

# 计算机网络服务质量优化方法研究综述

林 闯<sup>1)</sup> 李 寅<sup>1)</sup> 万剑雄<sup>2)</sup>

<sup>1)</sup>(清华大学计算机科学与技术系 北京 100084)

<sup>2)</sup>(北京科技大学信息工程学院 北京 100083)

**摘 要** 优化方法为设计更好的计算机网络服务质量保证机制提供了有力的理论支持. 相较于传统启发式的网络设计方法, 优化方法可以从理论上找到问题的最优解, 从而从根本上克服了启发式方法不能证明方案优劣程度的缺陷. 因此, 基于优化方法的机制设计与性能评价成为了当前网络服务质量领域中的一个前沿研究领域. 大量的研究着眼于从优化理论的角度重新建立网络模型, 按照优化理论给出的求解机制和实施方案设计网络协议. 计算机网络的优化可以划分为资源分配、任务调度、网络资源部署和系统参数配置等 4 方面问题. 对计算机网络服务质量的优化建模、求解、实施和评价成为当今研究的热点. 根据最新网络服务质量优化的研究进展, 文中对计算机网络服务质量研究中所涉及到的优化技术进行了研究与综述, 主要包括 4 个方面: 系统地描述了计算机网络模型优化算法的通用表达形式, 并将其按照不同的方式进行分类; 探讨了不同结构的优化模型对应的求解方案; 对比分析了不同优化算法的实施方案, 给出了方案之间的联系与区别; 归纳了计算机网络中优化方案的性能和代价评价方法. 最后, 对全文进行了总结, 并展望了进一步的研究方向.

**关键词** 服务质量; 优化模型; 资源配置; 任务调度; 性能评价

中图法分类号 TP393 DOI号: 10.3724/SP.J.1016.2011.00001

## Optimization Approaches for QoS in Computer Networks: A Survey

LIN Chuang<sup>1)</sup> LI Yin<sup>1)</sup> WAN Jian-Xiong<sup>2)</sup>

<sup>1)</sup>(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

<sup>2)</sup>(School of Information Engineering, University of Science & Technology Beijing, Beijing 100083)

**Abstract** Optimization approaches provide powerful theoretical support for designing mechanisms to guarantee quality of services (QoS) in computer networks. In contrast to heuristic methods for network design, optimization approaches can find out the optimal solutions theoretically, so that overcome the drawbacks of heuristic methods which cannot be proven to be optimal. It becomes the leading edge of the QoS-aware networking research field. Many researches begin to use optimization theory to reanalyze network models, and design protocols according to the solutions of optimization problems. Optimization approaches in networks can be categorized into several different subjects: resource allocation, task scheduling, resource deployment, and parameter configuration. The modeling, solving, implementing, and evaluating techniques for optimization of QoS in networks have become a hot research topic nowadays. According to the latest research progress, this paper delves into the optimization techniques for improving QoS in networks. It includes four parts: formulating the generic expressions for optimization of networks and classifying the optimization problems according to different perspectives; discussing solving methods to the models with different structures; comparing diverse ways of implementations by considering

收稿日期: 2010-11-26. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2010CB328105, 2009CB320504)、国家自然科学基金重点项目(60834004, 60932003, 60673187, 60702009, 60970001)资助. 林 闯, 男, 1948 年生, 博士, 教授, 博士生导师, 主要研究领域为计算机网络、系统性能评价、安全分析和随机 Petri 网. E-mail: chlin@tsinghua.edu.cn. 李 寅, 男, 1986 年生, 博士研究生, 主要研究方向为计算机网络性能评价、面向服务计算、Petri 网建模. 万剑雄, 男, 1982 年生, 博士研究生, 主要研究方向为计算机网络性能评价、航空电子网络、Petri 网建模.

the connection and difference between them; presenting methods for performance and cost evaluation of optimization schemes. Finally, we summarize the points of this paper and discuss the possible research challenges.

**Keywords** QoS; optimization models; resource allocation; task scheduling; performance evaluation

## 1 引言

随着计算机网络业务的逐步拓展,各种业务对网络的需求也在不断提升.为保证网络服务正常有效的部署,需要尽量提高或保障特定业务中关键的服务质量(QoS).相关的 QoS 评价指标主要包括网络的吞吐率、时延、丢包率等.它们可以被看成网络协议的效用.在资源有限的网络系统中,其对偶问题是最小化网络运行的代价.为了进一步提高网络的 QoS,不仅需要评价现有的网络协议,而且需要进行网络的再造工程,用优化理论对网络系统进行建模分析,得出使网络性能达到尽可能令人满意的机制.

不同的网络系统和业务的运行机制各异,但优化模型的初衷往往是相近的.如果将网络服务看成某种资源,则优化问题其实可以归结为优化资源配置.更细致的划分可将网络优化模型分为:

(1)资源分配.如链路带宽分配问题、队列空间分配问题等.

(2)任务调度.任务调度问题可以分为单队列多服务器调度、多队列单服务器调度和更一般的多队列多服务器的调度等.

(3)网络资源部署.如确保网络连通性的互联设备放置问题、在服务器集群中最大化资源利用率的软件放置问题以及最小化成本的服务器覆盖问题.

(4)系统参数配置.如传输节点的功耗配置、拥塞控制窗口的大小配置.

优化模型的结构决定了解决优化问题所采用的算法.算法中最重要的划分即是易解问题和 NP-难问题,这引导我们寻找优化方案的方向,抑或精确地保证最优的服务质量,抑或尽力而为得到次优的优化方案.如何将设计的优化算法部署到实际网络运营的控制中也是值得探讨的问题,因为算法的部署方式会直接影响网络运行的性能.它不仅是工程技术问题,也与优化理论本身相关.例如,我们更希望得到能够分布并行部署的算法,来减少每个网络节点的负载,并迅速地得到优化方案.网络的再造工程是一个循环的过程,评价-优化-评价-再优化地循环演进逐步提升网络系统的服务质量,而对网络优化

算法的评价为网络性能的不断改善提供了参考.

本文综述了优化方法在提高网络服务质量中的应用,主要包括 4 个方面:系统地描述了网络模型优化算法的通用表达形式,并将其按照不同的方式进行分类;探讨了不同结构的优化模型的求解方案;对比分析了不同网络优化算法的实施方案,给出了方案之间的联系与区别;归纳了网络优化方案的性能和代价评价方法.最后,对全文进行了总结,并提出了进一步的研究方向.

## 2 网络 QoS 优化模型表达与分类

网络系统旨在对用户尽可能高的服务质量,网络技术也在为可以接纳更多的潜在用户和提高用户感受而逐步演进.

图 1 描述了网络服务质量的优化目标和步骤.

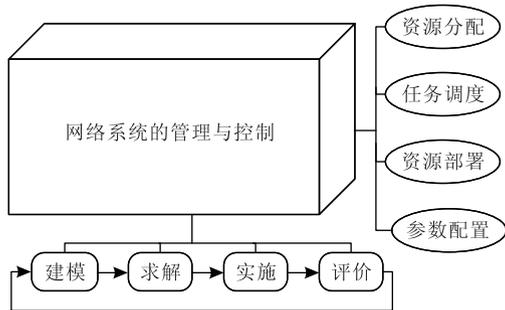


图 1 网络服务质量的优化目标和步骤

整个网络系统可以看作是一个拥有生产者和消费者的系统,生产者对应服务主体,提供相对有限的资源;消费者对应用户,利用生产者分配的资源.例如,对于无线资源而言,有限的带宽不能满足日益增长的无线业务和用户带来的需求.这不仅促进了新的策略和技术的提出,如针对受限无线带宽资源提出的动态频谱访问技术<sup>[1]</sup>,同时,在该技术背景下的资源配置问题得到了越来越多研究人员的重视,其更致力于解决资源供不应求的问题.资源配置问题蕴含一系列 QoS 指标,资源配置的有效性决定网络的利用率,分配的均衡程度决定网络的公平性,而这些 QoS 指标统一构成了每个用户和整个网络社会的效用.

网络中的任务调度是网络优化中的另一个关键

因素.任务执行和数据传输时序的不同对网络用户个体和网络整体的影响可能很显著.负载均衡调度是网络中最基本的调度问题,它可以被建模为以最小用户平均时延为目标的优化问题.实际上,调度问题可以被视为一种特殊的资源分配问题,例如以最大化网络吞吐率的网络流调度策略实际上在进行着网络信道的资源分配,亦可以将其看作为以时间为资源的分配问题,使分配的资源达到最少,资源的利用率最高.可见,资源分配和任务调度之间并没有不可逾越的鸿沟.可以将任务调度看作为在有限资源的约束下实现最大化 QoS 目标的决策过程,这恰恰对应着网络优化模型的建立.

网络资源部署可以归结为研究网络平面体系结构的问题,它主要探讨如何将资源合理地进行安置(Location/Placement),例如网络中服务器的放置、无线传感器网络中设备的布置等.该类问题的目标同样是面向更多的用户或提供更高的性能.

还有一类网络优化是针对网络系统的参数配置,从而指导网络参数的设计.在蜂窝网络中移动基站的传输功率大小将影响信噪比,如何决定功率的参数配置就是一个典型的网络参数配置优化问题.

## 2.1 基于四元素的网络 QoS 优化模型表达

网络 QoS 优化模型包含 4 个元素:优化目标、决策变量、约束条件和固定参数.

(1) 针对不同的研究对象,目标函数的表达不同,可以分为终端用户关心的目标和网络运营商关心的目标.

(2) 决策变量往往对应调度策略,变量的个数决定着网络系统设计的自由度.

(3) 约束条件限制了可供调度的资源容量或规定了一些策略上和技术上的限制.

(4) 固定参数是系统的常量,是网络中已经约定的不变因素.

文献[2]按网络中的不同层次列举了一些可能的无线网模型元素,如固定参数,包括传输能量、服务速率、源编码速率和包到达速率;目标函数有最小化总体能量、丢包率、延迟、传输成本,最大化总体吞吐量、利润等;约束条件包括传输能量的限制、信道访问的竞争、对延迟和安全的严格保证;而变量包括为每个用户分配的带宽、传输的功率等.除资源约束外,网络中数据传输的有序性保证需要也会对任务时序关系强加约束条件,例如无线传感器网络的数据聚合对数据时序的限制,文献[3]在该约束下试图最大化请求的性能和能量的效率,文献[4]在任务时序约束下最小化无线节点的能量消耗.图 2 描述了网络优化模型的对应元素.

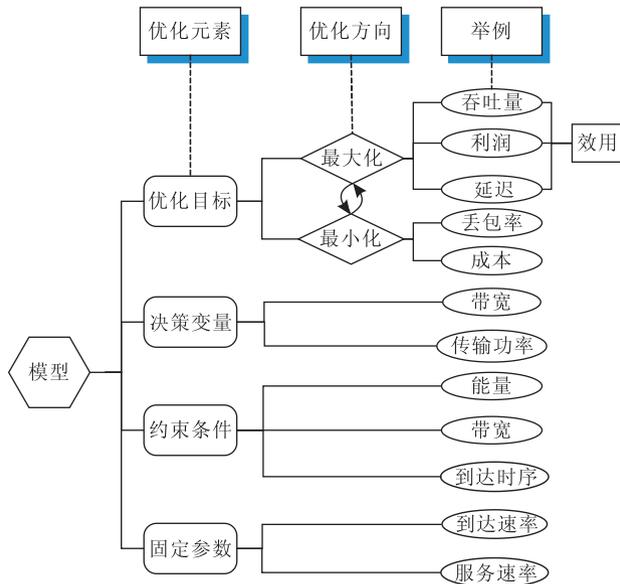


图 2 网络优化模型的元素

最通用的目标函数定义为效用函数(Utility Function),它可以是客观的物理指标,也有可能掺杂一定的主观因素.

(1) 客观的效用函数.这类效用函数均可以严格地使用数学表达式定义,如网络吞吐量、分组平均延迟、链路利用率以及网络的能源使用量等.

(2) 掺杂主观因素的效用函数.效用函数有时用来表达用户对网络服务的满意程度.由于不同用户对 QoS 的要求各异,如有些用户对延迟敏感,而有些用户可以容忍一定范围内的延迟,因此这类效用函数通常掺杂主观因素.

效用函数通常是单调递增的凹函数,即性能指标改善的边际效应随决策变量逐渐削弱.这满足边际效用逐渐减小的实际情况,也为将最大化效用问题转化为凸优化问题提供了理论基础.

通用的优化形式可以表达为

$$\begin{aligned} & \underset{x}{\text{minimize}} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i=1,2,\dots,m \\ & && h_i(x) = 0, \quad i=1,2,\dots,p \end{aligned} \quad (1)$$

或

$$\begin{aligned} & \underset{x}{\text{maximize}} && U(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i=1,2,\dots,m \\ & && h_i(x) = 0, \quad i=1,2,\dots,p \end{aligned} \quad (2)$$

事实上,上述两种表达形式是同构的,式(1)很容易化为式(2)的形式,只需令  $f_0(x) = -U(x)$ .利用下一节阐述的对偶方法也可以进行最大化 and 最小化目标的转化.

一类网络优化问题的目标函数有 max-min 或 min-max 的形式,例如研究网络负载均衡或公平调

度之类的问题<sup>[5-6]</sup>. 应用博弈论分析和设计网络调度策略时, 会经常遇到这种形式的目标函数, 它其实可以拆分为两个优化步骤. 例如, 自组织网络中的生存时间的优化可以表达为最大化生存时间, 而生存时间又可表达为网络所有节点中最小的生存时间<sup>[7]</sup>. 实际上, 这类最小最大优化目标通常可以通过引入新变量转化为通用的求最大或求最小值的优化问题.

文献[8]提到了很多优化模型表达的技巧, 许多技巧都在网络优化建模上得到了应用. 例如指示变量(indicator variable)<sup>[6]</sup>或函数(function)<sup>[9]</sup>可以表示事件的发生与否, 如是否发生信道干扰等. 指示变量往往辅以归一化条件, 表示有且仅有一个事件发生, 也可以方便地表达两个事件的互斥关系.

同一个优化指标可能对应有多种不同的目标函数, 其中的侧重点或需求各异. 例如, 文献[6]考虑了基于合作机会调度方法的频谱访问问题, 针对带宽分配给出了 3 种表达形式:

(1) 最大和带宽(Max-Sum-Bandwidth, MSB):

$$\max_{a_{n,m}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} a_{n,m} \cdot b_{n,m} \quad (3)$$

(2) 最大最小带宽(Max-Min-Bandwidth, MMB):

$$\max_{a_{n,m}} \min_{n < N} \sum_{m=0}^{M-1} a_{n,m} \cdot b_{n,m} \quad (4)$$

(3) 最大比例平均带宽(Max-Proportional-Fair, MPF):

$$\max_{a_{n,m}} \sum_{n=0}^{N-1} \log_{10} \left( \sum_{m=0}^{M-1} a_{n,m} \cdot b_{n,m} \right) \quad (5)$$

其中,  $a_{n,m}$  为指示变量, 为 1 代表将频带  $m$  分配给用户  $n$ ;  $b_{n,m}$  代表将频带  $m$  分配给用户  $n$  后用户  $n$  获得的收益.

## 2.2 优化模型方法分类

网络中的优化对象具有不同的特征, 对同一个研究对象也会存在多种模型方法. 依据不同的假设, 这些方法从不同的角度对网络性能指标进行改进. 随着数学理论不断发展, 求解同一模型的方法可能存在多种, 从而对网络机制产生了不同的解释. 这些求解算法也对应不同的资源分配、调度、部署和参数配置的方案. 下面依据模型、求解和实施方案 3 方面对网络优化模型进行分类.

### 2.2.1 按照模型的函数及变量类型分类

(1) 单目标优化与多目标优化

网络优化常常是对某一 QoS 指标而言, 将其它 QoS 属性作为限制, 也就是常见的单目标优化. 如果考虑的性能指标有多个, 就需要用多目标优化建模. 在许多组播应用中, 如声音、视频会议、在线游

戏等, 需要带宽、包延迟、包丢失率和传输成本等多种 QoS 需求. Donoso 等人<sup>[10-12]</sup>研究了将这些需求综合起来进行多目标优化. Wang 等人<sup>[13]</sup>研究了无线传感器网络多对一传输模式中的多目标 TDMA 调度模型, 考虑了平均延迟和能量消耗两个目标的帕累托前沿(Pareto Frontier). 单目标优化和多目标优化在某种意义上可以相互转化, 多目标通常利用归一加权或添加约束的方法转化为单一目标, 而带约束的单目标优化也可利用罚函数, 转换为多目标的加权形式.

(2) 连续变量优化与离散变量优化

模型中变量的类型会决定优化模型的种类, 在网络参数配置的优化问题中, 参数的范围通常是连续变化的, 如无线通信中的功率控制. 对于有连续取值的资源分配其变量也是连续的. 有时, 模型中的变量取值范围是一个有限可列集合, 这种变量称为离散变量. Awad 等人<sup>[14]</sup>研究了基于 OFDMA 的两跳合作中继网络的资源分配, 采用了二元整数规划对将多个中继站按对应关系分配给基站的问题进行建模. 整数规划可被应用到调度选择决策中. 例如, 路由选择问题可以将某条链路是否被包含在路由中设置为离散变量, 从而进行整数规划<sup>[15]</sup>. 资源部署也通常是离散优化问题, 因为表征是否在某个位置上部署资源常用到二元指示变量. 无线传感器的部署是其中的典型例子. Berry 等人<sup>[16]</sup>用整数规划的方法研究如何部署传感器检测供水管网系统中的恶意污染.

(3) 线性规划与非线性规划

目标函数和约束函数的类型会对优化模型的求解方法带来启发, 很多优化的研究工作按照函数类型将优化模型进行分类, 最简单的分类方式是线性规划(linear programming)与非线性规划(non-linear programming). 网络优化中的很多问题都可以归结为线性规划问题, 尤其是涉及流量控制的优化, 很早, 流量相关的 QoS 路由算法就开始利用线性规划模型, 并且在 20 世纪 70 年代就达到有效的结果<sup>[17]</sup>, 直到今天仍然被广泛运用. 整数线性规划的函数类型也是线性的, 但由于它的变量类型是离散的, 因此求解算法也不同于线性规划. 最近几年, 非线性在网络系统优化中越来越受到重视, 这一方面由于优化求解技术不断发展, 另一方面也因为简单的线性模型不足以表示网络元素的关系, 因此, 非线性优化广泛被应用到高速核心网络、无线自组织网络、编码与均衡理论、宽带访问、信息论和网络拓扑模型中<sup>[18]</sup>.

(4) 凸优化与非凸优化

随着优化理论研究的不断深化, 人们发现如果一个词来区分简单和难问题, 那么模型的凸性大

大概就是“分水岭”<sup>[19]</sup>. 幸运的是,大多数网络优化模型都保持了凸性,即在约束为凸集的条件下最小化一个凸目标函数或最大化一个凹目标函数. 线性规划就是保持这种凸性的简单模型. Luo 等人<sup>[20]</sup>综述了在通信方向的关于凸优化的研究,并且强调了一类锥优化(Conic Programming, CP)问题,包括二次锥规划(Second Order Cone Programming, SOCP)和半正定规划(Semidefinite Programming, SDP),启发了网络优化建模,尤其是无线通信网络的建模. 随着凸优化求解方法的逐渐成熟,越来越多的网络问题被建模为凸问题,很多不具备凸性质的模型也开始利用诸如变量转化等方法转化为凸优化,从而可以进行有效求解. 在一些应用中,函数的复杂性有时会破坏问题的凸性,Chiang<sup>[19]</sup>讨论了三类非凸优化问题的应用,包括利用最大化非凹性网络效用函数进行网络拥塞控制、通过几何和反曲规划进行无线网络功率控制以及利用分布式非凸优化进行 DSL 频谱管理. Chiang 将网络通信建模中的非凸优化归结为 4 类:最小化非凸函数(最大化非凹函数),例如非弹性应用的拥塞控制;非凸约束集,例如在低 SIR 范围的功率控制;整数约束,如单一路径路由和多用户检测;凸性不等式约束数量的指数增加,例如最优调度.

#### (5) 确定参数优化与随机参数优化

在确定性模型中,网络优化模型的参数是一个常数. 但是,在随机优化模型中,参数可以是一个随机变量. 目标和约束函数可能包含该随机变量相关的数字特征或概率特征. 对于具有突发性和随机性的网络而言,随机优化模型更能捕捉诸如包到达速率和包处理速率等不确定因素,但这又无疑带来了模型求解的复杂性. 李雅普诺夫稳定性常被用来研究随机网络优化问题,得出某些性能指标的界<sup>[21]</sup>.

网络优化模型按照目标和变量的类别分类如图 3 所示.

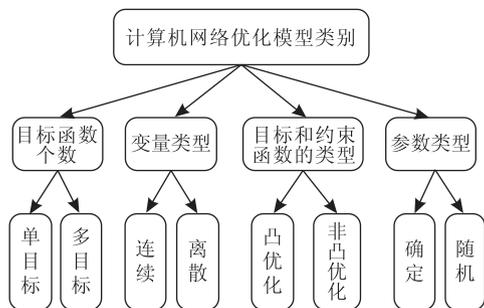


图 3 网络优化模型类别

#### 2.2.2 按照模型的求解方法分类

具有不同函数和变量类型的优化模型都对应不

同的求解算法. 下面我们按照求解算法对网络中的优化进行分类.

#### (1) 易解问题与难问题

在建立优化模型后,需要设计算法进行求解. 求解的效率是重要的指标. 我们通常要求解算法具有快速收敛的性质和多项式时间的复杂度. 因此,可以将优化模型按复杂程度分为 NP-难问题和易解问题. 凸优化问题没有对偶间隙,可以基于梯度算法几何地收敛到全局最优解,具有多项式时间求解算法. 而非凸优化问题可能存在严格的对偶间隙,通常是 NP-难问题. NP-难问题不可能在多项式时间找到全局最优解,往往采用现代优化算法对解空间进行搜索,或是采用近似算法求得一个满意解.

#### (2) 确定性算法与随机算法

一般优化问题的求解算法是确定的,在不能将问题划归到一个已知可解模型类别,或解空间过于庞大没有有效的求解方法时,常常利用随机算法. 现代优化方法中有很多含有随机的步骤. RED 算法<sup>[22]</sup>可以被看作是 TCP 流量控制的一个随机算法, Krishnan 等人<sup>[23]</sup>研究了基于 WLAN 结构下的吞吐量最大化问题,提出了一种叫做 wTOP-CSMA 的随机算法,能够在全连接网络中提供最优的吞吐量,即使隐藏终端存在.

#### 2.2.3 按照模型的实施方案分类

#### (1) 静态实施与动态实施

在确定性模型中,参数是固定的. 如果所有有关网络结构和动态特征都反映在这些参数上,并且所有参数都作为已知数据,那么总可以设计一种静态调度或分配算法. 其对应的实施方案也很简单,只需按照这种算法执行,而不考虑网络的即时状态. 这种优化算法是固定的,又称为离线算法. 它不随时间转移,因此不适用于动态随机性显著的网络环境. 在上文提到的随机优化模型中,通常需要设计动态算法,因为需要根据当前得到的状态信息,如排队长度、丢包率等进行调度决策,这些参数随时间不断变化,而并非在一开始就能准确地获得. 最著名的动态调度算法是在负载均衡或者最大吞吐率目标下的路由器调度,可以采用最短队列优先、最短预计完成时间优先等带有状态的算法. 动态的优化算法又称为在线算法. 需要说明的是,即使模型被表达为静态的,实际上参数也可能是不稳定的,只是在一段时间内近似有一个平稳的趋势. 当参数取值发生跳跃时,需要修改原来模型中的参数,重新求解优化模型. 这种折中方案综合了静态优化的简单易解和动态优化准确实时的优点.

## (2) 集中式实施和分布式实施

不仅优化算法的实施时间可能不同,而且其实施的位置也可以变化.可以采用集中式的方案实施优化算法,这通常需要统一地控制节点,监测网络的参数,并做出优化的分配或调度决策.这种集中式优化实施方案有两个缺点:实施算法的节点工作量大,使得算法运行的时间显著增加,从而带来整个网络系统的延迟增大;单节点的控制方案易受到攻击,存在单节点脆弱性.这使得人们在可能的情况下,尽力寻找分布式算法,即将整个优化算法并行地在多个网络节点上实施,每个节点只执行与该节点相关的优化子问题,然后通过相互传递子问题的解进一步得到原优化问题的最终方案.分布式算法以增加通信代价带来了合作并行求解优化模型的优势,提高了算法实施效率和安全性.

按照优化算法的实施方案分类如图 4 所示.

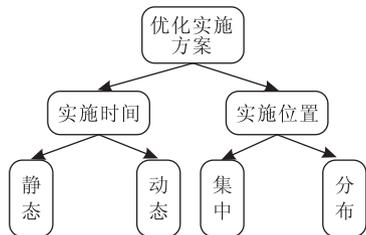


图 4 网络优化模型实施方案

## 3 网络优化模型的求解算法

由于网络节点的处理资源有限,许多网络部署又没有一个专门的集中控制节点,因此一般情况下,网络中的节点肩负着控制层面与数据层面的双重功能,负担较大.为达到实时有效的控制和性能优化,需要设计占用资源少、复杂性低的算法.尤其是本身携带能量较少,并且不便于频繁补充能量的节点,如无线传感器,或者实时性要求较高的网络业务.这就要求不仅提高网络性能需要优化算法,而且优化算法本身也需要优化.许多问题如果能将模型转化为众所周知的易解模型,如线性规划问题,即可以利用成熟的求解理论在多项式时间内找到最优方案;否则,很有可能属于 NP-难问题,这时只能去寻找近似算法,并保证近似算法本身开销少,而且近似方案一定程度上接近理论上的最优方案.

### 3.1 易解问题与多项式时间收敛算法

#### 3.1.1 凸优化方法

网络中的很多优化模型都可以归结为凸优化问题,如线性规划、二次规划等.随着研究的深入,人们发现判断一个问题是否具有凸性质在一定程度上决

定了寻找最优方案是否是易解的,这是因为对于凸优化问题而言,局部最优解即为全局最优解.有许多关于网络效用最大化(NUM)<sup>[24]</sup>的研究就是建立在求解凸优化问题的基础上.凸优化模型具有如式(1)的形式,且目标函数是凸函数、约束集是凸集,等式约束函数是仿射函数.对于最大化网络效用的模型,目标函数是凹函数.

对于无约束的凸目标函数优化来说,只需要利用一阶条件找到其一阶函数为 0 对应的变量值即为最优方案,若利用数值方法,则可以用梯度法或牛顿法逼近其最优解<sup>[25]</sup>.在有关调度和资源分配的问题中,由于有资源的限制,或者技术的要求,需要附加限制集.考虑如下较通用的调度控制模型<sup>[24]</sup>,网络中有  $L$  条链路,相应的容量为  $c_l$ ,  $S$  个源节点,每个以  $x_s$  的速率在一个固定的链路集合上传输,传输获得的效用函数为  $U_s(x_s)$ .假设每个节点容忍的最小传输速率为  $x_s^{\min}$ :

$$\begin{aligned} & \text{maximize}_{x_s} \quad \sum_s U(x_s) \\ & \text{subject to} \quad \sum_{s: l \in L(s)} x_s \leq c_l \quad \forall l \\ & \quad \quad \quad x_s \geq x_s^{\min} \quad \forall s \end{aligned} \quad (6)$$

若假定每个节点的效用函数都为凸函数,则整个目标函数即为凸函数,又易知线性约束形成的可行解的集合是一个多面体,因此上述模型是一个典型的凸优化模型,局部最优解即为全局最优解.在求解时,需要将其划归为无约束优化问题,从而方便地用梯度法或牛顿法进行迭代计算.障碍法(Barrier Method)<sup>[26]</sup>为每个不等式约束设计一个障碍函数,当满足约束不等式,该函数趋于 0,当约束趋于使等号成立时,函数趋向无穷,然后将障碍函数加到原目标函数中.为了适合用牛顿法求解目标函数,经常使用二次可微的对数障碍函数.上式(6)用障碍法可化为

$$\begin{aligned} & \text{maximize}_{x_s} \quad \sum_s U(x_s) + \\ & \quad \quad \quad \sum_l (1/t) \log(c_l - \sum_{s: l \in L(s)} x_s) + \\ & \quad \quad \quad \sum_s (1/t) \log(x_s - x_s^{\min}) \end{aligned} \quad (7)$$

可以证明当  $t$  很大时,该方法的解趋于原问题的最优解,事实上,与真实的理论最优值的差距不会超过  $m/t$  ( $m$  为约束条件中不等式的个数,这里  $m=s+l$ ).在求解过程中, $t$  的值应从小逐渐取到大,逐次进行迭代,折中收敛速度与求解式(6)的复杂度.将原问题进行修改,设链路容量依赖于所分配的资源,如传

输时槽、功率或带宽,则容量可以表示为资源的函数  $c_l(\varphi_l)$ <sup>[27]</sup>,资源的总和是固定的,则模型变为

$$\begin{aligned} & \underset{x_s}{\text{maximize}} \quad \sum_s U(x_s) \\ & \text{subject to} \quad \sum_{s: l \in L(s)} x_s \leq c_l(\varphi_l) \quad \forall l \\ & \quad \quad \quad \sum_l \varphi_l = \varphi_{\text{tot}} \\ & \quad \quad \quad x_s \geq x_s^{\min} \quad \forall s \\ & \quad \quad \quad \varphi_l \geq 0 \quad \forall l \end{aligned} \quad (8)$$

当将上式的不等式用对数障碍函数表示后,仍旧留有一个关于资源总和的等式约束,可以通过将梯度约束在等式成立的方向上实施梯度或牛顿法逼近(投影梯度法<sup>[26]</sup>).实际上,可以将式(8)中的等式约束变换为不等式约束,即所分配的资源不能超过资源上限,然后单纯地用障碍法去求解,很显然,得到的解仍然是原问题的最优解.可以看出,采用障碍法和投影梯度法进行迭代,每次更新的变量值都是一个可行解,并且逐渐逼近到最优值.障碍法关于  $t$  线性收敛,20世纪80年代提出了原对偶内点法(Primal-dual interior-point methods)<sup>[28]</sup>加速了求解的收敛,更为有效.它受障碍法的启发,将牛顿法运用到求解修改的 KKT 方程中,同时迭代原变量和对偶变量.值得一提的是,内点法迭代过程中得到的原变量和对偶变量不一定在可行域内,但最后收敛得到的解却是可行解.

由于网络中的实际凸优化问题有很多都在可行域内存在严格的可行解(即满足 Slater 条件<sup>[26]</sup>),因此具有强对偶性(strong duality),即原问题的解与对偶问题的解之间没有间隙(duality gap).于是,求解原问题就可以等价于求解相应的拉格朗日对偶问题.拉格朗日对偶放松了原问题的约束,并将其转化为加权的目标函数<sup>[24,26]</sup>,实施对偶时,有时只针对部分约束,如式(6)的拉格朗日函数为

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= \sum_s U_s(x_s) + \sum_l \lambda_l \left( c_l - \sum_{s: l \in L(s)} x_s \right) \\ &= \sum_s \left[ U_s(x_s) - \left( \sum_{l \in L(s)} \lambda_l x_s \right) \right] + \sum_l c_l \lambda_l \end{aligned} \quad (9)$$

对偶函数为

$$g(\boldsymbol{\lambda}) = \max_{\mathbf{x} \geq \mathbf{x}^{\min}} L(\mathbf{x}, \boldsymbol{\lambda}) \quad (10)$$

则对偶问题即为

$$\begin{aligned} & \text{minimize} \quad g(\boldsymbol{\lambda}) \\ & \text{subject} \quad \boldsymbol{\lambda} \geq 0 \end{aligned} \quad (11)$$

对偶问题可以用梯度法迭代求解

$$\lambda_l(t+1) = \left[ \lambda_l(t) - \alpha \left( c_l - \sum_{s: l \in L(s)} x_s^*(\lambda_l(t)) \right) \right]^+ \quad \forall l \quad (12)$$

$x_s^*$  为对偶函数(10)取到最优值时的变量值,  $\alpha > 0$  是一个足够小的正步长.

### 3.1.2 整数规划的多项式算法

网络资源的安置与选择等很多问题中都存在离散决策变量.由于离散的点分布破坏了可行域的凸性,一般意义上不能在多项式时间内找到全局最优解.而在通信网络中,由于资源的离散化配置,派生出许多的离散优化问题.网络中最经典的优化问题是路由选择问题,单纯考虑时延时,路由选择则可以模型化为一个最短路径问题.此外,桥式 LAN 可用最小生成树模型进行优化从而保证配置代价最小<sup>[29]</sup>. 这些问题通常可以归结为整数规划模型(Integer Programming),并写成上文提到的通用的表达方式.不同的是,模型中的决策变量全部或部分有整数限制.只有一部分变量为整数的优化模型被称为混合整数规划模型(Mixed Integer Programming). Jia 等人<sup>[30]</sup>分别用整数规划和混合整数规划模型分析了无线网络通过调节传输功率改变网络传输的连通性,在保证端到端流量需求和传输时延的约束下,最小化能量消耗.对于不能拆分的流量需求,用 0-1 变量表示某条链路是否被使用,一旦被使用,将传输所有流量,这样问题被描述成一个整数规划问题;而对于可以拆分的流量需求,不仅用 0-1 变量表示某条链路是否被使用,而且用一个实值变量表示链路上传输的流量,所有的流量需求会进行分流并通过不同的路径进行传输,在这种情形下,模型化为一个混合整数规划.

一般情况下,整数规划和混合整数规划问题很多是 NP-难问题,这与模型的结构相关.其中,有一类问题可以通过图论的技巧在多项式时间内找到最优解,例如网络流问题.最大流可用来求解在有向传输网络中,在链路容量限制和节点流守恒限制下,如何调度流量需求,可以使传输的流量总和达到最大.它常被用来求解网络的整体传输能力,并验证是否能够满足一定的流量需求.文献[31]用该方法检验多跳无线网整合节点的放置方案是否满足流量需求.最大流的求法是在残余网络中不断地寻找增广路径,直到增广路径不存在为止.当流量取离散整数值时,按照此方法得到的解也是整数的,并且是最优解.最小费用流是在链路容量限制和节点流守恒限制的基础上,给每条链路规定一个传输代价,并且预先指定每个源节点需要传输的最小流量和每个目的节点需要接收的最小流量.最小费用流算法在此基础上给出所花代价最小的方案,求解过程有消圈法、

连续最短路径法和对偶法. 如果不指定最小流量, 而是要求必须达到网络中可以传输的最大流量, 这时求最小费用问题就变成最小费用最大流问题. 网络中最短路径的求解方法根本上基于松弛算法, 如 RIP 路由协议采用 bellmen-ford 算法进行路由选择, 它是距离路径算法的一种, 而 OSPF 路由协议是一种典型的链路状态路由协议, 它采用 Dijkstra 算法完成最短路由的选择. 最小生成树算法可以应用在桥式 LAN 的配置问题上, 它基于拟阵结构, 可以设计出几种贪心算法, 其中包括 Kruscal 算法和 Prim 算法. 此外, 还有最大权二分图匹配算法, 它可以看成资源-消费者的分配问题, 每种资源只能分配给一个消费者, 每个消费者只能接收一种资源. 对于一对分配方案〈资源, 消费者〉, 消费者可以获得相应的收益. 求解整个资源的分配方案使所有消费者的总体收益最大即是最大二分图匹配算法, 它可以用匈牙利算法进行求解. 值得一提的是, 上述算法很多都可以归结为最小费用流问题: 当供给和需求流量为 1, 且链路传输代价也为 1 时, 最小费用流就化成了最短路径问题; 而当供给和需求流量为 1, 链路的代价为将资源分配给某个消费者的收益的负值时, 最小费用流问题转化为总收益的最大权匹配问题. 上述所有算法, 当可行域为整数点的集合时, 算法依然成立, 并可以证明算法在多项式时间内收敛到最优解.

### 3.2 NP 难问题与近似算法

网络中许多优化问题是 NP-难的, 在多项式时间内收敛的优化算法目前不存在. 文献[32]的报告提出了无线 Mesh 网中保证“完美”QoS 的挑战, 其中涉及了许多 NP-难问题: “完美”的零抖动分组调度是 NP-难问题, “完美”的无线 Mesh 网调度问题是 NP-难问题, “完美”的路由(达到 100% 的容量)是 NP-难问题……“完美”问题难于求解, 难于扩展, 也为我们力图找到“近似完美”的方案提供了在时间和效果上折中的动机. 解决此类问题可以分为两个步骤: 难问题的规约和难问题的近似.

#### 3.2.1 NP-难问题规约

图 5 给出了几个关于 NP 概念之间的关系. 如果能证明一个 NPC 问题能够在多项式时间内规约为一个给定问题, 则给定问题即为 NP-难问题<sup>[33]</sup>. 如果要求证明该给定问题也是 NPC 问题, 只需附加证明该问题是 NP 问题即可(即可以在多项式时间以确定性算法验证一个解). 从 Cook 最早证明电路可满足性问题属于 NPC 类开始, 已经发现许多 NPC 问题, 如背包问题、哈密尔顿路径、子集合问

题、团问题、顶点覆盖以及独立集问题等. 这些问题之间有很多本身可以用规约的方法证明其属于 NPC 类.



图 5 P、NP、co-NP、NP-难、NPC 之间的关系

在实际处理中, 只需证明某个已知的 NPC 问题能够规约为原问题的一个具体实例, 并且规约过程是在多项式时间内完成的. 文献[34]研究了如何在 CDN 网络中备份数据对象, 使得当用户从最近的 CDN 服务器中获取所需数据时, 遍历的自治系统数量最小. 每个 CDN 服务器的容量是有限的, 因此这个问题是典型的资源受限问题. 原问题可以由背包问题规约, 从而证明其属于 NPC 类. 事实上, 这类资源分配问题常常可以由类似的带约束的调度问题进行规约, 规约的过程也比较直观. 还有一类典型的规约问题是基于经典图论问题的, 例如, 文献[35]将探讨的无线传感器网络节能问题表达为一个连通性的图论问题: 在给定的一个传感器集合中, 计算每个传感器的传输功率, 使得每对传感器都是连通的, 且尽可能使总传输功率最小. 这个问题也是一个典型的带约束最优化问题. 文章使用了链式规约的技巧, 先由 4 平面顶点覆盖问题(4PlanarVC)规约到 3 平面顶点覆盖问题(3PlanarVC), 证明 3PlanarVC 是 NPC 问题, 由此规约 3rPlanarVC 问题(3PlanarVC 的特例, 保证每两个顶点之间的距离不小于 3)是 NPC 问题, 最后规约到原问题是 NPC 问题. 图论难问题的规约关键在于寻找一个与已知问题有类似结构的原问题的实例.

此外, 还有结论性更强的规约, 如 L-规约. 它被用在证明一个问题是 MAX SNP-难问题, 即不存在在多项式时间近似方案(PTAS)的问题. 为了证明一个优化问题 B 是 MAX SNP-难的, 给出从一个已知的 MAX SNP-难问题 A 到问题 B 的 L-规约就可以. 这个性质被用来证明原问题是一类很难被近似的问题. 文献[36]研究了 LTE 下行链路多输入多输出-频域分组调度问题的计算复杂性, 通过任务区间选择问题(JISPk)进行 L-规约, 证明了该调度问题也是 MAX SNP-难的. 此外, 文献[37]推荐了除 L-规约之外的两类近似保持规约: AP-规约和 PTAS-规约, 可以证明更多的结论.

#### 3.2.2 难问题的近似算法

为了在有限时间内实施优化算法, 通常采用近

似的方法. 固定模式的近似算法在实际中经常使用, 因为它们便于设计, 计算也简便. 最简单的一类近似算法是贪心算法, 它只关注于当前的局部优化, 并期望其对全局目标带来益处. 例如, 文献[36]采用了两种近似策略进行 LTE 上行链路的频域分配, 一是贪婪策略, 二是基于 Local Ratio 技术<sup>[38]</sup>设计的, 该技术给出了求解一类在满足特定约束下, 为达到最小或最大权重和而进行子集选择的 NP 问题的近似方法. 虽然有些近似方案看起来没有固定的章法, 但它的设计往往和规约的步骤和规约后的问题相关, 如由图论规约的难问题近似算法也常常涉及图论算法, 文献[35]就是基于最小生成树的算法得到了功率分配的近似方案.

很多 NP-难问题可以表示为整数线性规划问题, 于是与凸优化问题类似的原-对偶方法(Primal-dual)就被改造用来设计组合优化的近似算法. 由于原-对偶方法通用的模式, 它成为一种设计近似算法通用的技术. 采用其它方法, 如贪婪策略设计的很多近似算法也可以用原-对偶方法来解释<sup>[39]</sup>. 对于如下的经过变量松弛(去掉整数限制)后的原问题:

$$\begin{aligned} & \underset{x}{\text{minimize}} && \mathbf{c}^T \mathbf{x} \\ & \text{subject to} && \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ & && \mathbf{x} \geq \mathbf{0} \end{aligned} \quad (13)$$

和对偶问题

$$\begin{aligned} & \underset{y}{\text{maximize}} && \mathbf{b}^T \mathbf{y} \\ & \text{subject to} && \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \\ & && \mathbf{y} \geq \mathbf{0} \end{aligned} \quad (14)$$

二者对应的原互补松弛条件(primal complementary slackness conditions)为

$$x_j > 0 \Rightarrow \mathbf{A}^j \mathbf{y} = c_j \quad (15)$$

其中 $\mathbf{A}^j$ 为 $\mathbf{A}$ 的第 $j$ 列. 对偶互补松弛条件(dual complementary slackness conditions)为

$$y_j > 0 \Rightarrow \mathbf{A}_i \mathbf{x} = b_i \quad (16)$$

其中 $\mathbf{A}_i$ 为 $\mathbf{A}$ 的第 $i$ 行. 原-对偶近似算法基于如下事实: 当 $\mathbf{x}$ 和 $\mathbf{y}$ 同时满足两个互补松弛条件时, 它们分别是原问题和对偶问题的最优解<sup>[40]</sup>. 而原-对偶近似算法的步骤也是逐步增大对偶函数目标值, 并在此基础上寻找尽量满足原-互补松弛条件的原问题的可行解. 文献[41]研究了单播路由问题, 提出了一种有效的原-对偶启发式算法来找到一条路径, 平衡了路径代价、路径延迟和出错率.

此外, 还有基于分治算法、随机蒙特卡洛算法<sup>[39]</sup>的近似方案. 智能算法也是求解 NP-难问题的一种途径, 如模拟退火、遗传算法、蚁群算法、粒子群算法

等现代优化理论, 它们都是基于启发式迭代步骤的方法. Fei 等人<sup>[42]</sup>用遗传算法研究了带 QoS 约束的路由问题, 是智能算法在网络路由服务质量应用的经典文献.

图 6 给出了网络优化中 NP-难问题的规约经典问题和举例以及求解方法.

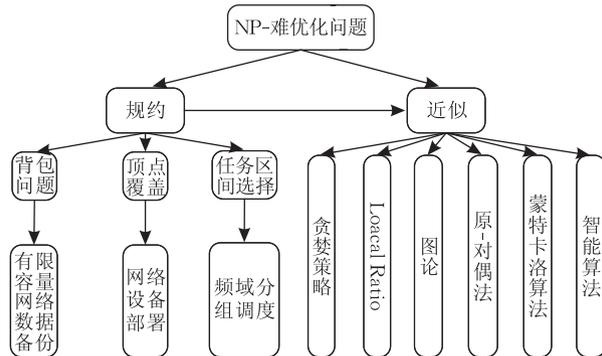


图 6 网络优化中 NP-难问题的规约经典问题和举例以及求解方法

## 4 优化实施方案

我们可以将优化算法的实施方案按照空间和时间归类.

按照优化算法实施的位置不同, 可以分为集中式优化实施方案和分布式优化实施方案. 顾名思义, 集中式优化需要有一个集中的控制器, 并用其计算的优化结果进行任务调度或资源分配. 而分布式算法的优化问题求解过程本身就是分布式并行完成的, 网络中的相关节点都参与了一部分优化过程, 并进行必要的信息传递, 最后每个节点利用信息对自己进行调度控制.

按照优化算法实施的时间不同, 可以分为静态优化和动态优化. 静态优化在算法过程中所有的参数都是锁定的, 而动态优化算法过程中要参考网络运行的状态做出决策, 在每个时间点可能做出不同的决策.

### 4.1 集中式优化与分布式优化

集中式优化与分布式优化算法的优缺点如表 1 所示.

表 1 集中式优化与分布式优化算法比较

	集中式优化	分布式优化
运行速度	慢(大量任务在同一机器上运行)	快(并行执行)
节点间通信量	小(只需控制器与每个节点交互)	大(可能存在广播式交互)
安全性	弱(易受单点攻击)	强
成本	高(需要新增设备)	低(只需利用现有设备)

例如,式(6)表述的流量优化控制问题,可以将其对应的拉格朗日函数写成如下形式:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_s L_s(x_s, \lambda^s) + \sum_l c_l \lambda_l \quad (17)$$

其中  $L_s(x_s, \lambda^s) = U_s(x_s) - \lambda^s x_s$ ,  $\lambda^s = \sum_{l \in L(s)} \lambda_l$ . 这实际上用对偶理论将问题进行了分解,每个网络节点通过分别求解最大化自己的拉格朗日函数来求得整体的对偶函数,即

$$x_s^*(\lambda^s) = \arg \max_{x_s \geq x_s^{\min}} [U_s(x_s) - \lambda^s x_s] \quad \forall s \quad (18)$$

然后通过梯度的方法最小化对偶函数. 整个算法是在每个节点上并行实施的. 由于可以把  $\lambda^s$  看作源节点  $s$  的排队延迟,  $\sum_{s, l \in L(s)} x_s$  可以用链路  $l$  的总流量衡量,因此此例中无需任何显示的信息交换<sup>[24]</sup>. 实际上可以把对偶变量看作资源的代价,则对偶的方法其实是一种基于代价的网络调度方案.

文献[24,27]给出了将求解过程分解的其它方法,如原分解法通过先固定一组变量求解最优值,再用关于这组变量的梯度进行更新,迭代求解最终的最优化问题. 原分解和对偶分解法可以嵌套使用,使求解过程层次化,相应的分解方案可被解释为在不同时间尺度上进行迭代. Chiang 等人<sup>[18]</sup>讨论了基于优化分解的分层蕴含的两种思想:一是可将网络看成一个优化器,网络的协议(如 TCP 协议)被看作分布式的优化解决方案. 虽然网络协议很多是启发式的设计,但可以通过逆向工程发现协议对应的基本数学问题,然后通过改进模型元素系统的方法改进协议. 二是通过优化分解进行网络层次的设计,网络的每层对应一个分解后的子问题,层次之间的接口即为优化迭代中需要进行的通信信息,通过寻找代价低、通信少的分解方法可以设计出比较合理的分层方式.

#### 4.2 静态优化与动态优化

静态优化与动态优化算法的优缺点如表 2 所示.

表 2 静态优化与动态优化算法比较

	静态优化	动态优化
易解程度	简单	复杂
优化结论	强,往往能找出最优解	弱,往往收敛到次优解
真实度	低	高
适应性	弱	强

动态优化可以采用基于状态的动态规划方法(或马尔可夫决策过程 MDP),其基本构成要素如下:

- (1) 决策者. 是做出决策的物理构件或人;
- (2) 系统状态空间与转移概率/强度. 系统状态是对当前系统状态的描述. 状态之间的可转移关系,

可以通过转移概率(离散时间模型)或转移强度(连续时间模型)来描述;

(3) 决策者在做决策时的信息集,即可观测集. 根据信息集的性质不同,可将 MDP 分为两类:一是决策者能得到系统的完全信息,即普通 MDP;二是决策者只能得到部分系统信息,这称为部分可观测马尔可夫决策过程(POMDP);

(4) 决策者在某个系统状态的行为集合. 决策者在每个状态的可选行为集合  $A$ ,依赖于当前状态  $x$ ,即可将该集合记为  $A(x)$ .

MDP 的求解可以使用标准的动态规划法<sup>[43]</sup>,而 POMDP 的求解方法可以参见文献[44]. POMDP 的求解耗费资源较大,实际应用中一般使用一些近似算法进行求解<sup>[45]</sup>.

网络中的动态队列管理优化问题常常会涉及到 MDP. 通常,MDP 的决策算法部署在控制器中.

图 7 描述了 3 种不同的队列管理模型,即单队列多服务器(SQMS)、多队列单服务器(MQSS)以及多队列多服务器(MQMS). 下面以图 7 中的模型 I 为例<sup>[46]</sup>,说明对动态优化问题如何进行建模分析. 考察一个  $M/M/K/(N-K)$  系统. 其中,有一个大小为  $N-K$  的缓冲队列, $K$  个服务器. 系统中可容纳的任务总数为  $N$  个. 任务的到达速率为  $\lambda$ . 服务器的服务速率各不相同. 令服务器  $i$  的服务速率为  $\mu_i$ . 控制行为可以在两种情形下发生:

(1) 在一个新任务到达时,控制器可以决定是将这个任务送至某个空闲服务器,或者在队列未满的情况下将这个新到达任务入队. 当队列已满,且所有服务器均在工作时,新到达的任务将会被丢弃;

(2) 在一个任务完成时,若队列非空,控制器可以决定是否取出一个任务,将其送至某个空闲服务器.

该系统可用一个可控过程  $\{Z(t)\} = \{X(t), U(t)\}$  来描述,其中  $\{X(t)\}$  是系统状态,  $\{U(t)\}$  是控制过程. 该可控系统的目标是通过一定的控制策略,使系统中的平均任务数最少. 其中,  $\{X(t)\} = \{Q(t), D(t)\}$ ,  $Q(t)$  为  $t$  时刻的队列长度,  $D(t) = (D_1(t), \dots, D_K(t))$  为  $t$  时刻各个服务器的状态,用二元指示变量表示. 系统中的任务数是一个随机过程,用  $L(t) = Q(t) + \sum_{1 \leq k \leq K} D_k(t)$  表示. 上述模型是一个连续时间马尔可夫决策过程. 系统的目标函数可以设定为最小化系统平均任务数. 令  $[0, t]$  内系统中的总任务数为

$$Y(t) = \int_0^t L(u) du \quad (19)$$

给定初始状态  $x$  与策略  $\delta$ , 其期望记为  $E_x^\delta$ , 则系统的优化目标为

$$\min_{\delta} g(x; \delta) = \min_{\delta} \lim_{t \rightarrow \infty} \frac{1}{t} E_x^\delta Y(t) \quad (20)$$

该 MDP 可以利用文献[47]中提供的算法进行求解。

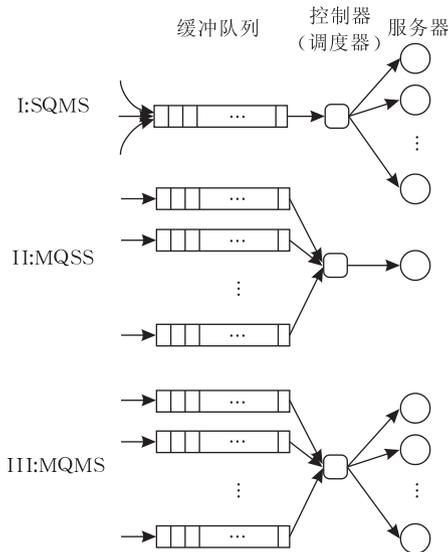


图 7 网络中的队列管理模型

需要说明的是, 静态优化与动态优化没有不可逾越的鸿沟, 动态优化往往也是建立在每一个时间槽中执行一次静态优化, 有时为了减少开销, 只有检测到系统参数发生明显变化时才重新计算, 因此需要考虑适应性与计算量之间的折中。

## 5 算法评价

一个优化算法的好坏取决于它在实际网络中实施后的效果。优化方案的复杂度越低, 需要的通信量越小, 代价越低; 优化方案的优化程度越高, 稳定性和适应性越强, 带来的性能改进越大。因此, 我们需要对优化方案的代价和性能进行评价。有些指标容易写出评估表达式, 如复杂度可以用经典的大  $O$  表达式描述, 而有些指标需要真正部署到实际应用中, 通过统计实验的方法进行评价, 如算法的适应性等, 很难给出一个显式的表达式。

### 5.1 优化算法的代价

优化算法可以用大  $O$  表示法给出其计算复杂度理论上限。在基于迭代方法的优化算法中, 也用该方法表示迭代的收敛速度。在解决凸优化问题时采用的原-对偶内点法通常有比线性更快的收敛速度, 例如, 对于线性规划, 内点法的收敛速度为  $O(\sqrt{n}L)$ , 其中  $L$  表示输入数据的二进制编码长度; 对于 SDP 规划, 内点法的收敛速度在一定条件限制下亦为

$O(\sqrt{n}L)^{[48-50]}$ 。这保证了凸优化问题的多项式时间可解性。在评价网络优化算法度时, 不仅应关注算法本身, 而且应考虑其实际应用中的平均性能。例如, 在文献[51]中, 先在算法的开始进行预处理, 求得网络的最大流量, 然后每进行一次最短路由选择, 采用一次最短路径的方法, 并且更新网络的容量。此时求网络最大流的复杂度被平摊到每次路由选择中, 使其可以忽略不计。评价算法的原则是关注于算法的复杂度对网络提供的服务带来的代价。

为了实现优化算法, 有时需要收集需要的参数信息, 尤其在分布式优化算法中, 每个节点都需要从其它节点获取需要的信息(即其余子问题的解), 这时不可避免地产生了大量通信代价。有些信息可以从系统状态中获知, 如前面所述的分解算法, 用排队延迟代替需要交换的对偶变量信息。评价通信代价可以用信息交互量表示, 文献[52]在最大化资源分配效用函数时采用两种分解方法: 对偶分解和高斯-塞德尔(Gauss-Seidel)分解, 并对比了在线性拓扑和树形拓扑情况下, 二者所需的通信量, 验证了采用高斯-塞德尔方法需要的通信代价较少。

### 5.2 优化算法的性能

在可以求得解析解的优化方案中, 优化算法完美地达到了使网络性能取得最优的目的, 如在网络路由协议中采用的最短路径算法, 在有限步骤之后精确地得到了最优方案。然而, 在只能用迭代算法逼近其最优值的数值求解时, 需要人为地设定与理论最优性能的容忍度  $\epsilon$ ,  $\epsilon$  通常是一个很小的常数, 其值的大小与迭代次数是负相关的, 因此需要折中考虑  $\epsilon$  的取法。在 NP-难问题中, 甚至不存在多项式时间内可以收敛到最优值的方案, 这就需要寻找近似算法, 以图达到性能与效率的折中。需要说明的是, 即使在存在多项式时间求解方案的情况下, 也可能鉴于需要尽可能地缩短优化算法运行的时间或将算法分布式实施而采用性能和复杂度的折中方案。例如文献[9]利用最大二分图匹配的方法进行动态信道分配, 在具体实施时采用贪心策略将匹配算法部署到所有请求用户上, 不仅达到了效率上的提高, 而且验证了其理论与理论最优值的差距小于 7%。对于 NP-难问题的近似方案, 由于不知道具体的最优解, 证明近似算法与理论最优解的差距就更为关键。令  $G$  为某一最大化问题, 近似算法的返回值为  $ALG(G)$ , 理论最优值为  $OPT(G)$ , 通常用近似比  $\rho(|G|)$  来描述近似程度, 即

$$OPT(G) \leq \rho(|G|) \cdot ALG(G) \quad (21)$$

$\rho(|G|)$  越小说明近似度越高, 当它与问题的规模无

关时,近似比为常数  $\rho$ ,特别地,若  $\rho=1$ ,则近似算法实际上找到了最优解.我们通常希望设计出具有与问题规模无关且尽量小的近似比的优化算法.文献[36]分别采用了贪婪算法和 Local-Ratio 技术得到两种不同的上行链路频域分组调度算法,并证明贪婪算法的近似比为  $O(\ln n)$ ,而基于 Local-Ratio 技术的算法达到了常数的近似比 2.

对于动态算法来说,常常要保证系统状态变量的稳定性,文献[53]给出了排队网络强稳定性的表达式,一个排队网络是强稳定的,当下式成立

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} E\{Q_q(\tau)\} < \infty, \quad q \in Q \quad (22)$$

其中,  $Q(t) = (Q_q(t))_{q \in Q}$  是描述  $t$  时刻队列长度的向量.可以用李雅普诺夫定理<sup>[21]</sup>证明稳定性.定义

非负的李雅普诺夫函数  $L(Q)$ ,例如  $L(Q) = \sum_{q=1}^L Q_q^2$ ,当存在常数  $B > 0, \epsilon > 0$ ,使得下式成立

$$E\{L(Q(t+1)) - L(Q(t)) | Q(t)\} \leq B - \epsilon \sum_{q=1}^L Q_q(t) \quad (23)$$

则网络是强稳定的,而且

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{q=1}^L E\{Q_q(\tau)\} < \frac{B}{\epsilon} \quad (24)$$

Neely 等人<sup>[21]</sup>利用李雅普诺夫优化,得到了目标性能与稳定性的折中 $[O(1/V), O(V)]$ ,李雅普诺夫优化已经应用到许多动态网络模型的优化中<sup>[54-56]</sup>.

## 6 总结与展望

网络协议机制往往从启发式设计开始,在真正部署之后需要不断地进行补丁式的改善,这是由于网络设计之初没有严格的理论支撑.网络机制与优化理论的偏离使得后续演进的实施代价增高,也为服务质量的保障设下了难题.从优化模型出发对网络系统进行建模分析,能够使其性能达到尽可能令人满意的程度,从设计到实施都有章可循.

本文从优化理论出发,结合网络中设计目标和约束与优化模型的对应关系,阐述如何应用优化理论辅助网络系统的设计.我们将网络优化模型分为资源分配、任务调度、网络资源部署和系统参数配置等 4 个方面,并依据不同的模型结构和易解程度给出了常用的解决方案.根据求解优化模型的算法不同,网络优化控制的实施方案也不同,因此应力图找到代价低、通信少、性能好和稳定性强的实施方法,设计出比较合理的部署方式.最后,对优化方案进行性能和代价的评价,以助于网络优化的再造工程.

网络中的优化理论将成为网络系统设计的基本出发点,它不仅能够严格导出网络中同一层次中资源分配和任务调度的最优决策,而且能够指导网络的跨层设计.很多网络机制基于启发式设计,尚待改进,因此需要网络优化理论的支持.不仅如此,网络优化理论本身也存在很多难点,值得关注,例如分布式优化中通信信息的传输,基于状态的动态规划状态空间的化简等.在实际中,网络中各节点的能量和资源有限,也亟需找到更有效的优化算法.此外,对优化算法的评价也是一个挑战,尤其是诸如适应性之类不易被量化的指标.优化算法的评价结果将对选择具有最优折中性能指标的算法提供参考.

## 参 考 文 献

- [1] Haykin S. Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications (JSAC)*, 2005, 23(2): 201-220
- [2] Hossain E, Niyato D, Han Z. *Dynamic Spectrum Access and Management in Cognitive Radio Networks*. Cambridge, UK: Cambridge University Press, 2009
- [3] Chipara O, Lu C, Stankovic J. Dynamic conflict-free query scheduling for wireless sensor networks//*Proceedings of the 14th IEEE International Conference on Network Protocols (ICNP'06)*. Santa Barbara, USA, 2006: 321-331
- [4] Xue C J, Xing G-L, Yuan Z-H, Shao Z-L, Sha E. Joint sleep scheduling and mode assignment in wireless cyber-physical systems//*Proceedings of the 2009 29th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'09)*. Montreal, QC, 2009:1-6
- [5] Radunović B, Le Boudec J-Y. A unified framework for max-min and min-max fairness with applications. *IEEE/ACM Transactions on Networking (TON)*, 2007, 15(5): 1073-1083
- [6] Zheng H-T, Peng C-Y. Collaboration and fairness in opportunistic spectrum access//*Proceedings of the 40th Annual IEEE International Conference on Communications (ICC'05)*. Seoul, Korea, 2005: 3132-3136
- [7] Zussman G, Segall A. Energy efficient routing in ad hoc disaster recovery networks//*Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM'03)*. San Francisco, USA, 2003: 682-691
- [8] Sarker R A, Newton C S. *Optimization Modeling: A Practical Approach*. Boca Raton, FL, USA: CRC Press, 2008
- [9] Jin J, Xu H, Li B-C. Multicast scheduling with cooperation and network coding in cognitive radio networks//*Proceedings of the 29th IEEE Conference on Computer Communications (INFOCOM'10)*. San Diego, USA, 2010: 1-9
- [10] Fabregat R, Donoso Y, Baran B, Solano F, Marzo J L. Multi-objective optimization scheme for multicast flows: A survey, a model and a MOEA solution//*Proceedings of the 3rd International IFIP/ACM Latin American Conference on Networking (LANC'05)*. New York, USA, 2005: 73-86

- [11] Donoso Y, Fabregat R, Solano F, Marzo J L, Baran B. Generalized multiobjective multitree model for dynamic multicast groups//Proceedings of the 40th Annual IEEE International Conference on Communications (ICC'05). Seoul, Korea, 2005; 148-152
- [12] Donoso Y, Fabregat R, Marzo J L. A multi-objective optimization scheme for multicast routing: A multitree approach. *Telecommunications Systems Journal*, 2004, 27(2-4): 229-251
- [13] Wang T, Wu Z, Mao J. A new method for multi-objective TDMA scheduling in wireless sensor networks using Pareto-based PSO and fuzzy comprehensive judgement//Proceedings of the High Performance Computing and Communications (HPCC'07). Lecture Notes in Computer Science 4782. Berlin, Heidelberg: Springer-Verlag, 2007; 144-155
- [14] Awad M K, Shen X-M. OFDMA based two-hop cooperative relay network resources allocation//Proceedings of the IEEE International Conference on Communications (ICC'08). Beijing, China, 2008; 4414-4418
- [15] Bley A. An integer programming algorithm for routing optimization in IP networks//Proceedings of the 16th Annual European Symposium on Algorithms (ESA 2008). Lecture Notes in Computer Science 5193. Springer, Berlin, Heidelberg, 2008; 198-209
- [16] Berry J, Hart W, Phillips C, Uber J, Watson J-P. Sensor placement in municipal water networks with temporal integer programming models. *Journal of Water Resources Planning and Management*, 2006, 132(4): 218-224
- [17] Leonardi E, Mellia M, Marsan M A, Neri F. Joint optimal scheduling and routing for maximum network throughput//Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05). Miami, USA, 2005; 819-830
- [18] Mung C, Steven H L, Luo Z-Q, Ness B S, Wei Y. Guest editorial: Nonlinear optimization of communication systems. *IEEE Journal on Selected Areas in Communications*, 2006, 24(8): 1421-1423
- [19] Chiang M. Nonconvex optimization in communication systems. *Advances in Mechanics and Mathematics*, 2006, 17(1): 1-60
- [20] Luo Z-Q, Yu W. An introduction to convex optimization for communications and signal processing. *IEEE Journal on Selected Areas in Communications (JSAC)*, 2006, 24(8): 1426-1438
- [21] Georgiadis L, Neely M J, Tassiulas L. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, 2006, 1(1): 1-149
- [22] Floyd S, Jacobson V. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking (TON)*, 1993, 1(4): 397-413
- [23] Krishnan S, Chaporkar P. Stochastic approximation algorithm for optimal throughput performance of wireless LANs//Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM'10). New Delhi, India, 2010; 409-410
- [24] Palomar D P, Chiang M. A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications (JSAC)*, 2006, 24(8): 1439-1451
- [25] Dennis J E, Schnabel R S. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ, USA; Prentice-Hall Inc., 1996
- [26] Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004
- [27] Johansson B, Johansson M. Mathematical decomposition techniques for distributed cross-layer optimization of data networks. *IEEE Journal on Selected Areas in Communications (JSAC)*, 2006, 24(8): 1535-1547
- [28] Mehrotra S. On the implementation of a primal-dual interior point method. *SIAM Journal on Optimization*, 1992, 2(4): 575-601
- [29] Cheng M X, Li Y, Du D-Z. *Combinatorial Optimization in Communication Networks*. Secaucus, NJ, USA; Springer-Verlag New York, Inc., 2006
- [30] Jia X, Li D, Du D. QoS topology control in Ad Hoc wireless networks//Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'04). Hong Kong, China, 2004; 1264-1272
- [31] Qiu L, Chandra R, Jain K, Mahdian R. Optimizing the placement of integration points in multi-hop wireless networks//Proceedings of the 12th IEEE International Conference on Network Protocols (ICNP'04). Berlin, Germany, 2004; 1092-1648
- [32] Szymanski T. Challenges in deploying novel systems and networks//Proceedings of the 5th International Conference on Systems and Networks Communications (CENTRIC). Nice, France, 2010
- [33] Cormen T H, Leiserson C E, Rivest L R, Clifford S. *Introduction to Algorithms*. 2nd Edition. Cambridge, MA, USA; The MIT Press, 2001
- [34] Kangasharju J, Roberts J, Ross K. Object replication strategies in content distribution networks. *Computer Communications*, 2002, 25(4): 367-383
- [35] Cheng X Z, Narahari B, Simha R, Cheng M X, Liu D. Strong minimum energy topology in wireless sensor networks; NP-completeness and heuristics. *IEEE Transactions on Mobile Computing*, 2003, 2(3): 248-256
- [36] Yang H K, Ren F Y, Lin C, Zhang J. Frequency-domain packet scheduling for 3GPP LTE uplink//Proceedings of the 29th Conference on Computer Communications (INFOCOM'10). San Diego, CA, USA, 2010; 2597-2605
- [37] Crescenzi P. A short guide to approximation preserving reductions//Proceedings of the 12th Annual IEEE Conference on Computational Complexity (CCC'97). Ulm, Germany, 1997; 262-273
- [38] Amotz B N, Reuven B Y, Ari F, Joseph N, Baruch S. A unified approach to approximating resource allocation and scheduling. *Journal of the ACM*, 2001, 48(5): 1069-1090
- [39] Hochbaum D. *Approximation Algorithms for NP-hard Problems*. Boston, MA, USA; PWS Publishing Company, 1997
- [40] Vazirani V V. *Approximation Algorithms*. Atlanta, GA, USA; Springer, 2001
- [41] Xue G L. Minimum-cost QoS multicast and unicast routing in communication networks. *IEEE Transactions on Communications*, 2003, 51(5): 817-824

- [42] Fei Xiang, Lou Junzhou, Wei Jieyi, Gu Guanqun. QoS routing based on genetic algorithm. *Computer Communications*, 1999, 22(15-16): 1394-1399
- [43] Bertsekas D P. *Dynamic Programming and Optimal Control*. New Hampshire, USA: Athena Scientific, 1995
- [44] Smallwood R, Sondik E. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 1971, 21(5): 1071-1088
- [45] Zhao Q, Tong L, Swami A, Chen Y. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework. *IEEE Journal Selected Areas in Communications: Special Issue on Adaptive, Spectrum Agile and Cognitive Wireless Networks*, 2007, 25(3): 589-600
- [46] Rykov V V, Efrosinin D. Optimal control of queueing systems with heterogeneous servers. *Queueing Systems*, 2004, 46(3-4): 389-407
- [47] Rykov V V. Monotone control of queueing systems with heterogeneous servers. *Queueing Systems*, 2001, 37(4): 391-403
- [48] Ye Y, Guler O, Tapia R A, Zhang Y. A quadratically convergent  $O(\sqrt{nL})$ -iteration algorithm for linear programming. *Mathematical Programming*, 1993, 59(2): 151-162
- [49] Freund R M, Mizuno S. Interior point methods: Current status and future directions. *High Performance Optimization*, 1996, 51: 1-9
- [50] Kojima M, Shida M, Shindoh S. Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs. *Mathematical Programming, Series A*, 1998, 80(2): 129-160
- [51] Acharya S, Gupta B, Chang Y, Risbood P, Srivastava A. Precomputing high quality routes for bandwidth guaranteed traffic//*Proceedings of the Global Telecommunications Conference(Globecom'04)*. Dallas, Texas, USA, 2004: 1202-1207
- [52] Kallitsis M, Kallitsis R, Callaway R D, Devetskiotis M, Michalidis G. A distributed algorithm for resource allocation for delay sensitive services//*Proceedings of the IEEE 14th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks(CAMAD'09)*. Pisa, Italy, 2009: 1-6
- [53] Neely M J, Urgaonkar R. Opportunism, backpressure, and stochastic optimization with the wireless broadcast advantage//*Proceedings of the Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA, 2008: 2152-2158
- [54] Neely M J. Delay-based network utility maximization//*Proceedings of the IEEE International 29th Conference on Computer Communications (INFOCOM'10)*. San Diego, CA, USA, 2010: 2669-2677
- [55] Neely M J. Universal scheduling for networks with arbitrary traffic, channels, and mobility. Technical Report: arXiv: 1001.0960. Cornell University, Ithaca, New York, USA, 2010
- [56] Urgaonkar K, Neely M J. Opportunistic scheduling with reliability guarantees in cognitive radio networks. *IEEE Transactions on Mobile Computing*, 2009, 8(6): 766-777



**LIN Chuang**, born in 1948, Ph.D., professor, Ph. D. supervisor. His research interests include computer networks, performance evaluation, network security analysis, and Petri net theory and its applications.

**LI Yin**, born in 1986, Ph. D. candidate. His research interests include QoS of computer networks, service computing and Petri nets.

**WAN Jian-Xiong**, born in 1982, Ph. D. candidate. His research interests include QoS of computer networks, avionics networks and Petri nets.

## Background

Computer network design using heuristic methods has been considered deficient to provide satisfactory performance. Optimization approaches provide powerful theoretical support for designing mechanisms to guarantee quality of services (QoS) in computer networks. The original intention of optimization approaches is to find out the optimal solutions theoretically, so that overcome the drawbacks by just adopting the rule of thumb. This paper categorizes optimization problems of computer networks into four subjects: resource allocation, task scheduling, resource deployment, and parameter configuration. It also surveys the techniques according to the optimization steps: modeling, solving, implementing, and evaluating. This work is partly supported by the National Basic Research Program (973 Program) of China

(Nos. 2010CB328105, 2009CB320504), National Natural Science Foundation of China (Nos. 60834004, 60932003, 60673187, 60702009, 60970001). These projects aim to provide better performance in computer networks and information systems. Our group has been working on the research of QoS in computer networks, and using optimization techniques to solve the network design problems, such as LTE Frequency-Domain Packet scheduling and power management in datacenter networks. Many good papers have been published in respectable international conferences and transactions, such as INFOCOM and IEEE Transactions on Parallel and Distributed Systems. This paper summarizes the optimization methods on the topic of QoS in computer networks.