ИНФОРМАТИКА

2012 **январь-март** **№ 1**

МАТЕМАТИЧЕСКОЕ И ЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 519.872

О.С. Дудина, С.А. Дудин

МНОГОКАНАЛЬНАЯ СИСТЕМА ОБСЛУЖИВАНИЯ КАК МОДЕЛЬ CALL-ЦЕНТРА С УСЛУГОЙ ПЕРЕЗВОНА АБОНЕНТУ

Рассматривается многолинейная система массового обслуживания с бесконечным буфером и нетерпеливыми запросами как модель call-центра с возможностью предоставления услуги перезвона абоненту в момент поступления вызова в случае занятости всех операторов. В систему поступает стационарный пуассоновский поток запросов. Время обслуживания запросов каждым прибором имеет показательное распределение. Во время ожидания запросы могут проявлять нетерпеливость и покидать систему. Получено условие существования стационарного распределения состояний системы. Приведен алгоритм для нахождения стационарного распределения, получены формулы для основных характеристик производительности системы.

Введение

Call-центр (центр обслуживания звонков) является неотъемлемой частью компаний, деятельность которых напрямую связана с постоянным контактом со своими клиентами по телефону. В условиях возрастающей конкуренции среди компаний качество обслуживания клиентов становится все более важным, поэтому от эффективной работы и производительности call-центра напрямую зависит не только имидж, но и прибыль компаний. На сегодняшний день кроме быстрого и квалифицированного ответа перед компаниями наиболее остро встает проблема качественного обслуживания большого количества вызовов от абонентов, которая может быть успешно решена путем использования call-центром сервисной технологии перезвона абоненту (услуга call-back). Клиент, который не хочет ждать на линии, может просто оставить номер своего телефона, и оператор свяжется с ним позже. Абоненты могут находиться в очереди, фактически не ожидая соединения с оператором. Эта услуга позволяет сохранить поступающие вызовы, избежать недовольства и раздражения клиентов, сделать загрузку операторов более равномерной и увеличить эффективность их работы. Как показывает практика, в call-центрах, не предоставляющих услугу перезвона абоненту, часть времени, которое затрачивает оператор на обслуживание клиента, уходит на выслушивание жалоб о длительном ожидании начала обслуживания, в то время как в call-центрах с услугой call-back клиенты реже жалуются, так как им предоставляется возможность выбора – ожидать ответ оператора или оставить номер для связи с ним. Таким образом, использование данной услуги позволяет также существенно снизить среднее время обслуживания клиента.

Для моделирования, описания функционирования и оптимизации современных call-центров широко применяются методы теории массового обслуживания. Адекватное математическое моделирование call-центров приводит к существенному увеличению их экономической эффективности, поскольку точное прогнозирование на этапе проектирования позволяет существенно снизить расходы на их содержание и обслуживание. Рассматривая call-центр как математическую модель, можно решить широкий круг задач по оптимизации его функционирования и улучшению качества обслуживания абонентов. Обширный обзор литературы по call-центрам приведен в работе [1].

В настоящей статье исследуется многолинейная система обслуживания с нетерпеливыми запросами и бесконечным буфером, которая может использоваться для моделирования и оптимизации функционирования call-центра с услугой перезвона абоненту. Предполагается, что в случае занятости всех операторов в момент поступления звонка абоненту предлагается подождать (при этом ему могут сообщить его номер в очереди и ориентировочное время ожидания – очередь в системе предполагается «видимой») или оставить свой номер, чтобы оператор перезвонил ему позже. В зависимости от номера в очереди абонент принимает решение, ожидать в очереди, получить обратный вызов (call-back) или уйти из системы без обслуживания (balk). Как показывает статистика, клиенты, получившие информацию о своем месте в очереди или времени ожидания, ожидают связи с оператором в 1,5–2 раза дольше, чем клиенты, которым эту информацию не сообщили, в результате чего число необслуженных клиентов значительно уменьшается, поэтому рассмотрение «видимой» очереди является важным моментом при моделировании современного call-центра.

В рассматриваемой моделе предполагается, что во время ожидания абонент может разрывать соединение из-за нетерпеливости, предусмотрены также потери абонентов в случае недозвона им оператора (абонент занят или не отвечает).

В работах [2, 3] представлен асимптотический анализ более простой модели call-центра с услугой перезвона абоненту в режиме большой загрузки при большом числе операторов. В настоящей статье приведен точный аналитический анализ рассматриваемой модели при отсутствии ограничений на число операторов.

1. Математическая модель

Рассматривается система обслуживания с бесконечным буфером, состоящая из N приборов (операторов), как модель call-центра с возможностью предоставления услуги перезвона абоненту в случае занятости всех операторов в момент поступления вызова.

Процесс поступления вызовов (запросов) описывается стационарным пуассоновским потоком с интенсивностью λ.

Если в момент поступления запроса есть свободный прибор, его обслуживание начинается немедленно. В противном случае абоненту предлагается ожидать обслуживания либо, в случае нежелания ждать на линии, он может оставить номер своего телефона, чтобы оператор перезвонил ему позже.

Таким образом, в математической модели все запросы будем разделять на реальные и виртуальные. Под реальным запросом будем понимать запрос абонента, который успешно дозвонился оператору без ожидания или который, в случае занятости всех операторов, занимает линию для ожидания в call-центре. Под виртуальным запросом будем понимать запрос абонента, который оставил свой телефон для того, чтобы оператор перезвонил ему позже.

Предполагается, что число абонентов, ожидающих обслуживание и занимающих линии для ожидания, не может превышать некоторый порог R.

Если в момент прихода запроса все приборы заняты и r, $r=\overline{0,R-1}$, реальных запросов находится в буфере, то этот запрос с вероятностью q покидает систему (теряется), а с дополнительной вероятностью запрос становится в буфер и ожидает обслуживания. При этом с вероятностью $p_r, r=\overline{0,R-1}$, он становится виртуальным, а с дополнительной вероятностью – реальным.

Если в момент прихода запроса все приборы и места для ожидания заняты (число реальных запросов в буфере равно R), то запрос с вероятностью δ покидает систему (теряется), а с дополнительной вероятностью становится виртуальным.

Время обслуживания запроса каждым прибором имеет показательное распределение с параметром μ, $0<\mu<\infty$.

Если в момент освобождения прибора в очереди нет реальных запросов, система начинает дозваниваться виртуальному запросу, стоящему первым. На время дозвона освободившийся прибор блокируется. Считаем, что время дозвона имеет показательное распределение с параметром $\tilde{\mu}$, $0<\tilde{\mu}<\infty$. Предполагается, что с вероятностью h система не дозванивается абоненту (абонент занят или не берет трубку), в этом случае виртуальный запрос теряется, а заблокированный прибор освобождается. С вероятностью $1-h$ система дозванивается до ожидающего обслуживания абонента, и прибор начинает обслуживание виртуального запроса.

Реальный запрос может проявлять нетерпеливость и покидать систему через экспоненциально распределенное с параметром $\alpha, 0<\alpha<\infty$, время с момента попадания в буфер, если он не попал на обслуживание.

2. Процесс изменения состояний системы

Процесс изменения состояний исследуемой системы описывается трехмерной цепью Маркова

$$\xi_t = \{i_t, r_t, n_t\},\ t \ge 0,$$

где i_t – число запросов в системе, $i_t \ge 0$;

r_t – число реальных запросов в буфере, $r_t = \overline{0, \max\{0, \min\{i_t - N, R\}\}}$;

n_t – число запросов на обслуживании, $n_t = \overline{0, \min\{i_t, N\}}$,

в момент времени $t, t \ge 0$.

Теорема 1. *Инфинитезимальный генератор* $Q = (Q_{i,j})_{i,j\ge 0}$ *цепи Маркова* ξ_t, $t \ge 0$, *имеет блочно-трехдиагональную структуру, его ненулевые блоки заданы следующим образом:*

$$Q_{0,0} = -\lambda;$$

$$Q_{i,i} = -\mu C_{i+1} - \tilde{\mu}\tilde{C}_{i+1} + (1-h)\tilde{\mu}\tilde{C}_{i+1}\hat{E}_{i+1} - \lambda I_{i+1},\ 1 \le i < N;$$

$$Q_{N,N} = -\mu C_{N+1} - \tilde{\mu}\tilde{C}_{N+1} + (1-h)\tilde{\mu}\tilde{C}_{N+1}\hat{E}_{N+1} - (1-q)\lambda I_{N+1};$$

$$Q_{i,i} = I_{i-N+1} \otimes (-\mu C_{N+1} - \tilde{\mu}\tilde{C}_{N+1} + (1-h)\tilde{\mu}\tilde{C}_{N+1}\hat{E}_{N+1}) - (1-q)\lambda I_{(i-N+1)(N+1)} - \alpha C_{i-N+1} \otimes I_{N+1},\ N < i < N+R;$$

$$Q_{i,i} = Q_1 = I_{R+1} \otimes (-\mu C_{N+1} - \tilde{\mu}\tilde{C}_{N+1} + (1-h)\tilde{\mu}\tilde{C}_{N+1}\hat{E}_{N+1}) - \lambda I_{(R+1)(N+1)} + \lambda\Delta \otimes I_{N+1} - \alpha C_{R+1} \otimes I_{N+1},\ i \ge N+R;$$

$$Q_{i,i-1} = \mu C_{i+1}E^-_{i+1} + h\tilde{\mu}\tilde{C}_{i+1}\tilde{E}^-_{i+1},\ 1 \le i \le N;$$

$$Q_{i,i-1} = \overline{I}_{i-N+1} \otimes (\mu C_{N+1}E_{N+1} + h\tilde{\mu}\tilde{C}_{N+1}) + (E^-_{i-N+1} - \overline{I}_{i-N+1}) \otimes (\mu C_{N+1} + h\tilde{\mu}\tilde{C}_{N+1}\hat{E}_{N+1}) +$$
$$+\alpha C_{i-N+1}E^-_{i-N+1} \otimes I_{N+1},\ N < i \le N+R;$$

$$Q_{i,i-1} = Q_0 = \hat{I} \otimes (\mu C_{N+1}E_{N+1} + h\tilde{\mu}\tilde{C}_{N+1}) + E_{R+1} \otimes (\mu C_{N+1} + h\tilde{\mu}\tilde{C}_{N+1}\hat{E}_{N+1}) + \alpha C_{R+1}E_{R+1} \otimes I_{N+1}, i > N+R;$$

$$Q_{i,i+1} = \lambda E^+_{i+1},\ 0 \le i \le N-1;$$

$$Q_{i,i+1} = (1-q)\lambda(I_{i-N+1} - \overline{P}_{i-N+1})E^+_{i-N+1} \otimes I_{N+1} + (1-q)\lambda\overline{P}_{i-N+1}\tilde{E}^+_{i-N+1} \otimes I_{N+1},\ N \le i < N+R;$$

$$Q_{i,i+1} = Q_2 = \lambda(\overline{\Delta}\hat{E}_{R+1} + \tilde{\Delta}) \otimes I_{N+1},\ i \ge N+R,$$

где I – *единичная матрица, в случае необходимости порядок матрицы указывается при помощи нижнего индекса*;

$\otimes$ – *операция Кронекерова произведения матриц*;

$C_l = \mathrm{diag}\{0, 1, \ldots, l-1\}$, $\tilde{C}_l = \mathrm{diag}\{l-1, l-2, \ldots, 0\}$, $l = \overline{2, \max\{N, R\}+1}$;

$\overline{P}_l = \mathrm{diag}\{p_0, p_1, \ldots, p_{l-1}\}$, $l = \overline{1, R}$;

$\Delta = \mathrm{diag}\{q, q, \ldots, q, \delta\}$;

$\overline{\Delta} = (1-q)\mathrm{diag}\{1-p_0, 1-p_1, \ldots, 1-p_{R-1}, 0\}$;

$\tilde{\Delta} = \mathrm{diag}\{(1-q)p_0, (1-q)p_1, \ldots, (1-q)p_{R-1}, 1-\delta\}$;

$E^-_l, \tilde{E}^-_l, l = \overline{2, \max\{N, R\}+1}$, – *матрицы размера* $l \times (l-1)$ *со всеми элементами, кроме элементов* $(E^-_l)_{0,0}, (E^-_l)_{i,i-1}$, $i = \overline{1, l-1}$, $(\tilde{E}^-_l)_{i,i}, i = \overline{0, l-2}$, *равных единице*;

$E^+_l, \tilde{E}^+_l, l = \overline{1, \max\{N, R\}+1}$, – *матрицы размера* $l \times (l+1)$ *со всеми нулевыми элементами, кроме элементов* $(E^+_l)_{i,i+1}, i = \overline{0, l-1}$, $(\tilde{E}^+_l)_{i,i}$, $i = \overline{0, l-1}$, *равных единице*;

$\hat{E}_l, l = \overline{1, \max\{N, R\} + 1},$ *– квадратные матрицы размера* l *со всеми нулевыми элементами, кроме элементов* $(\hat{E}_l)_{i,i+1}, i = \overline{0, l-1},$ *равных единице*;

$E_l, l = \overline{1, \max\{N, R\} + 1},$ *– квадратные матрицы размера* l *со всеми нулевыми элементами, кроме элементов* $(E)_{i,i-1}, i = \overline{1, l-1},$ *равных единице*;

$\bar{I}_l$ *– матрица размера* $l \times (l-1)$ *со всеми нулевыми элементами, кроме элемента* $(\bar{I}_l)_{0,0}$, *равного единице*;

$\hat{I}$ *– квадратная матрица размера* $R+1$ *со всеми нулевыми элементами, кроме элемента* $(\hat{I})_{0,0}$, *равного единице.*

Доказательство теоремы основывается на анализе переходов цепи Маркова $\xi_t, t \geq 0,$ за бесконечно малый интервал времени с последующей группировкой интенсивностей соответствующих переходов в блоки матрицы Q.

Рассматриваемая цепь Маркова $\xi_t, t \geq 0,$ является квазипроцессом гибели и размножения [4]. Согласно [4] условие эргодичности для цепей такого рода заключается в выполнении неравенства

$$\mathbf{y}Q_0\mathbf{e} > \mathbf{y}Q_2\mathbf{e}, \tag{1}$$

где вектор $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_R)$ является единственным решением следующей системы линейных алгебраических уравнений:

$$\mathbf{y}(Q_0 + Q_1 + Q_2) = \mathbf{0}, \mathbf{ye} = 1. \tag{2}$$

Здесь и далее $\mathbf{e}$ – вектор-столбец, состоящий из единиц, $\mathbf{0}$ – вектор-строка, состоящий из нулей.

Если размерность системы (2) невелика, она может быть решена стандартными методами на компьютере. В противном случае, принимая во внимание, что матрица $\bar{Q} = Q_0 + Q_1 + Q_2$ имеет блочную трехдиагональную структуру, для решения системы может быть использован численно устойчивый алгоритм, который состоит из следующих шагов:

Шаг 1. Вычисляем ненулевые блоки матрицы $\bar{Q}$:

$$\bar{Q}_{r,r} = -\lambda I_{N+1} + \lambda((1-\delta_{r,R})((1-q)p_r + q) + \delta_{r,R})I_{N+1} + \delta_{r,0}(\mu C_{N+1}E_{N+1} + h\tilde{\mu}\tilde{C}_{N+1}) -$$
$$-\mu C_{N+1} - \tilde{\mu}\tilde{C}_{N+1} + (1-h)\tilde{\mu}\tilde{C}_{N+1}\hat{E}_{N+1} - r\alpha I_{N+1}, r = \overline{0, R};$$
$$\bar{Q}_{r,r+1} = (1-q)(1-p_r)\lambda I_{N+1}, r = \overline{0, R-1};$$
$$\bar{Q}_{r,r-1} = \mu C_{N+1} + h\tilde{\mu}\tilde{C}_{N+1}\hat{E}_{N+1} + r\alpha I_{N+1}, r = \overline{1, R},$$

где $\delta_{i,j}$ – символ Кронекера, равный единице, если $i = j,$ и нулю в противном случае.

Шаг 2. Вычисляем матрицы $T_r, r = \overline{0, R-1},$ с помощью обратной рекурсии

$$T_r = -\bar{Q}_{r,r+1}(\bar{Q}_{r+1,r+1} + T_{r+1}\bar{Q}_{r+2,r+1})^{-1}, r = R-2, R-3, \ldots, 0,$$

при начальном условии

$$T_{R-1} = -\bar{Q}_{R-1,R}(\bar{Q}_{R,R})^{-1}.$$

Шаг 3. Вычисляем матрицы $F_r, r = \overline{0, R},$ по рекуррентным формулам

$$F_0 = I, F_r = F_{r-1}T_{r-1}, r = \overline{1, R}.$$

Шаг 4. Находим вектор $\mathbf{y}_0$ как единственное решение системы линейных алгебраических уравнений

$$\mathbf{y}_0(\bar{Q}_{0,0}+T_0\bar{Q}_{1,0})=\mathbf{0},\quad \mathbf{y}_0\sum_{r=0}^{R}F_r\mathbf{e}=1.$$

Шаг 5. Вычисляем векторы $\mathbf{y}_r=\mathbf{y}_{r-1}T_{r-1}=\mathbf{y}_0F_r, r=\overline{1,R}.$

Если условие эргодичности (1) цепи Маркова ξ_t выполнено, то существуют стационарные вероятности состояний системы, которые определяются следующим образом:

$$\pi(i,r,n)=\lim_{t\to\infty}P\{i_t=i,r_t=r,n_t=n\}, i\ge 0, r=\overline{0,\max\{0,\min\{i-N,R\}\}}, n=\overline{0,\min\{i,N\}}.$$

Сформируем векторы $\boldsymbol{\pi}_i, i\ge 0$, состоящие из вероятностей $\pi(i,r,n)$, перенумерованных в лексикографическом порядке компонент r,n:

$$\boldsymbol{\pi}(i,r)=(\pi(i,r,0),\pi(i,r,1),\ldots,\pi(i,r,\min\{i,N\})), r=\overline{0,\max\{0,\min\{i-N,R\}\}},$$

$$\boldsymbol{\pi}_i=(\boldsymbol{\pi}(i,0),\boldsymbol{\pi}(i,1),\ldots,\boldsymbol{\pi}(i,\max\{0,\min\{i-N,R\}\})), i\ge 0.$$

Известно, что векторы $\boldsymbol{\pi}_i, i\ge 0,$ удовлетворяют следующей системе линейных алгебраических уравнений:

$$(\boldsymbol{\pi}_0,\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_i,\ldots)Q=\mathbf{0},\ (\boldsymbol{\pi}_0,\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_i,\ldots)\mathbf{e}=1, \tag{3}$$

где Q – генератор цепи Маркова $\xi_t, t\ge 0$, определенный в теореме 1.

Для решения системы (3) может быть использован численно устойчивый алгоритм, который включает следующие шаги:

Шаг 1. Находим матрицу G как минимальное неотрицательное решение матричного уравнения

$$Q_2G^2+Q_1G+Q_0=O.$$

Шаг 2. Вычисляем матрицы G_i с помощью обратной рекурсии

$$G_i=-(Q_{i+1,i+1}+Q_{i+1,i+2}G_{i+1})^{-1}Q_{i+1,i},\ i=N+R-1,N+R-2,\ldots,0,$$

положив $G_{N+R}=G.$

Шаг 3. Вычисляем матрицы Φ_i с помощью рекуррентных соотношений

$$\Phi_0=I, \Phi_i=-\Phi_{i-1}Q_{i-1,i}(Q_{i,i}+Q_{i,i+1}G_i)^{-1},\ i=\overline{1,N+R-1};$$

$$\Phi_{N+R}=-\Phi_{N+R-1}Q_{N+R-1,N+R}(Q_1+Q_2G)^{-1};$$

$$\Phi_i=-\Phi_{i-1}Q_0(Q_1+Q_2G)^{-1}, i>N+R.$$

Шаг 4. Находим вектор $\boldsymbol{\pi}_0$ как единственное решение следующей системы линейных алгебраических уравнений:

$$\boldsymbol{\pi}_0(Q_{0,0}+Q_{0,1}G_0)=\mathbf{0},\boldsymbol{\pi}_0\sum_{i=0}^{\infty}\Phi_i\mathbf{e}=1.$$

Шаг 5. Вычисляем векторы $\boldsymbol{\pi}_i$ по формуле $\boldsymbol{\pi}_i=\boldsymbol{\pi}_0\Phi_i,\ i\ge 1.$

Алгоритм получен с использованием результатов, приведенных в статье [5], с учетом специальной структуры генератора Q.

Численная устойчивость приведенных в данной статье алгоритмов следует из того, что все обращаемые матрицы являются неприводимыми субгенераторами. Поэтому, как известно из теории матриц, обратные матрицы существуют и они неотрицательны.

3. Характеристики производительности системы

Вычислив векторы $\boldsymbol{\pi}_i, i \geq 0,$ можно найти различные характеристики производительности рассматриваемой системы. Соответствующие формулы приведены ниже.

Среднее число запросов в системе

$$L = \sum_{i=1}^{\infty} i\boldsymbol{\pi}_i \mathbf{e}.$$

Среднее число запросов (реальных и виртуальных), ожидающих обслуживание, вычисляется по формуле

$$N^{buffer} = \sum_{i=N+1}^{\infty} (i-N)\boldsymbol{\pi}_i \mathbf{e}.$$

Среднее число реальных запросов, ожидающих обслуживание, определяется как

$$N^{buffer}_{real} = \sum_{i=N+1}^{\infty} \sum_{r=1}^{\min\{i-N,R\}} r\boldsymbol{\pi}(i,r)\mathbf{e}.$$

Среднее число виртуальных запросов, ожидающих обслуживание, определяется как

$$N^{buffer}_{virt} = \sum_{i=N+1}^{\infty} \sum_{r=0}^{\min\{i-N,R\}} (i-N-r)\boldsymbol{\pi}(i,r)\mathbf{e}.$$

Среднее число приборов, занятых обслуживанием и заблокированных, равно

$$N^{server} = \sum_{i=1}^{\infty} \min\{i,N\}\boldsymbol{\pi}_i \mathbf{e}.$$

Среднее число приборов, занятых обслуживанием, определяется как

$$N^{server}_{service} = \sum_{i=1}^{\infty} \sum_{r=0}^{\max\{0,\min\{i-N,R\}\}} \sum_{n=1}^{\min\{i,N\}} n\pi(i,r,n).$$

Среднее число заблокированных приборов вычисляется по формуле

$$N^{server}_{block} = \sum_{i=1}^{\infty} \sum_{r=0}^{\max\{0,\min\{i-N,R\}\}} \sum_{n=0}^{\min\{i,N-1\}} (\min\{i,N\}-n)\pi(i,r,n) = N^{server} - N^{server}_{service}.$$

Вероятность того, что произвольный реальный запрос поступит в систему, когда все приборы заняты, число реальных запросов в буфере меньше R, и уйдет из системы на входе из-за нежелания ждать, вычисляется по формуле

$$P^{esc-loss}_{real} = q\sum_{i=N}^{\infty} \sum_{r=0}^{\min\{i-N,R-1\}} \boldsymbol{\pi}(i,r)\mathbf{e}.$$

Вероятность потери произвольного реального запроса на входе из-за наличия R реальных запросов в буфере вычисляется по формуле

$$P_{real}^{ent-loss} = \sum_{i=N+R}^{\infty} \pi(i,R)\mathbf{e}.$$

Вероятность того, что произвольный запрос поступит в систему, когда все приборы заняты, число реальных запросов в буфере меньше R, и станет виртуальным, вычисляется по формуле

$$P_{real}^{to-virt} = (1-q)\sum_{i=N}^{\infty}\sum_{r=0}^{\min\{i-N,R-1\}} p_r \pi(i,r)\mathbf{e}.$$

Вероятность того, что произвольный реальный запрос поступит в систему, когда все приборы заняты, число реальных запросов в буфере меньше R, станет в буфер и уйдет из системы из-за нетерпеливости, определяется формулой

$$P_{real}^{imp-loss} = (1-q)\sum_{i=N}^{\infty}\sum_{r=0}^{\min\{i-N,R-1\}}(1-p_r)\sum_{n=0}^{N}\pi(i,r,n)(1-z(r+1,n)),$$

где $z(r,n)$ – вероятность того, что за время ожидания реального запроса в буфере этот запрос не покинет систему из-за нетерпеливости при условии, что в момент его прихода n запросов находится на обслуживании и позиция этого запроса в буфере равна r, $r=\overline{1,R}$, $n=\overline{0,N}$.

Вероятности $z(r,n)$ определяются формулами

$$z(1,N) = (\alpha + N\mu)^{-1} N\mu;$$

$$z(1,n) = (\alpha + n\mu + (N-n)\tilde{\mu})^{-1}[n\mu + h(N-n)\tilde{\mu} + (1-h)(N-n)\tilde{\mu}z(1,n+1)], n = N-1, N-2, \ldots, 0;$$

$$z(r,N) = (r\alpha + N\mu)^{-1}[N\mu + (r-1)\alpha]z(r-1,N),\ r=\overline{2,R};$$

$$z(r,n) = (r\alpha + n\mu + (N-n)\tilde{\mu})^{-1}[(n\mu + (r-1)\alpha)z(r-1,n) + h(N-n)\tilde{\mu}z(r-1,n+1) +$$
$$+(1-h)(N-n)\tilde{\mu}z(r,n+1)],\ n = N-1, N-2, \ldots, 0,\ r=\overline{2,R}.$$

Вероятность потери произвольного реального запроса определяется как

$$P_{real}^{loss} = P_{real}^{esc-loss} + P_{real}^{ent-loss} + P_{real}^{to-virt} + P_{real}^{imp-loss}.$$

Интенсивность выходящего потока обслуженных запросов определяется формулой

$$\lambda^{out} = \mu N_{service}^{server}.$$

Вероятность потери произвольного запроса вычисляется как

$$P^{loss} = 1 - \frac{\lambda^{out}}{\lambda}.$$

4. Распределение времен пребывания и ожидания реального запроса

Пусть $V(x)$ – функция распределения времени пребывания произвольного реального запроса в системе и $v(s) = \int_0^{\infty} e^{-sx} dV(x)$, $\operatorname{Re} s > 0$, – ее преобразование Лапласа – Стилтьеса.

Теорема 2. *Преобразование Лапласа – Стилтьеса* $v(s)$ *распределения времени пребывания произвольного реального запроса в системе имеет вид*

$$v(s)=P_{real}^{ent-loss}+P_{real}^{esc-loss}+P_{real}^{to-virt}+\sum_{i=0}^{N-1}\pi_i\mathbf{e}\frac{\mu}{s+\mu}+(1-q)\sum_{i=N}^{\infty}\sum_{r=0}^{\min\{i-N,R-1\}}(1-p_r)\sum_{n=0}^{N}\pi(i,r,n)v(s,r+1,n), \quad (4)$$

где $v(s,r,n), r=\overline{1,R}, n=\overline{0,N},$ *вычисляются следующим образом:*

$$v(s,1,N)=(s+\alpha+N\mu)^{-1}(\alpha+N\mu\frac{\mu}{s+\mu}); \quad (5)$$

$$v(s,1,n)=(s+\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[\alpha+(1-h)(N-n)\tilde{\mu}v(s,1,n+1)+ \\ +(n\mu+h(N-n)\tilde{\mu})\frac{\mu}{s+\mu}], n=N-1,N-2,\ldots,0; \quad (6)$$

$$v(s,r,N)=(s+r\alpha+N\mu)^{-1}[\alpha+(N\mu+(r-1)\alpha)v(s,r-1,N)], r=\overline{2,R}; \quad (7)$$

$$v(s,r,n)=(s+r\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[\alpha+(1-h)(N-n)\tilde{\mu}v(s,r,n+1)+ \\ +(n\mu+(r-1)\alpha)v(s,r-1,n)+h(N-n)\tilde{\mu}v(s,r-1,n+1)], n=N-1,N-2,\ldots,0, r=\overline{2,R}. \quad (8)$$

Доказательство. Пометим произвольный реальный запрос и будем отслеживать процесс его обслуживания в системе. Учитывая вероятностный смысл преобразования Лапласа – Стилтьеса, $v(s)$ можно интерпретировать как вероятность того, что за время пребывания реального запроса в системе не наступит катастрофа из стационарного пуассоновского потока катастроф с интенсивностью s [6, 7].

В момент прихода реального запроса в систему могут произойти следующие события:

1. Запрос уходит из системы сразу же после прибытия. Вероятность этого события можно вычислить по формуле $P_{real}^{ent-loss}+P_{real}^{esc-loss}$. В этом случае вероятность ненаступления катастрофы равна единице.

2. Запрос поступит в систему, когда все приборы заняты, и станет виртуальным. Вероятность этого события равна $P_{real}^{to-virt}$. В этом случае вероятность ненаступления катастрофы также равна единице.

3. В момент прихода реального запроса в систему есть свободный прибор, и запрос сразу же идет на обслуживание. Вероятность этого события равна $\sum_{i=0}^{N-1}\pi_i\mathbf{e}$. В этом случае вероятность ненаступления катастрофы за время пребывания равна вероятности ненаступления катастрофы за время обслуживания и определяется как $\frac{\mu}{s+\mu}$.

4. В момент прихода запроса в систему все приборы заняты, и он становится в очередь.

Пусть $v(s,r,n)$ – преобразование Лапласа – Стилтьеса условного распределения времени пребывания помеченного запроса при условии, что в данный момент позиция помеченного запроса в очереди равна $r, r=\overline{1,R},$ и число запросов на обслуживании равно $n, n=\overline{0,N}$.

В случае нахождения преобразования Лапласа – Стилтьеса $v(s,r,n)$ вероятность того, что запрос будет принят в систему в момент, когда все приборы заняты, и за время его пребывания не произойдет катастрофа, запишется в виде

$$(1-q)\sum_{i=N}^{\infty}\sum_{r=0}^{\min\{i-N,R-1\}}(1-p_r)\sum_{n=0}^{N}\pi(i,r,n)v(s,r+1,n).$$

Учитывая вероятностный смысл преобразования Лапласа – Стилтьеса, можно получить следующую систему уравнений для $v(s,r,n),\ r=\overline{1,R},\ n=\overline{0,N}$:

$$(s+r\alpha+n\mu+(N-n)\tilde{\mu})v(s,r,n)=(1-h)(N-n)\tilde{\mu}v(s,r,n+1)+$$
$$+(1-\delta_{r,1})[((r-1)\alpha+n\mu)v(s,r-1,n)+h(N-n)\tilde{\mu}v(s,r-1,n+1)]+\delta_{r,1}(n\mu+h(N-n)\tilde{\mu})\frac{\mu}{s+\mu}+\alpha. \quad (9)$$

Легко доказать, что решение системы (9) можно записать в виде (5)–(8). Используя формулу полной вероятности, легко убедиться в справедливости утверждения теоремы. □

Вычислив преобразование Лапласа – Стилтьеса $v(s)$, можно найти распределение времени пребывания произвольного реального запроса, используя численные методы [8].

На практике важно знать среднее время пребывания произвольного реального запроса в системе, которое можно найти с помощью преобразования Лапласа – Стилтьеса $v(s)$.

Следствие 1. *Среднее время пребывания произвольного реального запроса в системе вычисляется по формуле*

$$V_{real}^{soj}=\mu^{-1}\sum_{i=0}^{N-1}\boldsymbol{\pi}_i\mathbf{e}-(1-q)\sum_{i=N}^{\infty}\sum_{r=0}^{\min\{i-N,R-1\}}(1-p_r)\sum_{n=0}^{N}\pi(i,r,n)v'(0,r+1,n), \quad (10)$$

где
$$v'(0,1,N)=-(\alpha+N\mu)^{-1}[1+N];$$
$$v'(0,1,n)=-(\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[1-(1-h)(N-n)\tilde{\mu}v'(0,1,n+1)+$$
$$+(n\mu+h(N-n)\tilde{\mu})\mu^{-1}],\ n=N-1,N-2,\ldots,0;$$
$$v'(0,r,N)=-(r\alpha+N\mu)^{-1}[1-(N\mu+(r-1)\alpha)v'(0,r-1,N)],\ r=\overline{2,R};$$
$$v'(0,r,n)=-(r\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[1-(1-h)(N-n)\tilde{\mu}v'(0,r,n+1)-$$
$$-(n\mu+(r-1)\alpha)v'(0,r-1,n)-h(N-n)\tilde{\mu}v'(0,r-1,n+1)],\ n=N-1,N-2,\ldots,0,\ r=\overline{2,R}.$$

Доказательство. Формула (10) для нахождения среднего времени пребывания произвольного реального запроса в системе следует из соотношения $V_{real}^{soj}=-v'(s)|_{s=0}$ с учетом того, что $v(s,r,n)|_{s=0}=1,\ r=\overline{1,R},\ n=\overline{0,N}$. Следствие доказано. □

На основании теоремы 2 преобразование Лапласа – Стилтьеса распределения времени ожидания произвольного реального запроса может быть найдено по формулам (4)–(8), если предварительно удалить из формул (4)–(6) множитель $\frac{\mu}{s+\mu}$.

Следствие 2. *Среднее время ожидания произвольного реального запроса в системе вычисляется по формуле*

$$V_{real}^{wait}=-(1-q)\sum_{i=N}^{\infty}\sum_{r=0}^{\min\{i-N,R-1\}}(1-p_r)\sum_{n=0}^{N}\pi(i,r,n)y'(0,r+1,n),$$

где
$$y'(0,1,N)=-(\alpha+N\mu)^{-1};$$
$$y'(0,1,n)=-(\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[1-(1-h)(N-n)\tilde{\mu}y'(0,1,n+1)],\ n=N-1,N-2,\ldots,0;$$
$$y'(0,r,N)=-(r\alpha+N\mu)^{-1}[1-(N\mu+(r-1)\alpha)y'(0,r-1,N)],\ r=\overline{2,R};$$
$$y'(0,r,n)=-(r\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[1-(1-h)(N-n)\tilde{\mu}y'(0,r,n+1)-$$
$$-(n\mu+(r-1)\alpha)y'(0,r-1,n)-h(N-n)\tilde{\mu}y'(0,r-1,n+1)],\ n=N-1,N-2,\ldots,0,\ r=\overline{2,R}.$$

Теорема 3. *Преобразование Лапласа-Стилтьеса* $w(s)$ *распределения времени пребывания в системе произвольного обслуженного реального запроса вычисляется следующим образом*

$$w(s)=\frac{1}{1-P_{real}^{loss}}[\sum_{i=0}^{N-1}\pi_i\mathbf{e}\frac{\mu}{s+\mu}+(1-q)\sum_{i=N}^{\infty}\sum_{r=0}^{\min\{i-N,R-1\}}(1-p_r)\sum_{n=0}^{N}\pi(i,r,n)w(s,r+1,n)], \tag{11}$$

где $w(s,r,n),\ r=\overline{1,R},\ n=\overline{0,N},$ *определяются формулами*

$$w(s,1,N)=(s+\alpha+N\mu)^{-1}N\mu\frac{\mu}{s+\mu}; \tag{12}$$

$$\begin{gathered}w(s,1,n)=(s+\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[(1-h)(N-n)\tilde{\mu}w(s,1,n+1)+\\+(n\mu+h(N-n)\tilde{\mu})\frac{\mu}{s+\mu}],\ n=N-1,N-2,\dots,0;\end{gathered} \tag{13}$$

$$w(s,r,N)=(s+r\alpha+N\mu)^{-1}(N\mu+(r-1)\alpha)w(s,r-1,N),\ r=\overline{2,R};$$

$$\begin{gathered}w(s,r,n)=(s+r\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[(1-h)(N-n)\tilde{\mu}w(s,r,n+1)+\\+(n\mu+(r-1)\alpha)w(s,r-1,n)+h(N-n)\tilde{\mu}w(s,r-1,n+1)],\ n=N-1,N-2,\dots,0,\ r=\overline{2,R}.\end{gathered}$$

Следствие 3. *Среднее время пребывания произвольного обслуженного реального запроса в системе вычисляется по формуле*

$$V_{real}^{soj-serv}=\frac{1}{1-P_{real}^{loss}}[\mu^{-1}\sum_{i=0}^{N-1}\pi_i\mathbf{e}-(1-q)\sum_{i=N}^{\infty}\sum_{r=0}^{\min\{i-N,R-1\}}(1-p_r)\sum_{n=0}^{N}\pi(i,r,n)w'(0,r+1,n)],$$

где
$$w'(0,1,N)=-(\alpha+N\mu)^{-1}((\alpha+N\mu)^{-1}N\mu+N);$$

$$\begin{gathered}w'(0,1,n)=-(\alpha+n\mu+(N-n)\tilde{\mu})^{-1}[(\alpha+n\mu+(N-n)\tilde{\mu})^{-1}((1-h)(N-n)\tilde{\mu}w(0,1,n+1)+\\+n\mu+h(N-n)\tilde{\mu})-(1-h)(N-n)\tilde{\mu}w'(0,1,n+1)+(n\mu+h(N-n)\tilde{\mu})\mu^{-1}],\ n=N-1,N-2,\dots,0;\end{gathered}$$

$$w'(0,r,N)=-(r\alpha+N\mu)^{-1}(N\mu+(r-1)\alpha)[(r\alpha+N\mu)^{-1}w(0,r-1,N)-w'(0,r-1,N)],\ r=\overline{2,R};$$

$$\begin{gathered}w'(0,r,n)=-(r\alpha+n\mu+(N-n)\tilde{\mu})^{-1}((r\alpha+n\mu+(N-n)\tilde{\mu})^{-1}\times\\\times[(1-h)(N-n)\tilde{\mu}w(0,r,n+1)+(n\mu+(r-1)\alpha)w(0,r-1,n)+\\+h(N-n)\tilde{\mu}w(0,r-1,n+1)]-(1-h)(N-n)\tilde{\mu}w'(0,r,n+1)-\\-(n\mu+(r-1)\alpha)w'(0,r-1,n)-h(N-n)\tilde{\mu}w'(0,r-1,n+1)),\ n=N-1,N-2,\dots,0,\ r=\overline{2,R}.\end{gathered}$$

Преобразование Лапласа – Стилтьеса распределения времени ожидания произвольного обслуженного реального запроса может быть найдено по формулам, приведенным в теореме 3, если предварительно удалить из формул (11)–(13) множитель $\frac{\mu}{s+\mu}$.

Следствие 4. *Среднее время ожидания произвольного обслуженного реального запроса в системе вычисляется по формуле*

$$V_{real}^{wait-serv}=V_{real}^{soj-serv}-\frac{1}{\mu}.$$

Заключение

В статье исследована многолинейная система массового обслуживания с бесконечным буфером и нетерпеливыми запросами. Построен процесс функционирования данной системы. Найдено условие существования стационарного распределения вероятностей состояний системы.

Приведен алгоритм для нахождения стационарного распределения, получены формулы для вычисления основных характеристик производительности. Полученные результаты могут применяться для оценивания производительности и оптимизации функционирования call-центров банков, экстренных служб, мобильных операторов, справочных служб.

Данная работа была частично поддержана Белорусским республиканским фондом фундаментальных исследований через грант Ф11М-003.

Список литературы

1. Aksin, O.Z. The modern call centers: a multi-disciplinary perspective on operations management research / O.Z. Aksin, M. Armony, V. Mehrotra // Production and Operation Management. – 2007. – Vol. 16. – P. 655–688.

2. Armony, M. On customer contact centers with a call-back option: customer decisions, routing rules, and system design / M. Armony, C. Maglaras // Operations research. – 2004. – Vol. 52, № 2. – P. 271–292.

3. Armony, M. Contact centers with a call-back option and real-time delay information / M. Armony, C. Maglaras // Operations research. – 2004. – Vol. 52, № 4. – P. 527–545.

4. Neuts, M. Matrix-geometric solutions in stochastic models – An algorithmic approach / M. Neuts. – USA : Johns Hopkins University Press, 1981. – 332 p.

5. Klimenok, V.I. Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory / V.I. Klimenok, A.N. Dudin // Queueing Systems. – 2006. – Vol. 54. – P. 245–259.

6. Kesten, H. Priority in waiting line problems / H. Kesten, J.Th. Runnenburg. – Amsterdam : Mathematisch Centrum, 1956. – 234 p.

7. Van Danzig, D. Chaines de Markof dans les ensembles abstraits et applications aux processus avec regions absorbantes et au probleme des boucles / D. van Danzig // Ann. de l'Inst. H. Pioncare. – 1955. – Vol. 14. – P. 145–199.

8. Abate, J. Numerical inversion of the Laplace transforms of probability distributons / J. Abate, W. Whitt // ORSA Journal on Computing. – 1955. – Vol. 7, № 1. – P. 36–43.

Поступила 06.01.12

Белорусский государственный университет,
Минск, пр. Независимости, 4
email: dudina_olga@email.com,
dudin85@mail.ru

O.S. Dudina, S.A. Dudin

MULTI-SERVER QUEUEING SYSTEM AS A MODEL OF CALL-CENTER WITH CALL-BACK OPTION

We consider multi-server queueing system with an infinite buffer and impatient customers as a model of call-center with call-back option when all servers are busy at arrival epoch. Arrival flow is described by the stationary Poisson process. Service times of customers by a server have an exponential distribution. The condition for the stable operation of the system has been obtained. An efficient algorithm for calculating the stationary probabilities of system states is proposed. The main performance measures are calculated.