

排队论及其在计算机 通信中的应用

盛友招 编著

北京邮电大学出版社

排队论及其在计算机 通信中的应用

盛友招 编著

北京邮电大学出版社

图书在版编目(CIP)数据

排队论及其在计算机通信中的应用/盛友招编著. - 北京: 北京邮电大学出版社, 1998

ISBN 7-5635-0306-4

I. 排… II. 盛… III. 排队论-应用-计算机通信
IV. TN91

中国版本图书馆 CIP 数据核字 (98) 第 14393 号

排队论及其在计算机通信中的应用

Paiduilun jiqi zai Jisuanji Tongxin zhong de Yingyong

编 著 盛友招

责任编辑 孙伟玲 王守平

*

北京邮电大学出版社出版

新华书店北京发行所发行 各地新华书店经售

河北省高碑店市印刷厂印刷

*

850 × 1168 毫米 1/32 印张 14.875 字数 385 千字

1998 年 9 月第一版 1998 年 9 月第一次印刷

印数: 1 - 1 500 册

ISBN 7-5635-0306-4/O · 16 定价: 22.50 元

内 容 简 介

本书分为排队论基础、排队论的基础及其应用以及排队论在计算机和现代电信网中的实际应用三大部分，书中除了注重基础知识的阐述和分析之外，还着重做到理论联系实际。本书适用于高校电子、计算机、通信、信息以及相关的管理等专业，作为研究生和本科生的教学或参考用书，也可为广大工程技术人员（包括正在攻读相关专业在职硕士的读者）酌情选用。

目 录

编者的话

主要公式汇集

主要符号汇集

主要缩写字母英汉对照

第一部分 排队论基础

第1章 绪论

§ 1.1 排队论的领域与特征	(3)
§ 1.2 队列的时间图符号	(7)
1.2.1 符号的定义	(7)
1.2.2 到达者与离去者	(11)
§ 1.3 指数分布以及单一服务装置的队列	(13)
1.3.1 指数分布的无记忆特性	(14)
1.3.2 具有指数到达间隔时间与服务时间队列的 分析	(15)
1.3.3 动态分析	(19)
1.3.4 泊松过程	(21)
§ 1.4 增与消过程	(22)
§ 1.5 统计平衡	(27)
§ 1.6 排队论中涉及的若干重要的概率分布	(33)
1.6.1 伯努里分布	(33)
1.6.2 二项式分布	(34)

1.6.3	多项式分布	(35)
1.6.4	几何分布	(35)
1.6.5	负二项式分布	(37)
1.6.6	均匀分布	(38)
1.6.7	负指数分布	(39)
1.6.8	泊松分布	(40)
1.6.9	爱尔兰分布	(40)
§ 1.7	离散时间马尔可夫链	(41)
本章小结		(44)
习 题		(45)
第 2 章	增与消过程及其排队模型	(49)
§ 2.1	概 述	(49)
§ 2.2	外部观察者的分布与到达顾客分布之间的关系	(53)
§ 2.3	爱尔兰损失系统——泊松输入、 s 服务装置、清除 阻塞的顾客	(55)
§ 2.4	爱尔兰时延系统——泊松输入、 s 服务装置(具有 指数服务时间)、时延阻塞的顾客	(58)
§ 2.5	准随机输入	(61)
§ 2.6	具有准随机输入的增与消系统到达顾客的 n 源 分布与外部观察者的 $(n-1)$ 源分布的等式	(64)
§ 2.7	准随机输入、 s 服务装置、清除阻塞的顾客 ——Engset 公式	(69)
§ 2.8	准随机输入、具有指数服务时间的 s 服务装置、 时延阻塞的顾客	(71)
本章小结		(73)
习 题		(74)

第3章 基本的单一服务装置的排队模型	(80)
§ 3.1 指数分布与具有单一服务装置的队列分析入门	(80)
§ 3.2 具有单一服务装置的一般队列	(82)
3.2.1 Little's 公式	(82)
3.2.2 到达与离去时刻	(84)
3.2.3 Little's 公式——对于更新过程 $N(t)$ 情况下 某些定数结果的延伸	(86)
3.2.4 发现队列为“空”的概率	(88)
3.2.5 Kendall's 符号	(90)
§ 3.3 M/GI/1 队列	(91)
§ 3.4 GI/M/1 队列	(100)
§ 3.5 GI/GI/1 队列	(107)
§ 3.6 应用举例	(112)
3.6.1 计算机系统的响应时间	(112)
3.6.2 磁盘访问的响应时间	(114)
3.6.3 一个“发送与等待”协议的模型	(115)
本章小结	(120)
习 题	(123)

第二部分 排队论的基础及其应用

第4章 非基本单一服务装置的排队模型	(129)
§ 4.1 通用的 M/M/1 队列	(129)
4.1.1 与状态有关的 M/M/1 队列	(129)
4.1.2 M/M/C 队列	(130)
4.1.3 M/M/1/m 队列	(132)
4.1.4 M/M/ ∞ 队列	(133)
4.1.5 M/M/m/m 队列	(134)

4.1.6	$M/M/C/M/K$ 队列	(135)
4.1.7	顾客可能离去的 $M/M/C$ 队列	(138)
§ 4.2	扩散过程近似法	(139)
4.2.1	概 述	(139)
4.2.2	用扩散过程近似表示离散过程	(143)
4.2.3	反射“障碍物”	(143)
4.2.4	吸收“障碍物”和瞬时的返回	(143)
4.2.5	连续过程的离散化	(143)
§ 4.3	特殊的队列	(152)
4.3.1	成批到达者的队列 $M^x/M/1$	(152)
4.3.2	成批服务的队列 $M/M^y/1$	(154)
§ 4.4	具有优先级的队列	(155)
4.4.1	不同顾客等级和绝对(抢占)优先级的 $M/M/1$ 队列	(156)
4.4.2	不同顾客等级和绝对(抢占)优先级的 $M/GI/1$ 队列	(158)
4.4.3	不同顾客等级和简单(非抢占)优先级的 $M/GI/1$ 队列	(160)
本章小结		(161)
习 题		(166)
第 5 章	排队网络基础	(177)
§ 5.1	最简单的排队网络	(177)
§ 5.2	开放排队网络的定数分析	(178)
5.2.1	有限转移的假设	(179)
5.2.2	局部依赖的假设	(180)
§ 5.3	闭合系统举例——中央服务装置系统	(183)
§ 5.4	一般闭合系统的解	(185)
§ 5.5	杰克逊(Jackson)定理	(187)

5.5.1	开放网络:Chapman-Kolmogorov 方程		(188)
5.5.2	开放网络:再生(更新)点方法的应用		(190)
5.5.3	闭合网络		(192)
§ 5.6	应用排队网络建模举例		(194)
5.6.1	指数排队网与乘积形式的解		(196)
5.6.2	开放排队网的全局平衡方程		(198)
5.6.3	闭合的排队网络		(203)
§ 5.7	BCMP 网络概要		(213)
5.7.1	BCMP 网络		(213)
5.7.2	服务时间分布		(214)
5.7.3	服务规则		(217)
5.7.4	BCMP 定理		(218)
5.7.5	BCMP 网络举例		(223)
5.7.6	分组交换网络的响应时间		(223)
本章小结			(225)
习 题			(226)
第 6 章	研究排队网络的近似方法		(229)
§ 6.1	分解方法		(231)
§ 6.2	平均值方法		(235)
6.2.1	乘积形式解的闭合网络		(236)
6.2.2	不同顾客等级的闭合网络		(238)
6.2.3	开放网络		(239)
§ 6.3	集合方法		(241)
6.3.1	集合方法的原理		(241)
6.3.2	集合方法举例——具有虚拟存储器系统的 研究		(243)
§ 6.4	隔离方法		(244)
6.4.1	隔离方法的原理		(244)

6.4.2 隔离方法的举例	(246)
§ 6.5 数据传输网的响应时间	(250)
§ 6.6 相互作用计算机模型举例	(254)
§ 6.7 虚电路的计算机网络的性能	(257)
本章小结	(261)
习 题	(262)

第三部分 排队论在计算机和现代电信网中的实际应用

第 7 章 计算机系统的性能分析	(265)
§ 7.1 概 述	(265)
7.1.1 定义与模型	(265)
7.1.2 已获得服务的分布	(268)
§ 7.2 成批处理算法	(270)
§ 7.3 循环(RR - Round Robin)调度算法	(272)
§ 7.4 后来先服务调度算法	(277)
§ 7.5 前后台调度算法	(279)
§ 7.6 多级处理机共享调度算法	(284)
§ 7.7 有限顾客源模型	(288)
本章小结	(292)
习 题	(293)
第 8 章 ATM 网络中的信源及其分析	(297)
§ 8.1 概 述	(297)
§ 8.2 常数比特速率业务	(301)
§ 8.3 可变比特速率业务	(304)
8.3.1 视频业务及其模型	(304)
8.3.2 语 音	(315)
8.3.3 数据应用	(321)
8.3.4 多媒体应用	(326)

§ 8.4 ATM 网络的服务质量的尺度	(326)
8.4.1 呼叫控制参数	(326)
8.4.2 信息传送参数	(330)
8.4.3 B-ISDN 应用中的信息传送尺度	(334)
本章小结	(340)
习 题	(341)
第 9 章 ATM 网络的阻塞控制	(345)
§ 9.1 概 述	(345)
§ 9.2 呼叫接纳控制	(352)
9.2.1 到达信息流的叠加	(354)
9.2.2 带宽分配	(356)
§ 9.3 用法参数控制与网络参数控制	(360)
9.3.1 UPC 功能	(360)
9.3.2 UPC 算法	(362)
§ 9.4 有选择丢弃信元	(370)
9.4.1 推 出	(371)
9.4.2 门 槛	(371)
§ 9.5 反应性的阻塞控制机制	(372)
9.5.1 终接节点通知技术	(374)
9.5.2 适配速率控制	(376)
9.5.3 内部呼叫参数协商	(376)
9.5.4 动态信源编码	(378)
§ 9.6 问题与思考	(379)
本章小结	(380)
习 题	(382)
第 10 章 ATM 交换技术及其性能分析	(388)
§ 10.1 概 述	(388)
§ 10.2 共享介质结构	(391)

10.2.1	ATM 输出缓冲模块交换单元	(392)
10.2.2	分组化自动选路集成系统	(393)
10.2.3	同步合成分组交换	(393)
§ 10.3	共享存储器结构	(393)
10.3.1	先导交换装置	(394)
10.3.2	Hitachi's 共享缓冲器交换装置	(396)
§ 10.4	空分结构	(397)
10.4.1	Banyan 网络	(399)
10.4.2	无阻塞空分结构	(410)
§ 10.5	ATM 交换结构的性能分析举例	(412)
10.5.1	综 述	(412)
10.5.2	输出队列的计算	(415)
10.5.3	输入队列的计算	(417)
10.5.4	共享介质的交换结构	(419)
本章小结		(441)
习 题		(442)
参考文献		(446)

第一部分

排队论基础

第1章 绪 论

§ 1.1 排队论的领域与特征

本课程在对计算机通信系统进行数学分析的基础上，对已有的系统或者准备新建的系统作出性能评价（或分析）。

众所周知，一个电话系统的功能是按照用户（或顾客）需求在主叫用户与被叫用户之间提供通信路由（或称通信信道，也称话路）。如果，在每一对电话用户之间提供一条固定的通信路由时，由于其成本接近天文数字，而不可能办到。为了解决这个问题，必须建立和保持通话的路由是“公用”的，即需要时选用，话毕时释放。这就隐含着这样一种可能性：由于某一时刻恰好没有空闲的路由可选用，这个呼叫就不能建立。亦即，不可能真正做到随时都能“按需”实现任意双方用户之间的通信要求，存在着（系统的）“供”和（顾客的或用户的）“需”之间的矛盾。

类似的问题不仅仅只限于上述电话系统，在许多系统的设计中都存在类似的矛盾。表 1.1 中列举了排队系统的实例。

在表 1.1 所列举的不同系统中，要能合理解决上述“供-需”之间的问题，涉及广泛的政治、经济以及技术方面的知识。但这些系统的共同特征是：在各种情况下，请求服务的次数以及要求服务的持续时间（除了统计意义上而言）一般是随机发生且难以预测的。

本课程力图开发和解释可用于设计和分析这些系统所需要的数学理论。虽然，这些系统通常是很复杂的，但可从所描述的系

统中提取一种数学模型进行分析从中获得有益的信息。

表 1.1 排队系统举例

系 统	顾 客	服务机构
计算机	作 业	CPU、磁盘、I/O
电 话	呼 叫	交换机
机 场	飞 机	跑 道
医 院	病 人	护 士
售票处	如足球爱好者	售票员
大规模运输	司 机	公共汽车、火车
旅馆服务台	顾客（旅客）	接待员
工作室	作 业	机器/工人
公路网	汽 车	交通信号灯
ATM 交换	ATM 信元	宽带信道

这个数学理论是应用概率论的一个分支，它有各种名称，如业务量（话务量）理论、排队论、阻塞理论、大量业务（服务）理论以及随机服务系统的理论。其中，业务量（话务量）理论通常应用于电话和通信业务量方面的理论，如同车辆交通流量理论。排队论通常用于描述更加专门化的等待队列的数学理论。但是，某些系统最有用的或最感兴趣的模型是不容许形成队列，所以，“排队论”的名称似乎不完全适宜，故称它为随机服务系统理论较确切一些。

从排队论的发展历史来看，它不但应用在电话业务量工程方面并已得到了很大的发展，而且还广泛地应用在工程领域、运筹学以及计算机科学等诸方面。排队论在计算机通信方面的应用正在日益受到关注。

现在考虑这样一个模型. 顾客请求使用一个特定的服务机构(服务装置). 如果某一个服务装置是可用的, 来的顾客便占用它并使用一定的时间后离去, 随后, 它将立即为其他来的顾客服务或者等待顾客来到. 如果来的顾客发现没有可利用的服务装置, 顾客可能采取等待或者离去. 这样的模型可以用三个特征来定义: 输入过程、服务机制以及排队规则.

输入过程: 描述有关服务请求的序列. 通常(举例说), 输入过程是由顾客相继到达时刻之间的间隔分布来确定的.

服务机制: 具有服务装置的数量以及顾客占用服务装置的时间长度等特征. 例如, 顾客可以用单个服务装置处理, 每一个顾客占用服务装置相同的时长.

排队规则: 确定对阻塞顾客的安排(当顾客发现所有服务装置全忙时). 例如, 它可以假设: 被阻塞的顾客立即离开系统, 或被阻塞的顾客在一个队列内等待服务并且按照 FCFS(先到先服务)方式给予服务.

设顾客按号码 1, 2, 3, ... 次序来到, 其到达时刻 $a_1 < a_2 < a_3 < \dots$, 如图 1.1 所示. 服务装置按某种次序给予服务: 顾客在队列中排队, 通常是对队列头中第一个顾客正在进行服务.

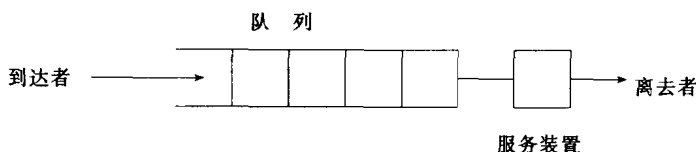


图 1.1 简单的队列

令 $N(t)$ 是在队列内等待的顾客数加上在 t 时刻正在进行服务的一个顾客, 考虑一个时间间隔 $[a, b]$, 这样 $N(a) = N(b) = 0$ (如图 1.2 所示).