УДК 519.872

О.С. ТАРАМИН

ДВУХФАЗНАЯ СИСТЕМА ОБСЛУЖИВАНИЯ С ГРУППОВЫМ МАРКОВСКИМ ВХОДЯЩИМ ПОТОКОМ

Two-phase queuing system with the batch markovian arrival process and multi-server second stage is investigated. The condition for the stable operation of the system is derived, the stationary distribution and performance measures are calculated.

Тандемные системы обслуживания широко используются при проектировании, анализе функционирования и оптимизации информационно-телекоммуникационных сетей, а также для моделирования реально существующих двухузловых сетей и проверки алгоритмов декомпозиции сетей общего вида [1].

Характерная особенность современных сетей – то, что потоки в них нестационарные, групповые и коррелированные. В настоящее время наиболее удачной математической моделью таких потоков является *BMAP* (Batch Markovian Arrival Process) (см., например, [2]). Эта модель, с одной стороны, хорошо описывает реальные потоки, с другой – позволяет получать аналитические результаты в наиболее простом и изящном виде.

Тандемные системы массового обслуживания с *BMAP*-потоком и однолинейными фазами рассматривались ранее в [3–5]. В данной статье анализируется двухфазная система обслуживания $BMAP/G/1 \to \cdot/M/N/R$, отличительная особенность которой – наличие нескольких приборов на второй фазе. Кроме того, в ней предусмотрены как потери запросов после обслуживания на первой фазе, так и блокировки обслуживающего прибора первой фазы этим запросом в случае отсутствия свободных мест в промежуточном буфере.

Математическая модель

Рассматривается система $BMAP/G/1 \to \cdot/M/N/R$. В нее поступает *BMAP*-поток запросов, описываемый управляющим процессом $\nu_t,\ t \ge 0,$ который является неприводимой цепью Маркова с непрерывным временем и пространством состояний $\{0,...,W\}$ и матричной производящей функцией $D(z) = \sum_{k=0}^{\infty} D_k z^k,\ |z| \le 1.$ Поступление групп запросов возможно только в моменты скачков управляющего процесса. Интенсивности переходов данного процесса, сопровождающиеся генерацией $k,\ k \ge 0,$ запросов, определяются элементами матриц $D_k,\ k \ge 1,$ и недиагональными элементами матрицы D_0. Матрица $D(1)$ является инфинитезимальным генератором управляющего процесса. Вектор стационарных вероятностей $\boldsymbol{\theta}$ этого процесса удовлетворяет системе линейных алгебраических уравнений $\boldsymbol{\theta} D(1) = \mathbf{0},\ \boldsymbol{\theta}\mathbf{e} = 1$ (здесь и далее $\mathbf{e}$ – вектор-столбец, состоящий из единиц). Средняя интенсивность λ поступления запросов в *BMAP*-потоке задается формулой $\lambda = \boldsymbol{\theta} D'(z)|_{z=1}\,\mathbf{e}.$ Вероятности поступления i запросов за время t в этом потоке описываются матрицами $P(i,t),$ которые являются коэффициентами разложения $e^{D(z)t} = \sum_{i=0}^{\infty} P(i,t) z^i$; (ν,ν')-й элемент матрицы $P(i,t)$ есть вероятность того, что в *BMAP*-потоке на интервале времени $(0,t]$ поступит i запросов и состояние управляющего процесса в момент времени t окажется равным ν' при условии, что $\nu_0 = \nu$. Средняя интенсивность поступления групп λ_b определяется как $\lambda_b = \boldsymbol{\theta}(-D_0)\mathbf{e}.$ Коэффициент вариации c_{var} длин интервалов между моментами поступления последовательных групп запросов задается формулой $c_{\text{var}}^2 = 2\lambda_b \boldsymbol{\theta}(-D_0)^{-1}\mathbf{e} - 1.$ Коэффициент корреляции c_{cor} длин двух соседних интервалов вычисляется по формуле $c_{\text{cor}} = (\lambda_b \boldsymbol{\theta}(-D_0)^{-1}(D(1) - D_0)(-D_0)^{-1}\mathbf{e} - 1))/c_{\text{var}}^2.$ Более подробную информацию о *BMAP*-потоке и его свойствах можно найти в [2].

Время обслуживания запросов на первой фазе характеризуется функцией распределения $B(t)$ с конечным первым моментом $b_1 = \int_0^{\infty} t dB(t).$ После обслуживания на первой фазе запрос поступает на вторую фазу рассматриваемой системы, которую образуют N независимых идентичных приборов. Между двумя фазами имеется буфер размера $R < \infty.$ В случае занятости всех приборов второй фазы запрос ожидает обслуживания в буфере. Предполагается, что при отсутствии свободного места в буфере в момент окончания обслуживания на первой фазе с вероятностью $\gamma,\ \ 0 \le \gamma \le 1,$ запрос уходит из системы недообслуженным (теряется), а с вероятностью $1-\gamma$ прибор первой фазы блокируется и не обслуживает следующий запрос, пока не освободится место в буфере. Время обслуживания запроса прибором второй фазы имеет показательное распределение с параметром μ.

Стационарное распределение вероятностей состояний системы

Пусть t_n – n-й момент окончания обслуживания на первой фазе. Рассмотрим процесс $\xi_n=\{i_n,r_n,\nu_n\}$, $n\ge1$, где i_n – число запросов на первой фазе в момент времени t_n+0, $i_n\ge0$; r_n – число запросов на второй фазе в момент времени t_n-0, $r_n=\overline{0,N+R}$; ν_n – состояние управляющего процесса ν_t в момент времени t_n, $\nu_n=\overline{0,W}$.

Процесс $\xi_n=\{i_n,r_n,\nu_n\}$, $n\ge1$, является неприводимой цепью Маркова. Перенумеруем состояния цепи $\xi_n, n\ge1$, в лексикографическом порядке и сформируем $(W+1)(N+R+1)\times(W+1)(N+R+1)$-матрицы $P_{l,m},\, l,m\ge0$, вероятностей переходов цепи $\xi_n, n\ge1$, из состояний со значением l первой компоненты в состояния со значением m этой компоненты.

Лемма 1. Матрицы одношаговых переходных вероятностей $P_{l,m}$ имеют вид

$$P_{l,m}=O,\ m<l-1,\ P_{l,l+i-1}=Y_i,\ i>0,\ l\ge0,\ P_{0,i}=V_i,\ i\ge0,$$

где матрицы $Y_i, V_i, i\ge0$, имеют блочную структуру $V_i=(V_i^{(r,r')})_{r,r'=\overline{0,N+R}}$, $Y_i=(Y_i^{(r,r')})_{r,r'=\overline{0,N+R}}$, блоки $V_i^{(r,r')}$, $Y_i^{(r,r')}$ соответствуют переходам процесса $r_n, n\ge1$, из состояния r в состояние r' и вычисляются следующим образом:

$$V_i^{(r,r')}=\sum_{l=r'}^{r+1}\Delta_{r+1,l}\sum_{k=1}^{i+1}D_k\Omega_{i-k+1}^{(l,r')},\ 0\le r'\le N+R-1,\ 0\le r\le N+R,$$

$$V_i^{(N+R,r')}=\gamma\sum_{l=r'}^{N+R}\Delta_{N+R,l}\sum_{k=1}^{i+1}D_k\Omega_{i-k+1}^{(l,r')}+(1-\gamma)\{N\mu(-D_0+N\mu I)^{-1}\times$$

$$\times\sum_{l=r'}^{N+R}\Delta_{N+R,l}\sum_{k=1}^{i}D_k\Omega_{i-k+1}^{(l,r')}+\sum_{k=1}^{i}\int_0^{\infty}P(k,t)N\mu e^{-N\mu}dt\Omega_{i-k+1}^{(N+R,r')}\},\ 0\le r'\le N+R,$$

$$Y_i^{(r,r')}=\Omega_i^{(r+1,r')},\ 0\le r'\le N+R-1,\ 0\le r\le N+R,$$

$$Y_i^{(N+R,r')}=\gamma\Omega_i^{(N+R,r')}+(1-\gamma)\sum_{k=0}^{i}\int_0^{\infty}P(k,t)N\mu e^{-N\mu}dt\Omega_{i-k+1}^{(N+R,r')},\ 0\le r'\le N+R,$$

где $\Omega_i^{(r,r')}=\int_0^{\infty}P(i,t)\delta_{r,r'}(t)dB(t)$, $\Delta_{r,r'}=\int_0^{\infty}e^{D_0t}\delta_{r,r'}(t)dt$, $r,r'=\overline{0,N+R}$,

$$\delta_{r,r'}(t)=\begin{cases}0,\ r<r';\\ C_r^{r'}e^{-\mu r't}(1-e^{-\mu t})^{r-r'},\ 0\le r'\le r\le N;\\ \int_0^t N\mu\dfrac{(N\mu\tau)^{r-N-1}}{(r-N-1)!}e^{-N\mu\tau}C_N^{r'}e^{-\mu r'(t-\tau)}(1-e^{-\mu(t-\tau)})^{N-r'},\ 0\le r'<N<r\le N+R;\\ \dfrac{(N\mu t)^{r-r'}}{(r-r')!}e^{-N\mu t},\ N<r\le N+R,\ N\le r'\le N+R.\end{cases}$$

Доказательство. Доказательство леммы основывается на анализе поведения системы в моменты окончания обслуживания на первой фазе и на вероятностном смысле функций и матриц. Здесь $\delta_{r,r'}(t)$ – вероятность того, что за время t число запросов на второй фазе изменится с r на r' при условии, что новые запросы на эту фазу не поступают; (ν,ν')-й элемент матрицы $\Omega_i^{(r,r')}=\int_0^{\infty}P(i,t)\delta_{r,r'}(t)dB(t)$ есть вероятность того, что за время обслуживания на первой фазе в систему поступит i запросов, управляющий процесс ν_t перейдет из состояния ν в состояние ν' и число запросов на второй фазе изменится с r на r'; (ν,ν')-й элемент матрицы $\Delta_{r,r'}D_k=\int_0^{\infty}e^{D_0t}\delta_{r,r'}(t)D_kdt$ есть вероятность того, что при условии, что процесс ν_t находится в состоянии ν и число запросов на второй фазе равно r, первой в $BMAP$-потоке поступит группа размера k, процесс ν_t окажется в состоянии ν' непосредственно после момента поступления группы и число запросов на второй фазе равно r' в этот момент; (ν, ν')-й элемент матрицы $\int_0^\infty P(i,t)N\mu e^{-N\mu}dt$ означает вероятность того, что за время блокировки прибора первой фазы в систему поступит i запросов и процесс ν_t перейдет из состояния ν в состояние ν'. □

Следствие 1. Процесс ξ_n, $n \ge 1$, является квазитеплицевой цепью Маркова.

Доказательство. Матрицы $P_{l,m}$ при $l > 0$ являются функциями разности $m - l$. Следовательно, согласно [6] рассматриваемый процесс ξ_n, $n \ge 1$, принадлежит классу многомерных квазитеплицевых цепей Маркова. □

Пусть $V(z) = \sum_{i=0}^{\infty} V_i z^i$, $Y(z) = \sum_{i=0}^{\infty} Y_i z^i$, $|z| \le 1$, – производящие функции матриц V_i и Y_i, $i \ge 0$. Представим эти матрицы в блочном виде $V(z) = (V_{r,r'}(z))_{r,r'=\overline{0,N+R}}$, $Y(z) = (Y_{r,r'}(z))_{r,r'=\overline{0,N+R}}$.

Следствие 2. Матрицы $V_{r,r'}(z)$, $Y_{r,r'}(z)$ имеют вид

$$V_{r,r'}(z) = \sum_{l=r'}^{r+1} \Delta_{r+1,l} \frac{1}{z}(D(z) - D_0)\Omega_{l,r'}(z),\ 0 \le r' \le N + R - 1,\ 0 \le r \le N + R,$$

$$V_{N+R,r'}(z) = \gamma \sum_{l=r'}^{N+R} \Delta_{N+R,l} \frac{1}{z}(D(z) - D_0)\Omega_{l,r'}(z) +$$

$$+(1-\gamma)\{N\mu(-D_0 + N\mu I)^{-1} \sum_{l=r'}^{N+R} \Delta_{N+R,l} \frac{1}{z}(D(z) - D_0)\Omega_{l,r'}(z) +$$

$$+\frac{1}{z} N\mu((-D(z) + N\mu I)^{-1} - (-D_0 + N\mu I)^{-1})\Omega_{N+R,r'}(z)\},\ 0 \le r' \le N + R,$$

$$Y_{r,r'}(z) = \Omega_{r+1,r'}(z),\ 0 \le r' \le N + R - 1,\ 0 \le r \le N + R,$$

$$Y_{N+R,r'}(z) = \{\gamma + (1-\gamma)N\mu(-D(z) + N\mu I)^{-1}\}\Omega_{N+R,r'}(z),\ 0 \le r' \le N + R,$$

где $\Omega_{r,r'}(z) = \int_0^\infty e^{D(z)t}\delta_{r,r'}(t)dB(t)$.

Теорема 1. Необходимым и достаточным условием эргодичности цепи Маркова ξ_n, $n \ge 1$, является выполнение неравенства

$$\rho = \lambda(b_1 + (1-\gamma)y_{N+R}(N\mu)^{-1}) < 1. \tag{1}$$

Здесь y_{N+R} – последняя координата вектора $\mathbf{y} = (y_0, ..., y_{N+R})$, который удовлетворяет системе линейных алгебраических уравнений

$$\mathbf{y}\tilde{Y}(1) = \mathbf{y},\ \mathbf{ye} = 1, \tag{2}$$

где $\tilde{Y}(1) = \left(\int_0^\infty \delta_{\min\{r+1,N+R\},r'}(t)dB(t) \right)_{r,r'=\overline{0,N+R}}$.

Доказательство. Матрица $Y(1)$ описывает одношаговые переходы конечных компонент (r_n, ν_n) цепи Маркова $\xi_n = \{i_n, r_n, \nu_n\}$ при условии, что счетная компонента $i_n > 0$. Эта матрица является неприводимой. В этом случае из теории многомерных квазитеплицевых цепей Маркова (см. [6]) следует, что необходимым и достаточным условием эргодичности цепи Маркова ξ_n, $n \ge 1$, является выполнение неравенства

$$\mathbf{x}Y'(1)\mathbf{e} < 1, \tag{3}$$

где вектор $\mathbf{x}$ – единственное решение системы линейных алгебраических уравнений

$$\mathbf{x}Y(1) = \mathbf{x},\ \mathbf{xe} = 1. \tag{4}$$

Представим вектор $\mathbf{x}$ в следующем виде:

$$\mathbf{x} = \mathbf{y} \otimes \boldsymbol{\theta}, \tag{5}$$

где $\mathbf{y}$ – инвариантный вектор неприводимой стохастической матрицы $\tilde{Y}(1)$, т. е. является единственным решением системы (2).

Принимая во внимание соотношение $\boldsymbol{\theta}e^{D(1)t} = \boldsymbol{\theta}$, прямой подстановкой можно проверить, что вектор $\mathbf{x}$ удовлетворяет системе (4). Подставив вектор $\mathbf{x}$ в виде (5) в неравенство (3) и используя соотношение $\boldsymbol{\theta}\sum_{r'=0}^{N+R}\Omega'_{r,r'}(z)|_{z=1}\,\mathbf{e} = \boldsymbol{\theta}D'(1)\mathbf{e}\int_0^\infty t dB(t) = \lambda b_1$, после некоторых алгебраических преобразований приводим неравенство (3) к виду (1). □

Далее будем считать, что неравенство (1) выполняется.

Пусть $\pi(i, r, \nu)$, $i \geq 0$, $r = \overline{0, N+R}$, $\nu = \overline{0, W}$, – стационарное распределение цепи Маркова $\xi_n = \{i_n, r_n, \nu_n\}$, $n \geq 1$. Сформируем вектор-строки стационарных вероятностей $\boldsymbol{\pi}(i, r) = (\pi(i, r, 0), \pi(i, r, 1), \ldots, \pi(i, r, W))$, $\boldsymbol{\pi}_i = (\boldsymbol{\pi}(i, 0), \boldsymbol{\pi}(i, 1), \ldots, \boldsymbol{\pi}(i, N+R))$, $i \geq 0$.

Для вычисления векторов $\boldsymbol{\pi}_i$, $i \geq 0$, используется алгоритм [6], разработанный для вычисления стационарного распределения многомерных квазитеплицевых цепей.

Вычислив стационарное распределение $\boldsymbol{\pi}_i$, $i \geq 0$, можно найти ряд стационарных характеристик производительности системы. Приведем некоторые из них.

1. Математическое ожидание и дисперсия среднего числа запросов на первой фазе в момент окончания обслуживания на этой фазе: $L_1 = \sum_{i=0}^{\infty} i\boldsymbol{\pi}_i\mathbf{e}$, $D^{(1)} = \sum_{i=0}^{\infty} i^2\boldsymbol{\pi}_i\mathbf{e} - L_1^2$.

2. Математическое ожидание и дисперсия среднего числа запросов на второй фазе в момент окончания обслуживания на первой фазе: $L_2 = \sum_{j=1}^{N+R} j \sum_{i=0}^{\infty}\boldsymbol{\pi}_i(\hat{\mathbf{e}}_j \otimes \mathbf{e}_{W+1})$, $D^{(2)} = \sum_{j=1}^{N+R} j^2 \sum_{i=0}^{\infty}\boldsymbol{\pi}_i(\hat{\mathbf{e}}_j \otimes \mathbf{e}_{W+1}) - L_2^2$, где $\hat{\mathbf{e}}_j$ – вектор-столбец, $(j+1)$-я координата которого равна единице, а остальные – нулю, $j = \overline{1, N+R}$.

3. Вероятность того, что произвольный запрос покинет систему после обслуживания на первой фазе $P_{\text{loss}} = \gamma\sum_{i=0}^{\infty}\boldsymbol{\pi}_i(\tilde{\mathbf{e}} \otimes \mathbf{e}_{W+1})$, где $\tilde{\mathbf{e}} = (0, \ldots, 0, 1)^T$.

4. Вероятность того, что произвольный запрос вызовет блокировку прибора первой фазы $P_{\text{block}} = (1-\gamma)\sum_{i=0}^{\infty}\boldsymbol{\pi}_i(\tilde{\mathbf{e}} \otimes \mathbf{e}_{W+1})$.

Численный эксперимент

В численном эксперименте исследуется влияние коэффициента корреляции входного потока на стационарные характеристики производительности системы.

Рассматриваются три *BMAP*-потока, у которых максимальный размер групп $k = 5$. $BMAP_1$ – это групповой пуассоновский поток, для которого $c_{\text{cor}} = 0$ и $c_{\text{var}} = 1$.

$BMAP_2$ и $BMAP_3$ имеют коэффициент вариации $c_{\text{var}} = 12,2732$ и коэффициенты корреляции $c_{\text{cor}} = 0,1$ и $c_{\text{cor}} = 0,2$ соответственно.

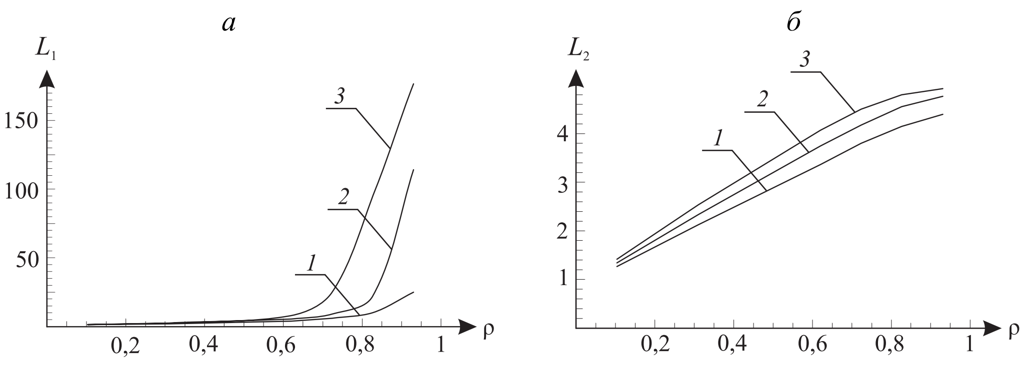

Рис. 1. Зависимости среднего числа запросов на первой (*а*) и второй (*б*) фазах от загрузки системы для различных *BMAP*-потоков. Здесь и на рис. 2, 3: *1* – $c_{\text{cor}} = 0$, *2* – $c_{\text{cor}} = 0,1$, *3* – $c_{\text{cor}} = 0,2$

Время обслуживания на 1-й фазе постоянное и равно $b_1 = 0{,}1$. Число приборов на второй фазе $N = 5$, размер буфера $R = 3$, вероятность блокировки после обслуживания на первой фазе $\gamma = 0{,}6$, интенсивность обслуживания на второй фазе $\mu = 2$.

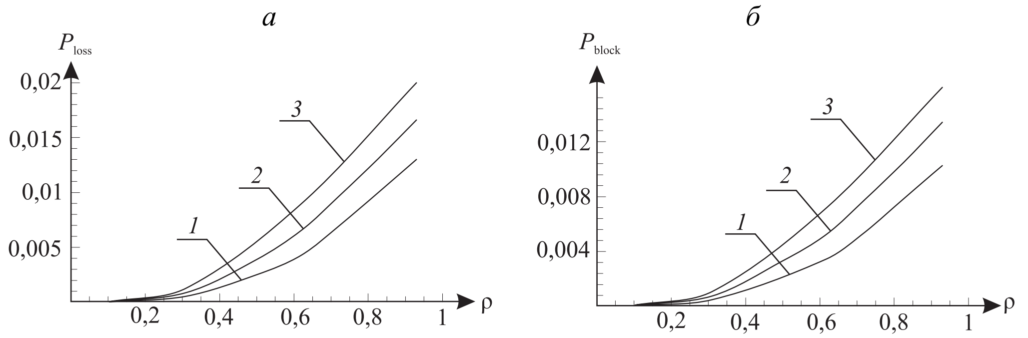

Рис. 2. Зависимости вероятностей отказа (*а*) и блокировки (*б*) от загрузки системы для различных *BMAP*-потоков

На рис. 1 представлены графики зависимостей среднего числа запросов на первой L_1 и на второй фазе L_2 от коэффициента загрузки системы ρ. Изменение величины ρ происходит за счет изменения средней интенсивности λ, которая, в свою очередь, изменяется путем нормирования элементов матриц D_k, $k = \overline{0,5}$. Зависимости вероятностей P_{loss} и P_{block} от ρ иллюстрирует рис. 2, а рис. 3 – зависимости дисперсий числа запросов на первой $D^{(1)}$ и на второй фазе $D^{(2)}$ от коэффициента загрузки ρ.

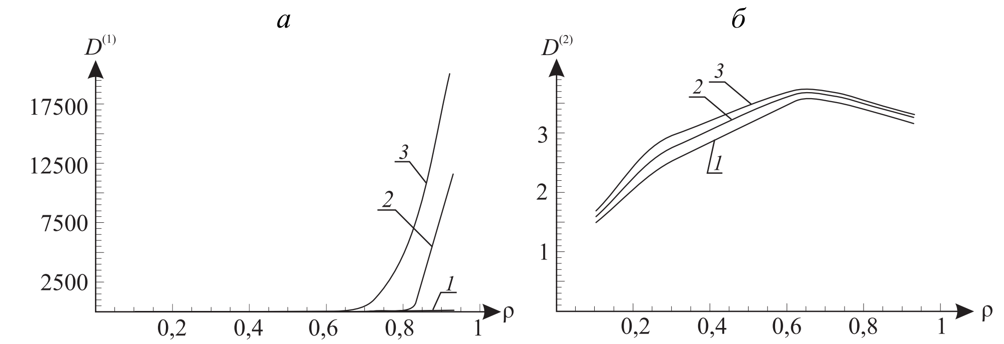

Рис. 3. Зависимости дисперсий числа запросов на первой (*а*) и второй (*б*) фазах от загрузки системы для различных *BMAP*-потоков

Из рисунков видно, что при одной и той же загрузке системы ρ значения величин L_1, L_2, P_{loss}, P_{block}, $D^{(1)}$ и $D^{(2)}$ возрастают с увеличением коэффициента корреляции. Из этого следует, что коэффициент корреляции является важной характеристикой входного потока, которую нельзя игнорировать при оценке производительности системы. Очевидно также, что при росте ρ дисперсия $D^{(2)}$ возрастает до некоторого значения, а затем начинает убывать. Это объясняется невозможностью дальнейшего роста дисперсии при увеличении числа запросов на второй фазе вследствие ограниченного числа мест на этой фазе.

1. Heindl A. // Performance Evaluation. 2003. Vol. 51. P. 117.
2. Lucantoni D.M. // Comm. Stat.-Stochastic Models. 1991. 7. P. 1.
3. Klimenok V.I., Breuer L., Tsarenkov G.V., Dudin A.N. // Performance Evaluation. 2005. Vol. 61. P. 17.
4. Gomez-Corral A., Martos M.E. // Performance Evaluation. 2006. 63. P. 910.
5. Klimenok V., Kim C.S., Tsarenkov G.V. et al. // Performance Evaluation. 2007. 64. P. 802.
6. Klimenok V.I., Dudin A.N. // Queueing Systems. 2006. Vol. 54. P. 245.

Поступила в редакцию 22.05.08.

Ольга Сергеевна Тарамин – аспирант кафедры теории вероятностей и математической статистики. Научный руководитель – доктор физико-математических наук, доцент кафедры теории вероятностей и математической статистики В.И. Клименок.