СПИСОК ЛИТЕРАТУРЫ

1. ISO/IEC 11172-2: Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – P. 2: Video: англ. – ISO/IEC, 1993.
2. Lee J.-B., Kalva H. The VC-1 and H.264 video compression standards for broadband video services: англ. – N.Y.: Springer publishing, 2008. – 496 p.
3. SMPTE 421M-2006: VC-1 compressed video bitstream format and decoding process: англ. – SMPTE, 2006.
4. MainConcept VC-1 Pro // MainConcept: Information. 2009. URL: http://www.mainconcept.com/site/prosumer-products-4/vc-1-pro-20250/information-20270.html (дата обращения: 14.10.2009).

Поступила 12.02.2010 г.

УДК 65.012.122

ОПТИМИЗАЦИЯ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С РЕЗЕРВНЫМ ПРИБОРОМ С УПРАВЛЕНИЕМ, ЗАВИСЯЩИМ ОТ ВРЕМЕНИ ОЖИДАНИЯ

Л.И. Самочернова

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

Изучена система обработки информации, функционирование которой описано математической моделью системы массового обслуживания с резервным прибором, управляемым по текущему 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

Значительное число работ [1–10] посвящено изучению управляемых систем массового обслуживания (УСМО). Это связано с тем, что именно УСМО удается описывать функционирование многих реальных технических систем, в частности, систем связи. Во многих работах изучаются системы массового обслуживания (СМО), в которых интенсивность обслуживания, моменты включения и отключения резервного прибора зависят от числа заявок в системе или от длины очереди. Однако системы, в которых стратегия управления резервным прибором зависит от времени ожидания, остались изученными очень слабо.

Данная статья посвящена оптимизации системы обработки информации, функционирование которой описано математической моделью УСМО с резервным прибором, управляемым по текущему времени ожидания заявки, находящейся первой в очереди.

1. Описание системы

По некоторому каналу связи поступают сообщения, которые должны обрабатываться ЭВМ. Эти сообщения посылаются большим количеством источников информации, так что общий поток таких сообщений можно считать пуассоновским. После предварительной обработки эти сообщения поступают для дальнейшей обработки на специализированную ЭВМ, которая в каждый момент времени может обрабатывать только одно сообщение. Оставшиеся сообщения записываются на диске и образуют очередь. Кроме основной имеется резервная ЭВМ, которая вводится, в основном, для повышения надежности всей системы. На обрабатываемые сообщения накладывается требование обработать их к определенному сроку. Поэтому, когда сообщений в очереди оказывается слишком много, и время ожидания сообщения, находящегося в очереди первым, достигает некоторой пороговой величины, резервная ЭВМ подключается к их обработке. Эту реальную систему опишем следующей математической моделью УСМО с резервным прибором.

Рассмотрим однолинейную систему массового обслуживания с простейшим входящим потоком интенсивности λ, к которой может подключаться резервный прибор. Обслуживание предполагается экспоненциальным с интенсивностями μ_1 и μ_2 соответственно для основного и резервного приборов. 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. Дисциплина обслуживания резервным прибором следующая: как только s – текущее время ожидания заявки, находящейся в очереди первой, достигает величины s_0 (s_0=const>0), подключается резервный прибор и берет на обслуживание заявку, стоящую первой в очереди. После обслуживания одной заявки он выключается, если $s<s_0$, но продолжает работать, если $s\geq s_0$. Необходимо найти такой оптимальный момент подключения s_0^{opt} резервного прибора, который бы минимизировал средние суммарные потери такой системы в единицу времени.

Опишем рассматриваемую систему массового обслуживания случайным процессом с компонентами $\{s(t), \nu_1(t), \nu_2(t)\}$, где $s(t)$ – текущее время ожидания заявки, находящейся первой в очереди, $\nu_1(t)$ – число требований, находящихся в момент времени t на обслуживании на основном приборе, а $\nu_2(t)$ – число требований, находящихся в момент времени t на обслуживании на резервном приборе, $\nu_1(t)=0,1$; $\nu_2(t)=0,1$. Кроме того, возможны еще четыре особых состояния: $\{\nu_1(t)=0,\ \nu_2(t)=0\}$ (система пуста); $\{\nu_1(t)=0,\ \nu_2(t)=1\}$ (очередь пуста и работает только резервный прибор); $\{\nu_1(t)=1,\ \nu_2(t)=0\}$ (очередь пуста и работает только основной прибор); $\{\nu_1(t)=1,\ \nu_2(t)=1\}$ (очередь пуста и работают оба прибора: и основной, и резервный).

Рассматривая возможные переходы за бесконечно малый промежуток времени в заданное состояние, можно показать аналогично тому, как сделано в [7], что вероятности переходов не зависят от предыстории, так как поток заявок простейший, а обслуживание экспоненциальное. Следовательно, процесс $\{s(t),\ \nu_1(t),\ \nu_2(t)\}$ с особыми состояниями $\{\nu_1(t)=0, \nu_2(t)=0\}$, $\{\nu_1(t)=0, \nu_2(t)=1\}$, $\{\nu_1(t)=1, \nu_2(t)=0\}$, $\{\nu_1(t)=1,\ \nu_2(t)=1\}$ является марковским случайным процессом. Предположим, что существует стационарное распределение вероятностей, которое, как известно, совпадает с финальным. Достаточным условием существования стационарного режима работы рассматриваемой системы массового обслуживания является условие [7]: $\lambda<\mu_1+\mu_2$.

Финальную плотность вероятностей $p(s,\ \nu_1,\ \nu_2)$ величины $(s,\ \nu_1,\ \nu_2)$ обозначим через $p_1(s)$ в области $0\leq s\leq s_0$, если $\nu_1=1$, $\nu_2=0$; $p_2(s)$ области $0\leq s\leq s_0$, если $\nu_1=1$, $\nu_2=1$; $p_3(s)$ в области $s\geq s_0$, для которой $\nu_1=1$, $\nu_2=1$. Обозначим через $\pi(0,0)$ – финальную вероятность особого состояния $\{\nu_1(t)=0,\ \nu_2(t)=0\}$ (вероятность того, что система пуста); $\pi(0,1)$ – финальную вероятность того, что система находится в состоянии $\{\nu_1(t)=0,\ \nu_2(t)=1\}$, когда очередь пуста, основной прибор не работает, а работает только резервный прибор; $\pi(1,0)$ – финальную вероятность того, что в очереди заявок нет, резервный прибор не работает, а работает только основной прибор, то есть вероятность особого состояния $\{\nu_1(t)=1,\ \nu_2(t)=0\}$; $\pi(1,1)$ – финальную вероятность того, что очередь пуста, а основной и резервный приборы обслуживают заявки, то есть вероятность особого состояния $\{\nu_1(t)=1,\ \nu_2(t)=1\}$.

Как показано в [7], финальные плотности вероятностей $\rho_i(s)$ $(i=\overline{1,3})$ и финальные вероятности особых состояний $\pi(0,0)$, $\pi(0,1)$, $\pi(1,0)$, $\pi(1,1)$ имеют следующий вид:

$$
\begin{gathered}
p_1(s)=C_1e^{(\lambda-\mu_1)s}+C_2-C_3e^{z_1s}-C_4e^{-z_2s},\\
p_2(s)=C_3e^{z_1s}+C_4e^{-z_2s},\\
p_3(s)=C_5e^{(\lambda-\mu_1-\mu_2)s}+C_6,
\end{gathered}
\tag{1}
$$

где

$$
z_1=\frac{\lambda-\mu_1-\mu_2+\sqrt{(\mu_1+\mu_2-\lambda)^2+4\lambda\mu_2}}{2},
$$

$$
z_2=\frac{\mu_1+\mu_2-\lambda+\sqrt{(\mu_1+\mu_2-\lambda)^2+4\lambda\mu_2}}{2}.
$$

При $\lambda\neq\mu_1$ C_i $(i=\overline{1,6})$, $\pi(0,0)$, $\pi(0,1)$, $\pi(1,0)$, $\pi(1,1)$ имеют следующий вид:

$$
\left.
\begin{gathered}
C_1=\frac{\lambda^2(\lambda-\mu_1)(\lambda-\mu_1-\mu_2)}{m}\times\\
\times[a_1e^{z_1s_0}(z_1+\mu_1+\mu_2)-a_2e^{-z_2s_0}(z_2-\mu_1-\mu_2)],\\
C_2=0,\\
C_3=\frac{\lambda^3(\lambda-\mu_1)(\lambda-\mu_1-\mu_2)a_1e^{(\lambda-\mu_1)s_0}}{m},\\
C_4=\frac{a_2C_3}{a_1},\\
C_5=\frac{\lambda^3(\lambda-\mu_1)(\lambda-\mu_1-\mu_2)}{m}\times\\
\times[a_1e^{(z_1+\mu_2)s_0}+a_2e^{(\mu_2-z_2)s_0}],\\
\pi(0,0)=\frac{\mu_1(\lambda-\mu_1)(\lambda-\mu_1-\mu_2)}{m}\times\\
\times[a_1(z_1+\mu_1+\mu_2)e^{z_1s_0}-a_2(z_2-\mu_1-\mu_2)e^{-z_2s_0}\\
-\lambda^2(z_1+z_2)e^{(\lambda-\mu_1)s_0}],\\
\pi(0,1)=\\
=\frac{\lambda^2\mu_1(\lambda-\mu_1)(\lambda-\mu_1-\mu_2)(z_1+z_2)e^{(\lambda-\mu_1)s_0}}{m},\\
\pi(1,1)=\frac{(\lambda+\mu_2)\pi(0,1)}{\mu_1},\\
\pi(1,0)=\frac{\lambda(\lambda-\mu_1)(\lambda-\mu_1-\mu_2)}{m}\times\\
\times\begin{bmatrix}a_1(z_1+\mu_1+\mu_2)e^{z_1s_0}-a_2(z_2-\mu_1-\mu_2)e^{-z_2s_0}-\\-\lambda(\lambda+\mu_2)(z_1+z_2)e^{(\lambda-\mu_1)s_0}\end{bmatrix},
\end{gathered}
\right\}
\tag{2}
$$

где

$$
\begin{gathered}
a_1=(\lambda+z_2)(\lambda+\mu_2)-\lambda\mu_1,\\
a_2=(z_1-\lambda)(\lambda+\mu_2)+\lambda\mu_1,\\
m=a_1e^{z_1s_0}\times\\
\times\left\{\begin{matrix}\lambda^2e^{(\lambda-\mu_1)s_0}\begin{bmatrix}(\lambda-\mu_1-\mu_2)(z_1+\mu_1+\mu_2)-\\-\lambda(\lambda-\mu_1)\end{bmatrix}-\\-\mu_1^2(\lambda-\mu_1-\mu_2)(z_1+\mu_1+\mu_2)\end{matrix}\right\}-\\
-a_2e^{-z_2s_0}\left\{\begin{matrix}\lambda^2e^{(\lambda-\mu_1)s_0}\begin{bmatrix}(\lambda-\mu_1-\mu_2)(z_2-\mu_1-\mu_2)+\\+\lambda(\lambda-\mu_1)\end{bmatrix}-\\-\mu_1^2(\lambda-\mu_1-\mu_2)(z_2-\mu_1-\mu_2))\end{matrix}\right\}.
\end{gathered}
$$

При $\lambda=\mu_1$ формулы (2) непригодны, так как имеется неопределенность, раскрывая которую находим

$$\left.\begin{aligned}
C_1 &= \frac{\mu_1\mu_2}{\bar{\bar{m}}}\begin{bmatrix} a_1 e^{z_1 s_0}(z_1+\mu_1+\mu_2)- \\ -a_2 e^{-z_2 s_0}(z_2-\mu_1-\mu_2)\end{bmatrix}, \\
C_2 &= 0, \\
C_3 &= \frac{\mu_1^2\mu_2 a_1}{\bar{\bar{m}}}, \\
C_4 &= \frac{a_2 C_3}{a_1}, \\
C_5 &= \frac{\mu_1^2\mu_2}{\bar{\bar{m}}}[a_1 e^{(z_1+\mu_2)s_0} + a_2 e^{(\mu_2-z_2)s_0}], \\
\pi(0,0) &= \frac{\mu_2}{\bar{\bar{m}}}\begin{bmatrix} e^{z_1 s_0} a_1(z_1+\mu_1+\mu_2)- \\ -e^{-z_2 s_0} a_2(z_2-\mu_1-\mu_2)- \\ -\mu_1^2(z_1+z_2)\end{bmatrix}, \\
\pi(0,1) &= \frac{\mu_1^2\mu_2(z_1+z_2)}{\bar{\bar{m}}}, \\
\pi(1,0) &= \frac{\mu_2}{\bar{\bar{m}}}\begin{bmatrix} e^{z_1 s_0} a_1(z_1+\mu_1+\mu_2)- \\ -e^{-z_2 s_0} a_2(z_2-\mu_1-\mu_2)- \\ -\mu_1(z_1+z_2)(\mu_1+\mu_2)\end{bmatrix}, \\
\pi(1,1) &= \frac{(\mu_1+\mu_2)\pi(0,1)}{\mu_1},
\end{aligned}\right\} \quad (3)$$

где

$$\bar{\bar{m}} = a_1 e^{z_1 s_0}[\mu_2(\mu_1 s_0+2)(z_1+\mu_1+\mu_2)+\mu_1^2]-$$
$$-a_2 e^{-z_2 s_0}[\mu_2(\mu_1 s_0+2)(z_2-\mu_1-\mu_2)-\mu_1^2].$$

2. Оптимизация системы

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

А. Потери от ожидания в очереди необработанных сообщений

Среднее значение потерь на ожидание можно записать в виде

$$L_1 = M\{f(s)+\sum_{i=1}^{n} f(s_i)\},$$

где s – текущее время ожидания сообщения, стоящего в очереди первым, а s_i ($i=\overline{1,n}$) – текущие времена ожидания последующих сообщений. Функция $f(s)$ бралась в виде: $f(s)=D_1 U(s-T)$, где D_1 – положительная константа, имеющая смысл потерь от ожидания одного сообщения в единицу времени;

$$U(x) = \begin{cases} x, \text{ если } x>0, \\ 0, \text{ если } x\le 0. \end{cases}$$

Величина T имеет смысл допустимого времени ожидания обработки сообщений.

Найдем функцию

$$F(s) = M\{\sum_{i=1}^{n} f(s_i)/s\} + f(s).$$

Если сообщение, стоящее в очереди первым, ждет время s, то за это время с вероятностью $\frac{(\lambda s)^n e^{-\lambda s}}{n!}$ поступило еще n сообщений. Обозначая через $s_1,\ldots,s_n$ их текущие времена ожидания, учитывая пуассоновский характер потока сообщений из условия нормировки получим, что $p(s_1,\ldots,s_n/s,n) = \frac{n!}{s^n}$.

Находя $p(s_i/s,n) = \frac{s_i^{i-1}(s-s_i)^{n-i}n!}{s^n(i-1)!(n-i)!}$, усредняя затем по n и s_i, получим

$$F(s) =$$
$$= f(s) + \sum_{n=1}^{\infty}\frac{(\lambda s)^n}{n!}e^{-\lambda s}\sum_{i=1}^{n}\int_0^s f(x)\frac{x^{i-1}(s-x)^{n-i}n!dx}{s^n(i-1)!(n-i)!} =$$
$$= f(s) + \lambda\int_0^s f(x)dx.$$

В частности, при $f(x)=D_1 U(x-T)$:

$$\int_0^s f(x)dx = \begin{cases} \frac{D_1}{2}(s-T)^2, \text{ если } s>T; \\ 0, \text{ если } s\le T. \end{cases}$$

Тогда

$$F(s) = D_1\left[U(s-T) + \frac{\lambda}{2}U^2(s-T)\right].$$

Таким образом, средние потери от ожидания сообщений в очереди имеют вид

$$L_1 = D_1\left\{\begin{aligned} &\int_T^{s_0}\left[\frac{\lambda(s-T)^2}{2}+s-T\right][p_1(s)+p_2(s)]ds + \\ &+\int_{s_0}^{\infty}\left[\frac{\lambda(s-T)^2}{2}+s-T\right]p_3(s)ds \end{aligned}\right\}, \quad (4)$$

где $p_1(s)$, $p_2(s)$, $p_3(s)$ при $\lambda\ne\mu_1$ и $\lambda=\mu_1$ имеют соответственно вид (1), (2) и (1), (3).

Б. Потери на амортизацию резервного прибора (резервной ЭВМ)

Потери на амортизацию резервной ЭВМ можно записать в виде

$$L_2 = D_2\left[\pi(0,1)+\pi(1,1)+\int_0^{s_0}p_2(s)ds+\int_0^{\infty}p_3(s)ds\right]. \quad (5)$$

где D_2 – потери от работы резервного прибора (резервной ЭВМ) в единицу времени, а $p_2(s)$, $p_3(s)$, $\pi(0,1)$, $\pi(1,1)$ при $\lambda\ne\mu_1$ и $\lambda=\mu_1$ имеют соответственно вид (1), (2) и (1), (3).

Таким образом, с учетом (4) и (5) средние суммарные потери системы массового облуживания в единицу времени примут вид

$$L(s_0) =$$

$$= D_1 \left\{ \begin{aligned} &\int_T^{s_0} \left[\frac{\lambda(s-T)^2}{2} + s - T \right] [p_1(s) + p_2(s)] ds + \\ &+ \int_{s_0}^{\infty} \left[\frac{\lambda(s-T)^2}{2} + s - T \right] p_3(s) ds \end{aligned} \right\} +$$

$$+ D_2 [\pi(0,1) + \pi(1,1) + \int_0^{s_0} p_2(s) ds + \int_{s_0}^{\infty} p_3(s) ds]. \quad (6)$$

Задача оптимизации такой системы массового обслуживания сводится к нахождению момента s_0^{opt} включения резервной ЭВМ, минимизирующего функцию потерь (6). Эта задача решена численно. Для проверки корректности решения задачи проведены численные расчеты по нахождению оптимального момента подключения s_0^{opt} резервной ЭВМ в зависимости от отношения D_1/D_2, T и параметров λ, μ_1, μ_2, полученных экспериментально.

Выводы

1. Предложена математическая модель процесса обработки информации в виде системы массового обслуживания с резервным прибором с управлением, зависящем от времени ожидания.
2. Проведена оптимизация системы при учете потерь на ожидание и амортизацию.
3. Результаты исследования могут быть использованы при оптимизации систем, где важна не длина очереди, а время ожидания заявок в очереди.

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. Коваленко И.Н. О СМО со скоростью обслуживания, зависящей от числа требований в системе, и периодическим отключением каналов // Проблемы передачи информации. – 1971. – Вып. 7. – № 2. – С. 106–111.
3. Поттосина С.А. Однолинейная система массового обслуживания с переменной интенсивностью обслуживания, зависящей от времени ожидания, функционирующая в случайной среде // В кн.: Управляемые системы массового обслуживания. – Томск: Изд-во ТГУ, 1984. – С. 100–105.
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. Сер. Системы связи. – 1990. – Вып. 7. – С. 3–8.
6. 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.
7. Зиновьева Л.И. Система массового обслуживания с гистерезисом и резервным прибором, управляемым временем ожидания // В кн.: Математическая статистика и ее приложения. – Томск: Изд-во ТГУ, 1980. – № 6. – С. 152–164.
8. Самочернова Л.И. Оптимизация системы массового обслуживания с переменной интенсивностью зависящей от времени ожидания // Известия Томского политехнического университета. – 2009. – Т. 315. – № 5. – С. 178–182.
9. Исследование двух однолинейных СМО с интенсивностью обслуживания, зависящей от времени ожидания / Самочернова Л.И.; Том. политехн. ун-т. – Томск, 2009. – 9 с. – Библиогр.: 7 назв. – Рус. – Деп. в ВИНИТИ 29.10.2009, № 659 – В 2009.
10. Переходной режим работы двухуровневой СМО. Дни науки: Сб. матер. научно-практ. конф. преподавателей и студентов. Вып. 8. Ч. 2 / Отв. ред. А.А. Маслак. – Славянск на Кубани: Издат. центр СГПИ, 2009. – С. 62–68.

Поступила 21.01.2010 г.