

基于流量总时延最小的并行 LSP 自适应流量分配

邵 旭, 丁 炜

(北京邮电大学培训中心, 北京 100876)

摘 要: MPLS 的引入使 IP 网络的流量工程成为可能. 如何平衡两个标记交换路由器的多条并行显式标记交换路径上的网络流量, 从而达到避免拥塞和优化网络性能的目的, 是流量工程研究的一个重要课题. 本文指出流量总时延最小是最有价值和最实用的流量工程优化目标, 通过分析平均分组时延和网络负载的约束关系, 证明了优化解的存在, 解出了其数学结果, 提出了一种自适应流量分配新方法, 理论分析和仿真结果都表明该方法具有简单、快速、收敛性强等优点.

关键词: IP QoS; 流量工程; 多协议标记交换; 自适应

中图分类号: TN913.24 **文献标识码:** A **文章编号:** 0372-2112 (2003) 01-0004-04

Adaptive Traffic Distribution on Parallel LSPs Based on the Minimization of Total Traffic Delay

SHAO Xu, DING Wei

(Training Center, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: The introduction of MPLS into Internet makes the Internet traffic engineering possible. It is important to balance the traffic load among several parallel explicit LSPs so as to avoid congestion and optimize the network capacity. The total traffic delay is the most valuable and practicable optimization object for traffic engineering. By analysing the constraint relationship between average packet delay and traffic load, the existence of optimum solution is proved and its mathematical result is gotten. A novel adaptive traffic contribution method based on such optimization object is proposed in the paper. Both theoretical analyses and simulation bear out the merits of simpleness, speediness and convergence.

Key words: IP QoS; traffic engineering; MPLS; adaptive

1 引言

MPLS(多协议标记交换)的日渐成熟,使 IP 网络流量工程的实现不再依赖于特定的链路层技术,如 ATM,从而使更加高效的 IP 网络的流量工程(Traffic Engineering)成为可能^[1]. 流量工程的主要目的是增加运行网络的效率,提高其可靠性,同时优化网络资源的使用和流量性能. 目前, IETF 的因特网流量工程工作组(TEWG)负责流量工程领域的标准化工作^[2]. 随着 ISP 网络规模的日益扩大和流量的激增,以及 ISP 之间的网络服务方面的激烈竞争,流量工程将变得越来越重要.

在多条并行的显式 LSP(标记交换路径)(可以通过 CR-LDP、扩展的 RSVP 或是手工配置的方式建立)之间合理的分配流量负荷是流量工程研究的一个重要课题,如图 1 所示. 如果两个节点间的流量超过了某条链路的负荷,那么通过流量工程技术就可以分开流量,然后将它们分别映射到并行的其它 LSP 上. 而且在很多情况下,合理的流量负荷分配能够优化整个网络的管理和性能. 如何在两个 LSR(标记交换路由器)的多条显式 LSP 之间合理的调配流量,是目前流量工程研究

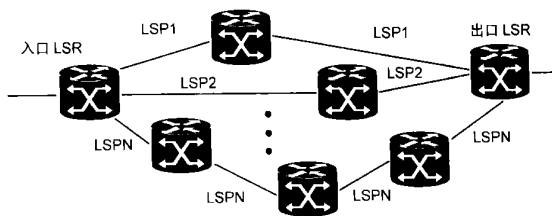


图 1 多条显式 LSP 之间的流量负荷分配问题的热点问题^[3-6,12]. Widjaja 等人首次提出了自适应流量工程的概念,但是只是简单地从避免拥塞的角度出发从高拥塞的 LSP 搬移流量到无拥塞的 LSP,仿真表明在整个网络负荷很小的时候能够很好的工作,随着网络负荷的增加,收敛性变得很差^[3-4]. Esmal Dinan 和 Awduche 等人给出了 MPLS 网络的动态区分模型和分析结果,但是对 Internet 的流量特征太理想化和模型化,得出的数学结果很难应用于实际,仿真的假设条件也过于简单^[5,6]. 本文通过详细考察 LSP 的时延和网络负载的约束关系,分析了多条 LSP 之间实现流量负荷合理分配所

收稿日期:2001-08-06;修回日期:2002-03-01

基金项目:自然科学基金资助项目(No. 69802003);国家“863 计划”资助项目(No. 863-317-9601-02)

应达到的目标,证明了优化解的存在,并从数学上得到了优化解的封闭结果,进一步论证了在实际网络工程中,不可能通过数学求解的方法得到优化解.最后作者提出了一种自适应流量工程的新方法,并通过详细设定仿真条件,对该方法进行了验证,并与其它方法进行了比较.

2 流量工程的目标和理论最优解

如图 1 所示,入口 LSR 和出口 LSR 之间共有 N 条 LSP, 预分配的总流量(即流量速率)为 λ , 分配到各条 LSP 上的流量分别为 x_1, x_2, \dots, x_N . 某条 LSP 上的流量是分配到该 LSP 的流量和 LSP 上固有的流量之和, 设为 $\mu_i, i = 1, \dots, N$, 它和该 LSP 的平均分组时延服从一定的函数关系. 例如, 对于单个网络节点的马尔可夫 $M/M/1$ 服务模型, 可以计算出平均时延为 $w(\mu) = 1 / (\mu - \mu)$, μ 为网络节点的平均分组处理能力^[7]. 当 μ 分别为 200 和 400 时, 平均时延和网络负载关系如图 2 所示. 虽然一条 LSP 有许多 LSR 组成, 流量特征复杂, 不再服从简单的 $M/M/1$ 模型, 但是大量实测数据和理论分析都可以证明, 随着网络负载增加时延急剧增加的基本特性是不变的^[8]. 设第 i 条 LSP 在没有分配流量之前流量为 t_i , 分配的流量为 x_i , 平均分组时延为 $y_i(x_i)$, 定义流量总时延为 $y(x_1, x_2, \dots, x_N) = \sum_{i=1}^N y_i(x_i) \cdot x_i$. 其中, 约束条件为 $\sum_{i=1}^N x_i = \lambda$. 图 3 仿真了 $N=3$, 总流量为 100, $\mu=200$, 三条 LSP 的平均时延和流量负载的关系服从图 2 的曲线时, 流量总时延和各个 LSP 流量的关系. 可以看到, 该曲面存在总时延的极小解. 为求函数 $y(x_1, x_2, \dots, x_N)$ 的极小值, 可以得到拉格朗日方程

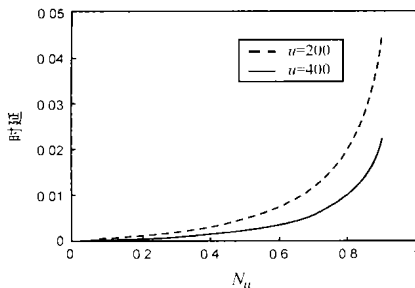


图 2 平均时延和流量负载的关系曲线

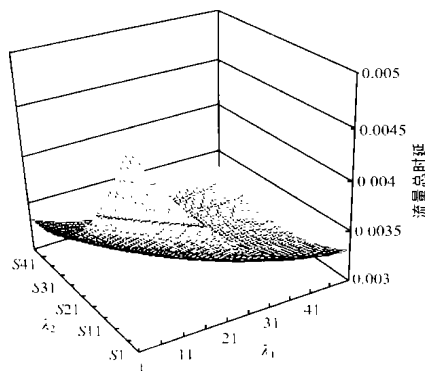


图 3 三条 LSP 情况下的流量总时延

$$\begin{cases} dy/dx_1 = 0 \\ dy/dx_{N-1} = 0 \\ = x_1 + x_2 + \dots + x_N \end{cases} \quad (1)$$

解这个方程, 可以解得当 $x_1 = \lambda_1, x_2 = \lambda_2, \dots, x_N = \lambda_N$ 时, 流量总时延最小, 最小时延为 $y_{\min} = y(\lambda_1, \lambda_2, \dots, \lambda_N)$.

流量的 QoS 参数包括时延, 分组丢失率等, 但是我们只选用了流量总时延最小作为流量工程的目标. 因为一条 LSP 的时延最容易获得, 只需要一条 LSP 两端的 LSR 发送检测分组即可, 而其它 QoS 参数很难用简单的方法实时测量到, 如分组丢失率^[9]. 另外, 目前, IP 网络中大部分业务是基本没有 QoS 要求的数据业务, 分组的丢失可以由 TCP 的重传机制做保障, 所以只有时延具有重要的意义. 流量总时延最小意味着能够更快的释放所占有的网络资源, 意味着网络具有更高转发效率. 所以选用流量总时延最小作为流量分配的目标具有理论和应用的意义.

但是, 实际网络工程中, 却无法使用我们得到的公式进行分配. 因为很难精确知道一条 LSP 的流量负载和时延的封闭数学函数, 而且一条 LSP 的负载为当前分配到该 LSP 的流量负载和其固有的流量负载之和, 而对于后者在实际中是不能实时得到的. 所以, 只能借助于其它方法.

3 自适应流量工程

3.1 算法

将流量分配到 LSP 上一般不是针对单个分组的, 目前普遍采用的方法是先把流量粗分类, 即相当于按照一定的规则分类为若干容器(bin), 称为流量过滤(filtering). 然后再将这些装有流量的容量按照一定的规则分配到各条 LSP 上, 称为流量发布(distribution). 之所以分成两个阶段, 一个重要原因是直接针对 IP 分组的各条 LSP 的流量分配会打乱 IP 分组在 TCP 中的顺序, 而且端 LSR 的处理负担很重. 流量过滤方法很多, 如基于源地址和目的地址, 或是其它方法使每个容器中具有更加均匀的流量^[4], 这不属于本文讨论范围. 流量发布是以容器为单位的, 独立于流量过滤的方法. 本文所讨论的其实是一种自适应流量发布方法. 但是, 为了叙述上的方便, 在不引起混淆的前提下, 本文一律称为流量的分配.

为达到流量总时延最小的流量分配优化目标, 本文提出了一种自适应流量分配算法.

第一步 在对各条 LSP 状态一无所知的情况下, 先平均分配流量, 即 $x_{i,0} = \lambda / N$.

第二步 入口 LSR 向各条 LSP 发送检测分组, 等检测分组从出口 LSR 返回后, 得到各条 LSP 的时延. 然后计算流量总时延. 为克服网络状况的随机性, 从而使算法具有较好的收敛性, 每次得到的各条 LSP 的时延都是在某一时间段得到的时延的平均值.

第三步 我们的思路是将流量从高时延的 LSP 转移到低时延的 LSP. 但是, 需要考虑两个限制条件: (1) 低时延的 LSP 由于新流量的引入会使时延增加, 高时延的 LSP 由于部分流量的转移会使时延减少; (2) 通过负载时延曲线可以看出, 不同的负载初始状态和 LSP 的负载时延特性的不同, 使流量的

增加或减少所引发的时延变化是不同的. 定义时延对负载的导数 = 时延/ 负载, 可以看到负载初始值越高, 导数越大. 所以, 对于那些比较大的 LSP, 在向其增加流量时一定要非常小心, 应该与比较小的 LSP 区别对待. 第 i 条 LSP 第 j 时刻的估算方法为 $y_{i,j} = (y_{i,j-1} - y_{i,j-1}) / x_{i,j}$. 如果前一个时刻没有流量的转移, 则用再往前时刻测得的数值, 依次类推.

把各个 LSP 的时延从小到大排列, 将时延最大的和时延最小的配对, 其次大的和其次小的配对, 以此类推, 设共有 M 对. 设从时延大的 LSP 到时延小的 LSP 转移的流量占总流量的百分比为 α_i , 流量转移逐渐递减, 即 $\alpha_1 > \alpha_2 > \dots > \alpha_M$. 由限制条件(1), 可知权值 α_i 不能取得过大, 需要采用自适应的方法, 逐渐逼近最优的结果. 由限制条件(2), 可知在 α_i 过大的时候需减小 α_i . 作者在这里提出了分段的处理方法, 先设定警戒门限 α_{max} , 测量的 α_i 与 α_{max} 相比较, 如果 $\alpha_i < \alpha_{max}$, 不修正 α_i ; 如果 $\alpha_i > \alpha_{max}$, 修正 α_i , 即 $\alpha_i = \alpha_i / k, k > 1$. α_{max} 和 k 需要根据实际情况而定.

计算完各对之间转移的流量后, 将其转换为各条 LSP 上转移的流量的百分比, 根据此百分比分配当前欲分配的总流量.

第四步 时延一定时间以后, 重新回到第二步.

该算法一直在运行, 不需要中止. 对于网络中固有流量和欲分配的流量的变动, 该算法都能够自适应调整, 并一直停留在最优的状态.

3.2 算法讨论

实际测量表明, 广域网络中的流量变化并没有人们想象的那么剧烈, 至少在 5 分钟内感觉不到明显的变化^[10]. 对于该算法, 只要测量完 LSP 分组平均时延后, 应用该时延进行优化时, 该时延没有显著变化就足够了, 显然是很容易满足的. 而文献[2]所提出的算法对此项条件的要求则苛刻得多.

如果算法开始时, 假设某条 LSP 时延最小, 而且它的 α_i 又比较小的话, 按照前述的自适应算法, 这种 LSP 每次将有较多的流量转移过来. 随着流量由时延大的向时延小的 LSP 的转移, 流量总时延在逐渐降低. 假设由于大量流量的涌入, 该 LSP 的时延虽然在所有 LSP 中还比较小, 但是已经不是最小, 此时算法将会减少流量的流入. 当该 LSP 的时延变得处于中等时, 其流量流入就几乎处于一种停滞状态. 如果 α_i 进一步上升而超过了 α_{max} , 由以上算法可以知道此时流入流量将进一步大幅度降低. 所以一系列的自适应保护机制, 保证了算法的收敛性, 还尽量使每次流量的调整都直接能指向最优结果, 即曲面沿梯度方向向流量总时延的最小值搜索^[11]. 由于时延和

的计算基本上只是与其上一个时刻的网络状况相关, 所以网络状况的改变并不会改变算法的精度. 而始终按照百分比来分配流量的方法使总流量的变动被各条 LSP 根据自己的状况按比例地承担, 流量分配做到了尽量与欲分配的流量的变化无关.

该算法不需要中间节点做任何工作, 出口 LSR 也只是需要简单地返回该 LSP 的时延检测分组, 算法完全是在入口 LSR 运行, 所以对整个 IP 网络和 MPLS 网络的运作机制没有任何影响.

4 仿真分析

参考其它文献所假定的仿真条件^[5]并尽量逼近真实的网络状况, 以 $N=4$ 进行仿真, 即一共有 4 个 LSP, 网络固有的流量速率分别为 80, 90, 100, 220, 服务速率分别为 200, 200, 200, 200, 400, 其时延和流量负载的关系如图 1 所示. 假设欲分配的总流量为 100, 网络状况变化非常缓慢, 欲分配的流量也没有变化, 然后按照此方法进行仿真, 结果如图 4~图 6 所示. 图 4 为各个 LSP 上的总流量与算法运行次数的关系曲线. 图 5 为各个 LSP 上分配的流量与运行次数的关系曲线. 图 6 为流量总时延的学习曲线. 仿真过程中 $\alpha_0 = 2.0\%$, $\alpha_1 = 1.0\%$. 可以看到程序运行不到 20 次就可以收敛到最优取值. LSP1~LSP3 虽然初始网络状态不同, 但是由于服从相同时延和负载关系曲线, 自适应算法最终使它们的总流量相同. LSP1 虽然开始时延很小, 但是受流量和时延的制约, 也不可能无限制的增加流量. 到达优化取值以后, 各条 LSP 的流量的转移仍就有所波动, 波动范围与 α_i 有关, 但是对流量总时延影响并不大.

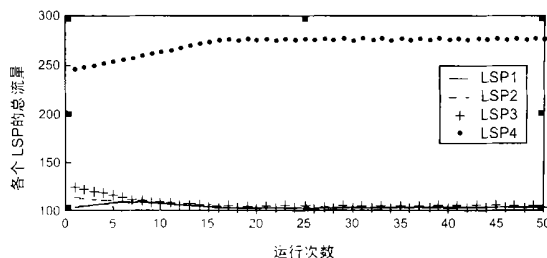


图 4 各个 LSP 的总流量

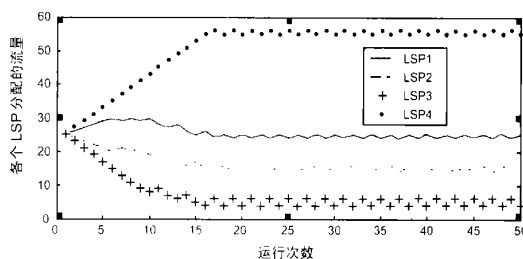


图 5 各个 LSP 分配的流量

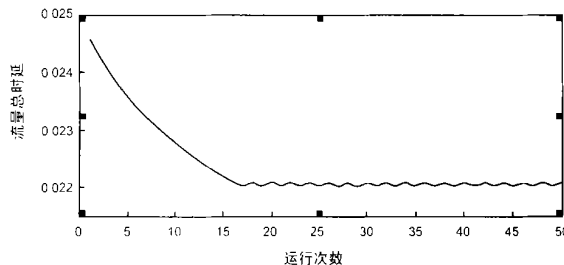


图 6 流量总时延的学习曲线

算法的运行周期取决于计算各条 LSP 时延的速度, 算法最快运行周期是所有 LSP 时延计算的最大值.

假设网络流量负载随时间而变动, 变动规律如图 7 所示. 假设所分配的流量总数为 100, 4 条 LSP 固有的流量速率分别为 70, 100, 110, 210, 其它条件与以上的仿真条件相同, 采用本

文提出的自适应流量分配算法对流量分配进行优化,流量分配结果如图 8 所示.图 9 为流量总时延,与不使用自适应流量分配情况做了比较.可以看到,自适应流量分配能够根据网络流量变化而自适应地调整,极大地优化网络性能.

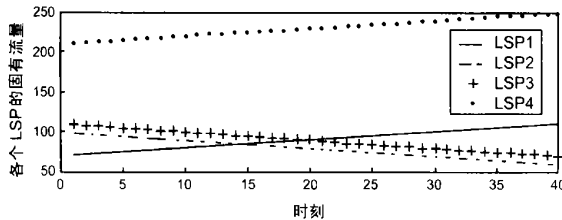


图 7 网络流量负载的变化

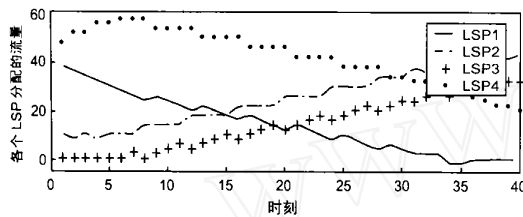


图 8 自适应流量分配的结果

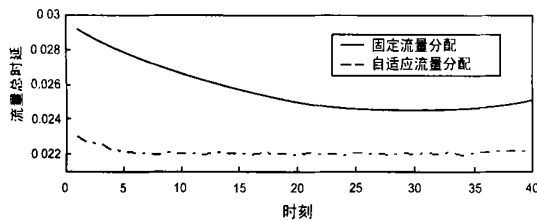


图 9 采用自适应流量分配和固定流量分配流量总时延比较

5 结论

本文所提出的算法以流量总时延最小为优化目标,成功地解决了多条并行 LSP 之间的流量负载平衡问题.理论分析与典型仿真表明,该算法具有收敛性强,收敛速度快、实施简单等优点.但是,该方法目前还是以无 QoS 要求的数据业务为支持的目标的,如何有效地支持 QoS 将是进一步研究的.

参考文献:

- [1] RFC 2702. Requirements for Traffic Engineering Over MPLS. [S].
- [2] A Framework for Internet Traffic Engineering. draft-ietf-tewg-framework-05.txt, Internet Draft [DB/ (DB/ OL)] [S].
- [3] Indra Widjaja, Anwar Elwalid. MATE: MPLS Adaptive Traffic Engineering, draft-widjaja-mpls-mate-01.txt, Internet Draft [DB/ (DB/ OL)] [S]. October, 1999.

- [4] A Elwalid, C Jin, S Low, I Widjaja. MATE: MPLS Adaptive Traffic Engineering [A]. Proc. INFOCOM 01 [C]. 2001.
- [5] E Dinan, D Awduche, B Jabbari. Analytical Framework for Dynamic Traffic Partitioning in MPLS Network [A]. IEEE International Conference on Communications [C]. New Orleans, Louisiana: IEEE, 2000.
- [6] E Dinan, D Awduche, B Jabbari. Optimal Traffic Partitioning in MPLS [A]. Networking 2000 (Net2000) [C]. Paris, France: Net2000, 2000.
- [7] 周炯磐. 通信网理论基础 [M]. 北京: 北京邮电学院出版社, 1992.
- [8] David McDysan. QoS & ATM Networks [M]. The McGraw-Hill Companies, Inc, 2000.
- [9] V Pason, J Mahdavi, A Adams, M Mathis. An architecture for large-scale Internet measurement [J]. IEEE Communication, 1998, 36(8).
- [10] K Thompson, GJ Miller, R Wilder. Wide-Area Internet Traffic Patterns and Characteristics [J]. IEEE Communications, 1997, 6(6).
- [11] 陈尚勤, 李晓峰. 快速自适应信息处理 [M]. 北京: 人民邮电出版社, 1995.
- [12] Hiroyuki Saito, Yasuhiro Miyao, Makiko Yoshida. Traffic engineering using multiple multipoint-to-point LSPs [A]. IEEE INFOCOM 2000 [C]. 2000.
- [13] 邓吉生, 王海兵, 张根度, 荆金华. 基于 MPLS 的流量工程——分布实时网络承载能力估计与分配模型 [J]. 电子学报, 2000, 28(11A): 126 - 129.

作者简介:



邵 旭 男, 1973 年 2 月生于河北省石家庄市, 1996 年和 1999 年分别在西安电子科技大学获得工学学士和工学硕士学位, 现为北京邮电大学 99 级通信与信息系统专业博士研究生, 发表论文 20 篇, 感兴趣的研究领域包括自适应信号处理、移动通信、IP QoS、流量工程等. e-mail: Shaoxuxu @263.net



丁 炜 男, 1935 年 5 月出生于云南省大理市, 北京邮电学院毕业, 教授、博士生导师, 长期从事 ISDN、ATM、IP 技术和宽带通信网接入技术研究, 出版 3 本专著, 发表学术论文百余篇, 享受政府特殊津贴, 被载入国际传记中心第九次编辑出版的世界文化名人录、中国专家大辞典、中华百年人物篇。