

СКАЛЯРНО-ВЕКТОРНЫЙ РЕКУРРЕНТНЫЙ АЛГОРИТМ НАХОЖДЕНИЯ СТАЦИОНАРНЫХ ВЕРОЯТНОСТЕЙ В ГЕТЕРОГЕННОЙ СИСТЕМЕ $M/(M_1, M_2)/(N_1, N_2)/\infty/FIFO$

Назаров А. А.^{1а}, Рындин А. В.^{2б}, Пакулова Е. А.^{3б}
Туренова И. А.^{4а}, Моисеева С. П.^{5а}

(^аНациональный исследовательский Томский
государственный университет, Томск)

(^бЮжный федеральный университет, Таганрог)

Предлагается обобщенная модель Эрланга с ожиданием, особенность которой заключается в том, что для обработки и передачи данных предоставлены два гетерогенных канала конечного объема и разной интенсивности обслуживания. При поступлении сообщения сначала обращаются в более «быстрый» канал, и если он обладает достаточным количеством единиц канального ресурса, то сообщение попадает на обслуживание. В противном случае при недостаточном количестве свободного ресурса для обслуживания сообщение обращается во второй, «медленный» канал. Если он обладает достаточным количеством свободного ресурса для его обслуживания, тогда сообщение попадает на обслуживание. В противоположном случае сообщение попадает в очередь. Сообщения, находящиеся в очереди, имеют срок жизни, после которого передавать их нет смысла. Предложен оригинальный скалярно-векторный алгоритм для расчета стационарных вероятностей состояний системы. Проведен анализ характеристик качества обслуживания. А именно, загрузка каналов, среднее время пребывания в системе, вероятность моментального обслуживания, характеристики задержки в обслуживании и среднее число отказов в обслуживании.

Ключевые слова: гетерогенная система массового обслуживания.

1. Введение

С повсеместным распространением беспроводных пользовательских устройств стала актуальна беспроводная передача данных различных приложений пользователей, кото-

¹ Анатолий Андреевич Назаров, д.т.н., профессор (nazarov.tsu@gmail.com).

² Артём Валерьевич Рындин, аспирант (artem_ryndin@mail.ru).

³ Екатерина Анатольевна Пакулова, к.т.н. (epakulova@sfnedu.ru).

⁴ Ирина Алексеевна Туренова, доцент (irenaturena@mail.ru).

⁵ Светлана Петровна Моисеева, д.ф.-м.н., профессор (smoiseeva@mail.ru).

рая порождает все больше трафика в беспроводной сети. Особенностью беспроводного устройства является то, что оно, как правило, не имеет постоянного подключения к какой-либо точке доступа (базовой станции), которая бы гарантировала необходимый уровень качества, поскольку обладает большой пространственной динамичностью. В таком случае пользователь может перемещаться в зоны с критически слабым сигналом сети или вовсе без него или в зоны с большим количеством пользователей, разделяющими пропускную способность канала.

Одним из решений обеспечения качества передачи данных для мобильных устройств пользователей является распределение данных приложений пользователя на несколько субпотоков передачи данных, организуя мультипотокую передачу данных с использованием нескольких доступных интерфейсов связи.

Мультипотокую передачу данных является одним из актуальных направлений современных исследований, которому посвящены стандарты протоколов мультипотковой передачи данных MPTCP (Multi-path TCP, RFC 6824[9]), SCTP (Stream Control Transmission Protocol, RFC 4960[20]), интернет-документов Multipath RTP (MPRTP) и Multi-Flow Real-Time Transport Protocol (MRTP), Multi-Path Extensions for QUIC (MP-QUIC) [4], а также большое количество научных работ как зарубежных, так и отечественных ученых: К. Chebrolu, D. Jurca, J. Nightigale, Q. De Coninck, O. Bonaventure, Р.Б. Трегубов, Е.А. Пакулова и т.д. [1, 3, 5, 3, 14, 16, 22]

В большинстве своем работы, посвященные проблематике мультипотковой передачи данных, построены на анализе параметров качества передачи данных на имитационных моделях. В данной статье предлагается аналитическая модель системы мультипотковой передачи данных в виде гетерогенной системы массового обслуживания с двумя каналами различной интенсивности обслуживания и ограничением на срок пребывания в очереди.

Многие явления реальной жизни, такие как компьютерные системы, коммуникационные сети, производственные системы, кассовые линии супермаркетов, а также структурные военные системы могут быть представлены с помощью моделей массо-

вого обслуживания. Системы массового обслуживания берут свое начало при изучении проблем проектирования АТС и впервые были проанализированы А.К. Эрлангом в начале 1900-х годов [7, 8]. В то время при проектировании телефонных систем одной из основных проблем, связанных с критериями производительности, было определение того, сколько линий необходимо обеспечить, чтобы гарантировать определенный уровень обслуживания. Подобные вопросы возникают во многих других случаях, например, при запросе максимального количества терминалов в компьютерной системе, допускающей удержание вероятности потери сообщений ниже заданного уровня, или при рассмотрении влияния скорости обслуживания на очереди ожидания и т.д. Изучая математические модели, раскрывающие вероятностную природу реальных систем, в теории массового обслуживания получены аналитические результаты для ключевых величин, которые характеризуют производительность систем. Такими ключевыми величинами являются, например, длина очереди и распределение времени ожидания, вероятности потерь, средние значения времени пребывания и пропускной способности и т.д.

В классической системе массового обслуживания $M/M/n/0$ предполагается, что исследуемая система состоит из n одинаковых серверов с отсутствием мест для ожидания. Известны результаты для таких систем с очередями и без, с входящим пуассоновским потоком вызовов, рекуррентным обслуживанием и неограниченным количеством мест ожидания ($M/G/N/\infty$). Достаточно много статей разных авторов посвящено изучению моделей такого типа, в том числе моделей с непуассоновскими входными потоками [1, 11–13, 15, 17, 21].

Вышеупомянутые работы позволяют получить характеристики работы системы, в основном используя численные, приближенные или асимптотические методы.

В то же время системы массового обслуживания с неоднородными приборами анализируются не так часто. Первые исследования систем массового обслуживания $M/M \rightarrow n/0$ с неоднородными приборами появились в [10, 18, 19].

При рассмотрении таких систем, предполагается, что приборы отличаются интенсивностью обслуживания, и более

медленный подключается при выполнении определенных условий, например, если очередь превышает заданное число.

В настоящей статье предлагается обобщенная модель Эрланга с ожиданием, особенность которой заключается в том, что для обработки и передачи данных предоставлены два гетерогенных ресурса конечного объема и разной интенсивности обслуживания. Предполагается, что поступившее сообщение может использовать любую свободную единицу канального ресурса (ЕКР) одного из каналов передачи информации, и все они идентичны с точки зрения возможностей и качества предоставления запрашиваемого сервиса. При поступлении сообщения сначала обращаются в первый канал, и если он обладает достаточным количеством ЕКР, то сообщение попадает на обслуживание. В противном случае при недостаточном количестве свободного ресурса для обслуживания сообщение обращается во второй канал. Если второй канал обладает достаточным количеством свободного ресурса для его обслуживания, тогда сообщение попадает на обслуживание. В противоположном случае сообщение попадает в очередь. Дисциплина обслуживания в очереди FIFO (First In – First Out). Что означает: первый пришел – первый поступил на обслуживание. Кроме того, сообщения, находящиеся в очереди, имеют срок жизни, после которого передавать их нет смысла.

Ставится задача определения характеристик качества обслуживания. А именно, загрузка каналов, среднее время пребывания в системе, вероятность моментального обслуживания, характеристики задержки в обслуживании и среднее число отказов в обслуживании.

2. Математическая модель

В качестве математической модели рассмотрим систему массового обслуживания (СМО) с двумя блоками обслуживания различной интенсивности, на вход которой поступает пуассоновский поток заявок с интенсивностью λ . Предполагается, что каналы имеют N_1 и N_2 ЕКР соответственно. Время обслуживания каждого сообщения является неотрицательной случайной величиной, имеющей

экспоненциальное распределение с параметрами μ_1 и μ_2 для первого и второго канала соответственно. Если свободных ресурсов нет, то сообщение поступает в очередь. Будем считать, что ограничений на очередь нет, но находящиеся в ней сообщения могут покинуть ее с интенсивностью α . Используя символику Д. Кендалла, запишем такую СМО в следующем виде: $M/(M_1, M_2)/(N_1, N_2)/\infty/\text{FIFO}$. Схематическое изображение системы представлено на рис. 1.

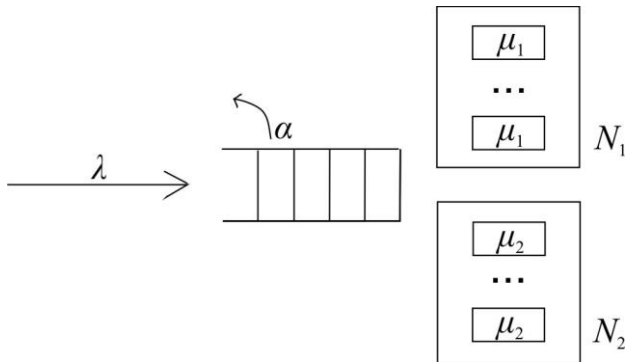


Рис. 1. Математическая модель

Следующий этап построения модели состоит в определении компонент случайного процесса, который будет использоваться для оценки показателей обслуживания сообщений. Для решения поставленных задач оценки качества обслуживания необходимо знать среднее число занятых ЕКР каждого канала, вероятность постановки на ожидание (задержка передачи), средний размер очереди, долю потерянных сообщений.

Обозначим $n_k(t)$ – число занятых ЕКР k -го канала в момент времени t , $i(t)$ – число сообщений, находящихся в очереди в момент времени t . Трехмерный случайный процесс $\{n_1(t), n_2(t), i(t)\}$ является марковским.

Определим $P(n_1, n_2, i) = P\{n_1(t) = n_1, n_2(t) = n_2, i(t) = i\}$ – стационарные вероятности числа занятых ЕКР каналов и сообщений в очереди, здесь $n_1 = 0, 1, \dots, N_1$; $n_2 = 0, 1, \dots, N_2$, $i = 0, 1, \dots, \infty$. Так как сообщения в очереди начинают накапливаться только в том случае, если все ЕКР обоих каналов заняты

($n_1 = N_1$ и $n_2 = N_2$), то мы будем иметь дело с вероятностями только вида $P(n_1, n_2, 0)$ и $P(N_1, N_2, i)$. Для компактности записи в дальнейшем будем опускать 0 в записи вероятностей числа занятых ЕКР при отсутствии сообщений в очереди, т.е. условимся записывать $P(n_1, n_2, 0) = P\{n_1(t) = n_1, n_2(t) = n_2, i(t) = 0\} = P_0(n_1, n_2)$. При этом стационарные вероятности числа сообщений в очереди, имеющие вид $P(N_1, N_2, i) = P\{n_1(t) = N_1, n_2(t) = N_2, i(t) = i\}$, будем записывать как $\pi(i)$.

Графическая иллюстрация интенсивностей и направлений переходов из произвольного состояния для рассматриваемого процесса показана на рис. 2, 3.

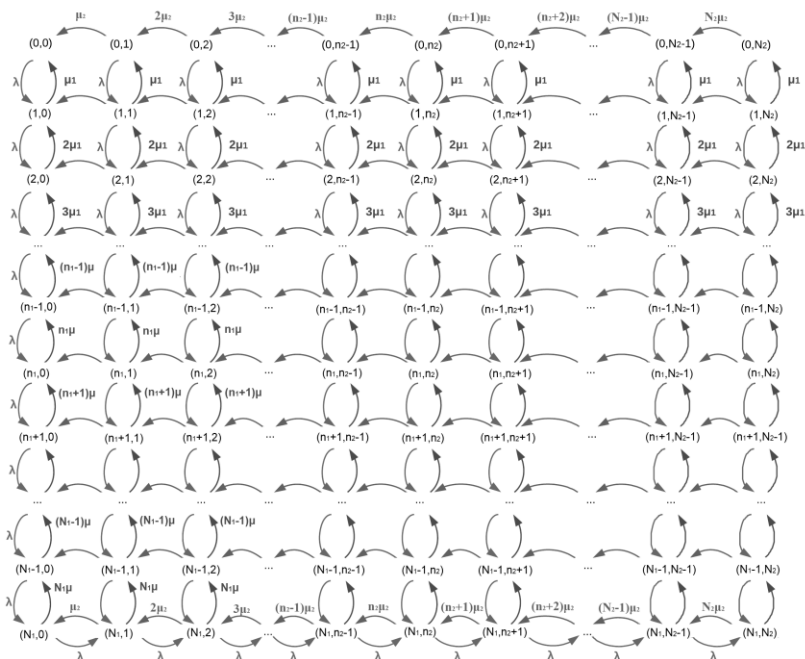


Рис. 2. Диаграмма переходов для $P_0(n_1, n_2)$

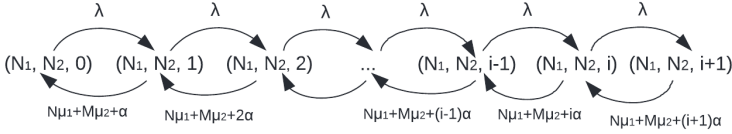


Рис. 2. Диаграмма переходов для $\pi(i)$

На основании графа переходов получаем следующие системы линейных уравнений равновесия для стационарного распределения вероятностей числа занятых ЕКР каналов:

$$\begin{aligned}
 (1) \quad & \left\{ \begin{aligned} & \lambda P(0,0) = \mu_1 P(1,0) + \mu_2 P(0,1), \\ & (\lambda + n_1 \mu_1) P(n_1, 0) = (n_1 + 1) \mu_1 P(n_1 + 1, 0) + \lambda P(n_1 - 1, 0) + \\ & \quad + \mu_2 P(n_1, 1), \\ & (\lambda + N_1 \mu_1) P(N_1, 0) = \lambda P(N_1 - 1, 0) + \mu_2 P(N_1, 1). \end{aligned} \right. \\
 (2) \quad & \left\{ \begin{aligned} & (\lambda + n_2 \mu_2) P(0, n_2) = \mu_1 P(1, n_2) + (n_2 + 1) \mu_2 P(0, n_2 + 1), \\ & (\lambda + n_1 \mu_1 + n_2 \mu_2) P(n_1, n_2) = (n_1 + 1) \mu_1 P(n_1 + 1, n_2) + \\ & \quad + \lambda P(n_1 - 1, n_2) + (n_2 + 1) \mu_2 P(n_1, n_2 + 1), \\ & (\lambda + N_1 \mu_1 + n_2 \mu_2) P(N_1, n_2) = \lambda P(N_1 - 1, n_2) + \lambda P(N_1, n_2 - 1) + \\ & \quad + (n_2 + 1) \mu_2 P(N_1, n_2 + 1). \end{aligned} \right. \\
 (3) \quad & \left\{ \begin{aligned} & (\lambda + N_2 \mu_2) P(0, N_2) = \mu_1 P(1, N_2), \\ & (\lambda + n_1 \mu_1 + N_2 \mu_2) P(n_1, N_2) = (n_1 + 1) \mu_1 P(n_1 + 1, N_2) + \\ & \quad + \lambda P(n_1 - 1, N_2), \\ & (N_1 \mu_1 + N_2 \mu_2) P(N_1, N_2) = \lambda P(N_1 - 1, N_2) + \lambda P(N_1, N_2 - 1). \end{aligned} \right.
 \end{aligned}$$

Для стационарного распределения вероятностей числа сообщений в очереди получим, обозначив для удобства $\mu = N_1 \mu_1 + N_2 \mu_2$:

$$(4) \quad (\mu + i\alpha)\pi(i) = \lambda\pi(i-1), \quad i = \overline{1, \infty}.$$

То есть

$$(5) \quad \pi(i) = \frac{\lambda}{(\mu + i\alpha)} \pi(i-1) = \pi(0) \prod_{v=1}^i \frac{\lambda}{(\mu + v\alpha)}, \quad i = \overline{1, \infty}.$$

Заметим, что при $\alpha > 0$ стационарный режим существует при любых значениях $\lambda, \mu_1, \mu_2, N_1$ и N_2 . При $\alpha = 0$ стационарный режим существует при $\lambda < N_1\mu_1 + N_2\mu_2$.

3. Численный скалярно-векторный алгоритм

Для численного решения составленной системы уравнений обозначим $V(n_1, n_2) = P(n_1, N_2 - n_2)$. Систему (1)–(3) перепишем в инверсионном порядке:

$$(6) \quad \begin{cases} (\lambda + N_2\mu_2)V(0,0) = \mu_1V(1,0), \\ (\lambda + n_1\mu_1 + N_2\mu_2)V(n_1,0) = (n_1 + 1)\mu_1V(n_1 + 1,0) + \\ \quad + \lambda V(n_1 - 1,0), \\ (N_1\mu_1 + N_2\mu_2)V(N_1,0) = \lambda V(N_1 - 1,0) + \lambda V(N_1,1). \end{cases}$$

$$(7) \quad \begin{cases} (\lambda + (N_2 - n_2)\mu_2)V(0,n_2) = (N_2 - n_2 + 1)\mu_2V(0,n_2 - 1) + \\ \quad + \mu_1V(1,n_2), \\ (\lambda + n_1\mu_1 + (N_2 - n_2)\mu_2)V(n_1,n_2) = (n_1 + 1)\mu_1V(n_1 + 1,n_2) + \\ \quad + \lambda V(n_1 - 1,n_2) + (N_2 - n_2 + 1)\mu_2V(n_1,n_2 - 1), \\ (\lambda + N_1\mu_1 + (N_2 - n_2)\mu_2)V(N_1,n_2) = \lambda V(N_1 - 1,n_2) + \\ \quad + \lambda V(N_1,n_2 + 1) + (N_2 - n_2 + 1)\mu_2V(N_1,n_2 - 1). \end{cases}$$

$$(8) \quad \begin{cases} \lambda V(0,N_2) = \mu_1V(1,N_2) + \mu_2V(0,N_2 - 1), \\ (\lambda + n_1\mu_1)V(n_1,N_2) = (n_1 + 1)\mu_1V(n_1 + 1,N_2) + \lambda V(n_1 - 1,N_2) + \\ \quad + \mu_2V(n_1,N_2 - 1), \\ (\lambda + N_1\mu_1)V(N_1,N_2) = \lambda V(N_1 - 1,N_2) + \mu_2V(N_1,N_2 - 1). \end{cases}$$

Полагая $V(0, 0) = C$, где C – некоторый скаляр, из (6) получаем:

$$(9) \quad V(1,0) = \frac{\lambda + N_2\mu_2}{\mu_1} C.$$

Из второго уравнения системы (6) с помощью рекуррентной формулы находим значения $V(n_1, 0)$ для всех $n_1=1, \dots, N_1 - 1$:

$$(10) \quad V(n_1 + 1,0) = \frac{(\lambda + n_1\mu_1 + N_2\mu_2)V(n_1,0) - \lambda V(n_1 - 1,0)}{(n_1 + 1)\mu_1}.$$

Сформируем из полученных значений два вектор-столбца:

(11) $\mathbf{v}_0 = \{V(0, 0), V(1, 0), \dots, V(N_1-1, 0), V(N_1, 0)\}^T$
 размерности $(N_1 + 1) \times 1$ и:

(12) $\hat{\mathbf{v}}_0 = \{V(0,0), V(1,0), \dots, V(N_1 - 1,0)\}^T$
 размерности $N_1 \times 1$.

Помним при этом, что (условие сшивания)
 $V(N_1, 0) = P(N_1, N_2, 0) = \pi(0)$.

Из последнего уравнения системы (6) найдем скаляр

$$(13) V(N_1,1) = \frac{1}{\lambda} \{ \mu V(N_1,0) - \lambda V(N_1 - 1,0) \},$$

который является последней компонентой вектора
 $\mathbf{v}_1 = \{V(0, 1), V(1, 1), \dots, V(N_1 - 1, 1), V(N_1, 1)\}^T$ размерности
 $(N_1 + 1) \times 1$.

Для нахождения значений остальных компонент вектора \mathbf{v}_1
 обозначим $\hat{\mathbf{v}}_1$ вектор размерности $N_1 \times 1$:

$$(14) \hat{\mathbf{v}}_1 = \{V(0,1), V(1,1), \dots, V(N_1 - 1,1)\}^T.$$

3.1. РЕШЕНИЕ СИСТЕМЫ (7)

Систему (7) перепишем в виде:

$$(15) \left\{ \begin{array}{l} (\lambda + (N_2 - n_2)\mu_2)V(0, n_2) - \mu_1 V(1, n_2) = \\ = (N_2 - n_2 + 1)\mu_2 V(0, n_2 - 1), \\ -\lambda V(n_1 - 1, n_2) + (\lambda + n_1\mu_1 + (N_2 - n_2)\mu_2)V(n_1, n_2) - \\ - (n_1 + 1)\mu_1 V(n_1 + 1, n_2) = (N_2 - n_2 + 1)\mu_2 V(n_1, n_2 - 1), \\ -\lambda V(N_1 - 2, n_2) + (\lambda + (N_1 - 1)\mu_1 + (N_2 - n_2)\mu_2)V(N_1 - 1, n_2) = \\ = N_1\mu_1 V(N_1, n_2) + (N_2 - n_2 + 1)\mu_2 V(N_1 - 1, n_2 - 1), \\ (\lambda + N_1\mu_1 + (N_2 - n_2)\mu_2)V(N_1, n_2) = \lambda V(N_1 - 1, n_2) + \\ + \lambda V(N_1, n_2 + 1) + (N_2 - n_2 + 1)\mu_2 V(N_1, n_2 - 1). \end{array} \right.$$

Введем трехдиагональные матрицы \mathbf{A}_{n_2} размерности $N_1 \times N_1$ при
 $n_2=0, \dots, N_2-1$, с элементами

$$(16) \left[\mathbf{A}_{n_2} \right]_{i,i} = \lambda + i\mu_1 + (N_2 - n_2)\mu_2,$$

$$\left[\mathbf{A}_{n_2} \right]_{i+1,i} = -\lambda,$$

$$\left[\mathbf{A}_{n_2} \right]_{i,i+1} = -(i+1)\mu_1, \quad i = \overline{0, N_1 - 1},$$

и векторы

$$(17) \hat{\mathbf{v}}_{n_2} = \{V(0, n_2), V(1, n_2), \dots, V(N_1 - 1, n_2)\}^T$$

размерности $N_1 \times 1$,

Тогда систему (15) запишем в матричном виде:

$$(18) \mathbf{A}_{n_2} \hat{\mathbf{v}}_{n_2} = (N_2 - n_2 + 1)\mu_2 \hat{\mathbf{v}}_{n_2-1} + \mathbf{e}N_1\mu_1 V(N_1, n_2),$$

где \mathbf{e} – вектор размерности N_1 вида $\mathbf{e} = \{0, 0, \dots, 1\}^T$.

Полагая в (18) известным вектор $\hat{\mathbf{v}}_{n_2-1}$ (при $n_2 = 1$) и скаляр

$V(N_1, n_2)$ (из (13) при $n_2 = 1$), найдем вектор

$$(19) \hat{\mathbf{v}}_{n_2} = \mathbf{A}_{n_2}^{-1} \{ (N_2 - n_2 + 1)\mu_2 \hat{\mathbf{v}}_{n_2-1} + \mathbf{e}N_1\mu_1 V(N_1, n_2) \}$$

размерности $N_1 \times 1$, добавляя к которому компоненту $V(N_1, n_2)$,

получим вектор размерности $(N_1 + 1) \times 1$:

$$(20) \mathbf{v}_{n_2} = \{ \hat{\mathbf{v}}_{n_2}, V(N_1, n_2) \}^T = \{ V(0, n_2), V(1, n_2), \dots, V(N_1 - 1, n_2), V(N_1, n_2) \}^T.$$

Из последнего уравнения системы (15) найдем значение скаляра $V(N_1, n_2 + 1)$:

$$(21) V(N_1, n_2 + 1) = \frac{1}{\lambda} \{ (\lambda + N_1\mu_1 + (N_2 - n_2)\mu_2)V(N_1, n_2) - \\ - \lambda V(N_1 - 1, n_2) - (N_2 - n_2 + 1)\mu_2 V(N_1, n_2 - 1) \},$$

который является последней компонентой вектора

$$(22) \mathbf{v}_{n_2+1} = \{ V(0, n_2 + 1), V(1, n_2 + 1), \dots, V(N_1 - 1, n_2 + 1), V(N_1, n_2 + 1) \}^T$$

размерности $(N_1 + 1) \times 1$.

3.2. РЕШЕНИЕ СИСТЕМЫ (8)

Систему (8) перепишем в виде

$$(23) \left\{ \begin{array}{l} \lambda V(0, N_2) - \mu_1 V(1, N_2) = \mu_2 V(0, N_2 - 1), \\ -\lambda V(n_1 - 1, N_2) + (\lambda + n_1\mu_1)V(n_1, N_2) - (n_1 + 1)\mu_1 V(n_1 + 1, N_2) = \\ = \mu_2 V(n_1, N_2 - 1), \\ -\lambda V(N_1 - 2, N_2) + (\lambda + (N_1 - 1)\mu_1)V(N_1 - 1, N_2) = \\ = N_1\mu_1 V(N_1, N_2) + \mu_2 V(N_1 - 1, N_2 - 1), \\ (\lambda + N_1\mu_1)V(N_1, N_2) = \lambda V(N_1 - 1, N_2) + \mu_2 V(N_1, N_2 - 1). \end{array} \right.$$

Три первых уравнения этой системы запишем в матричной форме:

$$(24) \mathbf{A}_0 \hat{\mathbf{v}}_{N_2} = \mu_2 \hat{\mathbf{v}}_{N_2-1} + N_1 \mu_1 \mathbf{e} V(N_1, N_2),$$

откуда получим вектор $\hat{\mathbf{v}}_{N_2}$ размерности $N_1 \times 1$:

$$(25) \hat{\mathbf{v}}_{N_2} = \mathbf{A}_0^{-1} \left\{ \mu_2 \hat{\mathbf{v}}_{N_2-1} + N_1 \mu_1 \mathbf{e} V(N_1, N_2) \right\},$$

добавляя к которому компоненту $V(N_1, N_2)$, получим вектор \mathbf{v}_{N_2} размерности $(N_1 + 1) \times 1$:

$$(26) \mathbf{v}_{N_2} = \left\{ \hat{\mathbf{v}}_{N_2}, V(N_1, N_2) \right\} = \\ = \left\{ V(0, N_2), V(1, N_2), \dots, V(N_1 - 1, N_2), V(N_1, N_2) \right\}^T.$$

3.3. СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ ЧИСЛА ЗАНЯТЫХ ЕКР И ЧИСЛА СООБЩЕНИЙ В ОЧЕРЕДИ

Так как $\pi(i)$ для всех $i \geq 0$ определяются системой (4), а для ненормированного значения $P(N_1, N_2) = V(N_1, 0)$ найдено значение, которое совпадает с $\pi_1(0) = V(N_1, 0)$, то ненормированные значения $\pi_1(i)$ составляют

$$(27) \pi_1(i) = \frac{\lambda}{(\mu + i\alpha)} \pi_1(i-1) = \pi_1(0) \prod_{v=1}^i \frac{\lambda}{(\mu + v\alpha)}.$$

Найдя нормирующий множитель

$$(28) d = \sum_{n_1=0}^{N_1} \sum_{n_2=0}^{N_2} V(n_1, n_2) + \sum_{i=0}^{\infty} \pi_1(i),$$

получим вероятности

$$(29) P(n_1, n_2) = \frac{1}{d} V(n_1, N_2 - n_2),$$

$$(30) \pi(i) = \frac{1}{d} \pi_1(i).$$

Таким образом, запишем рекуррентный векторный алгоритм.

Шаг 0: Определение векторов \mathbf{v}_0 и $\hat{\mathbf{v}}_0$ и скаляра $V(N_1, 1)$ согласно выражениям (9)–(13).

Шаг n_2 : Формирование матрицы \mathbf{A}_{n_2} с элементами (16).

Определение вектора $\hat{\mathbf{v}}_{n_2}$ согласно (19). Получение вектора \mathbf{v}_{n_2} вида (20) добавлением найденного на предыдущем шаге скаляра $V(N_1, n_2)$. Определение скаляра $V(N_1, n_2 + 1)$ согласно (21).

Данный шаг повторять для $n_2 = 1, \dots, N_2 - 1$.

Шаг N_2 : Определение матрицы \mathbf{A}_0 согласно (16). Получение вектора $\hat{\mathbf{v}}_{N_2}$ согласно (25) и вектора \mathbf{v}_{N_2} вида (26), добавляя скаляр $V(N_1, N_2)$, найденный на предыдущем шаге.

Шаг $N_2 + 1$: Определение ненормированных значений стационарных вероятностей числа сообщений в очереди $\pi(i)$, согласно равенству (27).

Шаг $N_2 + 2$: Нахождение нормирующего множителя согласно (28).

Шаг $N_2 + 3$: Определение стационарных вероятностей числа занятых ЕКР $P(n_1, n_2)$ и числа сообщений в очереди $\pi(i)$, согласно выражениям (29) и (30) соответственно.

4. Характеристики качества обработки сообщений

Выражения для вычисления показателей качества обработки сообщений следуют из их физического смысла и определяются через отношение интенсивностей анализируемых событий или суммирование стационарных вероятностей модели, к ним относятся: интенсивность потери сообщений; вероятность потери или доля сообщений, ушедших из очереди; вероятность положительной очереди; среднее количество сообщений в очереди; среднее время пребывания сообщения в очереди и в системе в целом; среднее число занятых ЕКР каждого из каналов.

При $\alpha > 0$ найдем интенсивность и вероятность потери заявок.

Интенсивность λ_1 потери заявок составляет:

$$(31) \lambda_1 = \sum_{i=0}^{\infty} \alpha i \pi(i).$$

Отношение интенсивности λ_1 потери сообщений и интенсивности λ поступления сообщений будем называть вероятностью P_0 потери сообщений или долей сообщений, ушедших из очереди:

$$(32) P_0 = \frac{\lambda_1}{\lambda} = \frac{\alpha}{\lambda} \sum_{i=0}^{\infty} i \pi(i).$$

При $\alpha = 0$ обозначим P_1 – вероятность положительной очереди, которая определяется равенством

$$(33) P_1 = \sum_{i=0}^{\infty} \pi(i).$$

Среднее количество сообщений в очереди определяется как

$$(34) \kappa = \sum_{i=0}^{\infty} i\pi(i).$$

Среднее время пребывания сообщения в очереди:

$$(35) T_o = \frac{\kappa}{\lambda}.$$

Тогда среднее время пребывания сообщения в системе:

$$(36) T_c = \frac{\kappa}{\lambda} + \frac{1}{\mu}.$$

Зная стационарное распределение вероятностей числа занятых ЕКР каналов $P(n_1, n_2)$, можно получить стационарные вероятности занятости каждого из каналов и их среднюю загрузку:

$$(37) P(n_1) = \sum_{n_2=0}^{N_2} P(n_1, n_2), \quad m_1 = \sum_{n_1=0}^{N_1} n_1 P(n_1);$$

$$(38) P(n_2) = \sum_{n_1=0}^{N_1} P(n_1, n_2), \quad m_2 = \sum_{n_2=0}^{N_2} n_2 P(n_2).$$

5. Пример численной реализации алгоритма

Проведем пример численной реализации разработанной модели мультипоточковой передачи данных, а также скалярно-векторного алгоритма для показателей качества обработки сообщений, полученных ранее: интенсивности потери сообщений; вероятности потери или доли сообщений, ушедших из очереди; вероятности положительной очереди; среднего количества сообщений в очереди; среднего времени пребывания сообщения в очереди и в системе в целом; среднего числа занятых ЕКР каждого из каналов.

Для этого примем следующие условия: время жизни пакетов $\alpha = 100$ мс, что соответствует трафику реального

времени, интенсивность прибытия пакетов λ составляет 4,3 Мбит/сек, что соответствует передаче видео с разрешением 1080 p (при применении H.264). Объем ресурсов сетей N_1 , N_2 варьируется от 1 Мбит/сек до 4 Мбит/сек (см. таблицу 1). Значения получены для стационарного режима.

Таблица 1. Численный расчет показателей качества передачи

Показатели качества передачи данных	Входные параметры модели			
	$\lambda = 4,3$ Мбит/сек			
	$N_1 = 1$ Мбит/сек	$N_1 = 1$ Мбит/сек	$N_1 = 2$ Мбит/сек	$N_1 = 4$ Мбит/сек
	$N_2 = 1$ Мбит/сек	$N_2 = 2$ Мбит/сек	$N_2 = 4$ Мбит/сек	$N_2 = 4$ Мбит/сек
Среднее количество сообщений в очереди	20	10,13	0,5741	0,1467
Интенсивность потерь сообщений	2	1,013	0,05741	0,0147
Вероятность потерь сообщений	50%	25%	1,4%	0,4%
Среднее время пребывания сообщения в СМО	5,499	2,866	0,3102	0,1617
Среднее время пребывания сообщения в очереди	4,999	2,533	0,1435	0,0367

6. Заключение

В настоящей статье представлена математическая модель мультипоточковой системы передачи данных в виде гетерогенной модели Эрланга с ожиданием и нетерпеливыми заявками. С помощью разработанного рекуррентного алгоритма получено распределение вероятностей (числа занятых приборов в каждом блоке и числа заявок, находящихся в очереди) и найдены технические характеристики системы, имеющие практическое значение для проектирования реальных информационных и телекоммуникационных систем. Вместе с тем следует отметить, что представленный алгоритм работает только для марковской модели, поэтому альтернативным подходом является применение метода асимптотического

анализа, что позволяет найти асимптотическое распределение вероятностей числа заявок. В дальнейших работах также планируется обобщение результатов для модели с разноприоритетными входящими потоками.

Литература

1. ПАКУЛОВА Е.А. *Алгоритмы и программные средства повышения эффективности передачи мультимедийных данных в беспроводных компьютерных сетях.* – Дисс. канд. техн. наук. – С.-Пб., 2016.
2. BOXMA O.J., COHEN J.W., HUFFELS N. *Approximations of the mean waiting time in an M/G/s queueing system* // Operations Research. – 1979. – No. 27. – P. 1115–1127.
3. CHEBROLU K., RAO R.R. *Bandwidth aggregation for real-time applications in heterogeneous wireless networks* // IEEE Trans. on Mobile Computing. – 2006. – Vol. 5. – No. 4. – P. 388–403.
4. CONINCK Q., BONAVENTURE O. *Multipath tester: Comparing mptcp and mpquic in mobile environments* // IEEE Network Traffic Measurement and Analysis Conference (TMA-2019). – 2019. – P. 221–226.
5. DE CONINCK Q., BONAVENTURE O. *Multipath Extensions for QUIC (MP-QUIC)* [электронный ресурс] <https://datatracker.ietf.org/doc/draft-deconinck-quic-multipath/>.
6. DE CONINCK Q., BONAVENTURE O. *Multipath quic: Design and evaluation* // Proc. of the 13th Int. Conf. on Emerging Networking Experiments and Technologies. – 2017. – P. 160–166
7. ERLANG A. *Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges* // The Post Office Electrical Engineers' Journal. – 1918. – No. 10. – P. 189–197.
8. ERLANG A. *The theory of probabilities and telephone conversations* // Nyt Tidsskrift for Matematik. – 1909. – 20(B). – P. 33–39.
9. FORD A. *TCP extensions for multipath operation with multiple addresses* // RFC 6824. – 2013. – P. 1–64.
10. GUMBEL H. *Waiting lines with heterogeneous servers* // Operations Research. – 1960. – No. 8(4). – P. 504–511.

11. KIMURA T. *Diffusion approximation for an M/G/m queue* // Operations Research. – 1983. – No. 31. – P. 304–321.
12. LISOVSKAYA E.Y., MOISEEVA S.P. *Study of the Queuing Systems M/GI/N/∞* // Communications in Computer and Information Science. – 2015. – No. 564. – P. 175–185.
13. NAZAROV A., MOISEEV A., MOISEEVA S. *Mathematical Model of Call Center in the Form of Multi-Server Queueing System* // Mathematics. – 2021. – 9(22):2877. – DOI: <https://doi.org/10.3390/math9222877>
14. SHORE H. *Approximations for the GI/G/c Queue-I: The Steady-State Probabilities* // Journal of the Operational Research Society. – 1988. – No. 39. – P. 279–284.
15. NIGHTINGALE J., WANG Q., GRECOS C. *Optimised transmission of H. 264 scalable video streams over multiple paths in mobile networks* // IEEE Trans. on Consumer Electronics. – 2010. – Vol. 56. – No. 4. – P. 2161–2169.
16. NOZAKI S.A., ROSS S.M. *Approximations in finite-capacity multi-server queues with Poisson arrivals* // Journal of Applied Probability. – 1978. – Vol. 15, Iss. 4. – P. 826–834.
17. PAKULOVA E., MILLER K., WOLISZ A. *Adaptive low-delay video streaming in heterogeneous wireless networks using MPRTTP* // IEEE 13th International Wireless Communications and Mobile Computing Conference (IWCMC-2017). – 2017. – P. 14–19.
18. SINGH V.P. *Markovian queues with three heterogeneous servers* // AIIE Transactions. – 1971. – Vol. 3 (1). – P.45–48.
19. SINGH V.P. *Two-server Markovian queues with balking: Heterogeneous vs. homogeneous servers* // Operations Research. – 1970. – No. 18(1). – P. 145–159.
20. STEWART R. *Stream control transmission protocol* // RFC 2960. – 2000. – 134 p.
21. WHITT W. *Comparison conjectures about the M/G/s queue* // Operations Research Letters. – 1983. – Vol. 2. – P. 203–209.
22. Патент РФ № 2636665, 20.02.2017.

**SCALAR-VECTOR RECURRENT ALGORITHM FOR
STATIONARY PROBABILITIES IN A HETEROGENEOUS
SYSTEM $M/(M_1, M_2)/(N_1, N_2)/\infty/\text{FIFO}$**

Anatoliy Nazarov, National Research Tomsk State University, Tomsk, Doctor of Science, professor (nazarov.tsu@gmail.com).

Artem Rindin, Southern Federal University, Taganrog, PhD student (artem_ryndin@mail.ru).

Ekaterina Pakulova, Southern Federal University, Taganrog, Cand.Sc. (epakulova@sfnu.ru).

Irina Turenova, National Research Tomsk State University, Tomsk, Cand.Sc., assistant professor (irenaturena@mail.ru).

Svetlana Moiseeva, National Research Tomsk State University, Tomsk, Doctor of Science, professor (smoiseeva@mail.ru).

Abstract: This article proposes a generalized Erlang model with queue, the peculiarity of which is that two heterogeneous channels of finite volume and different service intensity are provided for data processing and transmission. When a message arrives, it first turns to a "faster" channel, if it has a sufficient number of channel resource units, then the message gets serviced. Otherwise, if there is not enough free service resource, the message is sent to the second "slow" channel. If it has a sufficient amount of free service resource, then the message gets serviced. In the opposite case, the message gets into the queue. Messages in the queue have a lifetime, after which it makes no sense to transmit them. An original scalar-vector algorithm for calculating stationary probabilities of system states is proposed. The analysis of the characteristics of the quality of service is carried out. Namely, channel loading, average time spent in the system, the probability of instant service, the characteristics of service delays and the average number of service failures.

Keywords: heterogeneous queuing system, Erlang problem.

УДК 519.872

ББК 22.18

DOI: 10.25728/ubs.2022.98.1

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

Поступила в редакцию 01.07.2022.

Опубликована 31.07.2022.