

排队论在高校选课系统服务台模型设计中的应用

陈利平

(湖南工学院 计算机系, 湖南 衡阳 421002)

摘 要:在高校网络系统中,随着客户机数量和密集性任务的增加,单个 Web 服务器受到处理能力的限制,已经成为网络访问的新瓶颈。若增加 Web 服务器缓解资源的紧张,则可能造成成本增加,设备闲置。因此,Web 服务器具备高可用性将成为解决这一问题的最佳方法。在综合考虑选课系统中主要应用的算法基础上,以概率动态分布为基础,综合运筹学中的排队论原理,建立一种应用在高校选课系统中的多通道等待服务台模型。实践结果证明,文中提出的模型应用在高校选课系统中,减少运营成本,提高服务水平效果。

关键词:排队论;服务;选课

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2008)09-0216-03

Application of Queuing Theory in Design Server Model of University Course Selection

CHEN Li-ping

(Computer Department, Hunan Technology College, Hengyang 421002, China)

Abstract: In college network system, as the number of client and intensive tasks increase, single Web server has been dealing with capacity constraints, has become the new network access bottleneck. If additional Web servers ease the tension, it may cause increased cost of idle equipment. Therefore, the Web server with a high-availability solution to this problem would be the best way. In this paper, considered the main application of the system of selecting courses on the basis of the algorithm, based on the probability distribution of dynamic, integrated operations research in the theory of queuing theory, the establishment of a system used in college classes in the multi-channel model for desk. Practice results show that the model proposed in this paper application system in college classes, reduce operating costs, improve service quality, effectiveness.

Key words: queuing theory; service; course selection

0 引 言

随着对信息需求的不断深化,网络信息服务引起了人们的极大兴趣,同时对网络服务提出了更高的要求。一个突出的表现是,网络成为高校学习资源不可或缺的载体,起着越来越重要的作用,随着高校教学和管理平台建设的不断深入,网络逐渐承担起教室、教材、教师、管理者等多项职能而成为高校学习系统的要素之一,基于网络的选课系统即是其中应用最广泛的部分。然而大多数开设网上选课的高校,都有选课时资源和顾客(学生)之间无法协调的问题,造成学生抱怨选课系统承受力有限,无法顺利选课,出现排队等待现象,无法提供迅捷高效的服务。若增加服务器以缓

解资源的紧张,这样势必会造成成本增加,甚至出现因供大于求而使设施经常闲置,导致浪费。

因此运用排队论对选课系统中服务台设计与学生访问的流量进行定量分析,对避免盲目确定服务台的个数,提高选课系统的服务和管理水平,降低运营成本等有着重要的作用。

1 排队论原理

1.1 基本概念

排队论^[1,2]是研究排队系统(又称为随机服务系统)的数学理论和方法,是运筹学的一个重要分支。在日常生活中,人们会遇到各种各样的排队问题。如到餐馆就餐、到图书馆借书、在车站等车、去医院看病、去售票处购票、上工具房领物品等。在这些问题中,餐馆的服务员与顾客、公共汽车与乘客、图书馆的借书员与借阅者、医生与病人、售票员与买票人、管理员与工人

收稿日期:2007-12-17

基金项目:湖南省自然科学基金资助项目(06JJ5135)

作者简介:陈利平(1966-),女,湖南望城人,副教授,硕士,研究方向为计算机网络技术和数据库技术。

等均分别构成一个排队系统或服务系统。前者看成是服务机构,后者看成是顾客。排队问题的表现形式往往是拥挤现象。

排队系统的符号一般形式为: $X/Y/Z/A/B/C$

其中: X 表示顾客相继到达时间间隔的分布; Y 表示服务时间的分布; Z 表示服务台的个数; A 表示系统的容量,即可容纳的最多顾客数; B 表示顾客源的数目; C 表示服务规则。

描述一个排队系统的主要数量指标和相应的常用记号如下:

$N(t)$: 时刻 t 系统中的顾客数(又称为系统的状态);

$N_q(t)$: 时刻 t 系统中排队的顾客数,即排队长;

$T(t)$: 时刻 t 到达系统的顾客在系统中的逗留时间;

$T_q(f)$: 时刻 t 到达系统的顾客在系统中的等待时间。

排队论的基本问题是研究这些数量指标在瞬时或平稳状态下的概率分布及其数字特征,了解系统运行的基本特征;系统数量指标的统计推断和系统的优化问题等。

当系统运行一定时间达到平稳状态后,对任一状态 n 来说,单位时间内进入该状态的平均次数和单位时间内离开该状态的平均次数应相等,即系统在统计平衡下“流入,流出”。据此可得任一状态下的平衡方程如下:

$$\text{记: } C_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_{n-1}\mu_{n-2}\cdots\mu_0} \quad n = 1, 2, \cdots \quad (1)$$

则平稳状态的分布为:

$$P_n = C_n P_0 \quad n = 1, 2, \cdots \quad (2)$$

$$\text{由概率分布的要求: } \sum_{n=0}^{\infty} P_n = 1$$

$$\text{有: } [1 + \sum_{n=0}^{\infty} C_n] P_n = 1$$

$$\text{于是: } P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} C_n} \quad (3)$$

注意: 式(3) 只有当级数 $\sum_{n=0}^{\infty} C_n$ 收敛时才有意义,

即当 $\sum_{n=0}^{\infty} C_n < \infty$ 时,才能由上述公式得到平稳状态的概率分布。

1.2 计算公式

当学生选课时,如果不能得到及时服务,就需要等待。学生选课排队等待问题可以通过排队论的有关原理来解决,一般情况下,为了便于研究,可以把选课学生到达流看成泊松流,其服务时间一般服从指数分布。

因此,此时可以把选课系统的服务系统看作 $M/M/S$ 等待制多服务台模型。下面描述 $M/M/S$ 排队系统的计算公式。

设顾客单个到达,相继到达的时间间隔服从参数为 λ 的指数分布,系统中共有 S 个服务台,每个服务台的服务时间相互独立,且服从参数为 μ 的指数分布。当顾客到达时,若有空闲的服务台则可以马上接受服务,否则便排成一个队列等待,等待空间为无限。

下面讨论这个排队系统的平稳分布。记 $P = P\{N = n\} (n = 0, 1, 2, \cdots)$ 为系统达到平稳状态后队长 N 的概率分布,注意到对个数为 S 的多服务台系统,有:

$$\lambda_n = \lambda \quad n = 0, 1, 2, \cdots$$

$$\text{和 } \mu_n = \begin{cases} n\mu & n = 0, 1, 2, \cdots, s \\ s\mu & n = s, s+1, \cdots \end{cases}$$

记 $\rho_s = \rho/s = \frac{\lambda}{s\mu}$, 则当 $\rho < 1$ 时,由式(1)、式(2)和式(3),有:

$$C_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!} & n = 0, 1, 2, \cdots, s \\ \frac{(\lambda/\mu)^s}{s!} \left(\frac{\lambda}{s\mu}\right) = \frac{(\lambda/\mu)^n}{s! s^{n-s}} & n \geq s \end{cases} \quad (4)$$

$$\text{故: } P_n = \begin{cases} \frac{\rho^n}{n!} P_0 & n = 1, \cdots, s \\ \frac{\rho^n}{s! s^{n-s}} P_0 & n \geq s \end{cases} \quad (5)$$

$$\text{其中: } P_0 = \left[\sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s}{s!(1-\rho_s)} \right]^{-1}$$

式(4) 和式(5) 给出了在平衡条件下系统中顾客数为 n 的概率,当 $n \geq s$ 时,即系统中顾客数大于或等于服务台个数时,再来的顾客必须等待,因此记:

$$C(s, \rho) = \sum_{n=s}^{\infty} P_n = \frac{\rho^s}{s!(1-\rho_s)} P_0 \quad (6)$$

式(6) 称为Erlang等待公式,它给出了顾客到达系统时需要等待的概率。

对多服务台等待制排队系统,由已得到的平衡分布可得平均排队长 L_q 为:

$$L_q = \sum_{n=s+1}^{\infty} (n-s) P_n = \frac{P_0 \rho^s}{s!} \sum_{n=s}^{\infty} (n-s) \rho_s^{n-s} = \frac{P_0 \rho^s}{s!} \cdot \frac{d}{d\rho_s} \sum_{n=1}^{\infty} \rho_s^n = \frac{P_0 \rho^s \rho_s}{s!(1-\rho_s)^2} \quad (7)$$

$$\text{或: } L_q = \frac{C(s, \rho) \rho_s}{1 - \rho_s} \quad (8)$$

记系统中正在接受服务的平均数为 s ,显然也是正在服务的服务台的平均数,故

$$\bar{s} = \sum_{n=0}^{s-1} n P_n + s \sum_{n=s}^{\infty} P_n = \sum_{n=0}^{s-1} \frac{n \rho^n}{n!} P_0 + s \cdot$$

$$\frac{\rho^2}{s!(1-\rho_s)}P_0 = P_0\rho[\sum_{n=1}^{s-1} \frac{\rho^{n-1}}{(n-1)!} + \frac{\rho^{s-1}}{(s-1)!(1-\rho_s)}] = \rho^s \tag{9}$$

式(9)说明,平均在服务的服务台个数不依赖于服务台个数 S ,可得到平均 L_0 长。

L_0 = 平均排队长 + 正在接受服务的顾客的平均数 = $L_q + \rho$

对多服务台系统, Little 公式依然成立, 即有:

系统中车辆的平均逗留时间 W 为: $W = \frac{L}{\lambda}$ 。

系统中车辆的平均等待时间 W_q 为: $W_q = \frac{L_q}{\lambda} = W - \frac{1}{\mu}$ 。

2 应用举例及结果分析

采用多通道等待排队算法以前, 采用了 ASP.NET 开发的选课系统, 如果人数不超过 500 人同时在线选课, 系统基本可以流畅进行, 但是人数达到 2000 以上时, 系统出现瘫痪状态, 服务器无法处理数据, 学生的网页长时间打不开。针对这种情况, 曾经采取每次限制 500 个人的分批选课方案, 但引起选课的公正性问题。

根据采集到的高校选课系统的单向通道的学生到达流如图 1 所示, 并且服从负指数分布^[3~5]。因此, 采用多通道等待排队算法, 模拟并计算出最佳的服务器数量, 购入新服务器, 实现多通道选课, 并相应地采用负载均衡进行服务器的选择^[6~8]。在 2006~2007 学年的选课中, 一万多人同时在线, 服务器除了在最初的几分钟反应迟缓外, 基本能正常负荷, 使选课顺利完成。

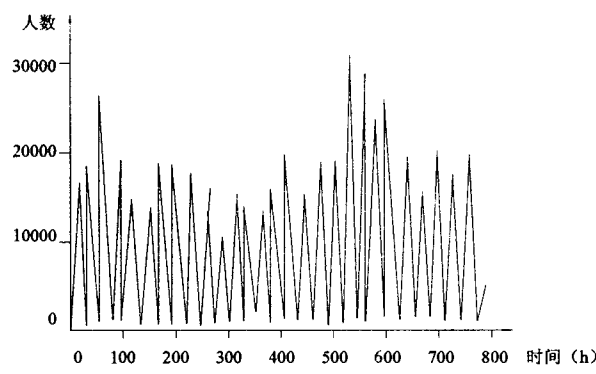


图 1 湖南工学院网站连续 720 小时请求数示意

表 1 所示为湖南工学院网站采用单通道(一台服务器)和多通道(用两台服务器负载均衡)平均请求数(Poisson 流参数倒数的估计值)表得出的两种算法的效果。

表 1 两种算法的比较

时间	单通道等待算法单位时间完成的请求数	多通道流量单位时间完成的请求数
8:00-9:00	1596	3639
9:00-10:00	2320	4053
10:00-11:00	3020	4959
11:00-12:00	2403	4129
12:00-13:00	1722	3902
13:00-14:00	1200	3082
14:00-15:00	4399	6534
15:00-16:00	5002	8023
16:00-17:00	4500	6834
17:00-18:00	2304	4523
18:00-19:00	3803	5349
19:00-20:00	2192	4350

3 结束语

文中运用排队论对高校选课系统中服务台的数量进行了分析, 得出系统选择应综合分析各种限制条件, 做好数据的统计分析, 如果增加服务台, 就要增加投资或发生空闲; 如果服务台太少, 排队现象就会严重, 会对顾客个人和社会带来不利影响。因此, 管理人员必须考虑两者间的平衡, 以提高服务质量, 降低成本。而采用合理的选课系统可缩短服务时间, 对减少运营成本, 提高服务水平, 效果非常显著。

运用本算法的选课系统顺利地完成了湖南工学院 2006~2007 学年第二学期的学生选课, 选课人数达 1 万余人, 在选课高峰期也没有出现系统响应超时等问题。系统的可靠性和稳定性得到了保障。该系统顺利地将选课的教务管理过程和计算机技术有机地结合起来, 使管理流程更加清晰, 管理更加简洁、方便灵活。

参考文献:

- [1] 盛友招. 排队论及其在计算机通信中的应用[M]. 北京: 北京邮电大学出版社, 1998.
- [2] 《运筹学》教材编写组. 运筹学[M]. 北京: 清华大学出版社, 2005.
- [3] 王 凌. 智能优化算法及其应用[M]. 北京: 清华大学出版社, 1989.
- [4] 王治和, 张 强. 一种基于排队论的系统性能评价软件设计[J]. 计算机科学, 2004, 31(增刊): 43-44.
- [5] 丛国超, 朱翼隽. 批量到达的多服务台排队模型求解[J]. 成都信息工程学院学报, 2007, 22(1): 98-100.
- [6] 孙庆宏, 温 渤. 排队论在生产过程时间组织中的应用[J]. 管理方略, 2003(11): 92-93.
- [7] 蒋 涛, 许 延, 陈清金. 网络环境下视频连续播放控制算法[J]. 计算机应用与软件, 2007, 24(12): 143-146.
- [8] 李 维. Delphi7 高效数据库程序设计[M]. 北京: 机械工业出版社, 2003.