

多路径网络资源分配 与业务管理

Resource Allocation and Service Management
for Multipath Networks

李世勇 孙微 著



科学出版社

多路径网络资源分配与业务管理

李世勇 孙 微 著

科 学 出 版 社

北 京

内 容 简 介

本书简要介绍多路径网络资源分配与业务管理的理论知识和主要成果,从经济学资源效用优化的角度,建立多路径网络资源分配模型,提出分布式的流量控制算法。内容主要包括多路径网络的资源公平分配与流量控制、并行多路径网络的资源公平分配与流量控制、动态主路径上的资源公平分配与流量控制、资源公平分配算法与稳定性分析、弹性服务的资源分配与流量控制、非弹性服务的资源分配与流量控制、多路径网络异构服务的资源分配、基于效用最优化的网络跨层映射、一体化网络服务层映射模型与资源分配。本书内容是作者近年来的研究成果。

本书可供计算机科学、自动化、系统科学、管理科学等专业的科研人员、高校教师和相关专业研究生使用,也可作为网络管理技术人员的参考书。

图书在版编目(CIP)数据

多路径网络资源分配与业务管理/李世勇,孙微著. —北京:科学出版社, 2018.1

ISBN 978-7-03-055258-7

I. ①多… II. ①李… ②孙… III. ①计算机网络-资源分配-研究 ②计算机网络-业务管理-研究 IV. ①TP393

中国版本图书馆CIP数据核字(2017)第275653号

责任编辑:任 静 董素芹/责任校对:孙婷婷

责任印制:张 伟/封面设计:迷底书装

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京九州迅驰传媒文化有限公司印刷

科学出版社发行 各地新华书店经销

*

2018年1月第 一 版 开本:720×1000 B5

2018年1月第一次印刷 印张:12 1/2

字数:232 000

定价:75.00元

(如有印装质量问题,我社负责调换)

作者简介

李世勇，博士，副教授，燕山大学经济管理学院教师，先后主持国家自然科学基金面上项目 1 项，国家自然科学基金青年项目 1 项，中国博士后科学基金特别资助项目 1 项，教育部高等学校博士学科点专项科研基金 1 项，教育部人文社会科学项目 1 项，中国博士后科学基金面上项目 1 项，河北省自然科学基金 1 项，以第一作者/通信作者身份在多个国际高水平学术期刊上发表学术论文 50 余篇，包括 *Electronic Commerce Research*、*Performance Evaluation*、*International Journal of Systems Science*、*International Journal of Communication Systems*、*Informatica*、*Asia-Pacific Journal of Operational Research*、*Applied Mathematics and Computation*、*Applied Mathematical Modelling*、*Top*、*Quality Technology & Quantitative Management* 等国际学术期刊，入选河北省第二批青年拔尖人才支持计划、河北省宣传文化系统“四个一批”人才工程、河北省“三三三人才工程”第三层次人选及河北省高等学校青年拔尖人才计划。

孙微，博士，副教授，燕山大学经济管理学院教师，先后主持国家自然科学基金青年项目 1 项，教育部人文社会科学项目 1 项，中国博士后科学基金面上项目 1 项（一等资助），河北省自然科学基金优秀青年基金 1 项，河北省自然科学基金青年基金 1 项，以第一作者/通信作者身份在多个国际高水平学术期刊上发表学术论文 40 余篇，包括 *European Journal of Operational Research*、*Asia-Pacific Journal of Operational Research*、*Performance Evaluation*、*International Journal of Systems Science*、*Electronic Commerce Research*、*Applied Mathematics and Computation*、*Top*、*Applied Mathematical Modelling*、*4OR: A Quarterly Journal of Operations Research*、*Central European Journal of Operations Research*、*Quality Technology & Quantitative Management* 等国际学术期刊，入选河北省第二批青年拔尖人才支持计划及河北省高等学校青年拔尖人才计划。目前担任中国运筹学会行为运筹与管理分会理事。

前 言

网络资源管理是计算机网络领域的重要研究内容,得到了国内外广大学者的广泛关注,许多资源分配模型和流量控制策略被相继提出。但由于网络资源的相对稀缺性,网络所能提供的传输能力毕竟是有限的,无法跟上网络中业务量的爆发式增长。

早期的资源分配机制和流量控制算法基本上都是为了提高网络资源的宏观利用率,如系统吞吐量、链路利用率等。而网络资源分配理应包含两个层面的含义:一是如何充分利用现有的有限网络资源,提供用户所需的数据传输服务,完成用户传输数据的需要;二是如何将网络的传输服务能力在要求服务的用户间进行分配。前者主要考虑资源使用的效率,后者更多地关注于资源分配的公平。早期的资源分配机制主要实现了第一个层面的内容,近年来学者才更为关注第二个层面的内容。由于网络最终是为广大用户的应用服务的,网络资源分配和流量控制必须也要考虑到微观层面,如用户满意度、网络系统效益等,从而以一定的指标将网络资源在需求用户间进行合理有效的分配,网络用户之间的公平性、网络服务的质量保证作为网络资源分配的重要指标引起了广大学者的兴趣。为了兼顾网络资源分配的效率和公平两方面的要求,同时满足用户的服务质量,需要从用户满意度的角度实现网络资源的最优分配。作为一个重要的理论成果,网络资源效用优化模型为资源分配和流量控制提出了有益的研究思路和方法。

效用指消费者从消费产品中所获得的主观上的满足,在经济学中通常用效用函数来定量描述消费量和效用值之间的关系。对于网络用户而言,当获得网络服务时,更为关心的是服务的实际效果,如多媒体视频是否连贯清晰、音频有无明显滞后或抖动等,用户对于所获得服务实际效果的这种满意度就可以用效用来刻画。显然,从网络系统的角度追求用户效用最大化,比系统吞吐量、链路利用率等传统的系统性能指标,更加贴近用户实际需求,这也就是面向服务、以用户为中心的 resource 分配目标。更重要的是,这一目标不仅体现了资源分配的效率要求,同时体现了一定的公平性思想,而且描述了用户获得服务的实际效果,因此得到了国内外学者的广泛关注。

本书简要介绍多路径网络资源分配与业务管理的理论知识和主要成果,从经济学资源效用优化的角度,建立多路径网络资源分配模型,提出分布式的流量控制算法。内容包括多路径网络资源公平分配、并行多路径网络资源公平分配、动

态主路径上资源公平分配、资源公平分配算法稳定性分析、弹性服务资源分配、非弹性服务资源分配、异构服务资源分配、网络跨层映射与资源分配。由于篇幅的限制，同时考虑到读者的广泛性，作者舍弃了不少更深入的理论问题和某些虽然重要但较特殊的资源分配模型。

阅读本书只需具备计算机网络、运筹学、微观经济学等基础知识，因此一些基础的预备知识在本书中不再赘述。对有志于从事网络资源管理的读者，这是一本较为合适的基础性书籍。

作者衷心感谢导师北京交通大学张宏科教授多年的鼓励和支持，感谢燕山大学的杨军教授、田乃硕教授、华长春教授在课题研究方面提供的无私帮助，感谢香港理工大学郭鹏飞教授的长期合作和指导。同时，感谢燕山大学经济管理学院李春玲教授、李泉林教授、张亚明教授的支持和鼓励。感谢鄂成国博士、刘佳博士、刘海鸥博士在课题研究过程中做的大量工作。最后感谢家人长年累月无私的支持和奉献，没有他们的一贯支持，这项工作是不可能完成的。

作者关于网络资源管理的研究得到了国家自然科学基金青年项目、面上项目的连续资助（批准号：71301139、71671159）。在此，对国家自然科学基金委员会表示感谢。

由于作者水平所限，不足之处在所难免，欢迎广大读者批评指正，以求改进。

李世勇 孙 微

2017年7月

于秦皇岛燕山大学

简略符号注释表

ACK	Acknowledgment	应答
AID	Access Identifier	接入标识
AIMD	Additive Increase and Multiplicative Decrease	加性增长/乘性递减
AQM	Active Queue Management	主动队列管理
ASR	Access Switching Router	接入交换路由器
AVQ	Adaptive Virtual Queue	自适应虚拟队列
BIC	Binary Increase Congestion Control	二元增式拥塞控制
CID	Connection Identifier	连接标识
CMT	Concurrent Multipath Transfer	并行多路径传输
cmpSCTP	concurrent multi-path SCTP	并行多路径 SCTP
GSR	General Switching Router	广义交换路由器
HSTCP	HighSpeed TCP	高速 TCP
HTTP	HyperText Transfer Protocol	超文本传输协议
IETF	Internet Engineering Task Force	互联网工程任务组
IP	Internet Protocol	网际协议
ISP	Internet Service Provider	互联网服务提供商
MCMP	Multiple Connection Management Protocol	多连接管理协议
MPTCP	Multipath TCP	多路径传输控制协议
MNUM	Multipath Network Utility Maximization	多路径网络效用最大化
M/TCP	Multipath Transmission Control Protocol	多路径传输控制协议
NACK	Negative Acknowledgment	否定应答
NGI	Next Generation Internet	下一代互联网
NUM	Network Utility Maximization	网络效用最大化

OSI	Open System Interconnect	开放系统互连（标准）
PI	Proportional Integral	比例积分
POA	Proximal Optimization Algorithms	近似优化算法
PSO	Particle Swarm Optimization	粒子群优化
pTCP	parallel TCP	并行 TCP
QoS	Quality of Service	服务质量
RED	Random Early Detection	随机早期检测
REM	Random Exponential Marking	随机指数标记
RID	Routing Identifier	路由标识
RTT	Round Trip Time	往返传输时间
SID	Service Identifier	服务标识
TCP	Transmission Control Protocol	传输控制协议
UDP	User Datagram Protocol	用户数据报协议
VoIP	Voice over Internet Protocol	网络电话
W-PR-SCTP	Westwood Stream Control Transmission Protocol with Partial Reliability	具有部分可靠性的 Westwood 流控制传输协议
XCP	eXplicit Control Protocol	显式控制协议
21CN	21st Century Network	21 世纪网络

目 录

前言

简略符号注释表

第 1 章 绪论	1
1.1 背景与意义	1
1.1.1 资源分配的研究进展	3
1.1.2 网络效用最大化理论	4
1.2 相关工作	7
1.2.1 多路径传输协议的研究	7
1.2.2 多路径网络的资源分配	8
1.3 本书主要内容	10
第 2 章 多路径网络的资源公平分配与流量控制	13
2.1 问题的提出	13
2.2 多路径网络资源公平分配模型	14
2.2.1 公平性与效用函数	14
2.2.2 资源公平分配模型	17
2.2.3 模型分析	18
2.3 分布式算法	21
2.3.1 算法描述	22
2.3.2 在网络中的实现	23
2.4 仿真与分析	25
2.4.1 比例公平性	26
2.4.2 调和平均公平性	27
2.5 本章小结	28
第 3 章 并行多路径网络的资源公平分配与流量控制	30
3.1 问题的提出	30
3.2 并行多路径网络的资源分配模型	31
3.2.1 并行多路径网络	31
3.2.2 资源公平分配模型	32
3.2.3 模型分析	34

3.3 分布式算法	37
3.3.1 算法描述	37
3.3.2 平衡点与稳定性	38
3.3.3 在网络中的实现	41
3.4 仿真与分析	42
3.4.1 比例公平性	43
3.4.2 调和平均公平性	45
3.5 本章小结	47
第4章 动态主路径上的资源公平分配与流量控制	48
4.1 问题的提出	48
4.2 资源公平分配模型	49
4.2.1 模型描述	49
4.2.2 模型分析	50
4.2.3 最优带宽分配	52
4.3 分布式算法	53
4.3.1 算法描述	53
4.3.2 平衡点与稳定性	54
4.3.3 在网络中的实现	55
4.4 仿真与分析	56
4.4.1 比例公平性	57
4.4.2 调和平均公平性	58
4.5 本章小结	60
第5章 资源公平分配算法与稳定性分析	61
5.1 问题的提出	61
5.2 多路径网络资源公平分配模型	62
5.2.1 模型描述	62
5.2.2 模型分析	63
5.3 无随机丢包的分布式算法	64
5.3.1 算法描述	64
5.3.2 算法局部稳定性	65
5.3.3 队列时延分析	68
5.4 具有随机丢包的分布式算法	69
5.4.1 模型描述	70
5.4.2 分布式算法	71
5.4.3 收敛速度	73

5.4.4 算法稳定性	73
5.4.5 ECN 标记	76
5.5 仿真与分析	78
5.5.1 无随机丢包的分布式算法	78
5.5.2 具有随机丢包的分布式算法	80
5.6 本章小结	81
第 6 章 弹性服务的资源分配与流量控制	82
6.1 问题的提出	82
6.2 弹性服务的资源分配模型	83
6.2.1 弹性服务与效用函数	83
6.2.2 资源分配模型	85
6.2.3 最优带宽分配	86
6.2.4 流量控制算法	88
6.3 并行多路径网络中弹性服务的资源分配	90
6.3.1 资源分配模型	90
6.3.2 最优带宽分配	91
6.4 仿真与分析	93
6.4.1 并行多路径网络	94
6.4.2 一般多路径网络	96
6.5 本章小结	97
第 7 章 非弹性服务的资源分配与流量控制	98
7.1 问题的提出	98
7.2 非弹性服务的资源分配模型	99
7.2.1 非弹性服务与效用函数	99
7.2.2 资源分配模型	100
7.3 软实时非弹性服务	101
7.3.1 资源分配模型	101
7.3.2 模型分析	103
7.3.3 最优带宽保障	105
7.3.4 并行网络模型	106
7.3.5 最优带宽分配	108
7.4 硬实时非弹性服务	109
7.4.1 资源分配模型	109
7.4.2 最优带宽保障	110
7.4.3 并行网络模型	113

7.4.4 最优带宽分配	114
7.5 基于 PSO 的非凸优化算法	116
7.5.1 PSO 算法	116
7.5.2 非凸优化算法描述	117
7.5.3 非凸优化算法实现	118
7.6 仿真与分析	119
7.6.1 弹性服务仿真与分析	120
7.6.2 软实时非弹性服务仿真与分析	120
7.6.3 硬实时非弹性服务仿真与分析	122
7.7 本章小结	124
第 8 章 多路径网络异构服务的资源分配	125
8.1 问题的提出	125
8.2 多路径网络异构服务的资源分配模型	126
8.2.1 资源分配模型	126
8.2.2 近似优化模型	126
8.2.3 模型分析	128
8.3 分布式算法	130
8.3.1 算法描述	130
8.3.2 算法性能分析	132
8.4 仿真与分析	133
8.4.1 弹性服务仿真与分析	133
8.4.2 异构服务仿真与分析	133
8.5 本章小结	136
第 9 章 基于效用最优化的网络跨层映射	137
9.1 问题的提出	137
9.2 相关工作	138
9.2.1 多连接多路径	138
9.2.2 网络效用优化	139
9.3 网络跨层映射	140
9.3.1 服务到连接的映射	141
9.3.2 连接到路径的映射	141
9.4 网络跨层映射模型	142
9.4.1 效用函数	142
9.4.2 映射模型	142
9.4.3 模型分析	143

9.4.4 协议栈与映射参数	147
9.5 映射算法	148
9.5.1 分布式算法	148
9.5.2 稳定性分析	149
9.5.3 具体实现	151
9.6 映射的性能分析	152
9.6.1 安全性	152
9.6.2 可靠性	153
9.7 仿真与分析	154
9.7.1 分布式算法	154
9.7.2 安全性	156
9.7.3 可靠性	157
9.8 本章小结	160
第 10 章 一体化网络服务层映射模型与资源分配	161
10.1 问题的提出	161
10.1.1 互联网体系架构研究	161
10.1.2 一体化网络体系架构	162
10.2 一体化网络服务层映射模型	163
10.2.1 服务层映射关系	163
10.2.2 基于效用最大化的映射模型	165
10.2.3 各类服务的资源分配	167
10.3 标识的设计与服务的 QoS 保证	169
10.3.1 标识的设计	169
10.3.2 服务的 QoS 保证	170
10.4 流量控制算法	171
10.4.1 弹性服务的流量控制算法	171
10.4.2 非弹性服务的流量控制算法	172
10.5 本章小结	173
参考文献	175

第1章 绪 论

1.1 背景与意义

近年来,互联网无论在规模上还是在业务上都获得了飞速发展,网络在社会的各个方面都得到了极为广泛的应用,成为社会 and 生活中越来越重要的部分。目前,互联网作为全球通信的重要网络基础设施,支持越来越多的不同种类的通信服务。伴随着互联网规模的迅猛发展、业务的爆炸式增长,网络正呈现出资源相对稀缺、服务形式多样化、资源分布化和应用商业化等特点^[1]。尤其是最近几年来用户利用多路径传输协议,可以通过多条路径同时传输数据,从而形成了多路径网络,数据流量更是迅猛增长,而且这种现象在可见的未来仍将持续,由此也产生一些新的问题。

(1) 网络通信资源的相对稀缺。互联网中的资源宏观上大致包括处理器(计算)资源、存储器资源、软件和数据资源、网络通信资源等。由于现在硬件的不断发展和性能的不断提高,一般来说网络节点的处理能力和计算能力不会成为网络传输的瓶颈。本书中的网络资源主要是指用于数据传输的通信资源,包括链路带宽和缓冲区,其中带宽为数据传输的基础资源,缓冲区为辅助资源,主要用于平滑数据流的突发状况,提高链路的利用率。由此,链路带宽和缓冲区资源的量以及它们的组合方式,决定了网络的数据传输能力,这种抽象的传输服务能力可以用丢包、时延、抖动等 QoS 特性进行描述。

互联网中一些新的带宽敏感的应用越来越多,如实时的多媒体视频、音频应用等,这些实时业务对网络带宽的需求越来越大。事实已经表明,网络资源的增长总是跟不上网络中业务量需求的增长,用户对网络资源的需求是无止境的,而网络资源对于用户来说永远是稀缺的,单纯靠增加网络基础设施解决不了网络资源分配中的根本问题,用这种途径来解决资源分配问题也是不恰当的。未来通信网络的一个重要问题是如何有效地分配有限的网络资源,从而调和用户对资源的极大需求与网络资源相对稀缺之间的矛盾,最大限度地实现资源的潜在价值。

(2) 用户之间的公平性问题。由于互联网本身是一个开放式、资源共享的网络,如何实现使用网络资源的用户之间的公平性成为互联网一直以来研究的一个重要内容,其中较为明显的一个公平性问题是采用不同传输协议所造成的不公平。当前网络流量中的主要协议类型为 TCP (Transmission Control Protocol) 与

UDP (User Datagram Protocol), 但是当二者同时竞争网络资源时, 突出问题就是 UDP 流会压倒 TCP 流, 抢占较多的网络资源^[2]。这是由于 UDP 本身没有拥塞控制机制, 当网络发生拥塞时, TCP 流根据拥塞机制进行退避, 而 UDP 流则没有拥塞退避机制。这种情况就造成对采用 TCP 的用户的不公平性。而且, 即使是采取相同传输协议的用户, 由于少量自私的用户为了最大化自己的利益, 通过建立多线程或者发送非 TCP 友好的流抢占资源, 也会对共享资源的其他用户产生不公平性。因此, 为了实现用户之间网络资源的公平分配, 设计一种公平有效的流量控制机制对于共享资源的用户尤为重要。

(3) 实时服务的质量保证问题。当前互联网的研究重点正在从传统的互联互通, 向现代服务和应用过渡, 高可用性成为网络发展的主要特征和必然要求, 其核心是如何合理有效地利用网络资源, 为用户提供所需而又可控同时具有一定质量保证的服务, 为网络的健康发展提供坚实的基础。网络中的服务按照服务所体现出来的效用特性大致分为两类^[3]: 非实时的数据服务和实时的多媒体服务(即本书研究的弹性服务和非弹性服务)。非实时的数据服务(如数据传输、E-mail 等)对网络的带宽和时延要求并不是非常敏感, 一般而言可以根据网络资源的实际情况进行动态调整。而实时的多媒体服务(如多媒体视频、音频等)对网络带宽和时延要求非常敏感, 当网络带宽较小时服务质量将会出现较大的抖动。例如, 多媒体视频服务通常要求较高的带宽, 对数据的丢包非常敏感; 多媒体音频服务要求较为稳定的吞吐量, 对时延非常敏感, 但允许少量的丢包。因此, 如何根据各类服务的不同特性, 分析网络中的带宽资源, 得到链路带宽应该满足的阈值, 从而保证实时多媒体服务一定的服务质量也是资源分配问题需要考虑的重要问题。

(4) 用户的支付与获得的满意度问题。互联网中的用户并没有真正的机会来表达他们对于获得的应用的真实价值, 几乎所有的用户应用都被看作平等的价值。在当前网络上所有的用户都以固定的速率支付(除了物理层上的不同接入, 如拨号上网和专线上网), 网络提供近似相等的网络容量。这种网络资源分配策略严重地限制了用户能够得到网络服务的满意度, 尤其是那些愿意为较高的速率和 QoS 支付更多费用的用户。下一代互联网中, 从用户角度来说, 用户已不再满足原有互联网的尽力而为型服务方式, 而是希望网络能够“按质估价”或者“据价给质”。用户希望获得高质量服务时, 愿意以较高的费用支付换取服务质量保证, 从而获得较高的满意度; 而在用户支付能力有限, 或者对服务质量不敏感时, 用户也愿意牺牲部分服务质量以换取支付费用的降低, 而仅获得一定程度的满意度^[4]。所以, 如何根据用户的支付动态地调整网络资源分配, 满足各类用户对于服务的不同满意度, 是当前网络资源分配要考虑的重要问题。

上述问题是互联网现有机制在发展进程中无法回避的问题, 对互联网的进一

步快速发展形成一定的挑战。实际上,这些问题基本都可以归结为如何以一定的评价指标更有效地实现网络资源的分配,即网络资源的最优分配问题。该问题得到了国内外广大研究学者的广泛关注,许多资源最优分配模型及相应的流量控制机制被相继提出,其中较为突出的就是基于微观经济学中的效用理论建立的模型,也就是本书的主要内容。

1.1.1 资源分配的研究进展

当前互联网的资源分配主要是通过源端的流量控制与链路端的拥塞预测或标记相互合作实现的。源端根据目的端反馈的网络拥塞状况,利用一定的流量控制算法调整其传输速率,而链路端根据其负载情况判断是否发生拥塞,并将该信息以反馈的 ACK 形式通告给源端。

对于采用具有拥塞控制机制的传输协议的源端,可以利用端到端的、分布式流量控制算法探索网络中的可用带宽资源,当网络中没有发生拥塞时,就提高传输速率以获得更多的可用资源,而当网络中出现拥塞后,就降低传输速率以避免更多的丢包。这种基于包丢失的 ACK 反馈机制,也称为“包平衡原理”或“自时钟机制”,这种“端到端”(end-to-end)的拥塞控制框架是互联网拥塞控制设计的基本原则^[5]。目前应用非常成功的 TCP 采用的就是基于 ACK 反馈机制的加性增长/乘性递减(Additive Increase and Multiplicative Decrease, AIMD)^[6]的窗口变化模型,同时后续的 TCP 改进版本基本上都沿用了这种传输模式,如 TCP Reno^[7]、TCP NewReno^[8]、TCP SACK (Selective Acknowledgment)^[9]、T/TCP (Transaction TCP)^[10]等,而随着网络资源的不断增加和网络用户对资源要求的不断提高,近年来出现了适用于高速网络的拥塞控制协议与算法,如 HSTCP^[11]、Scalable TCP^[12]、HTCP^[13]、Westwood TCP^[14]、Fast TCP^[15, 16]、BIC TCP^[17]、Compound TCP^[18]、EHSTCP^[19]等。为了提高用户的数据传输能力,目前已经在网络中得到应用的多路径传输协议也具有内在的拥塞控制机制,如 mTCP^[20]、MPTCP (Multipath TCP)^[21]、SCTP (Stream Control Transmission Protocol)^[22, 23]、CMT (Concurrent Multipath Transfer)^[24]、LS-SCTP (Load-Sharing SCTP)^[25]、cmpSCTP^[26]等,当在某条路径上发生拥塞时,源端就降低该路径上的传输速率,从而通过不断调整传输速率达到最优的资源分配。在网络的链路端,路由器除了为数据包选路和转发,还要根据途经链路的流量进行相应的拥塞预测。路由器通过当前的路由信息(队列及输入和输出速率)对网络拥塞进行预测与标记,在队列溢出之前随机选择包进行丢弃或标记,并以反馈的 ACK 形式及时通知源端来调整传输速率。路由器的这种根据负载或者队长对网络拥塞进行的预测也称为主动队列管理(Active Queue Management, AQM)^[27]。比较典型的 AQM 算法包括 RED (Random

Early Detection)^[28]、REM (Random Exponential Marking)^[29]、CHOCe^[30]、AVQ (Adaptive Virtual Queue)^[31]、PI (Proportional Integral)^[32]、BLUE^[33]、GREEN^[34]、YELLOW^[35]等。这样,通过源端的流量控制算法和链路端的拥塞预测,竞争网络资源的各个用户最终得到了最优资源分配。

资源分配机制和流量控制算法基本上都是为了提高网络资源的宏观利用率,如系统吞吐量、链路利用率等。而网络资源分配问题应该包含两个层面的含义:一是如何充分利用有限的资源,提供用户所需的数据传输服务;二是如何将网络的传输服务能力在要求服务的用户间进行分配。前者主要考虑资源使用的效率,后者更多地关注于资源分配的公平。早期的资源分配机制主要实现了第一个层面的内容,但较少深入分析第二个层面的内容。由于网络传输最终是为用户应用服务的,网络资源分配和流量控制必须要考虑到微观层面,如用户满意度、系统效益等,从而以一定的指标将网络资源在用户间进行合理有效的分配。为了达到资源分配的效率和公平两方面的要求,需要协调用户和网络之间的交互行为,实现网络资源优化分配的目标。作为一个重要的理论成果,网络效用最大化模型为资源分配和流量控制提出了新的研究思路和方法。

1.1.2 网络效用最大化理论

效用 (utility) 是经济学中的基本概念^[36],指消费者从消费产品中所获得的主观上的满足,在经济学中通常用效用函数来定量描述消费量与效用值之间的关系。对于网络中的用户,当获得网络中的服务时,真正关心的是服务的实际效果,如多媒体视频是否连贯清晰、音频有无明显滞后或抖动等,而用户对于所获得服务的这种满意度就可以用效用来刻画。

关于网络资源分配的优化目标,对于网络来说,通常以系统吞吐量、链路利用率或者无线网络的频谱利用率等作为网络系统的优化指标,其核心思想是实现网络资源最大限度的利用,可以认为是以网络为中心的资源分配目标。但是对用户来说,目的是利用网络资源传输数据完成所需要的服务,从而满足其通信和获取信息等方面的实际需求,至于链路利用率如何,并不是用户所直接关心的。效用概念能够很好地描述服务的实际效果或用户的满意程度。显然,从整个系统的角度追求用户总效用最大化,比系统吞吐量、链路利用率等传统的系统性能指标更加贴近实际需求,这也就是面向服务、以用户为中心的资源分配目标。更重要的是,这一目标不仅体现了资源分配的效率要求,还体现了一定的公平性思想。

1) 模型的经济学含义

为了实现面向服务、以用户为中心的资源分配目标,同时兼顾资源分配的效