基于 LER 支配集的 MPLS 网络拓扑聚合策略*)

杨宗凯　马娅婕　谭贤四　何建华

(华中科技大学电子与信息工程系　武汉 430074)

摘　要　MPLS 技术通过建立标签交换路径(LSP),并对具有不同转发等价类的数据流分配标签进行转发。聚合可以减少 LSP 的建立和维护开销,对于网络的扩展性具有重要的意义。本文提出了一种基于标签边界路由器(LER)支配集的拓扑聚合方案,采用分布式 LER 支配集构造法以减少建立 LSP 时出口节点的数量,从而在 MPLS 网络中形成一个聚合的虚拟骨干拓扑,可以对具有不同出口 LER 的 LSP 进行聚合,达到减少 LSP 的目的。同时支配 LER 的冗余性可以作为 LSP 的保护备份,因此使 MPLS 网络的具有较好的可扩展性和健壮性。

关键词　支配集,MPLS,聚合

A MPLS Network Aggregation Scheme Based on LER Dominating Set

YANG Zong-Kai　MA Ya-Jie　TAN Xian-Si　HE Jian-Hua

(Department of Electron and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074)

Abstract　MPLS builds the lable switching path and forwards the packet by assigning different labels to packets which have different Forwarding Equivalence Classes. Aggregation can reduce the overhead of establishing and maintaining the Label Switching Path (LSP). It is very important for the scalability of the MPLS networks. A network aggregation scheme based on Lable Edge Router (LER) dominating set is proposed. The distributed construction method of the LER dominating set is introduced. The aggregation scheme can reduce the amount of the egress nodes when establishing LSPs so as to form an aggregated virtual backbone topology. Then the amount of the LSPs in MPLS networks is reduced by aggregating the LSPs that have different egress LERs. At the same time, the redundancy of the dominating nodes can act as the backup of the working parh. It can efficiently improve the scalability and robusticity of the MPLS networks.

Keywords　Dominating set, Multi-Protocol Label Switching, Aggregation

1　引言

MPLS(Multi-Protocol Label Switching)技术是将分层网络的第二层交换和第三层路由结合起来的一种 L2/L3 集成数据传输技术[1]。在 MPLS 网络入口 LER(Label Edge Router)处将 IP 分组映射为特定的 FEC(Forwarding Equivalence Class),然后再将 FEC 用定长的标签编码表示。标签将插入到 IP 分组的首部,沿着标签交换路径 LSP(Label Switching Path)的后继点以输入分组首部的标签作为索引,查找表示下一跳的新标签,然后用新标签取代旧标签,将分组转发出去,直至 LSP 的出口。

LSP 是通过信号协议,如 Resource Reservation Protocol with Traffic Engineering (RSVP-TE)[2]或 Label Distribution Protocol(LDP)[3]来建立和维护的,MPLS 的网络节点可以根据需要选择其中任何一种信号协议。但是对于建立和维护 LSP 而言,网络节点在运行 RSVP-TE 或 LDP 时的开销对该网络的扩展性能产生很大的影响。由于 MPLS 是 IP 网络层的镜像,所以 MPLS 网络若对所有 n 个目的提供可达性,就需要为此创建 O(n)个交换路径(这里 n 的意义由交换路径的颗粒度决定)。因此建立点到点的径流的全网格状连接时,网络需要建立 $O(n^2)$条 LSP,当增加网络节点时,创建和维护新的 LSP 的开销将随着网络节点的数目的增加而急剧增长,这不利于 MPLS 的扩展。

提高 MPLS 网络的可扩展性一般可以采用这样两个基本的路由原理:(1)分组转发可以按照一棵反向树来进行,树的根在目的地路由器。例如文[5,6]在聚合算法下预先建立多点到点(mp2p)的 LSP,使从多个入口 LER 进入 MPLS 的分组通过预先建立的 mp2p 的 LSP 向同一个出口 LER 转发;文[7]提出了一种对于给定的 LSP 集合,运用启发式算法获得一个最小的 mp2p 树的方案。这些都是基于具有共同的出口 LER 的 LSP 聚合策略,或者算法复杂,或者需要各节点获得全网拓扑状态信息,因此对节点所维护的信息要求很高,扩展性不强。(2)通过路由聚合以减少目的地址的数目。例如文[8]为具有相同入口 LER 和出口 LER 的不同分组仅创建一条 LSP,即聚合了入口 LER 和出口 LER 所连接的局网内的不同目的地址间的路由。然而这种聚合方法只能对具有相同的入口和出口 LER 的交换路径进行聚合,因此适应性不强,不能达到较高的聚合率。

为了提高路由降合的效率和适应性,本文提出一种基于 LER 支配集的 MPLS 拓扑聚合策略,通过减少出口节点的数目以减少 MPLS 网络建立 LSP 的数量。该拓扑聚合策略采用分布式 LER 支配集构造法,每个 LER 独立确定是否为支配

*)国家自然科学基金项目(60202005)。**杨宗凯**　教授,博士生导师,主要研究方向为下一代互联网技术与网络安全。**马娅婕**　博士研究生,主要研究方向为 MPLS 流量工程与组播技术。

集中的节点，仅以支配集中的节点作为建立 LSP 时的出口节点，在 MPLS 网络中形成一个聚合的虚拟骨干拓扑，从而降低 MPLS 网络建立和维护 LSP 的开销。由于在聚合的拓扑中进行 MPLS 交换，因而无需预先建立 mp2p 的反向树，减少了对节点维护的拓扑信息的要求；同时也可以以具有不同入口和出口 LER 的交换路径进行聚合，因此聚合率高，适应性强，能有效地提高网络的可扩展性。

2 聚合算法

在图论中定义无向图 $G=(V,E)$（简称为图 G），其中 $V=\{v_1,v_2,\cdots\cdots,v_n\}$ 为 G 的点集合，$E=\{e_{ij}\}$ 为 G 的边集合，并且 $e_{ij}=\{v_i,v_j\}$ 是一个无序二元组。在图 G 中，一个点和边的交替序列 $(v_i,e_{ij},v_j,\cdots\cdots,v_k,e_{kl},v_l)$ 称为 G 中由 v_i 到 v_l 的一条路，记为 $\{v_i,v_l\}$ 路。定义图 G 的支配集 D，它是 G 的一个子集，满足：$\forall\ v\in V, v\in D$ 或 v 与 D 中的节点相邻。D 中的节点称为支配点。

定义了支配集之后，下面对基于 *LER* 支配集的拓扑聚合算法进行描述。

设网络拓扑图 $G=(V,E)$，$G_e=(V_e,E_e)$ 为 G 的一个子图，其中 V_e 为边界节点（LER）的集合，$V_e\subseteq V$。$\forall\ e\in E_e, e=\{n_i,n_j\}$，有 $n_i\in V_e, n_j\in V_e$。下面定义聚合算法中将用到的部分变量。

D：边界节点的支配点集合，$D\subseteq V_e$；

$|V_e|$：边界节点的数目，记为 l；

$|D|$：D 中节点的数目，记为 m；

d_i：与节点 i 邻接的边界节点数目，称为节点 i 的度数，由拓扑决定；

A：聚合度，$A=m/l$；

δ_i：支配点 i 的支配了被节点 i 支配的节点的数目；

$\bar{\delta}$：平均支配度，$\bar{\delta}=\frac{1}{m}\sum_{i=1}^{m}\delta_i$；

本文提出的聚合算法，主要包括支配集 D 的生成和支配节点的选择两个部分，下面是算法的具体步骤。

(1)N 中每个节点 i 维护其邻接的边界节点数目值 d_i，并以 1 跳为信息交换半径，周期性地和邻节点交换其 d_i、δ_i 的值以及支配点集的信息。

(2)初始时每个节点都不是支配节点，在每个周期内，各节点交换信息交换半径内的支配点的信息，计算新的节点支配信息状态，其计算规则为：

(a)若节点 i 的 $d_i=0$，则节点 i 为支配点。

(b)若节点 i 有 $d_i=1$ 的邻点节，则 i 为支配点。

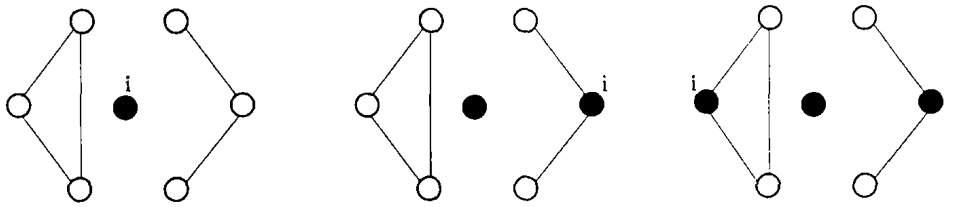
(c)若节点 i 的邻节点中无支配点，则节点 i 成为支配点。

以上三个计算规则形成了支配点集合 D。

(3)各非支配节点选择各自的支配节点，其选择规则为：

(*a*)若节点 j 的邻节点仅有一个支配点 i，则 j 受 i 支配；

(*b*)若节点 j 的邻节点有多于一个支配点 $1,\cdots,k$，则选择节点 t，满足 $\delta_t=\min(\delta_1,\cdots,\delta_k)$，作为其支配点。



规则 (a)，点 i 的度数为 0，点 i 为支配点

规则 (b)，点 i 具有度数为 1 的邻节点，点 i 为支配点

规则 (c)，点 i 的邻节点中没有支配点，点 i 为支配点

图 1 各条计算规则的图示说明

以上算法可以保证所生成的集合 D 为支配集。简要证明如下：

在子图 G_e 内，计算规则(a)保证了孤立的节点必为 D 内的元素，即支配点；计算规则(c)保证了若某一节点 $i\notin D$，则 i 的邻节点中必有一个是支配点。由此可知，G_e 内任一节点或者为支配点，或者与某一支配点相邻。因此上述算法生成的集合 D 为支配集。

同时，算法规则(a)和(b)可以起到加快网络节点确定其是否为支配点的作用，是对生成支配点集合的一种优化。

由于使用支配 LER 代表其支配域中的所有节点，因此可以通过仅以支配 LER 为出口节点来建立 LSP，从而达到聚合的目的。

3 性能评估

(1)*算法复杂度分析* 对于支配集的生成，有多种算法可以获得极小支配集，但是计算极小支配集的算法比较复杂，而且需要获得整个图的拓扑信息，如布尔代数算法。用极小支配集对网络拓扑进行聚合，其计算收敛速度慢，健壮性差，因此并不是最佳选择。该算法利用各节点及其相邻节点所维护的支配信息，在局部范围内进行交换和计算，各节点独立确定其支配状态，因此算法分布性强，计算复杂度低。而且可以利用支配点的冗余性设计网络的保护方案，例如当某支配节点或支配节点到被支配节点之间的链路出现故障时，可以让具有多重支配特性的被支配节点重新选择其支配节点，以提高网络的健壮性。

(2)*支配 LER 的处理开销* 支配 LER 对于输入的分组需要对其下一跳的目的地进行判断，因此需要检查 IP 层的首部以得到路由决策，如果是发往其被支配 LER 的分组，则需要重新进行标签封装。相比较传统的 LER 而言，这是一种额外的开销，对网络的性能，特别是时延将会产生影响。本算法在被支配 LER 选择其支配 LER 时，选取具有最小支配度的节点，使各支配节点的支配度尽量平均，即使得平均支配度 $\bar{\delta}$ 与聚合度 A 的值尽量接近，避免某些节点负担过重的情况产生。

(3)*LSP 和标签的数目* 对于没有采用聚合策略的 MPLS 网络，某节点 i 对所有 n 个出口节点提供可达性需要创建 n 个交换路径，若每个出口节点所连接的目的主机的平均数量为 d，则节点 i 平均需要 $n\times d$ 个标签对所有目的主机提供可达性。采用该算法聚合后，节点 i 平均仅需要创建 n/A

（下转第 47 页）

是指用户遭受的损失可以最小。

在BWFEP协议的第一步中，如果对协商的结果不满意，任意一方均可以退出协议，因而不存在不公平现象。步骤2中，由于传输的是加密数据，因此按照我们关于"算法不可破"的假设，虽然上方均得到了数据，但是由于没有解密密钥，因而恢复数据是不可能的。此外，即便某一方的计算能力比另外一方强，可以破解加密算法，得到 N 个明文数据块，但是由于 N 足够大，因而重组数据块也需要更多的计算开销。因此，第二步中也可以满足公平性的要求。

在第二步中，虽然协议对双方都是公平的，但并不排除发送虚假的数据，即第二步中可能存在欺骗现象。这个问题可以在第三步中被检查出来，这是因为发送方可以随机要求对方重组其中的任意两块。如果在第二步发送的数据为虚假数据，则重组后的结果自然不成立，用户可以立即发现该问题，并终止协议。第三步中另外一个问题是，如果用户不诚实，也有可能存在欺骗问题，从而导致协议的弱公平性。比如用户A如果不诚实，他可以发送不正确的解密密钥 K_{A_i} 或者 $K_{A_{i+1}}$，使得用户B不能解密或者重组数据，因此协议对B是不公平的。但是我们认为这只是弱公平性。理由如下：(1)在步骤3中，用户只可能蒙受一个数据块的损失，这是显而易见的，因为如果用户一旦发现重组后的数据不满意，可立即退出协议。(2)由于 N 足够大，数据块又是随机的，因而一个数据块的价值很小，这样的损失用户可以接受，即因为协议的不公平性，用户的损失可以最小化，保证了我们对"弱公平性"的要求。在步骤3中，唯一可能导致不公平的情况是最后一个数据块的交换，即一方得到了最后一个数据块，但却拒绝发送自己的最后一个数据块。在BWFEP协议中，没有考虑这一问题，我们拟在今后的问题中采用零知识证明的方法解决这个问题。

总结 对等计算模型在移动计算，移动自组网等均有着广泛的应用前景。本文设计的基本弱公平交换协议(BWFEP)，可在无可信第三方的条件下保证协议双方资源交换的弱公平性，从符合对等网络的基本要求和特征。对该协议的分析结果表明，BWFEP协议无法保证用户在最后一个数据块发送时实施欺骗行为，因此还需要对协议加以改进，从而满足用户对资源交换的强公平性的要求。

参考文献

1 Napster Home Page. http://www.napster.com/.
2 Yang B, Garcia-Molina H. Comparing Hybrid Peer-to-Peer Systems. Technical report. Stanford University, Feb. 2001. Available at: http://dbpubs.stanford.edu/pub/2000-35
3 Buyya R. Economic models for Management of the Resources in Peer-to-Peer and Grid Computing. Proc. of the Commercial Applications for High-Performance Computing Conf. 2001
4 Milojicic D S, et al. Peer-to-Peer Computing, HP Laboratories Palo alto: [Technique Report: HPL-2002-57]. March, 2002
5 王彩芬，葛建华．带脱线半可信第三方的公平非否认交换协议．电子学报，2002，30(2)：286～288
6 姬东耀，王育民．基于Pay Word的小额电子支付协议．电子学报，2002，30(2)：301～303
7 李志江，李明柱，杨义先，等．一个实用的公平电子合同协议．北京邮电大学学报，2002，25(2)：28～32
8 邓所云，詹榜华，胡正义，等．一个优化的公平的电子支付方案．计算机学报，2002，25(2)：1094～1098
9 Ray I, Ray I. Fair Exchange in E-commerce. ACM SIGecom Exchange, 2002, 3(2): 9～17
10 Levente B, Jean-Pierre H. Ratinal Exchange - A Formal Model Based on Game Theory. In: Proc. for the 2nd Intl. Workshop on Electronic Commerce (WELCOM 2001), Springer-Verlag, 2001
11 Shmatikov V, Mitchell J C. Finite-state analysis of two contract signing protocols. Theoretical Computer Science, 2002, 283(2): 419～450
12 Pfitzmann B, Schunter M, Waidner M. Optimal efficiency of optimistic contract signing. In: Proc. of the seventeenth annual ACM symposium on Principles of distributed computing, 1998. 113～122
13 周世杰，秦志光，张峰，等．基于P2P的信息存储技术．见：第12届全国信息存储技术学术会议(NCIS2002)，上海，2002

(上接第34页)

个交换路径，而所需要的标签数目为 $(n/A)\times d$。如果结合标签的聚合策略，例如令具有相同出口LER的分组(即具有相同CIDR前缀的分组)都使用相同的标签，则仅需要 n/A 个标签，这极大地提高了网络的可扩展性。当然，这是以支配LER的处理开销为代价的。

(4)时延分析 时延包括节点的时延和链路的时延。这里节点时延主要考虑支配LER的时延，因为支配LER的处理速度对该聚合网络的性能产生较大影响；而连路的时延主要考虑被支配LER到支配LER的链路的时延。如果图 $G=(V, E)$ 每个节点 $v\in V$ 具有一个非负的权 w，用 w_v 表示，那么考虑LER时延的支配集 D 的选取可以描述为：选择 D，使得 $W_D=\sum_{i\in D} w_i$ 最小。这是一个NP完全问题[9]。链路时延的减小可以考虑在链路最大时延要求下，被支配LER选择满足时延约束条件的支配LER作为其支配点，同时支配LER的选取也可以综合考虑节点的时延和链路的时延。这都将在后续工作中进行研究。

结论 本文提出了一种通过构建LER支配集对MPLS网络进行拓扑聚合的算法，证明了算法生成的支配集的正确性，并对算法和聚合网络的性能进行了分析。由于该聚合策略采用分布式算法，实现简单，对网络拓扑的变化具有较好的适应性，并能有效地减少MPLS网络的LSP和标签的数量，从而减少控制信息的通信量。同时支配LER的冗余性可以作为LSP的保护备份，因此使MPLS网络具有较好的可护展性和健壮性。

参考文献

1 Rosen E, Viswanathan A, Callon R. Multiprotocol Label Switching Architecture. RFC 3031, Jan. 2001
2 Awduche D, et al. RSVP-TE: Extensions to RSVP for LSP Tunnels. RFC3209, Dec. 2001
3 Andersson L, et al. LDP Specification, RFC3036, Jan. 2001
4 Fredette A, White C. Loa Andersson and Paul Doolan, Internet Draft, draft-fredette-mpls-aggregation-00. txt, Nov. 1997
5 Saito H, Miyao Y, Yoshida M. Traffic Engineering using Multiple Multipoint-to-Point LSPs. IEEE INFOCOM'2000, Tel Aviv, Israel, 2000, 2: 894～901
6 Urvoy-Keller G, Hébuterne G, Dallery Y. Traffic Engineering in a multipoint-to-Point Network. IEEE Journal on Selected Areas in Communications, 2002, 20(4): 834～849
7 Bhatnagar S, Ganguly S, Nath B. Label Space Reduction in Multipoint-point LSPs for Traffic Engineering. In: 2nd European Conf. on Universal Multiservice Networks, ECUMN'2002, Colmar, France, 2002. 29～35
8 Oh Y K, et al. Scalable MPLS Multicast using Label Aggregation in Internet Broadcasting Systems. In: 10th Intl. Conf. on Telecommunications, ICT'2003, Tahiti Papeete, French Polyn-esia, 2003
9 Alber J, et al. Fixed Parameter Algorithms for dominating set and related problems on planar graphs. Algorithmica, 2002, 33: 461～493