

МОДЕЛИРОВАНИЕ ГИБРИДНОГО ЦЕНТРА СВЯЗИ С СЕРВИСАМИ САМООБСЛУЖИВАНИЯ И ПОРОГОВЫМ УПРАВЛЕНИЕМ РАЗМЕЩЕНИЕМ ЗАЯВОК

Фархадов М. П.¹, Петухова Н. В.²

*(Учреждение Российской академии наук
Институт проблем управления РАН, Москва)*

Ефросинин Д. В.³

(Российский университет дружбы народов, Москва)

Семенова О. В.⁴

*(ЗАО НПО «Информационные и сетевые технологии»,
Москва)*

Моделируется гибридный центр связи и обработки вызовов современной архитектуры, имеющий сервисы самообслуживания на базе компьютерного распознавания речи. Центр связи представляется открытой экспоненциальной сетью массового обслуживания с двумя узлами и пороговым управлением размещением заявок. Один из узлов моделирует речевые серверы самообслуживания, а другой – операторскую группу. Оба узла рассматриваются как многолинейные системы массового обслуживания с идентичными приборами. Обслуживание заявки автоматом может быть успешным, и тогда заявка покидает систему, а может закончиться неудачей, и тогда заявка по-

¹ Маис Паша оглы Фархадов, кандидат технических наук, заведующий лабораторией (mais@ipu.ru).

² Нина Васильевна Петухова, старший научный сотрудник, (nuret@ipu.ru). (Москва, ул. Профсоюзная, д. 65, тел. (495) 334-87-10).

³ Дмитрий Владимирович Ефросинин, кандидат физико-математических наук, доцент (dmitriy_e@mail.ru).

⁴ Ольга Валерьевна Семенова, кандидат физико-математических наук, старший научный сотрудник, (olgasmnv@gmail.ru).

ступает в операторскую группу. Рассматривается задача оптимизации уровня порога. Определяются стационарные распределения вероятностей состояний сети, вероятностно-временные характеристики обслуживания.

Ключевые слова: гибридный центр связи и обслуживания вызовов, сервис самообслуживания, открытая экспоненциальная сеть, многолинейная система массового обслуживания, дисциплина обслуживания, модели очередей, распознавание речи, многофазное обслуживание.

1. Введение

Существует довольно много моделей для исследования центров обслуживания вызовов, но они ориентированы в основном на поиск экстенсивных путей повышения эффективности центров за счет загрузки операторского корпуса. В работах [3, 10, 11, 14, 15] приведены простые модели для описания функционирования центров обслуживания вызовов на основе «классической» теории массового обслуживания. С математическим аппаратом теории массового обслуживания, используемым в различных моделях, можно ознакомиться в [1, 2, 4, 16]. В [9] приведен перечень усложненных моделей центров обслуживания вызовов без определения области их возможного применения на практике. Модели в [18-26] учитывают такие факторы, как наличие нескольких классов вызовов, различные дисциплины обслуживания заявок, «терпеливость» пользователя, повторные вызовы и др. В [12] предложен материал о центрах обработки вызовов и приведен углубленный анализ различных математических моделей, описывающих процесс функционирования современных центров обработки вызовов. Задача оптимального управления марковской системой массового обслуживания с неоднородными приборами и общей очередью, известная в литературе как «проблема медленного прибора», рассмотрена в [13]. Однако в перечисленных материалах не исследованы центры обработки вызовов современной архитектуры с сервисами самообслуживания на основе речевых технологий.

Такие центры позволяют обеспечить в автоматическом режиме, без участия операторов, полное обслуживание большой доли входящих запросов. Операторы обслуживают оставшуюся часть входного потока, а также те заявки, которые не были обслужены автоматами по причине ошибок распознавателя, или по желанию клиента, или при занятости всех портов самообслуживания, или в случае запросов, не входящих в функционал автомата. Примерами систем обработки вызовов с сервисами самообслуживания на базе речевых технологий могут служить система «Автодиспетчер» для приема заявок на подачу такси [5], система «Автосекретарь» для диспетчеризации вызовов [6], и др. [7, 8]. В работе [17] представлена математическая модель открытой экспоненциальной сети массового обслуживания с двумя узлами, конечным буфером в первом узле и бесконечным числом мест для ожидания во втором узле, описывающая обслуживание заявок в центре обработки вызовов с сервисами самообслуживания. Вызовы клиентов поступают в центр обработки вызовов через сети разного вида: телефонную сеть общего пользования, сети операторов мобильной связи, сеть Интернет. Центр обработки вызовов включает в себя речевые серверы самообслуживания на базе компьютерного распознавания речи и операторскую группу. Коммутация вызовов может осуществляться по разным алгоритмам. В данной работе предлагается математическая модель, описывающая функционирование гибридного центра связи и обслуживания вызовов с бесконечным числом мест ожидания в узлах и с пороговым управлением очередью во втором узле.

2. Математическая модель

Рассматривается открытая экспоненциальная сеть массового обслуживания с двумя узлами. Узел i ($i = \overline{1,2}$) сети представляет собой многолинейную систему массового обслуживания типа $M/M/n_i$ с идентичными приборами. На рис. 1 приведена схема сети с двумя узлами и пороговым управлением размещением заявок.

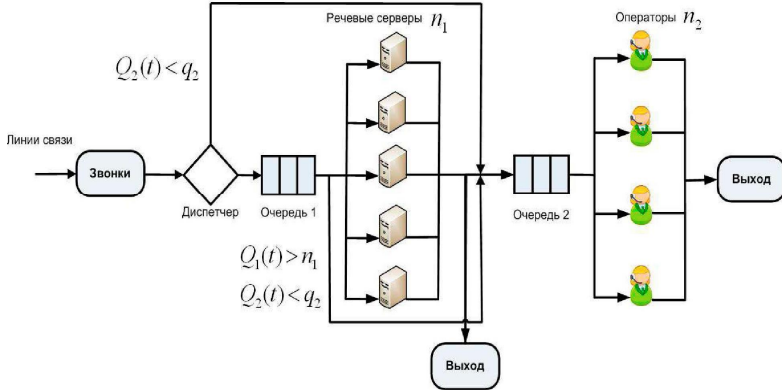


Рис. 1. Схема сети с двумя узлами и пороговым управлением размещением заявок

Поток заявок в сеть является простейшим с параметром λ . Заявка, прибывающая в сеть, поступает в узел 1 или 2 в зависимости от числа заявок в узле 2. Если последнее равно или превышает пороговый уровень q_2 ($q_2 > n_2$), то заявка направляется в узел 1, в противном случае – в узел 2.

Как только число заявок в узле 2 уменьшается до $q_2 - 1$, а в узле 1 есть ожидающие заявки, заявка, стоящая в начале очереди узла 1 переходит в конец очереди узла 2. Выбор данного вида управления обусловлен тем фактом, что время обслуживания в узле 1 может превышать время ожидания в узле 2, так как приборы узла 2 имеют более высокую скорость обслуживания.

Заметим, что приборы узла 1 являются ненадежными в связи с возможными ошибками при распознавании речи речевым сервером. При обслуживании заявки с вероятностью p может произойти отказ. В этом случае заявка покидает прибор узла 1 и переходит в узел 2.

Обозначим через $\{X(t)\}_{t \geq 0} = \{Q_1(t), Q_2(t)\}_{t \geq 0}$ случайный процесс, обозначающий состояние сети в момент времени t , где $Q_i(t)$ – число заявок в i -м узле сети в момент времени t , $i = 1, 2$.

Обозначим через E множество состояний процесса $\{X(t)\}_{t \geq 0}$,

$$E = \{x = (i, j) : i, j \geq 0\},$$

где состояние (i, j) случайного процесса означает, что в узле 1 находится i заявок, а в узле 2 – j заявок.

3. Стационарное распределение вероятностей состояний сети

Представленная сеть массового обслуживания относится к классу экспоненциальных сетей, таким образом, случайный процесс $\{X(t)\}_{t \geq 0}$, описывающий поведение этой системы, является однородным и марковским. Предположим, что существуют стационарные вероятности

$$\pi(i, j) = \lim_{t \rightarrow \infty} P[X(t) = (i, j)].$$

Условие эргодичности, при котором существуют эти вероятности, будет приведено позже. Система уравнений равновесия имеет следующий вид:

- (1) $\lambda\pi(0, 0) = (1 - p)\mu_1\pi(1, 0) + \mu_2\pi(0, 1),$
- (2) $(\lambda + i\mu_1 + \min\{j, n_2\}\mu_2)\pi(i, j) = \lambda\pi(i, j - 1) +$
 $+ (i + 1)(1 - p)\mu_1\pi(i + 1, j) + (i + 1)p\mu_1\pi(i + 1, j - 1) +$
 $+ \min\{j + 1, n_2\}\mu_2\pi(i, j + 1), \quad 0 \leq i < n_1, \quad 0 < j < q_2,$
- (3) $(\lambda + i\mu_1)\pi(i, 0) = (i + 1)(1 - p)\mu_1\pi(i + 1, 0) + \mu_2\pi(i, 1), \quad 0 \leq i < n_1,$
- (4) $(\lambda + n_1\mu_1)\pi(n_1, 0) = \mu_2\pi(n_1, 1),$
- (5) $(\lambda + i\mu_1 + \min\{j, n_2\}\mu_2)\pi(n_1, j) = \lambda\pi(n_1, j - 1) +$
 $+ \min\{j + 1, n_2\}\mu_2\pi(n_1, j + 1), \quad 0 < j < q_2,$
- (6) $(\lambda + n_1\mu_1 + n_2\mu_2)\pi(n_1, q_2) = \lambda\pi(n_1, q_2 - 1) +$
 $+ \lambda\pi(n_1 - 1, q_2) + n_1(1 - p)\mu_1\pi(n_1 + 1, q_2) +$
 $+ n_2\mu_2\pi(n_1 + 1, q_2) + n_2\mu_2\pi(n_1, q_2 + 1),$
- (7) $(\lambda + i\mu_1 + n_2\mu_2)\pi(i, j) = \lambda\pi(i - 1, j)I_{\{i > 0\}} +$
 $+ \lambda\pi(i, j - 1)I_{\{j = q_2\}} + \min\{i + 1, n_1\}(1 - p)\mu_1\pi(i + 1, j) +$
 $+ \min\{i + 1, n_1\}p\mu_1\pi(i + 1, j - 1) + n_2\mu_2\pi(i, j + 1),$
 $0 \leq i < n_1, \quad j \geq q_2, \quad (i, j) \neq (n_1, q_2),$

$$(8) \quad (\lambda + i\mu_1 + n_2\mu_2)\pi(i, j) = \lambda\pi(i-1, j) + n_1(1-p)\mu_1\pi(i+1, j) + n_1p\mu_1\pi(i+1, j-1) + n_2\mu_2\pi(i, j+1) + n_2\mu_2\pi(i+1, q_2)I_{\{j=q_2\}},$$

$$i > n_1, j \geq q_2.$$

Теорема 1. Если выполнено условие

$$\rho_1 = \frac{\lambda}{n_1\mu_1} < 1, \quad \rho_2 = \frac{p\lambda}{n_2\mu_2} < 1,$$

то процесс $\{X(t)\}_{t \geq 0}$ является эргодическим, а решение системы уравнений равновесия (1)–(8) представляется в форме

$$(9) \quad \pi(i, j) = \left(\frac{\lambda p}{n_2\mu_2}\right)^{j-q_2-1} \left(\frac{\lambda}{\mu_1}\right)^i \frac{1}{i!} \pi(0, q_2+1), \quad 0 \leq i < n_1, \quad j \geq q_2+1,$$

$$(10) \quad \pi(i, j) = \left(\frac{\lambda p}{n_2\mu_2}\right)^{j-q_2-1} \left(\frac{\lambda}{\mu_1}\right)^i \frac{1}{n_1! n_1^{i-n_1}} \pi(0, q_2+1), \quad i \geq n_1, \quad j \geq q_2+1,$$

$$(11) \quad \pi(i, q_2) = \left(\frac{\lambda}{\mu_1}\right)^i \frac{n_2\mu_2}{\lambda p} \frac{1}{n_1! n_1^{i-n_1}} \pi(0, q_2+1), \quad i \geq n_1+1,$$

$$(12) \quad \pi(i, j) = \left[\pi_{q_2} C_{q_2} - (n_1(1-p)\mu_1 + n_2\mu_2) \hat{G}_{n_2} \pi(n_1+1, q_2) \right] \times$$

$$\times \prod_{k=1}^{q_2-j} G_{q_2-k} e_i, \quad 0 \leq i \leq n_1, \quad 0 \leq j \leq q_2,$$

где вектор-строка

$\pi_{q_2+1} = (\pi(0, q_2+1), \pi(1, q_2+1), \dots, \pi(n_1, q_2+1))$. Матрицы G_i определяются рекуррентными соотношениями

$$G_0 = -Q_{1,0} Q_{0,0}^{-1},$$

$$G_j = -Q_{j+1,j} \hat{G}_j, \quad 1 \leq j \leq q_2,$$

$$\hat{G}_j = (Q_{j,j} + G_{j-1} Q_{j-1,j})^{-1}$$

где матрицы $Q_{i,j}, i \in \{j-1, j, j+1\}$, имеют вид

$$(13) \quad Q_{j-1,j} = \text{diag} \left\{ \underbrace{\lambda, \dots, \lambda}_{n_1+1} \right\} + \\
 + \text{diag}^- \{ p\mu_1, p2\mu_1, \dots, pn_1\mu_1 \}, j = \overline{1, q_2}, \\
 Q_{j,j} = \text{diag} \{ q_{0,j}, q_{1,j}, \dots, q_{n_1,j} \} + \\
 + \text{diag}^- \{ (1-p)\mu_1, 2(1-p)\mu_1, \dots, n_1(1-p)\mu_1 \} + \\
 + I_{\{j=q_2\}} \text{diag}^+ \{ \lambda, \dots, \lambda \}, j = \overline{0, q_2},$$

$$Q_{j+1,j} = \text{diag} \left\{ \underbrace{\min \{ j+1, n_2 \} \mu_2, \dots, \min \{ j+1, n_2 \} \mu_2}_{n_1+1} \right\}, j = \overline{0, q_2},$$

где $q_{j,j} = -(\lambda + i\mu_1 + j\mu_2)$. $\text{diag} \{ a_1, \dots, a_n \}$ – диагональная матрица размера n с диагональными элементами a_1, \dots, a_n , $\text{diag}^- \{ a_1, \dots, a_{n-1} \}$ и $\text{diag}^+ \{ a_1, \dots, a_{n-1} \}$ – соответственно, нулевые матрицы размера n с под- и наддиагональными элементами a_1, \dots, a_{n-1} . e_i обозначает нулевой вектор-столбец, $(i+1)$ -й элемент которого равен 1.

Вероятность $\pi(0, q_2 + 1)$ вычисляется из условия нормировки

$$(14) \quad \sum_{i=0}^{n_1} \sum_{j=0}^{q_2} \pi(i, j) + \sum_{i=n_1+1}^{\infty} \pi(i, q_2) + \sum_{i=0}^{n_1-1} \sum_{j=q_2+1}^{\infty} \pi(i, j) + \sum_{i=n_1}^{\infty} \sum_{j=q_2+1}^{\infty} \pi(i, j) = 1,$$

причем бесконечные суммы сходятся, если $p_i < 1$, $i = 1, 2$ и

$$\sum_{j=n_1+1}^{\infty} \pi(i, q_2) = \frac{n_2 \mu_2 \left(\frac{\lambda}{\mu_1} \right)^{n_1}}{p(n_1 \mu_1 - \lambda) n!} \pi(0, q_2 + 1), \\
 \sum_{i=0}^{n_1-1} \sum_{j=q_2+1}^{\infty} \pi(i, j) = \frac{n_2 \mu_2 e^{\frac{\lambda}{\mu_1}} \Gamma \left[n_1, \frac{\lambda}{\mu_1} \right]}{(n_2 \mu_2 - p\lambda)(n_1 - 1)!} \pi(0, q_2 + 1),$$

$$\sum_{i=n_1}^{\infty} \sum_{j=q_2+1}^{\infty} \pi(i, j) = \frac{n_2 \mu_1 \mu_2 \left(\frac{\lambda}{\mu_1}\right)^{n_1}}{(\lambda - n_1 \mu_1)(p\lambda - n_2 \mu_2)(n_1 - 1)!} \pi(0, q_2 + 1).$$

Доказательство. Заметим, что если число заявок во втором узле будет больше $q_2 + 1$, система будет функционировать как классическая сеть Джексона, имеющая два узла с интенсивностями поступления $\lambda_1 = \lambda$ и $\lambda_2 = p\lambda$ в первый и второй узел соответственно, поэтому предположим, что для состояний $x = (i, j)$, где $i \geq 0$, $j \geq q_2 + 1$, имеет место мультипликативное представление стационарных вероятностей

$$(15) \quad \pi(i, j) = \left(\frac{\lambda p}{n_2 \mu_2}\right)^{j-q_2-1} \left(\frac{\lambda}{\mu_1}\right)^i \frac{1}{i!} \pi(0, q_2 + 1), \quad 0 \leq i \leq n_1 - 1,$$

$$j \geq q_2 + 1,$$

$$\pi(i, j) = \left(\frac{\lambda p}{n_2 \mu_2}\right)^{j-q_2-1} \left(\frac{\lambda}{\mu_1}\right)^i \frac{1}{n! n_1^{i-n_1}} \pi(0, q_2 + 1), \quad i \geq n_1,$$

$$j \geq q_2 + 1.$$

Пусть $\pi_k(i)$ – стационарная вероятность того, что число заявок в узле k равно i , $i \geq 0$, $k = \overline{1, 2}$. Тогда в нашем предположении для $i \geq 0$, $j \geq q_2 + 1$ стационарные вероятности $\pi(i, j)$ будут иметь вид произведения

$$\pi(i, j) = \pi_1(i) \pi_2(j).$$

Вероятности $\pi_1(i)$, $\pi_2(j)$ для $i \geq 0$, $j \geq q_2 + 1$ можно получить, рассматривая отдельно первый и второй узел как системы массового обслуживания типа $M / M / n_1$ и $M / M / n_2$:

$$\pi(i) = \begin{cases} \pi_1(0) \left(\frac{\lambda}{\mu_1}\right)^i \frac{1}{i!}, & i = \overline{0, n_1 - 1}, \\ \pi_1(0) \left(\frac{\lambda}{\mu_1}\right)^i \frac{1}{n! n_1^{i-n_1}}, & i \geq n_1, \end{cases}$$

$$\pi_2(j) = \left(\frac{\lambda p}{n_2 \mu_2} \right)^{j-q_2-1} \pi_2(q_2+1), \quad j \geq q_2+1.$$

Таким образом, приходим к формулам (15), где $\pi(0, q_2+1) = \pi_1(0)\pi_2(q_2+1)$.

Далее преобразуем уравнения (7) и (8). Выражая из правой части этих равенств вероятности $\pi(i, q_2)$, $i \geq n_1+1$, а остальные вероятности выражая через $\pi(0, q_2+1)$, используя полученные выше формулы, после преобразований получим формулу (11).

Оставшиеся вероятности состояний ниже границы (n_1, n_2) вычисляются с помощью матричного подхода. Введем в рассмотрение следующие векторы стационарных вероятностей

$$\pi_j = (\pi(0, j), \pi(1, j), \dots, \pi(n_1, j)), \quad 0 \leq j \leq n_2.$$

Преобразуя уравнения (1)–(7) для $0 \leq i \leq n_1$ и $0 \leq j \leq q_2$ в векторно-матричную форму, получаем следующие уравнения для векторов π_j , $0 \leq j \leq q_2$:

$$\begin{aligned} \pi_0 Q_{0,0} + \pi_1 Q_{1,0} &= 0, \\ \pi_{j-1} Q_{j-1,j} + \pi_j Q_{j,j} + \pi_{j+1} Q_{j+1,j} &= 0, \quad 1 \leq j \leq q_2-1, \\ \pi_{q_2-1} Q_{q_2-1,q_2} + \pi_{q_2} Q_{q_2,q_2} + \pi_{q_2+1} Q_{q_2+1,q_2} + \\ &+ (n_1(1-p)\mu_1 + n_2\mu_2)e_{n_1}^t \pi(n_1+1, q_2) = 0, \end{aligned}$$

где матрицы $Q_{i,j}$, $i \in \{j-1, j, j+1\}$ имеют приведенную в утверждении форму. Осуществляя рекуррентную подстановку в последней системе, получим соотношения для векторов π_j в виде

$$\begin{aligned} \pi_j &= \pi_{j+1} G_j, \quad 0 \leq j \leq q_2-1, \\ \pi_{q_2} &= \pi_{q_2+1} Q_{q_2} - (n_1(1-p)\mu_1 + n_2\mu_2)e_{n_1}^t \hat{G}_{q_2} \pi(n_1+1, q_2). \end{aligned}$$

Из последней системы получаем

$$\pi_j = \left[\pi_{q_2+1} Q_{q_2} - (n_1(1-p)\mu_1 + n_2\mu_2)e_{n_1}^t \hat{G}_{q_2} \pi(n_1+1, q_2) \right] \prod_{k=1}^{q_2-j} G_{q_2-k}.$$

Отсюда непосредственно следует формула (12).

Таким образом, все вероятности системы представляются в виде функции, зависящей от вероятности $\pi(0, q_2 + 1)$, которая, в свою очередь, вычисляется из условия нормировки (14). Нетрудно показать, что для сходимости входящих в выражение бесконечных сумм необходимо, чтобы загрузки первого и второго узлов были меньше 1.

Дальнейшая подстановка полученных выражений в уравнения (1)–(8) подтверждает справедливость сделанных предположений относительно мультипликативной формы (15), что завершает доказательство теоремы.

4. Характеристики производительности системы

Вычислив стационарное распределение вероятностей состояний сети, далее можно получить характеристики производительности сети, которые приведены в следствиях 1–7.

Следствие 1. Среднее число занятых приборов \bar{C} вычисляется по формуле

$$\bar{C} = \bar{C}_1 + \bar{C}_2,$$

где \bar{C}_k обозначает среднее число занятых приборов узла $k \in \{1, 2\}$

$$\begin{aligned} \bar{C}_1 = & \sum_{i=0}^{n_1} \sum_{j=0}^{\infty} \min\{i, n_1\} \pi(i, j) + \sum_{i=n_1+1}^{\infty} \sum_{j=q_2}^{\infty} n_1 \pi(i, j) = \sum_{i=0}^{n_1} \sum_{j=0}^{q_2} \min\{i, n_1\} \pi(i, j) + \\ & + \frac{n_2 \mu_2 \left(\frac{\lambda}{\mu_1}\right)^{n_1} \left[n_1 n_2 \mu_1 \mu_2 + \lambda p (\lambda - n_1 \mu_1) \right] - n_2 \mu_2 (\lambda - n_1 \mu_1) e^{\frac{\lambda}{\mu_1}} \Gamma\left(n_1 + 1, \frac{\lambda}{\mu_1}\right)}{\mu_1 p (\lambda - n_1 \mu_1) (p \lambda - n_2 \mu_2) n_1!} \times \\ & \times \pi(0, q_2 + 1), \end{aligned}$$

$$\begin{aligned} \bar{C}_2 = & \sum_{i=0}^{n_1} \sum_{j=0}^{\infty} \min\{j, n_2\} \pi(i, j) + \sum_{i=n_1+1}^{\infty} \sum_{j=q_2}^{\infty} n_2 \pi(i, j) = \\ = & \sum_{i=0}^{n_1} \sum_{j=0}^{q_2} \min\{j, n_2\} \pi(i, j) + \end{aligned}$$

$$+ \frac{n_2^2 \mu_2 \left[n_2 \mu_2 \left(\frac{\lambda}{\mu_1} \right)^{n_1} - p(\lambda - n_1 \mu_1) e^{\frac{\lambda}{\mu_1}} \Gamma \left(n_1 + 1, \frac{\lambda}{\mu_1} \right) \right]}{p(\lambda - n_1 \mu_1) (p\lambda - n_2 \mu_2) n_1!} \pi(0, q_2 + 1),$$

$\Gamma(a, z) = \int_z^\infty t^{a-1} e_{-t} dt$ – неполная гамма-функция.

Следствие 2. Среднее число заявок \bar{Q}_k в очередях системы вычисляется по формуле

$$\bar{Q} = \bar{Q}_1 + \bar{Q}_2,$$

где \bar{Q}_k обозначает среднее число заявок в очереди узла $k = \overline{1, 2}$,

$$\begin{aligned} \bar{Q}_1 &= \sum_{i=n_1+1}^{\infty} \sum_{j=q_2}^{\infty} (i - n_1) \pi(i, j) = \\ &= \frac{n_2^2 \mu_2^2 \mu_1 \left(\frac{\lambda}{\mu_1} \right)^{n_1}}{p(\lambda - n_1 \mu_1)^2 (n_2 \mu_2 - p\lambda) (n_1 - 1)!} \pi(0, q_2 + 1) \\ \bar{Q}_2 &= \sum_{i=0}^{n_1} \sum_{j=n_2+1}^{\infty} (j - n_2) \pi(i, j) + \sum_{i=n_1+1}^{\infty} \sum_{j=q_2}^{\infty} (j - n_2) \pi(i, j) = \\ &= \sum_{i=0}^{n_1} \sum_{j=n_2+1}^{q_2} (j - n_2) \pi(i, j) + \\ &+ \frac{n_2^2 \mu_2^2 \left(\frac{\lambda}{\mu_1} \right)^{n_1} A_1 - n_2 \mu_2 p(\lambda - n_1 \mu_1) A_2 e^{\frac{\lambda}{\mu_1}} \Gamma \left(n_1 + 1, \frac{\lambda}{\mu_1} \right)}{p(\lambda - n_1 \mu_1) (p\lambda - n_2 \mu_2)^2 n_1!} \pi(0, q_2 + 1), \end{aligned}$$

$$A_1 = \lambda p(q_2 - n_2 - 1) - n_2 \mu_2 (q_2 - n_2),$$

$$A_2 = A_1 + \lambda p - n_2 \mu_2.$$

Следствие 3. Среднее число заявок \bar{N} в системе вычисляется по формуле

$$\begin{aligned} \bar{N} &= \bar{N}_1 + \bar{N}_2 = \sum_{k=1}^2 \bar{C}_k + \sum_{k=1}^2 \bar{Q}_k = \\ &= \sum_{i=0}^{n_1} \sum_{j=0}^{\infty} (i+j)\pi(i,j) + \sum_{i=n_1+1}^{\infty} \sum_{j=q_2}^{\infty} (i+j)\pi(i,j) = \\ &= \sum_{i=0}^{n_1} \sum_{j=0}^{q_2} (j-n_2)\pi(i,j) + \\ &+ \frac{n_2\mu_2 \left[\left(n_2\mu_2 \frac{\lambda}{\mu_1} \right)^{n_1} B_1 - p\lambda(\lambda - n_1\mu_1)^2 B_2 e^{\frac{\lambda}{\mu_1}} \Gamma \left(n_1 + 1, \frac{\lambda}{\mu_1} \right) \right]}{n_1\mu_1 p(\lambda - n_1\mu_1)^2 (p\lambda - n_2\mu_2)^2 (n_1 - 1)!} \times \\ &\times \pi(0, q_2 + 1), \end{aligned}$$

где

$$\begin{aligned} B_1 &= p^2 \lambda^2 (\lambda - n_1\mu_1)^2 - n_2\mu_2 p\lambda (\lambda^2 - \lambda\mu_1) \times \\ &\times (q_2 + 3n_1 - 1) + \mu_1^2 n_1 (q_2 + 2n_1) + \\ &+ n_2^2 \mu_2^2 \mu_1 (n_1\mu_1 (q_2 + n_1 + 1) - \lambda(q_2 + n_1)), \\ B_2 &= p\lambda(\lambda + q_2\mu_1) - n_2\mu_2(\lambda + \mu_1(1 + q_2)), \end{aligned}$$

\bar{N}_k обозначает среднее число заявок в узле $k = \overline{1,2}$:

$$\bar{N}_k = \bar{C}_k + \bar{Q}_k.$$

Следствие 4. Вероятность τ_k того, что заявка, поступив в систему, попадает в узел $k = \overline{1,2}$

$$\begin{aligned} \tau_1 &= \sum_{i=0}^{\infty} \sum_{j=q_2}^{\infty} \pi(i,j) = \\ &= \frac{n_2\mu_2 \left[\left(n_2\mu_2 \frac{\lambda}{\mu_1} \right)^{n_1} - p(n_1\mu_1 - \lambda) e^{\frac{\lambda}{\mu_1}} \Gamma \left(n_1 + 1, \frac{\lambda}{\mu_1} \right) \right]}{p(\lambda - n_1\mu_1)(p\lambda - n_2\mu_2)n_1!} \times, \\ &\times \pi(0, q_2 + 1), \tau_2 = 1 - \tau_1. \end{aligned}$$

С помощью формулы Литтла нетрудно получить среднее время ожидания (пребывания) заявки в сети.

Следствие 5. Среднее время ожидания (пребывания) \bar{W} (\bar{T}) заявки в сети вычисляется по формуле

$$\bar{W} = \frac{\bar{Q}}{\lambda}, \quad \bar{T} = \frac{\bar{N}}{\lambda}.$$

Следствие 6. Средний период занятости \bar{L} системы вычисляется по формуле

$$\bar{L} = \frac{1}{\lambda} \left(\frac{1}{\pi(0,0)} - 1 \right).$$

Следствие 7. Среднее число заявок \bar{N}_L , обслуженных за период занятости, определяется как

$$\bar{N}_L = \lambda \bar{L} + 1.$$

5. Задача оптимизации и численные примеры

В данном разделе рассмотрим задачу оптимизации порогового уровня q_2 и введем следующую структуру штрафов:

$c_{0,k}$ – стоимость ожидания заявки в очереди узла k в единицу времени,

$c_{u,k}$ – стоимость работы одного прибора в узле k в единицу времени.

Задача состоит в минимизации функционала потерь

$$(16) \quad \bar{V}(q_2) := \bar{V}(\lambda, \mu_1, \mu_2, p, n_1, n_2, q_2) \rightarrow \min_{q_2},$$

который в данном случае имеет вид

$$(17) \quad \bar{V}(q_2) = \sum_{k=1}^2 c_{u,k} \bar{C}_k + \sum_{k=1}^2 c_{0,k} \bar{Q}_k.$$

Замечание 1. Из формулы (17) следует, что:

1) если $c_{0,k} = c_{u,k} = 1$, $k = \overline{1,2}$, то задача сводится к минимизации среднего числа заявок в системе \bar{N} или времени пребывания \bar{T} .

2) если $c_{0,k} = 1$, $c_{u,k} = 0$, $k = \overline{1,2}$, то задача сводится к минимизации

ции среднего числа заявок в очереди \bar{Q} или времени ожидания \bar{W} .

Для проведения численного анализа рассмотрим два примера для случая $n_1 = 2$ и $n_2 = 4$. В табл. 1 представлены результаты вычисления оптимальных пороговых уровней q_2^* и соответствующих величин \bar{V} , \bar{N} , \bar{Q} и \bar{L} , принимающих при этих порогах минимальные значения, для различных значений стоимостей и интенсивностей обслуживания. Другие параметры системы принимают следующие значения:

$$\lambda = 0.9, p = 0.01, c_{0,1} = c_{0,2} = 2,5.$$

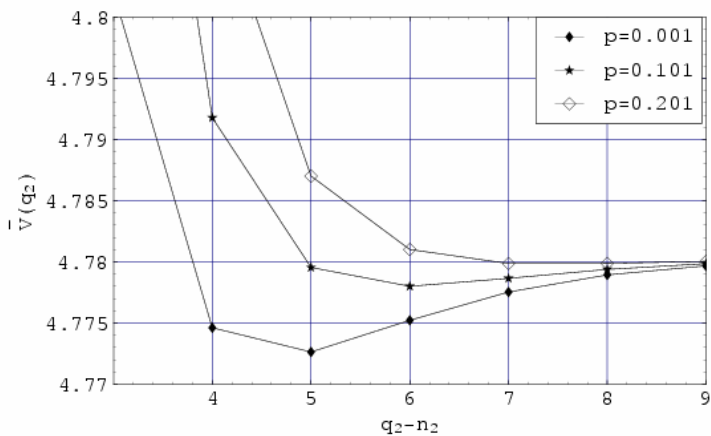
Из таблицы видно, что при уменьшении μ_1 и увеличении $c_{u,1}$, и, наоборот, при увеличении μ_2 и уменьшении $c_{u,2}$ значения оптимальных пороговых уровней для характеристик \bar{V} , \bar{N} и \bar{Q} увеличиваются.

Таблица 1. Значения функций $\bar{V}(q_2^*)$, $\bar{N}(q_2^*)$, $\bar{Q}(q_2^*)$ и $\bar{L}(q_2^*)$.

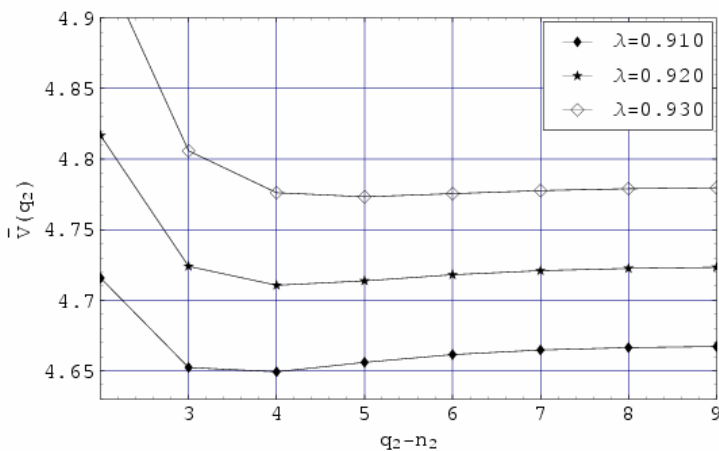
$c_{u,1}, c_{u,2}, \mu_1, \mu_2$	q_2^*	\bar{V}	\bar{N}	\bar{Q}	\bar{L}
0.5, 1.0, 0.5, 0.6	(6, 8, 6, 7)	1.5970	1.5419	0.0380	3.9147
0.5, 1.0, 1.5, 0.6	(4, 4, 4, 4)	1.4694	1.4891	0.0006	3.8491
0.5, 1.0, 0.5, 1.6	(6, 8, 5, 12)	0.5636	0.5630	0.0003	0.8391
0.5, 1.0, 1.5, 1.6	(5, 6, 4, 5)	0.5632	0.5629	0.0000	0.8391
2.5, 1.0, 0.5, 0.6	(8, 8, 6, 7)	1.6060	1.5419	0.0380	3.9147
2.5, 1.0, 1.5, 0.6	(5, 4, 4, 4)	1.5275	1.4891	0.0006	3.8491
2.5, 1.0, 0.5, 1.6	(7, 8, 5, 12)	0.5636	0.5630	0.0003	0.8391
2.5, 1.0, 1.5, 1.6	(6, 6, 4, 5)	0.5636	0.5629	0.0000	0.8391
0.5, 3.0, 0.5, 0.6	(6, 8, 6, 7)	4.5867	1.5419	0.0380	3.9147
0.5, 3.0, 1.5, 0.6	(4, 4, 4, 4)	4.3634	1.4891	0.0006	3.8491
0.5, 3.0, 0.5, 1.6	(6, 8, 5, 12)	1.6886	0.5630	0.0003	0.8391
0.5, 3.0, 1.5, 1.6	(5, 6, 4, 5)	1.6880	0.5629	0.0000	0.8391
2.5, 3.0, 0.5, 0.6	(7, 8, 6, 7)	4.6002	1.5419	0.0380	3.9147
2.5, 3.0, 1.5, 0.6	(4, 4, 4, 4)	4.4462	1.4891	0.0006	3.8491
2.5, 3.0, 0.5, 1.6	(8, 8, 5, 12)	1.6886	0.5630	0.0003	0.8391
2.5, 3.0, 1.5, 1.6	(5, 6, 4, 5)	1.6885	0.5629	0.0000	0.8391

Для следующего примера рассмотрим следующие значения параметров сети:

$$\mu_1 = 0.5, \mu_2 = 0.6, c_{0,1} = c_{0,2} = 2.5, c_{u,1} = 0.5, c_{u,2} = 3.0.$$



a)



b)

Рис. 2. Значение функции $\bar{V}(q_2)$ для различных значений q_2 , p и λ

В соответствии с выбранными значениями параметров полагаем, что узел 1, описывающий сервер с ненадежным сервисом распознавания речи, имеет более медленную скорость обслуживания, но при этом малые затраты на обслуживание. В то же время, узел 2 описывает работу операторов, являясь надежным, более быстрым и, одновременно, более дорогим по сравнению с узлом 1. На рис. 2 (а) для $\lambda = 0.9$ показано влияние вероятности p ошибки при обслуживании в узле 1, а на рис. 2 (b) – интенсивности λ поступления новых заявок для $p = 0.01$ на функционал потерь $\bar{V}(q_2)$. Очевидно, что увеличение значений параметров p и λ приводит к увеличению оптимального порогового уровня q_2 , так как в этом случае возрастает нагрузка на систему и возникает необходимость более интенсивного использования быстрого, но, при этом, дорогого узла 2.

6. Заключение

В работе рассмотрена математическая модель, которая позволяет исследовать характеристики телефонных систем массового обслуживания с комбинацией традиционных методов обслуживания и сервисов самообслуживания на основе компьютерных речевых технологий и современных интерактивных средств взаимодействия. Модель учитывает специфику, вносимую этими технологиями. Предложено моделирование гибридного центра связи экспоненциальной сетью массового обслуживания с пороговым управлением размещением заявок. Полученные результаты могут быть использованы для проектирования, исследования и моделирования работы центров связи и обслуживания вызовов с применением современных речевых интерфейсных технологий.

Литература

1. БОЧАРОВ П.П., ПЕЧИНКИН А.В. *Теория массового обслуживания*. // М.: УДН, 1995.

2. ВИШНЕВСКИЙ В.М. *Теоретические основы проектирования компьютерных сетей.* // М.: Техносфера, 2003.
3. ГОЛЬДШТЕЙН Б.С., ФРЕЙНКМАН В.А. *Call-центры и компьютерная телефония.* // СПб.: БХВ – Санкт-Петербург, 2002. – 372 с.
4. ДУДИН А.Н., КЛИМЕНОК В.И. *Системы массового обслуживания с коррелированными потоками.* // Минск: БГУ, 2000.
5. ЖОЖИКАШВИЛИ В.А., ПЕТУХОВА Н.В., ЗАЦЕПИН А.Н., АЗАРОВ В.В. *Современные технологии управления в диспетчерской службе такси* // Проблемы управления. 2006. №2. С. 32 – 34.
6. ЖОЖИКАШВИЛИ В.А., БИЛИК Р.В., ВЕРТЛИБ В.А., МЯСОЕДОВА З.П., ПЕТУХОВА Н.В., ФАРХАДОВ М.П. *Интеллектуальные телефонные услуги на основе речевых технологий* // Информационно-измерительные и управляющие системы. 2007. № 2. С. 75 – 78.
7. ЖОЖИКАШВИЛИ В.А., ПЕТУХОВА Н.В., ФАРХАДОВ М.П. *Компьютерные системы массового обслуживания и речевые технологии* // Проблемы управления. 2006. № 2. С. 3 – 7.
8. ЖОЖИКАШВИЛИ В.А., ПЕТУХОВА Н.В., ФАРХАДОВ М.П. *Мультисерверная архитектура интеллектуальных порталов самообслуживания* // IV Международная конференция по проблемам управления (МКПУ-IV). – Москва. – 2009. – С. 1744 – 1748.
9. ЗАРУБИН А.А. *Call- и контакт-центры: эволюция технологий и математических моделей* // Вестник связи. – 2003, – №8. – С. 85 – 88.
10. РОСЛЯКОВ А.В., САМСОНОВ М.Ю., ШИБАЕВА И.В. *Центры обслуживания вызовов (Call centre).* – М.: Эко-Трендз, 2002. – 272 с.
11. РОСЛЯКОВ А.В. *Современное состояние и прогнозы развития центров обслуживания вызовов.* // Инфосфера, 2001. №11.
12. РОСЛЯКОВ А.В. *Математические модели центров обслуживания вызовов* // М.: ИРИАС, 2006. – 336 с.
13. РЫКОВ В.В., ЕФРОСИНИН Д.В. *О проблеме медленного прибора* // Автоматика и телемеханика. 2009. №12. С. 81-91.

14. САМОЛЮБОВА А.Б. *Call Center на 100%*. // 2004, Москва: Альпина Бизнес Букс.
15. СОЛОНИН В., *Call-центры в современном бизнесе России*, 2005 // <http://www.cnews.ru/reviews/free/call-center/>.
16. ТАХА Х.А Введение в исследование операций. Пер. с англ. Изд. 6-е. // М.: Издательский дом "Вильямс 2001. – 912 с.
17. ФАРХАДОВ М.П., ПЕТУХОВА Н.В., ЕФРОСИНИН Д.В., СЕМЕНОВА О.В. *Математическая модель центра обслуживания вызовов с сервисами самообслуживания* // Proceedings of International Workshop "Distributed Computer and Communication Networks DCCN'2009". Sofia, Bulgaria. М.: R&D Company "Information and Networking Technologies 2009. P. 86 – 95.
18. ATAR R., A. MANDELBAUM AND M. I. REIMAN. 2004a. *Scheduling a multi-class queue with many exponential servers*. // Ann. Appl. Probab. 14, 1084 – 1134. MR2071417
19. ATAR R., A. MANDELBAUM AND M. I. REIMAN. 2004b. *Brownian control problems for queueing systems in the Halfin-Whitt regime*. // Ann. Appl. Probab. 14, P. 1084 – 1134. MR2071417.
20. BORST S., MANDELBAUM A., REIMAN M.I. *Dimensioning large call centers* // Operations research. – 2004. – Vol. 52-N I- P. 17 – 34.
21. CANS N., KOOLE G., MANDELBAUM A. *Telephone call centers: Tutorial, review, and research prospects* // Manufacturing and service operations management. – 2003. – N 5-P. 79 – 141.
22. GARNETT O., MANDELBAUM A., REIMAN M. *Designing a call center with impatient customers* // Manufacturing and service operations management. – 2002. – N 4. P. 208 – 227.
23. FEINBERG M.A. *Performance characteristics of automated call distribution systems* // IEEE. – GLOBECOM '90, 1990. – P. 415 – 419.
24. KOOLE G. *Call Center Mathematics, Version of January 26 of 2007*. // <ftp://www.math.vu.nl/~koole/ccmath/book.pdf>.
25. MANDELBAUM, A., ZELTYN S. *The Erlang-A/Palm queue, with applications to call centers*. // Working paper, The Tech-

nion, Haifa, Israel. 2005. Available at:
<http://iew3.technion.ac.il/serveng/References/references.html>

26. ZELTYN, S., MANDELBAUM A., *Call centers with impatient customers: many-server asymptotics of the M/M/n+G queue. // Queueing Syst.* 51 (2005), no. 3 – 4, P. 361 – 402.

MODELING HYBRID COMMUNICATION CENTER WITH SELF-SERVICES AND THRESHOLD QUEUING CONTROL

Mais Farkhadov, Institute of Control Sciences of RAS, Moscow, Cand.Sc., head of the laboratory (mais@ipu.ru).

Nina Petukhova, Institute of Control Sciences of RAS, Moscow, Senior research assistant (Moscow, Profsoyuznaya st., 65, (495)334-87-10).

Dmytry Efrosinin, Peoples' Friendship University of Russia, Cand.Sc., prof. assistant (dmitriy_e@mail.ru).

Olga Semenova, ZAO Research Development Company Information and Networking Technologies, Moscow, Cand.Sc., senior research assistant (olgasmnv@gmail.com).

Abstract: We consider a queuing model to estimate and optimize a modern architecture hybrid call center with a self-service facility based on computer speech recognition. This center is presented by a two-node network with threshold control of queuing. One of these nodes describes the self-service servers and the other one describes a group of operators. Both nodes are considered as multilinear queuing systems. When not served by the automated service, the customer goes to the group of operators. We obtain stationary probabilities of the system states and derive performance characteristics.

Keywords: hybrid call center, self-service, multilinear queuing systems, threshold control of queuing, open exponential network, speech recognition, multistage queuing system.

*Статья представлена к публикации
членом редакционной коллегии М. В. Губко*