

CloudEngine 12800 系列交换机

FCoE 和 DCB 技术白皮书

目 录

1 FCoE 简介	1
2 FCoE 的部署方案	4
3 FCoE 原理描述	5
3.1 FCoE 的基本概念	6
3.2 FCoE 的封装	9
3.3 FIP 协议	10
3.4 FCF	13
3.5 NPV	14
3.6 FSB	15
3.7 Zone	17
4 应用场景	19
4.1 FCF 场景	20
4.2 FCF+NPV 组网	20
4.3 FCF+FSB 组网	21
5 配置任务概览	23
6 配置注意事项	24
7 缺省配置	26
8 配置 FCF	27
8.1 创建 FCoE 接口和 FCF 实例	31
8.2 配置 Zone	32
8.3 （可选）配置 FCF 参数	33
8.4 （可选）配置 FIP Keepalive 报文的发送间隔	33
8.5 检查配置结果	34
9 配置 FSB	35
9.1 配置 FSB 实例	37
9.2 配置接口角色	37
9.3 （可选）配置 FIP 协议报文的交互超时时间	39
9.4 （可选）配置 FCoE 链路同步功能	39
9.5 检查配置结果	40

10 维护 FCoE.....	41
10.1 查看 FIP 报文的统计信息.....	42
10.2 清除 FIP 报文的统计信息.....	42
10.3 查看 FIP 和 FCoE 报文的统计信息.....	42
10.4 清除 FIP 和 FCoE 报文的统计信息.....	43
10.5 监控 FCoE 的运行状态.....	43
11 配置举例.....	44
11.1 配置 FCF 示例.....	45
11.2 配置 FCF+FSB 示例.....	48
12 DCB 简介.....	53
13 DCB 原理描述.....	54
13.1 PFC.....	55
13.2 ETS.....	57
13.3 DCBX.....	59
14 应用.....	63
15 配置任务概览.....	64
16 配置注意事项.....	65
17 缺省配置.....	66
18 配置 PFC 功能.....	67
19 配置 ETS 功能.....	69
19.1 配置 ETS 模板.....	70
19.2 应用 ETS 模板.....	71
19.3 检查配置结果.....	71
20 配置 DCBX 功能.....	72
21 维护 DCB.....	74
21.1 监控 DCB 的运行状态.....	75
21.2 清除 DCB 的统计信息.....	75
22 配置举例.....	76
22.1 配置 DCB 示例.....	77

1 FCoE 简介

介绍FCoE的定义和作用。

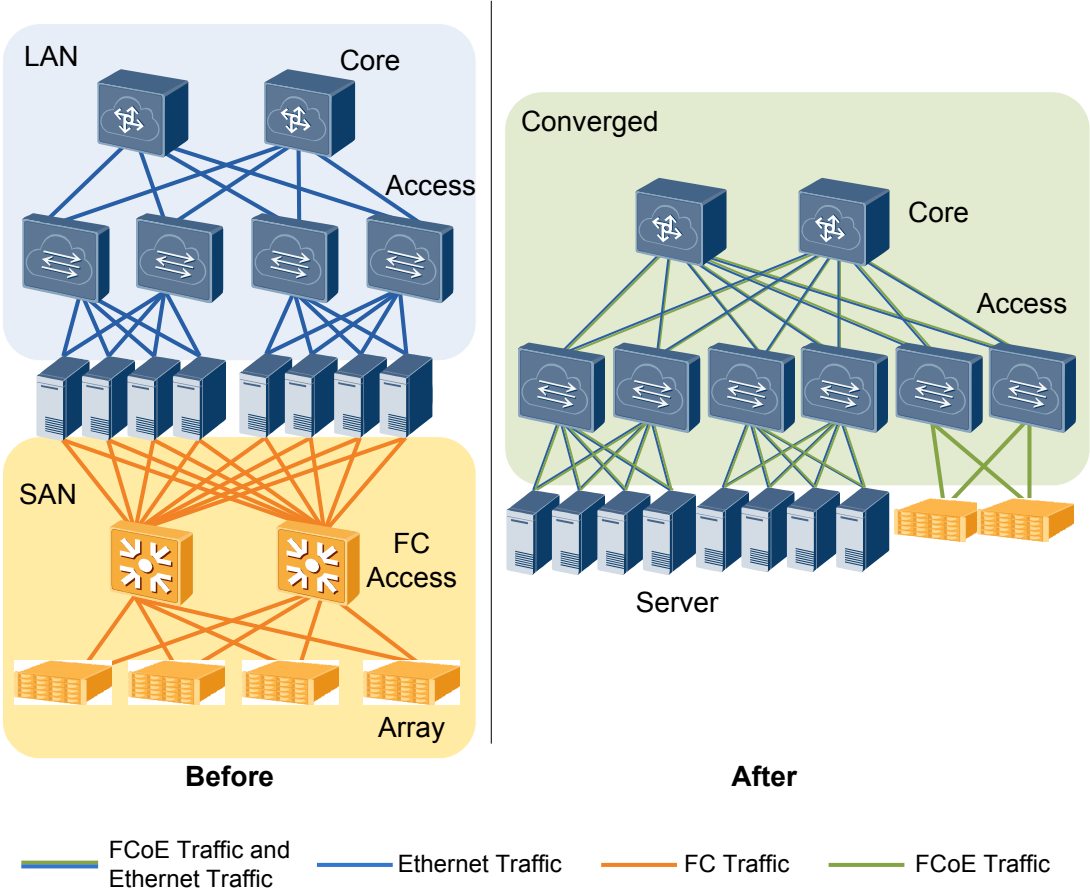
数据中心融合网络的发展趋势

如图1-1所示，传统数据中心组网中，以太网LAN（Local Area Network）用于服务器与服务器、客户端与服务器之间通信，存储区域网络SAN（Storage Area Network）用于服务器与存储设备之间通信，两者的部署和维护相互独立。

随着数据中心的飞速发展和服务器数量的激增，LAN和SAN的独立部署存在如下问题：

- 网络复杂：LAN和SAN的相互独立导致业务部署的灵活性差，网络扩展困难，网络的维护和管理成本高。
- 能效比低：服务器上至少配置4~6块网卡，包括用于接入LAN的网络接口卡NIC（Network Interface Card）和用于接入SAN的主机总线适配器HBA（Host Bus Adapter）。服务器的多类型的网卡使得整个数据中心的电力消耗和冷却成本增加。

图 1-1 数据中心网络融合前后对比



数据中心网络融合后，存储网络SAN和以太网LAN通过网络融合技术可共享同一个单一的、集成的网络基础设施，解决了不同类型网络共存所带来的问题，实现了网络基础设施整合、精简的目标。

本文所采用的网络融合技术是指以太网光纤通道FCoE（Fibre Channel over Ethernet）。

定义

FCoE是由美国国家标准委员会ANSI（American National Standards Institute）下属的T11（Technical Committee 11）委员会定义的一种融合网络技术，是以光纤通道FC（Fibre Channel）存储协议为核心的I/O整合方案。

目的

对比SAN和LAN网络模型可知，当网络融合后：

- FC流量无法正常转发。
- 现有以太网无法达到FC网络中无丢包转发的需求。

为解决上述问题，以太网光纤通道FCoE和数据中心桥接DCB（Data Center Bridging）应运而生。其中：

- FCoE实现了以太网帧承载FC帧，从而使得LAN和SAN能够共享网络资源，实现网络融合。

- DCB实现了在数据中心网络中构建一个无丢包以太网，使得传统以太网实现FC SAN网络中的拥塞控制，为FCoE融合业务提供了传输质量保证。

受益

FCoE可以为数据中心带来如下受益：

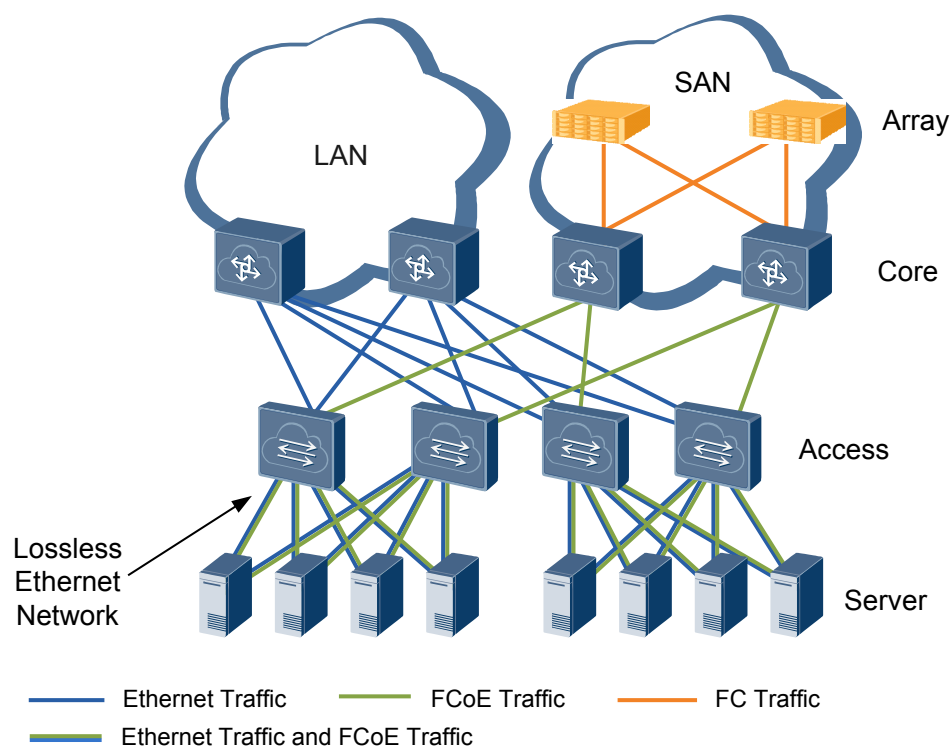
- 更低的总体拥有成本TCO（Total Cost of Ownership）：LAN和SAN网络通过FCoE技术共享网络资源，整合并更有效的利用以前分散的资源，减少对于SAN网络基础设施的投资，简化了网络复杂度，降低网络的管理和维护成本；服务器采用融合网络适配器CNA（Converged Network Adapter），减少数据中心的电力和冷却成本。
- 强大的投资保护：FCoE可以和数据中心现有的以太网及FC基础设施无缝互通，保护客户在现有FC SAN上的投资（如FC设备、工具成本和管理成本）。
- 增强的业务灵活性：FCoE使得所有的服务器共享存储资源，满足虚拟机迁移的需求，这样也提高了系统的灵活性和可用性。

2 FCoE 的部署方案

介绍FCoE在数据中心场景中的部署方案。

如图2-1所示，接入层部署FCoE实现了服务器I/O整合，简化服务器网络接入层的线缆设施。服务器上CNA支持FCoE功能，并连接到接入层交换机，接入层交换机再分别通过以太网链路连接到现有的LAN和SAN。

图 2-1 接入层部署 FCoE



3 FCoE 原理描述

关于本章

介绍FCoE实现原理。

[3.1 FCoE的基本概念](#)

[3.2 FCoE的封装](#)

[3.3 FIP协议](#)

[3.4 FCF](#)

[3.5 NPV](#)

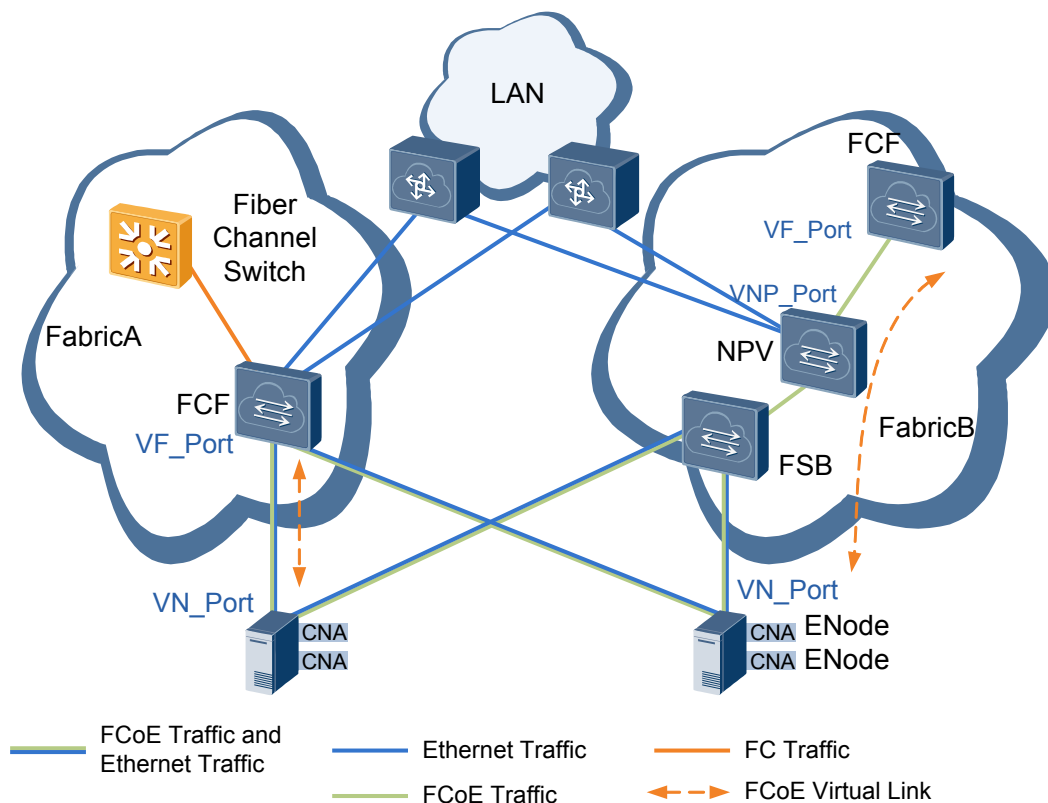
[3.6 FSB](#)

[3.7 Zone](#)

3.1 FCoE 的基本概念

如图3-1所示的FCoE组网中，FCoE中存在如下基本概念：**ENode**、**FCF**、**FIP协议**、**NPV**、**FSB**、**Fabric**、**FCoE虚链路（FCoE Virtual Link）**、**端口角色**、**FCoE VLAN**、**Zone**、**WWN**、**FC_ID**、**Domain ID**和**FC-MAP**。

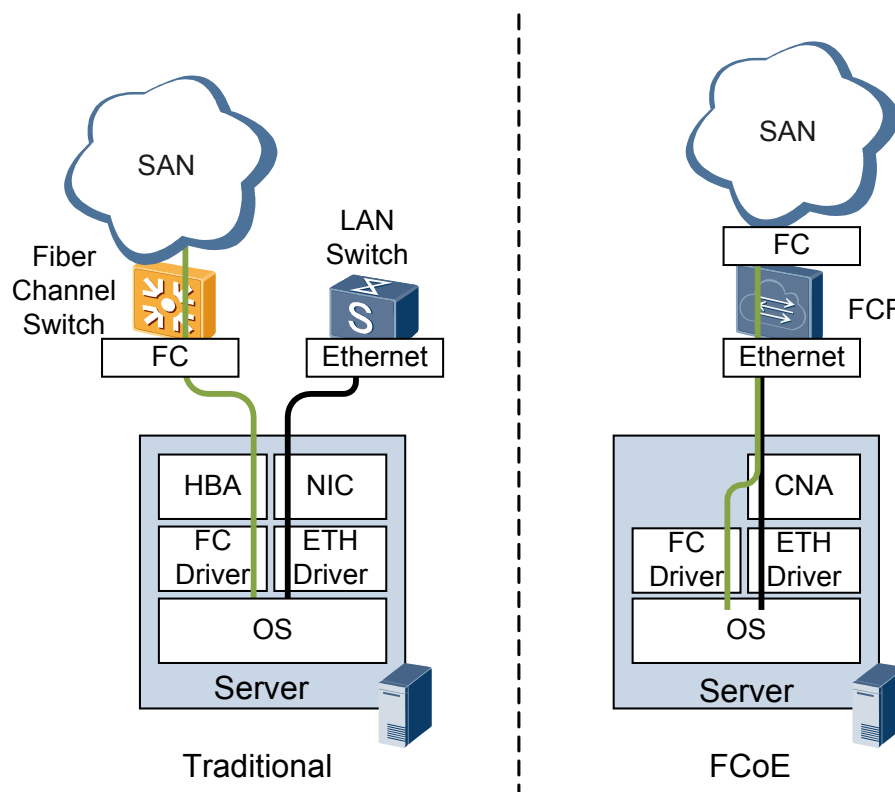
图 3-1 FCoE 组网



- ENode

ENode是指服务器上同时支持FCoE协议栈和FC协议栈的CNA。如图3-2所示，传统的服务器中，存在两块网卡：用于接入LAN的NIC和用于接入SAN的HBA。而ENode中的CNA替代了NIC和HBA的功能，在完成以太网数据转发的同时也完成FCoE报文的上层处理和FCoE封装/解封装。

图 3-2 传统服务器和 FCoE 服务器的区别



- FCF

FCoE数据转发器FCF（FCoE Forwarder）是指同时支持FCoE协议栈和FC协议栈的双协议栈交换机。FCF能够转发FCoE报文，主要用于连接传统SAN网络和LAN网络。

- FIP协议

FCoE初始化协议FIP(FCoE Initialization Protocol)主要用于FCoE网络中FC终端发现、Fabric登录和FCoE虚链路建立的二层协议。通过FIP协议，ENode可登录到Fabric，实现和目标FC设备通信。FIP协议同时也维护FCoE虚链路。

FIP协议的详细内容请参见[FIP协议](#)。

- NPV

NPV（NPort Virtualization）交换机位于Fabric网络边缘，处于节点设备与FCF之间，将节点设备的流量转发到FCF交换机。

- FSB

FSB（FIP Snooping Bridge）是指运行FIP Snooping功能的交换机，交换机本身不支持FC协议栈。交换机利用FIP Snooping功能侦听FIP协议，控制FCoE虚链路的建立，预防恶意攻击。

- Fabric

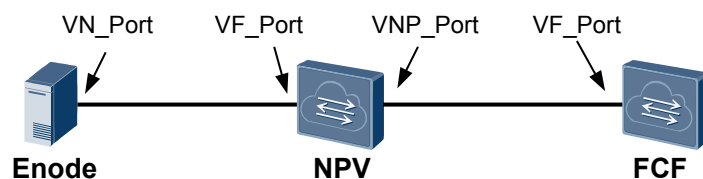
Fabric指网络节点通过一台或多台交换机互联的网络拓扑结构。

- FCoE虚链路（FCoE Virtual Link）

FCoE虚链路是指连接FCoE链路终端设备（如ENode和FCF）之间点到点的逻辑链路。因为FCF和ENode之间可能通过无丢包以太网连接起来，此时FCF和ENode之间的点到点关系将被破坏，由此引入FCoE虚链路。

- 端口角色

图 3-3 端口角色



在传统FC网络中，FC设备之间通过FC端口进行交互。其中FC端口分为N_Port和F_Port。

- N_Port (Node Port)：是指FC主机（如服务器或存储设备）和FC交换机之间的链路上FC主机侧的端口。
- F_Port (Fabric Port)：是指FC主机和FC交换机之间的链路上FC交换机侧的端口，主要是为FC主机提供Fabric接入服务。
- NP_Port(N_Port Proxy)：是指NPV交换机和FCF交换机之间的链路上NPV交换机侧的端口。

FCoE保留FC协议中的端口角色的概念，即在ENode和FCF之间的FCoE虚链路上，ENode侧的端口为VN_Port，FCF侧的端口为VF_Port。在NPV和FCF直连的链路上，NPV侧的端口为VNP_Port，FCF侧的端口为VF_Port。

- FCoE VLAN

FC-BB-5协议规定，FCoE的报文需要在特定的VLAN中进行转发。在FC协议栈中，FC设备支持多个VSAN（类似以太网中的VLAN），运行在不同的VSAN中的FC流量经过FCoE封装时需要用不同的FCoE VLAN来区分。这样做的好处是简化了处理，FCoE报文不携带VSAN的信息，仅携带FCoE VLAN的信息即可，通过FCoE VLAN区分不同的VSAN。

说明

- 虚链路仅存在于同一个FCoE VLAN中。
- FCoE VLAN仅承载FCoE流量，而不承载任何以太网流量（如IP流量）。
- Zone
创建Zone，并在Zone内根据不同的目的添加不同的N_Port成员，使不同Zone内的N_Port成员之间相互隔离，以达到访问控制的目的。
Zone set是Zone的集合，Zone是Zone member的集合。Zone member是Zone的成员，每个成员都是一个N_Port，称为N_Port成员。
- WWN
全球名字WWN（World Wide Name），用来标识Fabric网络和SAN网络中的实体，其中WWNN（World Wide Node Name）用来标识节点设备，WWPN（World Wide Port Name）用来标识设备端口。SAN网络中每个实体的WWN在设备出厂前就已分配好了。
- FC_ID
FC_ID又称为FC地址，在SAN网络中，FC协议通过FC地址访问网络中各个实体。一个FC地址可以唯一标识一台节点设备上的一个N_Port。
- Domain_ID

Domain_ID又称为域ID，在SAN网络中，一个域ID可以唯一标识一台FC交换机，在报文传输时，FC交换机之间的路由和转发使用的都是域ID。

- FC-MAP
FCoE帧在转发过程中使用的是本地唯一MAC地址（即仅在本地以太网内是唯一的），本地唯一MAC地址由FCF分配给ENode或者FCF响应使用ENode自身指定的本地唯一MAC地址。FCF分配给ENode本地唯一MAC地址的方式称为FPMA（Fabric Provided MAC Address）。FPMA实际上是将FC_ID加上一个24比特FCoE MAC地址前缀FC-MAP（FCoE MAC Address Prefix）的过程。

3.2 FCoE 的封装

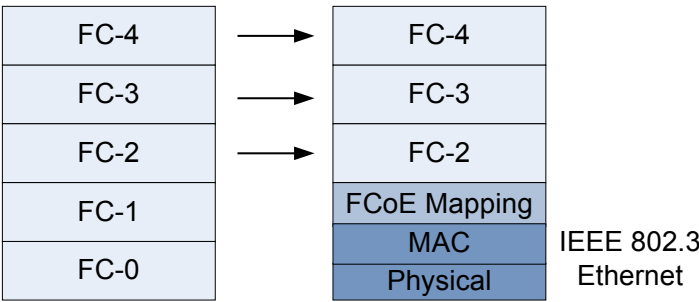
FCoE通过将FC帧封在普通以太帧中，从而实现FC流量在以太网中传输的功能。从FC协议的角度看，FCoE仅仅是将FC流量承载在另一种链路上传输；从以太网协议的角度看，FCoE仅仅是以太帧承载不同的上层协议。

FCoE 的协议栈

如图3-4所示，在传统的FC协议中，FC协议栈共分为5层：

- FC-0定义承载介质类型
- FC-1定义帧编解码方式
- FC-2定义分帧协议和流控机制
- FC-3定义通用服务
- FC-4定义上层协议到FC的映射

图 3-4 从 FC 到 FCoE 的映射关系

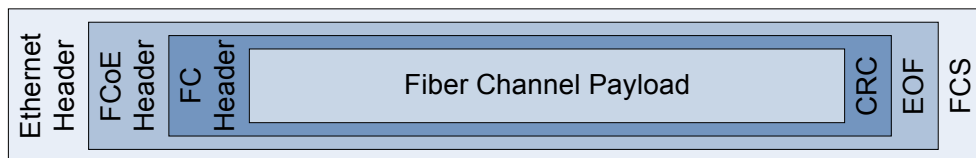


如图3-4所示，在FCoE协议栈中，FC-0和FC-1被映射成为IEEE 802.3 Ethernet协议的Physical和MAC，并添加了FCoE Mapping作为上层FC协议栈与底层Ethernet协议栈之间的适配层。

报文封装

FCoE协议将一个完整的FC帧封装在以太帧中，封装形式如图3-5所示。

图 3-5 FCoE 的报文封装



其中：

- Ethernet Header中指定了报文的源、目的MAC地址、以太帧类型和FCoE VLAN。
- FCoE Header指定了FCoE帧的版本号和控制信息。
- FC Header和传统FC帧相同，即指定了FC帧的源、目的地址等信息。

说明

在Ethernet Header中，以太帧类型字段取值为FCoE_TYPE(8906h)。

3.3 FIP 协议

基本原理

FCoE初始化协议FIP（FCoE Initialization Protocol）是FCoE的控制协议，用于成对FCoE设备（如ENode和FCF）之间FCoE虚链路的建立和维护。

虚链路建立阶段：

1. FIP发现FCoE VLAN和对端的FCoE虚接口。
2. FIP执行FCoE虚链路的初始化功能，如：FLOGI（Fabric Login）和FDISC（Fabric Discovery）。

当虚链路建立完成后，FIP协议还要履行FCoE虚链路的维护功能：

- 周期性检测FCoE虚链路两端的FCoE虚接口是否可达。
- FCoE虚链路的拆除FLOGO（Fabric Logout）。

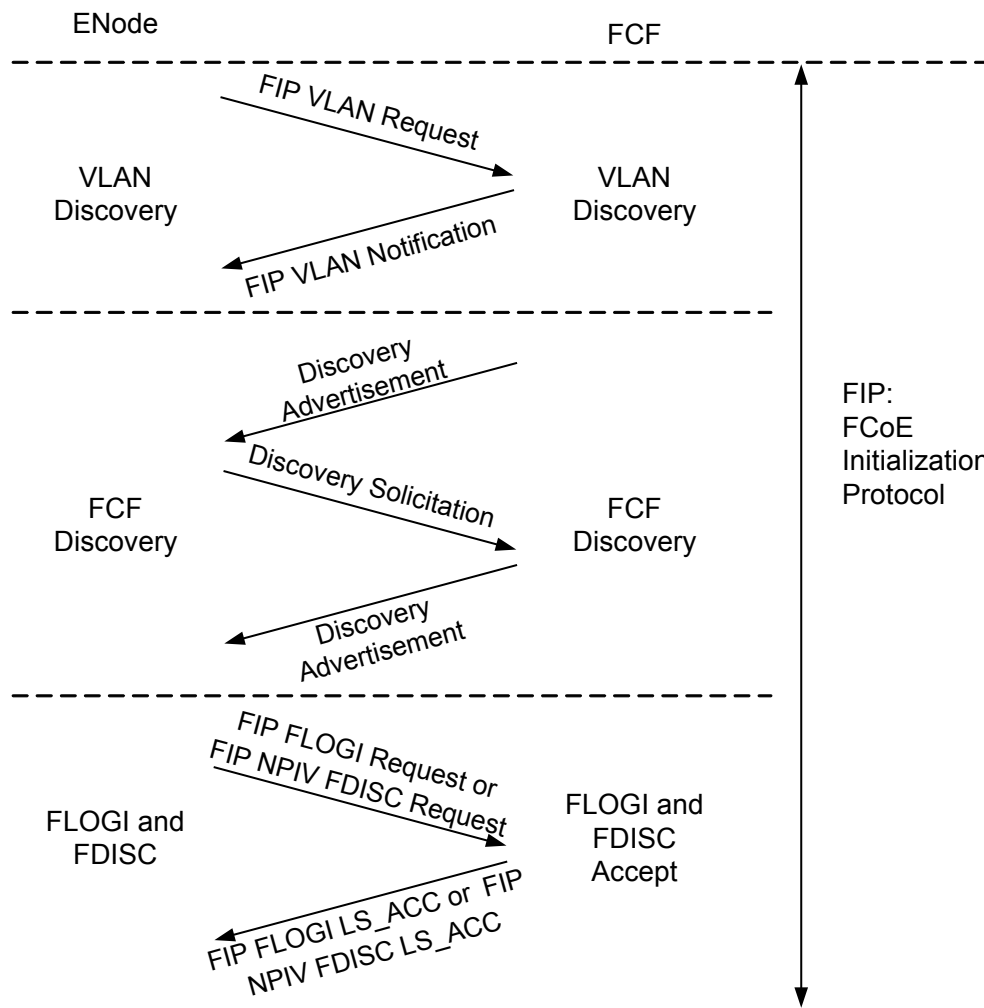
虚链路的建立

ENode和FCF之间的FCoE虚链路的建立过程如图3-6所示，其中用于建立虚链路的是FIP帧，当虚链路建立后传输的是FCoE帧。

说明

- FIP协议中，所有的协议报文均由ENode端发起，而FCF也会主动发出未被请求的FIP通告报文，将在FIP FCF Discovery说明。
- FIP帧和FCoE帧的以太帧类型（Type字段）和封装形式不同，因为FCoE是在传统的FC帧的基础上进行封装，而FIP在传统FC协议中没有对应的定义。

图 3-6 FCoE 虚链路的建立过程



虚链路的建立主要分为三个阶段，分别是**FIP VLAN Discovery**、**FIP FCF Discovery**以及**FIP FLOGI和FDISC**。其中FIP FLOGI和FIP FDISC的处理流程与传统FC协议中的FLOGI和FDISC类似。

1. FIP VLAN Discovery

FIP VLAN Discovery阶段主要功能是寻找用于承载FCoE帧的FCoE VLAN，使ENode发现所有潜在的FCoE VLAN，但是此时ENode并不能对FCF进行选择。

FIP VLAN Discovery的具体流程如下：

- ENode首先发送FIP VLAN Discovery的请求报文FIP VLAN Request。其中报文的目的MAC地址是一个组播地址，这个地址又被称为ALL-FCF-MAC（01-10-18-01-00-02），所有的FCF都会侦听这个组播地址的报文。
- 所有的FCF都能通过普通VLAN向ENode反馈一个或者多个FCoE VLAN，这些FCoE VLAN可被用于ENode的登录。

FC-BB-5协议规定，FIP VLAN Discovery是可选的。FCoE VLAN既可通过管理员手工配置，也可以通过FIP VLAN Discovery动态发现。

2. FIP FCF Discovery

FIP FCF Discovery主要用于ENode发现可以登录的FCF。

FIP FCF Discovery的具体流程如下：

- a. FCF周期性的在FCoE VLAN中发送Discovery Advertisement报文。报文的目的地MAC地址为组播地址ALL-ENode-MAC（01-10-18-01-00-01），以便所有ENode都能侦听到。报文携带的信息包括FCF的MAC地址、虚链路参数（包括FIP超时时间、FCF优先级等）。
- b. ENode从接收到的Discovery Advertisement报文中获取到可供登录的FCF信息，从中选择优先级最高的FCF后向其发送单播的Discovery Solicitation报文。
- c. FCF接收到Discovery Solicitation报文后，回应单播Discovery Advertisement报文，允许ENode登录。

除了接收周期性的Discovery Advertisement报文，新入网的ENode通常不需要等待所有的FCF发送Discovery Advertisement报文，协议规定ENode可向所有的FCF发送Discovery Solicitation报文，报文的目的地地址是ALL-FCF-MAC，FCF收到请求报文之后响应一个单播Discovery Advertisement报文给ENode。ENode从接收到的Discovery Advertisement报文中选择优先级较高的FCF与其建立虚链路。

3. FIP FLOGI和FDISC

当ENode选择一个FCF登录后，就会通过FIP FLOGI或FIP FDISC报文与该FCF的VF_Port建立虚链路，以便通过该虚链路传送FCoE帧。FIP FLOGI报文和FIP FDISC报文都是单播报文，分别替代FC协议中的FLOGI和FDISC报文，均用来为ENode分配MAC地址，以便其登录到Fabric。

FIP FLOGI和FIP FDISC的处理过程类似，两者唯一区别是：FIP FLOGI指ENode首次登录Fabric时建立虚链路的过程；而FIP FDISC指，当ENode上存在多个VM时，为每个VM建立虚链路的过程。下面以FIP FLOGI为例来进行介绍。

FIP FLOGI的具体流程如下：

- a. ENode向FCF发送请求报文FIP FLOGI Request。
- b. FCF响应ENode请求，并为ENode分配本地唯一MAC地址FPMA（Fabric Provided MAC Address）或者FCF响应使用ENode自身指定本地唯一MAC地址SPMA（Server Provided MAC Address）。

FCoE中的MAC地址

FCoE帧和FIP帧在转发过程中所使用的MAC地址是不同的，FIP帧在进行封装时用的是CNA在生产过程中指定的全球唯一MAC地址（又称ENode MAC地址），而FCoE帧则使用的是本地唯一MAC地址（即仅在本地以太网内是唯一的），本地唯一MAC地址由FCF分配给ENode或者FCF响应使用ENode自身指定的本地唯一MAC地址，分配过程参见FIP FLOGI和FDISC。

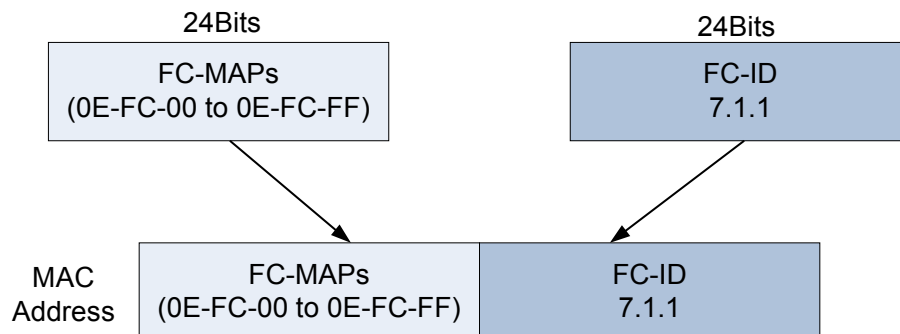
本地唯一MAC地址的分配方式为FPMA（Fabric Provided MAC Address）。

如图3-7所示，FPMA实际上是将FC_ID加上一个24比特FCoE MAC地址前缀FC-MAP（FCoE MAC Address Prefix）的过程。在协议FC-BB-5中为减轻FCoE的调度负担定义了256个FC-MAP。在大多数情况下，使用缺省的FC-MAP（0E-FC-00）即可满足用户需求。但当不同的Fabric或不同VSAN映射到同一个以太网VLAN中时，FC_ID可能会重叠，此时可采用不同的FC-MAP来解决上述问题。

说明

建议用户将一个FC Fabric映射到同一个以太网VLAN中。如果必须将多个FC Fabric运行在同一个以太网中，建议用户将其映射到不同的VLAN中。

图 3-7 FPMA 的地址结构



虚链路的维护

虚链路的监控

在传统FC网络中，如果物理链路发生故障，则FC协议可以马上感知到。但是在FCoE协议中，由于采用了以太封装，FC协议无法感知链路层的故障，为此FIP提供了一种简单的Keepalive机制。

虚链路的监控机制如下：

- ENode周期的向FCF发送ENode FIP Keepalive通告报文。如果FCF在2.5倍周期时间内没有收到已登录的ENode发送的ENode FIP Keepalive通告报文，则FCF认为该虚链路发生故障并断开该虚链路。
- FCF周期的以ALL-ENode-MAC为目的地址向所有的ENode发送组播Discovery Advertisement报文。如果ENode在2.5倍周期时间内没有收到FCF发送的组播Discovery Advertisement报文，则ENode认为该虚链路发生故障并断开该虚链路。

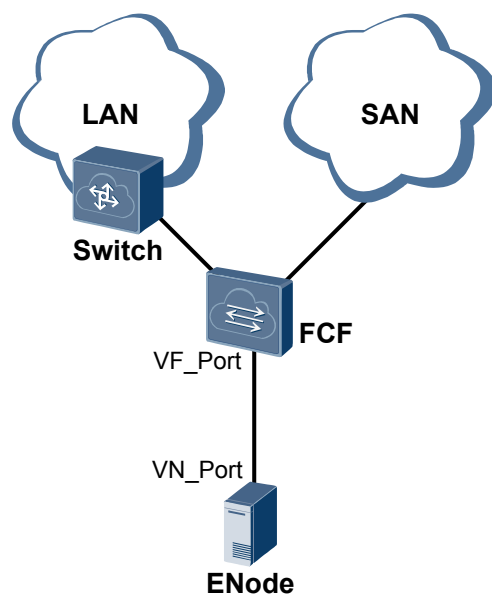
虚链路的拆除

如果FCF没有收到ENode发送的ENode FIP Keepalive通告报文，则向相应的ENode发送FIP Clear Virtual Link报文用于链路拆除。如果当ENode退出登录时，ENode也可通过向FCF发送Fabric Logout请求来删除虚链路。

3.4 FCF

FCF交换机同时支持FCoE协议栈和FC协议栈，使用以太帧封装FC报文的方式，实现了用FCoE虚链路替代物理的FC链路，主要用于连接传统SAN网络和LAN网络。

图 3-8 FCF 组网



如图3-8所示，FCoE虚链路连接ENode的VN接口到FCF交换机的VF接口。每台FCF交换机都会被分配一个域ID，而每个SAN网络最多支持239个域ID，因此也就限制了一个SAN网络中最多只能有239台交换机。

3.5 NPV

在SAN网络中，对于直接连接节点设备的边缘交换机的需求量很大，NPV交换机可以使边缘交换机突破一个SAN网络中239台最大交换机数目的限制。

说明

CE12800不支持NPV功能。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/936155004042010102>