

ИССЛЕДОВАНИЕ CALL-ЦЕНТРОВ КАК СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С ПОВТОРНЫМИ ВЫЗОВАМИ¹

Ключевые слова: *call-центр, СМО, повторные вызовы.*

Введение. Развитие инновационных технологий в сфере телекоммуникационных систем послужило причиной появления и активного внедрения почти во все сферы общественной и предпринимательской жизни таких систем, как call-центры (call centers). В теории и практике управления call-центры часто рассматриваются как синоним телефонного обслуживания клиентуры, как фактор успеха в стремлении фирмы добиться признания потребителей. Call-центры являются некоторым стандартом в области обслуживания и широко используются в телекоммуникационных и торговых фирмах, банках кредитно-карточной системы, отелях, социальных службах, службах экстренной помощи и т.д. Call-центры стали незаменимым средством связи и эффективного управления взаимоотношений с клиентами [1].

Большинство структур, в работе которых преобладает общение с людьми (а это могут быть как частные компании, так и государственные учреждения), реорганизовывают свою инфраструктуру, внедряя в нее от одного до нескольких call-центров. Такие системы предоставляют возможность наилучшим образом использовать имеющиеся в компании ресурсы (сотрудники, линии связи, оборудование, программное обеспечение) для обслуживания клиентских вызовов.

В последнее время call-центры стали называть контакт-центрами, поскольку в дополнение к телефонным средствам в них начали использовать возможности Интернета, электронной почты, факсимильной связи, чата, технологии баз данных и т.п. Эти факторы значительно влияют на функционирование call-центра и характер передаваемой информации, а также непосредственно на решения клиентов, поскольку последние могут чаще обращаться к call-центру и получать более разнообразную информацию. Call-центры — важное средство связи в современном мире, поэтому разработка адекватных математических моделей является одной из главных задач на этапе их проектирования.

В работах, где рассматриваются вопросы функционирования и администрирования call-центров, все исследования базируются на методах теории массового обслуживания [2]. Следует отметить, что наиболее адекватным методом для построения моделей call-центров является аппарат систем массового обслуживания (СМО) с повторениями [3]. Такими СМО можно учитывать первичный, вторичный и другие потоки вызовов, которые поступают на вход call-центра и существенно влияют на показатели его функционирования [4]. В большинстве публикаций по моделированию работы call-центров рассматриваются модели без учета потока повторных вызовов. В тех работах, где учитываются повторные вызовы, не рассматривается их зависимость от потока первичных вызовов.

Моделирование call-центров как систем с повторениями освещено в работах [5, 6], где показаны последствия игнорирования повторными вызовами при расчете оптимального количества операторов call-центра. В статье [5] call-центр моделируется как многоканальная СМО, в которой явно смоделированы следующие операции: клиентский отказ становится в очередь, беспокойство клиента и повторные вызовы. Полученная СМО проанализирована для стационарного и нестационарного режимов. Для стационарного режима применяется плавная аппроксимация (fluid approximation), которая облегчает анализ непрерывной во времени марковской цепи и которая используется для точного отображения систем больших call-цен-

¹ Работа выполнена при поддержке МОН Украины (НДР от 22.04.2009 г., договор № М/202-2009).

тров с большой нагрузкой. Методом плавной аппроксимации авторы исследуют явление повторных вызовов для реального call-центра. Модель применяется для оценки интенсивности поступления звонков на основании статистических данных, в которых невозможно отличить повторные звонки от первичных. Это общая проблема, с которой сталкиваются все call-центры. С помощью численных методов показано, что игнорирование явления повторных вызовов в call-центрах может привести к значительным погрешностям при последующем анализе прогнозирования и определении оптимального количества агентов.

В статье [6] call-центр моделируется как цепь Маркова с учетом состояния беспокойства клиентов и повторных вызовов. Показано, что игнорирование повторных вызовов может привести к недостаточному или, наоборот, избыточному (в зависимости от сделанного прогноза) укомплектованию персонала относительно оптимального.

Call-центр как СМО типа $M/M/c/0/L/H_j$. На практике количество абонентов call-центра представляет всегда конечное число, которое не превышает количества абонентов телефонной сети страны. Поэтому емкость орбиты также является конечным числом. Рассмотрим модель call-центра, которая имеет емкость орбиты, ограниченную заданной константой L . Если емкость орбиты равна L , то вызовы, которые поступают в систему, теряются и не влияют на функционирование системы.

Пусть вызов, который поступает в систему, после нескольких неудачных попыток уходит из системы. Пусть также H_j — вероятность того, что после j -й неудачной попытки произойдет еще $(j+1)$ -я попытка. Предположим, что вероятность повторных вызовов после неудачной повторной попытки не зависит от количества предыдущих попыток (т.е. $H_2 = H_3 = \dots$).

Пусть на c каналов обслуживания поступает пуассоновский поток первичных вызовов с интенсивностью λ . Если в момент прихода первичного вызова какой-либо из c каналов свободен, то вызов обслуживается и уходит из системы. В противном случае с вероятностью $1-H_1$ вызов уходит из системы без обслуживания и с вероятностью $H_1 > 0$ поступает на орбиту, если хотя бы одно из L мест орбиты свободно, или уходит из системы (вызов теряется), если все места на орбите заняты. Времена обслуживания распределены экспоненциально с параметром μ .

Вызовы на орбите представляют собой пуассоновский процесс с интенсивностью ν . Если в момент прихода повторного вызова какой-либо канал обслуживания свободен, то после обслуживания он уходит из системы и исчезает с орбиты. В противном случае с вероятностью $1-H_2$ вызов уйдет из системы или с вероятностью H_2 повторно сделает попытку обслужиться.

Построение аналитической модели. Функционирование системы $M/M/c/0/L/H_j$, как модели call-центра, можно описать с помощью двумерного процесса $(C(t), N(t))$, где $C(t)$ — число занятых каналов, $N(t)$ — число повторных вызовов на орбите в момент t . Процесс $(C(t), N(t))$ является марковским, определенным на множестве состояний $S^{(L)} = \{0, 1, \dots, c\} \times \{0, 1, \dots, L\}$. Его инфинитезимальные интенсивности $q_{(ij)(nm)}$ перехода из состояния (i, j) в состояние (n, m) задаются следующим образом:

при $0 \leq i \leq c-1, 0 \leq j \leq L$

$$q_{(ij)(nm)} = \begin{cases} \lambda, & \text{если } (n, m) = (i+1, j), \\ i\mu, & \text{если } (n, m) = (i-1, j), \\ j\nu & \text{если } (n, m) = (i+1, j-1), \\ -(\lambda + i\mu + j\nu), & \text{если } (n, m) = (i, j), \\ 0 & \text{в остальных случаях;} \end{cases}$$

при $i = c, 0 \leq j \leq L-1$

$$q_{(cj)(nm)} = \begin{cases} \lambda H_1, & \text{если } (n, m) = (c, j+1), \\ j\nu(1-H_2), & \text{если } (n, m) = (c, j-1), \\ c\mu, & \text{если } (n, m) = (c-1, j), \\ -(\lambda H_1 + j\nu(1-H_2) + c\mu), & \text{если } (n, m) = (c, j), \\ 0 & \text{в остальных случаях;} \end{cases}$$

при $i = c, j = L$

$$q_{(cL)(nm)} = \begin{cases} c\mu, & \text{если } (n, m) = (c-1, L), \\ L\nu(1-H_2), & \text{если } (n, m) = (c, L-1), \\ -(c\mu + L\nu(1-H_2)), & \text{если } (n, m) = (c, L), \\ 0 & \text{в остальных случаях.} \end{cases}$$

Поскольку множество состояний процесса $(C(t), N(t))$ конечно, то он всегда эргодичен [3]. Его стационарное распределение $p_{ij} = P(C(t) = i, N(t) = j)$ (p_{ij} — вероятность нахождения системы в состоянии (i, j)) может быть найдено как решение следующей системы уравнений:

$$(\lambda + i\mu + j\nu)p_{ij} = \lambda p_{i-1, j} + (j+1)\nu p_{i-1, j+1} + (i+1)\mu p_{i+1, j}, \quad 0 \leq i < c, \quad 0 \leq j < L; \quad (1)$$

$$(\lambda + i\mu + L\nu)p_{iL} = \lambda p_{i-1, L} + (i+1)\mu p_{i+1, L}, \quad 0 \leq i < c; \quad (2)$$

$$(\lambda H_1 + j\nu(1-H_2) + c\mu)p_{cj} = \lambda p_{c-1, j} + (j+1)\nu p_{c-1, j+1} + \lambda H_1 p_{c, j-1} + (j+1)\nu(1-H_2)p_{c, j+1}, \quad 0 \leq j < L, \quad 0 \leq j < L; \quad (3)$$

$$(c\mu + L\nu(1-H_2))p_{c, L} = \lambda p_{c-1, L} + \lambda H_1 p_{c, L-1}. \quad (4)$$

Эти уравнения удовлетворяют условию нормировки

$$\sum_{i=0}^c \sum_{j=0}^L p_{ij} = 1. \quad (5)$$

Наиболее важными показателями эффективности функционирования call-центра являются: а) стационарная вероятность занятости каналов $B = \lim_{t \rightarrow \infty} P\{C(t) = c\}$;

б) среднее число повторных вызовов $N = \lim_{t \rightarrow \infty} EN(t)$; в) среднее число занятых каналов обслуживания в стационарном режиме $Y = \lim_{t \rightarrow \infty} EC(t)$;

г) среднее время ожидания вызовов на орбите $W = \frac{N}{\lambda}$ (по формуле Литтла).

Используя аппарат производящих функций, получаем следующие формулы для среднего числа вызовов на орбите:

при $H_2 < 1$

$$N = \frac{\frac{\lambda}{\mu} H_2 + \frac{\lambda}{\mu} (H_1 - H_2) B - H_2 Y - \frac{\lambda}{\mu} H_1 p_{cL}}{\frac{\nu}{\mu} (1 - H_2)};$$

при $H_2 = 1$

$$N = \frac{1 + \frac{\nu}{\mu}}{\frac{\nu}{\mu} \left(c - \frac{\lambda}{\mu} H_1 \right)} \left[\frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu} \right)^2 - E(C(t))^2 - \frac{\lambda}{\mu} (1 - H_1) \left(\frac{\lambda}{\mu} + c + 1 - \frac{\frac{\lambda}{\mu} H_1}{1 + \frac{\nu}{\mu}} \right) \cdot B + \right. \\ \left. + \left(\frac{\lambda}{\mu} + c + 1 + \frac{L \frac{\nu}{\mu} + \frac{\lambda}{\mu} (1 - H_1)}{1 + \frac{\nu}{\mu}} \right) \frac{\lambda}{\mu} H_1 P_{cL}^{(L)} \right].$$

Модель call-центра как СМО типа $M/M/c/0//E_2$. Рассмотрим call-центр как многоканальную систему с пуассоновским входящим потоком вызовов,

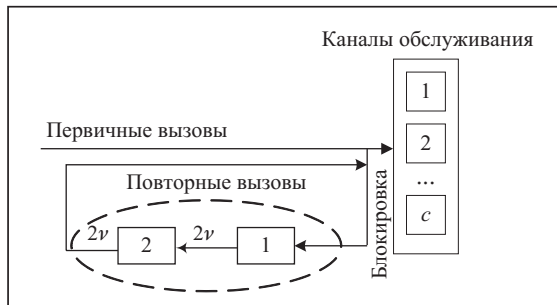


Рис. 1. Call-центр как СМО типа $M/M/c/0//E_2$

экспоненциально распределенным временем обслуживания, без мест ожидания, с неограниченной орбитой, без потерь и двухфазным эрланговским распределением потока требований на орбите (рис. 1).

Пусть на c каналов обслуживания поступает пуассоновский поток первичных вызовов с интенсивностью λ (плотность распределения $a(x) = \lambda e^{-\lambda x}$).

Если при поступлении вызова хотя бы один из каналов обслуживания свободен, то вызов немедленно занимает этот канал и, получив обслуживание, покидает систему. В противном случае вызов становится источником повторных вызовов.

Каждый такой источник создает двухфазный эрланговский процесс повторных вызовов с параметром ν (плотность распределения $d(x) = (2\nu)^2 x e^{-(2\nu)x}$). Если в момент поступления повторного вызова со второй фазы какой-либо канал обслуживания свободен, вызов обслуживается и затем покидает систему, а источник повторных вызовов исчезает. В противном случае вызов сделает попытку снова получить обслуживание.

Предполагается, что времена обслуживания распределены экспоненциально с параметром μ (плотность распределения $b(x) = \mu e^{-\mu x}$).

Построение аналитической модели. Функционирование системы можно описать с помощью трехмерного процесса $(X(t), Y(t), Z(t))$, где $X(t)$ — число занятых каналов обслуживания (для одноканальной системы — канал занят/канал не занят), $Y(t)$ — число вызовов на орбите на первой фазе, $Z(t)$ — число вызовов на орбите на второй фазе в момент t ; сумма $Y(t) + Z(t)$ представляет собой количество вызовов на орбите в момент времени t . Процесс $(X(t), Y(t), Z(t))$ определен на множестве состояний $S = \{0, 1, \dots, c\} \times \{0, 1, \dots\} \times \{0, 1, \dots\}$.

Построим диаграммы переходных состояний СМО $M/M/c/0//E_2$ как модели call-центра (рис. 2, 3). Выпишем интенсивности переходов процесса $(X(t), Y(t), Z(t))$ за интервал времени $(t, t + dt)$, $t \geq 0$. Из состояния (k, i, j) , $k = 0, c$, $i \geq 0$, $j \geq 0$, за время dt система может перейти в другое состояние с определенной вероятностью:

- λdt переходит в состояние $(k + 1, i, j)$, $k \neq c$ (поступил новый первичный вызов и он сразу же получил обслуживание);
- $j \cdot 2\nu dt$ переходит в состояние $(k + 1, i, j - 1)$, $k \neq c$, $j \geq 1$ (один из j повторных вызовов, которые находятся на второй фазе, выполнил удачную попытку обслужиться);

- $i \cdot 2\nu dt$ переходит в состояние $(k, i-1, j+1)$, $i \geq 1$ (один из i повторных вызовов, которые находятся на первой фазе, перешел во вторую фазу);
- $k\mu dt$ переходит в состояние $(k-1, i, j)$, $k \neq 0$ (закончилось обслуживание вызова, один из каналов стал свободным);
- λdt переходит в состояние $(k, i+1, j)$, $k = c$ (поступил новый первичный вызов и, найдя все каналы обслуживания занятыми, перешел на орбиту на первую фазу).

Тогда интенсивности переходов $q_{(k,i,j)(g,n,m)}$, $k, g = \overline{1, c}$, $i, j, n, m = \{0, 1, \dots\}$ (инфинитезимальные переходные интенсивности) процесса $(X(t), Y(t), Z(t))$ из состояния (k, i, j) в состояние (g, n, m) задаются следующим образом:

при $0 \leq k \leq c-1$

$$q_{(kij)(g,n,m)} = \begin{cases} \lambda, & \text{если } (g, n, m) = (k+1, i, j), \\ j \cdot 2\nu, & \text{если } (g, n, m) = (k+1, i, j-1), j \geq 1, \\ i \cdot 2\nu, & \text{если } (g, n, m) = (k, i-1, j+1), i \geq 1, \\ k\mu, & \text{если } (g, n, m) = (k-1, i, j), k \neq 0, \\ -(\lambda + j \cdot 2\nu + i \cdot 2\nu + k\mu), & \text{если } (g, n, m) = (k, i, j), \\ 0 & \text{в остальных случаях;} \end{cases} \quad (6)$$

при $k = c$

$$q_{(cij)(g,n,m)} = \begin{cases} \lambda, & \text{если } (g, n, m) = (c, i+1, j), \\ i \cdot 2\nu, & \text{если } (g, n, m) = (c, i-1, j+1), i \geq 1, \\ c\mu, & \text{если } (g, n, m) = (c-1, i, j), \\ -(\lambda + i \cdot 2\nu + c\mu), & \text{если } (g, n, m) = (c, i, j), \\ 0 & \text{в остальных случаях.} \end{cases} \quad (7)$$

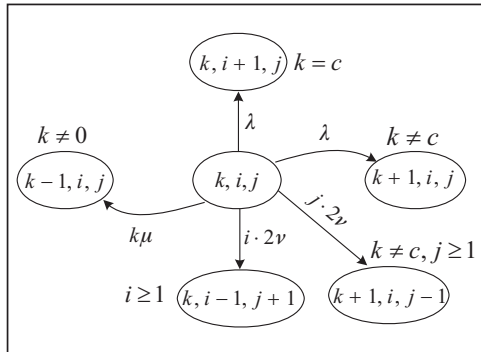


Рис. 2. Диаграмма переходов СМО типа $M/M/c/0//NL/E_2$ для перехода из состояния (k, i, j)

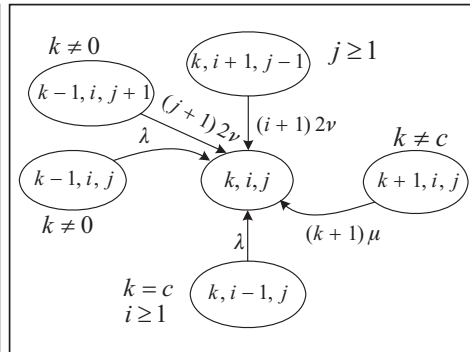


Рис. 3. Диаграмма переходов СМО типа $M/M/c/0//NL/E_2$ для перехода в состояние (k, i, j)

В стационарном режиме СМО типа $M/M/c/0//E_2$ удовлетворяет системе уравнений Колмогорова (см. рис. 2, 3) ($p_{kij} = 0$, если $\forall k, \forall i, \forall j < 0$):

$$(\lambda + j \cdot 2\nu + i \cdot 2\nu + k\mu)p_{kij} = \lambda p_{k-1, i, j} + (k+1)\mu p_{k+1, i, j} + (i+1)2\nu p_{k, i+1, j-1} + (j+1)2\nu p_{k-1, i, j+1} + \lambda p_{k-1, i, j}, \quad 0 \leq k \leq c-1, i \geq 0, j \geq 0; \quad (8)$$

$$(\lambda + i \cdot 2\nu + c\mu)p_{cij} = \lambda p_{c-1, i, j} + (i+1)2\nu p_{c, i+1, j-1} + (j+1)2\nu p_{c-1, i, j+1} + \lambda p_{c, i-1, j}, \quad k = c, i \geq 0, j \geq 0, \quad (9)$$

и условию нормировки

$$\sum_{k=0}^c \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{kij} = 1. \quad (10)$$

Модель call-центра как СМО типа $M/M/c/0/L/E_2$. Поскольку для системы $M/M/c/0///E_2$ довольно сложно получить аналитическое решение, ограничим емкость орбиты достаточно большой константой L (метод Уилкинсона). Данная модель будет адекватна процессу функционирования call-центра, поскольку емкость орбиты на практике — всегда конечное число.

Построение аналитической модели. С учетом введенных ограничений формулы (6), (7) преобразуются следующим образом:

при $0 \leq k \leq c-1, i+j \leq L$

$$q_{(kij)(g,n,m)} = \begin{cases} \lambda, & \text{если } (g, n, m) = (k+1, i, j), \\ j \cdot 2\nu, & \text{если } (g, n, m) = (k+1, i, j-1), j \geq 1, \\ i \cdot 2\nu, & \text{если } (g, n, m) = (k, i-1, j+1), i \geq 1, j < L, \\ k\mu, & \text{если } (g, n, m) = (k-1, i, j), k \neq 0, \\ -(\lambda + i \cdot 2\nu + j \cdot 2\nu + k\mu), & \text{если } (g, n, m) = (k, i, j), \\ 0 & \text{в остальных случаях;} \end{cases} \quad (11)$$

при $k = c, i+j \leq L$

$$q_{(cij)(g,n,m)} = \begin{cases} \lambda, & \text{если } (g, n, m) = (c, i+1, j), i \leq L-1, i+j < L, \\ i \cdot 2\nu, & \text{если } (g, n, m) = (c, i-1, j+1), i \geq 1, j < L, \\ c\mu, & \text{если } (g, n, m) = (c-1, i, j), \\ -(L \cdot 2\nu + c\mu), & \text{если } (g, n, m) = (c, i, j), i = L, j = 0, \\ -(\lambda + c\mu), & \text{если } (g, n, m) = (c, i, j), i = 0, j = L, \\ -(\lambda + i \cdot 2\nu + c\mu), & \text{если } (g, n, m) = (c, i, j), i \leq L-1, j \leq L-1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (12)$$

Для формул (8)–(10) согласно (11), (12) имеем систему ($p_{kij} = 0$, если $\forall k, \forall i, \forall j < 0$ или $\forall k, \forall i, \forall j > L$, или $i+j \leq L$):

$$(\lambda + j \cdot 2\nu + i \cdot 2\nu + k\mu)p_{kij} = \lambda p_{k-1,i,j} + (k+1)\mu p_{k+1,i,j} + (i+1)2\nu p_{k,i+1,j-1} + (j+1)2\nu p_{k-1,i,j+1} + \lambda p_{k-1,i,j}, \quad 0 \leq k \leq c-1, i \geq 0, j \geq 0, i+j \leq L; \quad (13)$$

$$(\lambda + i \cdot 2\nu + c\mu)p_{cij} = \lambda p_{c-1,i,j} + (i+1)2\nu p_{c,i+1,j-1} + (j+1)2\nu p_{c-1,i,j+1} + \lambda p_{c,i-1,j}, \quad k = c, 0 \leq i \leq L-1, 0 \leq j \leq L-1, i+j \leq L; \quad (14)$$

$$(L \cdot 2\nu + c\mu)p_{cL0} = \lambda p_{c-1,L,0} + 2\nu p_{c-1,L,1} + \lambda p_{c,L-1,0}, \quad k = c, i = L, j = 0; \quad (15)$$

$$(\lambda + c\mu)p_{c0L} = \lambda p_{c-1,0,L} + 2\nu p_{c,1,L-1}, \quad k = c, i = 0, j = L, \quad (16)$$

и условие нормировки

$$\sum_{k=0}^c \sum_{i=0}^L \sum_{j=0}^L p_{kij} = 1, \quad i+j \leq L. \quad (17)$$

Поскольку орбита ограничена константой L , то система уравнений (13)–(17) будет конечной, а СМО — эргодичной при любых условиях.

Численное решение системы. Система (13)–(17) может быть решена на компьютере с помощью стандартных процедур. Однако количество памяти, необходимое для хранения матрицы системы, достаточно объемно, хотя большинство ее элементов являются нулевыми. Поэтому решение системы (13)–(17) было выполнено с использованием технологии разреженных матриц (Sparse Matrixes) в среде Matlab.

Некоторые численные результаты. В табл. 1 получены результаты зависимости вероятности занятости каналов $B = \lim_{t \rightarrow \infty} P\{X(t) = c\}$ от емкости орбиты L для

Таблица 1

λ	L	Результаты	
		B Markovian	B Erlang2
0,5	1	0,4000	0,4044
	2	0,4400	0,4458
	3	0,4643	0,4697
	4	0,4790	0,4833
	5	0,4878	0,4908
	50	0,4999	0,5
0,9	1	0,5473	0,5534
	2	0,6018	0,6137
	3	0,6438	0,6604
	4	0,6771	0,697
	5	0,7040	0,7262
	50	0,8949	0,8978
2,0	1	0,7096	0,707
	2	0,7434	0,7396
	3	0,7705	0,7662
	4	0,7927	0,7882
	5	0,7112	0,8064
	50	0,9637	0,9557
	55	0,9667	0,9588
	70	0,9733	0,9659

систем $M/M/c/0/L/M$ (B Markovain) и $M/M/c/0/L/E_2$ (B Erlang2) при $\nu = 0,5$, $\mu = 1$, $c = 1$. Отметим, что с увеличением L разность между вероятностями занятости каналов уменьшается и представляет малую величину.

На рис. 4 получен график зависимости вероятности занятости каналов B от $\tau = 1/\mu$ и λ . Как видим, повторные вызовы влияют на вероятность занятости канала: с увеличением длительности между повторами вероятность занятости канала уменьшается.

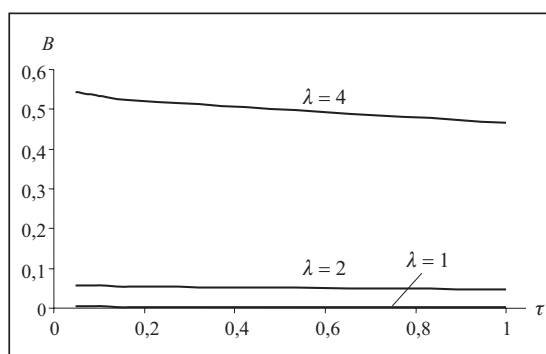


Рис. 4. График зависимости B от τ и λ при $c = 5$, $L = 50$, $\mu = 1$

Заключение. В статье были рассмотрены различные модели call-центров, учитывающие наличие повторных вызовов таких типов, как $M/M/c/0/L/H_j$, $M/M/c/0//E_2$ и $M/M/c/0/L/E_2$. Именно такими моделями можно адекватно описать функционирование call-центров, поскольку они учитывают вторичный, третичный и т.д. потоки вызовов, которые поступают в систему. Полученные численные результаты показали, что повторные вызовы влияют на показатели качества функционирования call-центров.

СПИСОК ЛИТЕРАТУРЫ

- Gans N., Koole G., Mandelbaum A. Telephone call centers: tutorial, review, and research prospects // Manufacturing and Service Operations Management (M&SOM). — 2003. — 5, N 2. — P. 79–141.
- Mandelbaum A. Call centers (centres): research bibliography with abstracts: Version 7. — 2006. — 198 p. — http://iew3.technion.ac.il/serveng/References/US7_CC_avi.pdf.
- Falin G.I., Templeton J.G.C. Retrial queues. — London: Chapman & Hall, 1997. — 395 p.
- Пустовая С.В. Зависимость показателей функционирования call-центра от распределения времени пребывания вызовов на орбите // Кибернетика и системный анализ. — 2009. — № 2. — С. 170–183.
- Aguir M.S., Karaesmen F., Aksin Z., Chauvet F. The impact of retrials on call center performance // Operations Research. — 2004. — 26. — P. 353–376.
- Aguir M.S., Aksin O.Z., Karaesmen F., Dallery Y. On the interaction between retrials and sizing of call centers // European J. of Operational Research. — 2008. — 191, N 2. — P. 398–408.

Поступила 24.11.2009