doi:10.3969/j.issn.1000-436x.2014.05.020

基于可编程路由技术的 MPLS 单标签分流传输算法

韩来权[1,2]，汪晋宽[1]，王兴伟[1,2]

(1. 东北大学 秦皇岛分校 计算机学院，河北 秦皇岛 066004；2. 东北大学 信息工程学院，辽宁 沈阳 110004)

摘 要：为提高数据传输的 QoS(服务质量)，运用 MPLS(多协议标签交换)、并发多路径和可编程路由技术，提出了 multipathMPLS 算法，实现了单个转发等价类标签进行多个标签交换路径并行分流的传输算法。NS2 仿真实验证明，该算法具备 MPLS 高速转发、并发多路径较高吞吐量、可编程路由器灵活部署等优点。

关键词：多协议标签交换；可编程路由技术；软件定义网络；标签交换路径

中图分类号：TP393.03 **文献标识码**：A **文章编号**：1000-436X(2014)05-0155-05

Flow-splitting algorithm of MPLS single label based on programmable router

HAN Lai-quan[1,2], WANG Jin-kuan[1], WANG Xing-wei[1,2]

(1.School of Electronic & Information, Northeastern University at Qinhuangdao,Qinhuangdao 066004,China;
2.School of Information Science & Engineering, Northeastern University, Shenyang 110004, China)

Abstract: To increase the QoS (quality of service) of data transmission, a novel flow-splitting algorithm, multipathMPLS, was proposed, which combined three technologies of MPLS (multi protocol label switching), concurrent multi-path and programmable router. MultipathMPLS implementes the flow-splitting transmission for the same FEC (forwarding equivalence class). Via the network simulation tool (NS2), this algorithm can obtain the high forwarding performance of MPLS, high throughput of concurrent multi-path and flexible configuration of programmable router.

Key words: MPLS; programmable router; software defined network; LSP

1 引言

多协议标签交换(MPLS)通过将分层网络的第二层(数据链路层)的交换和第三层(网络层)的路由技术很好地结合起来，以十分简洁的方式完成信息的传递，能够以无连接或显式路由的方式提供面向连接的业务，这使得它适用于动态隧道技术，能够保障数据传输业务的 QoS 需求[1]。虽然 MPLS 流量工程可以解决流量在网络中均匀分布的问题，但对同一 FEC(转发等价类)中的数据分组，其转发路径仍然是单一的。传统 IP 网络中的开放最短路径优先协议(OSPF)是典型的单约束算法[2]，只以跳数作为选路的条件，不考虑其他网络状况。而随着通信网络技术的快速发展，对于并发多路径的研究，在国内外逐渐成为热点。

理论上看，MPLS 和并发多路径转发技术各自具有优缺点：前者可以获得更短的数据转发时间，但是同一 FEC 中的数据分组转发路径相对单一，当某个 FEC 对应的 LSP 上数据流量过大时，路由器不能将该 FEC 数据分流成多条路径，可能导致网络拥塞；后者提供了灵活的转发路径，但未从根本上解决提高转发速率的问题，而且不同路径到达目的

收稿日期：2013-09-08；**修回日期**：2014-03-10

基金项目：国家杰出青年科学基金资助项目(61225012, 71325002)；高等学校博士学科点专项科研基金资助项目(20120042130003)；中央高校基本科研业务费专项基金资助项目(N110204003, N120104001)

Foundation Items: The National Science Foundation for Distinguished Young Scholars of China (61225012, 71325002); The Specialized Research Found for the Doctoral Program of Higher Education (20120042130003); The Special Found from the Central Collegiate Basic Scientific Research Bursary (N110204003, N120104001)

节点的数据分组会有不同的延迟[3]，可能使接收端得到一些乱序的分组，从而使得路由器造成错误的判断[4](如路由器会认为网络中产生了拥塞等)，带来新的问题。

2 相关工作

目前，已有多种多路径负载均衡方法：Zhao 等[5]提出了基于流的多路径，在 MPLS 域的入口路由器和出口路由器之间建立多条平行 LSP 用于均衡负载。由 Elwalid 等[6]提出的 MATE (MPLS adaptive traffic engineering)，引入了 MPLS 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将流量从高拥塞的链路转移到无拥塞的链路，仿真表明在整个网络负荷很小的时候能够很好地工作，随着网络负荷的增加，收敛性变得很差。

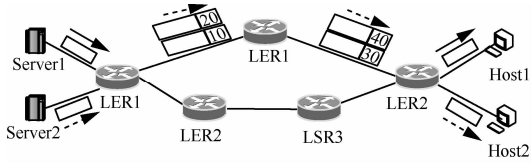
随着第一届可编程路由国际会议(PRESTO'08)在美国的召开，可编程路由技术出现飞速的发展[7]。根据网络的可编程特性，斯坦福大学的 Nick McKeown 教授和他的团队进一步提出了软件定义网络（SDN，software defined network）的概念。随着 CISCO 和 Juniper 相继公开自己的 IOS 和 JUNOS，研究人员可对路由模块进行更加可行的操作。随着国内外研究者们和路由器厂商的重视，可编程路由技术成为新的研究热点。

在经过大量分析比较之后，本文将 MPLS 和并发多路径有机结合，运用可编程路由器技术，在 MPLS 域中实现对 FEC 的多路径转发，提出了基于可编程路由器的 MPLS 的多路径分流算法：multipathMPLS，并通过 NS2 仿真[8]得到数据，再利用 Gawk、Gnuplot 等工具进行分析，结果表明此算法提高了 QoS。

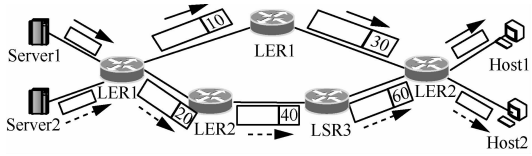
3 FEC 多路径分流算法原理

经过前期的研究，作者[3,9,10]提出了多种多路径数据传输算法。然而，先前的研究没有考虑到 MPLS 网络情况，本文研究内容是将这些算法与 MPLS 高效结合，以 MPLS 标签及交换路径作为优化依据，提出一种标签内分流的数据传输算法。

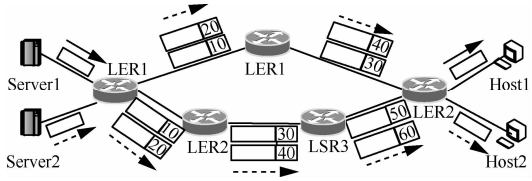
基本的 MPLS 转发原理如图 1 所示。图 1(a)所示，可将源端为 Server1、目的端为 Host1 的所有数据分组划分为同一 FEC 中，假设该 FEC 被 LER1 打上标签 10。对于每一组 FEC，都有与之唯一对应的 LSP(如：LER1→ LSR1→LER2)，由图 1(a)可知，普通的 MPLS 并不能得到很好的带宽利用率。引入流量工程的概念以后，MPLS 的转发原理如图 1(b)所示，网络管理员采用自定义某 FEC 的 LSP 就可以更好地管理网络的资源，FEC1(Server1→Host1)和 FEC2(Server2→ Host2)分别按照各自不同的 LSP 进行数据传输。



(a) 基本 MPLS 转发



(b) 引入流量工程的 MPLS 转发



(c) multipathMPLS 转发

图 1 3 种 MPLS 转发对比

本文考虑属于同一 FEC 的数据分组不止映射到唯一一条 LSP 上，而是实现“单 FEC 多 LSP”的映射。也就是说，FEC1 中的数据可分别映射到 2 条 LSP 上：LSP1(LER1→LSR1→LER2)和 LSP2 (LER1→LSR2→LSR3→LER2)，这些属于同一 FEC 中数据分组可分别通过这 2 条不同的 LSP 转发到 MPLS 区域的出口路由器 LER2，LER2 将标签去掉以后，再将这些来自不同路径但属于同一 FEC 中的数据分组整合转发到相同目的端 Host1，同理，FEC2 也有 2 条不同的 LSP 与之对应，如图 1(c)所示。

multipathMPLS 算法通过修改可编程路由器中的分类器模块和报文转发器模块来实现。算法流程描述如下，该算法由链路探测并建立标签交换表模块和报文转发器模块来实现。算法流程描述如下，该算法由探测更新模块(算法 1)和多 LSP 传输模块(算法 2) 2 部分组成。

算法 1 Probe & update Module

1) For each *LER* neighbor

2)　　Send probe packet *rtprotoDV*
3)　　If *FEC*∈available path Then
4)　　*SET A*←*LER* egress
5)　　End if
6)　End for
7)　For each element in *SET A*
8)　　*SET B*← get OSPF path
9)　　*SET B*← get alternative/ECMP path
10)　　Update *PHB* table
11)　End for

对应每一个 LER 的邻居，都要发出一个探测包 rtprotoDV(第 1)、2)行)，然后记录下某 FEC 所有可用路径，存入集合 *A*(第 3)、4)行)。对于集合 *A* 中的每一个元素(路径)，找出其中的最短路径，以及与该最短路径等值的路径(ECMP)，存入集合 *B*(第 7)行~9)行)。最后将 *B* 中的所有元素都作为待选 LSP，并更新 MPLS 域中各路由器的标签交换表。

算法 2　Multi-LSP transmission Module

1)　If packet arrived Then
2)　If labeling = *ture* Then
3)　*FEC* ← Get Address info from *SET B*
4)　*Next-hop* ← Lookup (*FEC*)
5)　Else
6)　If　labeling = Φ　Then
7)　*DA* ← Delete(*FEC*)
8)　*Next-hop* ← Lookup (*DA*)
9)　End If
10)　End If
11)　End If

对所有到达的数据分组，如果需要加上标签，则按照集合 *B* 中的 FEC 进行打包(第 1)行~4)行)；如果需要去掉标签，则按照传统的 OSPF 进行路由表查找，转发至下一跳(第 6 行~9)行)。

算法 1 和算法 2 易于在可编程路由器中灵活进行部署。

4　仿真实验

本文选用了 4 种具有代表性的转发技术进行仿真：OSPF(开放最短路径优先)、ECMP(等值多路径)、MPLS(多协议标签交换)、multipathMPLS。网络拓扑及各链路的代价(cost)如图 2 所示。其中，节点 0 发送 UDP 数据分组到节点 8，节点 1 发送 TCP 数据分组到节点 9。由图 3 可知，该 MPLS 域中，存在 5 条等值(cost 为 3)多路径：2→3→7、2→5→7、2→4→3→7、2→5→3→7、2→5→6→7。

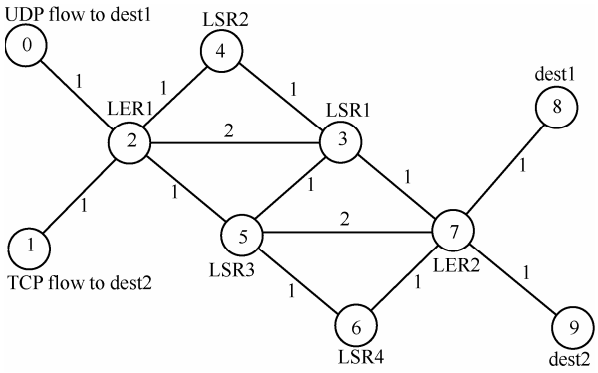


图 2　网络仿真拓扑

传统的 IP 网络采用最短路径优先算法，从节点 0 和节点 1 所发出来的数据分组都分别经过“2→5→7”后，到达各自的目的节点，当这条链路中途经的数据流逐渐增大时，这就可能导致这段最短路径过载。然而，当使用本文提出的 multipathMPLS 算法时，节点 0 发送的 UDP 数据分组和节点 1 发出的 TCP 数据分组均可分别经过 5 条不同的 LSP 到达各自的目的节点，实现了同一 FEC 的多路径分流。

5　理论分析

根据无向连通图 $G(V,E)$ 原理，其中 V 是节点的集合，E 是边的集合；$|V|=n$，$|E|=m$，设 $\forall v_i$，$v_j \in V$ $(i,j=1,2,\cdots,n)$ 之间存在 $Num(v_i,v_j)$ $\left(0 \leqslant Num(v_i,v_j) \leqslant m\right)$ 条边，对 v_i 与 v_j 间的任意一条边（设为第 k 条边）$e_{ij}^k \in E$ $\left(k=1,\cdots,Num(v_i,v_j)\right)$，可用带宽 bw 的取值区间为 $\left[bw_{ij}^{k-L}, bw_{ij}^{k-H}\right]$，服从均值为 $\mu_{Bw}\left(e_{ij}^k\right)=\dfrac{bw_{ij}^{k-L}+bw_{ij}^{k-H}}{2}$，方差为 $\sigma_{Bw}^2\left(e_{ij}^k\right)$ 的高斯分布，其中 bw_{ij}^{k-L} 和 bw_{ij}^{k-H} 分别表示 e_{ij}^k 上可用带宽的下限和上限。概率密度函数为

$$f_{Bw}(bw)=\frac{1}{\sqrt{2\pi\sigma_{Bw}^2\left(e_{ij}^k\right)}}e^{-\frac{\left(bw-\mu_{Bw}\left(e_{ij}^k\right)\right)^2}{2\sigma_{Bw}^2\left(e_{ij}^k\right)}} \tag{1}$$

对于具有带宽约束区间 $\Delta_{Bw}=[BW^L, BW^H]$ 的链路来说，用户在“在该链路上得到的带宽”的满意度函数为

$$S_{Bw}(bw)=\begin{cases}0, & bw<BW^L\\ e^{-\left(\frac{BW^H-bw}{bw-BW^L}\right)^2}+f_1, & BW^L\leqslant bw<BW^H\\ 1, & bw\geqslant BW^H\end{cases}\quad(2)$$

其中，修正函数为

$$f_1=\begin{cases}\varepsilon_1, & bw=BW^L\\ 0, & bw\neq BW^L\end{cases}\quad(3)$$

式（2）表示随着 bw 值的增大，用户对“在该链路上得到的带宽”越来越满意，当 bw 达到 BW^H，用户达到最大满意度 1。ε_1 是一个远小于 1 的正纯小数。

6　性能分析与比较

6.1　吞吐量及延时比较

由图 2 可知，节点 2 是整个网络中瓶颈节点之一，节点 2 的吞吐量性能直接关系到整个网络的健康状况。采用不同转发技术，节点 2 的吞吐量比较如图 3 所示，延时对比如图 4 所示。

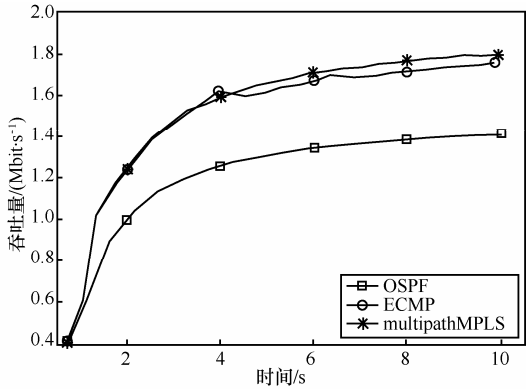


图 3　节点 2 吞吐量比较

由图 3 可知，multipathMPLS 和 ECMP 获得了几乎一样的吞吐量，且明显优于 OSPF。说明 multipathMPLS 算法继承了并发多路径的一大优点：高吞吐量。而传统的 OSPF 算法对于转发路径的选择不如多路径灵活，没有并发分流的功能，导致节点吞吐量不能达到最优值。

由图 4 可知，OSPF 平均延时相对较大; MPLS 延迟性能居中，因为普通 MPLS 没有引入分流功能。虽然路由器的转发速度快，但是链路状况不好。这就出现了链路 2→5→7 上的数据过载，整体的延迟性能提高较少。multipathMPLS 虽然抖动较大，但是平均延时很小;随着时间的推移，multipathMPLS 的平均延时成了三者中最小的，说明 multipathMPLS 具有 MPLS 高速转发的优点。

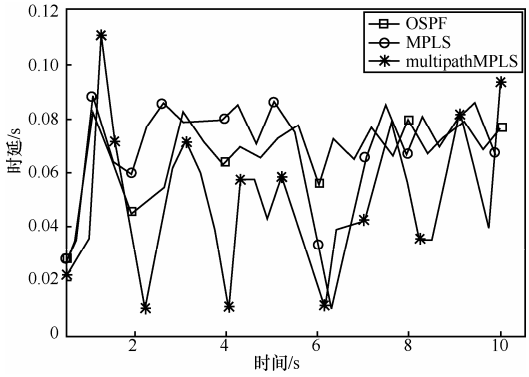


图 4　延时对比

6.2　分组丢失率比较

分组丢失率是衡量网络性能好坏的重要参数，本文对 4 种转发方式得到的仿真数据进行分析后，分组丢失率对比如表 1 所示。

表 1　　分组丢失率对比

算法	发分组总数	分组丢失数	分组丢失率
OSPF	1 842	74	4.02%
ECMP	2 172	18	0.83%
MPLS	1 862	94	5.05%
multipathMPLS	2 220	8	0.36%

可见，多路径转发时，单位时间内发分组数(吞吐量)要优于单路径(包括 OSPF、MPLS)，多路径分组丢失率要比单路径低。从以上数据中可以分析出，普通的 MPLS 由于提高了路由器的转发速度，导致数据分组很快都集中到某条单一的链路上，而该链路产生较大拥塞，数据不能很快到达下一跳，导致普通 MPLS 的分组丢失率成为了四者中最大的。multipathMPLS 算法由于采用了标签交换和并发多路径分流技术，既实现了高速转发，又降低了链路拥塞的概率，实现了四者中最低分组丢失率。

6.3　稳定性的比较

在一个 MPLS 网络中，路由器必须依赖于路由协议来准确地传播可达性信息和完成标签转发相关的工作。因此 MPLS 网络对路由协议的依赖性要高于 IP 网络。如果 MPLS 使用标签交换机制在路由和转发之间引入新一层的间接性，会导致 MPLS 网络的故障更加难以处理和排除。

为了比较稳定性，本文在网络仿真中做了如下设置：1) 引入了一次链路中断，使节点 2 和节点 5

之间的链路在 4 s 时刻突然断开，经过 1 s 后，在 5 s 时刻再自动恢复；2) 现实中一条链路上一般不可能只存在一条数据流，故文中在此处引入一个 Pareto 流量作为干扰流；3) 当目的节点 8 和节点 9 接收到的 10 Mbit 的数据(不包括干扰流)时，仿真结束。

有干扰情况下延迟性能比较如图 5 所示。图中 OSPF、multipathMPLS 分别表示无干扰情况下的各自曲线，OSPF-int、multipathMPLS-int 分别表示链路出现干扰以后各自的曲线，稳定性对比如表 2 所示。

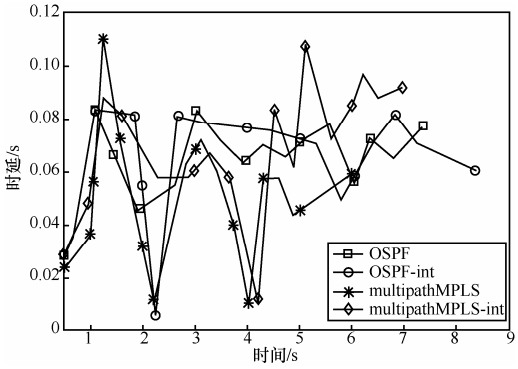


图 5　稳定性对比

表 2　　稳定性对比

算法	用时/s	平均延时/s	延时抖动
OSPF	7.341	0.060 5	较小
OSPF-int	8.375	0.062 0	小
multipathMPLS	6.008	0.052 8	大
multipathMPLS-int	6.949	0.058 6	大

在链路出现断路后，由于动态路由的结果，导致 OSPF 的转发路径由原来的 2→5→7 变成 2→3→7。从表 2 也可看出，不论链路中有无干扰存在，multipathMPLS 算法的平均延迟都很小。

表 2 还可以分析出，无故障时 OSPF 整个传输比 MPLS 多路径算法多用 1.333 s(7.341~6.008)的时间，这个差值说明了 MPLS 的转发速度快于 OSPF 算法。有故障时 OSPF 比 MPLS 多路径算法多用 1.426 s(8.375~6.949)的时间，这个差值说明了在遇到故障时，MPLS 多路径的反应时间好于 OSPF 算法。再对 2 个差值进行比较，它们之间相差 0.093 s (1.426~1.333)。这 93 ms 的差值，也反映了多路径算法具有更快的路由收敛时间。

7　结束语

本文提出了一种基于可编程路由器的 MPLS 的多路径分流传输算法：multipathMPLS，实现了“单 FEC 多 LSP”的分流转发算法，即单 MPLS 标签多个 LSP 路径并行分流转发，并通过 NS2 及相关工具，将该算法与 OSPF、ECMP、普通的 MPLS 分别进行了仿真，并作了详细对比分析，最终得出如下结论：multipathMPLS 算法，继承了多路径转发和 MPLS 两者各自的优点，实现了高吞吐量，低延时的转发效果，且易于在下一代网络的可编程路由器中部署使用。

参考文献：

[1] LEE Y, SEOK Y, CHOI Y, *et al*. A constrained multipath traffic engineering scheme for MPLS networks[A]. IEEE International Conference on Communications: ICC'02[C]. 2002.2431-2436.

[2] MOY J. OSPF Version 2. Internet RFC 2328[S]. 1998.

[3] 韩来权, 汪晋宽, 王翠荣. 基于流的跨层并发多路径转发算法[J]. 东北大学学报(自然科学版), 2009, 30(3):363-366.
HAN L Q, WANG J K, WANG C R. Flow-based cross-layer forward algorithm for concurrent multipath[J]. Journal of Northeastern University, 2009, 30 (3):363-366.

[4] YABANDEH M, ZARIFZADEH S, YAZDANI N. Improving performance of transport protocols in multipath transferring schemes [J]. Computer Communications, 2010, 30(17):3270-3284.

[5] ZHAO Z H, SHU Y T, ZHANG L F, *et al*. Flow-level multipath load blancing in MPLS network[A]. Proceedings of IEICE Transactions on Communications[C]. 2010. 2015-2022.

[6] ELWAILD A, JIN C, LOW S. MATE: MPLS adaptive traffic engineering [A]. Proceedings of INFOCOM'01[C]. 2001.89-93.

[7] ALBERT G, PARANTAP L, DAVID A, *et al*. Towards a next generation data center architecture: scalability and commoditization[A]. Proceedings of PRESTO 2008 [C]. 2008.57-62.

[8] NS2 network simulator[EB/OL]. http://www.isi.edu/ nsnam/ns/.

[9] HAN L Q, WANG J K, WANG X W, *et al.* Bypass flow-splitting forwarding in FISH networks[J]. IEEE Transactions on Industrial Electronics, 2011, 58(6):2197-2204.

[10] HAN L Q, WANG J K, WANG X W. A function migration algorithm based on programmable router of multipath networks[J]. Chinese Journal of Electronics, 2011, 20(1):170-174.

作者简介：

韩来权（1977-），男，黑龙江双鸭山人，博士，东北大学副教授，主要研究方向为未来互联网、云计算。

汪晋宽（1957-），男，辽宁沈阳人，东北大学教授、博士生导师，东北大学副校长，东北大学秦皇岛分校校长，主要研究方向为阵列天线、无线传感器网络。

王兴伟（1968-），男，辽宁盖州人，东北大学教授、博士生导师,主要研究方向为未来互联网、云计算、网络安全和信息安全。