Учитывая теперь соотношения (14), (16), (17), (21), (23)–(26), получим, что при $\theta$<<1 плотность распределения капитала фонда $P(S)$ имеет вид (после перехода к старому началу координат)

$$P(S) = \\ = \begin{cases} \dfrac{1-e^{-\theta\omega_0(S_2-S_1)}}{2(S_2-S_1)} e^{\theta\omega_0(S-S_1)} + o(\theta), & S < S_1, \\ \dfrac{2-e^{-\theta\omega_0(S-S_1)} - e^{\theta\omega_0(S-S_2)}}{2(S_2-S_1)} + o(\theta), & S_1 \le S \le S_2, \\ \dfrac{1-e^{-\theta\omega_0(S_2-S_1)}}{2(S_2-S_1)} e^{\theta\omega_0(S-S_2)} + o(\theta), & S > S_2. \end{cases} \qquad (28)$$

Построенная аппроксимация (28) решения системы уравнений (2), (4)–(6) может быть улучшена за счет учета дополнительных членов разложения функций $f_i(z,\theta)$ и $\psi_i(z,\theta)$ в ряд по степеням $\theta$.

### Заключение

Предложена и исследована математическая модель деятельности некоммерческого фонда при гистерезисном управлении его капиталом. Получены уравнения, определяющие плотность распределения капитала, найдено решение уравнений при экспоненциальном распределении поступающих в фонд премий и в случае малой нагрузки премии.

### СПИСОК ЛИТЕРАТУРЫ


1. Змеев О.А. Математическая модель фонда социального страхования с детерминированными расходами на социальные программы (диффузионное приближение) // Известия вузов. Физика. – 2003. – Т. 46. – № 3. – С. 83–87.
2. Лившиц К.И., Шифердекер И.Ю. Математическая модель деятельности некоммерческого фонда при релейном управлении капиталом // Вестник Томского государственного университета. Приложение. – 2006. – № 18. – С. 302–308.
3. Лившиц К.И., Шифердекер И.Ю. Диффузионная аппроксимация математической модели деятельности некоммерческого фонда при релейном управлении капиталом // Вестник Томского государственного университета. – 2006. – № 293. – С. 38–44.
4. Лившиц К.И., Сухотина Л.Ю., Шифердекер И.Ю. Пуассоновская модель деятельности некоммерческого фонда при релейном управлением капиталом // Вестник Томского государственного университета. Приложение. – 2006. – № 19. – С. 302–312.
5. Феллер В. Введение в теорию вероятностей и ее приложения. В 2 т. – М.: Мир, 1967. – Т. 1. – 498 с.
6. Глухова Е.В., Змеев О.А., Лившиц К.И. Математические модели страхования. – Томск: Изд-во ТГУ, 2004. – 180 с.


*Поступила 26.10.2009 г.*

УДК 65.012.122

# ОПТИМИЗАЦИЯ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ПЕРЕМЕННОЙ ИНТЕНСИВНОСТЬЮ, ЗАВИСЯЩЕЙ ОТ ВРЕМЕНИ ОЖИДАНИЯ

Л.И. Самочернова

Томский политехнический университет
E-mail: am@am.tpu.ru

*Изучена однолинейная система массового обслуживания с переменной интенсивностью обслуживания, зависящей от времени ожидания заявки, находящейся первой в очереди. Проведена оптимизация системы при учете потерь на ожидание и амортизацию.*

***Ключевые слова:***
*Система, обслуживание, время ожидания, амортизация, оптимальный момент.*
***Key words:***
*System, service, queuing time, depreciation, optimal moment.*

### Введение

Задача изучения управляемых систем массового обслуживания (СМО) является актуальной, поскольку функционирование многих реальных технических систем описывается с их помощью. Большое число работ посвящено изучению систем массового обслуживания, в которых интенсивность обслуживания, моменты включения и отключения резервных приборов, зависят от длины очереди или от числа заявок в системе [1–5]. Класс СМО с управлением по времени ожидания является пока мало изученным, хотя именно такие системы являются наилучшими моделями многих реальных объектов, в частности, вычислительных систем, используемых для обработки медико-биологической информации, систем связи. Существуют лишь отдельные работы, например [6–9], в которых изучены СМО с управлением по времени ожидания.

В данной работе рассмотрена однолинейная система массового обслуживания, аналогичная описанной в [9], но с интенсивностью обслуживания, зависящей от текущего времени ожидания заявки, находящейся первой в очереди.

### 1. Описание системы

Рассмотрим однолинейную систему массового обслуживания с простейшим входящим потоком интенсивности $\lambda$. Если заявка, находящаяся в некоторый произвольный момент времени $t$ первой в очереди поступила в систему в момент времени $t_0$, то величину $s(t)=t-t_0$ будем называть текущим временем ожидания заявки, находящейся первой в очереди. Будем считать, что в интервале времени $(t,t+\Delta t)$ обслуживание заявки закончится с вероятностью $\mu(s)\Delta t+o(\Delta t)$ независимо от предыстории, то есть обслуживание предполагается марковским с интенсивностью $\mu(s)$, где $\mu(s)$ – произвольная дифференцируемая функция. Таким образом, интенсивность обслуживания заявки, находящейся на приборе, меняется со временем и определяется временем ожидания заявки, которая по освобождении прибора поступит на обслуживание. Так как в простейшем потоке заявок отсутствует последействие, то $s(t)$ будет марковским процессом с особыми состояниями $\nu(t)=0$ и $\nu(t)=1$, где через $\nu(t)$ обозначено число требований, находящихся на обслуживании в момент времени $t$, если очередь в этот момент пуста. Обозначим через $p(s)$ финальную плотность вероятностей величины $s$, а через $\pi(\nu)$ – финальную вероятность того, что в очереди заявок нет, а на приборе обслуживаются $\nu$ заявок ($\nu=0$ или 1). Рассматривая переходы за интервал времени $\Delta t$, рассуждениями, аналогичными приведенными в [9], получим условия в нуле

$$p(0)=\lambda\pi(1),$$

$$\left.\frac{dp(s)}{ds}\right|_{s=0}-\lambda p(0)=-\lambda^2\pi(0)$$

и уравнение

$$p(s)=[1-\mu(s-\Delta t)\Delta t]p(s-\Delta t)+$$
$$+\lambda\Delta t\int_0^\infty e^{-\lambda v}\mu(s+v)\cdot p(s+v)dv,$$

где $v$ – значение интервала времени, через который поступило требование, следующее за находящимся в очереди первым. Это уравнение при $\Delta t\to 0$ преобразуется к виду

$$\frac{d^2p(s)}{ds^2}+[\mu(s)-\lambda]\frac{dp(s)}{ds}+\mu'(s)p(s)=0.$$

Общее решение этого дифференциального уравнения имеет вид

$$p(s)=e^{\lambda s-m(s)}[c_1+c_2\int e^{m(s)-\lambda s}ds],$$

где $m(s)=\int_0^s\mu(z)dz$, $c_1$, $c_2$ – произвольные константы.

Для определения четырех неизвестных констант $c_1$, $c_2$, $\pi(0)$, $\pi(1)$ имеем два условия в нуле, условие нормировки

$$\pi(0)+\pi(1)+\int_0^\infty p(s)ds=1$$

и естественное условие на бесконечности

$$\lim_{s\to\infty}p(s)=0.$$

Находя константы $\pi(0)$, $\pi(1)$, $c_1$, $c_2$, получим

$$p(s)=\frac{\lambda^2e^{\lambda s-m(s)}}{\lambda+\mu(0)+\lambda^2\int_0^\infty e^{\lambda s-m(s)}ds},\tag{1}$$

$$\pi(0)=\frac{\mu(0)}{\lambda+\mu(0)+\lambda^2\int_0^\infty e^{\lambda s-m(s)}ds},$$

$$\pi(1)=\frac{\lambda}{\lambda+\mu(0)+\lambda^2\int_0^\infty e^{\lambda s-m(s)}ds}.\tag{2}$$

Заметим, что для полученной финальной плотности вероятностей $p(s)$ и финальных вероятностей $\pi(0)$, $\pi(1)$ условие нормировки выполняется, если

$$\int_0^\infty e^{\lambda s-m(s)}ds<\infty.\tag{3}$$

Условие (3) является достаточным условием существования стационарного режима работы системы. Если $\int_0^\infty e^{\lambda s-m(s)}ds$ расходится, то финальное распределение вероятностей не существует.

### 2. Оптимизация системы

Рассмотрим случай, когда в описанной выше системе имеют место потери только двух видов.

#### А. Потери от ожидания заявок в очереди

Нахождение в очереди заявок, ожидающих обслуживания, приводит к потерям, которые будем называть *потерями от ожидания*. Пусть потери от ожидания определяются числом заявок $i$, находящихся в очереди. Будем считать, что если в очереди в течение единицы времени находится $i$ заявок, то потери от ожидания в единицу времени равны $F(i)$. Найдем математическое ожидание $MF(i)$ этих потерь. Пусть в какой-то момент времени заявка, находящаяся в очереди первой, ожидает обслуживания в течение времени $s$. За это время с вероятностью $\frac{(\lambda s)^{i-1}e^{-\lambda s}}{(i-1)!}$ пришла еще $i-1$ заявка. Таким образом, условное распределение числа заявок в очереди $i$ есть

$$p(i/s)=\frac{(\lambda s)^{i-1}e^{-\lambda s}}{(i-1)!},\quad i=1,2,...$$

В этом случае безусловное распределение $p(i)$ числа заявок в очереди имеет вид

$$p(i)=\int_0^\infty p(i/s)p(s)ds=\int_0^\infty p(s)\frac{(\lambda s)^{i-1}e^{-\lambda s}}{(i-1)!}ds.$$

Тогда среднее значение потерь на ожидание $L_1$ можно записать как

$$L_1=MF(i)=\int_0^\infty \tilde{F}(s)p(s)ds,$$

где

$$\tilde{F}(s)=\sum_{i=1}^\infty F(i)\frac{(\lambda s)^{i-1}e^{-\lambda s}}{(i-1)!}. \qquad (4)$$

В частности, если $F(i)=c_1 i$, то $\tilde{F}(s)=c_1(\lambda s+1)$, где $c_1$=const>0, имеющая смысл потерь от ожидания одной заявки в течение единицы времени.

Б. Потери на амортизацию обслуживающего устройства

Будем считать, что при работе прибора с интенсивностью $\mu(s)$ в системе в единицу времени возникают *потери на амортизацию*, равные $f(\mu(s))$, где $f(\mu)$ – монотонно возрастающая, трижды дифференцируемая функция. Тогда среднее значение потерь на амортизацию в единицу времени равно

$$L_2=\int_0^\infty f(m'(s))p(s)ds+\pi(1)f(m'(0)),$$

где учтено, что $\mu(s)=m'(s)$.

Таким образом, средние суммарные потери системы в единицу времени равны

$$L=L_1+L_2=\int_0^\infty \tilde{F}(s)p(s)ds+$$
$$+\int_0^\infty f(m'(s))p(s)ds+\pi(1)f(m'(0)).$$

Тогда, учитывая (1), (2), получаем, что задача оптимизации СМО сводится к минимизации функционала

$$L(m(s))=$$
$$=\frac{\int_0^\infty e^{\lambda s-m(s)}[\tilde{F}(s)+f(m'(s))]ds+f(m'(0))/\lambda}{\int_0^\infty e^{\lambda s-m(s)}ds+m'(0)/\lambda^2+1/\lambda} \qquad (5)$$

с начальными условиями $m(0)=0,\ m'(s)=\mu_1$.

Обозначим через $m_{opt}(s)$ такую функцию $m(s)$, которая доставляет минимум функционалу (5). Значение знаменателя в (5) при $m(s)=m_{opt}(s)$ будем обозначать через $\alpha^*$, т. е.

$$\alpha^*=\int_0^\infty e^{\lambda s-m_{opt}(s)}ds+(\lambda+\mu_1)/\lambda^2.$$

С учетом введенных обозначений задачу оптимизации функционала (5) можно сформулировать следующим образом. Найти минимум функционала

$$J[m(s]=\int_0^\infty e^{\lambda s-m(s)}[\tilde{F}(s)+f(m'(s))]ds+\frac{f(\mu_1)}{\lambda}$$

при дополнительном ограничении

$$\int_0^\infty e^{\lambda s-m(s)}ds+\frac{(\lambda+\mu_1)}{\lambda^2}=\alpha^*. \qquad (6)$$

Из вариационного исчисления известно, что для решения такой задачи надо написать уравнение Эйлера для вспомогательного функционала

$$V[m(s)]=\int_0^\infty e^{\lambda s-m(s)}[\tilde{F}(s)+f(m'(s))-\alpha]ds+$$
$$+\frac{f(\mu_1)}{\lambda}-\frac{\alpha(\lambda+\mu_1)}{\lambda^2}, \qquad (7)$$

где $\alpha$ – множитель Лагранжа. Учитывая, что $m'(s)=\mu(s)$, уравнение Эйлера для функционала (7) запишем в виде

$$\mu'(s)f''(\mu)+(\lambda-\mu(s))f'(\mu)+$$
$$+f(\mu)+\tilde{F}(s)-\alpha=0. \qquad (8)$$

Граничное условие имеет вид $\mu(0)=\mu_1$.

Заметим, что $\alpha^*$ заранее не известно и находится следующим образом. Полагая в (8) $\alpha$ произвольным, но фиксированным, найдем $\mu(s)$, соответствующее этому значению $\alpha$. Подставив в (5) найденное $\mu(s)$, получим значение $L$, соответствующее этому $\alpha$. Рассматривая $L$ как функцию $\alpha$, найдем такое значение $\alpha=\alpha_{min}$, при котором функция $L(\alpha)$ достигает минимума. Тогда $\alpha^*$ определится из (6) при подстановке соответствующей $\mu(s)$. Аналитически решить уравнение (8) не удается даже в случае, когда $f(\mu)$ и $\tilde{F}(s)$ достаточно простые функции. Эта задача решена численно.

Случай $f(\mu)=c_2\mu$ ($c_2$=const>0) является особым. При этом (8) вырождается в уравнение

$$\lambda c_2-\alpha+\tilde{F}(s)=0.$$

Это говорит о том, что решение вариационной задачи в этом случае находится в классе разрывных функций, для которых методика классического вариационного исчисления неприменима.

Этот особый случай рассмотрен с использованием принципа максимума Понтрягина [10]. Будем считать дополнительно, что $\mu(s)$ удовлетворяет условию $\mu_1\le\mu(s)\le\mu_2$, а $\tilde{F}(s)$ является монотонно возрастающей функцией от $s$. Применяя методику принципа максимума, получаем, что решение имеет вид

$$\mu(s)=\begin{cases}\mu_1 \text{ при } s\le s_0,\\ \mu_2 \text{ при } s>s_0,\end{cases}$$

т. е. оптимальной является двухуровневая СМО.

### Исследование двухуровневой СМО с интенсивностью обслуживания, зависящей от текущего времени ожидания заявки, находящейся первой в очереди

Обозначим финальную плотность вероятностей $p(s)$ в области $0\le s\le s_0$ через $p_1(s)$, а в области $s\ge s_0$ через $p_2(s)$. Из (1) получим при $\lambda\ne\mu_1$

$$p_1(s)=\frac{\lambda^2(\lambda-\mu_1)(\lambda-\mu_2)\exp[(\lambda-\mu_1)s]}{\lambda^2(\mu_1-\mu_2)\exp[(\lambda-\mu_1)s_0]-\mu_1^2(\lambda-\mu_2)},$$

$$p_2(s)=$$

$$=\frac{\lambda^2(\lambda-\mu_1)(\lambda-\mu_2)\exp[(\lambda s-\mu_1 s_0-\mu_2(s-s_0)]}{\lambda^2(\mu_1-\mu_2)\exp[(\lambda-\mu_1)s_0]-\mu_1^2(\lambda-\mu_2)}. \quad (9)$$

При $\lambda=\mu_1$, формулы (9) не пригодны, т. к. имеется неопределенность, раскрывая которую, получим

$$p_1(s)=\frac{\lambda(\lambda-\mu_2)}{(\lambda-\mu_2)(1+\lambda s_0)-\mu_2},$$

$$p_2(s)=\frac{\lambda(\lambda-\mu_2)\exp[(\lambda-\mu_2)(s-s_0)]}{(\lambda-\mu_2)(1+\lambda s_0)-\mu_2}.$$

Если считать, что работа обслуживающего прибора с большей интенсивностью $\mu_2$ вызывает в единицу времени потери, равные $c_2$, то функция потерь имеет вид

$$L(s_0)=\int_0^{s_0}\tilde{F}(s)p_1(s)ds+$$

$$+\int_{s_0}^{\infty}\tilde{F}(s)p_2(s)ds+c_2\int_{s_0}^{\infty}p_2(s)ds. \quad (10)$$

Эти потери зависят от параметра $s_0$, и оптимизация такой системы сводится к нахождению такого момента включения большей интенсивности обслуживания, который минимизирует функцию потерь (10).

Уравнения для определения оптимального момента включения большей интенсивности обслуживания имеют вид при $\rho_1\neq 1$:

$$c_2=\frac{(\rho_1-\rho_2)}{\rho_2^2(1-\rho_1)}\left\{\rho_1\rho_2(\rho_1-1)\int_0^{x_0}\tilde{F}\left(\frac{x}{\lambda}\right)e^{\frac{(\rho_1-1)x}{\rho_1}}dx+\right.$$

$$\left.+(\rho_2-1)\left(\rho_1^2 e^{\frac{(\rho_1-1)x_0}{\rho_1}}-1\right)e^{\frac{(1-\rho_2)x_0}{\rho_2}}\int_{x_0}^{\infty}\tilde{F}\left(\frac{x}{\lambda}\right)e^{\frac{(\rho_2-1)x}{\rho_2}}dx\right\};$$

при $\rho_1=1$:

$$c_2=\frac{(\rho_2-1)}{\rho_2^2}\left\{\rho_2\int_0^{x_0}\tilde{F}\left(\frac{x}{\lambda}\right)dx+\right.$$

$$\left.+(\rho_2-1)(2+x_0)e^{-\frac{(\rho_2-1)x_0}{\rho_2}}\int_{x_0}^{\infty}\tilde{F}\left(\frac{x}{\lambda}\right)e^{\frac{(\rho_2-1)x}{\rho_2}}dx\right\},$$

$$\text{где } \rho_i=\frac{\lambda}{\mu_i},\ (i=1,2),\ x_0=\lambda s_0.$$

Доказано, что если $\tilde{F}(s)$ положительная монотонно возрастающая функция, то эти уравнения либо не имеют положительного корня, либо имеют единственный, положительный корень, в котором $L(s_0)$ принимает свое минимальное значение. Отсутствие корня означает, что прибор должен все время работать с максимальной интенсивностью $\mu_2$.

**Пример.** Рассмотрим случай, когда потери от ожидания в очереди пропорциональны числу заявок, находящихся в очереди, т. е. $F(i)=c_1 i$, где $c_1$ – положительная константа, имеющая смысл потерь от ожидания одной заявки в течение единицы времени. Тогда из (4) получим, что $\tilde{F}(s)=c_1(\lambda s+1)$. В этом случае уравнения для определения оптимального момента включения большей интенсивности обслуживания $\mu_2$ при $\rho_1\neq 1$ и $\rho_1=1$ имеют соответственно вид:

$$\frac{c_1(\rho_1-\rho_2)}{(\rho_1-1)(\rho_2-1)}\{\rho_1^2(\rho_2-1)+(\rho_1-1)\times$$

$$\times[(\rho_2-1)(x_0+1)-\rho_2]+\rho_1^2(\rho_1-\rho_2)e^{\frac{(\rho_1-1)x_0}{\rho_1}}\}+$$

$$+c_2\rho_2(\rho_1-1)=0, \quad (11)$$

$$c_1\{4+x_0[2+(2+x_0)(1-\rho_2)]\}-2\rho_2 c_2=0. \quad (12)$$

Полагая в уравнениях (11), (12) $x_0=0$ и выражая $c_2/c_1$, получим при $\rho_1\neq 1$:

$$\frac{c_2}{c_1}=\frac{(1+\rho_1)(\rho_1-\rho_2)}{\rho_2(1-\rho_2)},$$

а при $\rho_1=1$:

$$\frac{c_2}{c_1}=\frac{2}{\rho_2}.$$

Если при $\rho_1\neq 1$ выполняется неравенство

$$\frac{c_2}{c_1}\le\frac{(1+\rho_1)(\rho_1-\rho_2)}{\rho_2(1-\rho_2)}, \quad (13)$$

то уравнение (11) не имеет положительного корня. Это означает, что при выполнении неравенства (13) прибор все время должен работать с большей интенсивностью $\mu_2$. Если неравенство (13) не выполняется, то уравнение (11) имеет единственный , положительный корень. Аналитически найти оптимальный момент включения большей интенсивности не удается даже в частном случае, когда $F(i)=c_1 i$. Уравнение (11) решено численно.

Совершенно аналогичными рассуждениями при $\rho_1=1$ получаем, что если

$$\frac{c_2}{c_1}\le\frac{2}{\rho_2}, \quad (14)$$

то нужно сразу же включать большую интенсивность обслуживания $\mu_2$. В случае, когда неравенство (14) не выполняется, существует положительный оптимальный момент включения большей интенсивности обслуживания $x_{opt}$, который находится из квадратного уравнения (12) и имеет вид

$$x_{opt}=\frac{2-\rho_2-\sqrt{(2-\rho_2)^2+2(1-\rho_2)(\rho_2 c_2/c_1-2)}}{\rho_2-1}.$$

Данная работа может быть полезна при оптимизации систем, где важна не длина очереди, а время ожидания заявок в очереди или время ожидания заявок в системе.

### Выводы

1. Получены явный вид финальной плотности вероятностей $p(s)$ и финальные вероятности $\pi(0)$, $\pi(1)$ особых состояний системы, а также явный вид функционала потерь для рассматриваемой системы массового обслуживания с управлением по текущему времени ожидания при учете потерь на ожидание и амортизацию обслуживающего прибора.
2. Доказано, что оптимальной является двухуровневая система массового обслуживания, получены уравнения для нахождения оптимального момента включения большей интенсивности обслуживания.
3. Для частного случая, когда потери от ожидания в очереди пропорциональны числу заявок, находящихся в очереди, получен оптимальный момент включения большей интенсивности обслуживания при $\rho_1=1$. Для случая $\rho_1\neq1$ выведено условие, когда прибор постоянно должен работать с большей интенсивностью обслуживания.

### СПИСОК ЛИТЕРАТУРЫ


1. Рыков В.В. Управляемые системы массового обслуживания // В кн.: Теория вероятностей. Математическая статистика. Теоретическая кибернетика. Итоги науки и техники. ВИНИТИ АН СССР. – М., 1975. – Т. 12. – С. 43–153.
2. Соловьев А.Д. Задача об оптимальном обслуживании // Техническая кибернетика. – 1970. – № 5. – С. 40–49.
3. Коваленко И.Н. О СМО со скоростью обслуживания, зависящей от числа требований в системе и периодическим отключением каналов // Проблемы передачи информации. – 1971. – Т. 7. – № 2. – С. 106–111.
4. Горцев А.М., Назаров А.А., Терпугов А.Ф. Управление и адаптация в системах массового обслуживания. – Томск: Изд-во ТГУ, 1978. – 208 с.
5. Горцев А.М., Катаева С.С. Оптимизация гистерезисного управления резервным каналом в вычислительной системе с двумя ЭВМ // Техника средств связи. Серия системы связи. – 1990. – Вып. 7. – С. 3–8.
6. Либура М. Об одноканальной системе обслуживания с простейшим входящим потоком и зависимостью времени обслуживания от времени ожидания в очереди // Проблемы передачи информации. – 1972. – Т. 8. – № 4. – С. 104–107.
7. Оптимизация СМО с симметричным резервным прибором / Самочернова Л.И., Храмова Ю.В.; Томск. политехн. ун-т. – Томск, 1996. – 8 с. – Рус. – Деп. в ВИНИТИ 26.06.1996, № 2127 – В96.
8. Переходной режим работы СМО с гистерезисной стратегией управления однотипным, симметричным резервным прибором / Самочернова Л.И., Назаренко Е.А.; Томск. политехн. ун-т. – Томск, 2000. – 10 с. – Рус. – Деп. в ВИНИТИ 2.08.2000, № 2158 – В00.
9. Зиновьева Л.И., Терпугов А.Ф. Однолинейная система массового обслуживания с переменной интенсивностью, зависящей от времени ожидания // Автоматика и телемеханика. – 1981. – № 1. – С 27–30.
10. Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф. Математическая теория оптимальных процессов. – М.: Наука, 1976. – 392 с.


*Поступила 15.10.2009 г.*