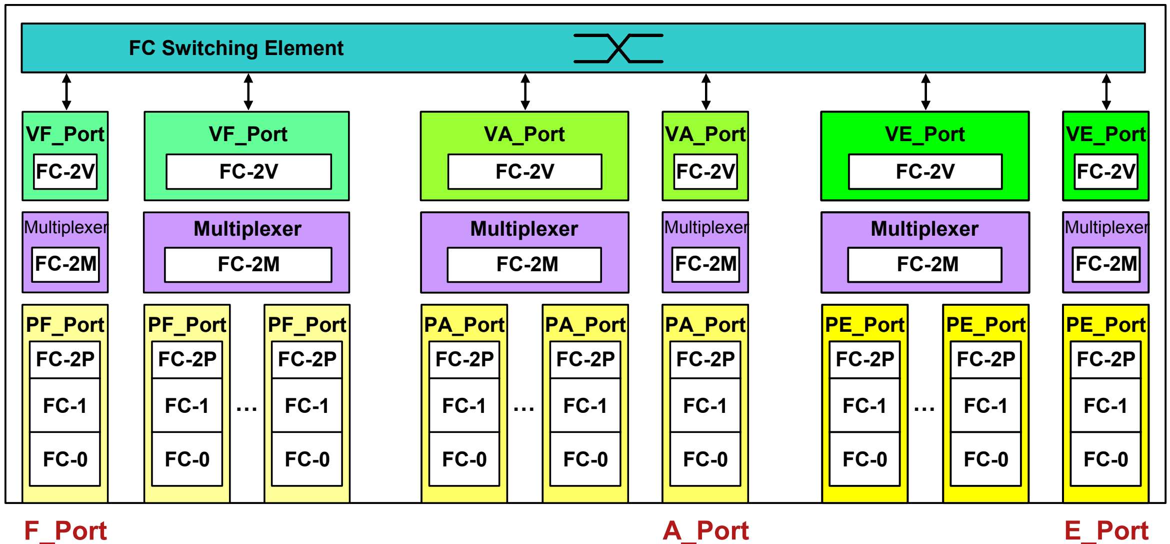

Figure 4 – Controlling Switch Functional Model

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).

1.4 FCDF Functional Model

Figure 5 shows the functional model of an FCDF.

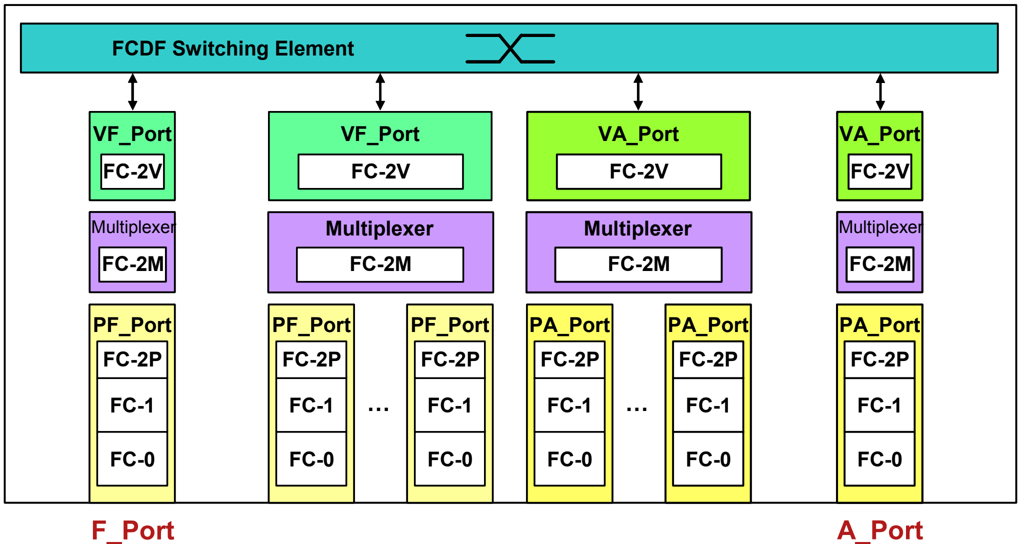

Figure 5 – FCDF Functional Model

An FCDF is a set of physical ports interconnected by an FCDF Switching Element. 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 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 6 shows the model of the FCDF Switching Element, composed by a Switch Construct, a Routing Table Update function, and a FCDF Controller function.

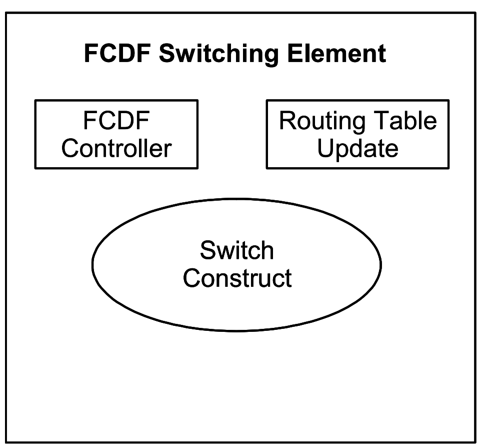

Figure 6 – FCDF Switching Element

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.

1.5 A_Port Operation

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 the Distributed 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 7.

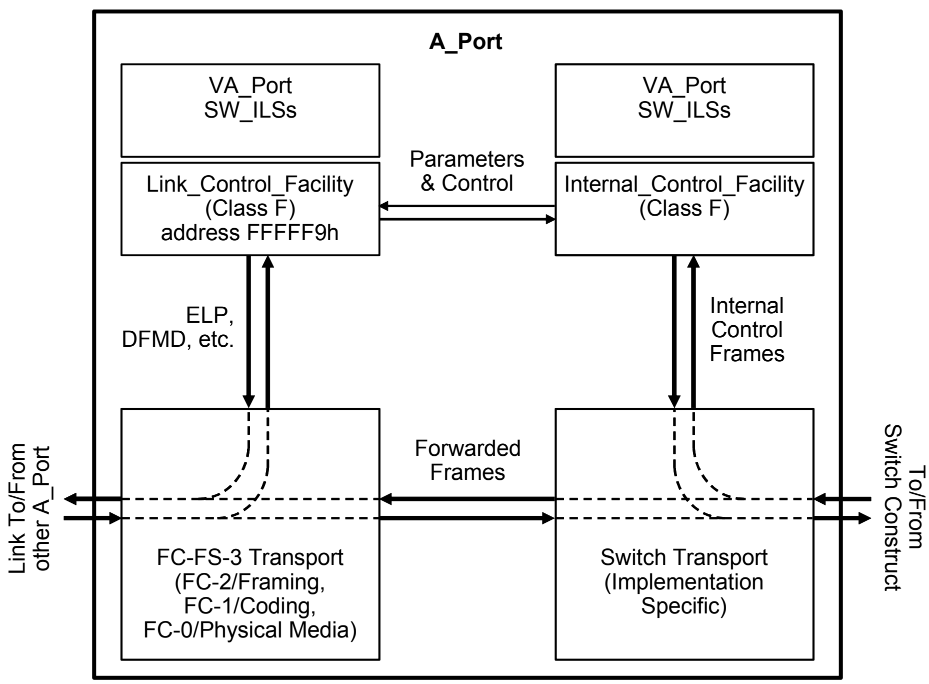

Figure 7 – A_Port Model

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.

1.6 Distributed Switch Protocols

1.6.1 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 flags in the ELP payload indicate if the originator of the ELP Request or SW_ACC is a Controlling Switch or an FCDF.

Bit 13 of the flags field of the ELP payload is the Controlling FCF/Switch flag. This bit set to one indicates that the originator of the ELP Request or SW_ACC is a Controlling Switch. This bit set to zero indicates that the originator of the ELP Request or SW_ACC is not a Controlling Switch.

Bit 12 of the flags field of the ELP payload is the FDF/FCDF flag. This bit set to one indicates that the originator of the ELP Request or SW_ACC is an FCDF. This bit set to zero indicates that the originator of the ELP Request or SW_ACC is not an FCDF.

A received ELP Request or SW_ACC having both these bits set to one is invalid and shall be ignored. Table 1 shows the meaning of the values of these bits.

Table 1 – 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, in accordance with FC-SW-5. 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 T11/11-224v0).

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 T11/11-224v0), otherwise (i.e., when the Controlling Switch is not yet operational) the Controlling Switch port shall go in 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 (see FC-SW-5). 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 T11/11-224v0).

The ports of an FCDF that has not yet received from the Primary Controlling Switch the Distributed Switch’s FCDF Set (see T11/11-225v0) 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 (see T11/11-225v0), the ports of an FCDF that have completed Link Intialization 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 reply, irrespective of the value of the Switch_Name field in the ELP Request payload. If the ELP is accepted then the FCDF Port behaves as an A_Port (i.e., an ASL is instantiated).

NOTE 1 – 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’. A port of an FCDF that received from the Primary Controlling Switch the Distributed Switch’s FCDF Set shall also reject a received ELP Request coming from a Controlling Switch other than the Controlling Switches that define its Distributed Switch, with Reason Code ‘Logical Error’ and Reason Code Explanation ‘Not Authorized’.

1.6.2 Distributed Switch Operations

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. The two Controlling Switches keep their state synchronized.

When becoming operational (i.e., when in state P2 or S2 of T11/11-224v0), a Controlling Switch instantiates ASLs with the FCDFs that are directly reachable and are part of its FCDF Set, as described in 1.6.1.

Upon instantiating an ASL with an FCDF, the Primary Controlling Switch shall initiate an FDRN Exchange (see T11/11-225v0) describing that link with the Secondary Controlling Switch 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 (see T11/11-225v0). 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). 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 (see T11/11-225v0).

Upon deinstantiating an ASL with an FCDF, the Primary Controlling Switch shall initiate an FDUN Exchange (see T11/11-225v0) describing that link with the Secondary Controlling Switch 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 Exchange with the Primary Controlling Switch, the FCDF becomes able to initiate ELP Requests to instantiate other ASLs with other FCDFs. Upon completing the NPRD Exchange with the Primary Controlling Switch, an FCDF becomes able to set up proper forwarding tables to forward FC frames inside and outside the Distributed Switch. At this point the FCDF enables its ports for logins from Nodes; any FLOGI received on a FCDF port before this point is responded by the FCDF by transmitting the NOS Primitive Sequence.

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 to keep the state synchronized. Upon completion of this FDRN 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 2 – 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 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.

Upon receiving on a port a FLOGI Request or a NPIV FDISC Request from a Node, an FCDF shall send a VNRN Request (see T11/11-225v0) 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. If the Primary Controlling Switch accepts the VNRN Request, it shall allocate to the newly reachable VN_Port an N_Port_ID from the Virtual Domain_ID, generate appropriate RSCN(s), and update the Fibre Channel Name Server. The Primary Controlling Switch shall also recompute the Zoning ACLs for the affected N_Port_IDs. The Primary Controlling Switch shall then distribute this information to the Secondary Controlling Switch and to each FCDF belonging to the Distributed Switch through an NPZD Exchange indicating an N_Port_ID allocation (see T11/11-225v0). The NPZD Requests sent to the Secondary Controlling Switch and to the FCDF that sent the VNRN Request shall include the FLOGI / NPIV FDISC LS_ACC Parameters (see T11/11-225v0); the NPZD Requests sent to the other FCDFs shall not include them. The NPZD Requests sent to the Secondary Controlling Switch shall carry zero Peering Entries. Upon receiving the NPZD Request including the FLOGI / NPIV FDISC LS_ACC Parameters, the FCDF shall accept the FLOGI Request or a FIP NPIV FDISC Request and complete the N_Port login.

When a VN_Port is logged out, an FCDF shall perform a VNUN Exchange (see T11/11-225v0) with the Primary Controlling Switch to inform it that the VN_Port is now unreachable. Upon completing a VNUN Exchange, the Primary Controlling Switch shall deallocate the N_Port_ID previously assigned to that VN_Port, generate appropriate RSCN(s), and update the Fibre Channel Name Server. The Primary Controlling Switch shall also recompute the Zoning ACLs for the affected N_Port_IDs. The Primary Controlling Switch shall then distribute this information to the Secondary Controlling Switch and to each FCDF belonging to the Distributed Switch through NPZD Exchanges indicating an N_Port_ID deallocation (see T11/11-225v0). These NPZD Requests shall not include the FLOGI / NPIV FDISC LS_ACC Parameters (see T11/11-225v0).

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_ID and distribute them to the FCDFs of the Distributed Switch through AZAD Exchanges (see T11/11-225v0).

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 its Domain_ID (i.e., not from the Virtual Domain_ID) if it accepts the received FLOGI or NPIV FDISC Request.