基于 MPLS L2VPN 技术的 VPWS 冗余备份的设计与实现①

孙文胜 肖 磊 (杭州电子科技大学 通信工程学院 浙江 杭州 310018)

摘 要: 目前在 VPN 应用领域, MPLS L2VPN 技术已成为最热门的 VPN 解决方案。为了增强基于 VPWS(Virtual Private Wire Service，虚拟专线业务，又称为 VLL)技术的 L2VPN 数据传输的可靠性，本文提出了 VPWS 冗余备份的设计方案。该方案采用状态机的设计方法，通过 C 语言中的函数指针和二维数组将状态和事件映射到各个处理函数上来实现。当网络链路发生故障时，该方案可以快速的切换 VPWS 冗余备份中主备 PW(pesiod –wire，虚链路)的状态以加快网络收敛，从而保证 L2VPN 业务的稳定性。

关键词: 多协议标签交换; 虚拟专线业务; 冗余备份; 状态机; 链路

Design and Implementation of VPWS Redundant Backup Based on MPLS L2VPN Technology

SUN Wen-Sheng, XIAO Lei

(School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: At present, MPLS L2VPN technology has been the most popular VPN solution in the field of VPN application. For enhancing the reliability of data transmission of L2VPN based on VPWS technology, this paper proposes a scheme of VPWS redundant backup. This solution adopts the state machine design method, which states and events are mapped to varieties of processing functions through function pointers and two-dimensional arrays realizes. When the network link goes wrong, the solution can switch the state of standby PW of VPWS redundant backup quickly to accelerate the network convergence, which ensures the stability of L2VPN service.

Keywords: multiprotocol label switching; virtual private wire service; redundant backup; state machine; pesiod – wire

传统的叠加式二层 VPN(Virtual Private Network，虚拟专网)技术依赖于专用的媒介(如 ATM 或 FR)，给网络维护和管理带来了沉重的负担。目前，采用 MPLS 技术的网络已经被普遍认为是下一代核心网络的发展方向，而 MPLS 技术最主要的优势之一就是能够提供基于 MPLS(Multiprotocol Label Switching，多协议标签交换)网络的二层 VPN 服务，使运营商可以在统一的 MPLS 网络上实现基于不同数据链路(包括 ATM、FR、VLAN、Ethernet、PPP 等)的二层 VPN 业务[1]。

VPWS 是以太网上点对点的二层 VPN 业务，它是以 MPLS 技术为支撑的[2]。本文在基于 Martini 方式的 VPWS 技术的基础上，提出了利用状态机来设计 VPWS 冗余备份的方案，并用 C 语言实现了 VPWS 冗余备份的功能。本文侧重介绍了如何利用状态机来维护 VPWS 冗余备份中主备 PW 的工作状态和处理各种相关的会话消息，MPLS 技术和扩展的 LDP 协议不作为本文讨论的重点。

① 收稿时间:2010-03-16;收到修改稿时间:2010-04-26

1　VPWS 技术与BFD技术原理

1.1 VPWS 技术原理简介

MPLS L2VPN 技术是在 MPLS 网络上透明传输用户二层数据。从用户的角度来看，MPLS 网络是一个二层交换网络，可以在不同节点间建立二层连接。VPWS 方式的 MPLS L2VPN 可以在公网中提供一种点到点的 L2VPN 业务，让两个站点之间的连接效果像直接用链路连接一样[3]。VPWS 标准存在两种解决方案，即 Martini 方式和 Kompella 方式，本文只讨论前者。

Martini 方式采用扩展的 LDP(Label Distribution Protocol，标签分发协议)信令方式来建立伪线路，该方式的 MPLS L2VPN 着重于在两个 CE(直接与服务提供商相连的用户边缘设备)之间建立 VC(Virtual Circuit，虚电路)。 Martini 方式采用 VC-TYPE 加上 VC ID 来标识一个 VC。VC-TYPE 表明 VC 的封装类型：ATM、VLAN 或 PPP；VC ID 则用于唯一标识一个 VC。同一个 VC-TYPE 的所有 VC 中，其 VC ID 必须在整个 PE(服务提供商网络上的边缘设备，与 CE 相连，主要负责 VPN 业务的接入。它完成报文从私网到公网隧道和报文从公网隧道到私网的映射与转发)中唯一。 连接两个 CE 的 PE 通过 LDP 交换 VC 标签，并通过 VC ID 绑定对应的 CE。当连接两个 PE 的 LSP(Lable Swtich Path，标签转发路径)建立成功，双方的标签交换和绑定完成后，一个 VC 就建立起来了，CE 之间可以通过此 VC 传递二层数据[4]。

PW 是两个 PE 设备之间的一条双向虚拟连接。它由一对方向相反的单向的 MPLS VC 组成，也称为仿真电路。PW 是 VPWS 在公网上的通信隧道，它建立在 MPLS(包括普通 LSP 和 CR-LSP)或 GRE 等隧道之上。创建 PW 需要：

(1) 首先在本端和对端 PE 之间建立 MPLS 或 GRE 等隧道。

(2) 确定对端 PE 的地址，可通过手工配置来指定对端 PE 地址。

(3) 利用 LDP 信令协议为 PW 分配 VC 标签，并将分配的 VC 标签通告给对端 PE，建立单向的 VC，从而创建 PW。如果 PW 建立在 MPLS 隧道之上，则 PW 上传输的报文将包括两层标签：内层标签为 VC 标签，用来判断报文属于的 VC，从而将报文转发给正确的 CE；外层标签为公网 MPLS 隧道标签，用来保证报文在 MPLS 隧道上的正确传输[5]。

1.2 BFD 原理简介

为了保护关键应用，网络中会设计有一定的冗余备份链路。网络发生故障时，要求网络设备能够快速检测出故障并将流量切换至备链路以加快网络的收敛速度。BFD 协议在上要求下产生，它提供了一个通用的标准化的，介质无关和协议无关的快速检测机制。BFD 的原理是通过在两台设备上建立会话，用来检测网络设备间的双向转发路径，为上层应用服务。会话建立后，会周期性的快速发送 BFD 报文，如果在检测时间内没收到 BFD 回应报文，则认为路径发生了故障通知被服务的上层应用进行相应的处理[6]。

BFD 技术具有如下优点：

(1) 能够对网络设备间任意类型的双向路径进行故障检测；

(2) 可以为不同的应用服务，提供一致的快速的故障检测时间；

(3) 提供小于 1 秒的检测时间，加快网络收敛，提高可靠性。

2　基于状态机的VPWS冗余备份设计方案

2.1 基于 Martini 方式的 VPWS 报文交互过程

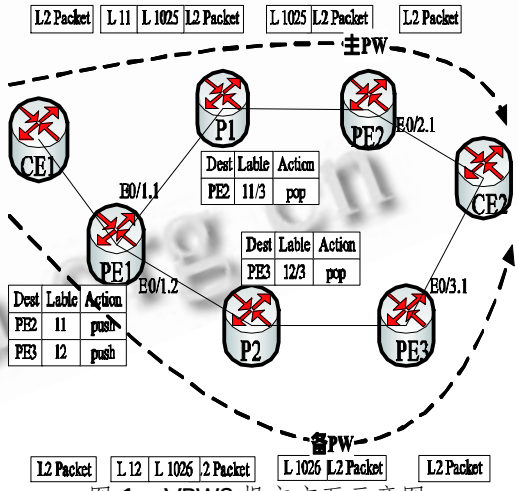

图 1　VPWS 报文交互示意图

如图 1(注：图 1 中虚线表示主备 PW 上报文的转发路径，图示中上方和下方的报文字段分别表示主备 PW 上报文的封装内容)所示，CE1 要发送报文到 CE3，当 PE1 收到 CE1 相关接口送到的二层报文后，根据 VC 的关联配置查找 VC-ID 对应的表项，如表 1 所示，得到下一跳为 PE2，VC 标签为 1025(作为 MPLS 报文字段的内层标签)。于是 PE1 在二层报文(L2 Packet)外封装 MPLS 头，Lable=1025。根据下一跳为 PE2 查找 LSP 隧道，得到外层标签为 11，增加第二个 MPLS

封装，Lable=11，并由出接口 E0/0.1 发送到 P 设备。

表 1 PE 设备上的 VPWS 报文表

Dev	VCID	Dest	State	Loc/Rmt-Lable	Interface
PE1	101	PE2	up	1024/1025	E0/1.1
PE1	102	PE3	block	1026/1027	E0/1.2
PE2	101	PE1	up	1025/1024	E0/2.1
PE3	102	PE1	up	1027/1026	E0/3.1

P 设备进行 MPLS 转发，查找 LSP 表后知道自己是次末中继，弹出外层标签，并发送到 PE2。 PE2 收到报文后，根据内层标签 1025 查找 VC-ID 对就的表项，如表 1 所示，得到出接口 E0/2.1，将内层标签进行弹出操作，然后将二层报文直接送达目的地 CE3[7]。

从上面的交互过程中可以看到外层的 LSP 隧道是被共享的，PE2 收到报文后根据内层标签的不同而映射到不同的 VC 上。

2.2 VPWS 冗余备份的设计要求

上节描述的 VPWS 报文交互采用的是单 PW 链路，为了增强基于 VPWS 技术的 L2VPN 数据传输的可靠性，需要在 PE 设备上将 VPWS 模块设计成具有冗余备份的功能，即具有双 PW 链路。 当网络发生故障时，PE 设备能够通过 BFD 协议快速检测出链路故障并通告给 VPWS 模块，使其将流量切换至备链路以加快网络的收敛速度，保证 L2VPN 数据传输不会被中断。

在具体的设计之前，先将 PW 的工作状态定义为三种类型：

(1) UP 状态：当一对方向相反的单向的 MPLS VC(Virtual Circuit，虚电路)在两台 PE 设备间建立起来后， 则称该 PW 处于 UP 状态，即该虚链路能够正常转发 L2VPN 报文；

(2) DOWN 状态：只要有一条单向的 MPLS VC 链路无法建立，则称该 PW 处于 DOWN 状态，无法转发 L2VPN 报文；

(3) BLOCK 状态：当 VPWS 具有备份 PW 时，后建立起来的 PW 链路将处于 BLOCK 状态，即备用状态。

假定 PE1-P1-PE2 的链路(如图 1 中的虚线所示) 为主 PW(pesiod －wire)，PE1-P2-PE3 的链路为备 PW。设此时主 PW 的工作状态为 UP，则备 PW 的工作状态为 BLOCK。 当 PE1 得知主 PW 不能工作时，即主 PW 的工作状态变为 DOWN 时。此时 PE1 会自动切换到备 PW，即备 PW 变为 UP 态来进行报文转发。其报文交互过程与上节描述的主 PW 类似。

关于 LSP 隧道如何建立，改进型的 LDP 如何通过交换 VC 标签来绑定相应 CE 的技术都已趋于成熟并广泛应用于各种大型网络设备中，因此本文重点讨论的是在 PE 设备上如何维护具有 VPWS 冗余备份功能的主备 PW 状态(分别为上文提到的 UP BLOCK DOWN 状态)，使其能够正确的响应 LDP 会话消息，BFD 会话消息以及用户手工倒换 PW 状态的命令[8]。对 VPWS 冗余备份模块的详细设计要求如下：

(1) 要在两台 PE 设备上建立 PW，就必须在双方之间交换信息。Martini 方式 MPLS L2VPN 着重于在两个 PE 之间建立 VC。Martini 方式对传统的 LDP 做了扩展(DP 标准为 RFC3036)用于交互 VC 信息，所以要能创建 PW 并正确的维护主备 PW 状态，就必须要求 VPWS 冗余备份模块能够处理扩展的 LDP 会话消息。

(2) 对于具有冗余备份的 PW 链路，当使用 BFD 来检测网络链路的故障时，要求 VPWS 冗余备份模块能正确处理 BFD 会话消息，从而维护主备 PW 的工作状态。

(3) 此外，还要求 VPWS 冗余备份模块能响应用户下达的手工切换命令来手动切换 PW 的工作状态。

2.3 状态机对 VPWS 冗余备份的描述

根据上述设计要求，本文采用状态机来设计。典型的状态机实现中需要考虑几个要素：状态、事件(也可称作消息)、事件处理函数以及系统上下文等。系统处于某个状态，收到某个事件后，解析出事件内容，然后调用相应的事件处理函数进行处理，这样系统就可能转变到另一新的状态，而且一个事件对不同的状态将产生不同的影响，有些状态接收某些事件可以不受任何影响，同一个状态转换到另一个状态也可以通过不同的事件[9]。

依据主备 PW 的工作状态和 LDP 以及 BFD 会话的消息类型，我们将上述情况抽象成 4 种状态和 9 种事件的状态机模型。4 种状态分别是：

(1) IDLE 状态，该状态下用户在 PE 设备上可能只创建了主 PW 或主备 PW 都被创建，但主备 PW 的工作状态都处于 DOWN，即不能转发任何 L2VPN 报文；

(2) NOBACKUP 状态，该状态下用户在 PE 设备上也可能只创建主 PW 或主备 PW 都被创建，但其中只有主 PW 处于 UP 态并用来转发 L2VPN 报文；

(3) NOSWITCH 状态，该状态下主备 PW 都被创

建，其中主 PW 的工作状态为 UP，进行 L2VPN 报文的转发，备 PW 为 BLOCK 态，处于备用状态；

(4) SWITCHOVER 状态，该状态下主备 PW 都被创建，其中备 PW 的工作状态为 UP，进行 L2VPN 报文的转发，主 PW 为 BLOCK 态，处于备用状态，该状态可以和 NOSWITCH 状态进行相互切换。

9 种事件分别是:

LDP 会话事件：LDP_MAIN_PW_UP, LDP_MAIN_PW_DOWN, LDP_BACKUP_PW_UP 和 LDP_BACKUP_PW_DOWN;

BFD 会话事件：BFD_MAIN_PW_UP, BFD_MAIN_PW_DOWN, BFD_BACKUP_PW_UP 和 BFD_BACKUP_PW_DOWN;

用户手工切换事件： USER_PW_SWITCH。

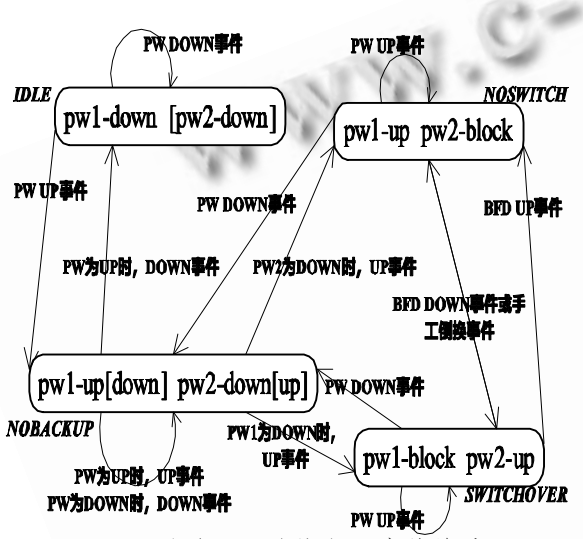

图 2　主备 PW 的状态--事件关系图

其中 LDP 会话和 BFD 会话事件中的 MAIN，BACKUP，UP，DOWN 等关键字分别表示主或备 PW 将要 UP 或 DOWN 的会话消息(事件)。本文依据上述 4 种状态和 9 类事件，结合 VPWS 的技术原理描绘出主备 PW 的状态--事件关系图，如图 2(注：图示中[]表示可选项)所示。

3　VPWS冗余备份的C语言实现

本文在 C 语言中采用二维数组结合函数指针的方法来实现状态机。根据上述 3.3 节中描述的 4 种状态和 9 类事件，定义如下的两类枚举：Enum PW_STATE 和 Enum PW_EVENT；再定义一个二维数组来存放处理函数的指针。利用二维数组的对应关系，将数组的两下标分别对应为枚举 PW_STATE 和 PW_EVENT；第一维表示四种状态，第二维表示每种状态下的 9 种不同的事件，定义如下[10]：

typedef unsigned long (*State_Maintain_Fun)(PW_INNODE pw_innode);

State_Maintain_Fung_pfPW_State_Maintain[4][9];

数组元素存放着不同处理函数的指针，每个处理函数对应一个事件和状态，用于处理当前状态下的事件。这样通过调用该二维数组的函数指针就可以将状态和事件映射到具体的处理函数上，并在相应的处理函数中将当前状态迁移到下一状态，从而实现状态机的功能。

对存储 36 个处理函数指针的数组，本文作进一步的优化处理：一方面，把与事件无关或不需要处理的状态的数组元素设为 NULL 指针，这样减少了差不多一半的处理函数；另一方面，对有着相似处理要求的状态和事件，只用一个函数来处理。这样在最后可以只留下十三个处理函数来组成状态机的实现机制[11]。

处理函数的实现各不相同，但一般都会将当前状态迁移到一个新的状态。下面们列出了状态处理函数(注：对每个处理函数的命名表示了这个函数所对应的状态和事件，如：IDLEState_Pro_LDP_MainUP 表示是在 IDLE 状态下响应 LDP_MAIN_PW_UP 事件，后续不再作重复说明。)的具体实现：

(1) IDLEState_Pro_LDP_MainUP：将当前主 PW 的状态置为 UP，并将状态更新为 NOBACKUP；

(2) IDLEState_Pro_LDP_BakUP 将当前备 PW 的状态置为 UP，并将状态更新为 NOBACKUP；

(3) NOBACKUPState_Pro_LDP_MainUP：先判定若当前主 PW 的状态为 UP，则不作任何处理，否则将主 PW 的状态置为 BLOCK，并将状态更新为 SWITCHOVER；

(4) NOBACKUPState_Pro_LDP_MainDOWN：先判定若当前主 PW 状态为 DOWN，则不作处理，否则将主 PW 状态置为 DOWN，并将状态更新为 IDLE；

(5) NOBACKUPState_Pro_LDP_BakUP：先判定若当前备 PW 状态为 UP，则不作处理，否则将主 PW 状态置为 BLOCK，并将状态更新为 NOSWITCH；

(6) NOBACKUPState_Pro_LDP_BakDOWN：先判定若当前备 PW 状态为 DOWN，则不作处理，否则将主 PW 状态置为 DOWN，并将状态更新为 IDLE；

(7) NOSWITCHState_Pro_LDP_MainDOWN：将主 PW 状态置为 DOWN，备 PW 状态置为 UP，并将状态更新为 NOBACKUP；

(8) NOSWITCHState_Pro_LDP_BakDOWN：仅将 PW 状态置为 DOWN，并将状态机的状态更新为 NOBACKUP；

(9) NOSWITCHState_Pro_BFD_MainDOWN：交换主备 PW 的状态，并将状态机的状态更新为 SWITCHOVER， 注：对于 NOSWITCH 状态下的 USER_PW_SWITCH 事件也使用该函数处理。

(10)SWITCHOVERState_Pro_LDP_MainDOWN：仅将主 PW 状态置为 DOWN，并将状态更新为 NOBACKUP；

(11) SWITCHOVERState_Pro_LDP_BakDOWN：将主 PW 状态置为 UP，备 PW 状态置为 DOWN，并将状态更新为 NOBACKUP；

(12) SWITCHOVERState_Pro_BFD_MainUP：将当前主 PW 状态置为 UP，备 PW 状态置为 BLOCK，并将状态更新为 NOSWITCH；

(13)SWITCHOVERState_Pro_BFD_BakDOWN：交换主备 PW 的状态，并将状态更新为 NOSWITCH，注：对于 SWITCHOVER 状态下的 USER_PW_SWITCH 事件也使用该函数处理。

本文设计的 PW 状态维护函数的伪代码如下：

```
int Maintain_PWState(enum PW_STATE curr
—ent_state, enum PW_EVENT event, PW_INNODE
pw_innode)
{
    int Ret;      /*定义变量并检查参数的合法性*/
    /*判定如果二维数组元素映射的处理函数名为
NULL，则不作任何处理，返回不处理信息*/
    if( g_pfPw PW_State_Maintain[current_
state][event] == NULL){  return UNPROESS; }
    /*调用二维数组元素映射的处理函数*/
    Ret=g_pfPwPW_State_Maintain[current_stat
e][event](pwinnode);
    return Ret;  /*向调用者返回处理结果*/
    }
```

在该函数中依据传入的 current_state 和 event 参数查找二维数组元素表，从而通过引用函数指针来达到调用相应处理函数的目的，具有映射快速和结构清晰等优点。 应用证明该状态机模块能正确的处理每个状态下的事件，从而正确的维护主备 PW 的工作状态，让 L2VPN 报文得到可靠的转发。通过在 PW 状态维护函数中添加断言以及状态和事件的调试信息，还可以有效的定位代码问题和跟踪主备 PW 的工作状态。

4 结语

本文采用状态机来设计具有 VPWS 冗余备份的 PW 链路，使得对复杂的 PW 状态迁移和类型繁多的会话消息的处理变得简化和高效。在充分考虑到状态机特点的基础上，本文在 C 语言中采用二维数组结合函数指针的方法来实现状态机，通过将状态和事件直接映射到相应的处理函数上，从而避免了在 C 语言中使用逻辑复杂且低效的条件判断语句。调试和检测表明，VPWS 冗余备份模块能够在检测到网络故障时，快速的切换到备 PW 链路以保证 L2VPN 数据的可靠传输。该设计方案已在华三通信公司的网络设备中得到应用，具有很强的使用价值和推广价值。这种状态机的设计思想还可以很好的推广到一些有着复杂状态和多种事件响应的环境中。

参考文献

1 甘朝钦.VPN 向下一代网络演进最现实合理的选择——MPLS L2VPN.电信技术, 2004,13(3):57—59.

2 宋庆.VPWS 系统的设计与实现[硕士学位论文].西安:西安电子科技大学, 2007.

3 金利忠,樊祥宁,冯军.二层 VPN 互联的研究.中兴通讯技术, 2005,19(6):30—33.

4 李丹荔. L2VPN-VPWS 的研究与实现[硕士学位论文].成都:电子科技大学, 2008.

5 樊自甫,万晓榆.T-MPLS 层网络和以太网间基于 PW 的互通模型.光通信研究, 2007,12(3):7—9.

6 周榜兰.基于 BFD 的 MPLS 网络自愈恢复技术的研究与实现[硕士学位论文].成都:西南交通大学, 2008.

7 华为技术有限公司.MPLS L2VPN 技术白皮书.[2008-07].http://www.huawei.com/cn/products/datacomm

8 陈麒帆, LUC DE GHEIN. MPLS 技术构架.北京:人民邮电出版社, 2008.387—405.

9 聂旭中.状态机设计研究[硕士学位论文].洛阳:洛阳师范学院学报, 2009.

10 NNETH REEK,徐波. C 和指针.北京:人民邮电出版社, 2008.91—115.

11 肖勇军,施荣华.基于 TMS 结构的 OSPF 邻居状态机设计与实现.长沙通信职业技术学院学报, 2005,12(3):49—52.