

文章编号: 1001-0920(2006)04-0381-04

基于遗传算法的应用层多播路由方案

程 鹏, 吴秋峰, 戴琼海
(清华大学 自动化系, 北京 100084)

摘 要: 通过分析应用层多播树的路由约束, 定义了 3 个适值函数, 分别考察应用层多播树在开销、平衡和网络层业务量均衡 3 方面的性能, 并根据生成的总适值函数进行遗传算法操作. 仿真结果表明, 与基于几何规则的应用层路由算法比较, 本文提出的算法对多播树的 3 方面性能都有所优化, 树的平衡和网络层业务量均衡性能的优化程度尤为明显.

关键词: 多播; 应用层多播; 多播路由; 遗传算法

中图分类号: TP18 **文献标识码:** A

Application Layer Multicast Routing Solution Based on Genetic Algorithms

CHEN G Peng, WU Qiu-feng, DA I Qiong-hai

(Department of Automation, Tsinghua University, Beijing 100084, China Correspondent: CHENG Peng, E-mail: chengpeng00@mails.tsinghua.edu.cn)

Abstract Application Layer Multicast routing is a multi-objective optimization problem. Three routing constraints, tree's cost, tree's balance and network layer load distribution, are analyzed. Three fitness functions are used to evaluate a multicast tree on the three indexes respectively and one overall fitness function is generated. A novel approach based on genetic algorithms is proposed. Numerical simulations show that compared with geometrical routing rules, the proposed algorithm improves all three indexes, especially on cost and network layer load distribution indexes.

Key words: Multicast; Application layer multicast; Genetic algorithms; Multicast routing

1 引 言

多播通信协议将数据源发出的信息同时发送到多个用户. 网络层多播协议在路由树的分权处进行信息复制, 需要网络层设备的支持, 在 Internet 上无法广泛应用. 近年来, 多播中的信息复制与转发机制逐渐被放置到网络终端上, 由应用层程序发起多播通信并维护多播组, 这样的机制称为应用层多播^[1,2]. 应用层多播不需要网络层设备和协议的支持, 在任何一对节点之间的数据通路都是普通的单播链路. 应用层多播协议将多播组的成员组织到两个拓扑中: 第一个拓扑是控制拓扑, 由多播成员之间

形成网状的连接, 以保持组的完整和组成员信息的及时更新; 另一个拓扑是数据拓扑, 是多播数据在组成员之间传输的通路, 一般是控制拓扑的树状子集, 称为应用层多播树, 由应用层多播路由算法在控制拓扑中建立.

无论是网络层多播还是应用层多播, 路由算法的目标都是形成一棵或一组满足一定性能要求的多播树^[3~11]. 目前的多播路由指标可分为两类: 树的开销与树的平衡. 优化前者的目的在于提高网络资源的使用效率, 减少多播通信的时延; 优化后者的目的在于避免树中部分节点因度数过大而性能下降, 导致多播通信性能下降甚至崩溃.

收稿日期: 2005-03-07; 修回日期: 2005-05-12

基金项目: 国家自然科学基金重点项目 (60432030).

作者简介: 程鹏 (1979—), 男, 山西阳泉人, 博士生, 从事多播自适应控制与复杂媒体网络的研究; 吴秋峰 (1936—), 男, 江苏宜兴人, 教授, 博士生导师, 从事企业网络、集成制造等研究.

上述两类指标可以较好地满足网络层多播的需求,但是不能全面体现应用层多播组的需求。应用层多播树的节点平衡只能优化多播业务在不同终端节点之间的分配,无法反映网络层业务量的均衡状况。这可能导致多播业务在底层网络的分布出现非常不均匀的状况,严重影响多播通信的效率和性能。因而,网络层业务的均衡程度也是衡量应用层多播路由方案性能的重要指标。

由于优化目标较多,目前的应用层多播协议大多放弃了对多播树进行优化的努力,而是以某些简单的基于几何的规则来建立一棵可用的多播树,且不考虑底层网络性能^[1,2,6]。遗传算法则可用于多目标约束下的路由。文献[3]使用遗传算法建立网络层多播树。[5]使用遗传算法优化了基于全连接控制拓扑的应用层多播树的开销和平衡性能。[7~10]利用遗传算法解决了在时延、带宽和树开销等服务质量约束下的网络层多播路由问题。

本文将网络层业务的均衡程度纳入路由算法优化目标,通过分析应用层多播树在开销、平衡和网络层业务量的均衡性能 3 方面的优劣程度,构造了描述多播树整体优劣程度的综合适值函数,提出了一个基于遗传算法的应用层多播路由方案,以获得同时满足这 3 方面约束的应用层多播树。

2 网络模型与问题提出

2.1 网络模型

在应用层多播中,终端自组织成为一个连通的控制拓扑,记为有向图 $M(H, P, c)$ 。控制拓扑由一个包含 $|H|$ 个终端的非空集合 H 和一个包含 $|P|$ 条单播路径的非空集合 P 构成, $P \subseteq H \times H$ 。每条单播路径的开销由函数 c 确定, $c: P \rightarrow R^+$ 。这些单播路径称作元路径,最终组成应用层多播树的路径是 P 的子集。集合 H 中包含数据源 s 和接收终端集合 $D, H = s \cup D$ 。当 H 中两两终端之间都有单播路径存在时,控制拓扑 M 是一个全连接图。

进一步考虑到底层网络的拓扑结构,应用层多播网络可以归纳为一个有向图 $G(H, N, L, E, c)$ 。集合 N 由 $|N|$ 个路由器构成,集合 E 由 $|E|$ 条连接路由器的链路构成,

$$E = \{e_i | i \in N, i \neq j \in N, e_i \in E\}, E \subseteq N \times N.$$

集合 L 由终端与路由器之间的链路组成,

$$L = \{l_j | j \in N, j \neq s, l_j \in L\}, |L| = |H|$$

E 和 L 中的所有链接和链路都由函数 c 确定其开销, $c: E \rightarrow R^+, c: L \rightarrow R^+$ 。 M 中的每条元路径都可以写成由 G 中的若干个链路和两个链接组成的一个序列。元路径的开销等于序列中所有链路和链接开销的总和。

2.2 应用层多播路由

应用层多播路由算法的目标是建立一棵覆盖所有终端的树 $T \subseteq M$, 即给定 G 和 M , 求 M 的扩展树 $T(H, F), F \subseteq P$ 。树应该在以下 3 方面的约束下建立:

(1) 开销约束

树 T 的开销记作 $C(T)$, 是组成树的所有元路径的开销之和。开销约束路由的目标是使得最终输出的多播树的开销 $C(T)$ 在所有可能的多播树中为最小。

(2) 平衡约束

对于一个终端而言,它的度数,即它在应用层多播树中的子节点的数量,要低于某个限定值,否则终端所具有的计算资源和网络资源将无法复制分发多播信息的需求,从而造成多播通信的性能下降。用 $d_{\max}(h)$ 描述终端 h 的度数上限,定义树 T 中节点 h 的剩余度数为

$$\text{res}_T(h) = d_{\max}(h) - d_T(h),$$

其中 $d_T(h)$ 为 h 在 T 中的度数。文献[4]中将平衡路由问题定义为:给定 G 和 M , 求扩展树 T , 满足:

$$1) d_T(h) \leq d_{\max}(h), \forall h \in H;$$

2) 所有节点的剩余度数中的最小值在所有树中最大。

这是个最大最小优化问题,有一定的计算复杂度^[5]。为了简化优化问题,构建变量

$$\text{rat}_{d_T}(h) = d_T(h) / d_{\max}(h),$$

上面提到的问题可以写成:给定 G 和 M , 求扩展树 T , 满足:

$$1) \text{rat}_{d_T}(h) \leq 1, \forall h \in H;$$

2) 树中所有节点的 $\text{rat}_{d_T}(h)$ 中的最大值在所有树中最大。

(3) 网络层业务量均衡约束

这一约束旨在限制多播树使用同一条物理链路的次数。与平衡约束路由类似,分别定义 $s_{\max}(e)$ 和 $s_T(e)$ 为树 T 中链路 e 被重复使用的次数上限和实际被重复使用的次数,构建变量

$$\text{rats}_T(e) = s_T(e) / s_{\max}(e).$$

网络层业务量均衡约束路由问题可以写成:给定 G 和 M , 求满足下列条件的扩展树 T :

$$1) \text{rats}_T(e) \leq 1, \forall e \in E;$$

2) 所有链路的 $\text{rats}_T(e)$ 中的最大值在所有树中最大。

3 基于遗传算法的应用层多播路由方案

本节具体描述了基于遗传算法的应用层多播路由方案的操作过程。

Step 1: 建立控制拓扑。给定 H, R, L, E, c, s , 每

个终端都随机选择若干其他终端作为自己在控制拓扑中的邻居, 并建立自己与所有应用层邻居之间的单播路径作为元路径。确保在任意一对终端之间都有路径存在。计算出所有元路径的开销与经过的底层链路序列。

Step2: 建立路由库。为 D 中每一个成员分别建立一个包含 R 条可选路由的路由库, 每条可选路由都由一条或若干条元路径组成, 这些可选路由按照其开销从小到大排列。

Step3: 染色体初始化。遗传算法中的每个染色体对应于一棵应用层多播树。对于接收终端集合 $D = \{h_1, h_2, \dots, h_k\}$, $k = |H| - 1$, 染色体由长度为 k 的整数序列表示。染色体中第 i 位基因的值 g_i 表示在这棵多播树中从 s 到 h_i 选用的路由在从 s 到 h_i 路由库中的位置。染色体的初始化过程随机地产生 P 个不同的染色体作为第零代初始种群。

Step4: 初始种群染色体适值计算。分别采用下列3个函数评估染色体在多播树开销、多播树平衡和网络层业务均衡性能方面的优化程度:

$$f_1 = 1 - C(T)/C(M),$$

$$f_2 = \begin{cases} \max_h \text{ratd}_T(h), & \text{if } \max_h \text{ratd}_T(h) < 1; \\ 0, & \text{otherwise} \end{cases}$$

$$f_3 = \begin{cases} \max_e \text{rats}_T(e), & \text{if } \max_e \text{rats}_T(e) < 1; \\ 0, & \text{otherwise} \end{cases}$$

其中 $C(M)$ 是 M 中所有元路径的开销之和。

染色体的总适值函数为

$$f = \lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3$$

Step5: 遗传进化操作。下一代种群通过选择、交叉和变异去除重复染色体3个阶段获得。

首先将当前的所有染色体根据适值从大到小排列, 适值较小的一半染色体被淘汰, 适值较大的一半染色体被选择, 拷贝到下一代, 成为下一代种群的成员, 适值最大的即为本代的输出。被选择的染色体通过两两交叉和变异操作, 得到另一半构成下一代种群的染色体。变异操作中每个基因发生变异的概率为 p , 基因变异后的值在 1 到 R 间随机选取。

如果经过选择、交叉和变异操作后, 新一代种群中出现了重复染色体, 需要生成新的随机染色体替代多余的重复染色体。

Step6: 判断是否已达到结束条件。当遗传操作的最大代数已经达到, 或目前算法的输出在过去的若干代内未改变时, 结束种群进化。当前种群中适值最大的染色体即为输出的路由方案。

4 仿真实验

仿真实验采用的拓扑由 GT-ITM^[12] 工具生成, 分布在 100×100 平面上。底层网络中共有 1 个主干

子网, 8 个主干路由器, 8 个接入子网, 32 个接入路由器, 终端数量为 20 个。每个终端都通过一个 4×4 的局域网与一个接入路由器相连。两个路由器之间的链路开销设置为两者之间的距离, 终端与路由器之间的链路开销为 1。两两终端之间元路径存在的概率为 0.4。遗传算法种群规模为 10。所有终端的度数上限为 4。主干网路由器之间的链路被使用次数上限设为 10, 接入网路由器之间的链路被使用次数上限设为 4, 主干网路由器与接入网路由器之间的链路被使用次数上限设为 8。 λ_1 , λ_2 和 λ_3 都设为 1。算法执行的最大代数为 500 代。在同一个控制拓扑中, 分别采用 CR 算法、ALM I 算法和本文提出的算法得到多播树, 以进行对比。

4.1 路由库规模 R 与变异概率 p 对算法的影响

仿真在一个具有 126 条元路径的控制拓扑上进行。在当前阶段的仿真中, 数据源确定为编号 1 的终端。 R 的取值分别为 5, 10, 15, 20, p 则从 0.1 到 0.9 变化。针对每组 R, p 参数进行 10 次计算。经过 500 代遗传操作, 所有的 360 个输出染色体的平衡适值和网络层业务量均衡适值都为 1, 综合适值的差异主要来自于开销适值。图 1 给出了每组 R, p 参数下输出适值的平均值。相同变异概率 p 下, 当 R 从 5 开始增大时, 算法的输出适值逐渐减小; 相同路由库规模 R 下, 当变异概率 p 增大时, 算法的输出适值总体上减小。当变异概率 p 处于 0.1 到 0.5 之间的时候, 树的开销得到了较好的优化。

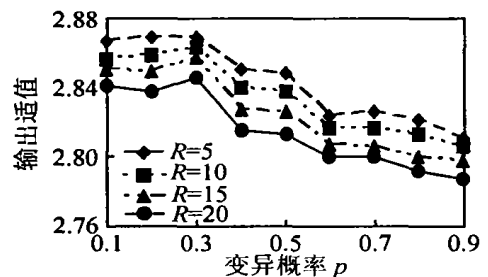


图1 R, p 对算法输出的影响

图 2 和图 3 描述了在路由库规模 R 分别为 5, 10, 15, 20, 变异概率 p 分别为 0.3 和 0.5 时, 路由算法输出适值的演化过程。可见, 当 p 为 0.3 时, 算法的输出方案适值高, 但是收敛较慢, 在 500 代时仍有较明显的提高趋势; 当 p 为 0.5 时, 算法的收敛速度较快, 200 代之后输出的适值变化已经很小, 但是最终输出的适值较小。

此拓扑中采用 CR 算法获得的以编号为 1 的终端为根的源树在开销、平衡和网络层业务均衡方面的适值分别是 0.834, 1 和 0; 采用 ALM I 协议得到的多播树在开销、平衡和网络层业务量均衡方面的适

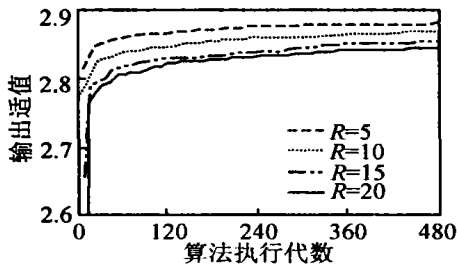


图2 p 为 0.3 时算法的输出

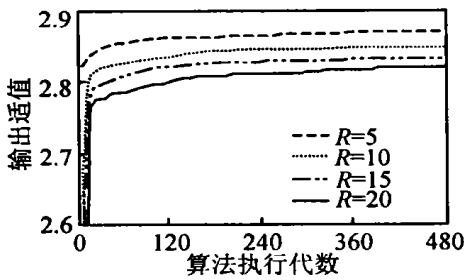


图3 p 为 0.5 时算法的输出

值分别为 0.794, 0.600 和 0.750。将 R 设为 5, 令 p 从 0.1 到 0.9 逐渐增大, 在每个 p 值下进行 20 次仿真。所得到的 180 棵多播树在平衡和网络层业务量均衡方面的适值都是 1, 只有 5 棵树 (2.78%) 的开销适值低于 CR 树。图 4 给出了这些树输出的总适值和开销适值的 90% 置信区间和均值。可见本文提出的算法改善了树的平衡性能和底层网络的均衡性能, 同时, 与 CR 算法和 ALM I 协议相比, 本文提出的算法也改善了树的开销。

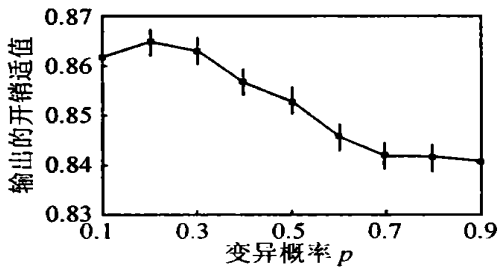


图4 $R = 5$ 时算法输出开销适值的 90% 置信区间与平均值

4.2 与其他路由协议的对比

当前阶段仿真中 R 设为 5, p 设为 0.2。将所有的终端逐个作为数据源, 分别进行 10 次路由。在得到的 200 棵多播树中, 只有 4 棵树 (2.00%) 的平衡适值不等于 1, 16 棵树 (8%) 的网络层业务量均衡适值不等于 1。相同数据源采用本文提出的算法所获得的 20 棵树在 3 方面的适值都高于采用 CR 构建的树。

所有输出树在平衡和网络层业务量均衡两方面的平均适值分别为 0.999 和 0.925, 而采用 CR 构建的树在这两方面的平均适值为 0.752 5 和 0, 这意味

着所有的 CR 树都不满足网络层业务量均衡约束。CR 树的平均开销适值为 0.847 3, 而通过本文提出的算法获得的树的平均开销适值为 0.882 6, 优化幅度达到 23%。表 1 给出了本文中提出的算法获得的树与通过 CR 算法和 ALM I 协议获得的树的对比。

表1 本文算法与 CR 算法和 ALM I 协议的对比

路由算法	开销适值	平衡适值	网络层业务量均衡适值
本文的算法	0.882 6	0.999 0	0.925 0
CR	0.847 3	0.752 5	0
ALM I	0.794 0	0.600 0	0.750 0

5 结 论

应用层多播的路由算法不仅需要考虑到树的开销、树的平衡等应用层性能, 还需要考虑到网络层业务量的均衡性能, 以避免底层物理链路成为通信的瓶颈。

本文将网络层业务量的均衡性能纳入路由方案的性能指标中, 将应用层多播路由归纳为多约束下的优化问题, 通过分析各个路由约束对树的影响, 构建了描述多播树在某一指标下优劣程度的适值函数, 并在此基础上提出了一种基于遗传算法的应用层多播路由方案。仿真试验表明, 与 CR 算法和 ALM I 协议相比, 本文提出的方案在保证树的平衡和网络层业务量均衡性能的同时, 保持了较小的开销, 从而改善了应用层多播树的综合性能。

本文提出的方案有一定的通用性, 可以应用于任意的控制拓扑组织协议, 还可以通过适值函数和权值的修正、更改、组合, 实现不同指标组合下的应用层多播路由。

参考文献 (References)

- [1] Liebeherr J, Nahas M, SiW S. Application-layer Multicasting with Delaunay Triangulation Overlays [J]. *IEEE J on Selected Areas in Communications*, 2002, 20(8): 1472-1488.
- [2] Pendarakis D, Shi S, Vema D, et al. ALM I: An Application Level Multicast Infrastructure [A]. *Proc of 3rd Usenix Symposium on Internet Technologies and Systems* [C]. San Francisco, 2001.
- [3] Hwang R H, Do W Y, Yang S C. Multicast Routing Based on Genetic Algorithms [J]. *J of Information Science and Engineering*, 2000, 16(6): 885-901.
- [4] Shi S Y, Turner J S. Routing in Overlay Multicast Networks [A]. *Proc of IEEE Infocam* [C]. New York, 2002.
- [5] Pan Y, Yu Z W, Wang L C. A Genetic Algorithm for the Overlay Multicast Routing Problem [A]. *Proc of ICCNM C'03* [C]. Shanghai, 2003.

(下转第 390 页)

建立更为准确和合理的评价决策方法具有重要意义

本文通过证据冲突在已知命题和未知命题之间的合理分配,提出了一种处理多属性决策问题的新ER算法,并证明了新方法完全满足4个合成公理通过运算比较,进一步验证了新方法的有效性和合理性

参考文献(References)

- [1] Huynh V N, Nakamori Y. A Satisfactory-oriented Approach to Multi-expert Decisionmaking under Linguistic Assessments [J]. *IEEE Trans on Systems, Man, and Cybernetics — Part B: Cybernetics*, 2005, 35(2): 184-196
- [2] Yang J B, Xu D L. On the Evidential Reasoning Algorithm for Multiple Attribute Decision Analysis under Uncertainty [J]. *IEEE Trans on Systems, Man, and Cybernetics — Part A: Systems and Humans*, 2002, 32(3): 289-304
- [3] Van Nam Huynh, Yoshiteru Nakamori, Tu Bao Ho, et al. Assessment Aggregation in the Evidential Reasoning Approach to MADM Under Uncertainty: Orthogonal Versus Weighted Sum [A]. *Advances in Computer Science — ASIAN 2004: Higher-level Decision Making* [C]. Thailand, 2004: 109-127,
- [4] 张山鹰, 潘泉, 张洪才. 一种新的证据推理组合规则[J]. *控制与决策*, 2000, 15(5): 540-544
(Zhang S Y, Pan Q, Zhang H C. A New Kind of Combination Rule of Evidence Theory [J]. *Control and Decision*, 2000, 15(5): 540-544)
- [5] Yager R R. On the Dempster-shafer Framework and New Combination Rules [J]. *Information System*, 1989, 41(2): 93-137.
- [6] 李弼程, 王波, 魏俊. 一种有效的证据理论合成公式[J]. *数据采集与处理*, 2002, 17(1): 33-36
(Li B C, Wang B, Wei J. An Efficient Combination Rule of Evidence Theory [J]. *J of Data Acquisition and Processing*, 2002, 17(1): 33-36)
- [7] 邢清华, 雷英杰, 刘付显. 一种按比例分配冲突度的证据推理组合规则[J]. *控制与决策*, 2004, 19(12): 1387-1390
(Xing Q H, Lei Y J, Liu F X. One Combination Rule of Evidence Theory Based on Distributing Confliction Proportion [J]. *Control and Decision*, 2004, 19(12): 1387-1390)
- [8] 林作铨, 牟克典, 韩庆. 基于未知扰动的冲突证据合成方法[J]. *软件学报*, 2004, 15(8): 1150-1156
(Lin Z Q, Mu K D, Han Q. An Approach to Combination of Conflicting Evidences by Disturbance of Ignorance [J]. *J of Software*, 2004, 15(8): 1150-1156)
- [9] Smets P, Kennes R. The Transferable Belief Model [J]. *Artificial Intelligence*, 1994, 66(2): 191-234
- [10] Li M, Yang J B. A Decision Model for Self-assessment of Business Process Based on the EFQM Excellence Model [J]. *Int J of Quality and Reliability Management*, 2003, 20(2): 164-188
- [6] Kranakis E, Singh H, Urrutia J. Compass Routing on Geometric Networks [A]. *Proc of 11th Canada Conf of Computational Geometry* [C]. Vancouver, 1999
- [7] Hamdan M, El-hawary M E. A Novel Genetic Algorithm Searching Approach for Dynamic Constrained Multicast Routing [A]. *Proc of CCECE 2003* [C]. Montreal, 2003
- [8] Haghghat A T, Faez K, Dehghan M, et al. GA-based Heuristic Algorithms for Bandwidth-delay-constrained Least-cost Multicast Routing [J]. *Computer Communications*, 2004, 27(1): 111-127.
- [9] Wang Z Y, Shi B X, Zhao E. Bandwidth - delay - constrained Least-cost Multicast Routing Based on Heuristic Genetic [J]. *Computer Communications*, 2001, 24(7-8): 685-692
- [10] Sun Q, Langendoerfer H. Computation of Constrained Multicast Trees Using a Genetic Algorithm [J]. *European Trans on Telecommunications*, 1999, 10(5): 513-516
- [11] Hamdan M, El-hawary M E. Hopfield-Genetic Approach for Solving the Routing Problem in Computer Networks [A]. *Proc of the 2002 IEEE Canadian Conf on Electrical and Computer Engineering* [C]. Winnipeg, 2002
- [12] Zegura E W, Calvert K, Bhattacharjee S. How to Model an Internetwork [A]. *Proc of IEEE Infocam '96* [C]. San Francisco, 1996: 594-602

(上接第384页)