

BGP MPLS VPN 模块设计与实现

李海华^{1,2}, 冯卫华³

(1 华北水利水电学院, 河南 郑州 450011; 2 河南工业贸易职业学院, 河南 郑州 450012;

3 河南司法警官职业学院, 河南 郑州 450002)

摘 要: 分析了 BGP MPLS VPN 模块的需求, 该模型包括三个主要子模块: LDP 模块、转发模块、转发表组装模块, 分析它们的工作原理, 提出了模块设计与实现思路, 将 VRF 表、FNT 表、IML 表三表合并为一个转发表, 由操作码决定查询的转发条目, 简化了对路由信息的管理, 提高了查表效率。

关键词: 多协议标签交换虚拟专用网; 需求分析; 模型; 子模块; 转发表

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2012)01-0138-05

Design and Implementation of BGP/MPLS VPN Modul

LI Hai-hua^{1,2}, FENG Wei-hua³

(1 North China University of Water Resources and Electric Power, Zhengzhou 450011, China;

2 Henan Industry And Trade Vocational College, Zhengzhou 450012, China;

3 Henan Judicial Police Vocational College, Zhengzhou 450002, China)

Abstract: Analyzes the needs of BGP MPLS VPN modul. The model include three main Sub-units, that are the unit of LDP, Forwarding and Forwarding Table unit. Meanwhile, analysis forward operation principle and puts thoughts of design and impletement. Besides, puts VRF-table, FTN-table and IML-table into a Fowarding-table of which the Forwarding-ietm is decided by Operation-cord, so that it simplifies the management of Route-information, improve the look-up table efficiency.

Key words: BGP MPLS VPN; requirement analysis; model; sub-unit; fowarding-table

1 引言

BGP MPLS VPN 采用对等网络模型, 构成网络的路由设备可以分为三类: P (Provider router, 骨干路由器)、PE (Provider edge routers, 服务提供商边缘路由器) 及 CE (Customer edge routers, 用户边缘路由器)。

CE 路由器经过一个或多个 PE 路由器的数据链路为用户提供访问服务, 它同直接相连的 PE 路由器建立邻居关系, 邻居建立以后, CE 路由器向 PE 路由器通告本地 VPN 路由, 并从 PE 路由器学习远端的路由。

PE 路由器用 RIPv2、OSPF 或 EBGp 协议形成

的静态路由跟与它相连的 CE 路由器交换信息, 一个 VPN 一个转发条目, 隔离了路由信息, 保证了数据转发的安全。

P 路由器是与 CE 路由器不相关联的提供商网络中的任一路由器。通过 MPLS 转发二层标签栈的业务分组, P 路由器仅要求维护提供商的 PE 路由器的路由, 不要求维护每一个用户站点的特定的 VPN 的路由信息^[1]。

2 BGP MPLS VPN 需求分析

BGP MPLS VPN 需求如下:

(1) 通过路由协议 (BGP、OSPF、RIP-2) 发布路由信息、更新路由表。

(2) 将用户路由重新分发到 MP-iBGP 协议中,用 MP-iBGP 协议利用导入导出规则更新对端 PE 路由器的 VPN 路由转发条目(VPN Routing Forwarding, VRF).

(3) 用标签分发协议(Label Distribution Protocol, LDP)协议分发标签,更新和管理标签信息库(Lable information Library, Lib).

(4) 创建标签转发表,该表包含 VRF 条目、转发等价类到下一跳标签转发条目映射(FEE To NHLFE, FTN)条目、入标签映射(INPUT LABEL Map, ILM)条目.

(5) 入口 PE, 据 VRF 转发条目为分组打上内层标签, 据 FTN 条目对分组打上外层标签; 在核心 P 中, 根据 ILM 条目进行交换标签操作; 在出口 P 中, 采用倒数第二跳机制, 根据 ILM 条目进行弹出外层标签操作, 在出口 PE 中, 弹出内层标签, 将分组恢复成 IP 数据报.

3 BGP/MPLS VPN 子模块划分

BGP/MPLS VPN 分为控制平面和数据平面. 控制平面生成标记转发信息库, 数据平面完成数据分组的转发处理, 如图 1 所示^[2].

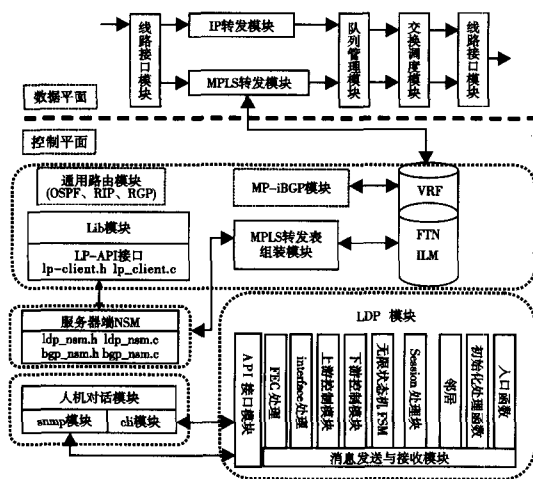


图 1 MPLS VPN 子模块划分

控制平面包含: LDP 模块、服务器端 NSM 模块、Lib 模块、通用路由模块、MPLS 转发表组装模块、MP-iBGP 模块、人机对话模块、转发表模块(包含 VRF 条目、FTN 条目、ILM 条目).

LDP 模块通过 UDP 协议发送 Hello 消息发现邻居, 建立对等实体, 双方建立会话实体, 会话实体通过 TCP 链接建立连接以发送/接收消息. 上游控制块通过调用会话, 将入标签跟一个接口、转发等价

类(Forwarding Equivalence Class, FEC)建立绑定关系. 下游控制块通过调用会话, 将出标签跟一个接口、FEC 建立绑定关系^[3].

LDP 模块它包含会话处理模块、邻居、有限状态机、上游控制模块、下游控制模块、接口处理模块、转发等价类处理模块、应用程序接口 API(Application Program Interface)模块, 涉及的主要程序有: ldp_session. h, ldp_session. c, ldp_fsm. h, Ldp_fsm. c, ldp_adjacency. h, ldp_adjacency. c, ldp_usm. h, ldp_usmk. c, ldp_dsm. h, ldp_dsm. c, ldp_interface. h, Ldp_interface. c, ldp_fec. h, ldp_fec. c.

LDP 模块中的 API 模块将绑定的标签信息通过服务器 NSM 模块存放到标签信息模块 Lib 中. Lib 模块涉及的主要程序有: lp_client. h lp_client. c, lib. h, lib. c.

服务器终端 NSM 模块通过服务器/终端方式提供 LDP 模块、MPLS 转发表组装模块、Lib 模块、通用路由模块之间的通信, NSM 的工作是以线程形式来完成的, 也就是通过调用线程机制, 需要做什么处理, 将有关参数提供给线程机制, 由线程机制产生线程进入线程队列, 根据优先级依次被调度执行. 涉及的主要程序有: nsm_api. h, nsm_api. c, nsm_server. h, nsm_server. c, ldp_nsm. h, Ldp_nsm. c.

MPLS 转发表组装模块根据通用路由模块提供的路由信息 Rib 和标签信息模块提供的 Lib 信息在转发表模块中创建 FEC 到 NHLFE 的映射 FTN 条目或输入标记映射 ILM 条目. 涉及的主要程序有: mpls_table. h, mpls_table. h.

人机对话模块通过命令行界面(command-line interface, CLI)或简单网络管理协议(Simple Network Management Protocol, SNMP)方式向路由器发出指令, 对 MPLS VPN 模块中的参数进行配置. 涉及的主要程序有: ldp_snmp. h, ldp_snmp. c, ldp_cli. h, ldp_cli. c show. c.

消息发送与接收模块包含: 数据的发送与接收处理、数据编码与解码, 数据的发送与接收处理模块的作用是接收和发送各种消息, 涉及的主要程序有: Ldp_network. h Ldp_network. c, 数据编码与解码处理消息的编码解码, 涉及的主要程序有: Ldp_packet. h Ldp_packet. c.

MP-iBGP 模块生成 VPN 标签, 创建转发表模块中的 VRF 条目.

数据平面有线路接口模块、IP 转发模块、MPLS 转发模块、队列管理模块、交换调度模块构成.

数据分组通过线路接口模块时,根据协议字段的值不同被分流,IP 业务分组送入 IP 转发模块完成向下一跳路由的转发,MPLS VPN 标签分组送入 MPLS 转发处理模块处理,根据分组标记进行查表操作,由查表结果进行 MPLS 标记交换和报头处理,最后由队列管理和交换调度模块将分组交换调度到输出接口,再由线路接口模块进行数据链路层封装后从出端口发送。

MPLS 转发表模块根据转发表模块中的转发条目完成数据报的转发工作,涉及的主要程序有:ldp_mpls.h ldp_mpls.c。

4 主要模块设计

4.1 LDP 分发协议模块描述

(1) LDP 模块功能

为标签信息库中的路由分配标签;将本地路由分配的标签转给上游。

(2) LDP 模块中的数据流图

整个结构分为上游有限状态机(upstream FS)和下游有限状态机(downstream FS)。针对一个 FEC,各种标记的处理信息记录保存在属性(Attr)中,上游的 Attr 保存与上游标记交换路由器相关的标记分发状态,下游的 Attr 保存与下游标记交换路由器相关的标记分发状态信息。上下游的标记分发完成以后,就可以形成 FEC→OUT 或 IN→OUT 的标记转发条目^[4]。

由图 2 可以看出 FEC 与上游的 FS 是一对多的关系,这是因为下游标签交换路由器(Label Switching Router, LSR)可以为不同的上游 LSRs 针对同一个 FEC 分配不同的标记;但是 FEC 与下游的 FS 是一一对应的。

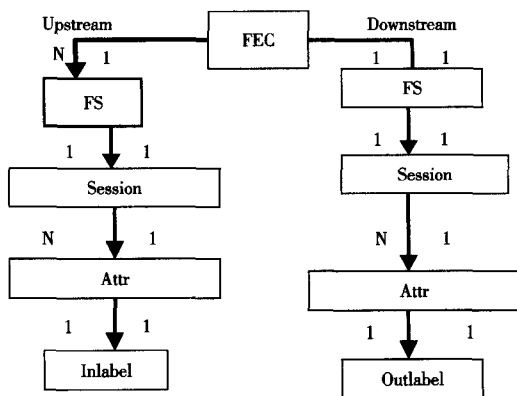


图 2 建立标记交换路径的操作对象关系图

在一对会话体(Session)之间,对于相同的 FEC

可能会根据不同的属性(Attr)分发不同的标记,所以 Session 与 Attr 的关系是一对多的关系。与 IP 地址有关的数据结构,如地址结构 addr_tree、转发等价类 fec_tree 和下一跳 nh_tree 信息都是用树结构来存储。

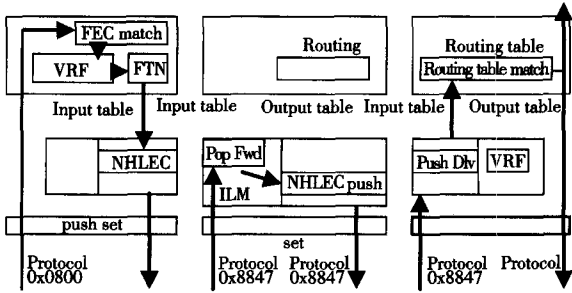
4.2 转发模块

图 3(a)中以太网层接收到在以太数据报头的协议字段值为 0x0800 的帧,它是 IPv4 的代码,以太网将输入的帧转送到 IP 层,MPLS 在 IP 路由表中搜索一个条目以作出路由决定,但因为这个条目匹配一个修改了的 FEC,因此,数据报从 IP 层被转送到 MPLS 层,包含在 IP 路由表中的附加信息是 VRF,根据 FEC 查到的 VRF 条目,将 IP 报打上内层标记,VRF 中有一个指针指向与它关联的 FTN,就是一个指针指向 MPLS 输出表条目,这个输出表条目包含两个结构的 NHLFE。一个 PUSH 结构定义数据报的标签号,一个 SET 结构定义数据报应发送到的端口。MPLS 层在数据报的开头增加一个在 NHLFE 中发现的标签,然后,将数据报发送到以太网层。以太网层产生一个协议字段设置为分派给 MPLS 单播数据报的 0x8847 的帧,把这个帧通过介质发送出去。

图 3(b)LSR 的以太网层接收一个在以太网头中协议字段设置为 0x8847 的帧,因为 0x8847 是分派给 MPLS 单播数据报在以太网中编码的代码,所以以太网层将这帧转送到 LSR 的 MPLS 层。MPLS 在输入表中搜索与数据报垫片头中嵌入标签相匹配的条目,输入表执行输入标签映射(ILM),告诉 MPLS 对数据报所何处理,输入表条目包含两个结构。POP 结构告诉 MPLS 移去 MPLS 头,而 FWD 结构指向 MPLS 输出表的一个条目。这个条目依次包含两个结构: PUSH 结构包含新的数据报标签,告诉 LSE 用这个新的标签在数据报中增加一个垫片头,而 SET 结构告诉 LSR 数据报应发送到以太网哪个接口,然后以太网层用协议字段代码 0x8847 组帧,通过这种定义,NHLFE 告诉 MPLS 路由器数据报头是弹出或是交换,在 MPLS-Linux 中 SWAP 操作通过弹出和压入一个垫片头来实现,这个结构要求弹出和压入一个在每下一个 MPLS 表中查找到的标签。这种情况下 NHLFE 同时包含在输入表和输出表中。

图 3(c)显示了在出口 LSR 一个标签是如何被弹出的。LSR 的以太网层接收一个在以太网头中协议字段设置为 0x8847 的帧,所以将这帧转送到

MPLS 层. 倒数第二层机制已将外层标记去掉,将剩下内层标记的分组送 VRF 处理,导出标记,恢复为 IP 报,从对应端口输入^[6].



(a)入口 PE 压入 (b)P 交换 (c)出口 PE 弹出
标签过程 标签过程 标签过程
图 3 FTN、ILM 在转发结构中所处的位置

4.3 转发表组装模块

4.3.1 转发表的设计

双层标签的结构如图 4 所示.

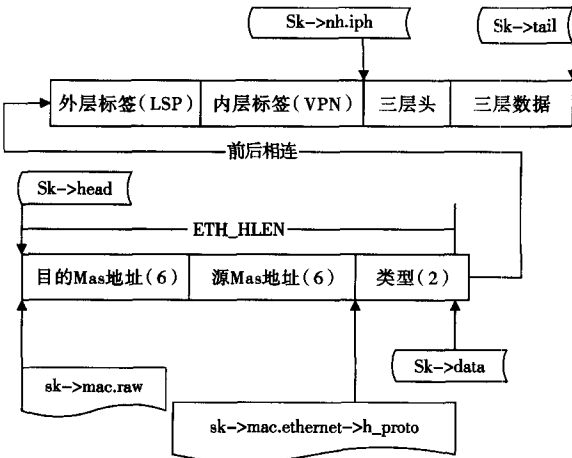


图 4 双层标签

数据打上标签后,垫片指针的位置 * shimhp = (struct shimhdr *)((uint8 *)sk->mac.raw + ETH_HLEN).

在图 5 中,在 BGP MPLS VPN 的入口边缘路由器 PE1 上,按照图 6 转发表中 VRF 表将分组打上内层 VPN 标签 8,利用 FTN 表将分组打上外层标签 7,在中间供应商路由器 P1 上,按 ILM 表交换外层标签(外层的标签 7 换成标签 9),在倒数第二跳供应商路由器 P3 上按 ILM 表去掉外层标签 9,在供应商出口路由器 PE2 上,按 VRF 表去掉内层 VPN 标签 8,然后按普通分组进行转发^[6].

在此设计中,考虑使用每平台范围的标记空间的情况,以及路由器的通用性,将 VRF 表、FTN 表、

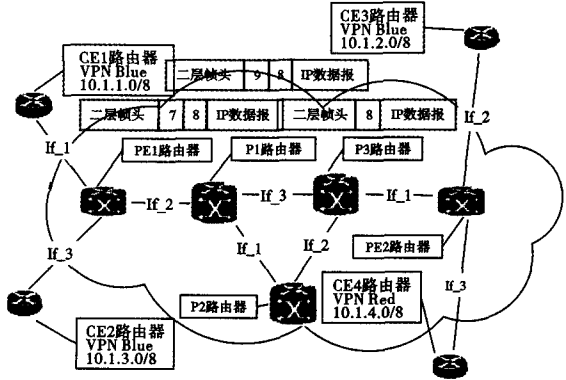


图 5 MPLS VPN 数据报转发过程

ILM 表三表合一进行设计,如图 7 所示. 由操作码决定查询的转发条目,缩短查表时间,提高查表效率,加快数据报的转发^[7].

VRF 表(PE1 路由器上)						
入端口 (IF)	转发等价类 (FEC, 目标主机)	入标签	出标签	出端口 (OIF)	下一路 (出口 PE 路由器)	操作码
If1	10.1.2.0/8	8	—	If2	PE2	PUSH_AND_LOOKUP

FTN 表(PE1 路由器上)						
入端口 (IF)	转发等价类 (FEC, 目标主机)	入标签	出标签	出端口 (OIF)	下一路 (出口 PE 路由器)	操作码
If2	10.1.2.0/8	—	7	If2	PE2	PUSH

ILM 表(P1 路由器上)						
入端口 (IF)	转发等价类 (FEC, 目标主机)	入标签	出标签	出端口 (OIF)	下一路 (出口 PE 路由器)	操作码
If2	10.1.2.0/8	7	9	If3	PE2	SWAP

ILM 表(P3 路由器上)						
入端口 (IF)	转发等价类 (FEC, 目标主机)	入标签	出标签	出端口 (OIF)	下一路 (出口 PE 路由器)	操作码
If3	10.1.2.0/8	9	—	If1	PE2	POP

VRF 表(PE2 路由器上)						
入端口 (IF)	转发等价类 (FEC, 目标主机)	入标签	出标签	出端口 (OIF)	下一路 (出口 PE 路由器)	操作码
If1	10.1.2.0/8	8	—	If2	PE2	POP

图 6 转发表(VRF、FTN、IML 表)

路由器	转发条目类型	入端口 (I,IF)	转发等价类 (FEC, 目标主机/出口 PE 路由器)	入标签	出标签	出端口 (OIF)	下一站 (出口 PE 路由器/P 路由器/CE 路由器)	操作码
PE1	VRF	if_1	10.1.2.0/8	—	11	—	PE2	PUSH_AND_LOOKUP_FOR_VC
	FTN	if_1	PE2	—	12	if_2	P1	PUSH
	VRF	if_1	10.1.3.0/8	—	16	if_3	PE1	PUSH_FOR_VC
	VRF	if_1	10.1.2.0/8	—	8	—	PE2	PUSH_AND_LOOKUP
	FTN	if_1	PE2	—	7	if_2	P1	PUSH
	VRF	if_1	10.1.3.0/8	—	18	if_3	PE1	PUSH_AND_LOOKUP
P1	FTN	if_1	PE1	—	—	—	—	—
	VRF	if_1	—	—	—	—	—	—
	ILM	if_1	10.1.3.0/8	16	—	if_3	10.1.3.0/8	POP_FOR_VC
P3	ILM	if_1	10.1.3.0/8	18	—	if_3	10.1.3.0/8	POP_FOR_VPN
	ILM	if_2	PE2	12	22	if_3	P3	SWAP
	ILM	if_2	PE2	7	9	if_3	P3	SWAP
PE2	ILM	if_2	—	—	—	—	—	—
	ILM	if_3	PE2	22	—	if_1	PE2	POP
	ILM	if_3	PE2	9	—	if_1	PE2	POP
PE2	ILM	if_3	—	—	—	—	—	—
	ILM	if_1	10.1.2.0/8	11	—	if_2	10.1.2.0/8	POP_FOR_VC
PE2	ILM	if_1	10.1.2.0/8	8	—	if_2	10.1.2.0/8	POP_FOR_VPN

图 7 VRF、FTN、IML 三表合一

转发表中的操作码指派对标签和非标签数据报

的处理如下:

(1) PUSH: 在 FTN 条目中找到这个操作码. 转发器在 IP 报的头创建一个垫片, 压入这个特定的标签, 当建立以太帧时, 转发器产生 MPLS 单播数据报, 设置帧类型到 ETH_P_MPLS_UC.

(2) POP: 在 ILM 条目中找到这个操作码. MPLS 转发器从标签数据报垫片中弹出特定的标签. 如果在协议栈中这个标签是最后一个标签, 将数据报转发到 IP 层, 如果协议栈中有多个标签, 转发器根据标签中的操作码作出决定.

(3) SWAP: 在 ILM 条目中找到这个操作码. MPLS 转发器用新的出标签交换特定的标签, 从而转发数据报.

(4) POP_FOR_VPN: 在 ILM 条目中找到这个操作码. 转发器从标签数据报垫片中弹出特定的标签. 这个标签是协议栈中的最后一个标签, 也就是非标签数据报依据在 ILM 条目中的目的(next-hop and/or out-going interface)转发.

(5) PUSH_AND_LOOKUP: 在 VRF 条目中找到这个操作码. 转发器在 IP 报的头创建一个垫片, 压入这个特定的标签. 接着, 它为全局 FTN 表中的请求 FEC 查找一个 FTN 条目. 如果没有 FTN 条目, 数据报被丢弃. 当建立以太帧时, 转发器产生 MPLS 单播数据报, 设置帧类型到 ETH_P_MPLS_UC.

(6) PUSH_AND_LOOKUP_FOR_VC: 在 FTN 条目中找到这个操作码. 转发器在以太帧的头部创建一个 MPLS 垫片, 并压入特定的标签. 接着, 在全局 FTN 表中为特定终端的 VC(Virtual Circuit)查找一个 FTN 条目. 如果找不至 FTN 条目, 丢弃数据报. 如果找到, 修改过的帧被编码后转发.

(7) PUSH_FOR_VC 在 FTN 条目中找到这个操作码. 转发器在以太帧的头部创建一个 MPLS 垫片, 并压入特定的标签. 修改过的帧被编码后转发.

(8) POP_FOR_VC: 在 ILM 条目中找到这个操作码. MPLS 转发器从标签数据报中弹出特定标签, 用 ILM 条目将所得到的 Layer-2 PDU 从特定的接口转送出去.

4.3.2 创建转发表

MP-iBGP 发布的消息, 获得邻居 PE 路由器中对等实体的路由条目信息及路由参数信息, 组装 VRF 表. 转发表安装过程类似, 组装 FTN 信息用函数 `ldp_mpls_ftn_label_install(ldp_fec_cb, ldp_label)` 来实现, 组装 ILM 信息用函数 `ldp_mpls_ilm_`

`label_install(ldp_fec_cb, ldp_label)` 来实现, 在这里用到两个结构: 转发等价类结构和标签结构, 这两个结构用图 8 表述.

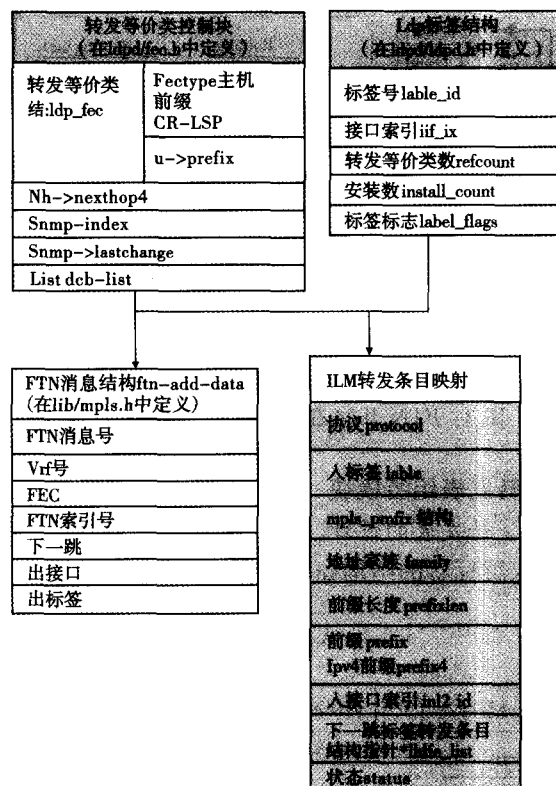


图 8 转发等价类结构和标签结构

5 结束语

构成 BGP MPLS VPN 网络的路由设备有: P 路由器、PE 路由器及 CE 路由器三类. 路由器的控制平面通过 LDP 子模块、转发表组装模块生成标记转发信息库, 数据平面通过数据转发子模块完成数据分组的转发处理. 设计的 BGP MPLS VPN 模块应用在快速自愈路由协议与试验系统中, 运行良好.

参考文献:

- [1] Puneet Agarwal, Bora A. Akyol. IETF working in progress, draft-ietf-mpls-ttl-00. TTL Processing in MPLS Networks[S]. 2002.
- [2] Juniper Networks. BGP/MPLS VPN Fundamentals. RFC 2547bis [S]. 2000.
- [3] E. Rosen, A. Viswanathan, R. Callon. Multiprotocol Label Switching. Architecture, IETF RFC-3031 [S]. 2001.

(下转第 147 页)

程序设计流程如图2所示。

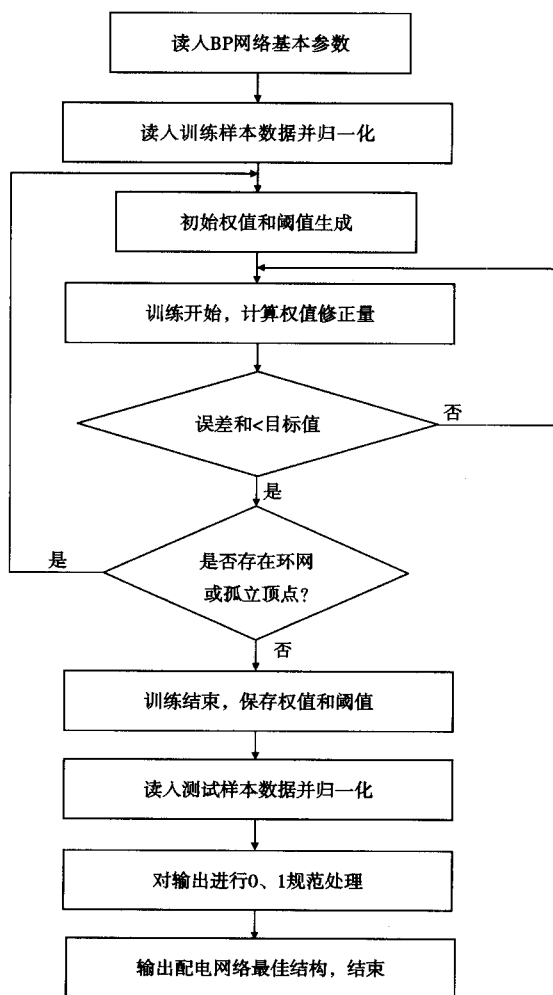


图2 配电网络重构BP训练流程图

5 结束语

将BP神经网络应用到配电网络重构算法的研究中,配电网络根据图论知识的具体拓扑结构要求,有效地改进与约束了配电网络,BP神经网络经过改

进后,其网络学习速度慢和收敛到局部最小等缺点得到了极大的改善。BP神经网络经过训练后,能针对具体配电网的负荷模式,输出最优的配电网开关状态,实现配电网最优重构,表现出BP神经网络输出速度快,能全局逼近的优点。进而实现了配电网故障定位,网络重构等,提高了供电的可靠性。

参考文献:

- [1] 刘健. 配电网理论及应用[M]. 北京:中国水利水电出版社,2007.
- [2] 孙红辉,王红霞,田涛. 一种基于不变距和BP网络的目标识别方法[J]. 微电子学与计算机,2011,28(3): 63-65.
- [3] 孙增圻. 智能控制理论与技术[M]. 北京:清华大学出版社,2003.
- [4] Fan Ji-Yuan, Zhang Lan, McDonald JD. Distribution network reconfiguration: single loop optimization[J]. IEEE Trans on Power Systems, 1996, 11(3): 1643-1647.
- [5] 徐俊明. 图论及其应用[M]. 北京:中国科学技术大学出版社,2010.
- [6] 闫志忠. BP神经网络模型的改进及其应用研究[D]. 长春:吉林大学,2003.
- [7] 戚涌,李千目,刘凤玉. 基于BP神经网络的网络智能诊断系统[J]. 微电子学与计算机,2004,21(10):10-13.

作者简介:

陈中孝 男,(1963—),副教授.研究方向为智能信息处理与智能控制.

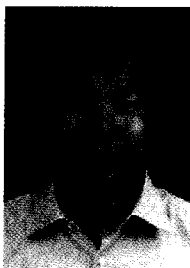
张晓斌 男,(1987—),硕士研究生.研究方向为检测技术与自动化装置.

李雷 男,(1985—),硕士研究生.研究方向为控制工程与控制理论.

(上接第142页)

- [4] 潘显志,李向丽,邱保志. 基于优先级的MPLS重路由转发机制[J]. 微电子学与计算机,2008(03): 149-151.
- [5] Ivan Pepelnjak, Jim Guichard 著,赵斌,陈文飞,徐鸿文译. MPLS和VPN体系结构[M]. 北京:人民邮电出版社,2003.
- [6] Andersson L, Doolan P, Feldman N, A. Fredette, B. Thomas. LDP Specification. RfC 3036, [S]. 2001.
- [7] 康银婵,汤小春. SSL VPN系统中自主签发证书的设计与实现[J]. 微电子学与计算机,2009(04):121-124.

作者简介:



李海华 男,(1965—),副教授.研究方向为网络协议与信息安全.

冯卫华 男,(1980—),讲师.研究方向为计算机软件与网络.