Теория массового обслуживания, сети телекоммуникаций и математическое моделирование

УДК 519.872

Система $MAP/G/1/\infty$ с инверсионным порядком обслуживания и вероятностным приоритетом, функционирующая в дискретном времени

А. В. Печинкин, И. В. Стальченко

Кафедра теории вероятностей и математической статистики
Российский университет дружбы народов
ул. Миклухо–Маклая, 6, Москва, Россия 117198

Рассматривается функционирующая в дискретном времени система массового обслуживания $MAP/G/1/\infty$ с дисциплиной обслуживания, при которой поступающая в систему заявка с некоторой вероятностью, зависящей только от её длины и (остаточной) длины заявки на приборе, становится на прибор, вытесняя обслуживавшуюся ранее на первое место в очередь, или с дополнительной вероятностью сама занимает первое место в очереди (инверсионный порядок обслуживания с вероятностным приоритетом). Для этой системы найдены основные стационарные характеристики функционирования. Приведены примеры расчётов, проведённых с помощью полученных аналитических соотношений.

Ключевые слова: система массового обслуживания, дискретное время, инверсионный порядок обслуживания с вероятностным приоритетом.

1. Введение

Для обработки информации в современных инфотелекоммуникационных системах широко используются различные варианты дисциплин разделения времени, что позволяет существенно улучшить качество их функционирования. Отметим, что во многих случаях практическая реализация таких дисциплин была обоснована расчётами, проведёнными с помощью математических моделей теории массового обслуживания.

На основе анализа систем массового обслуживания (СМО) в ряде случаев удалось определить и границы, до которых возможно улучшение качества обслуживания. Например, в довольно широком классе дисциплин оптимальной в смысле минимальной длины очереди является дисциплина преимущественного обслуживания заявки минимальной остаточной длины, или SRPT (*shortest remaining processor time*) [1,2]. Однако как практическое применение этой дисциплины, так и алгоритмы расчёта показателей функционирования даже СМО $M/G/1/\infty$ с дисциплиной SRPT весьма сложны [3–6], что сильно затрудняет её использование как в практическом, так и в теоретическом плане.

В работах [7–9] был введён сначала для частного случая, а потом и в общем виде инверсионный порядок обслуживания с вероятностным приоритетом (дисциплина LIFO PP) и найдены для СМО $M/G/1/\infty$ с этой дисциплиной основные стационарные показатели функционирования. Как оказалось, дисциплина LIFO PP обладает рядом интересных особенностей, что повлекло за собой появление большого числа работ по исследованию СМО с такой дисциплиной (см., например, [10–14]).

Статья поступила в редакцию 15 декабря 2009 г.
Работа выполнена при поддержке РФФИ (гранты N 08-07-00152 и N 09-07-12032).

В последнее время в связи с применением в инфотелекоммуникационных системах современных технологий оживился интерес к исследованию СМО, функционирующих в дискретном времени (см., например, [15–21]. Отметим, что во многих случаях анализ СМО в дискретном времени усложняется, поскольку, в отличие от непрерывного времени, в дискретном времени может одновременно произойти несколько изменений состояний. Для СМО в дискретном времени $Geo/G/1/\infty$ с геометрическим входящим потоком и дисциплиной обслуживания, введённой в [7], в [22] были найдены основные стационарные характеристики функционирования, а в [23] была рассмотрена СМО в дискретном времени $Geo/G/1/\infty$ с общим вариантом дисциплины LIFO PP, для которой также были найдены основные стационарные характеристики функционирования. В настоящей работе рассматривается обобщение СМО $Geo/G/1/\infty$ из [23] — функционирующая в дискретном времени система $MAP/G/1/\infty$ с марковским входящим потоком и дисциплиной LIFO PP.

2. Описание системы

Рассмотрим функционирующую в дискретном времени систему массового обслуживания $MAP/G/1/\infty$ с инверсионным порядком обслуживания и вероятностным приоритетом.

Входящий поток представляет собой (дискретный) марковский поток заявок (MAP — *Markov arrival process*) с числом состояний I, определяемый матрицами Λ и N, причём элементы λ_{ij}, $i,j=\overline{1,I}$, матрицы Λ задают вероятности таких изменений фаз генерации заявок, при которых заявки не поступают, а элементы n_{ij}, $i,j=\overline{1,I}$, матрицы N — вероятности изменений фаз генерации заявок, сопровождающихся поступлением заявок.

Обозначим через $\Lambda^*=\Lambda+\mathrm{N}$ матрицу вероятностей (общих) изменений фаз генерации заявок. Будем предполагать, что Λ^* — неприводимая матрица, N — ненулевая матрица. Это гарантирует невырожденность матрицы Λ. Кроме того, будем считать, что матрица Λ^* является матрицей вероятностей переходов непериодической цепи Маркова. Тогда эта цепь Маркова является эргодической и вектор-строка $\vec{\pi}_{\mathrm{a}}$ её предельных (стационарных) вероятностей состояний может быть найден из системы уравнений равновесия (СУР)

$$\vec{\pi}_{\mathrm{a}}=\vec{\pi}_{\mathrm{a}}\Lambda^*$$

с условием нормировки

$$\vec{\pi}_{\mathrm{a}}\vec{1}=1$$

(здесь и далее $\vec{1}$ — вектор-столбец размерности I из единиц).

Времена обслуживания заявок независимы в совокупности, не зависят от моментов поступления и одинаково распределены по произвольному дискретному закону с вероятностью b_k, $k\geqslant 0$, того, что обслуживание заявки продлится k тактов (предполагается, что $b_0=0$). Обозначим через $B_k=\sum\limits_{l=k}^{\infty}b_l$, $k\geqslant 1$, вероятность того, что обслуживание заявки продлится не менее k тактов, через $b=\sum\limits_{k=1}^{\infty}kb_k=\sum\limits_{k=1}^{\infty}B_k<\infty$ — среднюю длину (число тактов обслуживания) заявки и через $\beta(z)=\sum\limits_{k=1}^{\infty}z^kb_k$ — производящую функцию (ПФ) длины заявки.

Дисциплина обслуживания заключается в следующем. Предполагается, что в момент поступления заявки в систему становится известной её длина l, т.е. число тактов, необходимых для окончания обслуживания данной заявки. Эта длина сравнивается с (остаточной) длиной k заявки, находящейся на приборе. Вновь поступившая заявка с вероятностью d_{kl}, $k,l\geqslant 1$, зависящей только от длин l и k и не зависящей от предыстории функционирования системы, занимает первое место в очереди, а с дополнительной вероятностью $\overline{d}_{kl} = 1 - d_{kl}$, $k, l \geqslant 1$, сама становится на прибор, вытесняя обслуживавшуюся ранее заявку на первое место в очередь. Остальные заявки, находящиеся в системе, сдвигаются на одно место в очереди с сохранением порядка. Заявки с прерванным обслуживанием дообслуживаются. Если в момент поступления новой заявки на приборе оканчивается обслуживание заявки, то на прибор становится вновь поступившая заявка. Такая дисциплина называется инверсионным порядком обслуживания с вероятностным приоритетом (LIFO PP — *last in first out and probabilistic priority*).

Обозначим через $\lambda = \vec{\pi}_{\mathrm{a}} \mathrm{N} \vec{1}$ (стационарную) интенсивность марковского входящего потока заявок, т.е. среднее число заявок, поступающих в систему в стационарном режиме за один такт. Пусть также $\lambda_k = b_k \lambda$ — (стационарная) интенсивность входящего потока заявок длины k.

Будем предполагать, что загрузка системы $\rho = b\lambda$ меньше единицы. Это условие является необходимым и достаточным для существования стационарного режима функционирования системы $MAP/G/1/\infty$.

3. Стационарное распределение очереди

Введём следующие матрицы:

– матрицу G из вероятностей G_{ij}, $i, j = \overline{1, I}$, того, что сразу же после окончания периода занятости (ПЗ) рассматриваемой системы фаза генерации заявок будет j, при условии, что в начале ПЗ она была i;
– матрицу $\mathrm{G}(k)$, $k \geqslant 1$, из вероятностей $G_{ij}(k)$, $i, j = \overline{1, I}$, того, что сразу же после окончания ПЗ рассматриваемой системы фаза генерации заявок будет j, при условии, что в начале ПЗ она была i и ПЗ открылся заявкой длины k.

Поскольку вероятности изменений фаз генерации заявок за один ПЗ не зависят от дисциплин из класса консервативных дисциплин, то, воспользовавшись инверсионным порядком обслуживания с прерыванием обслуживания и дообслуживанием в системе $MAP/G/1/\infty$, имеем для матриц $\mathrm{G}(k)$ и $\mathrm{G}(k)$ матричные соотношения:

$$\mathrm{G}(k) = (\Lambda + \mathrm{NG})\mathrm{G}(k-1) = \ldots = (\Lambda + \mathrm{NG})^k, \quad k \geqslant 1, \tag{1}$$

$$\mathrm{G} = \sum_{k=1}^{\infty} b_k \mathrm{G}(k) = \sum_{k=1}^{\infty} b_k (\Lambda + \mathrm{NG})^k = \beta(\Lambda + \mathrm{NG}). \tag{2}$$

Матричное уравнение (2) может быть решено итерационным методом. При этом в качестве нулевой итерации удобно взять нулевую матрицу. Тогда в силу монотонной сходимости процедуры очень просто контролировать сходимость итерационного процесса, если учесть, что матрица G является стохастической.

Пусть:

$\vec{p}_0$ — вектор-строка из стационарных вероятностей p_{0i}, $i = \overline{1, I}$, того, что в системе отсутствуют заявки и фаза генерации заявок равна i;

$\vec{p}_1(k)$, $k \geqslant 1$, — вектор-строка, координатами которой являются стационарные вероятности $p_{1i}(k)$, $i = \overline{1, I}$, того, что в системе находится одна заявка остаточной длины k и фаза генерации заявок равна i;

$\vec{p}_n(k_1, \ldots, k_n)$, $n \geqslant 2$, $k_1, \ldots, k_n \geqslant 1$, — вектор-строка, координатами которой являются стационарные вероятности $p_{ni}(k_1, \ldots, k_n)$, $i = \overline{1, I}$ того, что в системе находится n заявок, фаза генерации заявок равна i, заявка на приборе имеет остаточную длину k_1, l-я заявка в очереди имеет длину k_{l+1}, $l = \overline{1, n-1}$.

Выпишем систему уравнений для векторов $\vec{p}_0$ и $\vec{p}_n(k_1, \ldots, k_n)$, $n \geqslant 1$. Для этого рассмотрим новую систему $MAP/G/1/n$, $n \geqslant 0$, с конечным общим числом n заявок, которые одновременно могут находиться в системе, отличающуюся от исходной только тем, что если в очереди находится n заявок, фаза генерации заявки равна i, на приборе обслуживается заявка длины k и поступает новая заявка длины l, то с вероятностью $d_{kl}G_{ij}(k)$ в системе (на приборе) остаётся вновь поступившая заявка, фаза генерации меняется с i-й на j-ю, а обслуживавшаяся ранее заявка покидает систему, и, наоборот, с вероятностью $\overline{d}_{kl}G_{ij}(l)$ систему покидает вновь поступившая заявка, фаза генерации меняется с i-й на j-ю, а находившаяся ранее не приборе заявка продолжает обслуживаться. При этом в системе $MAP/G/1/0$ (в этой системе, вообще, никогда не бывают заявки, даже на приборе) при поступлении заявки просто с вероятностью G_{ij} фаза генерации меняется с i-й на j-ю.

В силу метода исключения состояний (см. [24]) стационарные вероятности состояний в исходной и новой системах отличаются лишь на постоянный множитель. Это даёт возможность для составления уравнения для векторов $\vec{p}_0$ и $\vec{p}_n(k_1,\ldots,k_n)$, $n \geqslant 1$, стационарных вероятностей состояний воспользоваться введённой выше системой $MAP/G/1/n$ и получить следующие соотношения:

$$\vec{p}_0 = \vec{p}_0(\Lambda + \mathrm{NG}), \tag{3}$$

$$\vec{p}_1(k) = \vec{p}_1(k+1)\Lambda + b_k\sum_{m=0}^{\infty} d_{mk}\vec{p}_1(m+1)\mathrm{NG}(m) + \\ + \vec{p}_1(k+1)\mathrm{N}\sum_{m=1}^{\infty}\overline{d}_{km}b_m\mathrm{G}(m) + b_k\vec{p}_0\mathrm{N}, \quad k \geqslant 1, \tag{4}$$

$$\vec{p}_n(k_1,\ldots,k_n) = \vec{p}_n(k_1+1,k_2,\ldots,k_n)\Lambda + b_{k_2}d_{k_1k_2}\vec{p}_{n-1}(k_1+1,k_3,\ldots,k_n)\mathrm{N} + \\ + b_{k_1}\overline{d}_{k_2k_1}\vec{p}_{n-1}(k_2+1,k_3,\ldots,k_n)\mathrm{N} + b_{k_1}\sum_{m=0}^{\infty} d_{mk_1}\vec{p}_n(m+1,k_2,\ldots,k_n)\mathrm{NG}(m) + \\ + \vec{p}_n(k_1+1,k_2,\ldots,k_n)\mathrm{N}\sum_{m=1}^{\infty} b_m\overline{d}_{k_1m}\mathrm{G}(m), \quad n \geqslant 2, \quad k_1,\ldots,k_n \geqslant 1, \tag{5}$$

(здесь принято соглашение, что $d_{0k} = 1$, $k \geqslant 1$, и $\mathrm{G}(0) = \mathrm{E}$).

Для практических целей обычно вполне достаточно знать только лишь стационарные вероятности $\vec{p}_0$ и $\vec{p}_{nk}$, $n \geqslant 1$, $k \geqslant 1$, где координата p_{nki}, $i = \overline{1,I}$, вектора $\vec{p}_{nk}$ представляет собой стационарную вероятность того, что в системе находится n заявок, фаза генерации заявок равна i и (остаточная) длина заявки на приборе равна k. Тогда, суммируя уравнение (5) по всем возможным значениям аргументов $k_2,\ldots,k_n$, приходим к следующему уравнению:

$$\vec{p}_{nk} = \sum_{k_2,\ldots,k_n=1}^{\infty}\vec{p}_n(k,k_2,\ldots,k_n) = \sum_{k_2=1}^{\infty} b_{k_2}d_{kk_2}\vec{p}_{n-1,k+1}\mathrm{N} + \\ + b_k\sum_{k_2=1}^{\infty}\overline{d}_{k_2k}\vec{p}_{n-1,k_2+1}\mathrm{N} + \vec{p}_{n,k+1}\Lambda + b_k\sum_{m=0}^{\infty} d_{mk}\vec{p}_{n,m+1}\mathrm{NG}(m) + \\ + \vec{p}_{n,k+1}\mathrm{N}\sum_{m=1}^{\infty} b_m\overline{d}_{km}\mathrm{G}(m), \quad n \geqslant 2, \quad k \geqslant 1. \tag{6}$$

СУР (3), (4), (6) можно решить, последовательно определяя вектор $\vec{p}_0$, затем векторы $\vec{p}_{1k}$, $k \geqslant 1$, потом векторы $\vec{p}_{2k}$, $k \geqslant 1$, и т.д. При этом для нахождения $\vec{p}_{nk}$, $k \geqslant 1$, из уравнений (4) и (6) удобно использовать метод итераций.

СУР (3), (4), (6) позволяет определить стационарные распределения, связанные с числом заявок в системе, с точностью до постоянной. Для вычисления этой постоянной служит условие нормировки

$$\left(\vec{p}_0 + \sum_{n=1}^{\infty}\sum_{k=1}^{\infty}\vec{p}_{nk}\right)\vec{1} = 1.$$

Однако сейчас мы получим для нее более простое (обычное для систем с ожиданием) выражение.

Запишем СУР (3), (4), (6) в терминах ПФ

$$\vec{P}_k(z) = \sum_{n=1}^{\infty} z^n \vec{p}_{nk}, \quad k \geqslant 1.$$

Умножая при каждом k уравнение (4) на z, уравнения (6) на z^n и суммируя, приходим к системе уравнений

$$\vec{P}_k(z) = zb_k\vec{p}_0\mathrm{N} + \vec{P}_{k+1}(z)\Lambda + \vec{P}_{k+1}(z)\mathrm{N}\sum_{l=1}^{\infty} b_l(\overline{d}_{kl}\mathrm{G}(l) + zd_{kl}\mathrm{E}) + \\ + b_k\sum_{l=0}^{\infty}\vec{P}_{l+1}(z)\mathrm{N}(d_{lk}\mathrm{G}(l) + z\overline{d}_{lk}\mathrm{E}), \quad k \geqslant 1. \qquad (7)$$

Эта система также может быть решена методом итераций.

В частности, вектор $\vec{p}_k = \vec{P}_k(1) = \sum\limits_{n=1}^{\infty}\vec{p}_{nk}$, $k \geqslant 1$, координата p_{ki}, $i = \overline{1, I}$, которого представляет собой стационарную вероятность того, что в системе имеются заявки, фаза генерации заявок равна i и (остаточная) длина заявки на приборе равна k, можно определить из системы уравнений

$$\vec{p}_k = b_k\vec{p}_0\mathrm{N} + \vec{p}_{k+1}\Lambda + \vec{p}_{k+1}\mathrm{N}\sum_{l=1}^{\infty} b_l(\overline{d}_{kl}\mathrm{G}(l) + d_{kl}\mathrm{E}) + \\ + b_k\sum_{l=0}^{\infty}\vec{p}_{l+1}\mathrm{N}(d_{lk}\mathrm{G}(l) + \overline{d}_{lk}\mathrm{E}), \quad k \geqslant 1. \qquad (8)$$

Введём теперь стационарную вероятность $p_k = \vec{p}_k\vec{1}$, $k \geqslant 1$, того, что в системе имеются заявки и на приборе обслуживается заявка (остаточной) длины k. Тогда, учитывая, что $\mathrm{E}\vec{1} = \vec{1}$ и $\overline{d}_{kl} + d_{kl} = 1$, а также равенства $\mathrm{G}(l)\vec{1} = \vec{1}$ и $(\Lambda + \mathrm{N})\vec{1} = \vec{1}$, вытекающие из стохастичности матриц $\mathrm{G}(l)$ и $\Lambda^* = \Lambda + \mathrm{N}$, и равенство $\vec{p}_0 + \sum\limits_{l=1}^{\infty}\vec{P}_l(1) = \vec{\pi}_{\mathrm{a}}$, получаем из (8) после несложных арифметических преобразований:

$$p_k = \lambda B_k, \quad k \geqslant 1. \qquad (9)$$

Из формулы (9) следует, в частности:

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty}\lambda B_k = b\lambda = \rho,$$

откуда находим, что стационарная вероятность $p_0 = \vec{p}_0\vec{1} = 1 - \sum\limits_{k=1}^{\infty} p_k$ отсутствия заявок в системе имеет стандартный для систем с ожиданием вид

$$p_0 = 1 - \rho. \qquad (10)$$

Равенство (10) позволяет однозначно решать сначала уравнение (3), а затем и уравнения (4) и (6).

Дифференцируя ПФ $\vec{P}_k(z)$ в точке $z = 1$ соответствующее число раз, можно вычислять моменты любого порядка стационарного распределения числа заявок в системе. Так, вводя обозначение $\vec{p}_k^{\,\prime} = \vec{P}_k^{\,\prime}(1)$, $k \geqslant 1$, и дифференцируя (7) один раз, приходим к системе уравнений

$$\vec{p}_k^{\,\prime} = b_k\vec{p}_0\mathrm{N} + \vec{p}_{k+1}^{\,\prime}\Lambda + \vec{p}_{k+1}^{\,\prime}\mathrm{N}\sum_{l=1}^{\infty} b_l(\overline{d}_{kl}\mathrm{G}(l) + d_{kl}\mathrm{E}) + \sum_{l=1}^{\infty} b_l d_{kl}\vec{p}_{k+1}\mathrm{N}+$$
$$+ b_k\sum_{l=0}^{\infty}\vec{p}_{l+1}^{\,\prime}\mathrm{N}(d_{lk}\mathrm{G}(l) + \overline{d}_{lk}\mathrm{E}) + b_k\sum_{l=0}^{\infty}\overline{d}_{lk}\vec{p}_{l+1}\mathrm{N}, \quad k \geqslant 1.$$

Вычислив $\vec{p}_k^{\,\prime}$, нетрудно найти математическое ожидание Q стационарного распределения числа заявок в системе:

$$Q = \sum_{k=1}^{\infty}\vec{p}_k^{\,\prime}\vec{1}.$$

4. Стационарное распределение времени пребывания заявки в системе

Снова обратимся к ПЗ рассматриваемой системы, но теперь найдём распределение его длины.

Обозначим через $\gamma_{ij}(n)$, $n \geqslant 0$, $i, j = \overline{1, I}$, условную вероятность того, что ПЗ продлится n тактов и сразу же после его окончания фаза генерации заявок будет j, при условии, что в начале ПЗ фаза генерации была i. Через $\gamma_{ij}(n; k)$, $n \geqslant 0$, $k \geqslant 1$, $i, j = \overline{1, I}$, обозначим аналогичную вероятность, но при дополнительном условии, что открывает ПЗ заявка длины k. Соответствующие матрицы будем обозначать через Γ_n и $\Gamma_n(k)$. Очевидно, что $\Gamma_0 = 0$ и $\Gamma_n(k) = 0$ при $n < k$.

Воспользовавшись, как и раньше, свойством инвариантности ПЗ при всех консервативных дисциплинах и рассматривая инверсионную дисциплину обслуживания с прерыванием обслуживания и дообслуживанием, приходим к следующим соотношениям:

$$\Gamma_n(k) = \Lambda\Gamma_{n-1}(k-1) + \mathrm{N}\sum_{s=0}^{n-1}\Gamma_s\Gamma_{n-s-1}(k-1), \quad k \geqslant 1, \quad n \geqslant 1, \tag{11}$$

$$\Gamma_n = \sum_{k=1}^{\infty} b_k\Gamma_n(k), \quad n \geqslant 1. \tag{12}$$

Здесь для единообразия записи положено:

$$\Gamma_n(0) = \begin{cases}\mathrm{E}, & n = 0,\\ 0, & n \geqslant 1.\end{cases}$$

Введём (матричные) ПФ

$$\Gamma(z) = \sum_{n=1}^{\infty} z^n\Gamma_n \quad \text{и} \quad \Gamma(z; k) = \sum_{n=1}^{\infty} z^n\Gamma_n(k), \quad k \geqslant 1.$$

Тогда, умножая $\Gamma_n(k)$ на z^n в формуле (11) и суммируя по n, получаем:

$$\Gamma(z;k) = z(\Lambda + \mathrm{N}\Gamma(z))\Gamma(z;k-1) = \ldots = (z(\Lambda + \mathrm{N}\Gamma(z)))^k, \quad k \geqslant 1. \tag{13}$$

Аналогично, из формул (12) и (13) имеем:

$$\Gamma(z) = \sum_{k=1}^{\infty} b_k (z(\Lambda + \mathrm{N}\Gamma(z)))^k = \beta(z(\Lambda + \mathrm{N}\Gamma(z))).$$

Последнее уравнение численно может быть решено методом итераций.

Отметим, что $\Gamma(1) = \mathrm{G}$ и $\Gamma(1;k) = \mathrm{G}(k)$, $k \geqslant 1$.

Перейдём теперь к вычислению стационарного распределение времени пребывания заявки в системе.

Начнём с времени пребывания заявки на приборе. Обозначим через $\vec{\psi}(z;k)$, $k \geqslant 1$, ПФ времени обслуживания на приборе (с учётом возможных прерываний и фазы генерации заявок в начале обслуживания) заявки длины k.

Очевидно, что находящаяся на приборе заявка длины 1 на следующем такте (и независимо от фазы генерации) обязательно покинет систему, т.е. время её пребывания на приборе равно единице, а, значит,

$$\vec{\psi}(z;1) = z\vec{1}.$$

Находящаяся на приборе заявка длины k, $k \geqslant 2$, на следующем такте останется на приборе, если в систему не поступит новая заявка. Если же поступит новая заявка длины l, $l \geqslant 1$, то возможны два варианта. Во-первых, с вероятностью $d_{k-1,l}$ поступившая заявка встанет на первое место в очереди и продолжится обслуживание заявки на приборе. Во-вторых, поступившая заявка с вероятностью $\overline{d}_{k-1,l}$ прервёт обслуживание заявки на приборе на ПЗ, имеющий ПФ $\Gamma(z;l)$, и только после окончания этого ПЗ обслуживание выделенной заявки продолжится, но теперь уже при начальной длине $(k-1)$. Таким образом,

$$\vec{\psi}(z;k) = z\left(\Lambda + \mathrm{N}\sum_{l=1}^{\infty} b_l(d_{k-1,l} + \overline{d}_{k-1,l}\Gamma(z;l))\right)\vec{\psi}(z;k-1), \quad k \geqslant 2.$$

Обратимся к времени ожидания начала обслуживания заявки. Обозначая через $\Omega(z;k,l)$, $k,l \geqslant 1$, ПФ времени ожидания начала обслуживания на приборе (с учётом фаз генерации заявок) заявки длины k при условии, что в момент поступления её в систему на приборе находится заявка длины l, и учитывая, что при этом условии либо с вероятностью $\overline{d}_{lk}$ поступающая заявка сразу же попадает на прибор, либо с вероятностью d_{lk} ждёт начала обслуживания ПЗ с ПФ $\Gamma(z;l)$, получаем:

$$\Omega(z;k,l) = \overline{d}_{lk}\mathrm{E} + d_{lk}\Gamma(z;l), \quad k,l \geqslant 1.$$

Для того чтобы найти ПФ $\vec{\omega}(z;k)$, $k \geqslant 1$, стационарного распределения времени ожидания начала обслуживания на приборе (с учётом фазы генерации заявок в конце периода ожидания) заявки длины k, заметим, что интенсивность поступления заявок длины k с учётом фазы генерации после поступления можно представить в виде суммы вектор-строк $\vec{p}_0\mathrm{N}$ и $\vec{p}_{nl}\mathrm{N}$, $n,l \geqslant 1$. При этом если в системе на предыдущем такте отсутствовали заявки (с вероятностью $\vec{p}_0$) или если было n, $n \geqslant 1$, заявок, причём на приборе была заявка длины 1 (с вероятностью $\vec{p}_{n1}$), то поступающая заявка сразу же становится на прибор, а в противном случае она дополнительно ждёт начала обслуживания время с ПФ $\Omega(z;k,l)$. Поделив также на интенсивность λ_k входящего потока заявок длины k, имеем:

$$\vec{\omega}(z;k) = \frac{b_k}{\lambda_k}\left(\vec{p}_0\mathrm{N} + \sum_{n=1}^{\infty}\left(\vec{p}_{n1}\mathrm{N} + \sum_{l=2}^{\infty}\vec{p}_{nl}\mathrm{N}\Omega(z;k,l-1)\right)\right), \quad k \geqslant 1.$$

Наконец, суммируя времена ожидания начала обслуживания на приборе и собственно пребывания заявки на приборе (с учётом возможных прерываний), получаем для ПФ $\varphi(z;k)$, $k \geqslant 1$, ПФ стационарного распределения времени пребывания в системе (включая время ожидания начала обслуживания и собственно пребывания на приборе) заявки длины k выражение

$$\varphi(z;k) = \vec{\omega}(z;k)\vec{\psi}(z;k), \quad k \geqslant 1,$$

а для ПФ $\varphi(z)$ стационарного распределения времени пребывания в системе (включая время ожидания начала обслуживания и собственно пребывания на приборе) заявки произвольной длины — выражение

$$\varphi(z) = \sum_{k=1}^{\infty} b_k \varphi(z;k).$$

Дифференцируя полученные в этом разделе формулы в точке $z = 1$, можно найти моменты соответствующих характеристик.

5. Программная реализация

На основе полученных формул на языке Delphi была написана программа, позволяющая рассчитывать векторы $\vec{p}_0$ и $\vec{p}_{nk}$, $n \geqslant 1$, $k \geqslant 1$, стационарных вероятностей состояний, а также стационарные средние значения показателей функционирования системы связанных с распределением числа заявок в системе[1].

Приведём пример расчётов с помощью разработанной программы.

В систему поступает геометрический поток заявок с вероятностью a поступления заявки на такте. Каждая заявка с вероятностью $b_1 = 0{,}9$ обслуживается 1 такт и с вероятностью $b_{11} = 0{,}1$ — 11 тактов. Таким образом, средняя длина заявки (среднее число тактов обслуживания) $b = 2$. Расчёты проведены для 18 значений $a = 0{,}025;\ 0{,}05;\ \ldots;\ 0{,}45$, что соответствует значениям $\rho = 0{,}05;\ 0{,}1;\ \ldots;\ 0{,}9$ загрузки.

Рассматриваются следующие 5 вариантов дисциплины LIFO PP:

вариант 1 — после сравнения длины поступающей заявки с длиной заявки на приборе на прибор становится длинная заявка (т.е. $d_{kl} = 0$, если $k < l$, и $d_{kl} = 1$ в противном случае);

вариант 2 — на приборе остаётся прежняя заявка, а новая заявка занимает первое место в очереди ($d_{kl} = 1$). Этот вариант реализует инверсионный порядок обслуживания без прерывания обслуживания, с точки зрения стационарного распределения очереди эквивалентный дисциплине обслуживания в порядке поступления;

вариант 3 — независимо от длин заявок поступающая заявка с вероятностью 0,5 становится на прибор и с вероятностью $0,5$ — на первое место в очередь ($d_{kl} = 0{,}5$);

вариант 4 — новая заявка становится на прибор, а прежняя заявка занимает первое место в очереди ($d_{kl} = 0$). Это — инверсионный порядок обслуживания с прерыванием обслуживания;

вариант 5 — после сравнения длины поступающей заявки с длиной заявки на приборе на прибор становится короткая заявка ($d_{kl} = 1$, если $k < l$, и $d_{kl} = 0$ в противном случае).

На рис. 1 приведены полученные в результате расчетов значения среднего числа Q заявок в системе для указанных значений загрузки. Линии 1–5 соответствуют вариантам 1–5. Линия 6 даёт значения Q при тех же параметрах входящего потока, но при постоянной длине заявки, равной двум.

---

[1]Авторы благодарят В. В. Чаплыгина, оказавшего большую помощь в написании программы.

Отметим, что при большой загрузке вариант 5 дисциплины LIFO PP более чем в 3 раза уменьшает среднее число заявок в системе (и в силу формулы Литтла среднее время пребывания заявки в системе) по сравнению с обслуживанием заявок в порядке поступления (см. вариант 2). В первом случае $Q = 3{,}84$, а во втором — $Q = 12{,}04$.

6. Заключение

Таким образом, в настоящей работе получены математические соотношения для вычисления основных стационарных характеристик функционирующей в дискретном времени системы $MAP/G/1/\infty$ с марковским входящим потоком и инверсионным порядком обслуживания с вероятностным приоритетом (дисциплина LIFO PP).

Полученные соотношения использованы для написания программы, с помощью которой проведены численные расчёты. Расчёты, в частности, показали, что при загрузке, близкой к единице, от использования вариантов дисциплины LIFO PP можно получить многократное уменьшение среднего числа заявок в системе и среднего времени пребывания заявки в системе.

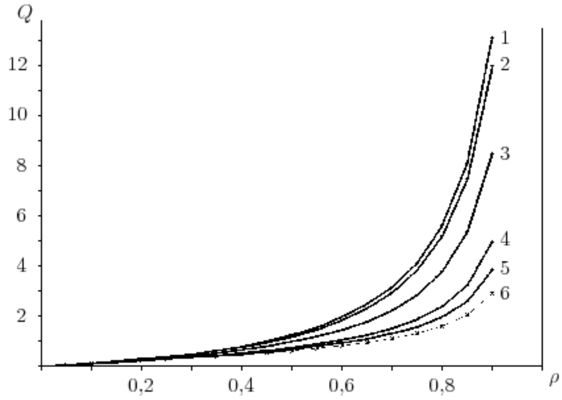

Рис. 1. **Среднее число заявок в системе для некоторых вариантов дисциплины обслуживания и различных значений загрузки**

Интересно отметить следующие факты.

Во-первых, как показано в [7], в функционирующей в непрерывном времени системе $M/G/1/\infty$ с пуассоновским входящим потоком и дисциплиной выбора на обслуживание короткой заявки (из поступающей заявки и заявки на приборе, см. вариант 5 из предыдущего раздела) стационарное распределение числа заявок при фиксированной загрузке ρ не зависит от распределения времени обслуживания заявки (и совпадает со стационарным распределением числа заявок в системе $M/D/1/\infty$ с постоянной длиной заявки и обслуживанием заявок в порядке поступления). Однако в аналогичной системе $Geo/G/1/\infty$ в дискретном времени такое свойство инвариантности отсутствует. Более того, от распределения времени обслуживания заявки зависит даже стационарное среднее число заявок в системе. Указанный факт не позволяет в дискретном времени, в отличие от непрерывного, находить точные нижние границы этих характеристик при разных распределениях времени обслуживания заявки.

Во-вторых, хорошо известно, что свойство инвариантности стационарного распределения числа заявок при фиксированной загрузке ρ имеет место для системы $M/G/1/\infty$ с инверсионным порядком обслуживания с прерыванием обслуживания и дообслуживанием (вариант 4). Это свойство справедливо и для системы $Geo/G/1/\infty$, но только если фиксированы не только загрузка, но и вероятность поступления заявки на такте (а, значит, и средняя длина заявки).

Литература

1. *Schrage L., Miller L.* The Queue $M/G/1$ with the Shortest Remaining Processing Time Discipline // Oper. Res. — 1966. — Vol. 14. — Pp. 670–684.
2. *Schrage L.* A Proof of the Optimality of the Shortest Remaining Processing Time Discipline // Oper. Res. — 1968. — Vol. 16. — Pp. 687–690.
3. *Бочаров П. П., Печинкин А. В.* Теория массового обслуживания. — М.: РУДН, 1995.
4. *Schassberger R.* The Steady-State Appearance of the $M/G/1$ Queue under the Discipline of Shortest Remaining Processing Time // Adv. Appl. Probab. — 1990. — Vol. 22. — Pp. 456–479.
5. *Grishechkin S. A.* On a Relationship Between Processor-Sharing Queues and Cramp–Mode–Jagers Branching Processes // Adv. Appl. Probab. — 1992. — Vol. 24. — Pp. 653–698.
6. *Печинкин А. В.* Система $MAP/G/1/\infty$ с дисциплиной SRPT // Теория вероятностей и ее прим. — 2000. — Т. 45, № 3. — С. 589–595.
7. *Печинкин А. В.* Об одной инвариантной системе массового обслуживания // Math. Operationsforsch. und Statist., Ser. Optimization. — 1983. — Т. 14, № 3. — С. 433–444.
8. *Нагоненко В. А.* О характеристиках одной нестандартной системы массового обслуживания. I // Изв. АН СССР. Технич. кибернет. — 1981. — № 1. — С. 187–195.
9. *Нагоненко В. А.* О характеристиках одной нестандартной системы массового обслуживания. II // Изв. АН СССР. Технич. кибернет. — 1981. — № 3. — С. 91–99.
10. *Pechinkin A., Svischeva T.* The Stationary State Probability in the $BMAP/G/1/r$ Queueing System with Inverse Discipline and Probabilistic Priority // Transactions of XXIV International Seminar on Stability Problems for Stochastic Models. — Jurmala, Latvia: 2004. — Pp. 141–147.
11. *Таташев А. Г.* Многоканальная система массового обслуживания с потерей кратчайших требований // Автоматика и телемеханика. — 1991. — № 7. — С. 187–189.
12. *Таташев А. Г.* Одна система массового обслуживания с инвариантной дисциплиной // Автоматика и телемеханика. — 1992. — № 7. — С. 92–96.
13. *Таташев А. Г.* Одна инверсионная дисциплина обслуживания в системе с групповым поступлением // Автоматика и вычисл. техника. — 1995. — № 1. — С. 53–59.
14. *Таташев А. Г.* Одна инверсионная дисциплина обслуживания в одноканальной системе с разнотипными заявками // Автоматика и телемеханика. — 1999. — № 7. — С. 177–181.
15. *Chaudhry M.* Invited talk: Queue-length and Waiting-Time Distributions of Discrete-Time $GI(X)/Geom/1$ Queueing Systems with Early and Late Arrivals // Queueing Systems: Theory and Applications. — 1997. — Vol. 25, No 1-4. — Pp. 307–324.
16. *Desert B., Daduna H.* Discrete Time Tandem Networks of Queues Effects of Different Regulation Schemes for Simultaneous Events // Performance Evaluation. — 2002. — Vol. 47, No 2. — Pp. 73–104.
17. *He J., Sohraby K.* An Extended Combinatorial Analysis Framework for Discrete-Time Queueing Systems with General Sources // IEEE/ACM Transactions on Networking (TON). — 2003. — Vol. 11, No 1. — Pp. 95–110.
18. *Chaudhry M., Gupta U., Goswami V.* On Discrete-Time Multiserver Queues with Finite Buffer: $GI/Geom/m/N$ // Computers and Operations Research. — 2004. — Vol. 31, No 13. — Pp. 2137–2150.
19. *Fiems D., Steyaert B., Bruneel H.* Discrete-Time Queues with Generally Distributed Service Times and Renewal-Type Server Interruptions // Performance Evaluation. — 2004. — Vol. 55, No 3-4. — Pp. 277–298.
20. *Atencia I., Moreno P.* A Discrete-Time $Geo/G/1$ Retrial Queue with General Retrial Times // Queueing Systems. — 2004. — No 48. — Pp. 5–21.

21. *Akar N.* A Matrix Analytical Method for the Discrete Time Lindley Equation using the Generalized Schur Decomposition // ACM International Conference Proceeding Series. — 2006. — Vol. 201, No 12.
22. *Печинкин А. В., Шоргин С. Я.* Система $Geo/G/1/\infty$ с одной «нестандартной» дисциплиной обслуживания // Информатика и ее применения. — 2008. — Т. 2, № 1. — С. 55–62.
23. *Pechinkin A. V., Shorgin S. Y.* The Discrete-Time Queueing System with Inversive Service Order and Probabilistic Priority // Proceedings of the 3rd International Workshop on Tools for Solving Structured Markov Chains (SMCTools 2008). — Athens, Greece: 2008.
24. Queueing Theory / P. P. Bocharov, C. D'Apice, A. V. Pechinkin, S. Salerno. — Utrecht, Boston: VSP, 2004.

UDC 519.872

The $MAP/G/1/\infty$ Discrete-Time Queueing System with Inversive Service Order and Probabilistic Priority

A. V. Pechinkin, I. V. Stalchenko

Department of Probability Theory and Mathematical Statistics
Peoples' Friendship University of Russia
Miklukho-Maklaya str., 6, Moscow, Russia 117198

This paper considers a discrete-time queueing system $MAP/G/1/\infty$ that is determined as follows. Upon arrival into the system of a new customer its length is compared with the (remaining) length of the customer in the device and with some probability which depends only on this two lengths, a new arrival will occupy the server while pushing out a servicing customer to the first place in the queue, and with the supplemental probability, alternatively, the newly arrived one occupies the first position in the queue (inversive service order and probabilistic priority). The main stationary characteristics of such system's behavior have been found. A number of numerical examples are presented according to found analytical formulae.

Key words and phrases: queueing system, discrete-time, inversive service order with probabilistic priority.