基于遗传算法的网络资源配置优化

张 英[1,2] 杨雅辉[2] 田浩然[2] 张 威[2] 董晶晶[3]
(中国科学院计算技术研究所 北京 100190)[1] (北京大学软件与微电子学院 北京 102600)[2]
(中国矿业大学(北京)机电与信息工程学院 北京 100083)[3]

摘 要 随着新一代网络的结构复杂化和应用多样化,如何进一步提高实时业务的服务质量,是急需解决的重要问题之一,而合理配置和优化网络资源是解决该问题的有效方法。针对多业务网络,提出一种以网络资源均衡度为目标、QoS为约束的优化模型;应用改进的遗传算法实现带宽和节点缓冲区资源的优化配置,以实现网络流量分布均衡,从而提高网络性能。通过分析遗传算法种群适应度的变化趋势,给出一种基于种群稳定性系数阈值的迭代终止方法,以提高算法收敛效率。通过实验验证了以上方法的有效性。

关键词 多业务网络,资源配置优化,遗传算法
中图法分类号 TP393.0 **文献标识码** A

Optimization of Network Resource Allocation Based on Genetic Algorithm

ZHANG Ying[1,2] YANG Ya-hui[2] TIAN Hao-ran[2] ZHANG Wei[2] DONG Jing-jing[3]
(The Institute of Computing Technology, the Chinese Academy of Sciences, Beijing 100190, China)[1]
(School of Software and Microelectronics, Peking University, Beijing 102600, China)[2]
(School of Mechanical Electronic & Information Engineering, China University of Mining & Technology, Beijing 100083, China)[3]

Abstract As the next-generation network architecture becomes more and more complex and the applications become more diverse, how to improve the performance of the network becomes a big problem that we must solve. One important effective method to solve the problem is to rationally allocate and optimize the network resource. Based on the multiservice networks, this paper proposed a new optimization model which is targeted at network resource leveling under the QoS restrictions, and balance of network traffic distribution by optimizing the allocation of the network bandwidth and buffers with an improved genetic algorithm so as to improve the performance of network. A method based on threshold of population stability coefficient to terminate the iteration of the genetic algorithm was given by analysing the variation trend of population stability coefficient. The experiments show that the method is effective to improve the efficiency of the algorithm.

Keywords Multiservice network, Optimization of allocation of network resource, Genetic algorithm

1 概述

随着新一代互联网的发展,实时业务飞速增长且业务呈现多样化。如何进一步提高服务质量,是我们面临解决的重要问题之一,而合理配置和优化网络资源已成为一种提高网络性能的有效方法。

为解决IP网络承载多媒体业务的服务质量问题,Internet工作任务组(IETF)提出了区分服务(DiffServ)和集成服务(IntServ)两种QoS体系结构。其中IntServ在数据路径上执行基于流的分类、调度和缓冲区管理,因此扩展性差、实施复杂;DiffServ将状态相关的控制转移到了网络边缘节点,对不同QoS业务进行分类,保持了网络内部的简单性,并得以广泛应用[2]。为此,国内外针对以上QoS体系从资源利用优化方面开展了一系列研究,包括接入控制、队列管理、调度算法等[7,8],其以提高网络资源的利用率或以承载业务量的最大化为目标,但大多是面向单一资源的。随着网络服务的多样化,应用要求资源配置需满足全网在承载预定业务状态下达到资源的合理配置和均衡使用。

基于以上问题,本文针对多业务网络,以资源均衡为目标、QoS为约束条件,建立优化模型;应用改进的遗传算法,达到网络带宽和节点缓冲区资源的优化配置,从而提高网络性能;通过对种群稳定性系数的分析给出一种基于阈值的迭代终止方法,从而提高算法的收敛效率。

2 网络描述

在一个IP网络中,分组从源端发送到目的端过程中,一

到稿日期:2011-06-20 返修日期:2011-09-18 本文受创新资金项目(BT2008-22)资助。

张 英(1951-),女,研究员,主要研究方向为计算机网络规划及性能评价,E-mail:ccss@ict.ac.cn;杨雅辉(1966-),女,博士,教授,主要研究方向为计算机网络规划及安全;田浩然(1982-),男,硕士生,主要研究方向为算法优化及软件;张 威(1986-),男,硕士生,主要研究方向为算法优化及软件,E-mail:beyond2004_ppa@163.com(通信作者);董晶晶(1983-),男,硕士,主要研究方向为网络优化算法。

般都会在中间节点的缓冲区中参加排队，等待节点提供的路由转发服务。因此，把网络描述为一个有多类业务流输入的随机服务系统，由此构成一个端到端的串行排队服务模型[1]。

2.1 网络资源描述

· 节点缓冲区

V 代表网络中的节点集，$|V|=n$ 表示网络中共有 n 个节点，每个节点的缓冲区长度为 H_i，$i=1,2,\cdots,n$。

· 链路带宽

L 代表网络中的链路集合，$L=\{l_{ij}\times x_{ij}\mid i,j=1,2,\cdots,n\}$，$U$ 代表链路带宽集合，则节点 i 与 j 的链路带宽为 u_{ij}，且 $u_{ij}=u_{ji}$。

· 缓冲区调度策略

本文假定在同一个网络系统中，所有节点的调度策略均相同。

2.2 网络业务描述

若网络所有业务的集合为 Y，则对于任意一个端到端的给定业务 y，$y\in Y$。

用 $y_{s,d,k}$ 表示端对 (s,d) 之间的第 k 种业务的业务流，$y_{s,d}$ 表示端对 (s,d) 之间的所有业务流，其中 s 表示源端，d 表示目的端，则网络中的所有业务流集合 S 为

$S=\sum y_{s,d}$，$s,d=1,2,3,\cdots,n$ 且 $s\neq d$

为此，端到端 (s,d) 时延和丢包率的要求为

时延需求：$TQOS_{s,d}=\min\{T_{qos,y_k},y_k\in y_{s,d}\}$

丢包率需求：$PQOS_{s,d}=\min\{P_{qos,y_k},y_k\in y_{s,d}\}$

3 优化目标及模型

优化模型目标通过衡量网络资源（链路带宽和节点缓冲区）均衡度，达到实现网络流量分布均衡和合理配置资源的目的。

3.1 使用均衡度

使用均衡度包括带宽空闲率和带宽使用均衡度，分别定义如下。

带宽空闲率：

$$U_{ij}=\frac{u_{ij}-M_{ij}\times l}{u_{ij}},i,j=1,2,3,\cdots,n$$

带宽使用均衡度为带宽空闲率的方差[4]：

$D(U)=E(U^2)-[E(U)]^2$

式中，u_{ij} 是链路 $L_{i,j}$（始于节点 i，止于节点 j）的带宽，M_{ij} 是链路 $L_{i,j}$ 上的流量，l 为平均分组长度。

3.2 等待均衡度

等待均衡度包括转发等待率和等待均衡度，分别定义如下。

定义 1 在节点 j 处的转发等待率为：

$$V_j=\frac{q_j\times l}{H_j},j=1,2,3,\cdots,n$$

定义 2 转发等待均衡度为数据转发等待率的方差[4]：

$D(V)=E(V^2)-[E(V)]^2$

式中，H_j 为第 j 个节点的缓冲区长度，q_j 为第 j 个节点需转发的平均分组个数，l 为平均分组长度。

3.3 目标函数

以上优化问题中，网络资源优化涉及链路带宽和节点缓冲区的均衡度及多目标优化问题，通常多目标之间需要进行权衡与折衷。可通过各目标之间的权重将多目标优化转化为单目标优化问题。此方法的优点在于可充分利用单目标优化问题求解，并在权重可以衡量的条件下给出问题的最优解。

由此，将网络资源均衡度定义为使用均衡度和等待均衡度的加权和[9]，以衡量网络资源使用的均衡情况。目标函数描述如下。

定义 3 网络资源均衡度 $S=\alpha D(U)+\beta D(V)$

其中，α 和 β 满足 $\alpha+\beta=1$，分别为使用均衡度和等待均衡度的权重系数。S 越小，表示网络流量分布越均衡，以此作为优化模型的目标函数。

优化模型为：

$\min S=\alpha D(U)+\beta D(V)$

约束条件：

$P_{s,d}\leqslant PQOS_{s,d}$

$T_{s,d}\leqslant TQOS_{s,d}$

3.4 模型求解

遗传算法是一种源于进化论的随机搜索算法[3]，它模拟生物进化过程，通过基因的遗传、杂交、变异等规律产生全局最优解，适合求解带有多参数、多变量、单目标或多目标的NP难优化问题[6]，以及大规模的带有约束条件的组合优化问题。因此本文应用该算法对上述模型进行求解。

遗传算法计算过程简单，具有良好的鲁棒性、灵活性、通用性，已在求解旅行商问题、布局优化、网络路由等具有NP难的组合优化问题中得以应用。

· 编码方式

本文模型用来求符合条件的最优带宽和缓冲区，最终解是由每条链路的带宽和每个节点的缓冲区组成的，即

$(U,H)=(B_1,B_2,B_3,\cdots,B_m;H_1,H_2,H_3,\cdots,H_n)$

这是一组实数组成的向量，若直接用实数编码方式，它的解空间将非常庞大（满足带宽和缓冲区范围的所有实数组成），搜索工作量也相当大。鉴于此，采用二进制的编码方式，每个带宽变量采用8位二进制数编码，每个缓冲区变量采用4位二进制数编码，然后对带宽和缓冲区进行变换，得出实际的编码串。

· 选择方法

采用排序选择法，根据适应度大小顺序对群体中个体排序，把事先设计好的概率表按序分配给个体作为各自的选择概率。选择概率仅与序号有关。

· 交叉和变异

本文采用分段交叉、单点变异的原则。交叉时，随机找到染色体的一个交叉点，对该交叉点后面的所有位进行交叉互换。变异时，按照一定的概率对种群中的部分染色体的随机一位进行变异。

· 适应度函数

由于遗传算法只能求解最大值问题，因而需要对上面模型当中的目标函数进行变换。适应度函数定为 $F=$定值$-S$，这样就可以通过遗传算法来求 F 最大值，从而求出 S 的最小值。

4 基于变异系数的种群稳定性分析及阈值计算

为提高算法的收敛效率，使适应度函数最优且种群处于稳定时迭代结束，通过实验分析获得了种群稳定性的判据。

4.1 定义

种群适应度：当前种群中各个体的适应度函数均值。

种群稳定性系数：在当前算法参数（种群大小、交叉变异率与算子）的设定下取种群适应度的变异系数的倒数（变异系数即标准差/均值）。该数值越大，种群越稳定。

种群稳定性系数的阈值：当种群稳定性系数大于阈值时，则认为种群达到稳定状态。

4.2 参数设置

设置内容	参数值
种群大小	32～512
选择策略	基于排序选择
交叉率	0.3
变异率	0.02

4.3 种群稳定性判据

选择种群大小为 512，保持其他参数不变，重复实验 96 次。用 $x_{i,j}$ 表示第 j 次试验中第 i 代时的种群稳定性系数，其中 $i,j\in N, i,j\geqslant 1, i\leqslant 2000$（最大迭代代数），$j\leqslant 96$（重复实验次数）。

根据 $xk,j(j=1,2,3,\cdots,96)$ 绘制种群稳定性系数的概率分布函数曲线，图 1 为第 2000 代时的曲线。横坐标为种群稳定性系数，纵坐标为概率分布函数值。每条曲线代表某一时刻的种群稳定性系数的概率分布函数情况，相邻曲线时间间隔为 100 代。

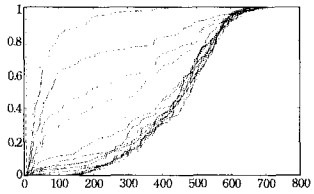

图 1 种群稳定性系数的概率分布函数对比图

到达第 1200 代以后，种群稳定性系数的概率分布情况趋于稳定。

根据种群稳定性系数稳定后的分布状况计算均值与方差。若第 k 代时种群稳定性系数分布已趋于稳定，则采用第 k 代时的种群稳定性系数描述种群稳定性系数的分布情况。

均值：$\text{mean}_k=\frac{1}{n}\sum_{j=1}^{n}x_{k,j}$

方差：$\text{var}_k=\frac{1}{n-1}\sum_{j=1}^{n}(x_{k,j}-\text{mean}_k)^2$

n 为重复实验次数，当前值为 96，计算得到 $\text{mean}_{2000}=451.07$，$\text{var}_{2000}=15003$。

4.4 阈值计算

根据正态分布理论，计算种群稳定性系数的阈值。设 2000 代时种群达到稳定的概率为 0.99，则 $P(\frac{x-\text{mean}_k}{\sqrt{\text{var}_k}}<z)=1-0.99$[4]，$\frac{x-\text{mean}_k}{\sqrt{\text{var}_k}}<-2.3263$。代入数据，得 $\frac{x-451.07}{\sqrt{15003}}<-2.3263$，$x<166.13$。

因此，在当前网络拓扑及算法参数设置条件下，种群大小为 512 时，阈值为 166.13。

4.5 阈值对算法性能的改进

将以上获得的阈值作为种群稳定性判据进行算法性能改进的实验。设算法按照最大迭代代数终止获得的历史最优解的目标函数值为 y，按照阈值判据终止获得的历史最优解的目标函数值为 y'，则精度损失率定义为 $|\frac{y'-y}{y}|$。对相同参数设置下多次获得的精度损失率求均值，可得到平均精度损失率，详见表 1。

表 1 算法性能的改进效果对照表

种群大小	重复试验次数	最小迭代代数	最大迭代代数	平均迭代代数	性能提升均值	平均精度损失率
64	82	104	233	131.80488	15.173945	0.433661939
128	117	104	693	191.01709	10.470267	0.343771133
200	119	104	2000	241.41176	8.2846004	0.188996519
256	102	135	653	305.77451	6.5407676	0.096118948
384	73	135	941	415.41096	4.8145095	0.217587367
512	96	183	2000	552.32292	3.6210701	0.075629445

由以上结果可以看出，将种群稳定性系数的阈值作为算法终止条件，对算法性能有较大提高。如种群大小为 512 时，性能比原来平均提高了 3.6210701 倍。但当种群较小时，精度损失率相对较大。这是因为对于小种群，种群间个体适应度变化相互抵消的作用相对较弱，因此通常难以获得恰当的种群稳定性阈值。为此，采用大种群确定种群稳定性阈值，而小种群则将该阈值作为遗传算法迭代终止判断条件，可以提高其精度损失率。

5 实验结果

5.1 实验参数配置

采用由 8 个节点组成的网络拓扑。设节点 1 为源节点，节点 8 为目的节点。业务包括 VOIP 和 VOD 两类。

VOIP 并发用户数为 3000，平均包长为 240B/p，发生速率为 33p/s；VOD 并发用户数为 4500，平均包长为 320B/p，发生速率为 150p/s。计算可知源端总负载为 240Mb/s。

各节点缓冲区为 512Mb，调度策略为先到先服务，管理策略为尾丢弃，路由采用最短路径策略。

5.2 实验结果分析

在网络配置设计中，带宽是根据业务量确定的。通常在分析网络带宽配置需求时，为满足高峰时所有业务的 QoS 要求，全网各链路的带宽容量要一致并取并发业务量的总和。另外考虑不可预计因素，一般带宽的最大利用率取 70%～80%。实验中，我们取 75%。实验结果如表 2 所列。

表 2 优化前后网络配置对照表

链路	配置带宽 (Mb/s)		带宽利用率 (%)		全网平均带宽利用率(%)	
	未优化	优化后	未优化	优化后	未优化	优化后
L1,5	320	184	25	43.5	22.5	48.7
L1,4	320	176	25	45.5		
L1,2	320	192	25	41.7		
L5,6	320	0	0	—		
L5,7	320	160	25	50		
L4,7	320	72	12.5	55.6		
L4,3	320	80	12.5	50		
L2,3	320	152	25	52.6		
L7,8	320	280	37.5	42.9		
L3,8	320	256	37.5	46.9		

由以上结果可以看出，为保障业务的 QoS，原有网络带宽配置存在浪费现象，优化后的全网带宽平均利用率提高了 2.16 倍。虽然路由采用最短路径策略，但网络流量的实际分布与理论分布会有差异。优化后的带宽是在保证 QoS 的前提下，按照各条链路的实际流量需求配置且利用率较合理，资源利用较均衡，因此不至于产生性能瓶颈，从而保证全网性能。

结束语 本文针对网络资源配置优化问题开展研究，提

出了一种基于多业务 QoS 保证的网络资源配置优化模型；应用遗传算法进行模型求解，达到网络带宽和节点缓冲区等网络资源的优化配置，从而提高网络性能；通过将种群稳定性阈值作为种群稳定性判据的方法，提高了算法的收敛效率。实验结果表明，本文模型给出的网络资源配置方案达到了资源均衡目标，全网带宽平均利用率提高了 2.16 倍。本方法有利于促进网络系统设计中资源配置的科学性。

参 考 文 献

[1] 张英，赵丽茹，谷新亮. 基于排队网络的多业务网络资源分析方法[J]. 计算机应用，2009，29(9)：2424

[2] 杨雅辉，张英，等. 网络规划与设计教程[M]. 北京：高等教育出版社，2008

[3] 席裕庚，柴天佑，恽为民. 遗传算法综述[J]. 控制理论与应用，1996，13(6)：697

[4] 同济大学概率统计教研组. 概率统计[M]. 上海：同济大学出版社，2009

[5] 林闯. 计算机网络和计算机系统的性能评价[M]. 北京：清华大学出版社，2001：290-293

[6] Wang Z，Crowcroft J. Quality-of-service routing for supporting multimedia applications[J]. IEEE Journal of Selected Areas in Communications，1996，14(7)：1228-1234

[7] Demers A，Shenker S. Analysis and Simulation of a Fair Queuing Algorithm[J]. Proceedings of ACM SIGCOMM，1989，19(4)：1-12

[8] Bennett J C R，Zhang H. WFQ：Worst-cast Fair Weighted Fair Queuing[C]// Proceedings of IEEE INFOCOM. Palo Alto，CA，August 1996：143-156

[9] 崔逊学. 多目标进化算法及其应用[M]. 北京：国防工业出版社，2006

[10] Liu De-rong，Cai Ying. A heuristic approach for measurement-based admission control with variable-size window[C]// IEEE. 2001：2537-2541

[11] 张英，张益，王冀鲁. 基于框图法的网络存储系统可靠性分析[J]. 计算机科学，2010，37(6)：102-105

(上接第 35 页)

复制成多份，分别放在网络的不同位置，请求节点只要访问到网络中的任何一份副本，都可以获得该数据，从而提高数据的可用性。但是网络的动态性增加了数据副本的维护开销，新节点的加入可能导致更好的副本存放位置，从而产生副本迁移，已有节点的退出可能导致副本的缺失。在动态环境下如何降低副本维护开销，尽量避免副本的迁移，是副本技术需要考虑的问题。在对目前已有的副本技术进行分析的基础上，提出了 ARMS 策略，它根据在线时长越长的节点，其剩余时长也往往越长的规律，在选择副本节点时，尽量选择年龄大的叶集节点，这些节点往往比较稳定，从而避免了副本的迁移。但是它也带来了副本分布过于集中的问题，使得网络中年龄大的节点往往保存了邻居中的所有副本，而年龄小的节点几乎不保存副本。本文基于此提出 RARMS 策略，该策略在选择副本节点时，首先随机选择 s 个叶集节点，然后从中选择年龄最大的非副本节点作为副本保存位置，如此循环 r 次。通过分析和模拟实验表明，该策略既能选择到稳定的副本节点，又能在一定程度上避免副本过于集中，达到预期的效果。另外，本文还对随机因子 s 的选择进行了分析，得出 s 的取值对 RARMS 策略的效果影响比较大，当 s 很小时，可以看成是随机邻居选择策略；而当 s 偏大时，可以认为是 ARMS 策略；而当 s 为 l/r 时，可以达到好的效果。

参 考 文 献

[1] Bhagwan R，Savage S，Voelker G M. Understanding Availability[C]// Proc. of IPTPS. 2003：256-267

[2] Chu J，Labonte K，Levine B N. Availability and Locality Measurements of Peer-To-Peer File Systems[C]// Proc. of ITCom：Scalability and Traffic Control in IP Networks. 2002：310-321

[3] Rhea S，Geels D，Roscoe T，et al. Handling Churn in a DHT[C]// Proc. of the USENIX Annual Technical Conference. 2004：127-140

[4] Legtchenko S，Monnet S，Sens P，et al. Churn-resilient replication strategy for peer-to-peer distributed hash-tables[C]// Proc. of SSS. 2009：485-499

[5] Ratnasamy S，Francis P，Handley M，et al. A scalable content-addressable network[C]// Proc. of the SIGCOMM'01. 2001：161-172

[6] Zhao B Y，Kubiatowicz J，Joseph A D，et al. Tapestry：An infrastructure for fault-tolerant wide-area location and routing[R]. UC Berkeley Technical Report：UCB/CSD-01-1141，2001

[7] Stoica I，Morris R，Karger D，et al. Chord：A scalable peer-to-peer lookup service for internet applications[C]// Proc. of the SIGCOMM'01. 2001：149-160

[8] Rowstron A，Druschel P. Pastry：Scalable，decentralized object location，and routing for large-scale peer-to-peer systems[C]// Proc. of the MiddleWare'01. 2001：329-350

[9] Dabek F，Kaashoek M F，Karger D，et al. Wide-area cooperative storage with CFS[C]// Proc. of SOSP. 2001：202-215

[10] Druschel P，Rowstron A. PAST：A large-scale，persistent peer-to-peer storage utility[C]// Proc. of HotOS VIII. 2001：75-80

[11] Yao Z，Wang X，Leonard D，et al. Node Isolation Model and Age-Based Neighbor Selection in Unstructured P2P Networks[J]. Proc. of IEEE/ACM Trans. on Networking，2009，17(1)：144-157

[12] Steiner M，En-Najjary T，Biersack E W. Long Term Study of Peer Behavior in the KAD DHT[J]. Proc. of IEEE/ACM Transactions on Networking，2009，17(6)：1371-1384

[13] Stutzbach D，Rejaie R. Understanding churn in peer-to-peer networks[C]// Proc. of the 6th ACM SIGCOMM on IMC. 2006：189-202

[14] Stutzbach D，Rejaie R，Duffield N，et al. On Unbiased Sampling for Unstructured Peer-to-Peer Networks[J]. Proc. of IEEE/ACM Trans. on Networking，2008，17(2)：377-390

[15] Sripanidkulchai K，Ganjam A，Maggs B，et al. The feasibility of supporting large-scale live streaming applications with dynamic application end-points[C]// Proc. of ACM SIGCOMM. 2004

[16] Sripanidkulchai K，Maggs B，Zhang H. An analysis of livestreaming workloads on the Internet[C]// Proc. of SIGCOMM IMC. 2004