

ДВУХФАЗНАЯ МОДЕЛЬ С НЕОГРАНИЧЕННЫМИ ОЧЕРЕДЯМИ ДЛЯ РАСЧЕТА ХАРАКТЕРИСТИК И ОПТИМИЗАЦИИ РЕЧЕВЫХ ПОРТАЛОВ САМООБСЛУЖИВАНИЯ

М.П. Фархадов, Н.В. Петухова, Д.В. Ефросинин, О.В. Семенова

Рассмотрена двухфазная модель с неограниченными очередями на каждой фазе, предназначенная для расчета и оптимизации телефонных речевых порталов самообслуживания на базе компьютерного распознавания речи. Определены стационарные распределения вероятностей состояний системы и вероятностно-временные характеристики обслуживания.

Ключевые слова: центр обслуживания вызовов, сервисы самообслуживания, открытая экспоненциальная сеть массового обслуживания, распознавание речи, многофазное обслуживание.

ВВЕДЕНИЕ

Компьютерные речевые технологии достигли в настоящее время достаточно высоких показателей качества, позволяющих начать их практическое применение. Речевые порталы с данными технологиями 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 может служить система «Автодиспетчер» для приема заявок на подачу такси [1], система «Автосекретарь» для диспетчеризации вызовов [2], и др. [3, 4].

В работе [5] дан обзор моделей для расчетов центров обслуживания вызовов классической архитектуры, а также приведена модель контакт-центра с сервисами самообслуживания с ограниченным числом мест для ожидания в первом узле и бесконечным числом мест ожидания во втором.

В данной работе исследуется открытая экспоненциальная многолинейная сеть массового обслуживания с бесконечным числом мест ожидания

в узлах, описывающая функционирование речевого портала современной архитектуры с сервисами самообслуживания на базе речевых технологий, архитектура которого приведена на рис. 1.

1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Рассмотрим открытую экспоненциальную сеть массового обслуживания с двумя узлами (рис. 2). Узел i сети представляет собой многолинейную систему массового обслуживания с n_i идентичными приборами, $i = 1, 2$. Число мест для ожидания в первом и втором узлах неограничено.

Поток заявок из внешнего источника является пуассоновским интенсивности λ и поступает в узел 1 или 2 в зависимости от числа свободных приборов в узле 2.

Если в узле 2 есть свободные приборы, то заявки из внешнего источника поступают на узел 2, в противном случае заявки поступают на узел 1. Если также в узле 2 освобождается прибор и нет заявок в очереди узла, то заявка, стоящая первой в очереди узла 1 (если очередь не пуста), переходит в узел 2 и занимает освободившийся прибор. Время обслуживания заявки в узле i экспоненциально распределено с параметром μ_i .

Предполагаем также, что с вероятностью $1 - p$ обслуживание заявки в первом узле успешно, и в момент завершения обслуживания эта заявка по-



Рис. 1. Архитектура центра обслуживания вызовов на основе современных технологий

кидает систему. В противном случае, с вероятностью p обслуживание неудачно (в ходе обработки запроса произошла ошибка, и запрос должен быть обработан оператором), и заявка переходит во второй узел.

Введем случайный процесс, описывающий состояние сети в момент времени t :

$$\{X(t)\}_{t \geq 0} = \{Q_1(t), Q_2(t)\}_{t \geq 0},$$

где $Q_i(t)$ — число заявок в i -м узле, $i = 1, 2$. Случайный процесс $\{X(t)\}_{t \geq 0}$ является однородным, марковским процессом. Условие его эргодичности будет приведено далее.

Рассмотрим стационарные вероятности

$$\pi(i, j) = \lim_{t \rightarrow \infty} P[X(t) = (i, j)], \quad i \geq 0, \quad j \geq 0.$$

Система уравнений равновесия (СУР) имеет следующий вид:

$$\lambda \pi(0, 0) = (1 - p) \mu_1 \pi(1, 0) + \mu_2 \pi(0, 1), \quad (1)$$

$$\begin{aligned} (\lambda + i \mu_1 + j \mu_2) \pi(i, j) = & \lambda \pi(i, j - 1) + (i + 1) \times \\ & \times (1 - p) \mu_1 \pi(i + 1, j) + (i + 1) p \mu_1 \pi(i + 1, j - 1) + \\ & + (j + 1) \mu_2 \pi(i, j + 1), \quad 0 \leq i < n_1, \quad 0 < j < n_2, \end{aligned} \quad (2)$$

$$(\lambda + i \mu_1) \pi(i, 0) = (i + 1)(1 - p) \mu_1 \pi(i + 1, 0) + \mu_2 \pi(i, 1), \quad 0 \leq i < n_1, \quad (3)$$

$$(\lambda + n_1 \mu_1) \pi(n_1, 0) = \mu_2 \pi(n_1, 1), \quad (4)$$

$$(\lambda + n_1 \mu_1 + j \mu_2) \pi(n_1, j) = \lambda \pi(n_1, j - 1) + (j + 1) \mu_2 \pi(n_1, j + 1), \quad 0 < j < n_2, \quad (5)$$

$$\begin{aligned} (\lambda + n_1 \mu_1 + n_2 \mu_2) \pi(n_1, n_2) = & \lambda \pi(n_1, n_2 - 1) + \\ & + \lambda \pi(n_1 - 1, n_2) + n_1(1 - p) \mu_1 \pi(n_1 + 1, n_2) + \\ & + n_2 \mu_2 \pi(n_1, n_2 + 1) + n_2 \mu_2 \pi(n_1 + 1, n_2), \end{aligned} \quad (6)$$

$$\begin{aligned} (\lambda + i \mu_1 + n_2 \mu_2) \pi(i, j) = & \lambda \pi(i - 1, j) I_{\{i > 0\}} + \\ & + \lambda \pi(i, j - 1) I_{\{j = n_2\}} + \min\{i + 1, n_1\} \mu_1 (1 - p) \times \\ & \times \pi(i + 1, j) + \min\{i + 1, n_1\} \mu_1 p \pi(i + 1, j - 1) + \\ & + n_2 \mu_2 \pi(i, j + 1) + (j + 1) \mu_2 \pi(i, j + 1), \\ & 0 \leq i \leq n_1, \quad j \geq n_2, \quad (i, j) \neq (n_1, n_2), \end{aligned} \quad (7)$$

$$\begin{aligned} (\lambda + n_1 \mu_1 + n_2 \mu_2) \pi(i, j) = & \lambda \pi(i - 1, j) I_{\{i > 0\}} + n_1(1 - p) \mu_1 \pi(i + 1, j) + \\ & + n_1 p \mu_1 \pi(i + 1, j - 1) I_{\{j > n_1\}} + n_2 \mu_2 \pi(i, j + 1) + \\ & + n_2 \mu_2 \pi(i + 1, n_2) I_{\{j = n_2\}}, \quad i > n_1, \quad j \geq n_2, \end{aligned} \quad (8)$$

где $I_{\{A\}}$ обозначает индикаторную функцию, принимающую значение 1, если выполнено условие A , и 0 — в противном случае.

Теорема. Если выполнено условие $\rho_1 = \frac{\lambda}{n_1 \mu_1} < 1$,

$\rho_2 = \frac{p \lambda}{n_2 \mu_2} < 1$, то процесс $\{X(t)\}_{t \geq 0}$ является эргодическим, а решение СУР представляется в форме

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

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

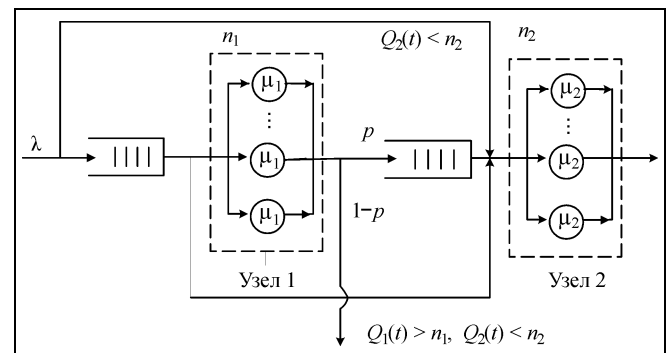


Рис. 2. Схема сети массового обслуживания с двумя узлами



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

$$i \geq n_1 + 1, \quad (9)$$

$$\pi(i, j) = \left[\pi_{n_2+1} G_{n_2} - (n_1(1-p)\mu_1 + n_2\mu_2) \hat{G}_{n_2} \pi \times \right. \\ \left. \times (n_1 + 1, n_2) \right] \prod_{k=1}^{n_2-j} G_{n_2-k} e_i, \\ 0 \leq i \leq n_2, \quad 0 \leq j \leq n_2, \quad (10)$$

где вектор-строка $\pi_{n_2+1} = (\pi(0, n_2 + 1), \pi(1, n_2 + 1), \dots, \pi(n_1, n_2 + 1))$.

Матрицы G_i определяются рекуррентными соотношениями

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

$$G_j = -Q_{j+1,j} \hat{G}_j, \quad 1 \leq j \leq n_2,$$

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

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

$$Q_{j-1,j} = \text{diag} \underbrace{\{\lambda, \dots, \lambda\}}_{n_1+1} = \text{diag} \{p\mu_1, p2\mu_1, \dots, pn_1\mu_1\}, \\ j = \overline{1, n_2},$$

$$Q_{j,j} = \text{diag}\{q_{0,j}, q_{1,j}, \dots, q_{n_1,j}\} + \\ + \text{diag} \{ (1-p)\mu_1, 2(1-p)\mu_1, p2\mu_1, \dots, n_1(1-p)\mu_1 \} + \\ + I_{\{j=n_2\}} \text{diag}^+ \{\lambda, \dots, \lambda\}, \quad j = \overline{0, n_2}, \\ Q_{j+1,j} = \\ = \text{diag} \underbrace{\{ \min\{j+1, n_2\}\mu_2, \dots, \min\{j+1, n_2\}\mu_2 \}}_{n_1+1}, \\ j = \overline{0, n_2},$$

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 — вектор-столбец, i -я координата которого (начиная с 0) равна 1.

Вероятность $\pi(0, n_2 + 1)$ вычисляется из условия нормировки

$$\sum_{i=0}^{n_1} \sum_{j=0}^{n_2} \pi(i, j) + \sum_{i=n_1+1}^{\infty} \pi(i, n_2) + \sum_{i=0}^{n_1-1} \sum_{j=n_2+1}^{\infty} \pi(i, j) + \\ + \sum_{i=n_1}^{\infty} \sum_{j=n_2+1}^{\infty} \pi(i, j) = 1,$$

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

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

Доказательство. Заметим, что если число заявок во втором узле будет больше $n_2 + 1$, система функционирует как классическая сеть Джексона, имеющая два узла с интенсивностями поступления $\lambda_1 = \lambda$ и $\lambda_2 = p\lambda$, поэтому предположим, что для состояний $x = (i, j)$, где $i \geq 0, j \geq n_2 + 1$, имеет место мультипликативное представление стационарных вероятностей

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

Далее преобразуем уравнения (7) и (8) СУР. Выражая из правых частей этих равенств вероятности $\pi(i, n_2)$, $i \geq n_1 + 1$, а остальные вероятности выражая через $\pi(0, n_2 + 1)$, используя полученные формулы, после элементарных преобразований получим выражение (9).

Оставшиеся вероятности состояний ниже границы (n_1, n_2) вычисляются с помощью матричного подхода. Введем в рассмотрение векторы стационарных вероятностей $\pi_j = (\pi(0, j), \pi(1, j), \dots, \pi(n_1, j))$, $0 \leq j \leq n_2$.

Преобразуя уравнения (1)–(7) СУР для $0 \leq i \leq n_1$ и $0 \leq j \leq n_2$ в векторно-матричную форму, получаем следующие уравнения для векторов π_j , $0 \leq j \leq n_2$:

$$\pi_0 Q_{0,0} + \pi_1 Q_{1,0} = 0, \\ \pi_{j-1} Q_{j-1,j} + \pi_j Q_{j,j} + \pi_{j+1} Q_{j+1,j} = 0, \\ 0 \leq j \leq n_2 - 1,$$

$$\pi_{n_2-1} Q_{n_2-1, n_2} + \pi_{n_2} Q_{n_2, n_2} + \pi_{n_2+1} Q_{n_2+1, n_2} + (n_1(1-p)\mu_1 + n_2\mu_2)e_{n_1}^t \pi(n_1+1, n_2) = 0,$$

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

$$\pi_j = \pi_{j+1} G_j, \quad 0 \leq j \leq n_2 - 1, \\ \pi_{n_2} = \pi_{n_2+1} G_{n_2} - (n_1(1-p)\mu_1 + n_2\mu_2) \times e_{n_1}^t \hat{G}_{n_2} \pi(n_1+1, n_2).$$

Из последней системы получаем

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

Отсюда непосредственно следует формула (10).

Таким образом, все вероятности системы представляются в виде функции, зависящей от вероятности $\pi(0, n_2+1)$, которая, в свою очередь, вычисляется из условия нормировки (11). Легко показать, что для сходимости входящих в выражение бесконечных сумм необходимо, чтобы величины ρ_1 и ρ_2 были меньше 1. Дальнейшая подстановка полученных выражений в СУР подтверждает справедливость сделанных предположений относительно мультипликативной формы (11), что завершает доказательство теоремы. ♦

2. ХАРАКТЕРИСТИКИ ПРОИЗВОДИТЕЛЬНОСТИ СИСТЕМЫ

Вычислив стационарное распределение вероятностей состояний системы, далее можно найти следующие характеристики производительности.

- Среднее число занятых приборов в первом и втором узлах

$$N_1 = \sum_{i=0}^{n_1} \sum_{j=0}^{\infty} \min\{i, n_1\} \pi(i, j),$$

$$N_2 = \sum_{i=0}^{n_1} \sum_{j=0}^{\infty} \min\{j, n_2\} \pi(i, j).$$

- Среднее число заявок в системе

$$L = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (i+j) \pi(i, j).$$

- Вероятности того, что заявка, поступив в систему, попадает в первый узел и второй узел, соответственно,

$$\tau_1 = \sum_{i=0}^{\infty} \sum_{j=n_2}^{\infty} \pi(i, j), \quad \tau_2 = 1 - \tau_1.$$

- Среднее время ожидания в системе может быть получено по формуле Литтла $W = L/\lambda$.
- Среднее время ожидания заявки во втором узле

$$W_2 = \frac{\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \pi(i, j)}{\lambda(1 - \tau_1(1-p))}.$$

3. ЗАДАЧА ОПТИМИЗАЦИИ

Введем следующую структуру штрафов за функционирование системы:

$c_{0,k}$ — стоимость в единицу времени ожидания в очереди узла k ;

$c_{u,k}$ — стоимость в единицу времени использования прибора в узле k ;

$c_{e,k}$ — стоимость в единицу времени простоя прибора в узле k ;

c_f — фиксированная стоимость включения системы мониторинга состояний.

Задача состоит в минимизации функционала потерь

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

который в данном случае имеет вид

$$\bar{V}(n_1, n_2) = c_{0,1} \bar{Q}_1 + c_{0,2} \bar{Q}_2 + c_{u,1} \bar{C}_1 + c_{u,2} \bar{C}_2 + c_{e,1} \bar{Z}_1 + c_{e,2} \bar{Z}_2 + c_f/B,$$

где $\bar{Q}_1 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} j p(i+n_1, j)$ и $\bar{Q}_2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} j \pi(i, j+n_2)$ — среднее число заявок в очереди k -го узла,

$\bar{C}_1 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \min\{i, n_1\} \pi(i, j)$ и $\bar{C}_2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \min\{i, n_2\} \pi(i, j)$ — среднее число занятых приборов в k -м узле, $\bar{Z}_1 = \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} (n_1-i) \pi(i, j)$ и $\bar{Z}_2 = \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} (n_2-i) \pi(i, j)$ — среднее число свободных приборов в k -м узле, $\bar{B} = 1/\lambda \pi(0, 0)$ — средняя длительность цикла между соседними посещениями состояния $(0, 0)$.

В качестве примера рассмотрим систему со следующим набором параметров: $(\lambda, \mu_1, \mu_2, p,$

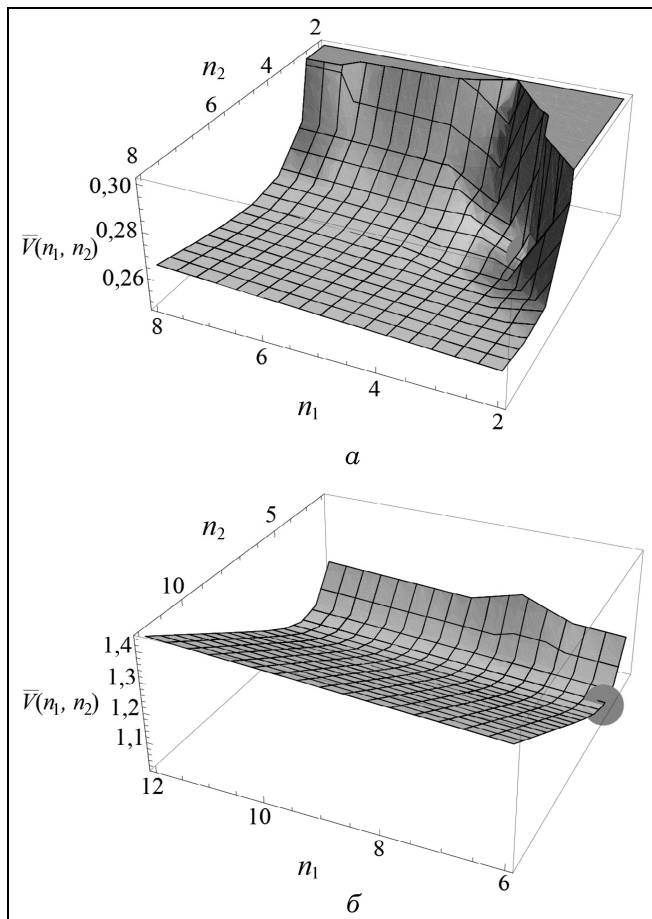


Рис. 3. Значение функции $\bar{V}(n_1, n_2)$ в зависимости от числа приборов

$c_{0,1}, c_{0,2}, c_{u,1}, c_{u,2}, c_{e,1}, c_{e,2}, c_p = (0,9; 0,5; 0,7; 0,001; 5,5; 2,5; 0,001; 0,005; 0,001; 0,005; 0,9)$. В этом случае оптимальное число приборов $(n_1^*, n_2^*) = (3, 4)$, и оптимальное значение функционала потерь $\bar{V}(n_1^*, n_2^*) = 0,247$ (рис. 3, а).

Для случая $(\lambda, \mu_1, \mu_2, p, c_{0,1}, c_{0,2}, c_{u,1}, c_{u,2}, c_{e,1}, c_{e,2}, c_p) = (2,5; 0,5; 0,7; 0,001; 5,5; 2,5; 0,01; 0,05; 0,01; 0,05; 2,9)$ оптимальные значения определяются как $(n_1^*, n_2^*) = (6, 4)$, $\bar{V}(n_1^*, n_2^*) = 1,011$ (рис. 3, б). Зависимость функции $\bar{V}(n_1, n_2)$ от числа приборов показана на рис. 3.

ЗАКЛЮЧЕНИЕ

Предложены математические модели, позволяющие исследовать характеристики телефонных систем массового обслуживания с комбинацией традиционных методов обслуживания и инновационного обслуживания: сервисов самообслужи-

вания на основе компьютерных речевых технологий и современных интерактивных средств взаимодействия. Предложенные модели учитывают специфику, вносимую новейшими технологиями. Исследована открытая экспоненциальная многолинейная сеть массового обслуживания с бесконечным числом мест ожидания в узлах. Определены стационарные распределения вероятностей состояний системы, найдены основные характеристики производительности системы.

Авторы надеются, что полученные результаты будут полезны специалистам для проектирования, исследования и организации высокоэффективной работы центров обслуживания и обработки вызовов современной архитектуры на основе новых телекоммуникационных и речевых технологий.

ЛИТЕРАТУРА

1. *Современные технологии управления в диспетчерской службе такси* / В.А. Жожикашвили, Н.В. Петухова, А.Н. Зацепин, В.В. Азаров // Проблемы управления. — 2006. — № 2. — С. 32—34.
2. *Интеллектуальные телефонные услуги на основе речевых технологий* / В.А. Жожикашвили, Р.В. Билик, В.А. Вертиб и др. // Информационно-измерительные и управляющие системы. — 2007. — № 2. — С. 75—78.
3. *Жожикашвили В.А., Петухова Н.В., Фархадов М.П.* Компьютерные системы массового обслуживания и речевые технологии // Проблемы управления. — 2006. — № 2. — С. 3—7.
4. *Жожикашвили В.А., Петухова Н.В., Фархадов М.П.* Мультисерверная архитектура интеллектуальных порталов самообслуживания // IV Междунар. конф. по проблемам управления (МКПУ-IV). — М., 2009. — С. 1744—1748.
5. *Математическая модель центра обслуживания вызовов с сервисами самообслуживания* / М.П. Фархадов, Н.В. Петухова, Д.В. Ефросинин, О.В. Семенова // Proc. of International Workshop Distributed Computer and Communication Networks DCCN'2009, Sofia, Bulgaria. — М., 2009. — P. 86—95.

Статья представлена к публикации членом редколлегии А.С. Манделем.

Фархадов Маис Паша оглы — зав. лабораторией, Институт проблем управления РАН, г. Москва, ☎ (495) 334-87-20, ✉ mais@ipu.ru,

Петухова Нина Васильевна — ст. науч. сотрудник, Институт проблем управления РАН, г. Москва, ☎ (495) 334-90-60, ✉ nvp@ipu.ru,

Ефросинин Дмитрий Владимирович — доцент, Российский университет дружбы народов, г. Москва, ☎ (495) 434-53-00, ✉ dmitriy_e@mail.ru,

Семенова Ольга Валерьевна — ст. науч. сотрудник, ЗАО Научно-производственное объединение «Информационные и сетевые технологии», г. Москва, ☎ (495) 515-57-92, ✉ olgasmnv@gmail.com.