以太网 OAM 技术白皮书

目 录

1 概述

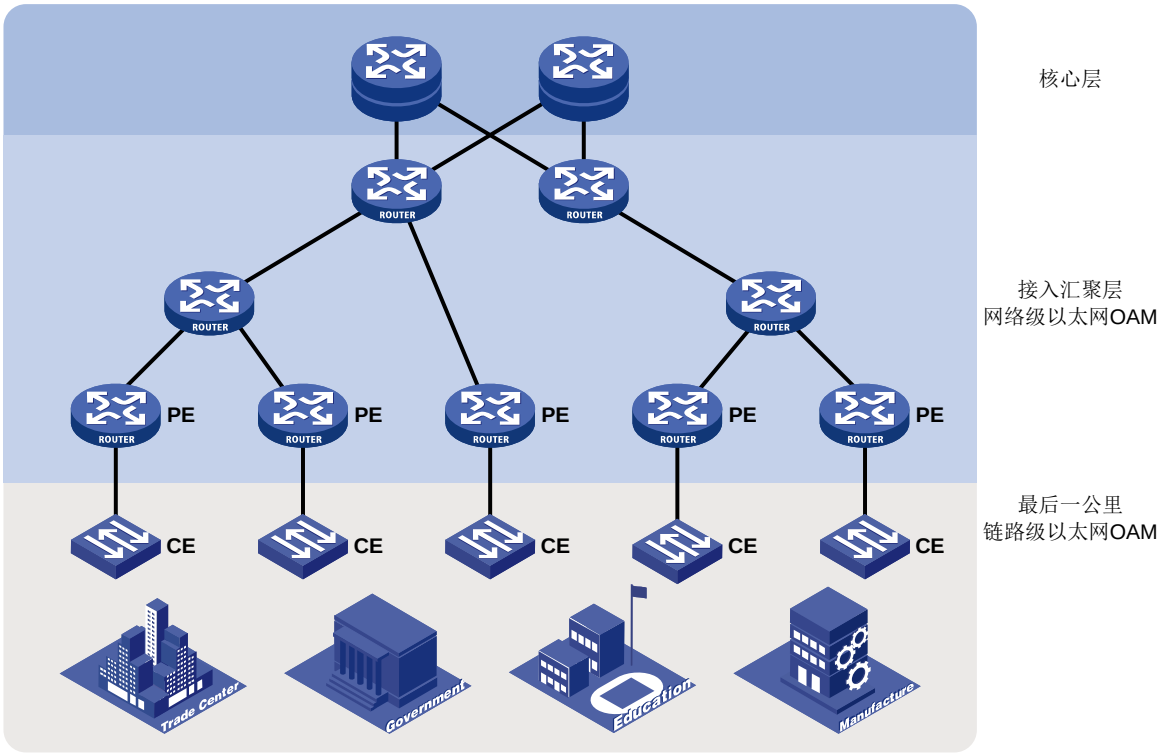
1.1 产生背景

以太网技术自诞生起，就以其简单易用和价格低廉的特点逐步成为局域网的主导技术。近年来，随着千兆、万兆以太网技术的相继应用，也促使网络运营商、设备制造商和标准化组织致力于将以太网技术向城域网和广域网领域推进。

以太网最初为局域网而设计，由于局域网本身已具备较高的可靠性和稳定性，因此在设计以太网之初并未建立管理维护的机制。而相对于局域网，城域网和广域网在链路长度和网络规模上都迅速扩大，有效管理维护机制的缺乏，已成为以太网技术在城域网和广域网应用的严重障碍。为此，在以太网上实现 OAM（Operation, Administration and Maintenance，操作、管理和维护）机制成为必然的发展趋势。以太网 OAM 技术可以有效提高对以太网的管理和维护能力，保障网络的稳定运行。

1.2 以太网OAM技术分级

图1 以太网 OAM 技术分级实现示意图



以太网 OAM 技术是分级实现的。如图 1所示，以太网 OAM 技术分为以下两个级别：

- 链路级以太网 OAM 技术：多应用于网络的 PE 设备—CE 设备—用户设备之间（也叫最后一公里）的以太网物理链路，用于监测用户网络与运营商网络之间的链路状态。

- 网络级以太网 OAM 技术：多应用于网络的接入汇聚层，用于监测整个网络的连通性、定位网络的连通性故障。

各级别上典型的以太网 OAM 协议如表 1所示。

表1 典型的以太网 OAM 协议

协议名称	应用级别	协议标准	说明
EFM OAM（Ethernet in the First Mile OAM，最后一公里以太网OAM）	链路级	IEEE 802.3ah	针对两台直连设备间的链路，提供链路性能监测、故障侦测和告警、环路测试等功能
CFD（Connectivity Fault Detection，连通错误检测）	网络级	IEEE 802.1ag ITU-T Y.1731	主要用于在二层网络中检测链路连通性，以及在故障发生时确认故障并定位

本文将对 EFM OAM 和 CFD 分别进行介绍。

2 EFM OAM 技术实现

2.1 概念介绍

2.1.1 OAM 实体

使能了 EFM OAM 功能的接口称为 EFM OAM 实体，简称 OAM 实体。

2.1.2 协议报文

EFM OAM 工作在数据链路层，其协议报文被称为 OAMPDU（OAM Protocol Data Units，OAM 协议数据单元）。EFM OAM 通过在设备之间定时交互 OAMPDU 来报告链路状态，使网络管理员能够对网络进行有效的管理。

图2 OAMPDU 报文格式示意图

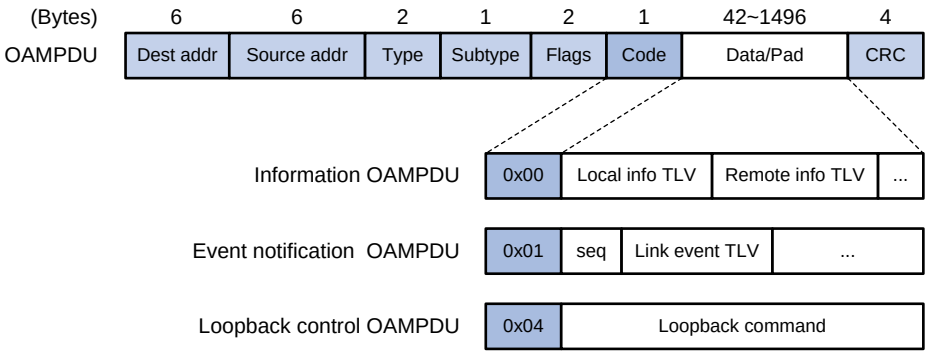


图 2所示为 OAMPDU 的报文格式和常见的 OAMPDU，OAMPDU 中重要字段的含义如表 2所示。

表2 OAMPDU 重要字段含义

字段	含义
Dest addr	目的MAC地址，为慢速协议组播地址：0x0180-C200-0002。慢速协议报文的特点是不能被网桥转发，因此无论是否具备OAM功能或OAM功能是否激活，EFM OAM报文都不能跨多跳转发
Source addr	源MAC地址，为发送端接口的MAC地址（若没有则采用该设备的桥MAC地址），是一个单播MAC地址
Type	协议类型，为0x8809
Subtype	协议子类型，为0x03
Flags	Flag域，包含了EFM OAM实体的状态信息
Code	消息编码，不同取值表示不同类型的OAMPDU，常见的OAMPDU如表3所示

表3 常见的 OAMPDU

Code 值	报文类型	中文含义	作用
0x00	Information OAMPDU	信息OAMPDU，也称为心跳报文	用于将OAM实体的状态信息（包括本地信息、远端信息和自定义信息）发给远端OAM实体，以保持以太网OAM连接
0x01	Event Notification OAMPDU	事件通知OAMPDU	一般用于链路监控，对连接本端和远端OAM实体的链路上所发生的故障进行告警
0x04	Loopback Control OAMPDU	环回控制OAMPDU	主要用于远端环回控制，用来控制远端设备的OAM环回状态，该报文中带有使能或去使能环回功能的信息，根据该信息开启或关闭远端环回功能

2.1.3 连接模式

EFM OAM 的连接模式有两种：主动模式和被动模式。EFM OAM 连接只能由主动模式的 OAM 实体发起，而被动模式的 OAM 实体只能等待远端 OAM 实体的连接请求；同处于被动模式下的两个 OAM 实体之间无法建立 EFM OAM 连接。这两种模式下设备的处理能力如表 4所示。

表4 两种工作模式下设备的处理能力

处理能力	主动模式	被动模式
初始化EFM OAM Discovery过程	可以	不可以
响应EFM OAM Discovery初始化过程	可以	可以
发送Information OAMPDU	可以	可以
发送Event Notification OAMPDU	可以	可以
发送不携带TLV的Information OAMPDU	可以	可以
发送Loopback Control OAMPDU	可以	不可以

处理能力	主动模式	被动模式
响应Loopback Control OAMPDU	可以，但需要远端为主动模式	可以

2.1.4 链路事件

EFM OAM 中定义的链路事件分为一般链路事件和紧急链路事件两大类。

1. 一般链路事件

一般链路事件用于链路性能监控，其包含的类型如表 5所示。

表5 一般链路事件

事件类型	描述
错误信号事件（Errored Symbol Event）	以收到设定数量的信号为检测窗口，在窗口期内检测到的错误信号数量如果达到或超过了检测阈值，就产生一次错误信号事件
错误帧事件（Errored Frame Event）	以设定的时间为检测窗口，在窗口期内检测到的错误帧数量如果达到或超过了检测阈值，就产生一次错误帧事件
错误帧周期事件（Errored Frame Period Event）	以收到设定数量的帧为检测窗口，在窗口期内检测到的错误帧数量如果达到或超过了检测阈值，就产生一次错误帧周期事件
错误帧秒事件（Errored Frame Seconds Event）	以设定的时间为检测窗口，在窗口期内检测到的错误帧秒（在某一秒内检测到至少一个错误帧，就称该秒为错误帧秒）数量如果达到或超过了检测阈值，就产生一次错误帧秒事件

2. 紧急链路事件

紧急链路事件用于远端故障检测，其包含的类型以及对应的 Information OAMPDU 发送频率如表6所示。

表6 紧急链路事件

事件类型	描述	OAMPDU 发送频率
链路故障（Link Fault）	远端链路信号丢失	每秒发送一次
致命故障（Dying Gasp）	不可预知的状态发生，比如电源中断	不间断发送
紧急事件（Critical Event）	不能确定的紧急事件发生	不间断发送

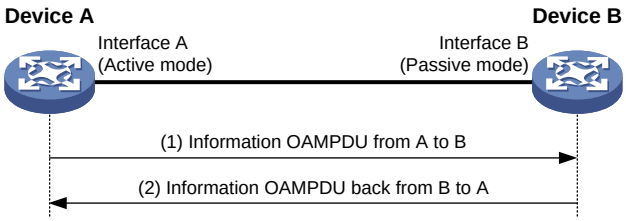
2.2 运行机制

2.2.1 EFM OAM 连接建立

EFM OAM 功能的实现建立在 EFM OAM 连接的基础之上，EFM OAM 连接的建立过程也称为 Discovery 阶段，即本端 OAM 实体发现远端 OAM 实体、并与之建立稳定对话的过程。

当设备的某个接口使能了 EFM OAM 功能时，如果该接口的 EFM OAM 工作模式为主动模式，便由该接口向远端发起 EFM OAM 连接。在建立 EFM OAM 连接的过程中，相连的 OAM 实体通过交互 Information OAMPDU 通报各自的 EFM OAM 配置信息。当 OAM 实体收到远端的配置参数后，决定是否建立 EFM OAM 连接。

图3 EFM OAM 连接示意图



如图 3所示，Device A 的接口 Interface A 工作在主动模式下，该接口上使能了 EFM OAM 功能后：

(1) Device A 向 Device B 发送 Information OAMPDU，其中包含 Device A 的 EFM OAM 配置信息。

(2) Device B 收到该 OAMPDU 后，与自己的 EFM OAM 配置进行匹配，然后向 Device A 回复 Information OAMPDU，其中除了包含 Device A 和 Device B 二者的 EFM OAM 配置信息外，还包含 Device B 对 Device A 的 EFM OAM 配置是否匹配的标志信息。

(3) Device A 收到 Device B 发来的 OAMPDU 后，判断 Device B 的 EFM OAM 配置与自己的配置是否匹配。

通过以上过程，如果双方的 EFM OAM 配置都匹配，EFM OAM 连接便建立起来。EFM OAM 连接建立后，两端的 OAM 实体会周期性地发送 Information OAMPDU 来检测连接是否正常。如果一端 OAM 实体在连接超时时间内未收到远端发来的 Information OAMPDU，则认为 EFM OAM 连接中断。

2.2.2 链路性能监控

当一端 OAM 实体监控到一般链路事件时，将向其远端 OAM 实体发送 Event Notification OAMPDU 进行通报，同时将监控信息记入日志并上报给网管系统；远端 OAM 实体收到该信息后，也将其记入日志并上报给网管系统。这样，管理员就可以通过观察日志信息动态地掌握网络的状况。

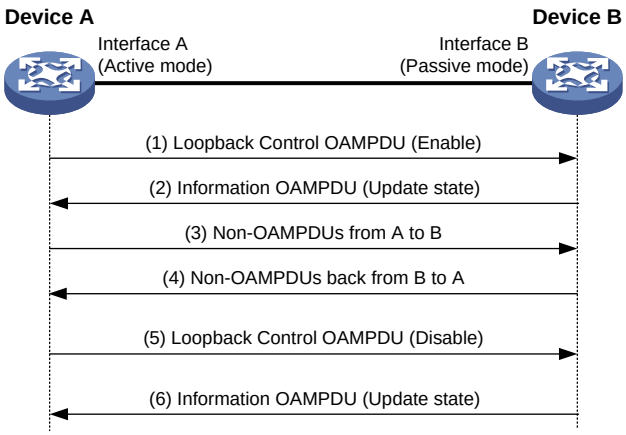
2.2.3 远端故障检测

当设备上发生紧急链路事件而导致流量中断时，故障端 OAM 实体通过 Information OAMPDU 中的 Flag 域将故障信息（即紧急链路事件类型）通知给远端 OAM 实体，同时将故障信息记入日志并上报给网管系统；远端 OAM 实体收到该信息后，也将其记入日志并上报给网管系统。这样，管理员就可以通过观察日志信息动态地了解链路状态，对相应的错误及时进行处理。

2.2.4 远端环回

远端环回功能是指主动模式下的 OAM 实体向远端发送除 OAMPDU 以外的所有其它报文时，远端收到报文后不按其目的地址进行转发，而是将其按原路返回给本端。该功能用于定位链路故障和检测链路质量：网络管理员通过观察非 OAMPDU 报文的返回情况，可以对链路性能（包括丢包率、时延、抖动等）作出评判。

图4 远端环回示意图



如图 4所示，Device A 的接口 Interface A 工作在主动模式下，在 Device A 与 Device B 之间的 EFM OAM 连接建立之后，使能该接口上的远端环回功能：

(1) Device A 向 Device B 发送带有使能信息的 Loopback Control OAMPDU，并等待回复。

(2) Device B 收到该 OAMPDU 后，向 Device A 回复状态改变的 Information OAMPDU，并进入环回状态（在此状态下，设备会把收到的非 OAMPDU 报文都按原路返回）。

(3) Device A 收到回复后，开始向 Device B 发送非 OAMPDU 的测试报文。

(4) Device B 收到测试报文后，将其按原路返回给 Device A。

(5) 当 Device A 需要停止远端环回时，向 Device B 发送带有去使能信息的 Loopback Control OAMPDU。

(6) Device B 收到该 OAMPDU 后便退出环回状态，并向 Device A 回复状态改变的 Information OAMPDU。

2.3 H3C实现的技术特色

EFM OAM 连接建立之后，两端的 OAM 实体会周期性地发送心跳报文（即 Information OAMPDU）来检测连接是否正常。如果一端 OAM 实体在连接超时时间内未收到远端 OAM 实体发来的心跳报文，则认为 OAM 连接中断。IEEE 802.3ah 中定义了心跳报文发送周期为 1 秒，连接超时时间为 5 秒。H3C 在协议规定的基础上，还允许用户对心跳报文的发送周期和连接超时时间进行配置。

对于 OAM 连接中断，用户可以在接口上配置是否关联 EFM OAM 动作：如果配置了关联，接口在收到远端以太网 OAM 事件时除了记录日志外，还会自动断开 OAM 连接，并设置该接口的链路层状态为 down；否则，只记录日志。

3 CFD 技术实现

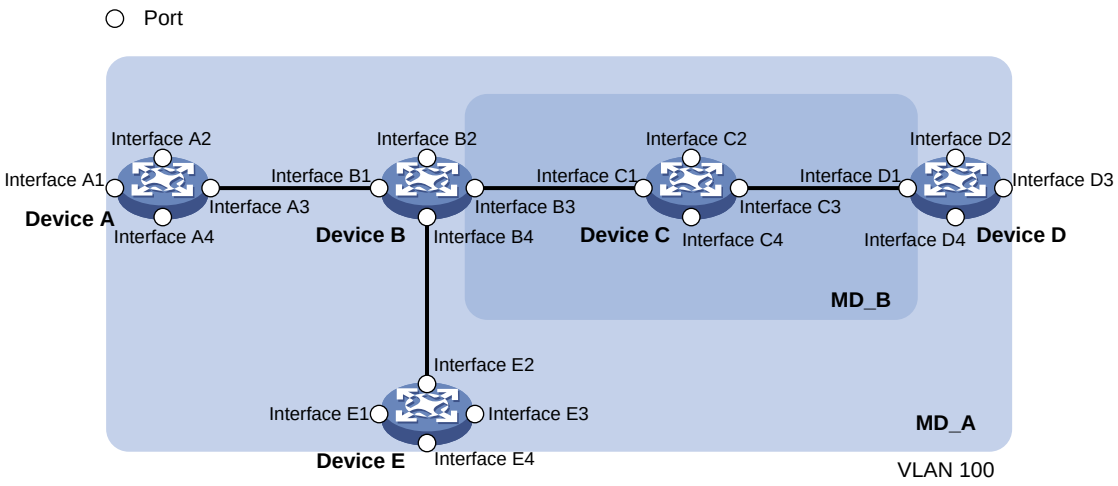
3.1 概念介绍

3.1.1 MD

MD（Maintenance Domain，维护域）是指连通错误检测所覆盖的一个网络或网络的一部分，它以“MD 名称”来标识。

为了准确定位故障点，在 MD 中引入了级别（层次）的概念。MD 共分为八级，用整数 0～7 来表示，数字越大级别越高，MD 的范围也就越大。不同 MD 之间可以相邻或嵌套，但不能交叉，且嵌套时只能由高级别 MD 向低级别 MD 嵌套，即低级别 MD 必须包含在高级别 MD 内部。低级别 MD 的 CFD PDU 进入高级别 MD 后会被丢弃；高级别 MD 的 CFD PDU 则可以穿越低级别 MD；相同级别的 MD 的 CFD PDU 不可以互相穿越。

图5 MD 嵌套示意图



在实际应用中，要对 MD 进行合理规划。如图 5所示，有 MD_A 和 MD_B 两个 MD，MD_B 嵌套在 MD_A 中。要在 MD_A 中进行连通性检测，就要求 MD_A 的 CFD PDU 能够穿越 MD_B。因此需要将 MD_A 的级别配置得比 MD_B 高。这样，MD_A 的 CFD PDU 就可以穿越 MD_B，从而实现了整个 MD_A 的连通性故障管理，而 MD_B 的 CFD PDU 则不会扩散到 MD_A 中。

3.1.2 MA

MA（Maintenance Association，维护集）是 MD 的一部分，一个 MD 可划分为一个或多个 MA。MA 以“MD 名称＋MA 名称”来标识。

一个 MA 服务于一个 VLAN，MA 中的 MP 所发送的报文都带有该 VLAN 的标签，同时 MA 中的 MP 可以接收由本 MA 中其它 MP 发来的报文。

3.1.3 MP

MP（Maintenance Point，维护点）配置在接口上，属于某个 MA，可分为 MEP（Maintenance association End Point，维护端点）和 MIP（Maintenance association Intermediate Point，维护中间点）两种。

1. MEP

MEP 确定了 MA 的边界，它以“MEP ID”来标识。

MEP 所属的 MA 确定了该 MEP 发出的报文所属的 VLAN。MEP 的级别等于其所属 MD 的级别，MEP 发出的报文的级别等于该 MEP 的级别。MEP 的级别决定了其所能处理的报文的级别：当 MEP 收到高于自己级别的报文时不会进行处理，而是将其按原有路径转发；而当 MEP 收到小于等于自己级别的报文时才会进行处理。

MEP 具有方向性，分为内向 MEP 和外向 MEP 两种：

- 内向 MEP 通过除其所在的接口以外的所有接口向外发送 CFD PDU，即在其所属 MA 所服务的 VLAN 中进行广播。
- 外向 MEP 则直接通过其所在的接口向外发送 CFD PDU。

2. MIP

MIP 位于 MA 的内部，不能主动发出 CFD PDU，但可以处理和响应 CFD PDU。MIP 可以配合 MEP 完成类似于 ping 和 tracert 的功能。当 MIP 收到不等于自己级别的报文时不会进行处理，而是将其按原有路径转发；只有当 MIP 收到等于自己级别的报文时才会进行处理。

3.1.4 CFD 分级配置

图6 CFD 的分级配置

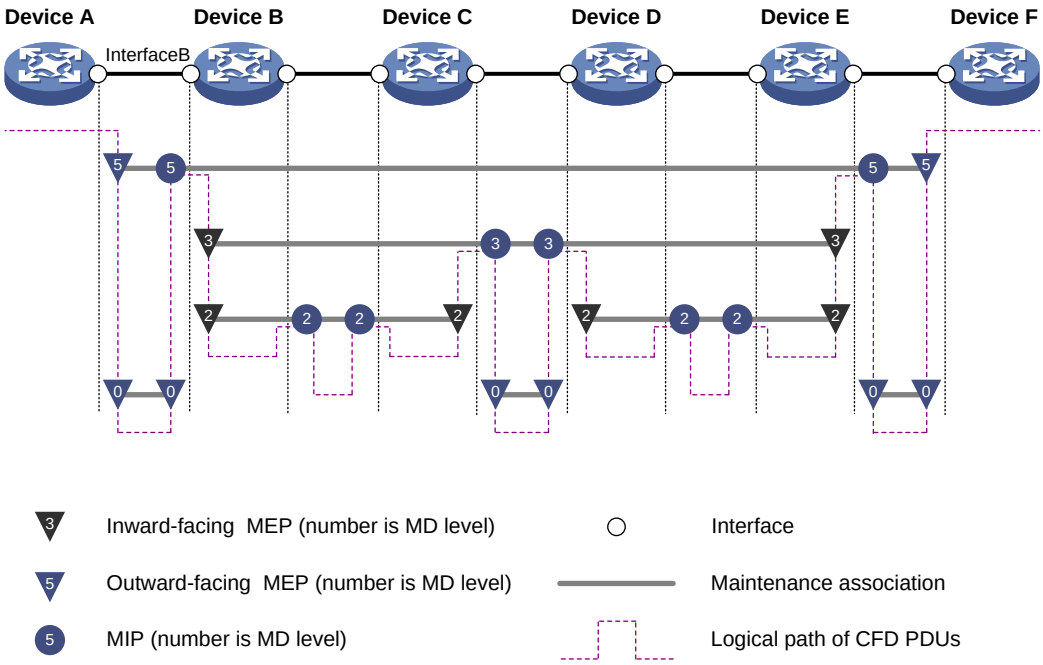


图 6所示为 CFD 的一种分级配置方式，图中共有 0、2、3、5 四个级别的 MD，标识号较大的 MD 的级别高、控制范围广；标识号较小的 MD 的级别低、控制范围小。在 Device A～Device F 的各接口上配置了 MP，譬如 Device B 的接口 Interface B 上配置有：级别为 5 的 MIP、级别为 3 的内向 MEP、级别为 2 的内向 MEP 和级别为 0 的外向 MEP。

3.1.5 协议报文

CFD 的协议报文被称为 CFD PDU。不同的 CFD PDU 具有相同的报文头，通过头部的类型字段来区分报文类型。

图7 CFD PDU 报文格式示意图

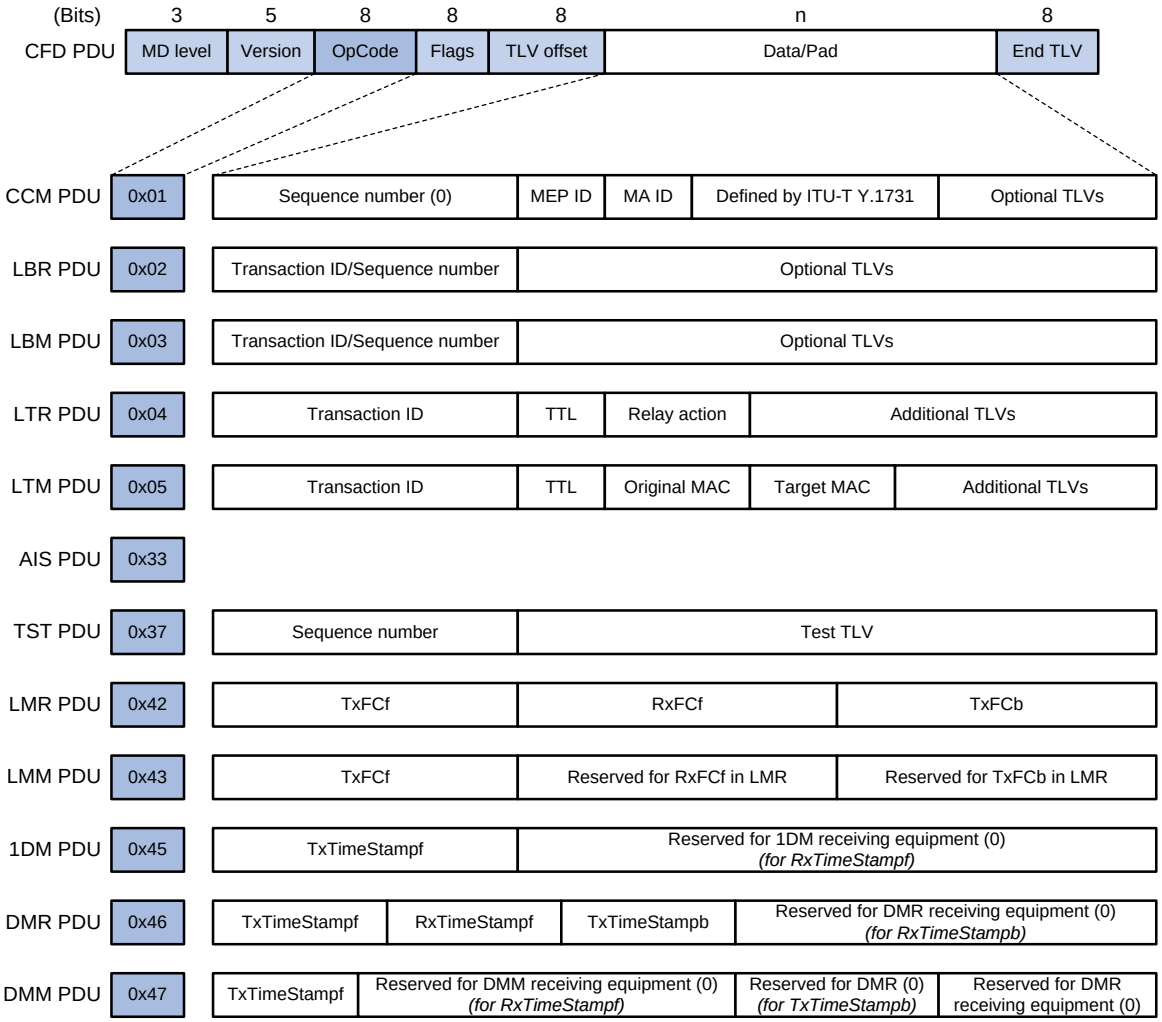


图 7所示为 CFD PDU 的报文格式和常见的 CFD PDU，CFD PDU 中重要字段的含义如表 7所示。

表7 CFD PDU 重要字段含义

字段	含义
MD level	MD的级别，取值范围为0～7，数值越大级别越高
Version	协议版本号，固定为0
OpCode	消息编码，不同取值表示不同类型的CFD PDU，常见的CFD PDU如表8所示
Flags	Flag域，在不同类型CFD PDU中表示不同的含义
TLV offset	TLV偏移量，表示第一个TLV相对于本字段的偏移数值

表8 常见的 CFD PDU

OpCode 值	报文类型	目标 MAC 地址	用途
0x01	CCM PDU	0180-C200-003x（1类组播地址，x取值如表9）	连续性检测
0x02	LBR PDU	发起端单播地址	环回测试
0x03	LBM PDU	目标端单播地址	
0x04	LTR PDU	发起端单播地址	链路跟踪
0x05	LTM PDU	0180-C200-003y（2类组播地址，y取值如表9）	
0x33	AIS PDU	0180-C200-003x（1类组播地址，x取值如表9）	告警抑制
0x37	TST PDU	目标端单播地址	比特错误测试
0x42	LMR PDU	发起端单播地址	单向丢包测试
0x43	LMM PDU	目标端单播地址	
0x45	1DM PDU	目标端单播地址	单向时延测试
0x46	DMR PDU	发起端单播地址	双向时延测试
0x47	DMM PDU	目标端单播地址	

表9 组播地址中 x 和 y 的取值

MD level	x 的取值	y 的取值
7	7	F
6	6	E
5	5	D
4	4	C
3	3	B
2	2	A
1	1	9
0	0	8

3.2 运行机制

说明

CFD 的有效应用建立在合理的网络部署和配置之上。它的功能是在所配置的 MP 之间实现的。

3.2.1 连续性检测功能

MEP 之间的连通失败可能由设备故障或配置错误造成，连续性检测（Continuity Check，CC）功能就是用来检测 MEP 之间的连通状态。该功能的实现方式是：由 MEP 周期性地发送 CCM PDU，相同 MA 的其它 MEP 接收该报文，并由此获知远端状态。若 MEP 在 3.5 个 CCM PDU 发送周期内未收到远端 MEP 发来的 CCM PDU，则认为链路有问题，会输出日志报告。通过 MD 中的多个 MEP 均发送 CCM PDU，就可以实现多点到多点之间的链路检测。

CCM PDU 中时间间隔域（Interval 域）的值、CCM PDU 的发送间隔和远端 MEP 的超时时间这三者之间的关系如表 10所示。

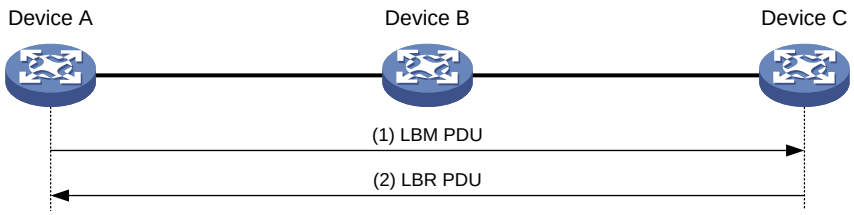
表10 参数关系表

CCM PDU 中时间间隔域的值	CCM PDU 的发送间隔	远端 MEP 的超时时间
1	10/3毫秒	35/3毫秒
2	10毫秒	35毫秒
3	100毫秒	350毫秒
4	1秒	3.5秒
5	10秒	35秒
6	60秒	210秒
7	600秒	2100秒

3.2.2 环回测试功能

环回测试（Loopback，LB）功能类似于 IP 层的 ping 功能，用于验证源 MEP 与目标 MP 之间的连接状态。该功能的实现方式是：由源 MEP 发送 LBM PDU 给目标 MP，并根据能否收到对端反馈的 LBR PDU 来检验链路状态。

图8 环回测试示意图



如图 8所示，在 Device A 与 Device C 之间进行环回测试的过程如下：

(1) Device A 向 Device C 发送 LBM PDU，其中携带该报文的发送时间。

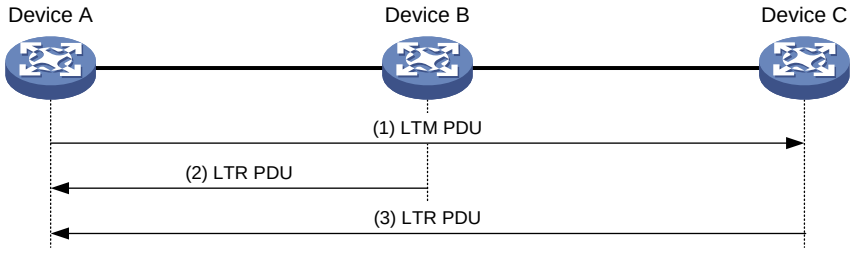
(2) Device C 收到该报文后，回复 LBR PDU 给 Device A，其中携带 LBM PDU 的发送和接收时间，以及 LBR PDU 的发送时间。

在超时时间内，如果 Device A 收到了 Device C 回应的 LBR PDU，则可以根据其中携带的时间信息计算出 Device A 到 Device C 的网络时延；否则，便认为 Device A 到 Device C 不可达。此外，通过连续发送多个 LBM PDU 并观察 LBR PDU 的返回情况，还可以了解网络的丢包情况。

3.2.3 链路跟踪功能

链路跟踪（Linktrace，LT）功能类似于 IP 层的 tracert 功能，用于确定源 MEP 到目标 MEP 的路径。该功能的实现方式是：由源 MEP 发送 LTM PDU 给目标 MEP，目标 MEP 及 LTM PDU 所经过的 MIP 收到该报文后，会发送 LTR PDU 给源 MEP，源 MEP 则根据收到的 LTR PDU 来确定到目标 MEP 的路径。

图9 链路跟踪示意图



如图 9所示，在 Device A 与 Device C 之间进行链路跟踪的过程如下：

(1) Device A 向 Device C 发送 LTM PDU，其中携带 TTL 值和目标 MEP 的 MAC 地址。

(2) Device B 收到该报文后，先将其 TTL 值减 1，再继续转发给 Device C，并回复 LTR PDU 给 Device A，其中也携带 TTL 值（等于 Device A 发送来的 LTM PDU 中的 TTL 值减 1）。

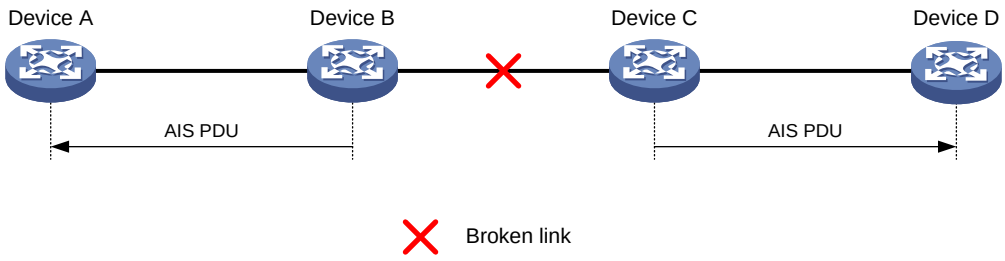
(3) Device C 收到该报文后，回复 LTR PDU 给 Device A，其中也携带 TTL 值（等于 Device B 转发来的 LTM PDU 中的 TTL 值再减 1）。由于根据 LTM PDU 中携带的目标 MEP 的 MAC 地址，Device C 可以判断出自己就是目标 MEP，因此不会再转发该报文。

如果 Device A 到 Device C 之间的路径存在故障，则故障点下游的设备将无法收到 LTM PDU，也不会回复 LTR PDU，据此可判定故障点的位置。例如，若 Device A 能收到 Device B 回复的 LTR PDU，但收不到 Device C 回复的 LTR PDU，就可以判定 Device B 和 Device C 之间的路径存在故障。

3.2.4 告警抑制功能

告警抑制功能是 ITU-T Y.1731 基于 CFD 定义的扩展功能，用来减少 MEP 故障告警的数量。如果 MEP 在 3.5 个 CCM PDU 发送周期内未收到远端 MEP 发来的 CCM PDU，便立刻开始周期性地发送 AIS（Alarm Indication Signal，告警指示信号） PDU，该报文的发送方向与 CCM PDU 相反。其它 MEP 在收到 AIS PDU 后，会抑制本端的故障告警，并转发 AIS PDU 给下游 MEP。此后，如果 MEP 收到了 CCM PDU，便停止发送 AIS PDU 并恢复故障告警。

图10 告警抑制示意图



如图 10所示，告警抑制的触发过程如下：

(1) Device B 和 Device C 之间的链路出现故障，Device B 和 Device C 之间的连续性检测失败，向用户发出故障告警信息。

(2) 检测到故障后，Device B 向 Device A 发送 AIS PDU，而 Device C 则向 Device D 发送 AIS PDU。

(3) Device A 和 Device D 收到 AIS PDU 后，都进入告警抑制状态。此后，如果 Device A 和 Device D 之间的连续性检测失败，将不再向用户发出故障告警信息。

AIS PDU 的缺省发送周期为 1 秒，由于告警抑制功能可以配置很多发送 VLAN，而报文发送数量过多将对设备 CPU 造成负担，因此在这种情况下可将 AIS PDU 的发送周期调整为 1 分钟，从而减少报文发送数量。

3.2.5 丢包测试功能

丢包测试（Loss Measurement，LM）功能用来检测 MEP 之间的丢包情况，分为单向丢包测试和双向丢包测试两种。

1. 单向丢包测试

单向丢包测试功能用来检测 MEP 之间的单向丢包情况。该功能的实现方式是：由源 MEP 发送 LMM（Loss Measurement Message，丢包测试报文） PDU 给目标 MEP，目标 MEP 收到该报文后，会发送 LMR（Loss Measurement Reply，丢包测试应答） PDU 给源 MEP，源 MEP 则根据两个连续的 LMR PDU 来计算源 MEP 和目标 MEP 间的丢包数，即源 MEP 从收到第二个 LMR PDU 开始，根据本 LMR PDU 和前一个 LMR PDU 的统计计数来计算源 MEP 和目标 MEP 间的丢包数。

如图 7所示，源 MEP 发送 LMM PDU 时会填充当前接口的发包统计计数 TxFCf，目标 MEP 收到该报文后，会获取当前接口的收发包统计计数，并向源 MEP 回应。LMR PDU 中携带以下统计值：

- TxFCf：从收到的最后一个 LMM PDU 中的 TxFCf 字段复制而来。
- RxFCf：收到最后一个 LMM PDU 时，当前接口的收包统计计数。
- TxFCb：发送 LMR PDU 时，当前接口的发包统计计数。

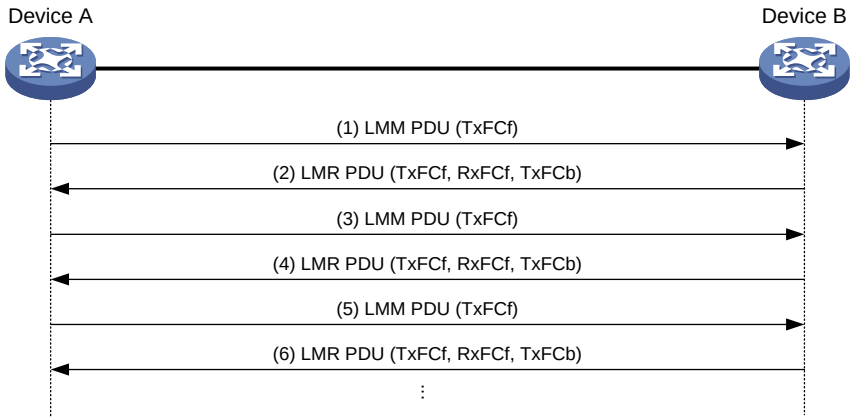
源 MEP 收到 LMR PDU 后（至少需要两次报文交互过程），进行丢包统计的计算公式如下：

- 远端丢包数＝$|TxFCb[t_c] - TxFCb[t_p]| - |RxFCb[t_c] - RxFCb[t_p]|$

- 本端丢包数＝|TxFCf[t_c] – TxFCf[t_p]| – |RxFCl[t_c] – RxFCl[t_p]|

其中，t_c 表示前一次收发的测试报文，t_p 表示当前一次收发的报文。

图11 单向丢包测试示意图



如图 11所示，Device A 与 Device B 之间的单向丢包测试过程如下：

(1) Device A 以 100 毫秒为间隔向 Device B 发送指定数量（缺省为 5 个）的 LMM PDU，报文中填充发送接口的发包统计计数 TxFCf。

(2) Device B 收到 LMM PDU 时，先获取当前接口的收包统计计数 RxFCf，并在向 Device A 回应 LMR PDU 时获取当前接口的发包统计计数 TxFCb。将 TxFCf、RxFCf 和 TxFCb 填充到 LMR PDU 后，将其发送给 Device A。

(3) Device A 收到 LMR PDU 时，先获取当前接口的收包统计计数 RxFCl。从收到第二个 LMR PDU 开始，Device A 就按前面的公式分别计算远端和本端的丢包数，并记录本次交互的计算结果。在本次测试完成后再对测试统计结果进行平均。

单向丢包测试结果的准确度依赖于发送 LMM PDU 和 LMR PDU，以及接收 LMM PDU 时，获取硬件统计计数的及时性和和准确性。一般需要硬件支持直接向报文中填充报文的收发统计计数。

2. 双向丢包测试

双向丢包测试功能的实现方式是：由源 MEP 发送 CCM PDU 给目标 MEP，目标 MEP 收到该 PDU 后，会在发送给源 MEP 的下一个 CCM PDU 中携带端口统计计数，源 MEP 则根据目标 MEP 发送的两个连续 CCM PDU 来计算源 MEP 和目标 MEP 间的丢包数，即源 MEP 从收到第二个 CCM PDU 开始，根据收到的连续两个 CCM PDU 中的统计计数来计算源 MEP 和目标 MEP 间的丢包数。

源 MEP 发送 CCM PDU 时会携带当前接口的发包统计计数 TxFCf。目标 MEP 收到该报文后，获取当前接口携带的收发包统计计数，并向源 MEP 回应，回应的 CCM PDU 中携带以下统计值：

- TxFCf：发送 CCM PDU 时，当前接口的发包统计计数
- RxFCf：收到最后一个 CCM PDU 时，当前接口的收包统计计数（对于第一个 CCM PDU 无意义）。

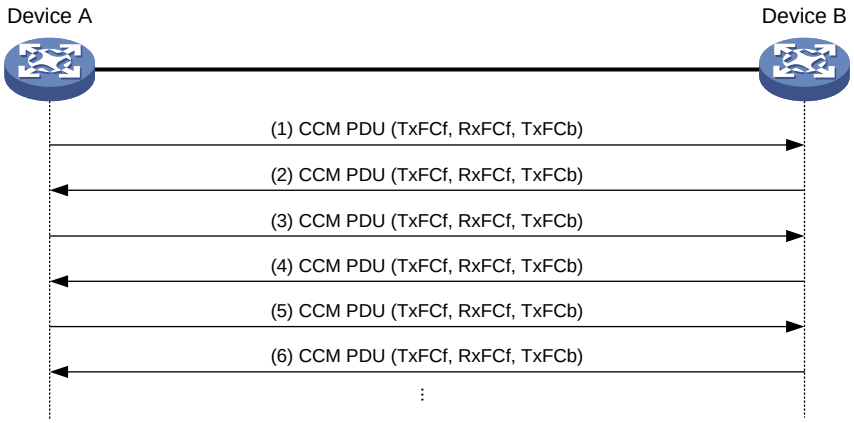
- TxFCb：从收到的最后一个 CCM PDU 中的 TxFCf 字段复制而来（对于第一个 CCM PDU 无意义）。

源 MEP 收到下一个 CCM PDU 时，当前端口统计收包计数为 RxFCb。源 MEP 收到 CCM PDU 后（至少需要收到两次 CCM 报文），进行丢包统计的计算公式如下：

- 远端丢包数＝$|TxFCb[t_c] – TxFCb[t_p]| – |RxFCb[t_c] – RxFCb[t_p]|$
- 本端丢包数＝$|TxFCf[t_c] – TxFCf[t_p]| – |RxFCf[t_c] – RxFCf[t_p]|$

其中，t_c 表示前一次收到的 CCM 报文，t_p 表示当前一次收到的 CCM 报文。

图12 双向丢包测试示意图



如图 12所示，Device A 与 Device B 之间的双向丢包测试过程如下：

(1) Device A 以 100 毫秒为间隔向 Device B 发送指定数量（缺省为 5 个）的 CCM PDU，报文中填充发送接口的发包统计计数 TxFCf。

(2) Device B 收到 CCM PDU 时，先获取当前接口的收包统计计数 RxFCf，并在向 Device A 回应 CCM PDU 时获取当前接口的发包统计计数 TxFCf。将 TxFCf、RxFCf 和 TxFCb 填充到 CCM PDU 后，将其发送给 Device A。

(3) Device A 收到 CCM PDU 时，先获取当前接口的收包统计计数 RxFCb。从收到第二个 CCM PDU 开始，Device A 就按前面的公式分别计算远端和本端的丢包数，并记录本次交互的计算结果。在本次测试完成后再对测试统计结果进行平均。

双向丢包测试结果的准确度依赖于发送和接收 CCM PDU 时，获取硬件统计计数的及时性和和准确性。一般需要硬件支持直接向报文中填充报文的收发统计计数。

3.2.6 帧时延测试功能

帧时延测试（Delay Measurement，DM）功能用来检测 MEP 之间报文传输的时延情况，分为单向时延测试和双向时延测试两种。

1. 单向时延测试

单向时延测试功能的实现方式是：源 MEP 发送 1DM（One-way Delay Measurement，单向时延测试） PDU 给目标 MEP，该报文中携带其发送时间。目标 MEP 收到该报文后记录其接收时间，并结合其发送时间来计算并记录链路传输的时延和抖动（即时延变化值）。

如图 7所示，1DM PDU 中的 TxTimeStampf 字段填充源 MEP 的发送时间 TxTimef，RxTimeStampf 字段填充目标 MEP 的接收时间 RxTimef。源 MEP 会以 100 毫秒为间隔发送指定数量（缺省为 5 个）的 1DM PDU。目标 MEP 收到该报文后，计算 RxTimef 与 TxTimef 的差值即为此次测试的时延。时延抖动则是本次测出的时延与已测出的最小时延的差值。

单向时延测试要求测试设备之间已完成时钟同步，否则测试出的时延有较大误差，此时只能进行时延抖动测试。

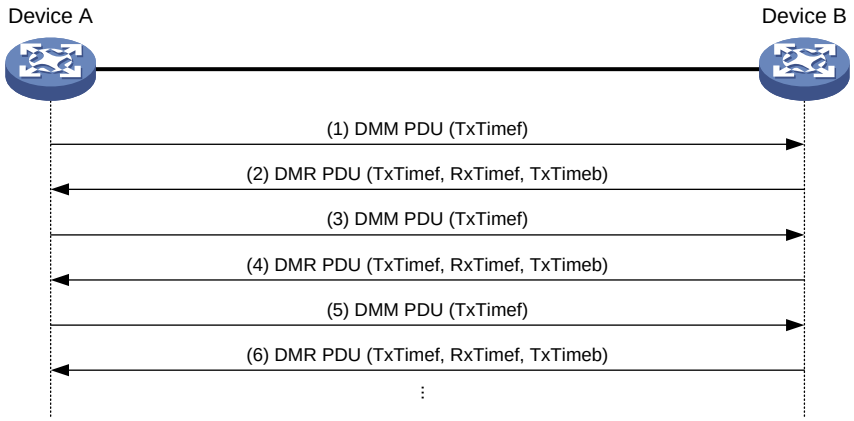
2. 双向时延测试

双向时延分为双向链路时延和双向报文时延两种。双向链路时延表示报文从本端发送到接收的一次往返过程中，报文在链路上消耗的时间；双向报文时延表示报文从本端发送到接收的一次往返过程中消耗的时间。该功能的实现方式为：源 MEP 向目的 MEP 发送测试请求报文，在接收到目的 MEP 的应答报文后，根据应答报文的接收时间和测试请求的发送时间的差值计算双向时延。

如图 7所示，DMM PDU 中的 TxTimeStampf 字段填充源 MEP 的发送时间 TxTimef，目标 MEP 收到此报文后直接将其中的 TxTimef 填充到 DMR PDU 中，并在 DMR PDU 中填充 DMM PDU 的接收时间 RxTimef 和 DMR PDU 的发送时间 TxTimeb。

源 MEP 在收到 DMR PDU 后，需要获取到 DMR PDU 的接收时间 RxTimeb。如果报文中填充了 DMM PDU 的接收时间和 DMR PDU 的发送时间，则可根据以下公式计算出报文往返消耗在链路上的双向链路时延＝ (RxTimeb – TxTimef) – (TxTimeb – RxTimef)；否则，只能计算出双向报文时延＝ RxTimeb – TxTimef。

图13 双向时延测试示意图



如图 13所示，Device A 与 Device B 之间的双向时延测试过程如下：

(1) Device A 以 100 毫秒为间隔向 Device B 发送指定数量（缺省为 5 个）的 DMM PDU，报文中填充发送时间 TxTimef。

(2) Device B 收到 DMM PDU 后向 Device A 回应 DMR PDU，其中填充 TxTimef、RxTimef 和 TxTimeb。

(3) Device A 收到 DMR PDU 后，计算出本此交互的双向链路时延。从收到第二个 DMR PDU 开始，还要计算出双向时延抖动和平均双向时延。

3.2.7 比特错误测试功能

比特错误测试功能用来测试 MEP 之间的报文比特错误。由源 MEP 发送指定数量的 TST（Test，比特错误测试） PDU 给目标 MEP（每个 TST PDU 都携带不同的递增序号），该报文中携带伪随机序列或全 0 值。目标 MEP 收到该报文后，根据其中指定的测试模式和测试比特内容进行计算和比较，从而确定报文是否有比特错误。测试的比特长度为 32 位，有以下几种测试模式：

- 全 0 比特不带校验和
- 全 0 比特带校验和
- 伪随机比特带校验和
- 伪随机比特不带校验和

3.3 H3C实现的技术特色

3.3.1 支持辅助 CPU 快速检测

由于 CCM PDU 的发送周期跨度很大，从 3.3 毫秒到 10 分钟。但是，3.3 毫秒的 CCM PDU 发送周期会对业务板上其它业务的性能产生影响，而其它业务对 CPU 的抢占也会影响 CCM PDU 的发送精度。因此，H3C 可采用单独的辅助 CPU 来处理这种快速报文的发送和接收，检测结果通过主 CPU 之间以及主 CPU 与辅助 CPU 之间的通信来通知 MEP 所在的业务板。

3.3.2 支持与 Smart Link 联动

Smart Link 实现了主备链路的冗余备份和快速迁移。在双上行组网中，当主用链路出现故障时，设备自动将流量切换到备用链路，这样就实现了主备链路的冗余备份。但是，对于传输链路上的设备或链路自身发生的故障（如光纤链路发生单通、错纤、丢包等故障）以及此类故障的恢复，Smart Link 本身是无法感知的。

H3C 通过将 Smart Link、Track 与 CFD 协议的连续性检测功能进行联动，可以对上述故障的发生或恢复进行检测。其原理如下：MEP 周期性地发送 CCM PDU，同一 MA 内的其它 MEP 收到该报文后便能获知远端 MEP 的状态。若 MEP 在 3.5 个发送周期内仍未收到该报文，便认为链路有问题，通知 Track 模块将与 CFD 连续性检测功能关联的 Track 项的状态置为 Negative，然后由 Track 模块通知 Smart Link 重新计算 Smart Link 的链路状态，以便进行链路切换。

3.3.3 支持 LTM PDU 自动发送

H3C 支持 LTM PDU 的自动发送，当本端 MEP 在 3.5 个 CCM PDU 发送周期内未收到远端 MEP 发来的 CCM PDU 时，便判定与远端 MEP 的连接中断。本端 MEP 会自动发送 LTM PDU，并通

过检测回应的 LTR PDU 来定位故障。这个过程会被记录下来，使网络管理员可以在事后查看故障的时间和路径等信息。

3.3.4 支持阈值告警功能

阈值告警功能用来监测链路的传输性能，传输性能包括链路的连通状态、时延、端口丢包率、比特错误率等。当链路的传输性能连续三次达到或超过阈值上限时，则提示用户已超过上限。当链路的传输性能连续三次达到或低于阈值下限时，则提示用户已低于下限。

3.3.5 支持端口联动功能

端口联动功能用来根据外向 MEP 检测到的链路故障结果，关闭或阻塞存在链路故障的端口，确保流量不在该端口丢失。

1. 端口联动功能的触发模式

在不同的端口联动触发模式下，触发条件有所不同：

- 连续性检测超时模式：当 CFD 连续性检测超时时，触发端口联动。
- 系统自动执行双向时延测试模式：当延时达到或超过上限阈值、达到或低于下限阈值时，触发端口联动。
- 远端故障标记模式：当收到携带远端故障标记的 CCM 报文时，触发端口联动。
- 系统自动执行单向丢包测试模式：当丢包率达到或超过上限阈值、达到或低于下限阈值时，触发端口联动。
- 系统自动执行比特错误测试模式：当发生比特错误的报文率达到或超过上限阈值、达到或低于下限阈值时，触发端口联动。

同一个接口上可以配置多种触发模式，满足任何一种情况，都会触发端口联动功能。

2. 端口联动功能的触发动作

如果链路一端配置了端口联动功能，则当端口上的外向 MEP 检测到链路故障后，该端口就会依据配置的联动触发动作来阻塞或关闭端口。

端口联动功能的触发动作分为阻塞端口和关闭端口两种：

- 阻塞端口：将端口的链路层协议状态置为 DOWN(CFD)，且不允许该端口继续收发数据报文。
- 关闭端口：将端口的物理状态置为 CFD DOWN，且不允许该端口继续收发数据报文和协议报文。

3. 故障链路的恢复

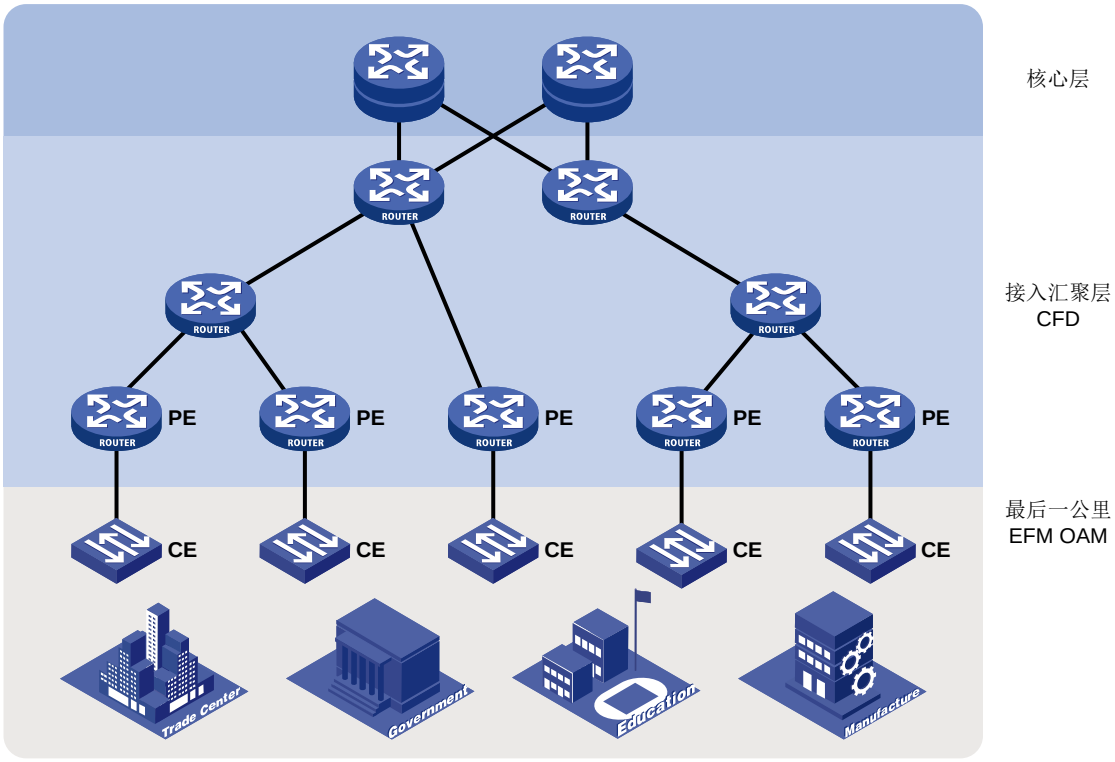
端口被阻塞或关闭后，若远端端口恢复正常：

- 端口联动触发的动作为阻塞端口：
  - 系统自动执行单向丢包测试模式：被阻塞的端口需要执行 **`undo cfd port-trigger slm action`** 命令或 **`cfd slm port-trigger up-delay`** 命令才能被重新开启。
  - 其它模式：被阻塞的端口会自动恢复正常。

- 端口联动触发的动作为关闭端口，需要执行 **undo shutdown** 命令或 **undo cfd port-trigger { cc-expire | dm | rdi | slm | tst } action** 命令才能被重新开启。

4 典型组网应用

图14 以太网 OAM 典型应用组网图



以太网 OAM 在城域网的典型应用如图 14所示，可分为以下两个层次进行部署：

- 在 CE 设备与 PE 设备之间的链路上部署 EFM OAM：通过 CE 设备与 PE 设备之间定时互发 Information OAMPDU 来检测用户业务接入链路的连通性。网络管理员可以通过观察错误帧的情况，来判断 CE 设备与 PE 设备之间链路的性能；通过远端环回功能可以检测链路的质量，或在发生链路故障时进行故障定位。
- 在接入汇聚层的网络中部署 CFD：先根据设备所属的 ISP 来划分 MD，把同一 ISP 管理下的设备划分在同一 MD 中；再根据业务来划分 MA，使每个 MA 对应一个 VLAN。CFD 通过 MA 内的各 MEP 定时互发 CCM PDU 来检测 MA 内网络的连通性。当检测到连通性故障后进行报警，网络管理员可以通过环回测试、链路跟踪、单向丢包测试等功能进行故障定位、路径查找或链路测试。

5 参考文献

- IEEE 802.3ah：Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications

- IEEE 802.1ag：Virtual Bridged Local Area Networks Amendment 5: Connectivity Fault Management
- ITU-T Y.1731：OAM functions and mechanisms for Ethernet based networks