

# 基于 CR-LDP TLV 的流量工程实现策略

杜 荔, 杨 琳

(东北大学信息科学与工程学院, 沈阳 110004)

**摘要:** 为了在 MPLS 网络中使流量分布趋于合理, 提出了一种基于 CR-LDP(限制路由标记分配协议)中 TLV(类型长度值)的流量工程实现策略及算法。该策略通过对网络资源属性和业务流属性的综合考虑, 依据对特定业务流进行裁决的判别函数, 将适于迁移的 TLV 结构业务流进行合理地及时迁移, 从而达到提高网络资源利用率的目的。仿真结果验证了该策略及算法的正确性和有效性。

**关键词:** 多协议标记交换; 流量工程; 限制路由标记分配协议; 显式路由标记交换路径

## Strategy for Implementing Traffic Engineering Based on CR-LDP TLV

DU Li, YANG Lin

(College of Information Science & Engineering, Northeastern University, Shenyang 110004)

**【Abstract】** In order to distribute the traffic in reason for MPLS networks, this paper proposes a new kind of strategy and algorithm based on the type length value in constraint route-label distribution protocol for implementing traffic engineering. Considering the attributes of network resource and the attributes of traffic flows synthetically, the strategy can transfer the suited traffic flows of TLV structure in time and enhance the utilization rate of network resources according to the estimation function for certain traffic flows. Furthermore, the validity of the strategy and algorithm is verified by simulation.

**【Key words】** MPLS; Traffic engineering; CR-LDP; ER-LSP

流量工程是采用适当方法对网络流量进行测量、建模、描述、优化和控制的一种网络工程技术<sup>[1]</sup>, 多协议标记交换(MPLS)<sup>[2,3]</sup>适合流量工程的最大特点之一就是拥有显式路由标签交换路径(ER-LSP)<sup>[4]</sup>。在QoS要求不十分严格的条件下, 使用一般基于MPLS的流量工程算法会造成网络资源的巨大浪费, 本文提出的基于CR-LDP TLV的动态迁移策略能够优化控制网络流量, 以实现网络资源的有效利用。

### 1 模型建立及 TLV 格式新定义

#### 1.1 模型建立

MPLS 流量工程的 3 个主要方面是相辅相成的: 流量中继属性表明业务数据对网络的要求, 资源属性描述网络提供给流量中继的资源情况, 限制路由则可以根据资源属性表现的资源状况和流量中继属性的要求来衡量或裁决现有网络资源能否满足业务的要求。

本文使用  $G = (V, E)$  表示网络的物理拓扑。其中,  $V$  代表网络中所有结点的集合,  $E$  代表所有结点间链路的集合; 使用  $Path(s, d)$  表示源结点  $s$  到目的结点  $d$  之间所有可达路径的集合。其中的每一元素  $Path_i(s, d)$  表示一条不重复的路径;  $Path_i(s, d)$  的数据结构包括:  $Pid$ (唯一标识路径的路径标识),  $Hop$ (该  $Path$  从源结点到目的结点所经过的链路数),  $N$ (该  $Path$  被选中的次数),  $TBw$ (该  $Path$  的总带宽),  $UBw$ (该  $Path$  上已使用过的带宽),  $PIL$ (该  $Path$  在业务流传输过程中所处地位的重要性),  $PTL$ (该  $Path$  目前所承担的负载状况);  $Flow_i(s, d)$  的数据结构包括:  $Fid$ (业务流标识),  $FBw$ (业务流所占用的带宽),  $LoP$ (业务流迁移优先级); 使用  $EBR$  表示网络等效空闲带宽比率, 其中  $n$  为网络拓扑中的路径总数。

$$EBR = \sum_{i=1}^n \Delta \times \frac{Bw_i(s, d) - UBw(s, d)}{Bw_i(s, d)} \quad (1)$$

本算法将 PIL、PTL 及 LoP 三者配合作为迁移策略的依据, 每次选择一条满足迁移策略条件的业务流对其进行迁移。为避免复杂的竞争机制, 同时加快算法的收敛性, 可将符合迁移条件的业务流先按所占带宽的大小降序排列, 然后依次尝试通过最后的判别函数, 满足判别函数的最大带宽占有者将成为最适合迁移的业务流。

#### 1.2 TLV 格式新定义

限制路由的标记分配协议(CR-LDP)沿用类型长度值(TLV)的编码结构。本文提出了一种新的“子”TLV, 需要扩展标准的 TLV——资源分类 TLV 和 FEC 要素 TLV。

##### 1.2.1 标记请求消息扩展

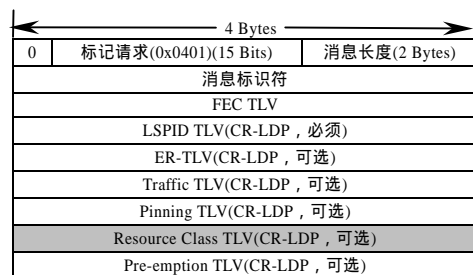


图 1 标记请求消息扩展编码

**基金项目:** 国家“863”计划基金资助项目(2002AA784030)

**作者简介:** 杜 荔(1962-), 女, 博士、副教授, 主研方向: 宽带通信领域中的 IP 交换技术, 网络管理技术; 杨 琳, 硕士

**收稿日期:** 2006-03-25 **E-mail:** duli@ise.neu.edu.cn

标记请求消息编码格式如图 1。其中资源分类 TLV 用来指定哪些链路可以被 CR-LSP 接受, 该信息考虑到网络拓扑的修改。资源分类 TLV 编码格式如图 2。

U	F	Type = 0x0822	Length = 4
RsCls			
PIL			
PTL			

图 2 资源分类 TLV 编码

在该 TLV 中定义了两个新的“子”TLV, 即:

(1)PIL 域: 表示该 Path 在业务流传输过程中所处地位的重要性, PIL 分为 3 个等级: Backbone, Minor, Normal。其中, Backbone 表示网络传输中的众多业务流都将经过的路径, 在路由选择策略中处于重要地位; Normal 表示普通的、无足轻重的路径; Minor 则表示其重要性介于 Backbone 和 Normal 之间的路径。在迁移策略中, 要尽量把业务流向 Normal 路径上面迁移, 其次是 Minor 路径, 使得 Backbone 尽可能空出足够的带宽以满足新业务流的要求; 同时, Normal 上的业务流不能向 Minor 上迁移, Normal 和 Minor 上的业务流也不能向 Backbone 上迁移, 这就从两方面保证 Backbone 带宽的充裕度, 使得高带宽的 Backbone 路径可接纳更多业务流, 并提供更完备的 QoS 保证。

(2)PTL 域: 表示该 Path 目前所承担的负载状况, PTL 分为 3 个等级: Heavy, Middle, Light。若某路径 PTL 值为 Heavy, 则不应再向该路径迁移业务流, 而应依次向 PTL 值为 Light 和 Middle 的路径上优先迁移。PTL 属性是动态的, 它从路径当前承担的负载程度方面来说明业务流是否应进行迁移, 随负载状况的变化而改变; 而 PIL 属性是静态的, 它从路径策略方面来指出业务流是否适合迁移。

### 1.2.2 FEC 类型长度值编码扩展

一个 FEC 由多个要素组成, FEC TLV 编码格式如图 3。FEC 要素的类型长度值编码格式如图 4。

U	F	FEC(0x0100)
长度(2 Bytes)		
FEC 要素 1(4 Bytes)		
...		
FEC 要素 n(4 Bytes)		

图 3 FEC TLV 编码

U	F	类型	长度
LoP			

图 4 FEC 要素 TLV 编码

图 4 中的 LoP 域为标准业务量参数 TLV 中的扩展子 TLV, 表示业务流的迁移优先级。LoP 分为 3 个等级: Forbid, Permit, Free。其中, Forbid 表示该类业务流优先等级最高, 禁止对该类业务流进行迁移, 适合于对 QoS 提出较高要求的各类业务; Free 表示该类业务流优先等级最低, 没有任何限制, 可以任意迁移, 适合于对 QoS 没有严格要求的数据业务; Permit 表示该类业务流的优先等级介于 Forbid 和 Free 之间, 可在一定条件下对该类业务流进行迁移。

## 2 判别函数及算法

### 2.1 判别函数

在本算法中对于特定业务流是否适合迁移到特定路径的最终判别要取决于判别函数。该判别函数预期达到的要求包括: 路由跳数 Hop, 带宽碎片要求 Bw, 空闲带宽比率  $\phi$ 。

$$Hop = \min\{Hop(Path_i(s,d))\}$$

$$Path_i(s,d) \in Path(s,d) \quad (2)$$

$$Bw = \min\{|FBw_i(s,d) - (Bw_i(s,d) - UBw_i(s,d))|\} \quad (3)$$

$$\phi = \max\left\{\frac{Bw_i(s,d) - UBw_i(s,d)}{Bw_i(s,d)}\right\} \quad (4)$$

判别函数计算出的迁移值 T 越大则表示越适合迁移, 迁移值 T 计算公式为

$$T = \frac{-\alpha|FBw_i(s,d) - (Bw_i(s,d) - UBw_i(s,d))| + \beta \lg\left(\frac{Bw_i(s,d) - UBw_i(s,d)}{Bw_i(s,d)}\right)}{Hop(Path_i(s,d))} * \gamma \quad (5)$$

其中,  $\alpha$  和  $\beta$  分别表示所需带宽与可用带宽之差和可用带宽比率的权重系数。  $\gamma$  表示滚降系数, 可根据如下公式计算:

$$\gamma = (\omega)^N \quad (6)$$

其中,  $\omega$  值应随路径被选中次数 N 的取值范围不同而不同。

$$\omega = \begin{cases} 0.7 & (N \leq 5) \\ 0.8 & (5 < N < 10) \\ 0.9 & (N \geq 10) \end{cases} \quad (7)$$

### 2.2 算法描述

算法描述如下:

步骤 1 根据网络流量的变化情况, 激发算法检测程序。

步骤 2 检测路径 PIL 和 PTL 属性, 判断该路径是否适合业务流的迁移。若适合, 则转到步骤 3; 否则, 转到步骤 8。

步骤 3 判断路径当前的空闲带宽是否能够满足当前某些业务流的带宽要求。若满足, 获得一个业务流集, 转到步骤 4; 否则, 转到步骤 8。

步骤 4 检测业务流 LoP 属性, 查看业务流集内各项是否适合迁移。若存在适合者, 转到步骤 5; 否则, 转到步骤 8。

步骤 5 对所有符合条件的业务流按所占带宽大小降序排列。

步骤 6 启动判别函数依次检测业务流, 若可以通过判别函数, 则转到步骤 7; 否则, 去除该业务流后转回步骤 6。

步骤 7 通过 ER-LSP 对该业务流进行迁移。

步骤 8 结束。

### 3 仿真及结果

仿真采用的网络拓扑如图 5 所示。路径 1(A-B-C-D)、路径 2(A-E-F-D)、路径 3(A-G-H-D) 的瓶颈带宽分别为 1Mbps、0.5Mbps 和 0.8Mbps。源结点发送 UDP 协议承载的 CBR 数据流。除端结点为普通结点外, 其余结点均为 MLPS 结点。各迁移属性设置如下: 为方便业务流的迁移, 参与仿真的业务流 LoP 值均为 Free; 路径 1 的 PIL 为 Backbone, 路径 2 的 PIL 为 Minor, 路径 3 的 PIL 为 Normal; 各路径的 PTL 值根据带宽占用比率的动态情况进行调整, 规定其调整的阈值如下:

$$PTL = \begin{cases} \text{Heavy} & (\theta > 80\%) \\ \text{Middle} & (45\% \leq \theta \leq 80\%) \\ \text{Light} & (\theta < 45\%) \end{cases} \quad (8)$$

$$\Delta = \begin{cases} 0.6 & (PIL = \text{Backbone}) \\ 0.3 & (PIL = \text{Minor}) \\ 0.1 & (PIL = \text{Normal}) \end{cases} \quad (9)$$

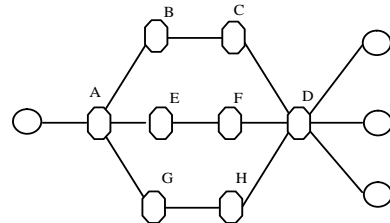


图 5 网络拓扑

本仿真中主要依据网络等效空闲带宽比率对算法质量进行评估。若网络等效空闲带宽比比较大, 则表明该网络中流量分布适当, 可以接收更大流量的业务流, 或者提供更高的 QoS 支持; 反之, 则表明网络已接近饱和, 或只能支持一些较低 QoS 要求的数据业务。本文对两种典型情况进行了仿真试验

(试验结果如图 6 和图 7 所示), 一种是网络中的大多数业务流都是较粗壮的, 即所占的带宽比较大; 另外一种是在网络中存在的业务流都是较纤细的, 即所占的带宽比较小。

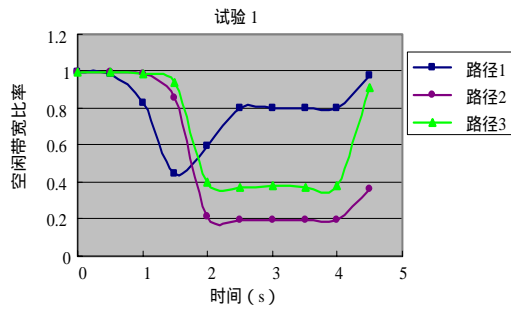


图 6 业务流为粗壮型时的空闲带宽比率

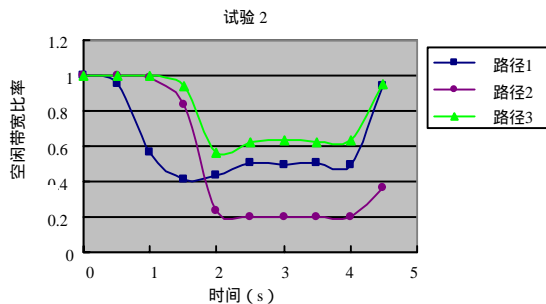


图 7 业务流为纤细型时的空闲带宽比率

从图 6、图 7 中可以看出, 伴随着两次试验中业务流的迁移, PIL 值为 Backbone 的路径 1 具有了更大的空闲带宽, 这有助于像路径 1 这样硬件条件优越的快速主干路径接纳数量更多、对带宽要求更高的业务流, 提供高质量的 QoS 保证。同时还可以看出, 尽管试验 2 中路径 3 的空闲带宽比率高于试验 1 中的, 但由于路径 3 的 PIL 等级低, 所以在图 8 两次试验的 EBR 对比中, 其 EBR 值仍要低于试验 1 约 20%。这说明当网络中的业务流多为粗壮型的情况下, 运用该算法比在业务流多为纤细型的情况下效果好得多。随着网络规模的

(上接第 109 页)

空, 则从链表上取下一个地址将表项添加进去; 若为空, 表明此段已写满有效表项。考虑到最长匹配要求, 先以前缀长度为  $m$  的段为中心分别向上和向下遍历各段, 找到最近的空闲链非空的段(设为段  $n$ )。

如图 2(b)所示, 若段  $n$  在段  $m$  的上方, 首先从前缀为  $n$  的段开始, 将前缀为  $(n-1)$  的段的最后一项移动到前缀为  $n$  的段的最后一项(即前缀  $n$  段的空闲链的末节点), 这样在前缀长为  $(n-1)$  的段就有了一个空闲节点, 然后将前缀为  $(n-2)$  的段的最后一项移动到前缀为  $(n-1)$  的段的最后一项, 使得前缀  $(n-2)$  的段有空闲节点, 依此类推, 直到前缀为  $m$  的段有空闲节点, 这时只需把表项添加到该空闲节点即可。若段  $n$  在段  $m$  的下方, 首先将段  $n$  的第一项移动到段  $n$  空闲链表的第一项, 然后将段  $(n+1)$  的第一项移动到段  $n$  的第一项, 依此类推, 直到前缀为  $(m-1)$  的段的第一项成为空闲节点, 将表项添加到该空闲节点即可。

由于采用选择移动, 因此表项移动的个数大大减少了, 提高了算法的运行效率。限于篇幅, 算法的具体实现不再给出。

扩大, 越来越多的新业务提出了严格 QoS、高带宽的要求, 该算法正适应了这一现实的演进。可以预见, 在业务流大多为粗壮型的未来, 本算法的优越性将得到充分的发挥。

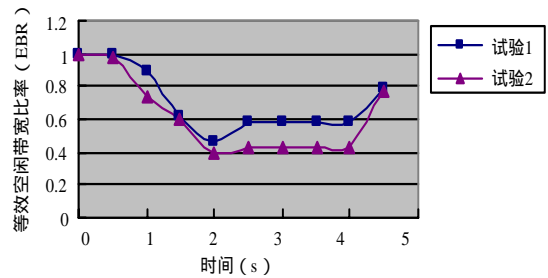


图 8 两次试验的 EBR 对比

## 4 结论

本文通过对网络资源属性和业务流属性的综合考虑, 定义了 PIL、PTL 和 LoP 3 个属性及判别函数, 并赋予 CR-LDP 中的 TLV 编码以新的定义, 提出了一种利用 MPLS 的显示路由技术实现流量工程的业务流迁移策略。通过将部分业务流进行及时的迁移, 使得网络资源可根据需要得到不同级别的利用, 从而达到在无须提升网络资源的条件下提高网络的吞吐能力, 使网络资源得到更加充分利用的目的。

## 参考文献

- 1 Juniper Networks, Traffic Engineering for the New Public Network [EB/OL]. 1999. <http://www.juniper.net>.
- 2 Awduche D O. MPLS and Traffic Engineering in IP Networks[J]. IEEE Communication Magazine, 1999, 37(12): 42-47.
- 3 Boutaba R, Szeto W, Iraqi Y. DORA, Efficient Routing for MPLS Traffic Engineering[J]. Journal of Network and Systems Management, Special Issue on Internet Traffic Engineering and Management, 2002, 10(3): 309-325.
- 4 Awduche D O, Malcolm J, O'Dell M. Requirements for TrafficEngineering over MPLS[Z]. 1998-10. draft-ietf-mpls-traffic-eng-00.txt.

## 4 总结

本文给出了一种基于 TCAM 的二级路由查找方案。该方案提出了一种基于比特选择的路由表项分类机制, 大大减小了转发查表引擎所消耗的功率。此外, 我们在国家“863”计划项目“可扩展到 T 比特的高性能 IPv4/v6 路由器基础平台及实验系统”上对该方案进行了试验, 取得了较好的结果。该方案的一大缺点就是提高了表项更新的难度, 但是快速更新算法的出现<sup>[1]</sup>使问题得到了解决。

## 参考文献

- 1 Shah D, Gupta P. Fast Updating Algorithms for TCAMs[J]. IEEE Micro, 2001, 21(1): 36-47.
- 2 Zane F, Narlikar G, Basu A. Power-efficient TCAMs for Forwarding Engines[C]//Proc. of IEEE Infocom'03. 2003.
- 3 Netlogic Microsystems Inc.. NSE5000 RA and RB Network Search Engine Devices 18MB, 9MB and 4.5 MB[EB/OL]. 2003-11-20. <http://www.netlogicmicro.com>.
- 4 Gupta P, Prabhakar B, Boyd S. Near-optimal Routing Lookups with Bounded Worst Case Performance[C]//Proceedings of Infocom'00. 2000: 1184-1192.