

Стохастические системы

© 2023 г. С.Н. СТЕПАНОВ, д-р техн. наук (s.n.stepanov@mtuci.ru),
М.С. СТЕПАНОВ, канд. техн. наук (m.s.stepanov@mtuci.ru)
(Московский технический университет связи и информатики)

ПРИБЛИЖЕННЫЙ МЕТОД ОЦЕНКИ ХАРАКТЕРИСТИК СОВМЕСТНОГО ОБСЛУЖИВАНИЯ ТРАФИКА РЕАЛЬНОГО ВРЕМЕНИ И ЭЛАСТИЧНОГО ТРАФИКА ДАННЫХ В МУЛЬТИСЕРВИСНЫХ УЗЛАХ ДОСТУПА

Построена и исследована математическая модель совместного обслуживания приоритетного трафика реального времени и эластичного трафика данных в мультисервисных узлах доступа. Приведены определения показателей качества совместного обслуживания поступающих запросов на информационное обслуживание. Сформирована система уравнений статистического равновесия и рассмотрено ее использование для вычисления точных значений введенных характеристик. Предложен метод приближенного расчета характеристик, основанный на построении системы упрощенных уравнений равновесия. Установлено, что полученные оценки показателей обслуживания заявок являются асимптотически точными в области больших и малых потерь. Показано использование разработанного метода для решения задачи оценки объема трафика, выгружаемого в ситуации перегрузки на другие узлы доступа или в другие диапазоны частот, с целью достижения заданных показателей QoS и задачи планирования необходимого по нагрузке объема ресурса передачи информации мультисервисного узла доступа.

Ключевые слова: мультисервисный трафик, трафик реального времени, эластичный трафик, рекурсивные алгоритмы, асимптотически точные характеристики, выгрузка избыточного трафика, планирование ресурса передачи.

DOI: 10.31857/S0005231023110065, EDN: OPBUYE

1. Введение

Развитие сетей связи идет по пути расширения числа сервисов. Набор услуг, получаемых абонентом, и их качество не должны зависеть от того, где пребывает абонент, как и с какой скоростью он передвигается и какие при этом применяются технологии доступа и пересылки информации. Это положение следует из тенденций развития телекоммуникационного рынка и закреплено в рекомендациях Международного союза электросвязи [1–6]. Понятно, что каждый вид услуг, будь то передача речи, видео или данных, имеет свои особенности, которые следует учитывать при исследовании условий

их совместного предоставления. В простейшем случае сервисы различаются только по величине требуемого ресурса передачи информации. В более сложных ситуациях необходимо учитывать детали формирования входных потоков заявок, условия доступа в сеть, наличие приоритета и возможности перераспределения ресурса в процессе обслуживания и т.д. Рассмотренные особенности функционирования действующих и перспективных систем связи анализируются в семействе так называемых мультисервисных моделей. Они имеют большое значение для практических приложений и являются предметом интенсивных исследований специалистов в области связи [4–10].

Важнейшим сегментом мультисервисной сети как фиксированной, так и подвижной связи является узел доступа, осуществляющий функцию концентрации абонентского трафика. Информационные потоки, попадающие в узлы доступа, можно разделить на две категории: трафик сервисов реального времени и трафик сервисов передачи эластичных данных. Передача трафика сервисов реального времени происходит с предварительным резервированием ресурса, который сохраняется на все время обслуживания заявки [1–3, 6]. Качество обслуживания характеризуется долей потерянных заявок и средним объемом занятого ресурса. Для пересылки эластичного трафика выделяется минимальный гарантированный ресурс, а также весь или часть ресурса, оставшегося свободным от пропуска трафика сервисов реального времени [8–11]. Качество обслуживания эластичных данных оценивается средним временем передачи файла и средней величиной получаемой пропускной способности узла доступа.

Обычно предполагается, что трафик реального времени имеет преимущество в занятии ресурса перед трафиком данных. Оно выражается в уменьшении скорости передачи эластичных данных до заранее заданного минимального значения, если выполнение этой процедуры способствует приему заявки на сервис реального времени. При появлении свободного ресурса скорость передачи данных возрастает. Перераспределение ресурса происходит в моменты времени, определяемые используемым механизмом адаптации скорости пересылки информации к условиям обслуживания заявок. Обычно это моменты поступления заявок и моменты окончания их обслуживания. Назовем подобную процедуру распределения динамической [7, 8, 10]. Пример ее реализации показан на рис. 1.

Анализируемый способ разделения ресурса позволяет значительно повысить его загрузку. Этот эффект особенно важен для сетей сотовой подвижной связи из-за ограниченности диапазона радиочастот, выделяемых для образования радиоканалов. По оценкам экспертов выигрыш может составить до нескольких десятков процентов от общего объема используемого ресурса. Отмеченный эффект получен в результате использования процедур управления трафиком, которые ускоряют передачу эластичных данных в ситуациях, когда уменьшается число пользователей услуг связи, находящихся на обслуживании. В сетях мобильной связи динамическое распределение ресурса выполняется диспетчером пакетов (англ., packet scheduler) и происходит под

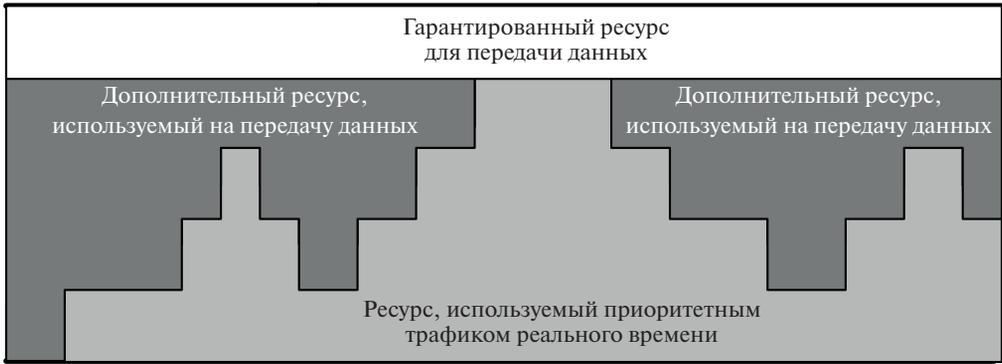


Рис. 1. Распределение пропускной способности узла доступа при совместной передаче приоритетного трафика сервисов реального времени и эластичных данных.

контролем программно-аппаратного комплекса RRM (англ., Radio Resource Management). Управляющие решения комплекса принимаются на основе информации о состоянии канала, числе и типах заявок, находящихся на обслуживании, уровне помех и т.д. [12–15].

Актуальность проблематики привлекла к ней внимание специалистов разной профессиональной подготовки, которая меняется от инженеров, занимающихся проектированием и эксплуатацией систем связи, до математиков. Давая общую характеристику опубликованным работам, следует отметить, что большинство исследований представляют из себя инженерные разработки, основанные на результатах имитационного моделирования и натурных экспериментов [1–3]. Теоретический анализ особенностей совместного обслуживания трафика сервисов реального времени и эластичных данных обычно заканчивается построением математической модели, определением характеристик качества обслуживания заявок и реализацией численных алгоритмов их оценки, основанных на решении системы уравнений равновесия каким-либо стандартным методом линейной алгебры [4, 5, 7–10, 13, 15, 16].

Для практических приложений, в частности для настройки процедур динамического распределения ресурса, необходимо построить несложные в реализации приближенные процедуры, имеющие приемлемую точность. К сожалению, таких методов в опубликованных работах представлено мало. Можно только упомянуть исследования [11, 14]. Особую важность имеют приближенные алгоритмы, которые

- основаны на общих принципах и могут легко обобщаться на другие модели формирования входных потоков заявок и процедуры распределения ресурса;
- имеют хорошую точность для значений входных параметров, которые соответствуют практическим приложениям, в частности для значений малых потерь, где происходит решение задачи планирования необходимого по нагрузке объема ресурса передачи информации, и для значений больших потерь, где решается задача оценки объема трафика, выгружаемого

на другие узлы доступа или в другие диапазоны частот, с целью достижения заданных показателей QoS;

- построены с использованием простых в реализации аналитических выражений и рекурсивных процедур.

Решению этих задач и посвящена данная работа. В разделе 2 рассмотрено математическое описание анализируемой модели узла доступа, на примере которой будут сформулированы принципы приближенной оценки характеристик качества совместного обслуживания трафика сервисов реального времени и эластичного трафика данных. В разделе 3 построена система уравнений равновесия и предложен алгоритм ее численного решения. Полученные результаты далее используются для оценки погрешности приближенных методов. В разделах 4 и 5 сформулированы принципы построения приближенных методов оценки характеристик рассматриваемой модели узла доступа и численно исследована их погрешность, в частности показано, что оценки являются асимптотически точными в области больших и малых потерь. В разделе 6 обсуждаются возможность использования модели для решения задачи определения объема выгружаемого трафика и задачи планирования необходимого объема ресурса. В последнем разделе сформулированы выводы по результатам выполненного исследования.

Новизна полученных результатов заключается в следующем:

- Сформулированы методы построения приближенных процедур оценки характеристик совместного обслуживания приоритетного трафика сервисов реального времени и эластичного трафика данных. Предлагаемые расчетные процедуры основаны на использовании системы упрощенных уравнений равновесия.
- Установлено, что полученные оценки характеристик являются асимптотически точными в области больших и малых потерь.
- Показано, что идеи, положенные в основу приближенных методов, носят общий характер и могут легко обобщаться на другие модели формирования входных потоков заявок и процедуры распределения ресурса.

2. Математическое описание модели

Обозначим через C пропускную способность мультисервисного узла доступа, создаваемую используемым стандартом фиксированной или беспроводной связи и выраженную в бит/с. Узел доступа обслуживает n пуассоновских потоков заявок на передачу трафика сервисов реального времени и один пуассоновский поток заявок на передачу эластичных данных. Заявки k -го потока на передачу трафика реального времени поступают с интенсивностью λ_k , требуют зарезервировать пропускную способность в объеме c_k бит/с на все время обслуживания, имеющее экспоненциальное распределение с параметром μ_k , $k = 1, \dots, n$. Заявки на передачу эластичных данных (файлов) поступают с интенсивностью λ_d . Объем файла имеет экспоненциальное распределение со средним значением F , выраженным в битах.

Введем понятие виртуального канала, который применим для численной оценки величины ресурса передачи информации, предоставляемой пользователям. Обозначим через c скорость передачи информации одного канала, выраженную в бит/с. Переход к виртуальным каналам упрощает моделирование процесса занятия ресурса заявками. Чем меньше значение c , тем точнее аппроксимация скоростей передачи. Однако в этой ситуации растет число состояний модели. Выбор величины c зависит от постановки задачи, например можно использовать такие выражения: $c = \min(c_1, \dots, c_n)$ или $c = \text{НОД}(c_1, \dots, c_n)$. В первом случае аппроксимация требований к битовой скорости более грубая, чем во втором. Общее число v имеющихся виртуальных каналов и число b_k виртуальных каналов, требуемых для обслуживания заявки k -го потока, находятся из соотношений:

$$v = \left\lfloor \frac{C}{c} \right\rfloor, \quad b_k = \left\lceil \frac{c_k}{c} \right\rceil.$$

Обслуживание эластичных данных происходит в соответствии с положениями дисциплины Processor Sharing. Рассмотрим реализацию этой процедуры. Для простоты предположим, что минимальный объем пропускной способности, который может быть использован для передачи файла, равен одному виртуальному каналу. Обозначим через μ_d параметр экспоненциального распределения времени передачи файла одним каналом. Пусть i число виртуальных каналов, занятых на обслуживание трафика реального времени, а $d > 0$ — число файлов, находящихся на передаче. Для их обслуживания используются $(v - i)$ каналов. Пусть $s = \lfloor \frac{v-i}{d} \rfloor$. Свободные каналы делятся между d файлами по следующему правилу. Для обслуживания каждого из $(v - i - sd)$ файлов используется $(s + 1)$ каналов, а для обслуживания каждого из оставшихся $((s + 1)d - (v - i))$ файлов используется s каналов. В результате выполнения этой процедуры все $(v - i)$ каналов заняты. Нетрудно показать, что в рассматриваемой ситуации время до окончания передачи одного из d файлов имеет экспоненциальное распределение с параметром $(v - i)\mu_d$.

Поступившая заявка на передачу трафика реального времени имеет приоритет в занятии ресурса, уменьшая при необходимости пропускную способность, используемую одним файлом, до одного канала. Пусть $i_k(t)$, $k = 1, \dots, n$ — число заявок k -го потока на передачу трафика сервисов реального времени, находящихся в момент времени t на обслуживании, а $d(t)$ — число файлов, обслуживаемых в момент времени t . Динамика изменения числа обслуживаемых заявок описывается марковским процессом $r(t) = (i_1(t), \dots, i_n(t), d(t))$, определенным на конечном пространстве состояний S , куда входят состояния (i_1, \dots, i_n, d) , с компонентами

$$(1) \quad \begin{aligned} i_1 = 0, 1, \dots, \left\lfloor \frac{v}{b_1} \right\rfloor; \quad \dots \quad i_n = 0, 1, \dots, \left\lfloor \frac{v - i_1 b_1 - \dots - i_{n-1} b_{n-1}}{b_n} \right\rfloor; \\ d = 0, 1, \dots, v - i_1 b_1 - \dots - i_n b_n. \end{aligned}$$

Качество обслуживания заявок k -го потока на передачу трафика сервисов реального времени определяется значениями доли потерянных заявок π_k и средним числом занятых виртуальных каналов m_k . Величина последней характеристики дает возможность рассчитать среднее число заявок k -го потока, находящихся на обслуживании, $y_k = m_k/b_k$ и занимаемую ими среднюю величину пропускной способности узла доступа $z_k = m_k c$. Обозначим через z_r среднюю величину пропускной способности узла, занимаемую трафиком реального времени $z_r = \sum_{k=1}^n z_k$. Качество обслуживания заявок на передачу эластичных файлов определяется значениями доли потерянных файлов π_d , средним числом занятых виртуальных каналов m_d , средним числом файлов, находящихся на обслуживании, y_d , средней величиной используемой пропускной способности узла доступа $z_d = m_d c$, средним временем передачи файла h_d , средним битрейтом, используемым для передачи файла c_d , средним числом виртуальных каналов b , используемых для передачи файла.

Введенные показатели можно определить и рассчитать, используя значения стационарных вероятностей $p(i_1, \dots, i_n, d)$ состояний $(i_1, \dots, i_n, d) \in S$. Обозначим для состояния (i_1, \dots, i_n, d) через i число виртуальных каналов, занятых на обслуживание трафика реального времени $i = i_1 b_1 + \dots + i_n b_n$. Приведем расчетные выражения

$$\begin{aligned}
 \pi_k &= \sum_{\{(i_1, \dots, i_n, d) \in S \mid i+d+b_k > v\}} p(i_1, \dots, i_n, d); \\
 m_k &= \sum_{(i_1, \dots, i_n, d) \in S} p(i_1, \dots, i_n, d) i_k b_k; \quad k = 1, \dots, n; \\
 \pi_d &= \sum_{\{(i_1, \dots, i_n, d) \in S \mid i+d+1 > v\}} p(i_1, \dots, i_n, d); \\
 m_d &= \sum_{\{(i_1, \dots, i_n, d) \in S \mid d > 0\}} p(i_1, \dots, i_n, d) (v - i); \\
 y_d &= \sum_{(i_1, \dots, i_n, d) \in S} p(i_1, \dots, i_n, d) d; \\
 h_d &= \frac{y_d}{\lambda_d (1 - \pi_d)}; \quad c_d = \frac{F}{h_d}; \quad b = \frac{m_d}{y_d}.
 \end{aligned}
 \tag{2}$$

3. Система уравнений статистического равновесия

Для оценки характеристик, заданных выражениями (2), необходимо найти значения $p(i_1, \dots, i_n, d) \in S^1$. Для этого достаточно построить и решить систему уравнений статистического равновесия, связывающую ненормированные

¹ Будем использовать строчные буквы для обозначения нормированных значений вероятностей состояний и характеристик и прописные буквы для обозначения их ненормированных значений.

вероятности модели. Она имеет следующий вид:

$$\begin{aligned}
 (3) \quad & P(i_1, \dots, i_n, d) \left\{ \sum_{k=1}^n \left(\lambda_k I(i + d + b_k \leq v) + i_k \mu_k \right) + \right. \\
 & \left. + \lambda_d I(i + d + 1 \leq v) + (v - i) \mu_d I(d > 0) \right\} = \\
 & = \sum_{k=1}^n P(i_1, \dots, i_k - 1, \dots, i_n, d) \lambda_k I(i_k > 0) + P(i_1, \dots, i_n, d - 1) \lambda_d I(d > 0) + \\
 & + \sum_{k=1}^n P(i_1, \dots, i_k + 1, \dots, i_n, d) (i_k + 1) \mu_k I(i + d + b_k \leq v) + \\
 & + P(i_1, \dots, i_n, d + 1) (v - i) \mu_d I(i + d + 1 \leq v), \quad (i_1, \dots, i_n, d) \in S.
 \end{aligned}$$

Здесь и далее $I(\cdot)$ — индикаторная функция, определяемая соотношением

$$(4) \quad I(\cdot) = \begin{cases} 1, & \text{если выполнено условие, сформулированное} \\ & \text{в скобках,} \\ 0, & \text{если это условие не выполнено.} \end{cases}$$

Полученные в результате решения системы (3) ненормированные значения вероятностей $P(i_1, \dots, i_n, d)$ необходимо нормировать.

Система уравнений (3) не обладает какими-либо специальными свойствами, которые обеспечивают рекурсивную оценку стационарных вероятностей. По этой причине для решения (3) используются стандартные методы линейной алгебры. Исходя из опыта решения подобных систем, для оценки вероятностей состояний рекомендуется использовать итерационный алгоритм Гаусса–Зейделя. Применение данного подхода позволяет рассчитывать стационарные вероятности для моделей систем связи с числом состояний до нескольких миллионов. Стандартная реализация рекурсии Гаусса–Зейделя для решения (3) не обладает гарантированной сходимостью, но в большинстве случаев она имеет место. Сходимость исследуется косвенными методами на основе анализа близости последовательных приближений и выполнения известных теоретических соотношений, связывающих величины характеристик исследуемой модели системы связи. К таким соотношениям относится формула Литтла. Детали реализации данного подхода можно найти в [10, 17–19]. Для обеспечения сходимости достаточно в (3) одну из неизвестных положить равной единице, убрать соответствующее уравнение (3) и перейти к решению неоднородной системы линейных уравнений. После проделанных преобразований рекурсия Гаусса–Зейделя всегда сходится (из-за наличия слабого диагонального преобладания), но требует для своей реализации существенно большего числа итераций.

Алгоритм Гаусса–Зейделя будет далее использоваться для оценки погрешности построенного в разделе 4 приближенного метода вычисления характе-

ристик качества совместного обслуживания поступающих заявок исследуемой модели узла. Достаточно обоснованные приближенные методы играют основную роль в разработке инженерных методик, направленных на решение задач оценки максимально достаточной нагрузки и минимально необходимого объема ресурса передачи информации мультисервисного узла доступа. Формулировке, анализу и примерам использования приближенных методов расчета характеристик построенной модели узла доступа будут посвящены последующие разделы работы.

4. Оценка характеристик с использованием упрощенных уравнений равновесия

Принцип упрощенных уравнений равновесия был впервые сформулирован и использован при построении границ урезанных пространств состояний, обеспечивающих заданную погрешность вычисления характеристик моделей с учетом эффекта повторных вызовов [20, 21]. Идея метода заключается в формировании системы соотношений для приближенного расчета стационарных вероятностей состояний модели исходя из требования о соблюдении для них локальных законов сохранения, которые для исследуемой модели выполняются в отдельных макросостояниях, выбираемых в соответствии со структурой используемого пространства состояний и физическим смыслом отдельных компонент состояния.

Для построенной модели к таким макросостояниям отнесем (i) , где объединены все состояния модели с числом занятых виртуальных каналов, равным i , $i = 0, 1, \dots, v$, и (d) , где объединены все состояния модели с числом передаваемых файлов, равным d , $d = 0, 1, \dots, v$. Действуя аналогичным образом, введем также макросостояние (i, d) и определим стационарную вероятность этого состояния из выражения

$$p(i, d) = \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} p(i_1, \dots, i_n, d),$$

$$i = 0, 1, \dots, v; \quad d = 0, 1, \dots, v - i.$$

Если известны значения $p(i, d)$, то можно рассчитать величины характеристик, заданных выражениями (2). Для этого достаточно воспользоваться соотношениями

$$(5) \quad \pi_k = \sum_{i=0}^v \sum_{d=0}^{v-i} p(i, d) I(i + d + b_k > v); \quad m_k = \frac{\lambda_k}{\mu_k} (1 - \pi_k) b_k;$$

$$\pi_d = \sum_{i=0}^v \sum_{d=0}^{v-i} p(i, d) I(i + d + 1 > v); \quad m_d = \sum_{i=0}^v \sum_{d=1}^{v-i} p(i, d) (v - i);$$

$$y_d = \sum_{i=0}^v \sum_{d=1}^{v-i} p(i, d) d.$$

Остальные характеристики обслуживания эластичного трафика определяются из выражений (2).

Получим вид локальных законов сохранения для макросостояний (i) , (d) , которые далее будут использоваться для приближенной оценки $p(i, d)$. Начнем с макросостояния (d) . Приравняв интенсивности выхода $r(t)$ из (d) к интенсивности перехода в макросостояние (d) , получаем систему соотношений следующего вида:

$$\begin{aligned}
 (6) \quad & \left(P(0, d) + P(1, d) + \dots + P(v - d - 1, d) \right) \lambda_d + \\
 & + \left(P(0, d)v + P(1, d)(v - 1) + \dots + P(v - d, d)d \right) \mu_d I(d > 0) = \\
 & = \left(P(0, d - 1) + P(1, d - 1) + \dots + P(v - d, d - 1) \right) \lambda_d I(d > 0) + \\
 & + \left(P(0, d + 1)v + P(1, d + 1)(v - 1) + \dots + P(v - d - 1, d + 1)(d + 1) \right) \mu_d, \\
 & \quad d = 0, 1, \dots, v - 1.
 \end{aligned}$$

Рассмотрев (6) последовательно при $d = 0, 1, \dots, v - 1$ и выполнив несложные алгебраические преобразования, приводим (6) к следующему виду:

$$\begin{aligned}
 (7) \quad & \left(P(0, d) + P(1, d) + \dots + P(v - d - 1, d) \right) \lambda_d = \\
 & = \left(P(0, d + 1)v + P(1, d + 1)(v - 1) + \dots + P(v - d - 1, d + 1)(d + 1) \right) \mu_d, \\
 & \quad d = 0, 1, \dots, v - 1.
 \end{aligned}$$

Теперь получим вид локальных законов сохранения для макросостояния (i) . Приравняв интенсивности выхода $r(t)$ из (i) к интенсивности перехода в макросостояние (i) , получаем систему соотношений следующего вида:

$$\begin{aligned}
 (8) \quad & \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \sum_{k=0}^n \left(P(i_1, \dots, i_k - 1, \dots, i_n, d) \lambda_k I(i_k > 0) + \right. \\
 & \left. + P(i_1, \dots, i_k + 1, \dots, i_n, d) (i_k + 1) \mu_k I(i + b_k + d \leq v) \right) = \\
 & = \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \sum_{k=0}^n \left(P(i_1, \dots, i_n, d) \lambda_k I(i + b_k + d \leq v) + \right. \\
 & \left. + P(i_1, \dots, i_n, d) i_k \mu_k I(i_k > 0) \right), \quad i = 0, 1, \dots, v.
 \end{aligned}$$

Перейдем к построению системы упрощенных уравнений равновесия, которую далее будем использовать для приближенного вычисления характеристик исследуемой модели узла доступа. Обозначим через $\hat{P}(i_1, \dots, i_n, d)$, $(i_1, \dots, i_n, d) \in S$ получаемые при этом оценки вероятностей стационарных

состояний модели. Сохраним для оценок введенных характеристик (2), вероятностей $P(i, d)$ и т.д., получаемых с использованием $\hat{P}(i_1, \dots, i_n, d)$, ранее введенные обозначения, только добавим в их запись символ $\hat{\cdot}$. Величины $\hat{P}(i_1, \dots, i_n, d)$ находятся из требования о выполнении для $\hat{P}(i_1, \dots, i_n, d)$ локального закона сохранения (6)

$$(9) \quad \left(\hat{P}(0, d) + \hat{P}(1, d) + \dots + \hat{P}(v - d - 1, d) \right) \lambda_d = \\ = \hat{P}(0, d + 1) v \mu_d + \hat{P}(1, d + 1) (v - 1) \mu_d + \dots + \hat{P}(v - d - 1, d + 1) (d + 1) \mu_d, \\ d = 0, 1, \dots, v - 1,$$

и соотношений

$$(10) \quad \hat{P}(i_1, \dots, i_n, d) i_k = \hat{P}(i_1, \dots, i_k - 1, \dots, i_n, d) I(i_k > 1) a_k, \\ (i_1, \dots, i_n, d) \in S, \quad a_k = \frac{\lambda_k}{\mu_k}, \quad k = 1, \dots, n.$$

Выражения (10) по форме совпадают с соотношениями детального баланса при формировании и обслуживании заявок на передачу трафика сервисов реального времени в мультисервисной модели Эрланга [7, 10, 22, 23]. Они определяют приближенный характер получаемых оценок, поскольку для исходной модели они не выполняются.

Воспользовавшись равенствами (10), можно показать, что для $\hat{P}(i_1, \dots, i_n, d)$ выполняется локальный закон сохранения (8), т.е. справедливы соотношения

$$(11) \quad \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \sum_{k=1}^n \hat{P}(i_1, \dots, i_k - 1, \dots, i_n, d) \lambda_k I(i_k > 0) + \\ + \hat{P}(i_1, \dots, i_k + 1, \dots, i_n, d) (i_k + 1) \mu_k I(i + b_k + d \leq v) = \\ = \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \sum_{k=1}^n \hat{P}(i_1, \dots, i_n, d) \lambda_k I(i + b_k + d \leq v) + \\ + \hat{P}(i_1, \dots, i_n, d) i_k \mu_k I(i_k \geq v), \quad i = 0, 1, \dots, v.$$

Далее покажем, что соотношения (9), (10) позволяют единственным образом определить значения $\hat{P}(i_1, \dots, i_n, d)$, $(i_1, \dots, i_n, d) \in S$, а с ними и введенные оценки показателей совместного обслуживания заявок в исследуемой модели узла доступа. Таким образом, оценки стационарных вероятностей $\hat{P}(i_1, \dots, i_n, d)$, $(i_1, \dots, i_n, d) \in S$, также как и их точные значения $P(i_1, \dots, i_n, d)$, $(i_1, \dots, i_n, d) \in S$ удовлетворяют одинаковым по форме локальным законам сохранения (7), (8) и (9), (11), что позволяет ожидать хорошую точность оценивания. Исходя из вероятностной интерпретации (9), (10), назовем эти соотношения системой упрощенных уравнений равновесия.

Построим рекурсивный алгоритм вычисления значений $\hat{P}(i, d)$ при фиксированном d . Введем вспомогательную характеристику

$$\hat{Y}_k(i, d) = \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \hat{P}(i_1, \dots, i_n, d) i_k, \\ i = 0, 1, \dots, v; \quad d = 0, 1, \dots, v - i.$$

Просуммировав (10) по всем $(i_1, \dots, i_n, d) \in S$, удовлетворяющим условию $i_1 b_1 + \dots + i_n b_n = i$, получаем равенство

$$(12) \quad \hat{Y}_k(i, d) = \hat{P}(i - b_k, d) a_k I(i \geq b_k).$$

Умножим (12) на b_k и просуммируем по $k = 1, 2, \dots, n$. Переставив слева порядок суммирования, получаем рекурсивное соотношение, связывающее последовательные значения $\hat{P}(i, d)$,

$$(13) \quad \sum_{\{(i_1, \dots, i_n, d) \in S \mid i_1 b_1 + \dots + i_n b_n = i\}} \hat{P}(i_1, \dots, i_n, d) \sum_{k=1}^n i_k b_k = \hat{P}(i, d) i = \\ = \sum_{k=1}^n a_k b_k \hat{P}(i - b_k, d) I(i \geq b_k), \quad d = 0, 1, \dots, v - 1; \quad i = 1, \dots, v - d.$$

Соотношения (7) и (13) позволяют построить рекурсивный алгоритм оценки значений $\hat{P}(i, d)$, $d = 0, 1, \dots, v$; $i = 1, \dots, v - d$. Перечислим последовательность действий при его реализации.

1. Положим $d = 0$, а значение $\hat{P}(0, 0) = 1$.
2. Выразим значения $\hat{P}(i, 0)$, $i = 1, \dots, v$, через $\hat{P}(0, 0)$, используя соотношение (13) при $d = 0$

$$\hat{P}(i, 0) = \frac{1}{i} \times \sum_{k=1}^n \hat{P}(i - b_k, 0) I(i \geq b_k)$$

и последовательно увеличивая i от 1 до v . При фиксированном i значения оценок $\hat{P}(i - b_k, 0)$, $k = 1, \dots, n$, либо уже представлены через $\hat{P}(0, 0)$ (для $i - b_k \geq 0$), либо равны 0 (для $i - b_k < 0$).

3. Положим $d = 1$, а значение $\hat{P}(0, 1) = x$.
4. Выразим значения $\hat{P}(i, 1)$, $i = 1, \dots, v - 1$, через x , используя соотношение (13) при $d = 1$

$$\hat{P}(i, 1) = \frac{1}{i} \times \sum_{k=1}^n \hat{P}(i - b_k, 1) I(i \geq b_k)$$

и последовательно увеличивая i от 1 до $v - 1$.

5. Положим в (9) $d = 0$. Получаем соотношение

$$(14) \quad \left(\hat{P}(0,0) + \hat{P}(1,0) + \dots + \hat{P}(v-1,0) \right) \lambda_d = \\ = \hat{P}(0,1)v\mu_d + \hat{P}(1,1)(v-1)\mu_d + \dots + \hat{P}(v-1,1)\mu_d,$$

которое позволяет выразить значение x через $\hat{P}(0,0)$. Воспользовавшись полученным соотношением и результатами п. 4, получаем выражения для $\hat{P}(i,1)$, $i = 0, 1, \dots, v-1$, через $\hat{P}(0,0)$.

6. Далее положим $d = 2, 3, \dots, v$ и, реализовав последовательность действий, сформулированных в перечисленных выше этапах реализации алгоритма, находим выражения для $\hat{P}(i,d)$, $d = 0, 1, \dots, v$; $i = 1, \dots, v-d$ через $\hat{P}(0,0)$.
7. Находим выражение через $\hat{P}(0,0)$ для нормировочной константы

$$N = \sum_{d=0}^v \sum_{i=0}^{v-d} \hat{P}(i,d).$$

8. Определяем нормированные значения оценок вероятностей $\hat{p}(i,d)$:

$$\hat{p}(i,d) = \frac{\hat{P}(i,d)}{N}, \quad d = 0, 1, \dots, v; \quad i = 1, \dots, v-d.$$

9. Используя $\hat{p}(i,d)$ и выражения (5), находим величины оценок характеристик совместного обслуживания заявок в исследуемой модели мультисервисного узла доступа.

На рис. 2 стрелками отмечена последовательность вычисления $\hat{p}(i,d)$ с помощью построенного алгоритма для $v = 5$.

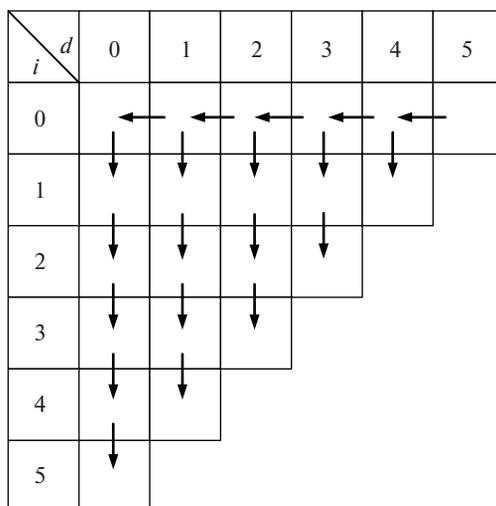


Рис. 2. Последовательность реализации алгоритма для $v = 5$.

С вычислительной точки зрения реализация построенного алгоритма не вызывает никаких затруднений. Трудоемкость процедуры сравнима с многократным использованием рекурсивного алгоритма оценки характеристик мультисервисной модели Эрланга [7, 10, 22, 23]. Сравним объемы вычислений при оценке характеристик модели методом Гаусса–Зейделя и построенным приближенным алгоритмом. На каждом шаге итерационного метода решения (3) определяются приближенные значения вероятностей всех состояний модели $(i_1, \dots, i_n, d) \in S$, принадлежащих пространству S , заданному соотношениями (1). В результате применения приближенного метода находятся оценки вероятностей $P(i, d)$ для состояний $(i, d) \in \hat{S}$, где $i = 0, 1, \dots, v$; $d = 0, 1, \dots, v - i$ (см. рис. 2) и расчеты выполняются только один раз. Отсюда следует примерная оценка эффективности разработанного метода. Объем вычислений по сравнению с итерационным методом уменьшается в число раз, равное числу итераций в реализации метода Гаусса–Зейделя, умноженному на отношение числа состояний в пространстве S к числу состояний в пространстве \hat{S} . В зависимости от параметров модели и условий счета эта величина может существенно превзойти несколько тысяч раз.

5. Погрешность оценки характеристик узла доступа

Проведем численное исследование точности приближенного метода оценки характеристик узла доступа, построенного в предыдущем разделе. Выберем следующие значения входных параметров: $C = 100$ Мбит/с; $n = 2$; $c_1 = 2$ Мбит/с; $c_2 = 5$ Мбит/с. Исходя из принятых допущений, получаем структурные параметры модели: $c = 1$ Мбит/с; $v = 100$ виртуальных каналов (в.к.); $b_1 = 2$ в.к.; $b_2 = 5$ в.к. Предположим, что $F = 80$ Мбит. Среднее время передачи файла с использованием одного канала равно 80 с. При проведении вычислений это время будет принято за единицу. Отсюда $\mu_d = 1$. Для сервисов реального времени параметры времени обслуживания выберем из выражений $\mu_1 = 0,5$ и $\mu_2 = 0,5$.

Введем параметр ρ , который будем использовать для оценки потенциальной загрузки одного виртуального канала. Определим ρ из выражения

$$(15) \quad \rho = \frac{a_1 b_1 + a_2 b_2 + \frac{\lambda_d}{\mu_d}}{v}.$$

Для эластичных данных потенциальную загрузку ресурса рассчитываем из условия об использовании для передачи файла одного канала². Будем предполагать, что все три потока заявок создают одинаковую потенциальную загрузку ресурса. Отсюда следуют выражения для оценки интенсивностей поступающих заявок:

$$(16) \quad \lambda_1 = \frac{v\rho\mu_1}{3b_1}; \quad \lambda_2 = \frac{v\rho\mu_2}{3b_2}; \quad \lambda_d = \frac{v\rho\mu_d}{3}.$$

² Наихудший сценарий обслуживания эластичных данных.

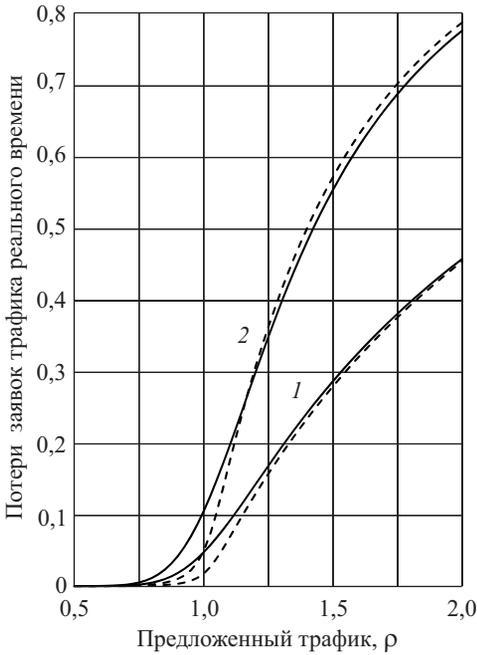


Рис. 3. Погрешность оценки доли потерянных заявок на обслуживание трафика сервисов реального времени.

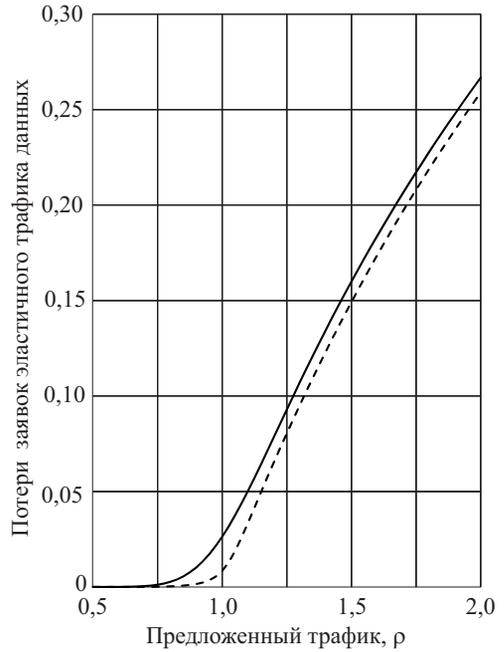


Рис. 4. Погрешность оценки доли потерянных заявок на обслуживание эластичных данных.

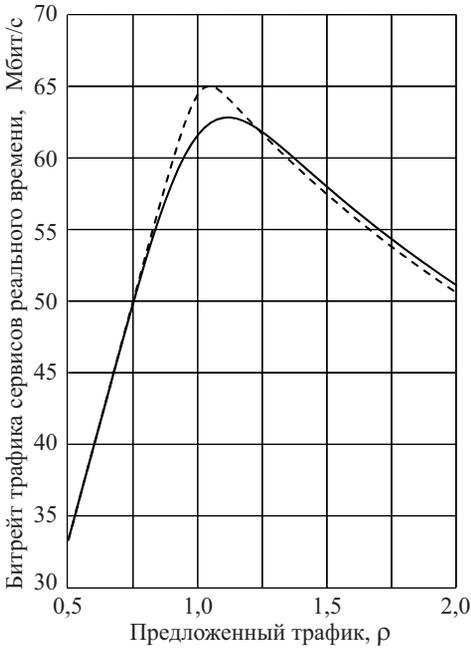


Рис. 5. Погрешность оценки среднего значения битрейта, использованного на обслуживание трафика сервисов реального времени.

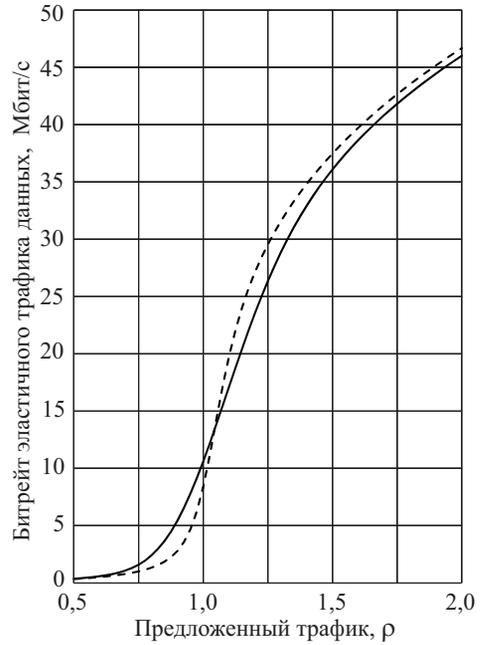


Рис. 6. Погрешность оценки среднего значения битрейта, использованного на обслуживание эластичных данных.

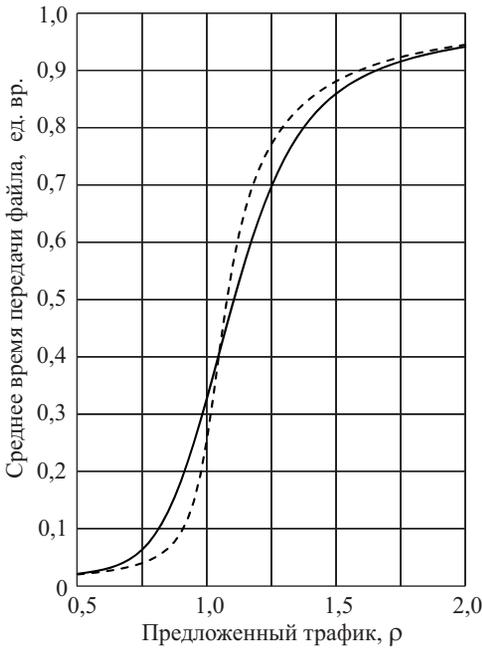


Рис. 7. Погрешность оценки среднего времени передачи файла.

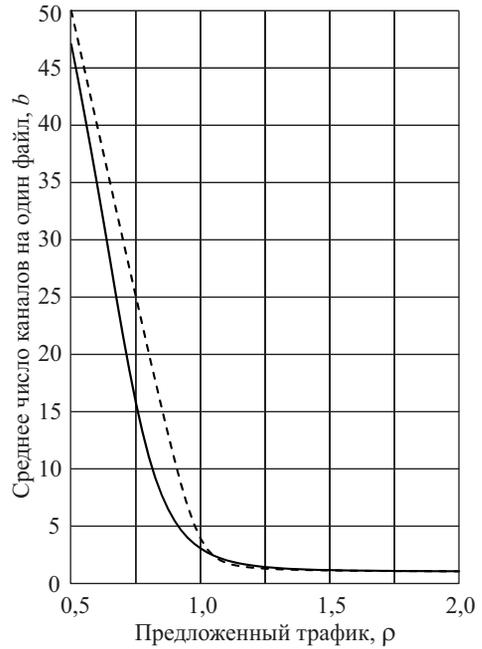


Рис. 8. Погрешность оценки среднего каналов, использованных на передачу одного файла.

Для выбранных значений входных параметров на рис. 3–8 показана зависимость точного и приближенного расчета основных характеристик модели от изменения ρ потенциальной загрузки виртуального канала. Точные значения характеристик получены из решения системы уравнений равновесия (3) итерационным методом Гаусса–Зейделя и использования определений (2). Приближенные величины находились из соотношений (5) после подстановки в них вместо $p(i, d)$ оценок $\hat{p}(i, d)$, полученных в результате реализации рекурсивного алгоритма, основанного на использовании упрощенных уравнений равновесия (9), (10) (см. раздел 4). Соответствующие кривые обозначены пунктиром. Цифра у кривой обозначает номер потока заявок сервисов реального времени. На рис. 3–4 оценивалась погрешность вычисления доли потерянных заявок на обслуживание трафика сервисов реального времени (рис. 3) и доли потерянных заявок на обслуживание эластичных данных (рис. 4). На рис. 5–6 оценивалась погрешность вычисления среднего значения битрейта, использованного на обслуживание трафика сервисов реального времени (рис. 5), и среднего значения битрейта, использованного на обслуживание эластичных данных (рис. 6). На рис. 7–8 оценивалась погрешность вычисления среднего времени передачи файла (рис. 7) и среднего каналов, использованных на передачу одного файла (рис. 8).

Из анализа численных данных следует, что полученные оценки обладают следующими свойствами.

1. Они имеют хорошую точность, которая возрастает при малой и большой нагрузке. Далее эти результаты будут исследованы более подробно.

2. С увеличением нагрузки на канал величины анализируемых характеристик увеличиваются за исключением среднего значения битрейта, используемого на обслуживание трафика сервисов реального времени (см. рис. 5). Это свойство объясняется тем, что с ростом ρ эластичные данные все чаще используют для передачи файла один канал, вытесняя тем самым из обслуживания трафик сервисов реального времени, требующий больших значений числа каналов.
3. Как правило, полученные приближенные значения характеристик при малой нагрузке дают оценку снизу для характеристик исследуемой модели узла доступа, а при большой нагрузке — оценку сверху.

Как было отмечено ранее, точность оценок характеристик возрастает в области больших и малых потерь. Приведем численные данные, подтверждающие этот вывод. В табл. 1 и 2 показаны величины характеристик качества обслуживания трафика сервисов реального времени и эластичных данных для модели узла доступа со значениями входных параметров, использованными при расчете кривых, показанных на рис. 3–8 и перечисленных в начале

Таблица 1. Точная и приближенная оценки характеристик модели узла доступа в условиях большой нагрузки

ρ	π_2		π_d		h_d		z_r		z_d	
	Точно	Прибл.	Точно	Прибл.	Точно	Прибл.	Точно	Прибл.	Точно	Прибл.
1,00	0,1051	0,0496	0,0264	0,0085	0,3270	0,2546	61,58	64,42	10,61	8,41
1,25	0,3484	0,3624	0,0930	0,0806	0,6972	0,7713	61,81	61,68	26,34	29,54
1,50	0,5538	0,5716	0,1600	0,1493	0,8594	0,8811	57,99	57,46	36,09	37,47
1,75	0,6877	0,7021	0,2173	0,2081	0,9157	0,9233	54,33	53,80	41,81	42,65
2,00	0,7760	0,7873	0,2668	0,2591	0,9414	0,9449	51,12	50,60	46,01	46,67
2,50	0,8778	0,8848	0,3493	0,3437	0,9649	0,9661	45,77	45,30	52,32	52,84
3,00	0,9290	0,9334	0,4154	0,4112	0,9756	0,9762	41,53	41,12	57,03	57,47
4,00	0,9723	0,9742	0,5147	0,5122	0,9853	0,9855	35,28	34,96	63,76	64,09
5,00	0,9875	0,9884	0,5853	0,5838	0,9897	0,9898	30,88	30,63	68,39	68,66
7,50	0,9974	0,9975	0,6956	0,6951	0,9942	0,9943	23,90	23,76	75,65	75,80

Таблица 2. Точная и приближенная оценки характеристик качества обслуживания эластичного трафика в условиях малой нагрузки

ρ	h_d		z_r		z_d		b	
	Точно	Прибл.	Точно	Прибл.	Точно	Прибл.	Точно	Прибл.
0,500	0,02124	0,02000	33,33	33,33	0,35401	0,33333	47,078	50,000
0,400	0,01717	0,01666	26,66	26,66	0,22899	0,22222	58,225	60,000
0,300	0,01451	0,01428	20,00	20,00	0,14512	0,14285	68,908	70,000
0,200	0,01259	0,01250	13,33	13,33	0,08398	0,08333	79,378	80,000
0,150	0,01182	0,01176	10,00	10,00	0,05912	0,05882	84,565	85,000
0,100	0,01114	0,01111	6,666	6,666	0,03714	0,03703	89,728	90,000
0,075	0,01083	0,01081	5,000	5,000	0,02708	0,02702	92,302	92,500
0,050	0,01054	0,01052	3,333	3,333	0,01756	0,01754	94,871	95,000
0,025	0,01026	0,01025	1,666	1,666	0,00855	0,00854	97,437	97,500

данного раздела. Приведенные численные данные подтверждают асимптотические свойства оценок. Погрешность оценок быстро уменьшается с увеличением ρ (табл. 1) и с его уменьшением (табл. 2). В области больших потерь асимптотические свойства проявляют себя для всех характеристик, перечисленных в (2). В области малых потерь — только для характеристик передачи эластичных данных.

Приведенные численные данные показали, что разработанный метод имеет хорошую точность для величин входных параметров, которые соответствуют практическим приложениям, в частности для значений малых потерь, где происходит решение задачи планирования необходимого по нагрузке объема ресурса передачи информации, и для значений больших потерь, где решается задача оценки объема трафика, выгружаемого на другие узлы доступа или в другие диапазоны частот, с целью достижения заданных показателей QoS. Приведем примеры решения сформулированных задач.

6. Практическое использование полученных результатов

Оценка требуемого ресурса узла доступа для заданной нагрузки и максимально допустимой нагрузки для заданной величины ресурса являются близкими задачами и решаются методом перебора. В первом случае начиная с некоего начального значения увеличивается пропускная способность узла доступа, пока не будет достигнут требуемый результат по качеству обслуживания, во втором — в ситуации перегрузки с этой же целью уменьшается входной поток заявок. Начнем с решения второй задачи.

Изменим в исследуемой модели мультисервисного узла доступа описание процедуры формирования входного потока. Предположим, что узел доступа обслуживает пуассоновский поток заявок интенсивности λ , разделенных на $n + 1$ сервисных категорий. Первые n категорий представляют из себя заявки на передачу трафика сервисов реального времени. Последняя ($n + 1$)-я категория — это заявки на передачу эластичного трафика данных. Примем, что $\lambda = \lambda_1 + \dots + \lambda_n + \lambda_d$. С вероятностью $p_k = \frac{\lambda_k}{\lambda}$, $k = 1, \dots, n$ заявка принадлежит k -й категории, требует c_k бит/с, $k = 1, \dots, n$ и занимает ресурс случайное время, имеющее экспоненциальное распределение с параметром μ_k . С вероятностью $p_d = \frac{\lambda_d}{\lambda}$ заявка принадлежит категории эластичного трафика и обслуживается по соответствующим правилам, введенным при описании исходной модели узла (см. раздел 2). Понятно, что рассмотренное изменение процедуры формирования входного потока не изменило математическое описание исследуемой модели узла доступа.

Предположим, что имеющийся или запланированный трафик приводит к перегрузке передаточных возможностей узла доступа. С хорошей точностью эту ситуацию можно отследить, используя метод оценки характеристик качества обслуживания заявок, основанный на использовании упрощенных уравнений равновесия (см. раздел 4 и табл. 1). Допустим, что в сложившейся ситуации нельзя просто увеличить пропускную способность узла доступа. С возникшими трудностями можно справиться, перенаправив (говорят так-

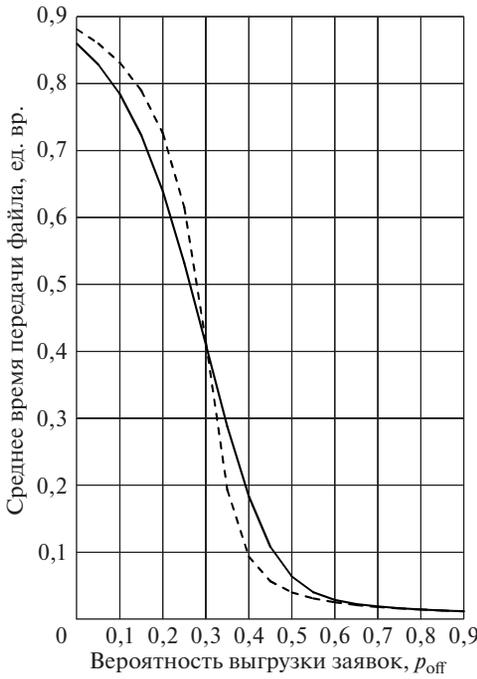


Рис. 9. Зависимость оценки среднего времени передачи файла от p_{off} .

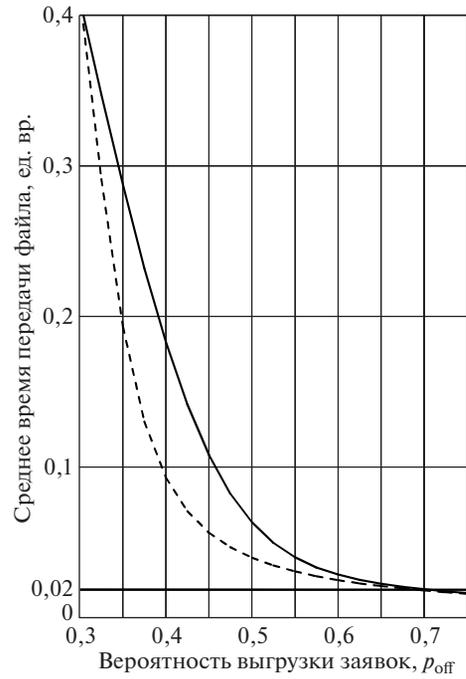


Рис. 10. Оценка доли выгружаемого трафика для обеспечения заданных показателей QoS.

же выгрузив) какую-то часть входного потока на другие узлы доступа с тем, чтобы характеристики качества обслуживания оставшейся части трафика на имеющемся ресурсе не превышали нормативных значений [24]. Обозначим через p_{off} вероятность выгрузки поступившей заявки.

Формальная постановка задачи выглядит следующим образом. Необходимо определить величину p_{off} , которая обеспечивает заданный уровень качества обслуживания оставшихся заявок в форме выполнения неравенств

$$(17) \quad \pi = \max(\pi_1, \dots, \pi_n, \pi_d) < \pi^{\text{norm}}, \quad h_d < h_d^{\text{norm}}.$$

Здесь π^{norm} и h_d^{norm} — требуемые по условиям соглашения об обслуживании величины соответственно максимальных потерь заявок и среднего времени передачи эластичных данных.

Рассмотрим модель узла доступа для значений входных параметров, использованных при расчете содержания табл. 1. Величины характеристик обслуживания заявок при $\rho = 1,5$ говорят о том, что узел находится в состоянии перегрузки и надо уменьшать поступающий поток заявок. Возьмем следующие значения нормативных показателей: $\pi^{\text{norm}} = 0,01$, $h_d^{\text{norm}} = 0,2$.

Задачу оценки вероятности выгрузки решим в два этапа. Вначале найдем величину p_{off} , которая обеспечивает выполнение неравенства (17) для среднего времени передачи файла. Отметим, что выбор $h_d^{\text{norm}} = 0,2$ означает, что

среднее время передачи файла выбирается в 50 раз меньшим, чем среднее время передачи файла на одном канале, и составляет 1,6 с. На рис. 9 показано изменение h_d с увеличением p_{off} в диапазоне от 0 до 0,9. Приводятся точное значение характеристики, найденное в результате решения системы уравнений равновесия (3) итерационным методом Гаусса–Зейделя, и ее оценка, найденная с использованием упрощенных уравнений равновесия (кривая выделена пунктиром). Отметим, что начальное и конечное значения диапазона изменения p_{off} принадлежат области, где используемые оценки характеристик являются асимптотически точными. Это обуславливает высокую точность вычисления характеристик исходной модели, свидетелем чего являются кривые, приведенные на рис. 9. Более детально решение задачи определения p_{off} показано на рис. 10, где установлено, что $p_{\text{off}} \approx 0,7$.

Вторая часть вычисления доли выгружаемого трафика заключается в необходимости выполнения неравенства $\pi \leq 0,01$. Чтобы установить этот факт, построим верхнюю оценку для π , предположив, что эластичные файлы передаются только с использованием одного канала. Это означает, что все заявки, поступающие в узел доступа, обслуживаются по правилам трафика сервисов реального времени. В данной ситуации величины характеристик можно рассчитать с использованием рекурсии, построенной для мультисервисной модели Эрланга [7, 10, 22, 23]. Полученные результаты подтверждают выполнение неравенства $\pi \leq 0,01$. Таким образом, решение задачи достигнуто для вероятности выгрузки $p_{\text{off}} \approx 0,7$. Аналогичным образом решается задача в ситуации, когда выгрузке подвергается только трафик сервисов реального времени или только эластичный трафик.

Оценка требуемого ресурса узла доступа для заданной нагрузки решается такой же последовательностью, поэтому эту часть исследования опустим.

7. Заключение

Разработана и исследована математическая модель совместного обслуживания трафика реального времени и эластичного трафика данных в мультисервисном узле доступа. Поступление заявок всех сервисных категорий подчиняется пуассоновскому закону, длительности обслуживания заявок на передачу трафика реального времени имеют экспоненциальное распределение, как и объемы передаваемых эластичных данных. Обслуживание эластичных данных происходит в соответствии с положениями дисциплины Processor Sharing. Поступившая заявка на передачу трафика реального времени имеет приоритет в занятии ресурса, уменьшая при необходимости пропускную способность, используемую одним файлом, до заранее заданного минимального значения. Построен марковский процесс, описывающий изменение состояний модели. Приведены определения показателей качества совместного обслуживания поступающих запросов на информационное обслуживание. Сформирована система уравнений статистического равновесия и рассмотрено ее использование для оценки введенных характеристик. Предложен метод приближенной оценки характеристик, основанный на построении системы упрощен-

ных уравнений равновесия. Показано использование разработанных методов для решения задачи оценки объема трафика, выгружаемого в ситуации перегрузки на другие узлы доступа или в другие диапазоны частот, с целью достижения заданных показателей QoS и задачи планирования необходимого по нагрузке объема ресурса передачи информации мультисервисного узла доступа. Полученные результаты также применимы к анализу пропускной способности узлов концентрации абонентской нагрузки в системах спутниковой связи [25], дата-центрах [26], в справочно-информационных центрах [21].

Отметим положительные характеристики разработанного приближенного метода.

- Формирование упрощенных уравнений равновесия основано на использовании локальных законов сохранения, которые выполняются в отдельных макросостояниях модели, выбираемых в соответствии со структурой используемого пространства состояний и физическим смыслом отдельных компонент состояния. Тем самым можно утверждать, что данный подход основан на общих принципах функционирования моделей систем связи, описываемых марковскими процессами, что дает возможность обобщать полученные результаты на другие модели формирования входных потоков заявок и процедуры распределения ресурса, в частности на модели с групповым поступлением заявок и ожиданием.
- Полученные оценки характеристик качества обслуживания поступающих заявок имеют хорошую точность для значений входных параметров, которые соответствуют практическим приложениям, в частности, для значений малых потерь, где происходит решение задачи планирования необходимого по нагрузке объема ресурса передачи информации, и для значений больших потерь, где решается задача оценки объема трафика, выгружаемого на другие узлы доступа или в другие диапазоны частот, с целью достижения заданных показателей QoS.
- Используемые в процессе реализации метода вычислительные алгоритмы основаны на использовании простых рекурсивных процедур.

СПИСОК ЛИТЕРАТУРЫ

1. Study on scenarios and requirements for next generation access technologies. 3GPP Technical Report (TR) 138.913 version 15.0.0 Release 15. 2018.
2. Network Slice Selection Services. 3GPP Technical Specification (TS) 129.531 version 15.5.0. Release 15. 2019.
3. *Evans J., Filsfils C.* Deploying IP and MPLS QOS for multiservice networks. Theory and Practice. Morgan Kaufmann Publishers. Elsevier, 2007. 456 p.
4. *Ross K.W.* Multiservice loss models for broadband telecommunication networks. London: Springer, 1995. 343 p.
5. *Iversen V.B.* Teletraffic Engineering and Network Planning. Technical University of Denmark, May 2010. 370 p.
6. Broadband network traffic. Performance evaluation and design of broadband multiservice networks. Final report of action COST 242 / James Roberts (ed). Lecture notes in computer sciences. Springer, 1996. 584 p.

7. *Степанов С.Н.* Основы телетрафика мультисервисных сетей. М.: Эко – Трендз, 2010. 392 с.
8. *Bonald T., Fewillet M.* Network Performance Analysis. United Kingdom, London, Wiley, 2011. 253 p.
9. *Basharin G.P., Gaidamaka Yu.V., Samouylov K.E.* Mathematical Theory of Teletraffic and Its Application to the Analysis of Multiservice Communication of Next Generation Networks // Autom. Control Comput. Sci. 2013. V. 47. No. 2. P. 62–69.
10. *Степанов С.Н.* Теория телетрафика: концепции, модели, приложения. М.: Горячая линия – Телеком, 2015. 868 с.
11. *Степанов С.Н., Степанов М.С.* Планирование ресурса передачи при совместном обслуживании мультисервисного трафика реального времени и эластичного трафика данных // АиТ. 2017. № 11. С. 79–93.
Stepanov S.N., Stepanov M.S. Planning Transmission Resource at Joint Servicing of the Multiservice Real Time and Elastic Data Traffics // Autom. Remote Control. 2017. V. 78. No. 11. P. 2004–2015.
12. *Shorgin S., Samouylov K., Gaidamaka Y., Chukarin A., Buturlin I., Begishev V.* Modeling Radio Resource Allocation Scheme with Fixed Transmission Zones for Multiservice M2M Communications in Wireless IoT Infrastructure // Lecture Notes Comput. Sci., Springer, Cham. 2015. V. 9012. P. 473–483.
13. *Begishev V., Petrov V., Samouylov A., Moltchanov D., Andreev S., Koucheryavy Y., Samouylov K.* Resource Allocation and Sharing for Heterogeneous Data Collection over Conventional 3GPP LTE and Emerging NB-IoT Technologies // Comput. Communicat. 2018. V. 120. No. 2. P. 93–101.
14. *Naumov V., Gaidamaka Yu., Yarkina N., Samouylov K.* Matrix and Analytical Methods for Performance Analysis of Telecommunication Systems. Springer Nature, Switzerland, 2021.
15. *Kallos G.A., Vassilakis V.G., Moscholios I.D., Logothetis M.D.* Performance modelling of W-CDMA networks supporting elastic and adaptive traffic // Proc. of 4 Int. Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs '06), Ilkley, 2006.
16. *Gudkova I.A., Samouylov K.E.* Modelling a radio admission control scheme for video telephony service in wireless networks // Lecture Notes Comput. Sci. 2012. V. 7469. P. 208–215.
17. *Bonald T., Virtamo J.* A recursive formula for multirate systems with elastic traffic // IEEE Communicat. Lett. 2005. V. 9. No. 8. P. 753–755.
18. *Степанов С.Н., Степанов М.С.* Планирование ресурса передачи информации соединительных линий мультисервисных иерархических сетей доступа // АиТ. 2018. № 8. С. 66–80.
Stepanov S.N., Stepanov M.S. Planning the Resource of Information Transmission for Connection Lines of Multiservice Hierarchical Access Networks. Autom. Remote Control. 2018. V. 79. No. 8. P. 1422–1433.
19. *Степанов С.Н., Степанов М.С.* Эффективный алгоритм оценки требуемого объема ресурса беспроводных систем связи при совместном обслуживании гетерогенного трафика устройств интернета вещей // АиТ. 2019. № 11. С. 108–126.
Stepanov S.N., Stepanov M.S. Efficient Algorithm for Evaluating the Required Volume of Resource in Wireless Communication Systems under Joint Servicing of

- Heterogeneous Traffic for the Internet of Things // Autom. Remote Control. 2019. V. 80. No. 11. P. 2017–2032.
20. *Степанов С.Н., Степанов М.С.* Методы оценки необходимого объема ресурса мультисервисных узлов доступа // *АиТ*. 2020. № 12. С. 129–152.
Stepanov S.N., Stepanov M.S. Methods for Estimating the Required Volume of Resource for Multiservice Access Nodes // Autom. Remote Control. 2020. V. 81. No. 12. P. 2244–2261.
 21. *Stepanov S.N.* Markov models with retrials: The calculation of stationary performance measures based on the concept of truncation // *Math. Comput. Model.* 1999. No. 30. P. 207–228.
 22. *Stepanov S., Stepanov M.* Estimation of the Performance Measures of a Group of Servers Taking into Account Blocking and Call Repetition before and after Server Occupation // *Mathematics*. 2021. V. 9. No. 21. 2811. P. 1–24.
<https://doi.org/10.3390/math9212811>
 23. *Fortet R., Grandjean Ch.* Congestion in a Loss System When Some Calls Want Several Devices Simultaneously // *Electr. Communicat.* 1964. V. 39. No. 4. P. 513–526.
 24. *Kaufman J.S.* Blocking in a shared resource environment // *IEEE Transact. Communicat.* 1981. V. 29. No. 1. P. 1474–1481.
 25. *Chen J., Chang Z., Guo X., Li R., Han Z.* Hamalainen, T. Resource Allocation and Computation Offloading for Multi-Access Edge Computing With Fronthaul and Backhaul Constraints // *IEEE Transactions on Vehicular Technology*. 2021. V. 70. No. 8. P. 8037–8049.
 26. *Stepanov S.N., Andrabli U.M., Stepanov M.S., Ndayikunda J.* Reservation Based Joint Servicing of Real Time and Batched Traffic in Inter Satellite Link // *Proc. of 2020 Systems of Signals Generating and Processing in the Field of on Board Communications*. Moscow. Russia. 2020. P. 1–5.
<https://doi.org/10.1109/IEEECONF48371.2020.9078542>
 27. *Volkov A.O., Korobkina A.V., Stepanov S.N.* Development of Model and Algorithms for Servicing Real-Time and Data Traffic in a Cloud Computing System // *Proc. of 2022 Systems of Signals Generating and Processing in the Field of on Board Communications*. Moscow. Russia. P. 1–6. 2022.
<https://doi.org/10.1109/IEEECONF53456.2022.9744289>

Статья представлена к публикации членом редколлегии А.И. Ляховым.

Поступила в редакцию 19.07.2023

После доработки 29.08.2023

Принята к публикации 04.09.2023