

覆盖网络的链路优化问题研究

张艳梅

(中央财经大学信息学院, 北京 100081)

摘 要: 在综合考虑各个影响覆盖网络路由性能因素的基础上, 提出针对覆盖链路选取问题(OLSP)的多目标优化模型, 利用覆盖链路对物理链路的重用度因素使模型能以最少的维护代价获得最高的覆盖路由性能。采用 Pareto 占优理论和免疫克隆算法对 OLSP 进行求解, 通过数字仿真实验验证该模型的合理性和 ICA 的有效性。

关键词: 覆盖链路选取问题; 覆盖网络; 链路选取; 拓扑优化; 多目标优化; 免疫克隆算法

Research on Link Optimization Problem of Overlay Network

ZHANG Yan-mei

(School of Information, Central University of Finance and Economics, Beijing 100081, China)

【Abstract】 A multi-goal optimization model for Overlay Link Selection Problem(OLSP) is proposed based on the consideration of all factors influencing on Overlay network performance. It aims at achieving the best routing performance and the least IP link overlap degree with the least maintaining costs. Pareto optimality theory and Immune Clone Algorithm(ICA) are adopted to solve this problem and the digital simulation experimental result shows the availability of this model and the convergence of ICA.

【Key words】 Overlay Link Selection Problem(OLSP); Overlay network; link selection; topology optimization; multi-goal optimization; Immune Clone Algorithm(ICA)

DOI: 10.3969/j.issn.1000-3428.2011.02.087

1 概述

目前, 关于覆盖网络构造和优化的研究大多关注覆盖节点的部署问题^[1-2], 很少考虑覆盖链路选取问题(Overlay Link Selection Problem, OLSP)。对于一个有 n 个节点的覆盖网络, 存在 $\frac{n(n-1)}{2}$ 条覆盖链路。如果采用全连通的拓扑结构, 则每个节点需要维护 $(n-1)$ 个节点的状态信息, 随着网络中节点数的增加, 这将给覆盖节点造成严重的负担。而且有些节点之间的直接连通路由比间接连通路由的代价更大。许多研究已经证明, Overlay 网络中覆盖链路的选择对路由服务的性能影响很大, 因此, 在给定覆盖拓扑维护开销的情况下, 如何选择覆盖链路对构造高性能的覆盖网络拓扑结构具有非常重要的意义。

在覆盖网络中, 2 个覆盖节点通过覆盖链路相连接, 实际上覆盖链路重用了底层的 IP 链路, 每个覆盖路由由一个或多个 IP 链路组成。2 个覆盖链路可能重用相同的 IP 链路, 覆盖节点无法直接控制底层物理链路的资源。要想得到覆盖链路的最新信息, 只能通过测量的方法, 随着覆盖链路的增加, 探测开销变得非常大。因为覆盖网络在进行路由时不易充分探测到底层物理链路的性能信息, 如果有多条覆盖链路使用的 IP 链路全部或者部分相同, 可能导致被重用的 IP 链路繁忙, 甚至发生拥塞, 而其他 IP 链路处于闲置的状态, 使 IP 层网络负载不均衡, 降低了网络的路由性能。也就是说, 如果覆盖链路对底层 IP 链路的重用度太高, 会导致网络性能下降。文献[3]以可靠性作为约束条件, 研究了覆盖链路的选取问题, 但没有考虑 IP 链路重用度对覆盖网络性能的影响。文献[4]的研究表明, 邻接拓扑结构和物理拓扑感知的 K 个最小生成树结构(K-MST)因为利用 IP 层的物理拓扑信息选择覆盖链路, 所以覆盖网络的性能大大提高。但以上研究只通过实

验验证了这个事实, 未给出这 2 种拓扑结构的数学模型。

2 覆盖链路选取问题模型

在一个给定覆盖节点的覆盖网络中, 维护的覆盖链路越多, 维护代价方面的开销就越大。但是如果覆盖链路选取过少, 则会影响覆盖网络的性能。通过以往的研究^[4]可知: 在选择链路时, 采用物理拓扑感知的 K-MST 拓扑结构能够充分考虑底层物理拓扑信息, 提高覆盖网络性能。而提高覆盖网络性能与减小覆盖网络拓扑维护代价是相互矛盾的目标, 应尽量实现最高的性价比。不失一般性, 本文给定覆盖拓扑的维护代价, 并在给定的维护代价基础上, 求解使覆盖网络路由性能达到最佳的一些链路。下面给出覆盖链路选择问题的形式化描述和模型。

设无向图 $G_{ON}=(V, E)$, 其中, V 表示覆盖节点的集合; E 表示任意 2 个覆盖节点之间边的集合。对于任意节点 $i, j \in V$ 、边 $e(i, j) \in E$, 边权函数 $e_D(i, j) \rightarrow R^+$ 表示物理链路的延迟, 边权函数 $e_B(i, j) \rightarrow R^+$ 表示物理链路的带宽。 $V^o \subset V$ 表示覆盖节点的集合, 对于任意覆盖节点 $u, v \in V^o$, $e^o(u, v)$ 表示两节点之间的覆盖链路, 边权函数 $e_D^o(u, v) \rightarrow R^+$ 表示覆盖链路的延迟, 边权函数 $e_B^o(u, v) \rightarrow R^+$ 表示覆盖链路的带宽。

定义 1 IP 链路重用标记函数

$$\varphi^{e^o(u,v)} e(i, j) = \begin{cases} 0 & \text{覆盖链路 } e^o(u, v) \text{ 使用了物理链路 } e(i, j) \\ 1 & \text{覆盖链路 } e^o(u, v) \text{ 没有使用物理链路 } e(i, j) \end{cases}$$

定义 2 IP 链路重用度定义为该 IP 链路在 K-MST 覆盖拓

基金项目: 中央财经大学“211 工程”三期基金资助项目(2009211041)

作者简介: 张艳梅(1976 -), 女, 副教授、博士, 主研方向: 覆盖网络, 服务计算, 服务组合

收稿日期: 2010-06-30 **E-mail:** jilzym0309@sina.com

扑中使用的次数总和：

$$w(i, j) = \sum_{e^o(u, v) \in KMST} \varphi^{e^o(u, v)} e(i, j)$$

其中, $KMST$ 表示 K-MST 结构。

定义 3 覆盖网络的物理链路重用度定义为所有 IP 链路在 K-MST 覆盖拓扑中被重复使用的次数总和：

$$olap(KMST) = \sum_{e(i, j) \in G_{ON}} w(i, j)$$

$$\text{其中, } w(i, j) = \begin{cases} w(i, j) & \text{if } w(i, j) = 1 \\ w(i, j) - 1 & \text{if } w(i, j) > 1 \end{cases}$$

定义 4 覆盖链路的带宽为组成该覆盖链路的所有物理链路中的带宽最小值： $e_B^o(u, v) = \min(e_B(i, j)) \times \varphi^{e^o(u, v)} e(i, j)$ 。整个覆盖网络为 K-MST 拓扑结构, 其链路带宽为各覆盖链路带宽之和, 表示为：

$$bandwidth(KMST) = \sum_{e^o(u, v) \in KMST} e_B^o(u, v) \quad (1)$$

定义 5 覆盖链路的延迟为组成该覆盖链路的所有物理链路的延迟之和, 即：

$$e_D^o(u, v) = \sum (e_D(i, j)) \times \varphi^{e^o(u, v)} e(i, j)$$

同理, 各覆盖网络的链路延迟为各覆盖链路延迟之和, 表示为：

$$delay(KMST) = \sum_{e^o(u, v) \in KMST} e_D^o(u, v) \quad (2)$$

本文在解决 OLSP 时, 采用了 K-MST 结构, 借鉴图论中的概念以及服务组合覆盖网络的相关知识, 给出定义 6。

定义 6 K-MST 的覆盖网络拓扑结构是由 K 棵覆盖链路重叠度最小的生成树(MST)组成的。

由 MST 的基本性质可知, 在具有 n 个节点的覆盖网络图 G 中, 若选择 K 个最小生成树的覆盖拓扑结构, 并且希望覆盖链路完全没有重复, 则 K 的取值应小于或等于 $\left\lfloor \frac{n}{2} \right\rfloor$ 。因此,

得到 K 的取值范围为： $1 \leq K \leq \left\lfloor \frac{n}{2} \right\rfloor$ 。

为了保证覆盖网络的正常运行, 需要周期性地发送探测消息, 收集物理网络的 QoS 属性和资源状态。覆盖拓扑维护开销可以被定义为为了覆盖网络拓扑图, 每个时间单位所需维护的消息数量。假设 G_{ON} 中覆盖节点 V^o 的个数为 n , 维护覆盖拓扑的开销限制在每个时间单位不超过 M 个消息, 覆盖拓扑的维护周期是每个时间单位触发 I 次, 因此, 可以维护的边的数量为 $k = \left\lfloor \frac{M}{I} \right\rfloor$ 。维护一个生成树拓扑的消息数量为

$$n-1, \text{ 这样, K-MST 中生成树的最大数量为 } K = \left\lfloor \frac{M}{(n-1) \cdot I} \right\rfloor。$$

综上, 得到链路选择的优化模型：

$$\max(bandwidth(KMST)) \quad (3)$$

$$\min(delay(KMST)) \quad (4)$$

$$\min(olap(KMST)) \quad (5)$$

s.t.

$$topology(G_{ON}) = KMST, \quad K \leq \left\lfloor \frac{M}{(n-1) \cdot I} \right\rfloor \quad (6)$$

其中, 式(3)、式(4)表示覆盖网络的路由性能最好, 即最大的网络带宽和最小的延迟; 式(5)表示覆盖网络的物理链路重用度最小; 式(6)表示采用 K-MST 覆盖拓扑的约束条件, K 的值根据维护开销约束 M 以及状态维护周期 I 求得。

3 求解覆盖链路选择问题的免疫克隆算法

免疫克隆算法是一种模拟自然免疫系统功能来求解问题

的自适应搜索算法, 它能够兼顾全局搜索和局部搜索, 快速收敛到最优解。

OLSP 属于多目标优化问题, 往往通过加权等方式转化为单目标问题, 然后用数学规划的方法求解, 每次只能得到一种权值情况下的最优解, 且它们对于权重值或目标给定的次序较敏感。借鉴文献[5]将经济学中的 Pareto 理论与进化算法结合起来求解多目标优化问题的新思路, 本文采用 Pareto 占优理论和免疫克隆算法求解 OLSP。

对于多目标优化问题, 一个解对于某个目标来说可能是较好的, 而对其他目标来讲可能较差, 因此, 存在一个折中解的集合, 称为 Pareto 最优解集^[5]。进化算法通过在代与代之间维持由潜在解组成的种群来实现全局搜索, 这种从种群到种群的方法对于搜索多目标优化问题的 Pareto 最优解集是很有用的。因此, 其求解目的是找到一个分布均匀、多样性好而且能够快速向着最优 Pareto 前沿面逼近的最优解集。

结合覆盖链路选取问题的特点, 求解 OLSP 问题的多目标免疫克隆算法 ICA_OLSP 设计如下：

(1) 编码问题

本算法采用二进制编码方式, 抗体 a_i 的编码可以表示为 0-1 对称矩阵 $\mathbf{a}_i = (a_{11}^{[i]}, a_{12}^{[i]}, \dots, a_{1n}^{[i]}; a_{21}^{[i]}, a_{22}^{[i]}, \dots, a_{2n}^{[i]}; \dots; a_{n1}^{[i]}, a_{n2}^{[i]}, \dots, a_{nn}^{[i]})$ 。第 k 个抗体的基因 $a_{ij}^{[k]} \in \{0, 1\}$ 表示是否选取覆盖节点 i, j 之间的覆盖链路, 如果选取该链路, 则 $a_{ij}^{[k]} = 1$, 否则, $a_{ij}^{[k]} = 0$ 。设初始抗体群规模为 m , 即 $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ 。每个抗体 a_i 都对应于一个候选解 K-MST, 抗原对应于如下目标函数： $f(\mathbf{a}_i) = \{f_1(\mathbf{a}_i), f_2(\mathbf{a}_i), f_3(\mathbf{a}_i)\}$, 其中, $f_1(\mathbf{a}_i) = \frac{1}{bandwidth(\mathbf{a}_i)}$; $f_2(\mathbf{a}_i) = delay(\mathbf{a}_i)$; $f_3(\mathbf{a}_i) = olap(\mathbf{a}_i)$ 。

(2) 初始抗体群的生成方法

根据上文, K 的取值范围是已知的, 因此, 选择在可行解空间内生成初始抗体群。可以通过变换边权值的比重得到一组 K-MST 拓扑结构, 即初始抗体群。初始抗体群的质量决定了算法求解的质量和算法收敛速度。在进化求解过程中, 免疫基因操作可能会打破这种初始生成的 K-MST 结构, 形成类 K-MST 结构。求解的目的是得到 Pareto 最优解集, 而且只要边数在约束范围内, 其解仍为可行解。

(3) 克隆选择操作

克隆选择操作(Clonal Selection Operation, CSO)是由亲和度诱导的抗体随机映射, 实际上包括了克隆操作、免疫基因操作和克隆选择操作。对一个抗体群进行克隆选择操作后将得到一个新的抗体群, 其抗体群的状态转移可以表示成如下的随机过程：

$$\begin{aligned} A(i) &\xrightarrow{\text{克隆操作 } T_c} B(i) \xrightarrow{\text{免疫基因操作 } T_r / T_m} \\ C(i) \cup B(i) &\xrightarrow{\text{克隆选择操作 } T_s} A(i+1) \end{aligned}$$

克隆操作、免疫基因操作和克隆选择操作的细节参见文献[6]。算法的时间复杂性为 $O(m^2)$, m 为抗体群规模。

4 仿真模拟及结果分析

实验使用 Brite 工具生成网络拓扑图, 基于 Waxman 的拓扑生成算法 $P_e(u, v) = \beta \exp \frac{-l(u, v)}{L\alpha}$ 产生具有实际网络特性的图的模型。覆盖节点数为 50 个, 表示组合代理和征用节点的集合。平均每个节点的度数 $m=5$, $\alpha=0.15$, $\beta=0.2$ 。边的权值分为 3 种：(1)权值 $e_D \in [0, 1]$, 表示链路的延迟代价; (2)权值 $e_B \in [1, 10]$, 表示链路的带宽代价; (3)为了模拟覆盖链路对物理链路的重用问题, 为包含 n 个节点的完全图的每条边从 1~

$\frac{n(n-1)}{2}$ 进行编号, 然后给出一个边权 $e_o \in [0, 0.3]$ 表示对物理链路的重用度。

初始抗体群规模为 10, 备份抗体群规模为 10。因为初始抗体群的质量对免疫克隆算法的解的质量和收敛速度影响很大, 且 OLSP 问题的抗体是 K-MST(K=5)拓扑结构, 所以初始抗体群产生方法如下: 分别以延迟和带宽的 5 种不同比例权重的综合指标为代价各产生 2 个抗体, 共 20 个抗体作为初始抗体群; 分别以延迟和带宽单项指标为代价各产生 5 个抗体作为备份抗体群。其中, 克隆比例为 3; 个体重组概率为 0.6, 个体变异概率为 0.5。算法的最大进化代数数为 100, 每种实验运行 20 次, 取平均值作为最终结果。

本文通过建立 Pareto 最优前端来获得各优化指标, 同时近似最优的解结果, 寻优收敛过程的空间点图形表示见图 1~图 3, 它们分别表示在可行时间内各优化指标对应的 Pareto 最优前端图形, 即各指标的最优解随进化代数变化的情况, 其反映了多目标免疫克隆搜索的优化和收敛过程。算法执行完毕产生的一组最终最优解集如图 4 所示。取得的最终解从单个优化指标角度看可能并不是最优的, 但同时考虑 3 个子目标时是最优的。将最终的这些最优解提供给用户, 用户即可根据通信业务需要从中选取较好的解, 例如在实时视频传输时更关注的是端到端时延, 则可将时延作为一种硬约束条件, 选择符合该约束或者此指标最小的路由, 这也是多目标求解的优势。

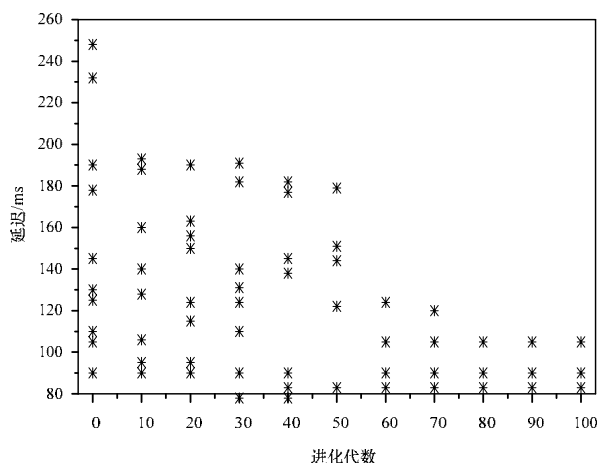


图 1 最优延迟随进化代数的变化

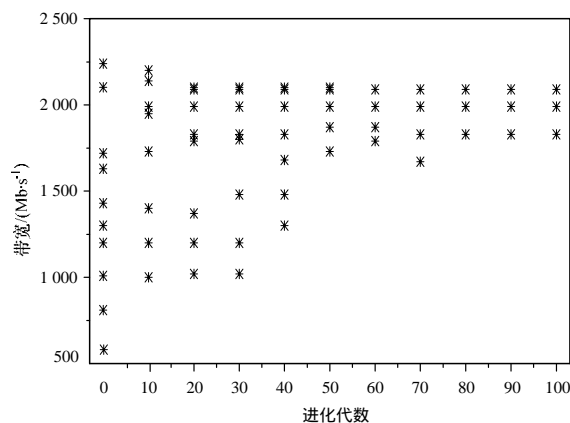


图 2 最优带宽随进化代数的变化

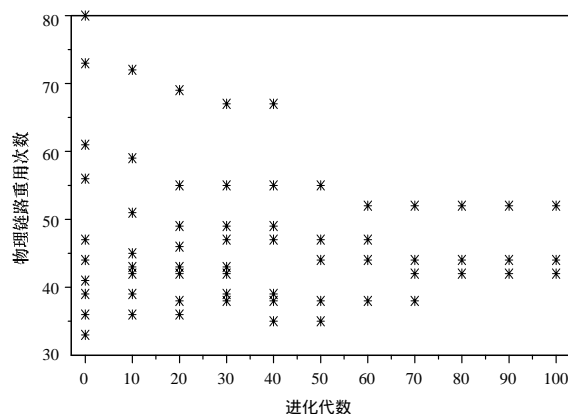


图 3 最优物理链路重用度随进化代数的变化

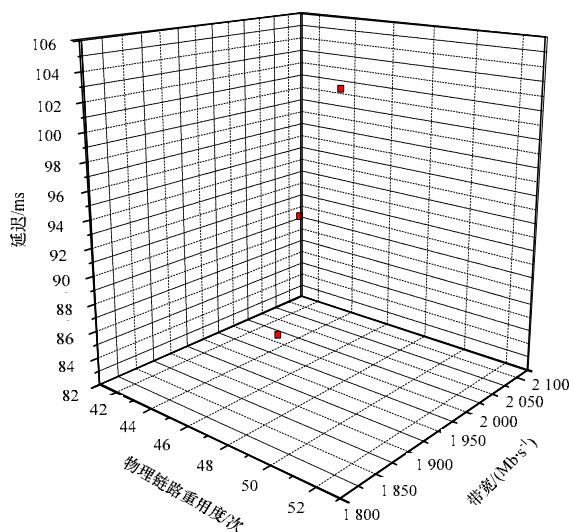


图 4 Pareto 最优解集

5 结束语

本文研究了覆盖网络中覆盖链路的优化选取问题。在覆盖链路选取过程中充分考虑了可能影响覆盖网络性能的各种因素, 并提取重要影响因子建立了多目标的数学优化模型。在求解过程中, 通过经济学中的 Pareto 占优理论和智能优化算法相结合的方法来求 Pareto 最优解集。仿真结果验证了本文模型的合理性和求解算法的有效性。

参考文献

- [1] Wang Wenjie, Jin Cheng, Jamin S. Network Overlay Construction under Limited End-to-end Reachability[C]//Proc. of INFOCOM'05. Miami, USA: IEEE Press, 2005.
- [2] 张艳梅, 余镇危. 面向服务组合的覆盖服务网络设计[J]. 计算机工程, 2007, 33(22): 137-139.
- [3] 王德志. 基于 Overlay Network 应用层与网络层协同组播机制研究[D]. 北京: 中国矿业大学, 2007.
- [4] Li Zhi, Mohapatra P. The Impact of Topology on Overlay Routing Service[C]//Proc. of INFOCOM'04. [S. l.]: IEEE Press, 2004.
- [5] 尚荣华, 焦李成, 马文萍. 免疫克隆多目标优化算法求解约束优化问题[J]. 软件学报, 2008, 19(11): 2943-2956.
- [6] 张艳梅. 基于覆盖网络的服务组合关键技术研究[D]. 北京: 中国矿业大学, 2009.

编辑 张 帆