

一种应用于负载均衡的模拟退火遗传算法

贾艳萍, 孟相如, 麻海圆

(空军工程大学 电讯工程学院, 陕西 西安 710077)

摘要:提出了一种基于模拟退火遗传优化算法,求解流量工程中的网络负载均衡问题。这种新型算法不仅能够均衡网络业务流量,相对于其它遗传算法,还具有收敛速度快、简单高效的特点。通过理论分析详尽说明算法的设计思想和相对于现有算法的优越性。

关键词:流量工程;负载均衡;遗传算法;模拟退火算法

中图分类号: TP393 **文献标识码:** A **文章编号:** 1009-3516(2007)01-0059-04

当前的 Internet 必须从提供单一的尽力而为业务发展为能够保证多种类型的业务质量,以满足日益增长的需求。流量工程(TE:Traffic Engineering)是实现这一目标的关键技术之一。传统的 Internet 路由协议默认的总是基于目的地址的最短路径选路方式,常常导致网络上的流量分布不平衡,使得网络上有些链路因为过负荷产生拥塞现象,而另一些链路资源却处于闲置状态,增加丢包率和恶化资源利用率^[1]。TE 的主要目的就是优化资源利用率,提高网络性能,增加网络的健壮性,负载均衡是其中的重要功能。

TE 的实质,是在复杂约束条件下的多业务多目标网络流问题,这是一个 NP-hard 问题,是不能用任何已知多项式算法求解的问题。这个问题不一定存在所有目标函数共同的极大点,其优化求解主要采用两大类研究方法:启发式算法和遗传算法。前者是枚举法的改进,往往有很强的剪枝手段,但其计算量太大因此不适用于大规模流量工程问题。遗传算法以其鲁棒性、并行性及高效性等优点得到了广泛的应用^[2-4],遗传算法能从概率意义上以随机的方式寻求到最优解,但在实际应用中会出现一些问题,主要是容易产生早熟现象、局部寻优能力较差和过早收敛等,为了找到最优解或次优解,本文提出了一种基于模拟退火遗传算法^[5]优化的负载均衡算法。

1 负载均衡问题建模

1.1 基本函数定义

把网络定义为一个有向加权图 $G=(V,E)$,其中 V 是网络有限的节点集, E 是网络有限的链路集,设 $|V|=n, |E|=m$,边 $(i,j) \in E$ 上有链路带宽 b_{ij} 、链路时延 d_{ij} 和费用 c_{ij} 3 个参数。 b_{ij} 、 d_{ij} 、 c_{ij} 分别表示节点 i 到 j 的可预留带宽、延迟、费用。用集合 $R=(S,B,D,T)$ 表示网络中的 LSP 请求,对任意 $r \in R$,用四元组 (s_r, b_r, d_r, t_r) 表示,其中 s_r, t_r 分别是入口节点和出口节点; b_r 是业务的最小带宽要求; d_r 是业务最小时延要求,设:

$$\rho'_{ij} = \begin{cases} 1 & \text{如果 LSP } r \text{ 经由 } (i,j) \in p \\ 0 & \text{其它} \end{cases}; \quad h(p) = \sum_{i=1}^n \sum_{j=1}^n \rho'_{ij}; \quad \text{Cost}(p) = \sum_{i=1}^n \sum_{j=1}^n c_{ij} \rho'_{ij}$$

其中 p 是一条从源节点 s 到目的节点 t 的路径, $h(p)$ 是 LSP 请求 r 的路径长度, $\text{Cost}(p)$ 是 r 的费用。在 QoS 度量参数中,带宽属于凹性度量参数,而时延属于加性度量参数,因此路径 p 的带宽 $\text{Bandwidth}(p)$ 和时延 $\text{Delay}(p)$ 分别为

$$\text{Bandwidth}(p) = \min b_{ij} \rho'_{ij}; \quad \text{Cost}(p) = \sum_{i=1}^n \sum_{j=1}^n d_{ij} \rho'_{ij}$$

收稿日期:2006-07-17

作者简介:贾艳萍(1982-),女,四川南充人,硕士生,主要从事宽带网络技术研究;

孟相如(1963-),男,陕西蓝田人,教授,博士生导师,主要从事宽带通信网络和信号处理技术研究。

则所选择的通路应该同时满足下面的式子:

$$\text{Bandwidth}(p) \geq B \quad (1) \quad ; \quad \text{Delay}(p) \leq D \quad (2)$$

$$\sum_{j:(i,j) \in E} \rho_{ij}^r - \sum_{j:(j,i) \in E} \rho_{ji}^r = 0, \quad i \neq s_r, t_r \quad (3)$$

$$\sum_{j:(i,j) \in E} \rho_{ij}^r - \sum_{j:(j,i) \in E} \rho_{ji}^r = 1, \quad i = s_r \quad (4)$$

$$\sum_{j:(i,j) \in E} \rho_{ij}^r - \sum_{j:(j,i) \in E} \rho_{ji}^r = -1, \quad i = t_r \quad (5)$$

式(3)是对中间节点的限制,流入的流量应等于流出的流量;式(4)和(5)分别是对入口和出口节点的流量限制。

1.2 资源消耗函数

网络资源利用情况可以从绝对和相对两方面来考虑,绝对是指从业务请求对网络资源的占用情况考虑,希望某一请求选择占用资源最少的路径;相对是指从整体网络资源被最大利用的角度来考虑,希望网络资源被均衡使用,避免由于不合理的路径选择造成的瓶颈效应。对于一条所选路径,一个业务流占用的网络资源是沿途各链路为该路径预留资源 B 的总和与该业务流占用该路径时间的乘积加上此路径上的费用,即资源消耗函数 Expense 为^[4]

$$\text{Expense}(p) = \sum_{i=1}^n \sum_{j=1}^n B \rho_{i,j}^r \cdot \text{Delay}(p) + \text{Cost}(p) = h(p) \cdot B \cdot \text{Delay}(p) + \text{Cost}(p) \quad (6)$$

由式(6)可以看出,跳数、延迟和费用越少的路径消耗的网络资源越少。由此,资源消耗函数 $\text{Expense}(p)$ 的最小化就体现了最小跳数、最小延迟、最小费用的优化准则。

负载分布情况可以用接入新的连接后对原有链路可用带宽的影响来衡量。节点 i 到节点 j 间链路利用率 U_{ij} 、平均链路利用率 EU 和链路利用率方差 $\text{Equal}(p)$ 分别为

$$U_{ij} = \frac{B}{b_{ij}} \quad (7) \quad ; \quad \text{EU} = \frac{\sum_{i=1}^n \sum_{j=1}^n U_{ij} \rho_{i,j}^r}{h(p)} \quad (8) \quad ; \quad \text{Equal}(p) = \sum_{i=1}^n \sum_{j=1}^n (U_{ij} - \text{EU})^2 \rho_{i,j}^r \quad (9)$$

文献[6]认为,路径的可用带宽越大,该路径的安全性就越好。公式(7)中,当 B 接近于 b_{ij} ,即 $U_{ij} \rightarrow 1$ 时,说明链路 (i,j) 的可用带宽与业务所要求的带宽接近,那么选择这条链路不安全,而且如果这一链路接纳此连接后,就无法再接纳其他经过该链路的连接请求了。当 B 远远小于 b_{ij} 时,链路 (i,j) 的可用带宽比较大,这条链路比较安全,接纳新的连接后,剩余带宽比较多,不会影响接纳新的连接请求。路径上链路利用率的方差是定义负载分布的均衡程度。

1.3 建立目标函数

对于一个有最小带宽和时延约束的业务请求,寻找一条既能满足约束条件又能体现充分利用网络资源、实现负载均衡的路径问题是一个多目标优化问题,这种问题最显著的特点是目标之间的不可公度性和目标产生的矛盾性。所谓目标之间的不可公度性,是指各个目标没有统一的度量标准。如在 QoS 路由的选择目标中,通常希望得到端到端的延迟最小、跳数最小、瓶颈带宽最大、所占用的网络资源最少、使用网络的费用最小等。在多数情况下上述 5 个目标之间是相互冲突的,不可能存在一个解使它们同时达到最优,为此构造目标函数如下:

$$F = \mu \cdot \text{Expense}(p) + v \cdot \text{Equal}(p) \quad (10)$$

其中, μ 和 v 为正的实数系数,选择合适的系数可以使目标函数 F 的两个部分占有大致相等的比例。这样,就可以把体现网络资源合理分配、网络负载均衡的优化问题描述为:寻找一条路径 p ,在满足跳数、带宽、延迟、费用约束条件的基础上,使得函数 F 最小化。

2 基于模拟退火遗传算法的求解

遗传算法(GA:Genetic Algorithm)是模拟生物进化过程的一种并行优化算法,适用于在复杂而庞大的搜索空间中寻找最优解或次优解,目前已广泛用于解决 NP-hard 问题。但是遗传算法也存在一些不足:比如它的局部搜索能力较差、收敛过早等;模拟退火算法是 1982 年 Kirkpatrick 等将固体退火思想引入组合优化领域,提出了一种解大规模组合优化问题,特别是 NP 完全组合优化的有效近似算法。它具有较强的局部搜索

能力,并能使搜索过程避免陷入局部最优解,但模拟退火算法不适合整个搜索空间,不能使搜索过程进入最有希望的搜索区域,从而使模拟退火算法的运行效率不高。因此,本文采用引入模拟退火机制的新型遗传算法结合有效的自适应遗传算子来求解负载均衡问题。

算法实现方法是:以遗传算法运算流程作为主体流程,把模拟退火机制融入其中,用以进一步调整优化群体。整个算法的执行过程由两部分组成,首先通过遗传算法的进化操作(侧重全局搜索)产生出较优良的一个群体,再利用模拟退火算法的退火操作(侧重局部搜索)来进行进一步基因个体的优化调整。其中模拟退火操作设计针对一定规模群体中的每个基因个体,对每个基因个体实施一定演变而产生新解的设计引用了遗传算法中的变异和倒位思想。这种算法思想策略,从对全局最优解的搜索角度和算法的进化速度上来提高遗传算法的性能。运行过程反复迭代,直到满足某个终止条件为止,其流程如图 1 所示。

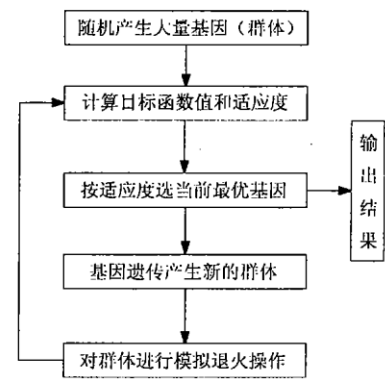


图 1 算法流程图

2.1 编码机制

以 $n \times n$ 位一维二值编码作为遗传算法的编码机制,表示为 $\chi_1, \chi_2, \dots, \chi_n \in \{0, 1\}$ 。 χ_k 表示从节点 i 到节点 j 的链路 (i, j) 是否被选中 ($\chi_k = 1$ 表示被选中, $\chi_k = 0$ 表示未被选中, 这里 $i = k/n, j = k\%n$, “/” 是取整操作, “%” 是取余操作)。在编码过程中,如下两条规则可以简化编码:① 某些基因的值是确定的,如若两节点之间不存在链路,则与之对应的基因值一定为 0,因此这些基因可以去掉,在解码过程中,将相应位置设置为 0 即可;② 业务量流入了源节点或流出了目的节点的链路在编码时也不考虑,设置为 0,因为如果出现了这种情况就表示一定有环路。

解码时要将 $n \times n$ 的一维编码转换成二维表示形式,即先将 $n \times n$ 位长的二进制编码串切断为两个 $n \times n/2$ 位长的编码串,然后将其转换成对应的十进制整数代码。

2.2 适应度函数

适应度函数评估是选择操作的依据,它的设计直接影响到遗传算法的性能,因此适应度函数的设计要结合求解问题来定。根据给出的目标函数,考虑到在遗传算法中适应度函数通常是求最大值的形式,建立的适应度函数为 $fitness = 1/F$ 。

$$f_i = e^{f_i/T} / \sum_{i=1}^{Popsize} e^{f_i/T}, \quad T = T_0(0.99^{g-1})$$

式中, f_i 为第 i 个个体的适应度, g 为遗传代数, T 为温度, T_0 为初始温度。

2.3 选择操作

采用最佳个体保存法和比例选择法相结合的选择方法。即在群体交叉、变异之前,先选出最佳个体,直接复制到下一代中;其余个体的选择采用最基本的比例选择法(赌轮选择法),即各个体被选择的概率与其适应度成比例。

2.4 交叉和变异

采用单点交叉和均匀变异算子,交叉、变异概率采用自适应的 P_{cross} 和 P_{mutate} , 2 者计算表达式如下:

$$P_{cross} = \begin{cases} \frac{k_1(f_{max} - f')}{f_{max} - f_{avg}} & f' \geq f_{avg} \\ k_2 & f' < f_{avg} \end{cases}; \quad P_{mutate} = \begin{cases} \frac{k_3(f_{max} - f)}{f_{max} - f_{avg}} & f \geq f_{avg} \\ k_4 & f < f_{avg} \end{cases}$$

式中, f_{max} 是群体中最大的适应度值, f_{avg} 是每代群体的平均适应度值, f' 是要交叉的 2 个个体中较大的适应度值, f 是要变异个体的适应度值,且 $0 < k_1, k_2, k_3, k_4 \leq 1$ 。

2.5 模拟退火操作

1) 算法中对群体中的个体执行个体跳变操作的思想基于遗传算法的“变异”操作,通过两步实现:第一是对个体中的基因按一定的概率进行随机单点置反操作;第二是把个体基因串按较小的概率 $q (q < 0.01)$ 进行前后等长交换操作。这样的演变操作对维持群体中个体的多样性,防止算法陷入局部最优有着重要作用。

2) 应选取充分大的控制参数初值,即可达到准平衡的论证。实际上,为使算法进程一开始就达到准平衡,应让初始接受率 α 接近于 1,由 Metropolis 准则 $\exp(f/T_0) \approx 1$ 可推知 T 值很大。参数选取的法则是:要

求算法进程在合理的时间里搜索尽可能大的解空间,只有足够大的 T_0 才能满足这个要求。

3) 控制参数终值 T_f 选取。由于算法收敛于最优解集是随控制参数 T 值的缓慢减小逐渐进行,只有在控制参数终值 T_f 充分小时,才有可能得出高质最终解。

3 算法分析

本文提出的基于遗传算法的负载均衡算法能够收敛到全局最优解,因为在算法中采取了以下几种策略:①以自适应概率 $P_{\text{cross}}, P_{\text{mutate}} \in (0, 1)$ 进行交叉、变异操作;②引入模拟退火操作;③按比例选择且在选择前保存最优个体。文献[5]证明了满足此3个条件的遗传算法能够收敛到最优解。但是在实际求解时,求出全局最优解的时间复杂度很大,本算法通过限定总的迭代次数,求出一个性能较好的可行解。

算法的复杂度主要由参数的选择、交叉和变异操作决定,因此本算法的复杂度为 $O(\text{generation} \times \text{popsize} \times n^2)$,其中 generation 为总迭代次数, popsize 为群体规模, n 为网络节点总数。

该算法和现有的其他算法比较,具有以下特点:①基于遗传算法,进行全空间并行搜索,具有较强的鲁棒性;②对遗传操作产生的新个体进行模拟退火操作,提高了局部搜索能力避免算法陷入局部最优解;③自适应的交叉变异算子,并结合最优个体保存进行选择,提高了算法的全局搜索性能;④对适应度拉伸的方法,可以避免在温度高时个别好的个体的后代充斥整个种群,造成早熟,同时在温度下降后使适应度相近的个体适应度差异放大,从而使优秀的个体优势更明显。

参考文献:

- [1] Xiao Xi Peng. Traffic Engineering With MPLS in the Internet [J]. IEEE Networking, 2000, (4): 29 - 33.
- [2] Juidette H, Youlal H. Fuzzy Dynamic Path Planning Using Genetic Algorithms [J]. Electronics Letters, 2000, 36(4): 374 - 376.
- [3] Man K F, Tang K S, Kwong S. Genetic Algorithms [M]. London: Springer, 1999.
- [4] 陈燕, 宋玲, 李陶深. 基于遗传算法的网络负载均衡的选播路由算法 [J]. 计算机工程, 2005, 31(8): 93 - 95.
- [5] 康立山, 谢云, 尤矢勇, 等. 非数值并行算法(第一册)—模拟退火算法 [M]. 北京: 科学出版社, 1995.
- [6] Vegesna S. IP 服务质量 [M]. 北京: 人民邮电出版社, 2001.
- [7] 王小平, 曹立明. 遗传算法—理论、应用与软件实现 [M]. 西安: 西安交通大学出版社, 2002.

(编辑: 门向生)

A Simulated Annealing Genetic Algorithm for Load Balancing

JIA Yan - ping, MENG Xiang - ru, MA Hai - yuan

(The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077, Shaanxi, China)

Abstract: A simulated annealing genetic algorithm for load balancing in traffic engineering is proposed in this paper. The new algorithm can well balance the traffic and converge more quickly when compared with current genetic algorithms. Design guidelines and advantages over other algorithms are explained in detail.

Key words: traffic engineering; load balancing; genetic algorithm; simulated annealing algorithm