

排队过程的仿真研究

严刚峰¹, 黄显核¹, 李思明²

(1. 电子科技大学, 四川 成都 610054; 2. 成都理工大学, 四川 成都 610059)

摘要:运用事件调度法, 设计了排队系统的仿真方案, 并进行了仿真研究。采用 M/M/1 等待制服务系统为例进行了理论计算, 验证了仿真结果的正确性。通过比较 M/M/n 等待制服务系统在相同服务规则以及不同服务规则下的仿真数据, 表明了 M/M/n 等待制服务系统在不同服务规则下可以得到较好服务效率。对于复杂排队服务系统的设计方案的优化提供了仿真验证的方法。

关键词:排队系统; 随机变量; 队列; 计算机仿真

中图分类号: TP391

文献标识码: A

1 引言

现实生活中广泛存在着排队过程。火车站的售票系统, 超市的收银系统乃至对于通信网资源, 当多个用户发出随机的资源使用的请求过程等等都是排队系统的实例。简单地说, 排队过程就是当一个系统对某一群体提供某种服务时, 该群体占用此服务系统时所呈现的状态。排队过程研究的主要内容是排队系统的运行指标和排队系统的优化问题。排队系统虽然已有比较成熟的随机过程理论加以描述, 特别是对较为简单的单服务台排队系统, 采用随机过程的理论可以得到统计平均意义的平稳分布。若想采用随机过程理论来进行系统的模拟, 特别是复杂的多服务台排队系统的动态模拟, 还是相当困难的。首先运用随机过程理论对常见的 M/M/1 等待制服务系统的运行指标进行了计算, 然后对 M/M/1 等待制服务系统进行了仿真, 根据产生的仿真数据来验证仿真方案的正确性。最后在 M/M/1 等待制服务仿真系统的基础上, 设计并比较了 M/M/n 等待制服务仿真系统在相同服务规则与不同服务规则的运行指标。证明采用不同服务规则的运行指标较相同服务规则的运行指标为佳。为复杂的多服务台排队系统的最优化设计提供了仿真验证的方法。

2 典型排队系统的模型

一个排队系统主要是由 4 个要素组成: 顾客的到达过程、排队的规则、服务的时间以及服务系统的结构。对于顾客的到达过程一般有以下儿种达到规律: 一般到达、Poisson 到达、Erlang 到达、等间隔到达、Bernoulli 到达以及预约到达等。而所谓排队规则用来表明来到的顾客是按照怎样的规则接受服务的。常见规则有: 损失制、先到先服务、后到先服务、随机选择服务、优先服务以及批量服务等。服务时间是指在同一时刻有多少服务台接纳顾客以及每位顾客接受服务的时间所服从的分布情况, 常见的分布有指数分布、Erlang 分布以及定长分布等等。服务系统的结构一般是指服务台的数目而言。一个典型的排队系统模型^[1]如图 1 所示。

3 M/M/1 等待制服务系统的仿真验证

3.1 理论计算的结果

为了理论计算的简便, 不失一般性, 考虑比较常见的 M/M/1 等待制系统, 即顾客按照泊松过程到达, 平均率(强度)为 λ , 服务时间服从指数分布, 强度为 μ , 只有一个服务台的排队系统, 系统按照先到先服务的规则进行服务, 当顾客来到服务台时, 若服务员正在给前一位顾

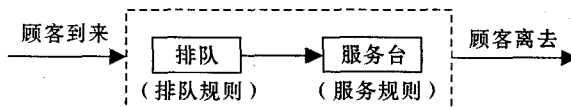


图 1 典型的排队模型

客服务时,该顾客就排队等待。

设 $X(t)$ 为系统在 t 时刻的顾客人数,则 $\{X(t), t \geq 0\}$ 显然满足^[2]:

(1) $X(0) = 0$;

(2) 具有独立增量;

(3) 对任意 $0 \leq s < t$, $X(t) - X(s)$ 服从参数为 $\lambda(t-s)$ 的泊松分布,即:

$$P\{X(t) - X(s) = k\} = \frac{[\lambda(t-s)]^k}{k!} e^{-\lambda(t-s)}, \quad k = 0, 1, 2, \dots$$

其状态空间为

$$E = \{0, 1, 2, \dots\},$$

状态转移速度图如图2所示。状态转移速度矩阵为:

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \mu & -(\lambda + \mu) & \lambda & 0 & \dots \\ 0 & \mu & -(\lambda + \mu) & \lambda & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

由平稳分布

$$\Pi Q = 0$$

可得:

$$\begin{cases} \pi_0 = \left[\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^k \right]^{-1} = \left[\frac{1}{1 - \frac{\lambda}{\mu}} \right]^{-1} = 1 - \frac{\lambda}{\mu} \\ \pi_k = \left(\frac{\lambda}{\mu} \right)^k \pi_0 = \left(\frac{\lambda}{\mu} \right)^k \left(1 - \frac{\lambda}{\mu} \right) \end{cases}$$

在 $\frac{\lambda}{\mu} < 1$ 时, M/M/1 等待制系统的相关运行指标为:

服务台空闲的概率

$$\pi_0 = 1 - \frac{\lambda}{\mu}$$

服务台忙期的概率

$$\pi = 1 - \pi_0 = \frac{\lambda}{\mu}$$

顾客在服务台等待人数的平均数

$$L_q = \sum_{k=1}^{\infty} (k-1) \pi_k = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

顾客在服务台平均等待时间

$$W_q = \frac{1}{\lambda} \cdot \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\lambda}{\mu(\mu - \lambda)}$$

考虑某超级市场的顾客按泊松过程来到单服务台,为了方便,设每单位时间到来4人,计价收款时间服从指数分布,平均每个顾客需2.5个单位时间,这就是 M/M/1 等待制系统。按照性能指标计算式,可知:

$$\lambda = \frac{1}{4}, \quad \mu = \frac{1}{2.5} = \frac{2}{5}, \quad \frac{\lambda}{\mu} = 0.625$$

因此,服务台空闲的概率

$$\pi_0 = 1 - 0.625 = 0.375$$

服务台忙期的概率

$$\pi = 0.625$$

顾客在服务台等待人数的平均数

$$L_q = \sum_{k=1}^{\infty} (k-1) \pi_k = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = \frac{\lambda^2}{\mu(\mu - \lambda)} \approx 1.042$$

顾客在服务台平均等待时间

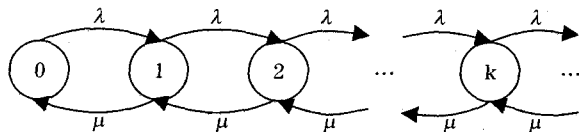


图2 M/M/1 等待制系统的状态转移速度图

$$W_q = \frac{1}{\lambda} \cdot \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\lambda}{\mu(\mu - \lambda)} \approx 4.167$$

3.2 仿真实现

直接进行 M/M/1 等待制系统的仿真显然不容易,考虑到相邻两顾客到达服务台的时刻之差是服从指数分布的,下面首先说明这一点。

设 $\{X(t), t \geq 0\}$ 是参数为 λ 的泊松过程, $\{T_n, n = 1, 2, \dots\}$ 为时间间距序列。 T_1 表示事件第一位顾客到达前所需要的时间,所以 $\{T_1 > t\}$ 表示在 $[0, t)$ 内顾客还未到达,因此,事件 $\{T_1 > t\}$ 的发生当且仅当没有顾客在 $[0, t)$ 时间内出现,于是,对于 $t \geq 0$, 有

$$P\{T_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

$$P\{T_1 \leq t\} = 1 - P\{T_1 > t\} = 1 - e^{-\lambda t}$$

所以 T_1 的分布函数

$$F_{T_1}(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

即 T_1 服从参数为 λ 的指数分布。

T_2 为第 1 位顾客到达第 2 位顾客到达的时间间距,考虑到各位顾客到达事件具有增量独立性,所以:

$$P\{T_2 > t | T_1 = s_1\} = P\{X(s_1 + t) - X(s_1) = 0\} = e^{-\lambda t}$$

$$\begin{aligned} P(T_2 > t, T_1 > s) &= \int_s^\infty \int_t^\infty f(x, y) dx dy = \int_t^\infty \int_s^\infty f_{T_2|T_1}(x|y) f(y) dx dy \\ &= \int_s^\infty P(T_2 > t | T_1 = y) f(y) dy = \int_s^\infty e^{-\lambda t} f(y) dy = e^{-\lambda t} P(T_1 > s) \\ &= P(T_2 > t) P(T_1 > s) \end{aligned}$$

当 $s = 0$ 时,可见 T_2 也是服从参数为 λ 的指数分布,而且 T_1 与 T_2 是独立的,因而 T_1 与 T_2 是独立同分布的。采用数学归纳法,由完全相同的求取方法可以得到时间间距序列 $\{T_n, n = 1, 2, \dots\}$ 满足: $T_n, n = 1, 2, \dots$ 相互独立,而且都为服从参数为 λ 的指数分布。

有上面的结果,就可以通过产生顾客到达的时间间距序列,从而产生顾客的到达时间进行仿真。由变换法即可产生这样一个随机序列。下面简述产生原理,设 U 是区间 $[0, 1]$ 上的满足均匀分布的随机变量,要产生的随机变量的 X 的概率分布函数为 $F(x)$,考虑到 $F(x)$ 是区间 $[0, 1]$ 上单调不减的函数,显然有:

$$P\{X \leq x\} = P\{F^{-1}(U) \leq x\} = P\{U \leq F(x)\}$$

又因为,如果 $m \in [0, 1]$, 则 $P\{U \leq m\} = m$, 所以:

$$P\{X \leq x\} = F(x)$$

因此生成概率分布函数为 $F(x)$ 的随机变量 X 的方法为:

首先产生 $[0, 1]$ 上的均匀分布随机变量 U , 然后令 $X = F^{-1}(U)$, 则所生成的 X 即为满足所需概率分布函数为 $F(x)$ 的随机变量。

由于顾客到达的时间间距序列满足相互独立,都为服从指数分布,概率分布函数为:

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

令 $U = F_T(t)$, 则可解得:

$$T = -\frac{1}{\lambda} \ln(1 - U)$$

所以产生顾客到达的时间间距序列就可以先产生 $[0, 1]$ 上的均匀分布随机变量 U , 再通过 $f(U) = -\frac{1}{\lambda} \ln U$ 即产生了满足要求的顾客到达的时间间距序列。

采用事件调度法^[3], 产生 M/M/1 等待制系统仿真的具体过程为: 首先产生仿真事件列表, 由产生的时间间隔序列确定事件的发生时刻, 比较事件的发生时刻, 确定要发生的事件, 对应进入相应的仿真子系统, 循环直至仿真结束, 并输出所需的仿真数据。具体的仿真流程图如图 3 所示。

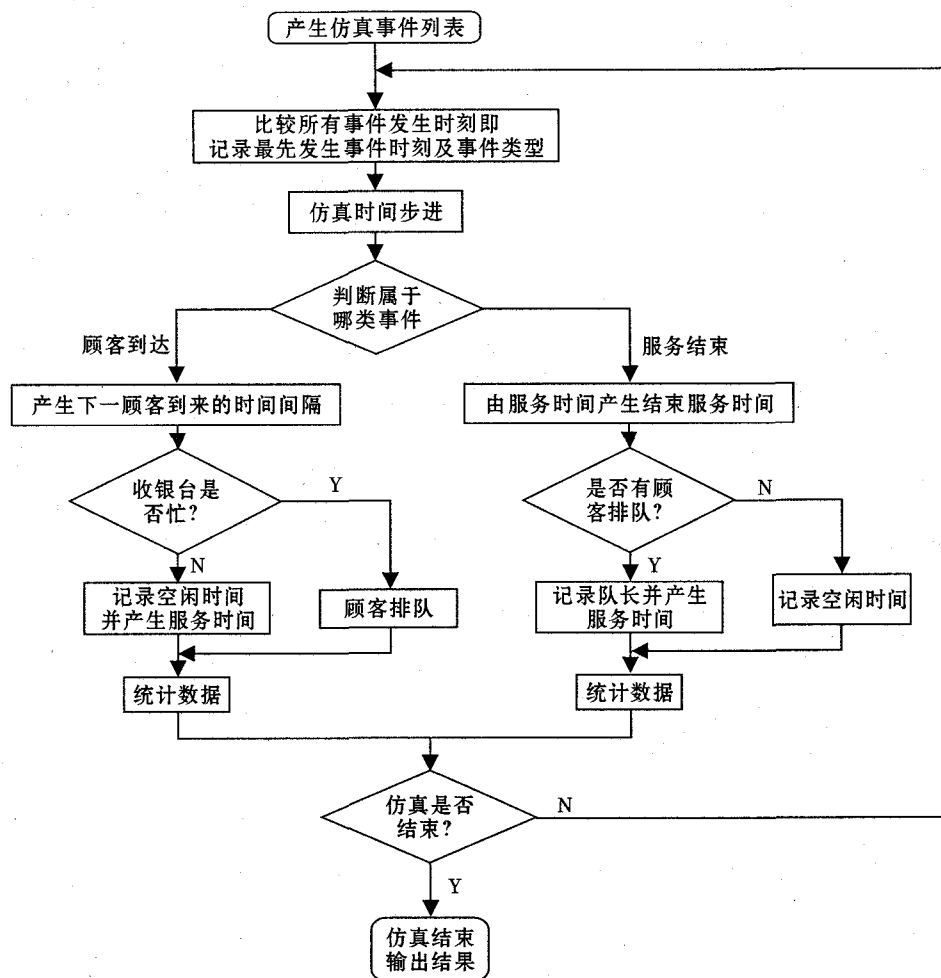


图3 仿真流程图

由图3,可得M/M/1等待制服务系统的过程仿真结构如图4所示。其中服务台的输入分别是:顾客的到达时间间隔序列和服务台对顾客的服务时间,并且顾客到达时间间隔服从参数为4的指数分布,服务台对顾客的服务时间服从参数为2.5的指数分布。

服务台仿真结构如图5所示。

产生1200个到达时间间隔序列以及1200个服务所需时间,即仿真1200位顾客通过M/M/1等待制服务系统的仿真结果。

选取其中一段数据,如表1所示,可见完全符合实际情况。

总等待时间为:4750.949,每位顾客的平均等待时间为 $4750.949/1200 = 3.959$;

服务台的空闲率为: $1 - 3133.879/(3133.879 + 2083.343) = 0.399$

当然也可以选取不同随机数发生序列,重复多次仿真后再取统计平均。但考虑到1200个数据已足够大,其结果应接近平稳分布所计算的数据,事实上也正是如此。

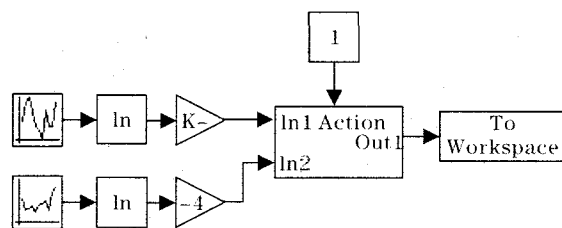


图4 M/M/1等待制服务系统的过程仿真结构图

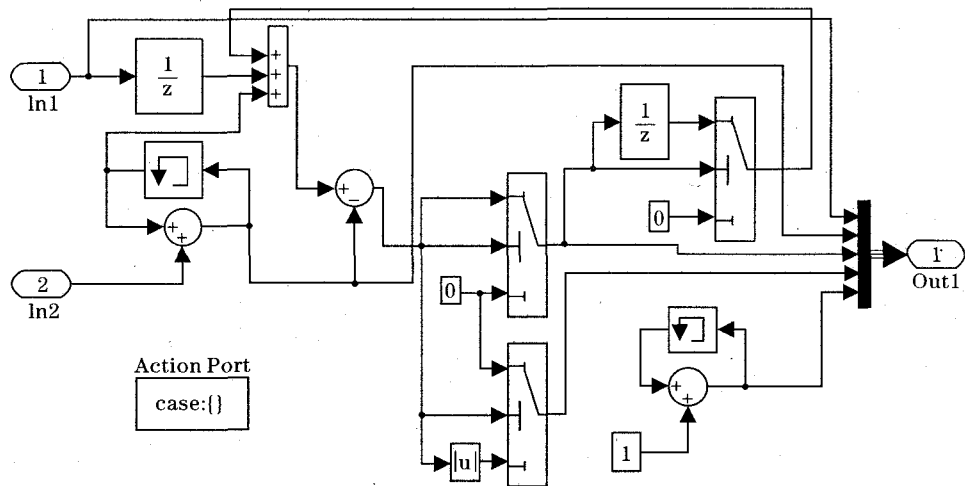


图 5 服务台仿真结构图

表 1 顾客通过 M/M/1 等待制服务系统的部分仿真数据

顾客达到序号	顾客到达时间	顾客所需服务时间	顾客等待时间	服务台空闲时间
1	1.9369	6.3952	0	1.9369
2	9.6788	0.67656	0	1.3468
3	17.174	5.6978	0	6.8181
4	17.798	1.4041	5.0732	0
5	18.206	1.2686	6.0693	0
6	22.044	3.4122	3.5003	0
7	34.564	0.9067	0	9.1081
8	35.975	1.8955	0	0.50425
9	40.492	3.8368	0	2.6215
10	45.273	4.0016	0	0.94386
11	46.535	4.6949	2.7392	0
12	52.206	0.47138	1.7627	0
13	56.965	0.34949	0	4.2875
14	60.067	2.9522	0	2.7521
15	60.249	0.075412	2.7698	0

4 M/M/n 在相同与不同服务规则下的仿真

等待制多服务台系统的模型如图 6 所示。

可见其结构为多个服务台并联,每个服务台的结构并没有变化,产生一个均匀分布的随机数来控制顾客所到达的服务台即可仿真这样一个多服务台相同服务规则下的排队系统。以 M/M/6 等待制排队服务系统为例,可设计系统仿真结构图如图 7 所示。

其中,顾客到达时间间隔服从参数为 4 的指数分布,服务台对顾客的服务时间服从参数为 2.5 的指数分布。选取不同随机数序列,仿真 6 次 1200 位顾客通过 M/M/6 等待制服务系统的仿真结果。

统计平均后的仿真结果如表 2 所示。

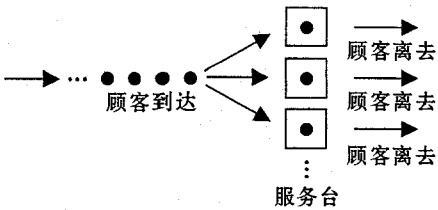


图 6 等待制多服务台系统的模型

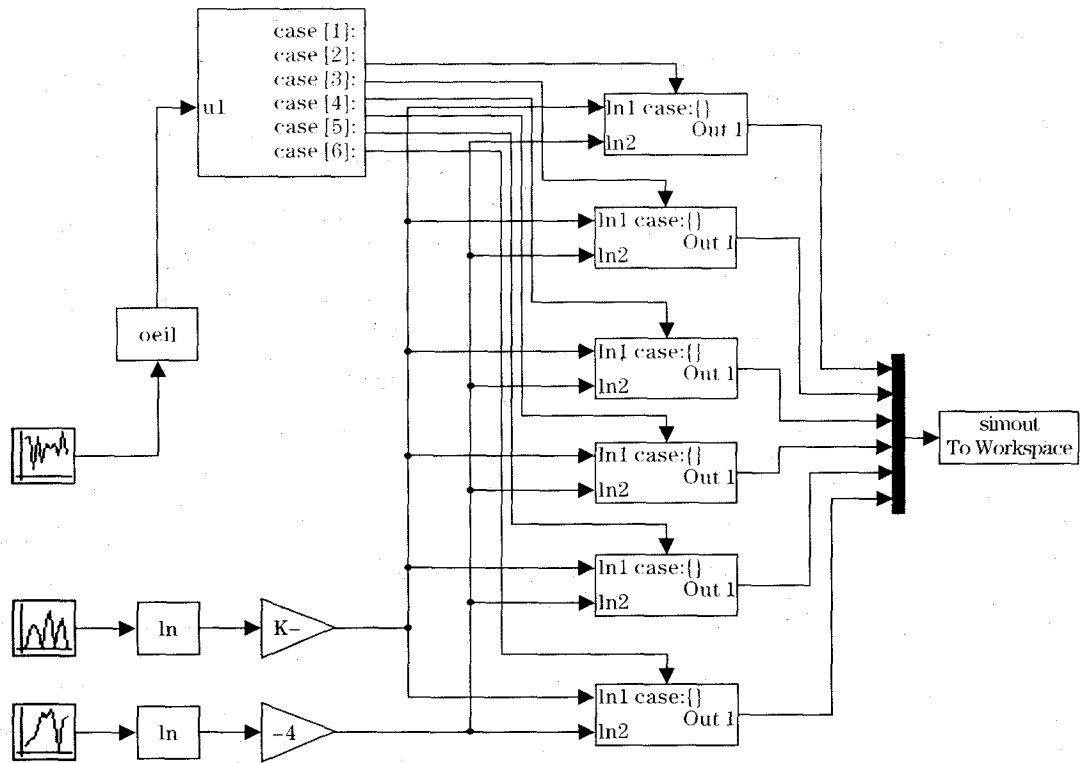


图 7 M/M/6 等待制系统的过程仿真结构图

表 2 1200 位顾客通过 M/M/6 等待制服务系统的仿真数据

	服务台 1	服务台 2	服务台 3	服务台 4	服务台 5	服务台 6
平均服务人数	191	203	211	194	205	196
平均总服务时间	452	495	516	468	488	504
平均总空闲时间	382	393	405	369	414	289
顾客平均等待时间	2.52	3.21	3.41	2.87	2.75	4.51
服务台空闲率	0.45	0.44	0.44	0.44	0.45	0.36

从表 2 可以看出,对于 M/M/6 等待制排队服务系统,各服务台服务的顾客数在 200 左右,与单服务台系统相比较,顾客的平均等待时间较单服务台系统明显减少,服务台的空闲率变大。这说明仿真系统比较客观地反映了实际情况。

考虑到大多数服务系统中对服务台设置不同的服务规则,比如说绝大多数大型的超市收银台系统设置所购物品小于 5 件的快速通道,以及刷卡消费通道。可以这样来进行模拟,所购物品小于 5 件的快速通道相当于设置专门服务台对服务时间较少的顾客提供服务,而刷卡消费通道则相当于设置专门服务台对服务时间较长的顾客提供服务。因此可以产生的随机服务时间序列值的大小来设置服务台,对应的系统仿真结构图如图 8 所示。其他参数与相同服务规则的 M/M/6 等待制服务系统一致。结构图中,服务台 1 为快速通道服务台,服务台 2 为刷卡消费服务台。

统计平均后的仿真结果如表 3 所示。

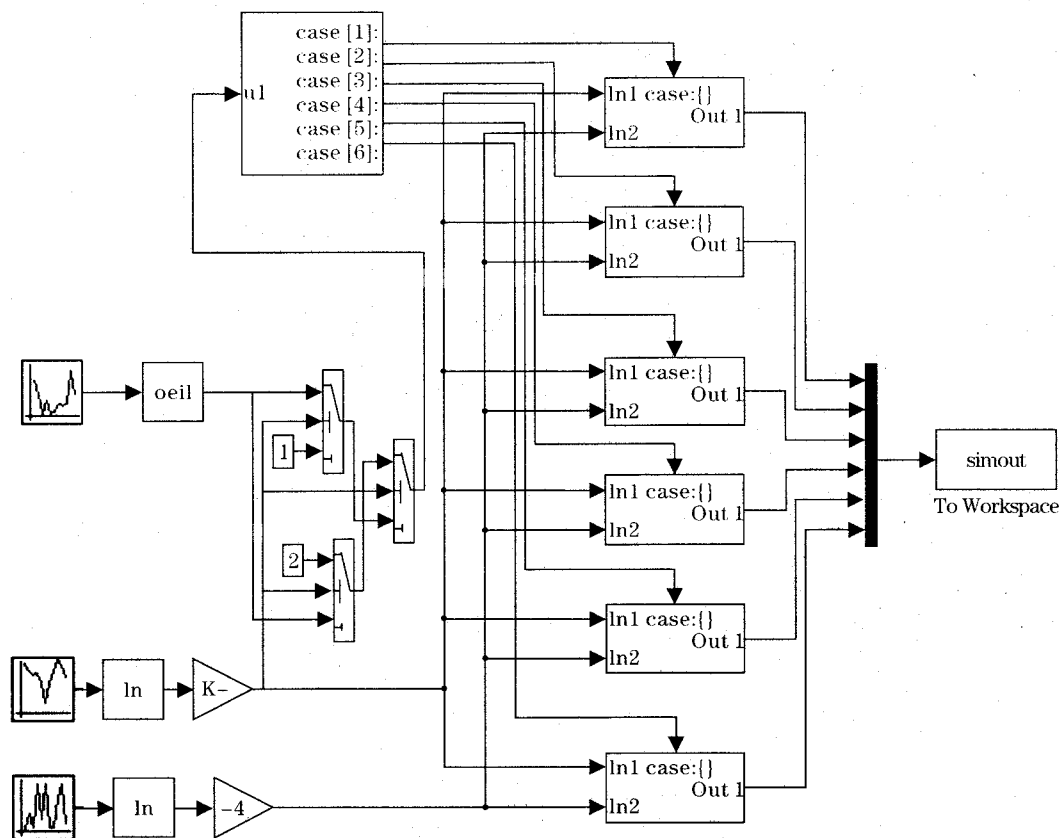


图 8 M/M/6 等待制不同服务规则系统的过程仿真结构图

表 3 1200 位顾客通过 M/M/6 等待制不同服务规则系统的仿真数据

	服务台 1	服务台 2	服务台 3	服务台 4	服务台 5	服务台 6
平均服务人数	247	78	221	226	206	222
平均总服务时间	61	573	477	450	428	474
平均总空闲时间	668	149	451	506	419	535
顾客平均等待时间	0.02	4.18	1.27	1.37	1.54	1.26
服务台空闲率	0.89	0.19	0.48	0.52	0.15	0.53

分析表 2 和表 3 的数据可以看出,采用不同服务规则系统时,顾客的平均等待时间减少了,提高了服务效果,而且服务台的平均空闲率也有所增加。其中快速通道服务台的空闲率较大,是因为产生的随机服务时间列中很小服务时间的值较多所致,这可以通过改变产生的随机数序列加以改善。所以采用不同服务规则时,服务机构可适当减少服务台的数量。从而减少了服务机构的服务成本。

5 结论

排队服务系统应用广泛,对于顾客而言,总是希望能够进入服务系统就立即得到服务,而且在系统中等待的时间越短越好,因而服务台数越多越好,这样顾客花费的时间就少,服务效果就好。但是增加服务台数就会增加服务机构的服务成本,减少其利润。因此,对于一个排队服务系统的设计就需要优化。不可避免地需要对排队服务系统的模型进行求解,然而,对于复杂排队服务系统模型的解析方法求解和数值方法求解常常很难得到。采用此仿真方案,可以对较复杂排队服务系统的设计方案进行仿真,得到较为客观的仿真数据,为服务机构的正确决策提供参考,因而具有较强的实用价值。

参考文献:

- [1] 华兴. 排队论与随机服务系统[M]. 上海: 上海翻译出版公司, 1987.
- [2] 陈良均, 朱庆棠. 随机过程及应用[M]. 北京: 高等教育出版社, 2003.
- [3] 肖田元, 张燕云, 陈加栋. 系统仿真导论[M]. 北京: 清华大学出版社, 2000.

Simulation research of queue model

YAN Gang-feng¹, HUANG Xian-he¹, LI Si-ming²

(1. University of Electronic Science & Technology of China, Chengdu 610064, China; 2. Chengdu University of Technology, Chengdu 610059, China)

Abstract: A simulation project of the queuing system based on the event dispatch method is designed and the simulation research is made. The simulation project is verified by the theory calculation method using M/M/I queuing service system as an example. The preferable service efficiency of the M/M/n queuing service system under different service rules is shown by the compared simulated data of the M/M/n queuing service system under the different service rules and the same service rules. It provides a method of the simulation verification for the optimized design of the complicated queuing service system.

Key words: queuing system; random variable; computer simulation