

基于标签堆栈的一种提高 MPLS-VPN 性能的机制^{*}

孙琼 耿彦辉 俞能海

(中国科学技术大学信息处理中心 合肥 230027)

摘要 近几年,Internet 已经经历了指数级的增长,各种各样的应用对当前的 IP 网络提出了挑战。基于 MPLS 的 VPN 对于未来的 Internet 服务是个很好的解决方案,并且 MPLS 技术已经被一些大的 Internet 服务供应商采用,来提供 VPN 服务以及在他们 VPN 传输网络的顶层提供一些增值应用。本文提出了一个基于使用标签堆栈的机制来有效降低 MPLS-VPN 网络所需的标签数量。通过这种方法可以减小转发表的大小,提高转发速度,从而使 MPLS-VPN 的性能得到有效的提高。仿真结果也证实了所提方法的优点和有效性。

关键词 多协议标签交换,虚拟专用网,标签堆栈,标签交换路由器

A Mechanism to Improve the Performance of MPLS-VPN Based on Label Stack

SUN Qiong GENG Yan-Hui YU Neng-Hai

(Information Processing Center, USTC, Hefei 23027)

Abstract Internet has experienced an exponential growth in recent years, and various kinds of application are raising critical challenges to current IP network. The MPLS-based VPN technology is now becoming a good solution for the delivery of future Internet services, and Multiprotocol Label Switching (MPLS) technology is being adopted by some largest Internet service providers to offer VPN and some value-added applications on top of their VPN transport networks. In this paper, a novel mechanism based on using label stack properly is proposed to reduce the number of labels needed efficiently. As the result, the size of forwarding table can be reduced and then the forwarding speed can be improved efficiently. In this way, the performance of MPLS-based VPN can be improved. And our simulation results validated our approach, demonstrating that MPLS-VPN established by our approach has a better performance.

Keywords MPLS, VPN, Label stack, LSR

1 引言

目前,很多大公司的网络由分散在各地的子网络组成。通常,这些网络之间的连接是由服务供应商提供的专用传输线路提供的,从而保证了可靠性和安全性,但租用专用线路的费用非常昂贵。VPN (Virtual Private Network; 虚拟专用网) 技术是此类问题一个很好的解决方案^[1]。VPN 技术通过公共基础设施^[2]为企业分散的网络提供了安全的点到点连接,就好象在使用一个专用线路。这个方法大大降低了租用专用线路的花费。

MPLS (Multiprotocol Label Switching; 多协议标签交换)^[3]技术是由 IETF^[4]在 1997 年提出的新一代高速骨干网络交换标准,它被推广的目的是提高转发性能。MPLS 的基本思想就是根据一个短的定长标签来转发数据包。当数据包进入 MPLS 域的入口节点,一个标签就被插入 IP 包头,然后数据包就通过标签交换沿着 LSP (Label Switching Path; 标签交换路径) 转发,这样就不用像传统 IP 路由器在每一跳寻找最大地址匹配。MPLS 技术最显著的优点是不需要在每一跳进行包头信息的分析,而是仅仅需要在数据包进入 MPLS 网络时分析包头并分配标签。入口路由器可以使用一些除目的 IP 地址之外的信息(比如数据包的来源和数据包的类型),来给数据流指定不同的路径,以满足不同的 QoS (Quality of Service; 服务质量) 要求。比如,多媒体应用数据要求低延时

和低延时抖动,那么就可以把这些特别的需求信息编码到包头中,由入口路由器进行分析,并进行满足要求的传输。此外,通过使用 MPLS 所支持的显示路由,相对于传统的 IP 网络更容易实现流量工程。所有这些优点使 MPLS 成为先进并具有吸引力的技术被应用在骨干网中。高性能的 MPLS 骨干网也使得基于 MPLS 的 VPN 具有更优异的性能。

为了达到 MPLS-VPN 网络的最优性能,本文提出了一个全新的基于使用标签堆栈的方法,以减小需要的标签的数量。通过这种方法,可以有效地提高 MPLS-VPN 网络的整体性能。本文第 2 部分,指出并分析传统 MPLS-VPN 网络存在的问题;第 3 部分,提出基于使用标签堆栈来减少标签数量及提高转发速度的方法;第 4 部分,给出所提机制的仿真结果及性能分析。最后是结论。

2 问题分析

图 1 给出了 MPLS-VPN 的网络结构。MPLS-VPN 网络由 CE (Customer Edge router; 客户边缘路由器)、PE (Provider Edge router; 供应商边缘路由器) 和 LSR (Label Switching Router; 标签交换路由器) 组成,CE 可以使用 IP 协议传输包,服务在网络边缘执行。实际上,VPN 只是存在于服务供应商网络的边缘。核心路由器不参与实际的 VPN 工作,只是负责在 LSP (Label Switching Path; 标签交换路径) 上转发包。客户路由器也不参与 VPN 工作,也只是通过客户建立的寻

^{*} 本文获中国科学院知识创新工程(项目编号:cxj1-91)资助和国家 863 计划项目(项目编号:2002AA783043)资助。

址和路由^[6]机制来为 IP 数据包寻路。

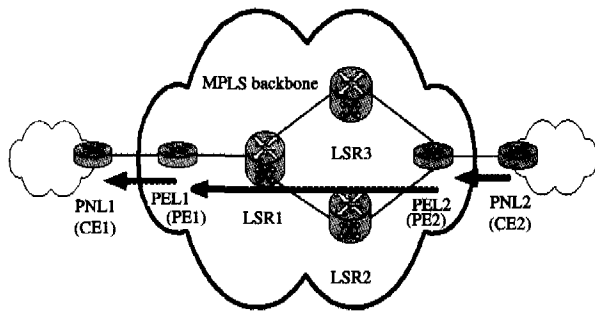


图1 MPLS-VPN 的网络结构图

这里不再讨论 MPLS-VPN 网络的优点,而是将重点放在 MPLS-VPN 网络存在的问题。目前最大的问题在于提供基于 MPLS 骨干网的 VPN 服务对大量的客户变得越来越复杂和困难,因为一些大的网络服务供应商不得不建立和管理用于连接 VPN 端点的数以千计的 MPLS LSP,而且这种基于覆盖模型的 VPN 现在被大量地采用^[6]。这就产生了问题,因为每个标签堆栈条目(32 比特)中只有 20 比特被用来为标签编码,于是标签的数量就被限制在 2^{20} 以内。所以考虑到网络的扩展性,必须节约使用标签。同时,随着使用标签数量的下降,用来完成转发数据包任务的 LSR 所维护的转发表尺寸也可以变得很小,结果最终 LSR 可以更快地转发数据包。

本文下面部分将提出基于使用标签堆栈的方法,它可以降低所需标签的数量并且达到提高 MPLS-VPN 网络性能的最终目标。

3 基于使用标签堆栈的机制

由第 2 部分的分析可知,为了使 MPLS-VPN 能有更好的性能,小标签数量非常重要。然而,通过下面的例子可以看出,在同一时刻,小标签数量和小标签堆栈深度是不能兼得的。小标签数量是以大标签堆栈深度的代价得到的,但太大的标签堆栈深度同样不可取,因为这将占用 IP 包头中更多的空间。所以必须在标签数量和标签堆栈深度之间寻求一个平衡^[7],以达到提高 MPLS-VPN 性能的目标。

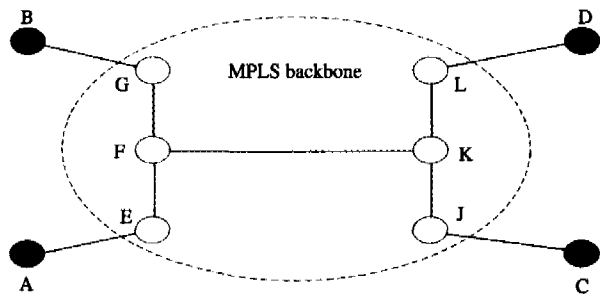


图2 网络拓扑结构

现在通过一个具体的例子来说明标签尺寸和标签堆栈深度之间的关系。考虑一个如图 2 所示的网络结构,图中 A, B, C, D 是四个有相同的输入带宽和输出带宽的 VPN 端点。这里采用 Hose Model(点到多点模型)^[8]提供的 VPN 模型。在这种模型中,输入和输出带宽必须由每个 VPN 端点来指定,它们是允许的可流入 VPN 端点及输出到 VPN 端点的最大流量的总和。

每个 VPN 端点来指定,它们是允许的可流入 VPN 端点

及输出到 VPN 端点的最大流量的总和。

首先考虑标签堆栈深度为 1 时的场景。共需要 4 个标签,因为需要给每个 VPN 端点分配 1 个标签,以区分它们(A, B, C, D 分别对应标签 1, 2, 3, 4)。这样,当网络中的每个节点收到输入的数据包时,根据数据包的标签就可以判断出目的 VPN 端点,根据转发规则把数据包发送到相应输出接口即可,并不需要对标签进行入栈出栈操作。表 1 给出了收到数据包时,节点 F 和节点 K 的输出接口及执行的操作。

表 1 标签堆栈深度为 1 时的转发表

节点	输入接口	包的标签	输出接口	节点动作
F	K-F, G-F	1	F-E	None
	E-F, K-F	2	F-G	None
	E-F, G-F	3, 4	F-K	None
K	F-K, L-K	3	K-J	None
	F-K, J-K	4	K-L	None
	J-K, L-K	1, 2	K-F	None

从表 1 中可以清楚看出,当一个 VPN 端点要发送数据包到其它 VPN 端点时,只需要给数据包加上一个和目的 VPN 端点一一绑定的标签。例如,当节点 F 在 E-F 或 G-F 接口收到一个带有标签 3 的数据包,它把包转发到 F-K 接口,不执行任何堆栈操作。然后,当节点 K 收到节点 F 发送过来并带有标签 3 的数据包,把数据包发送到 K-J 接口,最终由节点 J 把数据包发送到 VPN 端点 C。这意味着在这个场景中,也即标签堆栈深度为 1 的情况下,只需要 4 个标签(一个标签对应一个 VPN 端点)。

表 2 标签堆栈深度为 2 时的转发表

节点	输入接口	包的标签	输出接口	节点动作
E	A-E	*	E-F	None
	F-E	*	E-A	None
F	E-F	1	F-G	None
	G-F	1	F-E	None
	E-F, G-F	2	F-K	Pop
	K-F	1	F-E	None
	K-F	2	F-G	None
G	F-G	*	G-B	None
	B-G	*	G-F	None
J	J-K	*	C-J	None
	C-J	*	J-K	None
K	J-K	1	K-L	None
	L-K	1	K-J	None
	J-K, L-K	2	K-F	Pop
	F-K	1	K-J	None
L	F-K	2	K-L	None
	K-L	*	L-D	None
	D-L	*	L-K	None

下面考虑标签堆栈深度为 2 的场景。在这种情况下,为使每个 VPN 端点都能和其它 VPN 端点通信,只需要 2 个标签(1 和 2)。表 2 为节点 E, F, G, J, K, L 的转发表(* 代表 1 或 2)。例如,如果 VPN 端点 A 想和端点 B 通信,它只需要为数据包加入标签 1。更复杂的情况,如果端点 A 想和端点 C 通信,端点 A 需要把标签 1 和 2 先后压入数据包的标签堆栈。这样,根据转发表的规则,当节点 E 从 E-F 接口收到带有标签(1, 2)的数据包时,它将把数据包发送到 E-F 接口,而

不进行对标签堆栈的操作;然后节点 F 会把数据包发送到 F-K 接口并同时把标签 2 从标签堆栈的顶部弹出。这时,标签堆栈中只有一个标签(1)并位于标签堆栈顶部,然后节点 K 把来自节点 F 的数据包发送到 K-J 接口,最终由节点 J 把数据包转发到 VPN 端点 C。

在这个具体的例子中,当标签堆栈深度由 1 增加为 2 时,需要的标签也同时由 4 减少为 2 了。

由上面的讨论可以看出,基于标签堆栈的方法能有效地减少网络需要的标签数量,同时用来给标签编码需要的空间(比特位)也相应地减少了。关于标签尺寸和标签堆栈深度之间的量化关系,基于信息论,可以给出如下结论:

1. 如果标签堆栈深度为 1,那么在一个有 n 个端点的网络中,至少需要 n 个标签,否则甚至不能区分网络中的目标端点。

2. 如果标签堆栈深度为 s ,那么至少需要 $L \geq n^{1/s}$ 个标签,其中 n 是网络中目的端点的个数。

3. 由第 2 条可以推出,如果标签数量 L 是给定的,那么标签堆栈深度至少应为 $s \geq \log n / \log L$ 。

可以看出,标签数量和标签堆栈深度之间是有一定限制的。但如果在限制条件内存在多个标签数量和标签堆栈深度的组合时,哪一种才是最优的?因为在不同的骨干网络模型、不同的 VPN 模型下,结果是不同的。所以这里借助于网络仿真,在不同的网络规模以及在不同的标签数量和标签堆栈组合下,只对树状模型的 MPLS-VPN 的性能进行估计。

4 仿真实验

我们使用网络仿真工具 NS2(Network Simulation2)进行了不同网络规模下的仿真实验,仿真结果给出了在不同标签数量和标签堆栈深度组合下 MPLS-VPN 的性能。

4.1 小尺度 MPLS-VPN 网络下的仿真

图 3 显示了一个小规模 MPLS-VPN 网络模型。图中 8 个黑色的节点是 VPN 端点、4 个白色节点组成 MPLS 核心网络。如果采用深度为 1 的标签堆栈,所需的标签数量为 8,并需要 3 比特的空间来为这 8 个标签编码($2^3=8$)。当标签堆栈的深度增加到 2,就只需要 3 个标签及 2 比特的编码空间。但是当监测节点 A 的吞吐量,可以注意到标签深度的增加没有带来任何区别,因为 1 比特的差别实在是太小了。

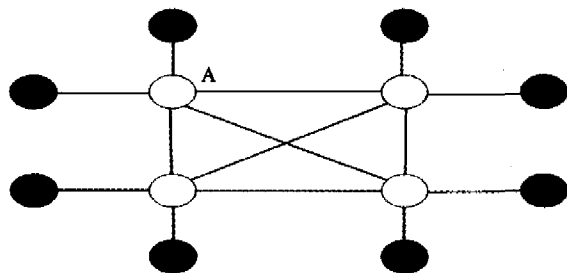


图 3 小规模 MPLS-VPN 网络的拓扑结构图

4.2 中尺度 MPLS-VPN 网络下的仿真

在这种场景下,建立了一个有 1000 个端点的 MPLS-VPN 网络。可以发现,当把标签堆栈深度从 1 增加到 3 时,所需标签的数量从 1000(为 1000 个标签编码,需要 10 比特空间)下降到 70(为 70 个标签编码,需要 7 比特空间)。这意味着相对于标签堆栈深度为 1 的情况,节约了超过 90% 的标

签。虽然节省了大量的标签,但节省的 3 比特空间并没有使 LSR 的性能得到显著的提升。另外,LSR 可以很容易保持 1000 个标签,所以只有当几百个 VPN 簇存在于同一个网络里时,这种节约才会变得非常有意义。

4.3 大尺度 MPLS-VPN 网络下的仿真

本实验场景创立了 100 个 VPN 簇(每个簇有 1000 个 VPN 端点),并且使它们在一个 MPLS 骨干网中互连。在拥有 100000 个端点的大型网络中,本文所提机制的优点和有效性就变得很明显了。当标签堆栈深度为 1 的时候,LSR 需要维持 100000 个标签(20 比特的编码空间);而当标签堆栈深度增加到 3 的时候,标签的数量只要 7000 了(13 比特的编码空间)。可以看出,标签数量和编码比特位空间的节约都是显著的。因此,LSR 可以更快和更有效地转发数据包。仿真实验中一个被监测的 LSR 的吞吐量和延迟如图 4 和图 5 所示。

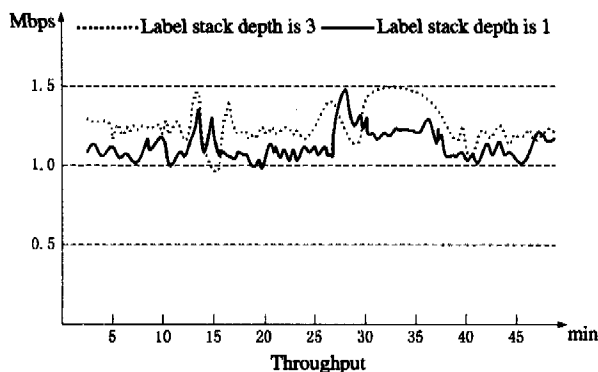


图 4 被监测 LSR 的吞吐量

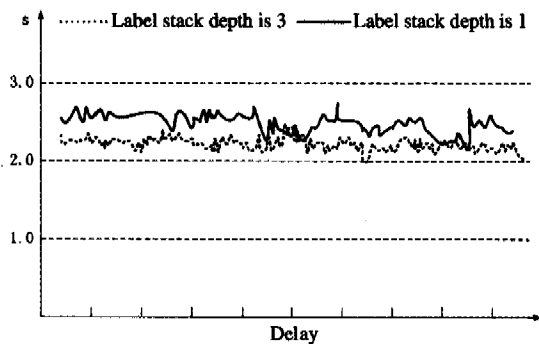


图 5 被监测 LSR 的延时

从图上可以看出,当标签堆栈深度由 1 增加到 3 时,被监测 LSR 的吞吐量增大而延时减小了。这意味着本文所提机制可以使 LSR 更快地转发数据包,MPLS-VPN 网络的整体性能也因此得到了提高。

结论 为了使 MPLS-VPN 网络达到更好的性能,本文提出了一种全新的基于使用标签堆栈的机制。用这种机制,网络需求的标签数量可以得到有效的降低,并且减小了每个 LSR 所维护的转发表的尺寸。而比较小的转发表可以使 LSR 更快和更有效地转发数据包,从而最终提高了 MPLS-VPN 网络的整体性能。除了理论上的研究和分析,我们还做了相应的仿真实验,实验结果也验证了所提机制的优点和有效性。

(下转第 51 页)

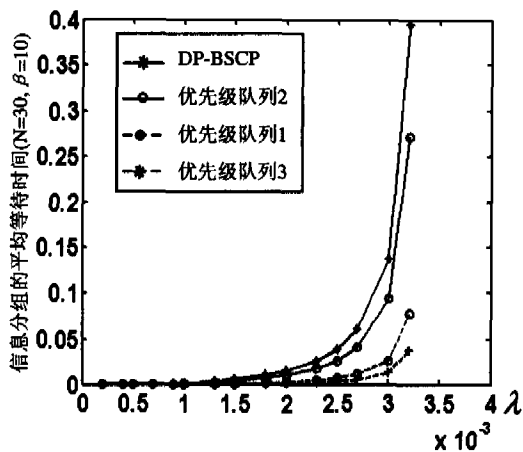


图6 $N=30$ 时 DP-BSCP 中各优先级队列及系统的 $E[L]$

表1 DP-BSCP 各优先级队列的平均等待时间(时隙)
($N=20, \beta=10$)

系统达率 λ	优先级 1 队列 平均等待时间	优先级 2 队列 平均等待时间	优先 3 级队列 平均等待时间	系统平均 等待时间
0.0005	0.486	0.498	0.876	0.620
0.001	1.280	1.275	1.176	1.244
0.0015	2.195	2.176	2.523	2.298
0.002	3.308	3.277	3.296	3.294
0.0025	5.130	4.926	4.953	5.003
0.003	7.508	7.387	7.237	7.377
0.0035	12.009	11.7776	11.506	11.764
0.004	20.444	20.128	19.642	20.071
0.0045	45.431	44.063	42.983	44.159

表2 DP-BSCP 各优先级队列平均等待时间(时隙)
($N=30, \beta=10$)

系统达率 λ	优先级 1 队列 平均等待时间	优先级 2 队列 平均等待时间	优先 3 级队列 平均等待时间	系统平均 等待时间
0.0005	0.876	0.810	1.079	0.922
0.001	2.189	2.196	2.233	2.206
0.0015	4.055	3.867	4.101	4.007
0.002	7.730	7.521	7.589	7.613
0.0025	15.387	15.123	14.919	15.143
0.003	44.742	43.873	43.187	43.934
0.0032	120.780	118.757	118.082	119.206

的要求压力比 WTRP 大大降低;同时,DP-BSCP 在到达率 λ

<0.0025 的区域内接近线性,而 WTRP 系统在 $\lambda > 0.0015$ 时,曲线就已经急剧上升,因而 DP-BSCP 的稳定性能较之 WTRP 也大幅度提高。此外,由于 DP-BSCP 充分考虑了分组本身的优先级和在缓冲区中的等待时延,采用动态优先级的策略进行服务,使不同类型的分组均能得到较好的服务,不会出现“饿死”现象,从而能够满足较高的 QoS 需求。今后将进一步研究 DP-BSCP 在 Ad hoc 网络及移动分布式操作系统通信中的应用,例如可靠性、系统内部通信、数据服务等。

参考文献

- 1 Bharghavan V. Performance evaluation of algorithms for wireless medium access. Proc IEEE Performance and Dependability Symposium '98, Raleigh, NC, 1998
- 2 Tang Z, Garcia-Luna-Aceves J J. Hop-reservation multiple access (HRMA) for ad-hoc networks. In: IEEE INFOCOM '99, New York, 1999. 194~201
- 3 Garces R, Garcia-Luna-Aceves J J. Collision avoidance and resolution multiple access for multichannel wireless networks[A]. In: IEEE INFOCOM'2000, Tel Aviv, Israel, 2000. 595~602
- 4 M J-NG, Lu I. Spread spectrum medium access protocol with collision avoidance in mobile ad-hoc wireless networks[A]. In: IEEE INFOCOM'99. NY, March 1999. 776~783
- 5 Li Jiandong, et al. User-dependent perfect scheduling multiple access(UPMA) for wireless internet access. In: Future Telecommunications Forum, Beijing; Beijing University of Posts and Telecommunications, 1999. 59~63
- 6 Li Jiandong, et al. User-dependent perfect scheduling multiple access(UPMA) for wireless ad hoc Internet access. In: Proc WCC 2000/ICCT 2000, Beijing; Publishing House of Electronics Industry, 2000. 1714~1717
- 7 张文柱,李建东,刘凯. 依据用户妥善安排的多址接入协议在分布式无线网络中的性能研究. 电子学报, 2002, 30(6): 1175~1179
- 8 刘凯,李建东,翁继伟. 移动分布式无线网络中具有 Qos 保证的 UPMA 协议. 电子与信息学报, 2003, 25(3): 370~376
- 9 Ergen M, Varaiya P, Walrand J. WTRP-wireless token ring protocol. URL; http://wow.eecs.berkeley.edu/wtrp, 2002, 12~108
- 10 Crow R P, et al. IEEE 802. 11 Wireless Local Area Networks. IEEE Commun. Mag, 1997, 35(9): 116~126
- 11 Sharon O, Altman E. An efficient polling MAC for wireless LAN. IEEE/ACM Trans Networking, 2001, 9(4): 439~451
- 12 赵东风,郑苏民. 周期查询式门限服务排队系统中信息分组的延迟分析. 通信学报, 1994, 15(2): 18~23
- 13 赵东风,李必海,郑苏民. 周期查询限定服务排队系统研究. 电子科学学刊, 1997, 19(1): 44~49

(上接第 44 页)

参考文献

- 1 Ferguson P, Huston G. What is VPN. The Internet Protocol Journal, 1999
- 2 Palmieri F. VPN scalability over high performance backbones. In: Proceedings ISCC 2003, 2003. 975~981
- 3 Rosen E C, Viswanathan A, Callon R. MultiProtocol Label Switching architecture. RFC 3031, 2001
- 4 Rosen E C, Tappan D, Rekhter Y, et al. MPLS label stack encoding. RFC 3032, 2001

- 5 Frederickson G N, Janardan R. Designing networks with compact routing tables. Algorithmica, 1998, 3: 171~190
- 6 Davie B, Rekhter Y. MPLS: Technology and Applications, Morgan Kaufmann Publishers, 2000
- 7 Peleg D, Upfal E. A trade-off between space and efficiency for routing tables. J Assoc Comput Mach, vol. 36, no. 3, 1989, 36(3): 510~530
- 8 Kumar A, Rastogi R, Silberschatz A, et al. Algorithms for provisioning virtual private networks in the hose model. Proceedings ACM SIGCOMM, 2001