

支持时延-带宽约束的动态层次组播路由

胡虹雨 陆慧梅 曹元大 饶思敏

(北京理工大学计算机科学技术学院智能信息技术北京市重点实验室 北京 100081)

摘 要 层次网络及层次路由成为解决大规模网络 QoS 路由可扩展性问题的一个主要手段. 文中对 PNNI 层次网络模型下的时延-带宽多 QoS 约束的动态组播路由问题进行了全面研究; 在已提出支持时延-带宽约束的拓扑聚集算法(Stair)的基础上, 进一步对组播树节点需维护的组播树状态信息及其聚集问题进行研究, 并提出“伪树上边界节点”模式的域内组播树状态信息的聚集方法, 最后设计了基于聚集拓扑信息和组播树状态信息的动态层次组播路由算法. 仿真结果显示, 该路由由不仅大量压缩了存储和扩散的拓扑信息和组播树状态信息, 同时还保持了与平面网络近似的路由效率, 实现了大规模网络情况下组播路由的扩展.

关键词 时延; 带宽; 组播树状态信息; 组播路由; 层次路由

中图法分类号 TP393

DOI 号: 10.3724/SP.J.1016.2009.00564

Dynamic Hierarchical Multicast Routing with Delay-Bandwidth Constraints

HU Hong-Yu LU Hui-Mei CAO Yuan-Da RAO Si-Min

(Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology,
Beijing Institute of Technology, Beijing 100081)

Abstract In order to solve the scalability problem of QoS routing in large networks, large networks are structured hierarchically by grouping router nodes into different domains. Routing algorithms based on hierarchical networks are investigated and become one of the most effective solutions for the scalability problem. The authors conduct a comprehensive study on dynamic hierarchical multicast routing with delay-bandwidth constraints based on PNNI hierarchical network model in this paper. A novel topology aggregation method, Stair method, has been proposed in the authors' previous work, which is the first step for hierarchical routing and in which two QoS parameters (delay and bandwidth) are considered. On this basis, the authors further study the Multicast Tree State Information (MTSI) that tree nodes need to maintain for multicast routing and the aggregation problem of internal MTSI of a domain. Pseudo Tree Border Node (PTBN) mode for MTSI aggregation is then proposed in this paper. Finally Delay-Bandwidth constrained Hierarchical Multicast Routing (DBHMR) algorithm using aggregated topology information and multicast tree state information is designed, which supports group members to join and leave multicast group dynamically. Simulation results show that the whole routing system achieves large compression of topology information and multicast tree state information not only stored in but also exchanged by nodes and at the same time gains the similar route efficiency compared with the flat network. Scalable multicast routing in large networks is realized eventually in this paper.

Keywords delay; bandwidth; multicast tree state information; multicast routing; hierarchical routing

收稿日期: 2008-09-20; 最终修改稿收到日期: 2009-01-06. 本课题得到国家自然科学基金(60503050, 60773045)、北京理工大学基础研究基金项目(200511F4212)和北京市重点学科建设项目资助. 胡虹雨, 女, 1976 年生, 博士研究生, 主要研究方向为 QoS 路由、IP 组播. E-mail: hu_hong_yu@163.com. 陆慧梅(通信作者), 女, 1974 年生, 博士, 副教授, 主要研究方向为下一代互联网体系结构、组播通信. E-mail: luhuimei@bit.edu.cn. 曹元大, 男, 1944 年生, 教授, 博士生导师, 研究领域包括人工智能、计算机网络和网络安全. 饶思敏, 女, 1986 年生, 硕士研究生, 主要研究方向为网络与分布式计算.

1 引言

当今 Internet 规模正以指数速度增长. 在大规模网络情况下, 路由特别是 QoS 路由计算对路由器内存、CPU 以及网络带宽资源的消耗急剧增多, 路由面临可扩展性问题. 将网络进行层次划分、域内扩散原始网络拓扑信息、域间扩散聚集后的信息, 并在聚集信息的基础上实现层次路由成为解决路由可扩展性问题的一个主要手段^[1-2].

在层次网络模型下支持 QoS 的动态组播路由需要解决的主要问题有: (1) 如何高效地实现域内拓扑压缩, 以较少信息失真获得较高信息压缩. 我们已在文献[3]中对此问题进行了深入研究, 提出了全新的 Stair 拓扑聚集算法. Stair 算法采用“规则楼梯”来模拟两边界节点间多条通路的 QoS 属性, 聚集后的域拓扑信息复杂度为 $O(|B|)(|B| \text{ 为域的边界节点的个数})$. 文献[3]通过聚集前后的“QoS 支持区域”的变化情况(即面积偏差)对 Stair 算法进行了评价, 结果显示, 与其它聚集算法^[4-5]相比, Stair 算法获得了较小的面积偏差, 从而获得了较小的信息失真. (2) 动态组播路由所需的组播树状态信息有哪些, 这些信息是否需要聚集以及如何聚集. 文献[6-7]讨论了当网络中启动较多组播组时, 路由器中组播树转发信息的聚集问题. 它们大多是通过在多个组播组之间建立共享树来达到减少组播路由条数的目的. 目前鲜有文献对域的组播树状态信息聚集问题展开研究, 仅文献[2]对此进行了初步讨论. 文献[2]仅考虑了单 QoS(带宽)约束情况下组播树节点的 QoS 能力信息, 域的组播树状态信息由边界节点的最大可接收能力信息表示. (3) 如何依据聚集后的拓扑信息和组播树状态信息组织跨域路由, 并保持与平面网络情况近似的路由效率. 文献[1, 2, 5]对此进行了研究; 文献[1-2]提出的层次组播路由以 HMR^[8]层次网络模型为基础, 通用性不强; 文献[5]仅研究了 PNNI^[9]模型下的单播路由.

本文针对加性(时延)和凹性(带宽)两个 QoS 约束, 全面研究了 PNNI 层次网络模型下的动态组播路由问题, 整体框架结构如图 1 所示. 整个系统以层次网络模型为基础, 以组播路由算法为中心, 所有节点维护自己所在域的详细的以及其它域聚集形式的拓扑信息和组播树状态信息, 随着组成员加入请求(/退出请求)的提出, 这些信息将作为组播路由算法的输入, 以计算是否存在满足用户时延-带宽需求

的一条可行嫁接路径. 域的拓扑信息聚集部分的研究工作(图 1 中斜线部分)已在文献[3]中介绍, 本文不再赘述. 本文重点研究: (1) 域的组播树状态信息的聚集和压缩(图 1 中网格部分). 对组播树节点的组播树状态信息进行定义, 将域内树上节点的组播树状态信息限制在本域内, 同时配置非树上边界节点为伪树上节点, 通过其对域内组播树状态信息进行聚集并向外扩散, 达到屏蔽内部细节和减少域间扩散的组播树状态信息的效果; (2) 基于聚集形式的拓扑信息和组播树状态信息的组播路由(图 1 中阴影部分), 设计时延-带宽约束下的动态层次组播算法(Delay-Bandwidth constrained Hierarchical Multicast Routing, DBHMR), 经过跨域分支计算、中间域内分支细化以及目的地域内嫁接点优化等步骤, 成功构建满足用户时延、带宽请求的动态跨域组播树. 仿真结果显示, 该系统在实现拓扑信息和组播树状态信息大量压缩的情况下, 仍保持了与平面近似的路由效率, 实现了大规模网络下组播路由的扩展.

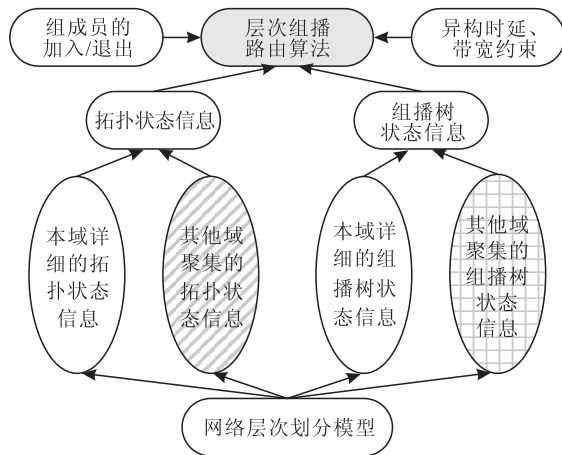


图 1 层次组播路由框架

本文第 2 节介绍相关模型; 第 3 节讨论组播树状态信息聚集与扩散, 引入伪树上边界节点聚集与扩散模式; 第 4 节全面介绍了 DBHMR 动态层次组播路由算法; 第 5 节为仿真结果; 第 6 节对全文进行了总结.

2 PNNI 层次网络模型及 QoS 属性

PNNI(The Private Network-to-Network Interface)^[9]作为网络层次划分模型事实上的标准, 不仅适合 ATM 网络, 还适合 IP 等其它网络. 本文讨论 PNNI 二层网络, 整个网络可表示为 $N=(A, L)$,

其中 A 为网络中所有域的集合, L 为域间物理链路的集合. 假设第 i 个域表示为 $A_i = (V_i, B_i, E_i)$, 其中 V_i 为域 A_i 的所有节点的集合, $B_i \subseteq V_i$ 为边界节点的集合, E_i 为域 A_i 内物理链路的集合. 定义 V 表示网络中所有节点的集合. 定义 E 表示网络中所有物理链路的集合.

图 2 为一个 PNNI 二层网络的示意图, 其中节点被划分至 A、B、C、D、E 5 个域中. 由于存在拓扑信息聚集和屏蔽, 域 D 中的节点看到其它域的拓扑信息为聚集形式的 (即由抽象的逻辑链路构成), 而收集到的本域拓扑信息是详细的 (即由原始的物理链路构成). 组播树源节点 v_r 在域 A 中, 已构造的组播树如图 2 中细虚线所示. 若域 E 中的节点 v_0 欲加入到组播树中, 它将就近选择域 D 中的已在树上的节点作为嫁接点进行加入. 随着组成员的动态加入和退出, 最终将构建一棵跨域的组播树.

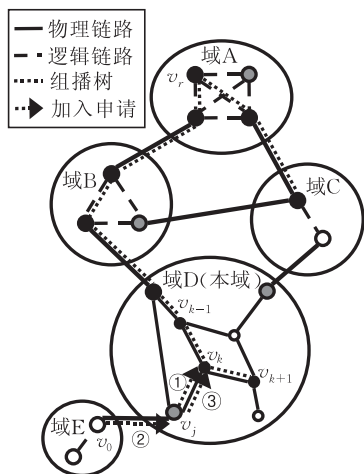


图 2 PNNI 二层网络及跨域组播树示意图

对于任意一条物理链路 $e \in E$, 定义两种 QoS 度量函数, 分别为: (1) 时延函数 $d(e): E \rightarrow R^+$; (2) 可利用带宽函数 $w(e): E \rightarrow R^+$, R^+ 表示正实数集. 对于节点 v_0 到 v_k 之间的一条物理路径 $p(v_0, v_k)$, 设其对应 QoS 参数为 (d_p, w_p) , 根据这两个参数的特性可推出:

$$d_p = \sum_{i=0}^{k-1} d(v_i, v_{i+1}) \quad (1)$$

$$w_p = \min_{i=0}^{k-1} w(v_i, v_{i+1})$$

3 组播树状态信息的聚集与扩散

在支持时延-带宽约束的动态组播路由中, 当加入节点 v_0 在选择嫁接点 (设为 v_k) 时, 需要考察 v_0 到

v_k 以及 v_k 到组播源 v_r 两段子路径的时延和带宽信息. 如果组播树上节点 v_k 预先已获得其到 v_r 子路径的 QoS 信息并扩散给节点 v_0 , 则将能加快申请节点 v_0 的计算分析速度. 树上节点 v_k 维护的其到 v_r 子路径的 QoS 属性体现了该节点对外提供的嫁接能力, 并间接体现了用户以 v_k 为嫁接点嫁接到组播组后能够获得的服务质量. 因此每个成功加入组播组的节点都应维护其到组播源的组播分支的 QoS 状态信息, 即组播树状态信息.

3.1 组播树状态信息定义

组播树节点维护的组播树状态信息包括组播树分布情况以及组播树上节点的 QoS 能力.

定义 1. 组播树覆盖关系 (Multicast, M) 确定了组播树分布情况. M 表示组播树覆盖的网络节点的集合, $M \subseteq G \times V$, G 表示所有组播组的集合, $G = \{G_i | i = 1, 2, \dots, k\}$. $(G_a, v_k) \in M$ 表示路由器 v_k 位于组播组 G_a 的组播树 T_{G_a} 上.

定义 2. 组播树上节点的目前累计时延 (Accumulated Delay, AD) 是指数据从组播源沿组播树分支到达该树上节点所需时延: $AD: M \rightarrow \{R^+, 0\}$. 设组播组 G_a 的组播源为 v_r , 树上节点 v_k 的目前累计时延为

$$AD(G_a, v_k) = \sum_{i=r}^{k-1} d(i, i+1) \quad (2)$$

定义 3. 组播树上节点的最大可接收能力 (Max Receivable Layers, MR) 是指该节点所能收到组播应用数据的最高层数: $MR: M \rightarrow I^+$, I^+ 表示正整数集. 如果某节点的最大可接收能力为 i , 则表明其可以接收 $1, 2, \dots, i$ 层的组播数据. 节点的最大可接收能力并不等于其自身当前的接收层数, 而是指在当前接收层数的基础上, 加上从源到该节点的剩余可利用带宽所能容许传送的最大层数:

$$MR(G_a, v_k) = \min(MR(G_a, v_{k-1}),$$

$$ToLayer(GW_{G_a}(v_{k-1}, v_k) + w(v_{k-1}, v_k))) \quad (3)$$

其中, $MR(G_a, v_{k-1})$ 表示上游节点 v_{k-1} 的最大可接收能力, $GW_{G_a}(v_{k-1}, v_k)$ 表示 v_{k-1} 至 v_k 间物理链路上已经为组播组 G_a 预留的带宽, $w(v_{k-1}, v_k)$ 表示 v_{k-1} 至 v_k 间物理链路上的剩余可利用带宽. $ToLayer(w)$ 表示将带宽 w 转换成对应的组播数据层数, 我们称之为“层宽”.

定义 4. 在时延-带宽约束下, 组播树上节点的组播树状态信息 (Multicast Tree State Information, $MTSI$) 定义为二元组 (AD, MR) .

$MTSI$ 体现了该树上节点当前属于哪个组播组, 以及时延、带宽方面的 QoS 能力信息. 这些信息

向外扩散后,成为其它节点加入计算时是否选择其作为嫁接点的判断依据。

3.2 组播树状态信息聚集——伪树上边界节点聚集模式

随着组规模的增大,成功加入组播组的节点越来越多,最极端情况下其数目达到 $|V|$.如果各树上成员节点的 MTISI 不经过处理在全网扩散,网络中转发的和节点存储的 MTISI 数目将急剧增多,组应用同样面临规模可扩展性问题.另一方面,网络经域的划分后,“域内信息需对外屏蔽”,域的 MTISI 信息需同域的拓扑信息一样遵循该原则。

为减少和屏蔽域的 MTISI 信息,我们首先想到对不同类型节点的组播树状态信息扩散范围进行区分和限制:树上节点如果是边界节点则在全网扩散其 MTISI,如果是域内部节点则限制其 MTISI 在本域内扩散,全面描述见“3.3 节组播树状态信息的扩散”.通过该限制,域内信息得到屏蔽同时域间扩散的 MTISI 大大减少.但是域 A_i 外部的节点只能获得 A_i 边界节点的 MTISI 而无法获得 A_i 内部节点的 MTISI,外部节点嫁接到更优节点的机会将损失。

进一步我们观察到,图 2 中,域外节点 v_0 欲嫁接到该域某内部树上的节点 v_k 上,其嫁接分支必先经过该域的某非树上边界节点 v_j .当 v_0 成功加入后, v_j 也将成为 v_k 的下游节点,即外部节点的加入将触发非树上边界节点的加入.受此启发,我们考虑将加入触发顺序反过来,如图 2 中①、②、③加入顺序所示:①让非树上边界节点 v_j 预先成为内部树上节点 v_k 的下游节点;②域外节点 v_0 在实际加入时将选择 v_j 作为嫁接目标;③当加入请求到达 v_j 后, v_j 再在域内确定一个最优的树上节点并细化域内嫁接分支.最终域外节点可成功嫁接到域内树上的节点上,从而建立最优的嫁接分支和组播树。

通过步骤①,边界节点 v_j 将获取 MTISI 信息,该信息间接体现了 v_k 的 MTISI 信息以及其嫁接能力.因此可以认为是非树上边界节点代替内部树上节点向外扩散其嫁接能力,从而达到屏蔽域内细节和减少域间 MTISI 的效果.由于该非树上边界节点 v_j 并没有真正连接到树上,我们称这种“预先嫁接并获得内部组播树信息”的边界节点为伪树上边界节点(Pseudo Tree Border Node, PTBN).图 2 中,黑色节点为树上节点,灰色节点为 PTBN,白色节点为非树上节点。

定义 5. 伪树上边界节点(Pseudo Tree Border Node, PTBN). 对于任一组播组 G_a 来说,如果域 $A_i(V_i, B_i, E_i)$ 的某边界节点 v_j 不是组播组 G_a 的

树上节点,同时域内存在组播组 G_a 的树上节点(不含树上边界节点),则节点 v_j 配置自己为伪树上边界节点.记 PM 为组播树与伪节点的覆盖关系, $PM \subseteq G \times V$,则 $(G_a, v_j) \in PM$ 表示 v_j 是组播组 G_a 的一个伪节点。

从以上定义可以看出,只要域内部存在树上节点,则域的所有非树上边界节点都将启动为 PTBN,因此可以推出以下定理。

定理 1. 对于任意一个域 A_i ,如果内部存在关于组播组 G_a 的树上节点,则 A_i 的所有边界节点要么真正位于 G_a 组播树上,要么为 G_a 的伪树上节点。

根据定理 1,树上边界节点和伪树上边界节点一起,将全面提供域的各个方向上的组播树状态信息,从而向外界反映域的全方位接入能力,起到屏蔽内部细节和压缩信息的作用.我们称这种组播树状态信息聚集模式为伪树上边界节点模式。

当非树上边界节点启动 PTBN 配置时,主要对内部树上节点组播树状态信息进行收集.由于一个域内部可能存在多个树上节点,因此边界节点配置为 PTBN 的过程分为两个子步骤:“收集一个树上节点的 MTISI”和“聚集多个节点的 MTISI”。

3.2.1 PTBN 收集一个树上节点的 MTISI

边界节点 v_j 收集内部树上节点 v_k 的 MTISI,是假定自己嫁接到 v_k 上成为其下游节点从而获得自己的 MTISI.当 v_j 嫁接到 v_k 上时,可能会存在多条嫁接分支 b_1, b_2, \dots, b_z ,设这些嫁接分支的 QoS 能力为集合 $bqos_{v_j, v_k} = \{(d_{b_i}, w_{b_i}) | i=1, 2, \dots, z\}$.结合 v_k 的组播树状态信息 (AD_{v_k}, MR_{v_k}) , v_j 通过 $bqos_{v_j, v_k}$ 中不同分支成为 v_k 的下游节点后,其 MTISI 也是一个集合,记为 $ptqos_{v_j, v_k}$:

$$\begin{aligned} ptqos_{v_j, v_k} = \{ & (AD_{v_j, v_k}, MR_{v_j, v_k}) | AD_{v_j, v_k} = AD_{v_k} + d_{b_i}, \\ & MR_{v_j, v_k} = \min(MR_{v_k}, ToLayer(w_{b_i})), \\ & (d_{b_i}, w_{b_i}) \in bqos_{v_j, v_k} \} \end{aligned} \quad (4)$$

以图 2 域 D 进行举例说明, v_k 和 v_{k+1} 为域内部树上节点, v_j 为非树上边界节点, v_j 对 v_k 的组播树状态信息收集过程说明如下:设 v_k 的 MTISI 为 (d_k, L_2) ,假定 v_j 到 v_k 之间有 5 条分支,分支的 QoS 属性分别为 $(d_{k1}, w_{k1}), (d_{k2}, w_{k2}), (d_{k3}, w_{k3}), (d_{k4}, w_{k4}), (d_{k5}, w_{k5})$.我们将这些分支的带宽转换成对应媒体分层的层数,设转换后的 QoS 属性为 $(d_{k1}, L_1), (d_{k2}, L_1), (d_{k3}, L_2), (d_{k4}, L_3), (d_{k5}, L_3)$,如图 3 所示, $ToLayer^{-1}()$ 为 $ToLayer()$ 的逆函数.结合 v_k 的 MTISI,则 v_j 通过 v_k 成为树上节点后,获得的 MTISI 为 $(d_k + d_{k1}, L_1), (d_k + d_{k2}, L_1), (d_k + d_{k3},$

$L_2), (d_k + d_{k4}, L_2), (d_k + d_{k5}, L_2).$

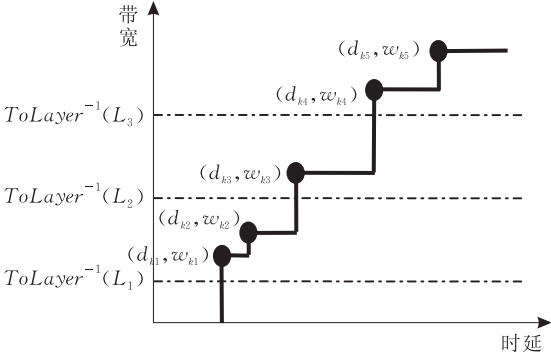


图 3 带宽折合成层宽示意图

3.2.2 PTBN 聚集并过滤多个树上节点的 MTSI

如果域内存在多个树上节点 v_0, v_1, \dots, v_{n-1} , 则 v_j 需将针对不同树上节点 $v_k (0 \leq k \leq n-1)$ 形成的组播树状态信息 $ptqos_{v_j v_k} = \{(AD_{v_j v_k}, MR_{v_j v_k})\}$ 进行聚集. 我们首先将它们进行合并, 形成集合 $ptqos_{v_j} = \bigcup_{k=0}^{n-1} ptqos_{v_j v_k}$.

在未知用户 QoS 需求的情况下, 我们需保留 $ptqos_{v_j}$ 中最好的 MTSI 并向外扩散. 但是在时延、带宽两个约束条件下, 在 $ptqos_{v_j}$ 中确定一个最优的 MTSI 将十分困难. 同“两点间多条通路可以过滤出代表点通路”^[3] 一样, 多个 MTSI 也可以过滤出其代表点信息, 代表点信息能无损体现原多个嫁接点及多条嫁接分支的 QoS 能力. 依据代表点的特性, $ptqos_{v_j}$ 的代表点集合定义如下:

$$ptqos'_{v_j} = \{(AD_y, MR_y) | (AD_y, MR_y) \in ptqos_{v_j}, \\ \forall (AD_x, MR_x) \in ptqos_{v_j}, (AD_y < AD_x) \text{ or } (MR_y > MR_x)\} \quad (5)$$

代表点集合 $ptqos'_{v_j}$ 即为伪树上节点 v_j 的 MTSI. 由于我们在 $ptqos_{v_j v_k}$ 中已将嫁接分支的带宽属性 w_{b_i} 转换成对应媒体数据的层数 $ToLayer(w_{b_i})$, 因此 $ptqos'_{v_j}$ 中的任意 MR 都将小于 $|MaxLayer|$ ($|MaxLayer|$ 为组播应用的最大媒体分层数), 从而 $|ptqos'_{v_j}| \leq |MaxLayer|$. 最终一个域在域间扩散的 MTSI 将不超过 $|B| \times |MaxLayer|$, 域内组播树信息得到有效压缩.

我们称以上聚集多个组播树状态信息的方式为“层宽代表点”方式, 其为“伪树上边界节点”聚集模式的有效组成部分.

3.3 组播树状态信息的扩散

根据树上节点类型的不同(域内树上节点、树上边界节点和伪树上边界节点), 我们对它们组播树状态信息的扩散模式进行总结并区分. 具体地, 对于某

域 A_i 内任意节点 v_k :

- (1) 如果 v_k 为域内树上节点, 则它将自身组播树状态信息在域 A_i 内扩散;
- (2) 如果 v_k 为该域边界节点且为树上节点, 则它将自身组播树状态信息在域 A_i 及其它域 $A_j (A_j \neq A_i)$ 扩散;
- (3) 如果 v_k 配置为该域伪树上边界节点, 则它将聚集得到的组播树状态信息向其它域 $A_j (A_j \neq A_i)$ 扩散;

由于仅仅是边界节点在域间扩散组播树状态信息, 因此组播树状态信息大大减少, 能够适应大规模网络的情况.

4 时延-带宽约束的动态层次组播路由算法(DBHMR)

本章将全面设计 PNNI 层次网络模型下的满足时延-带宽约束的动态层次组播路由算法 DBHMR. 尽管两个加性参数约束的路由问题被认为是 NP 难的问题, 但是时延-带宽约束的路由问题却可以在多项式时间内解决^[5]. 随着组播成员的加入和退出, 组播树将动态地进行嫁接和剪枝, 因此组播路由的过程实际上就是组播树嫁接和剪枝的过程. 本文将详细介绍组成员加入计算的过程, 组成员退出过程将不细述.

在已有文献中, 文献[10]提出的 QHMR-1 是一个源组播路由算法不支持层次网络, 同时它仅支持一个 QoS 约束(带宽). 文献[5]提出的 LSRA 是一个 PNNI 网络模型下的层次单播路由算法, 它支持时延和带宽两个约束, 但单播路由与我们要设计的组播路由存在很大不同: 单播路由仅仅是形成源节点到目的地节点两点间的一条通路, 而组播路由则要形成源到各成员节点间的一个组播树. 树型结构的一个显著特性是没有环路. 因此, 在设计层次组播路由时, 需要考虑所构造的组播树不能存在环路.

综合以上考虑, 我们结合 QHMR-1 和 LSRA 设计了 DBHMR 算法, 其主要思想是: 加入节点首先进行“跨域路由”计算, 得到嫁接点以及一条到达该嫁接点的满足用户时延和带宽请求的跨域路径. 然后加入节点沿该路径发送加入请求消息, 加入请求消息不仅携带了该跨域路径信息, 同时将记录沿路经过链路的实际的时延、带宽信息. 由于加入节点在计算跨域路径时, 并未获得其它域的详细的拓扑信息, 因此基于聚集信息计算出来的跨域路径为一

条粗略路径,当加入请求消息到达某中间域的边界节点,该节点需要启动“中间域内路由”计算来细化通过该区域的路径.细化的域内路径将补充入加入请求消息中,同时加入请求消息将继续前递.当加入请求消息到达目地域的边界节点时,将启动“目地域内路由”计算来细化一个具体且更优的嫁接点.整个加入计算过程可分为:跨域路由、中间域内路由、目地域内路由、带宽调整、连接建立以及路由回退等子过程.

设组播源节点 v_r 发起一个组播应用 G_a ,其组播数据编码成 L 层.组播源节点初始化组播树为 v_r 本身,其 MTISI 配置为: $AD(G_a, v_r) = 0$, $MR(G_a, v_r) = L$. v_r 依据其节点类型在相应范围内扩散以上信息.如果 v_r 为一个内部节点,则它会触发其它不在树上的边界节点成为“伪树上边界节点”,并将以上信息在全网域内扩散.最终网络中的其它节点都将收到不同详细程度的组播源 v_r 的 MTISI.

假定某路由器 v_0 要求加入组播 G_a ,且用户的带宽要求为 W ,时延要求为 D ,因此加入计算的任务就是找到一个树上节点 v_s , v_s 的最大可接收能力不小于 W ,且从 v_s 到 v_0 路径上的可利用带宽不小于 W ,同时从 v_r 到 v_s 再到 v_0 的路径的时延不超过 D .

4.1 跨域路由算法

“跨域路由算法”为一个集中式算法,加入节点 v_0 需维护全局的拓扑信息(包括本域的详细拓扑信息以及其它域 Stair 算法聚集后的拓扑信息)以及全局 MTISI 信息(包括本域的详细 MTISI 和其它域边界节点的 MTISI),并对所有树上节点一一进行可行性分析.加入节点和某树上节点一旦建立嫁接分支,该分支应不能经过其它树上节点,以免引起组播树环路,因此我们在对一个树上节点进行分析时,将其它树上节点临时删除,这样通过 Dijkstra 算法计算得到加入节点到某树上节点的通路,将不含其它树上节点(除终点).经过 MTISI 聚集后,树上边界节点(含伪节点)的 MTISI 已经全面表示了域内部的和边界的组播树状态信息,因此在删除一个树上节点时,我们统一将其所属域一并删除.

算法 1. 跨域路由 (N, M, PM, W, D, v_0) .

1. 将收集到的其它域的聚集拓扑信息恢复成全相连结构;

Stair 拓扑聚集算法解压将得到规则楼梯标识各逻辑链路的全相连结构.

2. 确定全网树上节点集合 V_s ;

依据 PTBN 模式的 MTISI 聚集和扩散的结果, V_s 包含计算节点 v_0 所在域的所有树上节点(边界的以及内部的)以

及其它域的树上边界节点及伪树上边界节点;

3. 逐一对树上节点集合中的每个树上节点 v_s ($v_s \in V_s$) 进行分析,以判断其是否能成为嫁接点:

3.1. 如果 v_s 是真正的树上节点,则判断这个树上节点 v_s 的最大可接收能力是否满足用户带宽请求 W ,如果不满足则继续分析其它树上节点;

3.2. 如果 v_s 为伪树上节点,则判断 v_s 的 MTISI 数组中是否存在大于或等于用户带宽请求 W 的最大可接收能力,如果没有则继续分析其它树上节点;

3.3. 删除网络中所有其它树上节点 v'_s ($v'_s \in V_s, v'_s \neq v_s$). 如果 v'_s 与 v_0 以及 v_s 都不在同一个域,则删除该节点时,一并删除 v'_s 所在域;

3.4. 删除网络中所有带宽小于 W 的逻辑链路;

由于此时其它域的逻辑链路都由规则楼梯来标识,此时根据规则楼梯最高点的带宽值进行判断:如果该逻辑链路的最大带宽都不满足 W ,则该逻辑链路被删除;否则,表明存在大于等于 W 的物理链路,因而该逻辑链路将被保留;

3.5. 确定剩余逻辑链路的时延值;

经 Stair 压缩算法还原后,各域全相连结构中逻辑链路的 QoS 能力表示为时延-带宽平面上的一个规则楼梯,因此根据纵坐标带宽值 W 可以找到对应的最小的时延值 D_{\min} ,如图 4 所示.此最小时延值即为该逻辑链路的时延值.

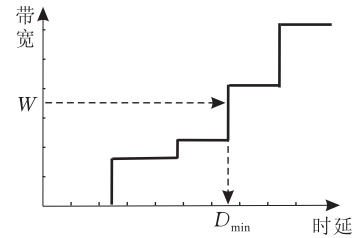


图 4 确定逻辑链路时延值示意图

3.6. 删除网络中所有带宽小于 W 的物理链路;

3.7. 在剩余网络中应用 Dijkstra 算法,来寻找一条从 v_0 到 v_s 的时延值最小的路径.如果计算所得路径的时延值与 v_s 目前累计时延值之和没有超过用户时延请求 D ,则此路径成为备选路径,加入备选路径集;

4. 如果备选路径集不为空,则从中选择时延最小的一条路径作为输出.此路径即为计算得到的跨域路径,对应的树上节点 v_s 即为 v_0 的嫁接点;

5. 计算节点 v_0 将“跨域路径”封装入加入请求消息,并沿此路径发送加入请求消息.

定理 2. 跨域路径上有且仅有目地嫁接点 v_s 为树上节点(含伪树上边界节点).

证明. 根据以上跨域路由算法,计算节点 v_0 确定 v_s 为嫁接点后,计算其到 v_s 的路径时删除了网络中所有已知的其它树上节点(包括 v_0 所在域的树上节点、 v_s 所在域(目地域)的边界树上节点以及伪树上边界节点、存在树上节点)的中间域(即非 v_0 和 v_s 所在的域),因此计算得到的跨域路径上除 v_s 外不

可能再出现其它树上节点(含伪树上边界节点)。

证毕。

通过跨域路由计算,不仅得到了路由器 v_0 嫁接到组播树 G_a 满足用户时延 D 和带宽 W 请求最优的节点 v_s ,即嫁接点,同时还确定了两条之间的一条跨域路径,即嫁接分支.依据定理 2,该嫁接分支将不会对已有组播树构成环路.根据跨域路径,路由器 v_0 知道如何到达本域的一个边界节点,同时知道需穿越哪些中间域到达嫁接点所在目地域.

4.2 中间域内路由算法

当加入请求消息到达某域 $A_i(V_i, B_i, E_i)$ 的边界节点 v_i 、跨域路径指示的下一跳为该域的另一边界节点 v'_i ,且 v'_i 不为跨域路径的终点时,表明加入请求消息到达一个中间域的边界,则由 v_i 节点在本地启动“中间域内路由”算法来细化通过该区域的路径.细化的域内路径将补充到加入请求消息的跨域路径中,同时加入请求消息将会被继续前递.

定理 3. 跨域路径经过的中间域 $A_i(V_i, B_i, E_i)$ 内不存在真实树上节点,即 $\forall v_i \in V_i, (G_a, v_i) \notin M$.

证明. 我们将通过反证法对此进行证明.假设中间域 A_i 内存在组播组 G_a 的成员节点 $v_i (\exists v_i \in V_i, (G_a, v_i) \in M)$,分两种情况进行讨论:

情况 1. 如果 v_i 是边界节点,则 v_i 的组播树状态信息会扩散到加入请求节点 v_0 . v_0 在加入计算时,如果没有选择 v_i 作为嫁接点,则计算嫁接分支时会将没成为嫁接点的其它树上节点 v_i 及其所属域 A_i 一并删除,因此域 A_i 不可能出现在嫁接路径上;

情况 2. 如果 v_i 是内部节点,则 v_i 的组播树状态信息在域内进行扩散,这将会促进伪树上边界节点的形成,同时 v_i 通过伪树上边界节点向外扩散其组播树状态信息.如果伪树上边界节点没有被选择作为嫁接点,则 v_0 计算嫁接分支时同样会将伪树上边界节点及其所属域 A_i 删除,此时域 A_i 同样不可能出现在嫁接路径上;

因此,反过来,我们可以得出结论:出现在嫁接路径上的中间域内一定不存在真实树上节点(也不会存在伪树上边界节点).

证毕。

由于 A_i 内不存在组播树上节点,因此穿过该域的路径不会连接到现有组播树上,不会对组播树现状进行改变,因而也不可能使组播树形成环路,此时的域内路由可以直接参考文献[11]中的 CBDRA (Centralized Band-Delay Routing Algorithm) 算法,

来寻找一条 v_i 到 v'_i 的带宽不小于 W 且时延最小的域内路径。

4.3 目地域内路由算法

当加入请求消息沿跨域路径到达目地域 A_d 的第一个节点 v_d 时, v_d 根据以下 3 种情况进行相应的“目地域”内部处理:

(1) 如果 v_d 是树上边界节点,根据定理 2,则 v_d 必定是跨域路径计算的终点嫁接点.由于 v_d 是一个真实的组播树上节点, v_d 将立即转入“4.4 节嫁接点带宽判断及调整”过程;

(2) 如果 v_d 是伪树上边界节点,根据定理 2, v_d 同样必定是终点嫁接点.由于 v_d 不是一个真实的组播树节点,它是收集了域内组播树节点的信息而形成,因此 v_d 将启动“目地域内路由算法”,以确定一个具体的嫁接点和剩余嫁接路径;

(3) 如果 v_d 是普通边界节点,而其下一跳为嫁接点 v_s (这种情况发生在嫁接点 v_s 所在域内部没有树上节点,树上节点位于该域边界上,而嫁接分支恰好要“穿越”该域才能到达位于边界的嫁接点 v_s . 根据定理 2 只有嫁接分支的终点才为树上节点,因此嫁接分支经过该域的第一个边界节点 v_d 将为非树上节点). 此时 v_d 的下一跳为嫁接点 v_s , 如果直接使用“中间域内路由算法”来细化 v_d 到 v_s 在该域的路径,由于目的域不同于中间域可能存在其它边界树上节点,因此得到的细化路径可能包含其它边界树上节点,这将对组播树构成环路. 另一方面,由于跨域路由计算时,使用的拓扑信息是聚集形式的,有可能在该域存在比 v_s 更优的嫁接点 v'_s (如时延更小). 因此,我们令 v_d 启动“目地域内路由算法”,以在目的域内确定一个更优的嫁接点 v'_s (有可能最终还是 v_s).

“目地域内路由算法”与“跨域路由算法”相似,“跨域路由算法”在全网范围内确定一个嫁接点和嫁接分支,而“目地域内路由算法”将在目地域范围内确定一个更优的嫁接点和剩余嫁接分支,将算法 1 的输入稍作修改即可,包括:将“全网拓扑 N ”替换为“目地域拓扑 A_d ”;将“将全网树上节点集合及其 MTSI”替换为“目地域树上节点集合及其 MTSI”;将“加入节点 v_0 ”替换成“边界节点 v_d ”. 该算法在计算时目地域其余树上节点被删除,因此该段域内分支上将不含树上节点(除新嫁接点 v'_s 外),不会对组播树构成环路。

目地域内路由算法完成后, v_d 将沿计算得到的域内路径向新嫁接点 v'_s 发送加入请求消息. v'_s 收到

加入请求消息后,将转入“4.4 节嫁接点带宽判断及调整”过程。

4.4 嫁接点带宽判断及调整

为支持组成员异构带宽的需求,我们设计组播树上节点向外扩散的带宽信息为“最大可接收能力”^[2],其并不等于树上节点的当前接收层数。因此嫁接点根据以下两种情况执行不同的操作:

(1) 如果嫁接点的当前接收层数能满足用户的请求,则嫁接点直接转入“连接建立”操作,并沿着加入请求消息走过的逆向路径发送“连接建立消息”,以建立组播树分支;

(2) 如果嫁接点的当前接收层数低于用户的请求,但是其最大可接收能力能够满足用户的请求,则嫁接点将沿组播树向上游前递加入请求消息,直到某树上节点的当前接收能力满足用户的请求。随后由该树上节点向下游触发“连接建立”的过程。下游已在树上的节点将为该组播应用增加带宽分配,来满足新用户带宽请求,下游未在树上节点则为组播应用分配全新的带宽,建立全新组播树分支。

4.5 连接建立

收到“连接建立消息”的节点启动连接建立工作,主要有创建或更新组播路由表项、预留建立连接(分支)所需带宽、触发变更后的拓扑信息和组播树状态信息的扩散、继续向下游传递“连接建立消息”等。

4.6 路由回退操作

当加入请求消息在前递的过程中发现链路带宽或累计时延已不满足用户请求,此时发生路由失败,引发路由回退(Crankback)操作,具体为

(1) 发现路由失败的节点 v_k 发送“重试请求消息”回退给前一节点 v_{k-1} ;

(2) v_{k-1} 启动“跨域路由”计算一个从自身到组播树满足用户需求的嫁接点及嫁接路径;

(3) 如果计算成功,则将新的嫁接路径替换原嫁接路径 v_{k-1} 以后的部分,并沿新的嫁接路径向新的嫁接点发送加入请求消息;

(4) 如果计算不成功,则 v_{k-1} 将继续向后回退,继续发送“重试请求消息”给自己的加入请求消息的前一跳 v_{k-2} 。一个加入请求的重试回退操作将不超过 MAX_RETRY_NUM 次;

(5) 如果最终无法找到满足用户需求的嫁接点和分支,则加入申请节点将收到“加入失败消息”,此次加入申请被拒绝。

5 仿真与结果分析

对拓扑信息的聚集效果以及失真情况,而未给出它在真实路由中的信息压缩效果以及对路由的影响。本文将结合 Stair 拓扑聚集算法、PTBN 的组播树状态信息的聚集和扩散模式以及实用的组播路由算法(DBHMR 算法)进行仿真实验,来全面评价层次网络中动态组播路由的拓扑信息、组播树状态信息的压缩效果以及两者对组播路由效率的影响。

5.1 待比较模式

我们设计以下 3 种模式进行仿真实验:

(1) 平面网络模型;

(2) PNNI 二层网络模型启动 Stair 拓扑聚集但不启用 PTBN 功能;

(3) PNNI 二层网络模型启动 Stair 拓扑聚集同时启用 PTBN 功能。

5.2 评价参数

我们从信息压缩特征和路由性能特征两方面对 3 种模式的运行结果进行评价。

信息压缩特征包括:(1) 某一时刻,节点维护的网络拓扑信息数量;(2) 某一时刻,节点维护的组播树状态信息数量;(3) 节点间交换的网络拓扑信息分组数目;(4) 节点间交换的组播树状态信息分组数目。

路由性能特征包括:(1) 路由成功率,为“被接受的加入请求的总数”与“加入请求总数”的比。由于用户在发起加入申请时有时延、带宽约束,因此根据网络当前的负载状态进行路由计算,会客观存在加入成功或不成功的情况。(2) 路由接受率,为“被接受的加入请求的总数”与“跨域路由计算可行的加入请求的总数”的比。由于存在信息失真,原本在跨域加入计算时可行的加入请求在细化找路过程中可能未真正连接成功,因此路由接受是指跨域路由计算成功并且最终建立连接的情况。平面情况下,路由接受率为 100%。

5.3 仿真环境

仿真实验在 Fedora Core 8 操作系统及 NS 2.33 网络仿真平台^[12]下完成。仿真实验所用拓扑由 GT-ITM 拓扑生成器^① N-level hierarchical graph 模型随机生成。所得二层网络拓扑图满足 PNNI 网络分层模型,且上层和下层的域内构造都遵循 Waxman 模型(如下面公式),公式中具体参数含义见文献^[13]:

① GT-ITM: Georgia Tech Internetwork Topology Models.
http://www.cc.gatech.edu/projects/gtitm/

$$p(u,v) \propto \alpha e^{-\frac{d(u,v)}{\beta L}}, \alpha > 0, \beta \leq 1 \tag{6}$$

实验中,下层域的 α 和 β 设置为 0.6 和 0.6,上层域的 α 和 β 设置为 0.9 和 0.1. 下层域内的链路,设置其时延即为该链路的长度(0~15 单位之间),设置其带宽为 100~1000 的随机数. 上层域内的链路,设置其时延为 1 层域内链路时延的 2 倍,并设置其带宽为 1000~5000 的随机数. 满足以上约束的一幅随机拓扑图如图 5 所示.

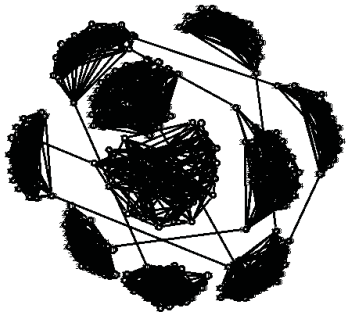


图 5 随机二层拓扑图

5.4 结果分析

在信息压缩方面,在网络规模为“10 个域×20 个节点/域”的网络拓扑中,在完成“10 组×100 加入请求次/组”运算后,统计不同模式下节点维护和交换的拓扑信息量和组播树状态信息量,如表 1 和表 2 所示,数据为 5 幅随机拓扑图的平均结果. 表 1 中,

表 1 不同模式下节点维护的信息量比较

| | 所有节点存储的 拓扑信息总数 | 每个节点存储的 拓扑信息数目 | 所有节点存储的 组播树状态信息总数 | 每个节点存储的 组播树状态信息数目 |
|----------|-------------------|-------------------|----------------------|----------------------|
| 平面 | 273600 | 1368 | 122743.2 | 613.716 |
| 分层不启动伪节点 | 52516.2 | 262.581 | 32981.2 | 164.906 |
| 分层启动伪节点 | 52516.2 | 262.581 | 50594.4 | 252.972 |

表 2 不同模式下节点间交互的信息量比较

| | 所有节点间交换的 拓扑信息分组总数 | 每个节点间交换的 拓扑信息分组数目 | 所有节点间交换的 组播树状态信息分组总数 | 每个节点间交换的 组播树状态信息分组数目 |
|----------|----------------------|----------------------|-------------------------|-------------------------|
| 平面 | 1357602 | 6788.01 | 605311.6 | 3026.558 |
| 分层不启动伪节点 | 812276.4 | 4061.382 | 164789.6 | 823.948 |
| 分层启动伪节点 | 926082.8 | 4630.414 | 487243.6 | 2436.218 |

在路由性能方面,由于平面模式基于完整、详细的信息进行路由计算,因此其路由性能最好,同时我们考察二层模式以及二层伪节点模式取得可扩展性的同时,其路由性能是否受到影响. 图 6(a)~(d)显示了 3 种模式下路由成功率分别随用户带宽请求平均值、用户时延请求平均值、启动的组应用数以及每组发起加入的组成员数的变化情况. 图 6(a)和(b)的其它测试参数为:每个拓扑图随机启动 10 个组应用,针对每个组应用随机选择不重复的 100 节点(不

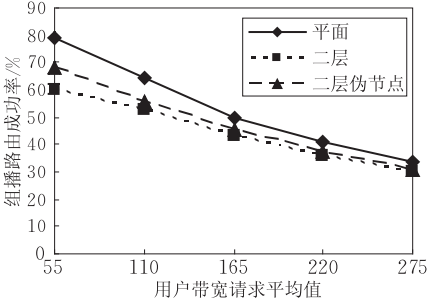
分层情况下节点维护的拓扑状态信息数量约是不分层情况下的 1/5,节点维护的拓扑信息数量大大减少. 二层和二层伪节点模式维护的拓扑状态信息数量一致,是因为两者基于相同的网络层次划分、相同的统计时间点且网络拓扑在算法运行期间未发生变化. 而由于 DBHMR 算法引入伪节点功能以及实现组播树状态信息聚集,分层情况下节点维护的组播树状态信息数目也大大压缩. 分层不启动伪节点约是不分层情况下的 1/4,而分层启动伪节点则稍多,是不分层情况下的 1/2 到 1/3,原因是伪节点在一定程度上协助域内树上节点进行组播树状态信息扩散,增加了组播树状态信息数量. 表 2 显示的是节点间交互的信息量:就拓扑信息分组数目而言,二层和二层伪节点模式约是平面模式的 6/10 和 7/10,节点间交换的拓扑信息量得到压缩,同时二层和二层伪节点模式两者交换的拓扑信息分组数目稍有不同,原因是两者在加入请求结果方面存在不同,因此两者在运行过程中网络的负载变化也将不同,触发的拓扑更新次数也将不同;而就交互的 MTSI 分组数目而言,二层模式约是平面模式的 1/4,而二层伪节点模式约是平面模式的 4/5,节点间交换的 MTSI 数量减少,同时它与三个模式的加入请求结果相关,加入请求成功率越高,组播树状态信息的更新将更加频繁,节点间交换的组播树状态信息数量将越大.

同于本组的源节点)执行加入请求. 图 6(c)和(d)的其它测试参数为:用户时延请求随机分布在 20~200 单位之间;用户带宽请求随机分布在 30~300 单位之间. 其它测试参数为:所有组应用支持的带宽随机分布在 100~5000 单位之间;所有数据均为 5 幅随机拓扑图下的平均值. 图 6 显示平面模式的路由成功率为最好,网络分层后(即二层模式)路由成功率有所下降,但引入“伪节点功能”后,由于域内部组播树状态信息得到聚集和扩散,因此二层伪节

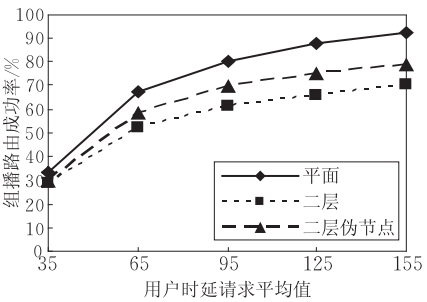
点模式下路由成功率又得到提高,介于平面模式与二层之间.同时路由成功率随各参数的变化而变化,如图 6(a)中当用户带宽请求逐渐增大,网络中的可利用的带宽资源越来越少,因此路由成功率不断降低;图 6(c)和(d)中当组应用数或每组加入成员数越来越多,网络负载越来越重,路由成功率也呈下降趋势.

图 7(a)~(d)显示了 3 种模式下路由接受率随各参数的变化情况.图 7 的测试场景与图 6 相同.

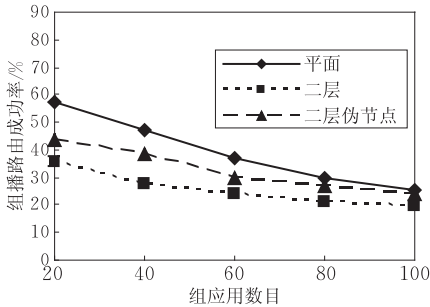
图 7 显示:二层模式和二层伪节点模式的路由接受率都保持在 99.5%以上,与平面情况下的路由接受率 100%相差不大,发生路由回退的概率较小.同时,二层伪节点模式的路由接受率要稍高于二层模式,这是由于伪节点的组播树状态信息在一定程度上体现了域内组播节点嫁接能力,组播树状态信息在更大范围内扩散,有助于路由成功率和路由接受率的提高.



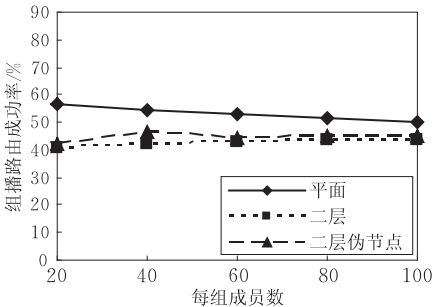
(a) 路由成功率随用户带宽请求的变化(用户时延请求平均值: 110)



(b) 路由成功率随用户时延请求的变化(用户带宽请求平均值: 55)

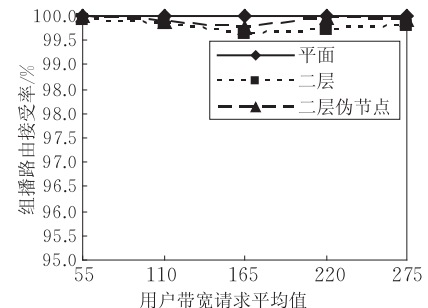


(c) 路由成功率随组数的变化(每组成员数: 10)

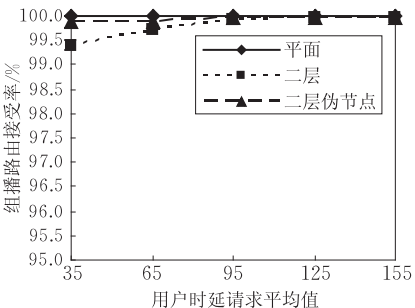


(d) 路由成功率随组成员数的变化(组应用数: 10)

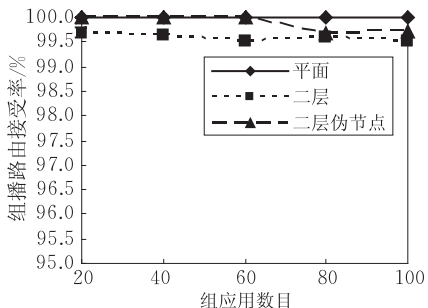
图 6 不同模式下路由成功率的比较



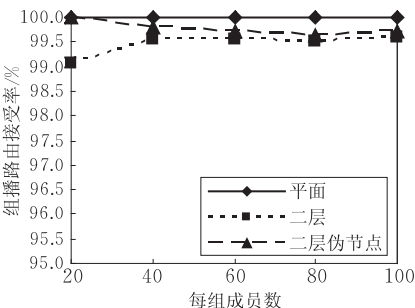
(a) 路由接受率随用户带宽请求的变化(用户时延请求平均值: 110)



(b) 路由接受率随用户时延请求的变化(用户带宽请求平均值: 55)



(c) 路由接受率随组数的变化(每组成员数: 10)



(d) 路由接受率随组成员数的变化(组应用数: 10)

图 7 不同模式下路由接受率的比较

总的来说,在 Stair 拓扑聚集算法、伪节点聚集域内组播树状态信息功能和 DBHMR 算法的共同作用下,分层模式与平面模式相比,拓扑信息存储压缩率达到 5~6 倍,同时组播树状态信息存储压缩率达到 2~4 倍,而路由成功率和路由接受率并没有明显下降.在引入 PTBN 功能后,节点扩散的和存储的组播树状态信息稍有增多,但其路由成功率和路由接受率又有提高,因此二层伪节点模式是平面模式和二层模式两者之间的一个较好的平衡.

6 结 论

针对大规模网络下 QoS 路由的可扩展性问题,在已设计支持时延-带宽约束的 Stair 拓扑聚集算法的基础上,本文对 PNNI 层次网络模型下的动态组播路由问题进行了全面研究,包括:对组播树状态信息进行定义,并提出了聚集域内部组播树状态信息的伪树上边界节点,实现了组播树状态信息的有效压缩;进一步设计了支持组成员动态加入和退出的 DBHMR 层次组播路由算法,成功建立层次化组播动态树.仿真实验表明,Stair 聚集算法获得了较高的拓扑信息压缩,PTBN 聚集模式实现了组播树状态信息的有效压缩,同时层次路由保持了和平面网络近似的路由成功率和路由接受率,实现了大规模网络情况下组播路由的扩展.

参 考 文 献

- [1] Lu Hui-Mei, Xiang Yong, Shi Mei-Lin. QoS-based hierarchical multicast routing. *Chinese Journal of Computers*, 2004, 27(6): 772-781(in Chinese)
(陆慧梅,向勇,史美林.支持 QoS 的层次组播路由算法框架 QHMR. *计算机学报*, 2004, 27(6): 772-781)
- [2] Lu Hui-Mei, Xiang Yong, Cao Yuan-Da. Heterogeneous bandwidth-constrained dynamic hierarchical multicast routing. *Chinese Journal of Computers*, 2006, 29(6): 898-905 (in Chinese)
(陆慧梅,向勇,曹元大.异构带宽约束的动态层次组播路由. *计算机学报*, 2006, 29(6): 898-905)
- [3] Lu Hui-Mei, Hu Hong-Yu, Xiang Yong, Yan Yong-Hang. Delay-bandwidth constrained topology aggregation algorithm. *Journal on Communications*, 2007, 28(3): 93-99(in Chinese)
(陆慧梅,胡虹雨,向勇,闫永航.基于时延-带宽约束的新型层次拓扑聚集算法. *通信学报*, 2007, 28(3): 93-99)
- [4] Korkmaz Turgay, Krunk Marwan. Source-oriented topology aggregation with multiple QoS parameters in hierarchical ATM networks. *ACM Transactions on Modeling and Computer Simulation*, 2000, 10(4): 295-325
- [5] Lui King-Shan, Nahrstedt Klara, Chen Shi-Gang. Routing with topology aggregation in delay-bandwidth sensitive networks. *IEEE/ACM Transaction on Networking*, 2004, 12(1): 17-29
- [6] Ben Ali N, Belghith A, Moulhierac J, Molnar M. QoS multicast aggregation under multiple additive constraints. *Computer Communication*, 2008, 31(15): 3564-3578
- [7] Ben Ali Naouel, Moulhierac Joanna, Belghith Abdelfettah, Molnar Miklos. mQMA: Multi-constrained QoS multicast aggregation//*Proceedings of the IEEE GLOBECOM*. Washington, 2007: 1927-1932
- [8] Xiang Yong. Research on group communication for computer supported cooperative work [Ph. D. dissertation]. Department of Computer Science and Technology, Tsinghua University, Beijing, 1998(in Chinese)
(向勇.计算机协同工作中的群组通信研究[博士学位论文].清华大学计算机科学与技术系,北京,1998)
- [9] The ATM Forum. Private Network-Network Interface Specification Version 1.0. March, 1996
- [10] Lu Hui-Mei. Research on scalable QoS-based multicast routing [Ph. D. dissertation]. Department of Computer Science and Technology, Tsinghua University, 2004(in Chinese)
(陆慧梅.可扩展的 QoS 组播路由研究[博士学位论文].清华大学计算机科学与技术系,北京,2004)
- [11] Wang Zheng, Crowcroft Jon. Bandwidth-delay based routing algorithms//*Proceedings of the IEEE GLOBECOM*. Singapore, 1995, 3: 2129-2133
- [12] Xu Lei-Ming, Pang Bo, Zhao Yao. NS and Network Simulation. Beijing: Posts & Telecom Press, 2003(in Chinese)
(徐雷鸣,庞博,赵耀. NS 与网络模拟.北京:人民邮电出版社,2003)
- [13] Waxman B M. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 1988, 6(9): 1617-1622



HU Hong-Yu, born in 1976, Ph. D. candidate. Her current research interests include QoS routing and IP multicast.

LU Hui-Mei, born in 1974, Ph. D., associate professor. Her current research interests include computer network architecture, QoS-based multicast routing and application layer multicast.

CAO Yuan-Da, born in 1944, professor, Ph. D. supervisor. His major research interests include artificial intelligence, computer network and network security.

RAO Si-Min, born in 1986, M. S. candidate. Her current research interests include computer network and distrib-

uted computing.

Background

With the fast development of hardware technologies, as well as with the increasing demands for quality of services (QoS) fueled by continuously emerging multimedia applications, offering guaranteed and better services than best effort services becomes an important and challenging issue for the next generation networks, among which QoS-based unicast/multicast routing is the core and indispensable technology. Although many QoS-based multicast routing schemes have been proposed, most of them are not practical in the Internet environment because many issues are involved in them, such as, excessive computation overhead, large amount of the knowledge of the global network state, only one QoS parameter considered, no support to the dynamic group membership and the heterogeneous QoS requirements imposed by group members. This paper conducted a comprehensive study of all of these problems, which get support by the National Natural Science Foundation of China (Nos.60503050, 60773045), Basic Research Foundation of Beijing Institute of Technology (No.200511F4212) and Key Discipline Con-

struction Project of Beijing.

The authors have proposed several QoS multicast routing algorithms to solve the scalability problem; QHMR^[1] and HBHMR^[2]. QHMR focus on the hierarchical multicast routing problem with only one QoS parameter constraint (bandwidth) based on HMR hierarchical network model. HBHMR resolved the dynamic hierarchical multicast routing problem with the heterogeneous bandwidth requirements imposed by different group members which is as same as based on HMR model. Further more, the authors deeply discussed the aggregation problem of the topology information of a domain and Stair method was proposed in 2007. With the basis of their previous work^[1-3], this paper conducted a detailed study on the aggregation problem of the multicast tree state information of a domain with more QoS parameter constraints (delay and bandwidth) and the dynamic multicast routing using the aggregated topology information and multicast tree state information based on PNNI model which is a more popular hierarchical network model.