

ПРИБЛИЖЕННЫЙ АНАЛИЗ ДЛИТЕЛЬНОСТИ ПРЕБЫВАНИЯ ТРЕБОВАНИЙ В СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ С ДЕЛЕНИЕМ И СЛИЯНИЕМ ТРЕБОВАНИЙ

Постнова О. С.¹, Тананко И. Е.², Рогачко Е. С.³

(Саратовский национальный исследовательский
государственный университет имени Н.Г. Чернышевского,
Саратов)

Рассматривается открытая сеть массового обслуживания с делением и слиянием требований. При поступлении в сеть каждое требование разделяется на два фрагмента. Фрагменты требований обслуживаются в одноприборных системах обслуживания с неограниченной очередью, образующих ациклическую структуру рассматриваемой сети. После завершения обслуживания в сети фрагменты требований поступают в выделенную систему сборки требований, где требования собираются из своих фрагментов и покидают сеть. Предлагается приближенный метод вычисления математического ожидания длительности сборки требований и математического ожидания длительности пребывания требований в сети массового обслуживания с делением и слиянием требований при предположениях, что входящий в сеть поток требований является пуассоновским, а длительности обслуживания фрагментов требований являются экспоненциально распределенными случайными величинами. Для оценки точности приближенного метода вычисления соответствующих характеристик сети полученные результаты сравниваются с результатами имитационного моделирования сети. Предложенный метод может применяться при невысокой нагрузке сети массового обслуживания, функционирующей в стационарном режиме. Представленная сеть массового обслуживания может использоваться в качестве моделей современных многопроцессорных вычислительных систем, а также других систем с параллельным и распределенным принципом функционирования.

Ключевые слова: сеть массового обслуживания с делением и слиянием требований, ациклическая сеть массового обслуживания, обслуживание фрагментов требований, приближенный анализ, длительность пребывания требований в сети.

¹ Оксана Сергеевна Постнова, ассистент (oksana.karpenko.2000@mail.ru).

² Игорь Евстафьевич Тананко, к.ф.-м.н., доцент (tanankoie@info.sgu.ru).

³ Екатерина Сергеевна Рогачко, к.ф.-м.н., доцент (rogachkoes@info.sgu.ru).

1. Введение

Одним из способов оперативной обработки больших объемов мало связанной информации является разделение этой информации на части с дальнейшей их одновременной обработкой высокопроизводительными вычислительными системами.

В качестве математических моделей систем параллельной или распределенной обработки информации для решения задач эффективного их использования часто используют системы и сети массового обслуживания с делением и слиянием требований или, как они называются в англоязычной литературе, «fork-join queueing systems». Поступая в такую систему, требование разделяется на фрагменты, которые затем занимают приборы и обслуживаются. Требование считается обслуженным, когда завершится обслуживание последнего из родственных фрагментов требования. Поэтому одним из основных результатов анализа систем fork-join с различной структурой и особенностями функционирования считается нахождение математического ожидания (м.о.) длительности пребывания требований в системе [11, 12].

Первые результаты исследования системы fork-join были представлены для случая разделения требования на два фрагмента с последующим одновременным началом их обработки двумя приборами с различающимися параметрами экспоненциальной функции распределения длительностей обслуживания в работе [7]. Верхние и нижние оценки м.о. длительности пребывания требований в системе с произвольным числом параллельных приборов, пуассоновским входящим потоком и экспоненциальной длительностью обслуживания фрагментов требований получены в работе [15].

Естественным развитием теории и методов анализа систем обслуживания типа fork-join является исследование двух и более параллельно связанных систем массового обслуживания с очередями, образующих сеть массового обслуживания с отдельной обработкой фрагментов требований. Некоторые представления об основных структурах, свойствах и характеристиках таких математических объектов можно получить на основе работ [6, 20].

В [3] получены условие существования стационарного режима и аналитические выражения для нижней и верхней границ м.о. длительности пребывания в fork-join сети из двух параллельных систем обслуживания с неоднородными приборами и общими предположениями о распределении длительностей обслуживания фрагментов.

Для оценки математического ожидания и дисперсии времени реакции fork-join сети обслуживания с различными типами функций распределения длительностей обслуживания фрагментов предлагается подход [9, 10, 21], использующий методы машинного обучения, методы оптимизации и имитационное моделирование. Дальнейшие исследования выявили зависимость между длительностью пребывания фрагментов в системах сети обслуживания и временем реакции этой сети [8].

Работы [18, 19] посвящены изучению fork-join сетей параллельных систем массового обслуживания в условиях высокой загрузки. Например, для сети обслуживания с дискретным временем [19] получен предел при числе систем, стремящемся к бесконечности, непрерывной функции, зависящей от максимально возможной загрузки сети обслуживания. В работе [18] решается задача определения функции распределения времени реакции fork-join сети обслуживания с использованием верхней и нижней границ рабочей нагрузки, определяемой как объем работы, которую должен выполнить прибор системы обслуживания, прежде чем он станет свободным при отсутствии поступления очередного фрагмента.

В работах [14, 17] решается задача оптимального распределения фиксированного количества ресурсов между параллельными системами fork-join сети, в которой обслуживаются фрагменты однородных [14] или неоднородных [17] требований. Предложенные алгоритмы распределения ресурсов минимизируют время реакции рассматриваемой сети обслуживания. В работе [16] решается задача синтеза структуры и алгоритма обслуживания требований fork-join сети массового обслуживания с заданными средним и дисперсией времени реакции.

Рассмотренные fork-join системы и сети массового обслуживания имеют параллельную структуру и могут использоваться в качестве математических моделей многопроцессорных вычислительных систем, оперативной обработки больших объемов информации, поддержки принятия решений с использованием машинного обучения и нейронных сетей. При этом актуальной остается задача анализа и эффективного использования систем передачи информации, особенно сетей с коммутацией пакетов и систем многопутевой маршрутизации. Связь между источником и получателем данных в таких системах осуществляется посредством сети передачи данных с произвольной топологией. В качестве математической модели таких систем может использоваться сеть массового обслуживания произвольной топологии, состоящая из бесконечноприборных систем обслуживания [4].

В данной работе рассматривается открытая сеть массового обслуживания с делением и слиянием требований, включающая системы массового обслуживания, объединенные в ациклическую сеть обслуживания, и выделенную систему сборки фрагментов требований. Эта сеть может быть использована в качестве математической модели, например, распределенной системы обработки и передачи данных, характеризующейся разделением данных на пакеты. Такой способ передачи данных отличается достаточно равномерной загрузкой всех технических устройств на широком диапазоне значений внешней нагрузки на систему. Для рассматриваемой сети массового обслуживания выполняются следующие условия. Входящий в сеть поток требований является пуассоновским. Поступающие требования делятся на два фрагмента. Длительности обслуживания фрагментов требований во всех системах имеют экспоненциальное распределение. Описывается модель системы сборки фрагментов требований. Решается задача приближенного вычисления математических ожиданий длительности сборки требований и длительности пребывания требований в сети. Для оценки точности приближения полученных результатов используется их сравнение с результатами имитационного моделирования рассматриваемой сети.

2. Описание модели

Рассматривается сеть массового обслуживания N^f с делением и слиянием требований. Сеть N^f включает L систем обслуживания S_i типа $M/M/1$ с интенсивностями обслуживания μ_i , $i = 1, \dots, L$, и выделенную систему S_{L+1} , называемую системой сборки.

Из источника S_0 в сеть обслуживания с интенсивностью λ_0 поступает пуассоновский поток требований одного класса. В момент поступления требования в сеть это требование разделяется на два фрагмента, называемых родственными, и оба фрагмента независимо друг от друга поступают в одну из систем сети S_i , $i = 1, \dots, L$, с которыми связан источник.

Фрагменты одного требования независимо друг от друга и независимо от фрагментов других требований переходят между системами сети S_i , $i = 1, \dots, L$, до тех пор, пока не завершится их обслуживание в сети. После этого фрагменты требований поступают в систему сборки S_{L+1} . Родственные фрагменты, ранее принадлежавшие одному требованию, собираются в единое требование, которое покидает систему сборки и сеть N^f . Переходы фрагментов между системами обслуживания сети, источником и системой сборки осуществляются в соответствии с маршрутной матрицей $\Theta = (\theta_{ij})$, $i = 0, 1, \dots, L$, $j = 1, \dots, L + 1$, где θ_{ij} – вероятность того, что после пребывания в системе S_i фрагмент перейдет в S_j .

Сеть массового обслуживания с введенными ранее параметрами систем, в которой обслуживаются фрагменты требований, без системы сборки обозначим через N . Предполагается, что сеть N – ациклическая, т.е. структура сети представляется ориентированным ациклическим графом. Элементы маршрутной матрицы $\bar{\Theta} = (\bar{\theta}_{ij})$, $i, j = 0, 1, \dots, L$, сети N определяются как $\bar{\theta}_{ij} = \theta_{ij}$, если $j \neq 0$, и $\bar{\theta}_{i,0} = \theta_{i,L+1}$. Предполагается, что в сети N число смежных для S_i , $i = 0, 1, \dots, L$, (кроме S_i , для которых $\bar{\theta}_{i,0} = 1$) систем обслуживания S_j , в которые возможен переход фрагментов из S_i , намного больше двух (числа фрагментов

требования), и вероятности $\bar{\theta}_{ij} > 0$, $j = 0, 1, \dots, L$, сравнимы. В этом случае выходящий из сети N поток фрагментов требований можно считать пуассоновским [5] с интенсивностью $2\lambda_0$.

Для сети N необходимым условием существования стационарного режима является

$$(1) \quad \lambda_0 < \frac{1}{2} \min_{i=1, \dots, L} \frac{\mu_i \omega_0}{\omega_i} = a,$$

где $\omega = (\omega_j)$, $j = 0, 1, \dots, L$, – вектор относительных интенсивностей потоков фрагментов требований в сети N – находится как решение уравнения $\omega = \omega \Theta$ с условием нормировки $\sum \omega_j = 1$.

Назначение системы сборки S_{L+1} – сбор требований из родственных фрагментов, поступающих из сети N . Первый из двух поступающих в систему сборки родственных фрагментов будем называть «первым» фрагментом требования. В момент поступления «второго» из родственных фрагментов мгновенно происходит объединение фрагментов в единое требование, которое покидает систему сборки. Таким образом, длительность сборки требования из родственных фрагментов определяется как время пребывания в системе сборки «первого» фрагмента, то есть длительность интервала времени между моментами выхода родственных фрагментов из сети N .

3. Приближенный метод анализа

Математическое ожидание $E\tau_0^f$ длительности пребывания требований в сети N^f определяется как сумма математического ожидания длительности пребывания требований в сети обслуживания N и математического ожидания длительности сборки требований из фрагментов в системе сборки S_{L+1} . Для вычисления последней величины опишем приближенную модель системы сборки требований.

Определим параметры системы сборки в соответствии с символикой Дж. Кендалла [1].

Заметим, что если в систему сборки поступает фрагмент, который не имеет среди уже находящихся фрагментов в этой системе родственного фрагмента, то вновь прибывший фрагмент

занимает свободный прибор. Следовательно, система сборки содержит неограниченное число приборов.

Известно, что выходящий из сети N поток фрагментов требований является пуассоновским [5]. Поэтому в систему сборки также поступает пуассоновский поток фрагментов с интенсивностью λ_0 , так как обслуживаются в системе только «первые» фрагменты.

Результаты имитационного моделирования показывают [2], что длительность сборки требования из родственных фрагментов можно представить экспоненциально распределенной случайной величиной с математическим ожиданием $1/\mu$, а также то, что изменение параметра μ приближенно описывается убывающей линейной функцией интенсивности λ_0 .

Таким образом, систему сборки представим в нашей модели системой массового обслуживания типа $M/M/\infty$.

Определим функцию $\mu = \mu(\lambda_0)$, $\lambda_0 \in (0, a)$. Графиком функции $\mu(\lambda_0)$ является прямая, проходящая через две точки $(0, \nu)$ и $(a, 0)$, где ν – максимальное значение μ , получаемое при $\lambda_0 = \varepsilon$, где $\varepsilon > 0$ – малое число. Действительно, при $\lambda_0 \rightarrow 0$ требование, поступающее в сеть N^f , с вероятностью, стремящейся к 1, заставит сеть пустой. Тогда оба фрагмента требования успевают выйти из сети N , прежде чем в сеть поступит новое требование, и длительность ожидания в системе сборки «первым» фрагментом требования «второго» фрагмента будет минимальной. В этом случае параметр μ , очевидно, принимает максимальное значение ν . Следовательно,

$$(2) \quad \mu = -\frac{\nu}{a}\lambda_0 + \nu, \quad \lambda_0 \in (0, a).$$

Найдем ν . При $\lambda_0 = \varepsilon$ вероятность того, что между двумя поступающими в систему сборки родственными фрагментами требования поступит фрагмент или фрагменты других требований, близка к нулю. Тогда математическое ожидание длительности сборки требований из фрагментов или, иначе, математическое ожидание длительности интервала времени между родственными фрагментами в выходящем из сети N потоке фрагментов совпадает со среднеквадратическим отклонением σ_0 длительно-

сти реакции τ_0 сети N (длительности пребывания фрагментов требований в сети N). Тогда

$$(3) \quad \nu = 1/\sigma_0, \quad \sigma_0 = \sqrt{E\tau_0^2 - (E\tau_0)^2},$$

где $E\tau_0$ и $E\tau_0^2$ – первый и второй начальные моменты величины τ_0 .

Из формулы Литтла [1] для сети N следует, что

$$(4) \quad E\tau_0 = \frac{q}{2\lambda_0}, \quad q = \sum_{i=1}^L \frac{\psi_i}{1 - \psi_i},$$

где q – математическое ожидание числа фрагментов требований в сети N , $\psi_i = \lambda_i/\mu_i$ – коэффициент использования системы S_i , $\lambda_i = 2\lambda_0\omega_i/\omega_0$ – интенсивность потока фрагментов требований в систему S_i , $i = 1, \dots, L$.

Чтобы найти второй момент $E\tau_0^2$, введем в рассмотрение τ_i – длительность пребывания в сети N фрагментов требований, поступивших в сеть через систему S_i , $i = 1, \dots, L$, с первым и вторым начальными моментами $E\tau_i$ и $E\tau_i^2$ соответственно, и представим $E\tau_0^2$ в виде:

$$(5) \quad E\tau_0^2 = \sum_{i=1}^L \bar{\theta}_{0i} E\tau_i^2.$$

Известно [13], что величины $E\tau_i$ могут быть найдены из системы уравнений:

$$(6) \quad E\tau_i = \frac{1}{\mu_i(1 - \psi_i)} + \sum_{j=1}^L \bar{\theta}_{ij} E\tau_j, \quad i = 1, \dots, L,$$

а величины $E\tau_i^2$ – из следующей системы уравнений:

$$(7) \quad E\tau_i^2 = \frac{2E\tau_i}{\mu_i(1 - \psi_i)} + \sum_{j=1}^L \bar{\theta}_{ij} E\tau_j^2, \quad i = 1, \dots, L.$$

Введем обозначения векторов и матриц: $\mathbf{E}\boldsymbol{\tau} = (E\tau_i)$, $\mathbf{u} = (1/(\mu_i - \lambda_i))$, $i = 1, \dots, L$, $\bar{\boldsymbol{\Theta}}' = (\bar{\theta}_{ij})$, $i, j = 1, \dots, L$; $\mathbf{I}$ – единичная матрица порядка L , тогда из системы уравнений (6) имеем

$$\mathbf{E}\boldsymbol{\tau} = \mathbf{u}(\mathbf{I} - \bar{\boldsymbol{\Theta}}')^{-1}.$$

Умножив уравнения (7) на λ_i и просуммировав их по i , $i = 1, \dots, L$, получим

$$\sum_{i=1}^L \lambda_i E\tau_i^2 = 2 \sum_{i=1}^L \frac{\lambda_i}{\mu_i(1-\psi_i)} E\tau_i + \sum_{i=1}^L \lambda_i \sum_{j=1}^L \bar{\theta}_{ij} E\tau_j^2,$$

$$\sum_{i=1}^L \lambda_i E\tau_i^2 = 2 \sum_{i=1}^L \frac{\lambda_i}{\mu_i(1-\psi_i)} E\tau_i + \sum_{j=1}^L (\lambda_j - 2\lambda_0 \bar{\theta}_{0j}) E\tau_j^2,$$

или

$$2\lambda_0 \sum_{i=1}^L \bar{\theta}_{0i} E\tau_i^2 = 2 \sum_{i=1}^L \frac{\lambda_i}{\mu_i(1-\psi_i)} E\tau_i,$$

откуда с учетом (5) имеем

$$(8) \quad E\tau_0^2 = \frac{1}{\lambda_0} \sum_{i=1}^L \frac{\psi_i}{1-\psi_i} E\tau_i.$$

Окончательно, подставив найденные по формулам (4) и (8) значения в формулу (3), найдем значение ν , используемое в формуле (2).

Обозначим P_k – стационарная вероятность пребывания системы сборки в состоянии k , $k = 0, 1, 2, \dots$, где k – число фрагментов в системе сборки. Имеем

$$P_k = \frac{(\lambda_0/\mu)^k}{k!} e^{-\lambda_0/\mu}, \quad k = 0, 1, 2, \dots,$$

где μ вычисляется по формуле (2). Математическое ожидание числа фрагментов требований в системе сборки $n = \lambda_0/\mu$.

Определим математическое ожидание $E\tau_0^f$ длительности пребывания требований в сети N^f . Очевидно, что математическое ожидание числа требований в сети N^f равно $(q + n)/2$. С учетом формулы Литтла и найденного математического ожидания $1/\mu$ длительности сборки требований величина $E\tau_0^f$ приближенно определяется выражением

$$(9) \quad E\tau_0^f = \frac{q - n}{2\lambda_0} + 1/\mu.$$

4. Результаты моделирования и точность приближения

Для проверки точности полученных аналитических приближенных результатов была разработана имитационная модель сети N^f [2]. В этой модели реализовано разделение требований на два фрагмента при поступлении требований в сеть, независимая маршрутизация фрагментов в сети N , ожидание «первыми» фрагментами в системе сборки поступления именно родственных «вторых» фрагментов требований.

Рассматривается сеть N^f с $L = 30$ системами обслуживания при различных топологиях и маршрутных матрицах сети N : сеть параллельных систем в три слоя (тип сети « p »): $\bar{\theta}_{0i} = \bar{\theta}_{ij} = \bar{\theta}_{jk} = 0,1$, $\bar{\theta}_{k0} = 1$, $i = 1, \dots, 10$, $j = 11, \dots, 20$, $k = 21, \dots, 30$; сеть с древовидной структурой в три слоя (тип сети « t »): $\bar{\theta}_{0i} = 0,1$, $i = 1, \dots, 10$, $\bar{\theta}_{10} = \bar{\theta}_{20} = 1$, $\bar{\theta}_{ij} = 0,1$, $i = 3, \dots, 10$, $j = 11, \dots, 20$, $\bar{\theta}_{19,0} = \bar{\theta}_{20,0} = 1$, $\bar{\theta}_{jk} = 0,1$, $\bar{\theta}_{k0} = 1$, $j = 11, \dots, 18$, $k = 21, \dots, 30$.

Приводятся результаты вычислительных экспериментов при двух вариантах векторов интенсивностей обслуживания в системах сети N : вектор одинаковых интенсивностей обслуживания (вектор « e »): $\mu_i = 4$, $i = 1, \dots, 30$; вектор различных интенсивностей обслуживания (вектор « d »): $\mu_i = \mu_{i+10} = \mu_{i+20} = 4 + 0,1(i-1)$, $i = 1, \dots, 10$. При таких параметрах сети N согласно условию (1) существования стационарного режима $a = 20$.

Будет использоваться обозначение сети N (тип топологии, тип вектора интенсивностей обслуживания), где «тип топологии» принимает значения из $\{p, t\}$, а «тип вектора интенсивностей обслуживания» принимает значения из $\{e, d\}$.

На рис. 1 представлена интенсивность обслуживания μ в системе сборки при значениях $0 < \lambda_0 < a$, удовлетворяющих условию (1), по результатам имитационного моделирования (ИМ) и результатам аппроксимации (АП) по формуле (2). Результаты имитационного моделирования здесь и далее получены с доверительной вероятностью 0,95 и шириной доверительного интервала 0,01.

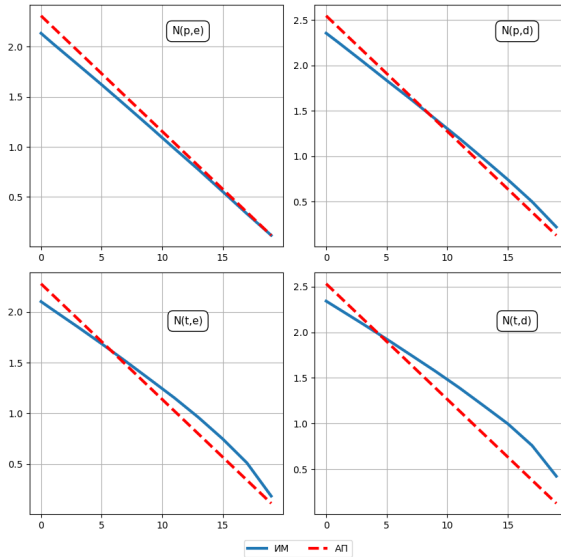


Рис. 1. Интенсивность обслуживания в системе сборки в зависимости от λ_0

Введем величину $\delta = |\tilde{\mu} - \mu|/\tilde{\mu}$, где $\tilde{\mu}$, μ – оценки интенсивности обслуживания в системе сборки, полученные в результате имитационного моделирования и аппроксимации соответственно. Из рис. 1 видно, что для сетей с различными параметрами при $\lambda_0 \rightarrow a$, когда $\mu \rightarrow 0$, результаты имитационного моделирования и аппроксимации могут существенно различаться и погрешность аппроксимации, определяемая величиной δ , может быть большой. Например, при $\lambda_0 \geq 0,55a$ для сетей $N(t,*)$ имеем $\delta > 0,1$. При указанных значениях λ_0 выполняется условие $\lambda_0 \gg \mu$ или, иначе, $\bar{n} \gg 1$. Далее случай $\lambda_0 > 0,55a$ не рассматривается, так как в этом случае наблюдается низкая точность аппроксимации μ .

В таблице 1 представлены оценки $\tilde{n}$ и n математического ожидания числа фрагментов требований в системе сборки, полученные в результате имитационного моделирования и аппроксимации соответственно, и величина $\delta = |\tilde{n} - n|/\tilde{n}$ при λ_0 , изменяющейся от 0,05a до 0,55a с шагом 0,1a.

Таблица 1. Математическое ожидание числа фрагментов требований в системе сборки

Тип сети	λ_0	$N(*, e)$			$N(*, d)$		
		$\tilde{n}$	n	δ	$\tilde{n}$	n	δ
$N(p, *)$	1,0	0,493	0,456	0,075	0,444	0,413	0,069
	3,0	1,643	1,528	0,070	1,467	1,385	0,056
	5,0	3,087	2,887	0,065	2,727	2,617	0,040
	7,0	4,965	4,663	0,061	4,308	4,227	0,019
	9,0	7,517	7,086	0,057	6,375	6,423	0,007
	11,0	11,200	10,585	0,055	9,199	9,594	0,043
$N(t, *)$	1,0	0,496	0,462	0,067	0,443	0,416	0,061
	3,0	1,617	1,550	0,041	1,432	1,395	0,026
	5,0	2,966	2,928	0,013	2,596	2,634	0,015
	7,0	4,630	4,730	0,022	4,002	4,256	0,063
	9,0	6,758	7,187	0,064	5,715	6,466	0,131
	11,0	9,530	10,736	0,127	7,912	9,660	0,221

Рис. 2 показывает результаты оценки плотности распределения вероятностей длительности обслуживания фрагментов требований в системе сборки. Результаты приведены для сетей рассмотренных типов при $\lambda_0 = 0,55a$. Из результатов имитационного моделирования следует, что длительность сборки требований – экспоненциально распределенная случайная величина с параметром μ . Соответствующие расстояния Колмогорова указываются далее в таблице 2.

По результатам имитационного моделирования число фрагментов требований в системе сборки имеет распределение Пуассона. На рис. 3 показаны результаты для сетей рассмотренных типов при $\lambda_0 = 0,25a$.

В таблице 2 приводятся расстояния Колмогорова для оценки точности аппроксимации распределения вероятностей длительности сборки и распределения вероятностей числа фрагментов в системе сборки. Расстояние Колмогорова для оценки точности аппроксимации распределения вероятностей длительности сборки определяется

$$\Delta_1 = \sup_x |F(x) - \tilde{F}(x)|,$$

где $F(x)$ – функция распределения экспоненциально распределенной случайной величины с найденным параметром μ , $\tilde{F}(x)$ – эмпирическая функция распределения по результатам имитационного моделирования. Расстояние Колмогорова для оценки точности аппроксимации распределения вероятностей числа фрагментов в системе сборки определяется

$$\Delta_2 = \max_{i \geq 0} \left| \sum_{k=0}^i P_k - \sum_{k=0}^i \tilde{P}_k \right|,$$

где P_k и $\tilde{P}_k$ – вероятности состояния k , $k = 0, 1, 2, \dots$, системы сборки, вычисленные аналитически и по результатам имитационного моделирования соответственно.

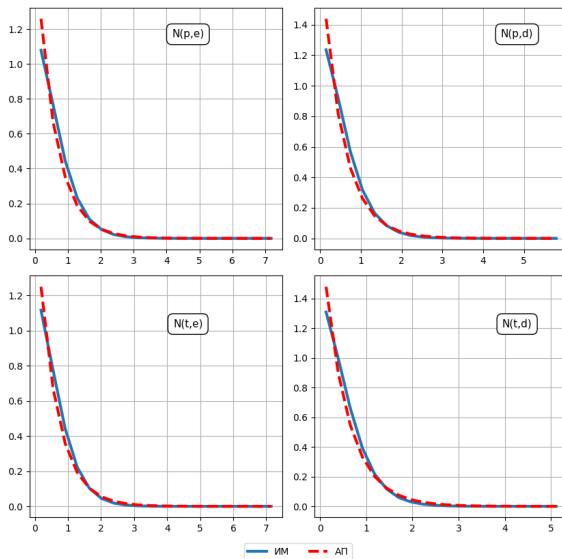


Рис. 2. Плотность распределения вероятностей длительности обслуживания фрагментов в системе сборки при $\lambda_0 = 0,55a$

На основе приведенных результатов можно сделать вывод, что полученная в работе приближенная формула (2) дает достаточно точную точность аппроксимации при значениях $0 < \lambda_0 < 0,5a$, т.е. для сетей N с невысокой нагрузкой.

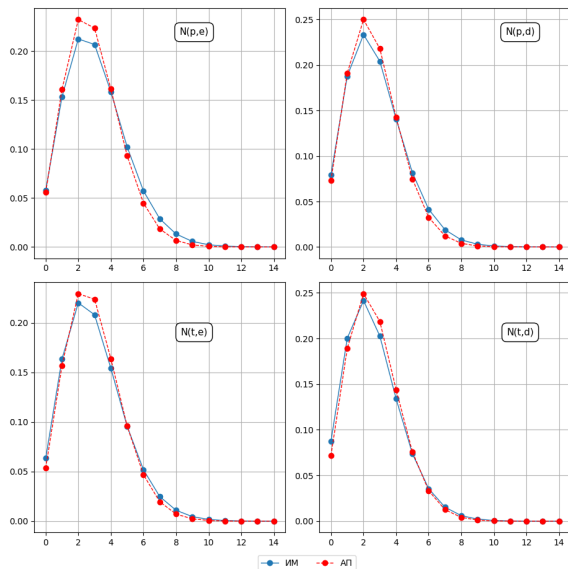


Рис. 3. Распределение вероятностей числа фрагментов требований в системе сборки при $\lambda_0 = 0,25a$

Таблица 2. Расстояния Колмогорова

Тип сети	λ_0	Δ_1		Δ_2	
		$N(*, e)$	$N(*, d)$	$N(*, e)$	$N(*, d)$
$N(p, *)$	1,0	0,082	0,078	0,019	0,016
	3,0	0,078	0,073	0,033	0,025
	5,0	0,074	0,066	0,045	0,030
	7,0	0,071	0,058	0,058	0,031
	9,0	0,070	0,048	0,074	0,053
	11,0	0,068	0,035	0,092	0,098
$N(t, *)$	1,0	0,078	0,075	0,017	0,014
	3,0	0,069	0,063	0,021	0,014
	5,0	0,056	0,049	0,018	0,027
	7,0	0,046	0,034	0,046	0,068
	9,0	0,032	0,048	0,095	0,136
	11,0	0,046	0,071	0,174	0,240

В таблице 3 представлены оценки $E\tilde{\tau}_0^f$ и $E\tau_0^f$ математического ожидания длительности пребывания требований в се-

ти N^f , полученные в результате имитационного моделирования и аппроксимации по формуле (9) соответственно, и величина $\delta = |E\tilde{\tau}_0^f - E\tau_0^f|/E\tau_0^f$ при λ_0 , изменяющейся от 0,05а до 0,55а с шагом 0,1а.

Таблица 3. Математическое ожидание длительности пребывания требований в сети N^f

Тип сети	λ_0	$N(*, e)$			$N(*, d)$		
		$E\tilde{\tau}_0^f$	$E\tau_0^f$	δ	$E\tilde{\tau}_0^f$	$E\tau_0^f$	δ
$N(p, *)$	1,0	1,060	1,017	0,040	0,953	0,916	0,039
	3,0	1,183	1,137	0,039	1,053	1,014	0,036
	5,0	1,338	1,289	0,037	1,173	1,137	0,031
	7,0	1,542	1,487	0,036	1,328	1,294	0,025
	9,0	1,821	1,757	0,035	1,535	1,502	0,021
	11,0	2,225	2,148	0,035	1,814	1,792	0,012
$N(t, *)$	1,0	0,904	0,868	0,040	0,811	0,780	0,038
	3,0	0,989	0,957	0,032	0,879	0,854	0,028
	5,0	1,093	1,067	0,024	0,961	0,945	0,017
	7,0	1,225	1,206	0,015	1,062	1,058	0,004
	9,0	1,396	1,391	0,004	1,189	1,205	0,013
	11,0	1,630	1,648	0,011	1,358	1,404	0,034

Из приведенных результатов видно, что погрешность вычислений по формуле (9) не превосходит 4%. Это свидетельствует о применимости приближенной формулы для оценки математического ожидания длительности пребывания требований в сети N^f при невысокой нагрузке сети.

5. Заключение

Предложен приближенный метод вычисления математического ожидания длительности пребывания требований в сети массового обслуживания с делением и слиянием требований. Особенность рассматриваемой сети состоит в том, что, во-первых, поступающие в сеть требования делятся на два фрагмента. Во-вторых, обслуживание фрагментов требований осуществляется не в параллельных системах обслуживания, а в ациклической сети систем обслуживания. После завершения обслуживания

ния фрагментов их объединение в требования выполняется в системе сборки требований. Описаны приближенная модель системы сборки и способ вычисления ее стационарных характеристик. Темой дополнительных исследований может быть определение соотношения длительности сборки и длительности пребывания требований в сети.

Разработанный метод может быть использован для решения задач анализа и оптимизации систем многопутевой маршрутизации, систем последовательно-параллельной обработки информации, производственных систем, моделируемых сетями обслуживания с делением и слиянием требований. Метод наиболее применим для вычисления характеристик систем с нагрузкой не выше средней, но может быть также использован для выявления потенциально узких мест в системах.

Литература

1. ВИШНЕВСКИЙ В.М. *Теоретические основы проектирования компьютерных сетей*. – М.: Техносфера, 2003. – 512 с.
2. КАРПЕНКО О.С., ТАНАНКО И.Е., РОГАЧКО Е.С. *Исследование имитационной модели открытой сети массового обслуживания с делением и слиянием требований* // Системы управления, информационные технологии и математическое моделирование : Материалы V Всероссийской научно-практической конференции с международным участием. – Омск: ОмГТУ, 2023. – С. 209–214.
3. КЛИМЕНОК В.И. *Характеристики производительности системы массового обслуживания с расщеплением запросов* // Информатика. – 2023. – Т. 20, №3. – С. 50–60.
4. ОСИПОВ О.А., ТАНАНКО И.Е. *Сети массового обслуживания произвольной топологии с делением и слиянием требований: случай бесконечноприборных систем обслуживания* // Вестник Тверского государственного университета. Серия: Прикладная математика. – 2017. – №4. – С. 43–58.
5. ЦИЦИАШВИЛИ Г.Ш., ОСИПОВА М.А. *Стационарные потоки в ациклических сетях массового обслуживания* //

- Дальневосточный математический журнал. – 2016. – №2. – С. 223–228.
6. ENGANTI P., ROSENKRANTZ T., SUN L. et al. *ForkMV: Mean-and-variance estimation of fork-join queuing networks for datacenter applications* // IEEE Int. Conf. on Networking, Architecture and Storage (NAS-2022). – Philadelphia, PA, USA, 2022. – P. 1–8.
 7. FLATTO L., HAHN S. *Two parallel queues created by arrivals with two demands I* // SIAM Journal on Applied Mathematics. – 1984. – Vol. 44, No. 5. – P. 1041–1053.
 8. GORBUNOVA A.V., LEBEDEV A.V. *Correlations of the sojourn times of subtasks in fork-join queueing systems with $M/M/1$ -type subsystems* // Advances in Systems Science and Applications. – 2024. – Vol. 24, No. 2. – P. 1–18.
 9. GORBUNOVA A.V., LEBEDEV A.V. *Nonlinear approximation of characteristics of a fork-join queueing system with Pareto service as a model of parallel structure of data processing* // Mathematics and Computers in Simulation. – 2023. – Vol. 214. – P. 409–428.
 10. GORBUNOVA A.V., LEBEDEV A.V. *On estimating the characteristics of a fork-join queueing system with Poisson input and exponential service times* // Advances in Systems Science and Applications. – 2023. – Vol. 23, No. 2. – P. 99–114.
 11. KEMPER B., MANDJES M. *Mean sojourn times in two-queue fork-join systems: bounds and approximations* // OR Spectrum. – 2012. – Vol. 34, No. 3. – P. 723–742.
 12. KO S.S., SERFOZO R. *Response times in $M/M/s$ fork-join networks* // Advances in Applied Probability. – 2004. – Vol. 36, No. 3. – P. 854–871.
 13. LEMOINE A.J. *Networks of queues – a survey of equilibrium analysis* // Management Science. – 1977. – Vol. 24, No. 4. – P. 464–481.
 14. MARIN A., ROSSI S., SOTTANA M. *Dynamic resource allocation in fork-join queues* // ACM Trans. on Modeling

- and Performance Evaluation of Computing Systems. – 2020. – Vol. 5, No. 1. – Article No. 3. – P. 1–28.
15. NELSON R., TANTAWI A.N. *Approximate analysis of fork/join synchronization in parallel queues* // IEEE Trans. on Comp. – 1988. – Vol. 37, No. 6. – P. 739–743.
 16. NGUYEN M., ALESAWI S., LI N. et al. *A black-box fork-join latency prediction model for data-intensive applications* // IEEE Trans. on Parallel and Distributed Systems. – 2020. – Vol. 31, No. 9. – P. 1983–2000.
 17. OZKAN E. *Control of fork-join processing networks with multiple job types and parallel shared resources* // Mathematics of Operations Research. – 2022. – Vol. 47, No. 2. – P. 1310–1334.
 18. RAAIJMAKERS Y., BORST S., BOXMA O. *Fork-join and redundancy systems with heavy-tailed job sizes* // Queueing Systems. – 2023. – Vol. 103, No. 1. – P. 131–159.
 19. SCHOL D., VLASIOU M., ZWART B. *Large fork-join queues with nearly deterministic arrival and service times* // Mathematics of Operations Research. – 2021. – Vol. 47, No. 2. – P. 1335–1364.
 20. THOMASIAN A. *Analysis of fork/join and related queueing systems* // ACM Computing Surveys. – 2014. – Vol. 47, No. 2. – Article No. 17. – P. 1–71.
 21. VISHNEVSKY V.M., KLIMENOK V.I., SOKOLOV A.M. et al. *Investigation of the fork-join system with Markovian arrival process arrivals and phase-type service time distribution using machine learning methods* // Mathematics. – 2024. – Vol. 12, No. 5. – Article No. 659. – P. 1–22.

AN APPROXIMATION FOR THE RESPONSE TIME IN FORK-JOIN QUEUEING NETWORKS

Oksana Postnova, Saratov State University, Saratov, Assistant
(oksana.karpenko.2000@mail.ru).

Igor Tananko, Saratov State University, Saratov, Cand.Sc.,
Associate Professor (tanankoie@info.sgu.ru).

Ekaterina Rogachko, Saratov State University, Saratov, Cand.Sc.,
Associate Professor (rogachkoes@info.sgu.ru).

Abstract: An open fork-join queueing network is considered. An arriving job is split to be serviced into two tasks. The tasks are serviced independently at multiple service-nodes. Each service-node is a queueing system with one server and infinite capacity queue. Service-nodes form a queueing network with an acyclic topology. Two tasks associated with the job synchronize at a join-node before they leave the network. Approximations for the mean time spent by a task in the join-node and the mean response time in the fork-join queueing network are derived by assuming that jobs arrive according to a Poisson process and task service times have an exponential distribution. The accuracy of these approximations is demonstrated by comparing approximate results to simulation results. These approximations can be applied to the analysis of queueing networks with relatively small steady-state workload. The results can be used for the performance analysis of multiprocessor systems and other modern distributed computing systems.

Keywords: fork-join queueing network, acyclic queueing network, task service, approximation, response time.

УДК 519.872

ББК 22.18

*Статья представлена к публикации
членом редакционной коллегии А.В. Горбуновой.*

Поступила в редакцию 14.01.2025.

Дата опубликования 31.05.2025.