

© 2020 г. С.Н. СТЕПАНОВ, д-р техн. наук (stpnsrg@gmail.com),
М.С. СТЕПАНОВ, канд. техн. наук (mihstep@yandex.ru)
(Московский технический университет связи и информатики)

МЕТОДЫ ОЦЕНКИ НЕОБХОДИМОГО ОБЪЕМА РЕСУРСА МУЛЬТИСЕРВИСНЫХ УЗЛОВ ДОСТУПА

Построена и исследована математическая модель распределения ресурса передачи информации мультисервисного узла доступа. В модели рассматривается произвольное число потоков мультимедийного трафика, которые различаются интенсивностью поступления заявок, величиной ресурса, используемого для обслуживания одной заявки, и временем занятия ресурса. Интервалы времени между поступлением заявок имеют экспоненциальное распределение с параметром, зависящим от числа заявок рассматриваемого потока, находящихся на обслуживании. Построен рекурсивный алгоритм оценки характеристик. Установлены соотношения между интегральными и потоковыми характеристиками качества обслуживания заявок. Построен эффективный алгоритм оценки объема ресурса, требуемого для обслуживания заданных потоков трафика с необходимым качеством. Эффективность расчетной процедуры достигается в результате организации рекурсии по объему ресурса и использования нормированных значений вероятностей состояний. Рассмотрено решение задачи оценки необходимого объема ресурса для модели мультисервисного узла, допускающего использование механизмов резервирования ресурса и его динамического распределения при обслуживании эластичного трафика. Приведены численные примеры, иллюстрирующие особенности реализации построенных расчетных процедур.

Ключевые слова: мультисервисный узел доступа, марковские модели, система уравнений равновесия, рекурсивные алгоритмы, резервирование, эластичный трафик.

DOI: 10.31857/S0005231020120089

1. Введение

Одной из основных задач моделирования сетей и систем связи является разработка научно обоснованных средств оценки необходимого объема ресурса передачи информации [1–8]. Результаты моделирования также играют существенную роль в обосновании действий оператора, направленных на повышение эффективности работы сети и улучшение качества обслуживания пользователей [5–8]. К таким действиям относятся: дифференцированное обслуживание пользователей, резервирование ресурса, динамическое распределение ресурса и т.д. Особенно важно найти решение перечисленных задач для мультисервисных узлов доступа, реализующих функцию концентрации возникающих информационных потоков. Оцениваемый ресурс определим через величину скорости передачи информационного потока, необходимой для

обеспечения заказанного сервиса. В сетях фиксированной связи использование этого определения не вызывает особых затруднений [9–13]. В сетях беспроводной связи битовая скорость является функцией технологии и условий передачи информации, в частности расстояния до базовой станции, играющей роль мультисервисного узла доступа [14–16]. На уровне поступления заявок распределение ресурса для обоих видов связи исследуется схожими методами, поэтому для простоты далее будем предполагать, что рассматривается фиксированная связь. Разработка методов оценки объема ресурса включает в себя решение ряда задач, возникающих из-за необходимости учета специфики формирования и обслуживания мультисервисных информационных потоков. Обсудим их и дадим им более точную формулировку.

В мультисервисных сетях по определению рассматривается процесс совместного обслуживания нескольких потоков трафика. Наличие большого числа потоков с разными характеристиками усложняет задачу планирования необходимого объема ресурса из-за неопределенности в выборе нормативных показателей. Для этих целей обычно рассматривают значения интегральных характеристик. В их числе: максимальная доля потерянных заявок, доля потерянного трафика и т.д. При этом качество обслуживания отдельных потоков не оценивается. Возникающие трудности можно устранить, если найти зависимость между интегральными характеристиками и характеристиками обслуживания отдельных потоков. Решение этой задачи служит основанием для выбора и использования метрики при оценке достаточности ресурса.

Следующая задача относится к построению эффективного алгоритма оценки величины ресурса. Эффективность в данном контексте означает достижение минимальных затрат вычислительных ресурсов на реализацию алгоритма и его стабильность. Все эти требования необходимы для применения алгоритма в программно-аналитических продуктах типа калькуляторов сетевой инфраструктуры. Этого результата можно добиться, если организовать рекурсию по объему ресурса и при проведении вычислений использовать только нормированные значения вероятностей состояний.

Результаты исследований [5–8] показывают, что совместное занятие ресурса несколькими информационными потоками, с одной стороны, может привести к его перераспределению и ухудшению качества обслуживания отдельных потоков, а с другой стороны, может быть использовано для повышения эффективности его занятия, например при пересылке эластичного трафика. Отмеченные положительные и отрицательные стороны совместного обслуживания трафика должны быть учтены при построении модели мультисервисного узла доступа и в дальнейшем приняты во внимание при разработке методов оценки необходимого объема ресурса.

Решения перечисленных задач будут рассмотрены в настоящей работе на примере модели мультисервисного узла доступа, в которой интервалы времени между поступлением заявок зависят от числа заявок рассматриваемого потока, находящихся на обслуживании. Помимо базового варианта модели будут рассмотрены ее обобщения, допускающие использование механизмов резервирования ресурса и его динамического распределения при пересылке эластичного трафика. Решение отдельных задач исследовалось в более ранних публикациях в основном для модели с пуассоновскими потоками зая-

вок. Так, рекурсивные алгоритмы оценки характеристик рассматривались в [1–4], вопросы планирования ресурса исследовались в [9–13], использование механизмов резервирования анализировались в [5, 8, 16], особенности обслуживания эластичного трафика изучались в [8, 10, 11], совместная передача трафика реального времени и данных представлена в [8, 11, 4–18]. Здесь эти результаты будут получены в более общей постановке и более эффективными средствами.

Работа имеет следующую структуру. В разделах 2 и 3 построены соответственно функциональная и математическая модели мультисервисного узла доступа. В следующем разделе сформулированы определения основных характеристик качества обслуживания поступающих потоков заявок. Рекурсивный алгоритм их оценки рассмотрен в разделе 5. Там же установлены соотношения между интегральными и потоковыми характеристиками качества обслуживания заявок, которые упрощают выбор метрики при решении задач планирования ресурса узла доступа. Эффективный алгоритм оценки требуемого числа каналов передачи информации построен в разделе 6. Эффективность расчетной процедуры достигается в результате организации рекурсии по объему ресурса и использовании нормированных значений вероятностей состояний. В разделе 7 решение задачи оценки необходимого числа каналов рассмотрено для модели мультисервисного узла, допускающего использование механизмов резервирования ресурса и его динамического распределения при пересылке эластичного трафика.

2. Функциональная модель

Мультисервисный узел доступа реализует функции агрегации и передачи данных от разнообразных источников информационной нагрузки. К такому можно отнести абонентские терминалы, мультимедийные компьютеры, видеоборудование и т.д. По запросу пользователя устанавливается виртуальное соединение (ВС) и осуществляется передача данных в виде последовательности пакетов. В большинстве случаев эта последовательность представляет локально периодический поток. За периодом генерации пакетов, обычно происходящей с максимальной для рассматриваемого источника скоростью, следует период времени, когда пакеты не поступают. Такая структура потока позволяет реализовать принцип статистического мультиплексирования и уменьшить потребности в ресурсе передачи информации. Виртуальное соединение выполняет пересылку информационных потоков, относящихся к некоторому конечному набору сервисов. Сюда входят передача речи и разных форм видео, обмен файлами и т.д. Будем предполагать, если это не оговаривается особо, что рассматриваются коммуникационные сервисы реального времени, требующие для своего обслуживания фиксированную скорость передачи информации.

Важными составными частями мультисервисного узла доступа являются: процедура контроля доступа (CAC — Call Admission Control), буфер и высокоскоростная линия концентрации трафика. Процедура CAC может быть реализована с использованием протокола RSVP (Resource reSerVation Protocol) в рамках архитектуры управления ресурсами IntServ (Integrated Ser-

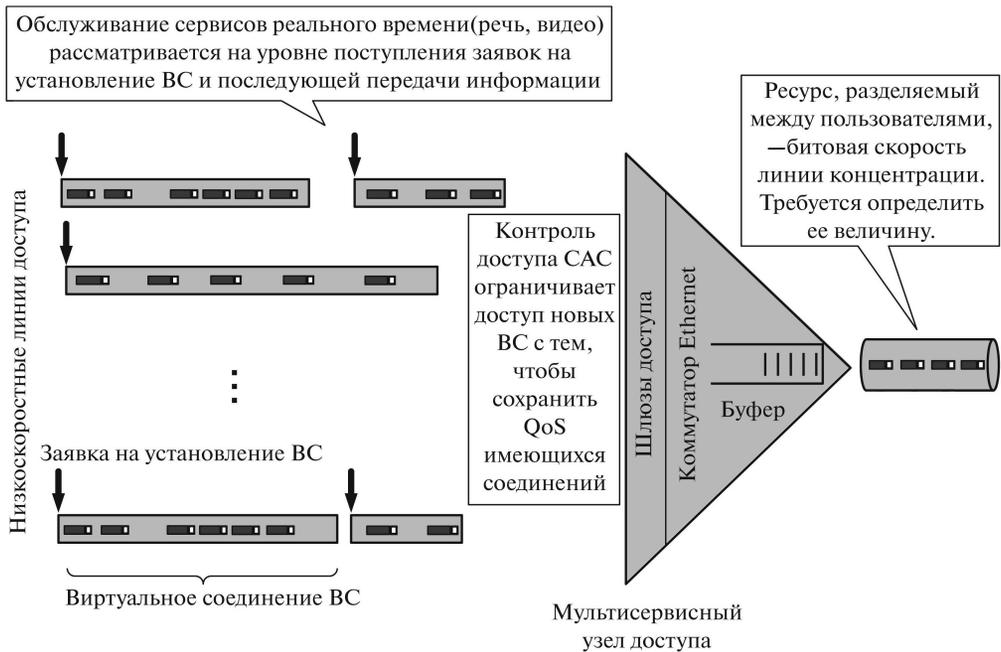


Рис. 1. Функциональная модель мультисервисного узла доступа.

vices). Формализованное описание САС основано на введении дескриптора трафика, который задает потребность в ресурсе передачи информации у поступившей заявки, и сравнимого дескриптора, оценивающего свободный ресурс [19]. Решение о приеме заявки принимается после сравнения значений обоих показателей. В качестве дескриптора трафика можно взять значение пиковой интенсивности поступления информации в потоке, ассоциированном с обслуживанием заявки. Это решение отличается простотой, однако существенно переоценивает потребности в ресурсе, не учитывает возможности статистического мультиплексирования трафика на уровне пакетов и поэтому неэффективно. Эти недостатки отчасти устраняет использование в качестве дескриптора понятия эффективной скорости передачи информации [20]. Ее величина лежит между средней и пиковой скоростями передачи информации анализируемого потока и учитывает свойство статистического мультиплексирования. Заявка принимается, если сумма значений используемого ресурса r_o и ресурса, требуемого для обслуживания поступившей заявки r_n , не превосходит имеющегося ресурса C . Таким образом, для приема заявки необходимо выполнение неравенства: $r_o + r_n < C$. Буфер получает пакеты из установленных соединений и передает их один за другим по линии. Он служит интерфейсом между поступающими потоками пакетов и высокоскоростной линией и сглаживает всплески информационной нагрузки. Функциональная модель мультисервисного узла доступа показана на рис. 1.

Построенная функциональная модель мультисервисного узла в упрощенной форме описывает процесс занятия и использования ресурса передачи информации. Она освобождена от конкретики технологий и протоколов. Их

влияние переводится с микро- на макроуровень в форме параметров, зависящих от состояния сети и действия разного рода механизмов, управляющих процессом обслуживания абонентов. Она служит своего рода интерфейсом между исследуемой системой связи и моделью, с помощью которой далее будут найдены показатели качества обслуживания заявок, используемые для оценки требуемой по нагрузке скорости линии. Отметим, что могут применяться и более сложные модели формализованного описания условий приема заявок в соответствии с процедурой САС. Часть из них, основанная на применении процедуры резервирования ресурса, будет рассмотрена в разделе 7.

3. Математическая модель

Обозначим через C скорость передачи информации, выраженную в битах в секунду, обеспечиваемую линией концентрации абонентского трафика (см. рис. 1). Назовем единицей ресурса u минимальное требование к скорости от поступающих заявок (скорость одного виртуального канала). Предположим, что значение C нацело делится на u , и обозначим через $v = \frac{C}{u}$ скорость передачи линии, выраженную в виртуальных каналах. Линия используется для обслуживания n потоков заявок на передачу трафика сервисов реального времени. Интервал времени между последовательным поступлением заявок k -го потока имеет экспоненциальное распределение с параметром $\lambda_k(i_k) = \alpha_k + i_k \beta_k$, $k = 1, \dots, n$, где i_k — число заявок k -го потока, находящихся на обслуживании. Для допуска заявки k -го потока требуются b_k виртуальных каналов, которые резервируются на случайное время, имеющее экспоненциальное распределение с параметром μ_k , и используются для передачи информационного потока, ассоциированного с рассматриваемой заявкой.

В литературе [4, 6–8] рассмотренная модель входного потока заявок носит название *поток ВРР* (Bernoulli-Poisson-Pascal) по трем известным частным случаям. Для потока Бернулли $\lambda_k(i_k) = (s_k - i_k)\zeta_k$, где s_k — число пользователей, создающих k -й поток заявок, а ζ_k — параметр экспоненциального распределения времени между последовательными поступлениями заявок от одного пользователя. Для пуассоновского потока $\lambda_k(i_k) = \lambda_k$ и значение λ_k не зависит от числа абонентов, находящихся на обслуживании. Для потока Паскаля $\lambda_k(i_k) = (s_k + i_k)\zeta_k$, где s_k — положительное целое число и $\zeta_k > 0$.

Наличие двух параметров в модели входного потока дает возможность с большей точностью аппроксимировать поступление заявок. Процесс поступления заявок зависит от числа пользователей, находящихся на обслуживании. Тем самым учитывается зависимость поступления заявок от загрузки линии связи. Рассмотренную модель рекомендуется использовать в тех ситуациях, когда необходимо выделить небольшие группы так называемых *тяжелых* абонентов, создающих существенный объем потенциального трафика. Необходимость учета этой особенности формирования входного потока особенно актуальна для перспективных сетей подвижной связи. Эти системы работают в диапазоне высоких частот. Здесь размеры соты невелики, и число активных абонентов влияет на величину интенсивности поступающего потока заявок. Аналогичная ситуация наблюдается в местах городской застройки при обслуживании небольших территорий (офисы, отдельные здания), где

используются малые соты (фемтосоты, пикосоты, метросоты и т.д.) с предельно малым (до нескольких десятков) числом абонентов. В дальнейшем, не теряя общности изложения материала, ограничимся рассмотрением модели входного потока Бернулли. По аналогии с моносервисным случаем эта модель называется мультисервисной моделью Энгсета.

4. Характеристики обслуживания заявок

Достаточность ресурса линии оценим долей потерянных заявок, а эффективность его использования — средним числом занятых единиц ресурса. Для оценки этих характеристик достаточно знать долю времени пребывания линии в состоянии с известным числом заявок каждого потока, находящихся на обслуживании. Выбор характеристик определяет состояние модели в виде вектора (i_1, \dots, i_n) , где i_k — число обслуживаемых заявок k -го потока. Значения i_k ограничены пропускной способностью линии $\sum_{k=1}^n i_k b_k \leq v$ и числом абонентов в каждой группе $i_k \leq s_k$, $k = 1, \dots, n$. Векторы (i_1, \dots, i_n) , удовлетворяющие приведенным неравенствам, определяют пространство S состояний модели.

Динамика изменения состояний модели во времени описывается случайным марковским процессом $r(t) = (i_1(t), \dots, i_n(t))$, где $i_k(t)$ — число заявок k -го потока, находящихся в момент t на обслуживании. Пусть U_k — множество состояний $(i_1, \dots, i_n) \in S$, удовлетворяющих условию $i_1 b_1 + \dots + i_n b_n + b_k > v$. В каждом из состояний U_k поступившая заявка k -го потока получает отказ. Обозначим через $p(i_1, \dots, i_n)$ стационарную вероятность состояния (i_1, \dots, i_n) . Она имеет интерпретацию доли времени пребывания $r(t)$ в (i_1, \dots, i_n) , что позволяет использовать $p(i_1, \dots, i_n)$ для оценки характеристик модели.

Из обратимости марковского процесса $r(t)$ следует свойство мультипликативности. Для всех $(i_1, \dots, i_n) \in S$

$$(1) \quad p(i_1, \dots, i_n) = \frac{1}{N} \times \frac{\prod_{j=0}^{i_1-1} (s_1 - j) \gamma_1^{i_1}}{i_1!} \dots \frac{\prod_{j=0}^{i_n-1} (s_n - j) \gamma_n^{i_n}}{i_n!},$$

где $\gamma_k = \zeta_k / \mu_k$ — среднее число заявок, поступающих от одного абонента k -го потока за среднее время обслуживания заявки $1/\mu_k$. Далее это время будет принято за единицу и отдельно в модели не рассматривается, а N — нормировочная константа

$$N = \sum_{(i_1, \dots, i_n) \in S} \frac{\prod_{j=0}^{i_1-1} (s_1 - j) \gamma_1^{i_1}}{i_1!} \dots \frac{\prod_{j=0}^{i_n-1} (s_n - j) \gamma_n^{i_n}}{i_n!}.$$

Мультипликативное соотношение и алгоритмы, полученные на его основе, не зависят от вида функции распределения времени обслуживания заявки и функции распределения времени между поступлением заявок от одного

абонента [21—23]. Требуется только, чтобы соответствующие времена не зависели друг от друга. Это существенно расширяет область применения полученных расчетных выражений.

Качество обслуживания заявок k -го потока оценим долей времени недоступности ресурса передачи $\pi_{t,k}$; долей потерянных заявок $\pi_{c,k}$; средним числом потенциальных соединений $a_k = \frac{s_k \gamma_k}{1 + \gamma_k}$ (среднее число соединений в отсутствие потерь); средним числом занятых единиц ресурса m_k ; средним числом обслуживаемых заявок $y_k = m_k / b_k$; долей потерянных соединений $\pi_{\ell,k} = (a_k - y_k) / a_k$ и интенсивностью входного потока заявок $\Lambda_k = (s_k - y_k) \gamma_k$:

$$(2) \quad \begin{aligned} \pi_{t,k} &= \sum_{(i_1, \dots, i_n) \in U_k} p(i_1, \dots, i_n); \\ \pi_{c,k} &= \frac{\sum_{(i_1, \dots, i_n) \in U_k} p(i_1, \dots, i_n)(s_k - i_k)}{\sum_{(i_1, \dots, i_n) \in S} p(i_1, \dots, i_n)(s_k - i_k)}; \\ m_k &= \sum_{(i_1, \dots, i_n) \in S} p(i_1, \dots, i_n) i_k b_k. \end{aligned}$$

Из формулы Литтла следует $y_k = \Lambda_k (1 - \pi_{c,k})$. Из определений характеристик получаем $\pi_{c,k} = \pi_{\ell,k} (1 + \gamma_k) / (1 + \pi_{\ell,k} \gamma_k)$. Если для k -го потока заявок известно значение одной из введенных ранее характеристик y_k , m_k , $\pi_{c,k}$, $\pi_{\ell,k}$, Λ_k и величины входных параметров s_k , a_k (или s_k , γ_k), то, используя приведенные выше выражения, можно найти значения оставшихся характеристик. Величина a_k определяет потенциальное число соединений и будет далее использоваться при описании входных параметров модели в процессе проведения вычислений. Предположим, что в результате измерений или расчетов стали известны значения $\pi_{c,k}$, s_k , a_k . Величины оставшихся характеристик могут быть найдены из выражений:

$$(3) \quad \begin{aligned} \gamma_k &= \frac{a_k}{s_k - a_k}; \quad \pi_{\ell,k} = \frac{\pi_{c,k}(s_k - a_k)}{s_k - a_k \pi_{c,k}}; \quad y_k = \frac{a_k s_k (1 - \pi_{c,k})}{s_k - a_k \pi_{c,k}}; \\ m_k &= \frac{a_k b_k s_k (1 - \pi_{c,k})}{s_k - a_k \pi_{c,k}}; \quad \Lambda_k = \frac{a_k s_k}{s_k - a_k \pi_{c,k}}. \end{aligned}$$

5. Оценка характеристик и их свойства

Алгоритм оценки характеристик модели основан на использовании показателей пребывания $r(t)$ во множестве агрегированных состояний модели S_i , $i = 0, 1, \dots, v$, в которое включены состояния $(i_1, \dots, i_n) \in S$, удовлетворяющие условию $i_1 b_1 + \dots + i_n b_n = i$. Пусть $p(i)$ — вероятность занятости i единиц ресурса, а $y_k(i)$ — среднее число заявок k -го потока, находящихся на обслуживании в ситуации, когда заняты i единиц ресурса

$$(4) \quad p(i) = \sum_{(i_1, \dots, i_n) \in S_i} p(i_1, \dots, i_n), \quad y_k(i) = \sum_{(i_1, \dots, i_n) \in S_i} p(i_1, \dots, i_n) i_k.$$

Используя $p(i)$ и $y_k(i)$, можно оценить все введенные ранее характеристики модели. Из определений $\pi_{t,k}$ и $\pi_{c,k}$ следует

$$(5) \quad \pi_{t,k} = \sum_{i=v-b_k+1}^v p(i); \quad \pi_{c,k} = \frac{\sum_{i=v-b_k+1}^v (p(i) s_k - y_k(i))}{\sum_{i=0}^v (p(i) s_k - y_k(i))}.$$

Значения y_k , m_k , $\pi_{\ell,k}$, Λ_k находятся с помощью полученных ранее формул (3) их косвенной оценки через значения $\pi_{t,k}$, $\pi_{c,k}$ и величины s_k , a_k . Построим рекурсивный алгоритм оценки $p(i)$ и $y_k(i)$, $i = 0, 1, \dots, v$. Из обратимости $r(t)$ для $(i_1, \dots, i_n) \in S$ следует соотношение детального баланса

$$p(i_1, \dots, i_k - 1, \dots, i_n)(s_k - i_k + 1)\gamma_k = p(i_1, \dots, i_k, \dots, i_n) i_k.$$

Просуммируем полученное выражение по всем $(i_1, \dots, i_n) \in S_i$. Используя определения S_i , $p(i)$, $y_k(i)$, находим искомую рекурсию

$$(6) \quad Y_k(i) = P(i - b_k) s_k \gamma_k - Y_k(i - b_k) \gamma_k, \quad k = 1, \dots, n;$$

$$P(i) = \frac{1}{i} \sum_{k=1}^n b_k Y_k(i).$$

Положим значение¹ $P(0) = 1$. По определению $Y_k(0) = 0$, $k = 1, \dots, n$. Выразим значения $Y_k(i)$, $k = 1, \dots, n$, $P(i)$ через $P(0)$, используя рекурсию (6) и последовательно увеличивая i от 1 до v . Находим величину нормировочной константы $N = \sum_{i=0}^v P(i)$, нормированные значения вероятностей $p(i) = P(i)/N$ и вспомогательных функций $y_k(i) = Y_k(i)/N$, $i = 0, 1, \dots, v$. Рассчитываем значения характеристик с помощью (5), определения и формулы (3). При реализации рекурсии для больших значений v может теряться точность вычислений. Это происходит из-за наличия отрицательного слагаемого в (6). В данной ситуации для расчета характеристик рекомендуется использовать алгоритм свертки [7].

Воспользуемся разработанными алгоритмами для анализа зависимости характеристик качества обслуживания заявок от параметров возникающих информационных потоков и условий допуска заявок к занятию ресурса. Обозначим через $\rho = \frac{1}{v} \sum_{k=1}^n a_k b_k$ потенциальную загрузку одного виртуального канала (в.к.). Будем предполагать, что b_k занумерованы в порядке возрастания их значений. Рассмотрим модель узла для следующих значений входных параметров: $v = 100$; $n = 4$; $b_1 = 1$; $b_2 = 5$; $b_3 = 10$; $b_4 = 20$. Величины v и b_k выражены в виртуальных каналах. Примем, что $a_k = v\rho/nb_k$, $s_k = \lfloor a_k \rfloor + 10$, $k = 1, 2, 3, 4$. Отсюда следует, что $\gamma_k = \frac{a_k}{s_k - a_k}$ и потенциальная загрузка ресурса $a_k b_k$ у всех потоков одинакова и равна $v\rho/n$. На рис. 2 показана зависимость $\pi'_{c,k} = \pi_{c,k}/b_k$ от увеличения ρ .

В ситуации, когда $\rho \approx 1$, выполняется соотношение $\pi_{c,i}/b_i \approx \pi_{c,j}/b_j$. Кроме этого, если $\rho < 1$ и $b_i > b_j$, то $\pi_{c,i}/b_i > \pi_{c,j}/b_j$, если $\rho > 1$ и $b_i > b_j$, то

¹ Здесь и далее прописные буквы используются для обозначения ненормированных величин вероятностей и характеристик, а строчные — для нормированных.

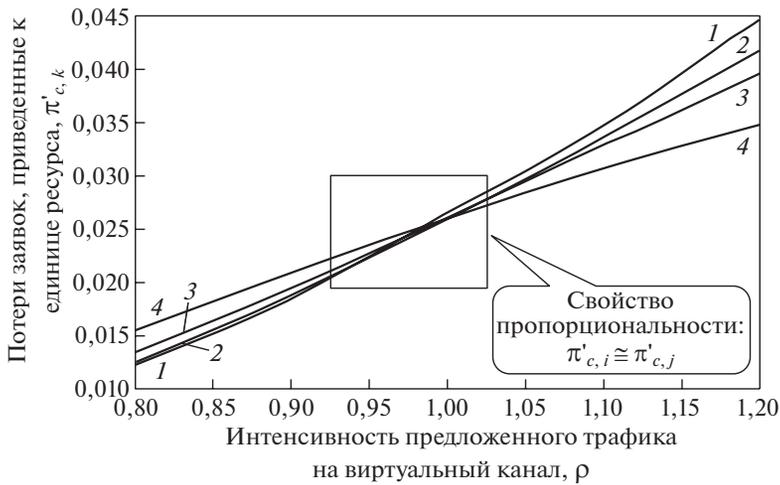


Рис. 2. Зависимость потерь заявок, приведенных к единице ресурса, от потенциальной загрузки единицы ресурса.

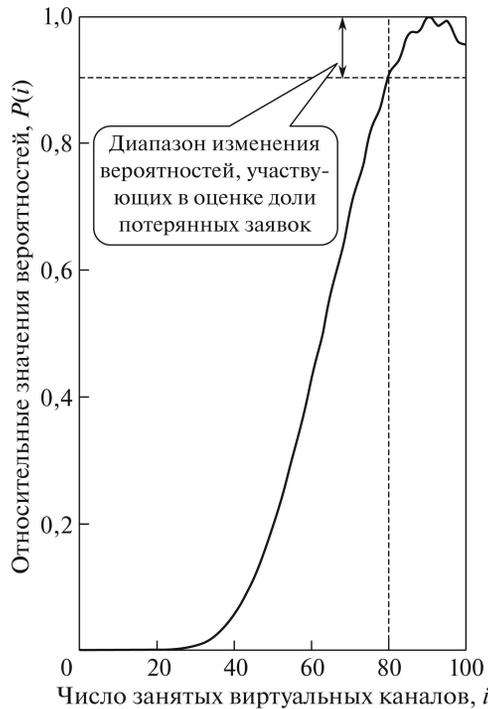


Рис. 3. Зависимость относительных значений $P(i)$ от i при $\rho = 1$.

$\pi_{c,i}/b_i < \pi_{c,j}/b_j$. Дадим пояснение этим результатам. Используя формулу мультипликативности и определение $\pi_{c,k}$, можно показать, что величина характеристики совпадает со значением $\pi_{t,k}$ доли времени недоступности ресурса передачи для заявок k -го потока, рассчитанным для тех же параметров,

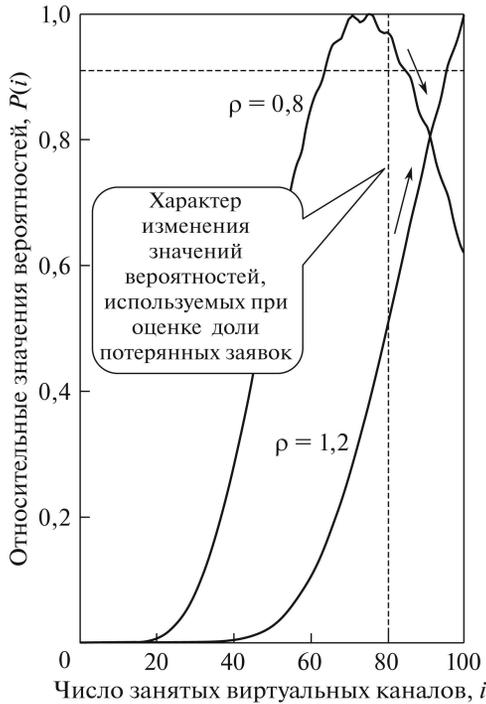


Рис. 4. Зависимость относительных значений $P(i)$ от i при $\rho = 0,8$ и $\rho = 1,2$.

но с числом абонентов, формирующих k -й поток, равным $s_k - 1$, т.е. на единицу меньше, чем в анализируемой модели. Отмеченное изменение числа абонентов не сказывается сильно на свойствах и величине $\pi_{c,k}$. По этой причине зависимость поведения $\pi_{c,k}$ от ρ можно исследовать на примере зависимости $\pi_{t,k}$ от ρ . По определению $\pi_{t,k} = p(v) + p(v-1) + \dots + p(v-b_k+1)$. В этой сумме представлены b_k значений вероятностей состояний с максимальным занятым ресурсом. Зависимость $P(i)$, отнормированных по максимальному значению $P(i)$, показана на рис. 3 ($\rho = 1$) и рис. 4 ($\rho = 0,8$ и $\rho = 1,2$).

Из результатов расчетов видно, что в области $\rho \approx 1$ величины $P(i)$, участвующие в определении $\pi_{t,k}$, примерно равны, а для ρ , больших или меньших единицы, выполняются свойства монотонного убывания (для $\rho < 1$) и возрастания (для $\rho > 1$). Эти свойства и определяют характер поведения $\pi'_{c,k}$, отмеченный на рис. 2.

Рассмотрим, какие практические приложения имеет полученный результат. При обслуживании потоков мультисервисного трафика возникают проблемы с выбором метрик для оценки достаточности ресурса. Трудности связаны с тем, что обычно нормируются интегральные характеристики качества обслуживания заявок, например, такие, как максимальная доля потерь заявок, а какие при этом величины потерь у отдельных потоков, остается неясным. Воспользуемся отмеченными выше свойствами совместного обслуживания заявок и покажем, что существуют простые соотношения, которые могут помочь решить сформулированную проблему.

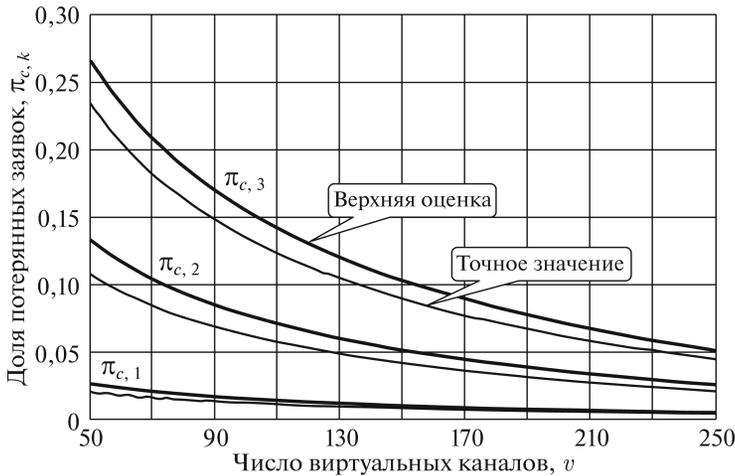


Рис. 5. Точные и приближенные значения потерь в зависимости от v .

Обычно в задачах планирования $\rho \lesssim 1$. Примем далее это допущение. Предположим, что достаточность ресурса планируется исходя из ограничения на максимальную долю потерянных заявок $\max_k \pi_{c,k} = \pi_{c,n} \leq \pi$. Из свойства пропорциональности для $\rho \lesssim 1$ находим оценки сверху для потерь заявок всех потоков

$$(7) \quad \pi_{c,k} \lesssim \pi_{c,n} \frac{b_k}{b_n} = \pi \frac{b_k}{b_n}.$$

На рис. 5 показаны результаты анализа погрешности приближенного расчета потерь при заданном значении максимальной доли потерянных заявок. Исходные данные те же, что были использованы при вычислении данных, представленных на рис. 3, за исключением v , которое менялось от 50 до 300. Нагрузка на канал $\rho = 0,8$. Приближенные значения $\pi_{c,k}$ найдены из соотно-

Анализ погрешности приближенного расчета потерь исходя из известного значения максимальной доли потерянных заявок

| v ед.рес. | π_1 | | π_2 | | π_3 | |
|----------------|----------|----------|----------|----------|----------|----------|
| | Точно | Прибл. | Точно | Прибл. | Точно | Прибл. |
| 50 | 0,032887 | 0,034599 | 0,170260 | 0,172995 | 0,348522 | 0,345990 |
| 75 | 0,029118 | 0,029235 | 0,144857 | 0,146175 | 0,293293 | 0,292350 |
| 100 | 0,026561 | 0,026081 | 0,129868 | 0,130405 | 0,260794 | 0,260809 |
| 125 | 0,024229 | 0,023859 | 0,118881 | 0,119296 | 0,238268 | 0,238592 |
| 150 | 0,022739 | 0,022160 | 0,111309 | 0,110798 | 0,222043 | 0,221596 |
| 175 | 0,021829 | 0,020960 | 0,105821 | 0,104798 | 0,209939 | 0,209596 |
| 200 | 0,020937 | 0,019936 | 0,101177 | 0,099679 | 0,200654 | 0,199357 |
| 225 | 0,019929 | 0,019057 | 0,096631 | 0,095287 | 0,191666 | 0,190574 |
| 250 | 0,019425 | 0,018426 | 0,093898 | 0,092132 | 0,185390 | 0,184264 |
| 275 | 0,018968 | 0,017774 | 0,091270 | 0,088869 | 0,179780 | 0,177738 |
| 300 | 0,018323 | 0,017197 | 0,088150 | 0,085986 | 0,173754 | 0,171971 |

шений (7) и выделены на рисунке жирной линией. Полученные приближенные значения потерь отличаются высокой точностью особенно для больших значений ресурса передачи информации.

Особенно высокую точность оценки (7) имеют в области $\rho \approx 1$. Соответствующие результаты приведены в таблице для тех же значений параметров, что использовались для рис. 5.

6. Эффективный алгоритм оценки объема ресурса

Рассмотрим использование рекурсии (6) для оценки v скорости линии концентрации мультисервисного трафика, требуемой для обслуживания заданных потоков трафика с требуемым качеством. Предложенный трафик характеризуется величинами a_k , b_k , s_k , фиксированными на время решения задачи. Достаточность ресурса оценивается из достижения требуемых значений $\pi = \max_k \pi_{c,k}$. Понятно, что сформулированная задача может быть решена методом перебора. Этот подход имеет два существенных недостатка. Во-первых, для каждого промежуточного значения объема ресурса r вычисляются все ненормированные вероятности состояний $P(i)$, $i = 0, 1, \dots, r$, которые затем нормируются. Хотя для оценки π достаточно знать только $b = \max_k b_k$ нормированных вероятностей состояний с максимальным занятым ресурсом. Тем самым значительно увеличивается объем вычислительной работы при определении v . Во-вторых, реализация рекурсии (6) может привести к вычислительным сложностям из-за применения в (6) относительных значений $P(i)$, выраженных через $P(0)$. Значение $p(0)$ быстро стремится к нулю с ростом v . Эта характеристика (6) затрудняет реализацию рекурсии в программно-аналитических продуктах типа калькуляторов сетевой инфраструктуры.

Построим алгоритм оценки v , свободный от перечисленных недостатков. Рекурсия будет вестись по объему имеющегося ресурса r , и при ее реализации будут использоваться только b нормированных вероятностей состояний модели с максимальным числом занятых каналов. Обозначим через r переменное значение ресурса линии, а через $p_r(i)$, $y_{k,r}(i)$ обозначим зависимость от r для вероятностей состояний $p(i)$ и вспомогательных функций $y_k(i)$. Аналогично через $y_k(r)$ и $\pi_{c,k}(r)$ обозначим зависимость от r для среднего числа обслуживаемых заявок y_k и доли потерянных заявок $\pi_{c,k}$. Реализация алгоритма включает в себя следующие шаги.

1. Положим $p_0(0) = 1$. Из (4) следуют соотношения $y_{k,0}(0) = 0$, $k = 1, \dots, n$.
2. Для каждого фиксированного $r = 1, 2, \dots$ находим $\min(b, r + 1)$ нормированных вероятностей $p_r(i)$, $i = r, r - 1, \dots, \max(r - b + 1, 0)$ и значений вспомогательных функций $y_{k,r}(i)$, $k = 1, \dots, n$, $i = r, r - 1, \dots, \max(r - b + 1, 0)$ с максимальным занятым ресурсом. Для этого используются аналогичные вероятности и вспомогательные функции, полученные на предыдущем шаге. Вначале находится нормировочная константа

$$1 + \frac{1}{r} \sum_{k=1}^n \left\{ p_{r-1}(r - b_k) s_k - y_{k,r-1}(r - b_k) \right\} b_k \gamma_k = 1 + S_{r-1},$$

затем нормированные значения вероятностей

$$p_r(r) = \frac{S_{r-1}}{1 + S_{r-1}}, \quad i = r;$$

$$p_r(i) = \frac{p_{r-1}(i)}{1 + S_{r-1}}, \quad i = r-1, r-2, \dots, \max(r-b+1, 0)$$

и вспомогательных функций

$$y_{k,r}(i) = \frac{y_{k,r-1}(i)}{1 + S_{r-1}}, \quad k = 1, \dots, n; \quad i = r, r-1, \dots, \max(r-b+1, 0);$$

$$y_k(r) = \frac{y_k(r-1)}{1 + S_{r-1}} + y_{k,r}(r), \quad k = 1, \dots, n.$$

3. Далее рассчитывается значение функционала, используемого для оценки достаточности ресурса. В рассматриваемом случае таковым является $\pi = \max_k \pi_{c,k}$. Выражение для потерь заявок k -го потока для объема ресурса r находится с использованием соотношения (5) и имеет вид

$$\pi_{c,k}(r) = \frac{\sum_{i=r-b_k+1}^r (p_r(i) s_k - y_{k,r}(i))}{s_k - y_k(r)}.$$

Величина π сравнивается с его нормативным значением, и если требуемый уровень потерь не достигнут, то число виртуальных каналов увеличивается на единицу и расчеты повторяются начиная с п. 2.

Для расчета характеристик и выполнения следующего шага достаточно знать b вероятностей состояний и соответствующее число значений вспомогательных функций с максимальным числом занятых каналов. Количество используемых параметров не зависит от объема ресурса. При проведении вычислений не возникает проблем с переполнением или исчезновением порядка, поскольку расчеты выполняются только с нормированными значениями вероятностей состояний модели с максимальным числом занятых каналов. По условиям анализа модели они имеют наибольшие относительные значения. По сравнению с традиционным методом оптимизированный алгоритм сокращает расчетную работу при оценке требуемого числа каналов мультисервисного узла доступа примерно в $v/(2 + \frac{b}{n})$ раз.

7. Оценка ресурса в условиях дифференцированного обслуживания

В исследуемой модели мультисервисного узла доступа все потоки имеют одинаковые условия использования ресурса передачи информации. Встроим в модель узла процедуры дифференцированного обслуживания потоков трафика и воспользуемся появляющимися возможностями для решения следующих двух задач. Первая из них связана с устранением отрицательных последствий неконтролируемого перераспределения ресурса в пользу отдельных потоков. Вторая — относится к анализу процедур управления ресурсом,

направленных на повышения эффективности его занятия. Рассмотрим, как в этих условиях решается задача оценки объема ресурса, необходимого по нагрузке и качеству обслуживания пользователей услуг связи, а также как происходит выбор параметров управления ресурсом.

Начнем с первой задачи. Необходимость ее решения возникает при совместной передаче неоднородного трафика в единой транспортной среде. Заявки с малыми требованиями к ресурсу вытесняют из процесса обслуживания заявки, имеющие существенные требования к скорости передачи. Самый простой способ выравнивания показателей обслуживания или в предоставлении преимущества в занятии ресурса заключается в использовании механизмов его резервирования для выделенной группы потоков. Интерес представляют те способы, которые допускают возможность относительно простой реализации и оценки эффективности применения. Этими качествами обладает процедура резервирования, основанная на фильтрации поступающих потоков заявок, зависящей от номера потока и степени загрузки ресурса.

Вернемся к математическому описанию модели мультисервисного узла доступа, изложенному в разделе 3, и внесем в него изменения, относящиеся к реализации процедуры резервирования. Обозначим через $\varphi_k(i)$ вероятностную функцию, которая будет применяться для фильтрации процесса доступа поступающих заявок к ресурсу в зависимости от значений k номера потока и i общего числа занятых единиц ресурса. Заявка k -го потока, поступившая в момент занятости i единиц ресурса, принимается к обслуживанию с вероятностью $1 - \varphi_k(i)$, а с противоположной вероятностью $\varphi_k(i)$ заявка получает отказ и не возобновляется. В остальном схема функционирования узла не меняется. Сохраним для параметров и характеристик обобщенной модели те же обозначения, что были использованы в разделе 3.

Динамика изменения состояний (i_1, \dots, i_n) описывается многомерным марковским процессом $r(t) = (i_1(t), \dots, i_n(t))$ с компонентами, введенными в разделе 4, и изменениями в пространстве состояний S , вытекающими из применения функций $\varphi_k(i)$. Значения $p(i_1, \dots, i_n)$ стационарных вероятностей $r(t)$ связаны системой уравнений равновесия, имеющей вид

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

В (8) для состояния $(i_1, \dots, i_n) \in S$ значение i определяет величину занятого ресурса $i = i_1 b_1 + \dots + i_n b_n$, а $I(\cdot)$ — индикаторная функция события. Найденные значения $P(i_1, \dots, i_n)$ необходимо нормировать.

Характеристики качества обслуживания заявок определяются по аналогии с (2). Приведем несколько определений:

$$(9) \quad \begin{aligned} \pi_{t,k} &= \sum_{(i_1, \dots, i_n) \in S} p(i_1, \dots, i_n) \varphi_k(i); \\ m_k &= \sum_{(i_1, \dots, i_n) \in S} p(i_1, \dots, i_n) i_k b_k; \quad y_k = m_k / b_k; \\ \pi_{c,k} &= \frac{\sum_{(i_1, \dots, i_n) \in S} p(i_1, \dots, i_n) (s_k - i_k) \varphi_k(i)}{\sum_{(i_1, \dots, i_n) \in S} p(i_1, \dots, i_n) (s_k - i_k)}; \quad \pi_{\ell,k} = (a_k - y_k) / a_k. \end{aligned}$$

Для оценки объема ресурса и выбора параметров его резервирования необходимо решить систему уравнений (8). Сделать это можно итерационным методом Гаусса–Зейделя. Для оценки характеристик можно также использовать и достаточно обоснованные приближенные методы.

Результаты расчетов показывают, что при совместной передаче неоднородного трафика в единой транспортной среде заявки с малыми требованиями к ресурсу имеют существенно меньшие потери и ограничивают тем самым допуск к ресурсу для других заявок. Наличие механизма резервирования дает возможность выравнивать потери заявок при обслуживании неоднородного трафика, а также создать условия для использования определенного объема ресурса только заявками выделенной группы потоков. Покажем это. Для решения первой задачи достаточно выбрать значения функций фильтрации из соотношений (предполагается, что b_k занумерованы в порядке возрастания значений):

$$(10) \quad \varphi_k(i) = 1, \quad i = v - b_n + 1, v - b_n + 2, \dots, v; \quad k = 1, \dots, n.$$

В этой ситуации доли времени занятости ресурса для всех потоков будут одинаковыми, поскольку выполняется соотношение

$$(11) \quad \pi_{t,k} = \sum_{(i_1, \dots, i_n) \in S} p(i_1, \dots, i_n) \varphi_k(i) = \sum_{i=v-b_n+1}^v p(i), \quad k = 1, \dots, n.$$

Допустим, ставится задача предоставить в эксклюзивное пользование z виртуальных каналов потокам заявок с номерами $2, 3, \dots, n$. В этой ситуации функцию фильтрации заявок выберем из соотношений

$$(12) \quad \begin{aligned} \varphi_k(i) &= 1, \quad i = v - b_1 - z + 1, v - b_1 - z + 2, \dots, v; \quad k = 1; \\ \varphi_k(i) &= 1, \quad i = v - b_k + 1, v - b_k + 2, \dots, v; \quad k = 2, 3, \dots, n. \end{aligned}$$

В результате такого выбора функций блокировки заявки потоков с номерами $k \geq 2$ используют все доступные каналы узла. Причем z каналов из v имеющихся могут использоваться только этими потоками. Заявки 1-го потока не могут занять более $v - z$ каналов. Увеличивая значение z , можно

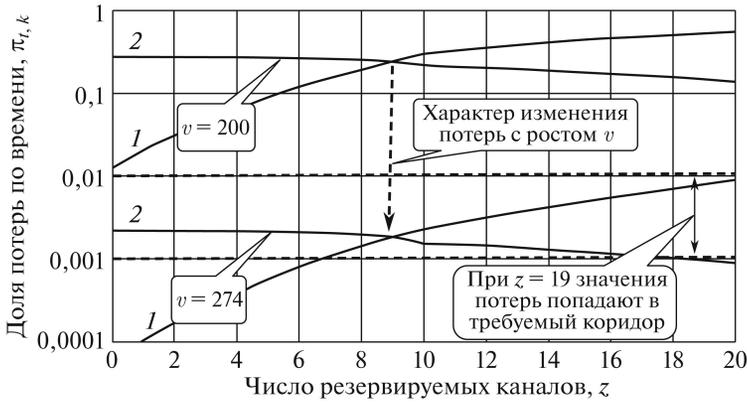


Рис. 6. Оценка величин ресурса и уровня резервирования, обеспечивающих заданные характеристики обслуживания заявок мультисервисного узла доступа.

уменьшить потери потоков с номерами $k \geq 2$ до требуемых величин. Правда, этот результат достигается за счет ухудшения качества обслуживания заявок 1-го потока. По этой причине наряду с использованием резервирования для выравнивания потерь заявок необходимо также увеличивать и общий объем ресурса с тем, чтобы качество обслуживания всех потоков было в пределах нормы.

Рассмотрим такую постановку задачи. Предположим, что требуется найти значения v и z такие, чтобы потери «легкого» трафика (1-й поток) были менее π_1 , а потери «тяжелого» трафика (потоки с номерами $k \geq 2$) — менее π_2 . Для решения задачи поступим следующим образом. После выравнивания потерь только «тяжелых» заявок с использованием правила (10) для них определяется число z резервируемых каналов, обеспечивающее требуемое соотношение потерь для всех потоков трафика. Если значение z с такими свойствами не находится, то v увеличивается на единицу и расчеты повторяются. Поскольку можно ожидать, что с ростом v при фиксированном z потери уменьшаются, а с увеличением z при фиксированном v потери «легкого» трафика увеличиваются, «тяжелого» — падают, то всегда можно найти требуемое решение. Иллюстрацией этого подхода служат данные, приведенные на рис. 6, где показаны зависимости потерь заявок $\pi_{t,k}$ от увеличения v и z .

Характеристики найдены для следующих значений входных параметров: $v = 200$; $n = 3$; $b_1 = 1$; $\mu_1 = 1$; $b_2 = 5$; $\mu_2 = 1$; $b_3 = 10$; $\mu_3 = 1$. Величины v , b_k и z выражены в виртуальных каналах. Примем, что $a_k = v\rho/nb_k$, $s_k = \lfloor a_k \rfloor + 10$, $k = 1, 2, 3$. Здесь ρ — потенциальная нагрузка одного в.к., $\zeta_k = \frac{a_k}{s_k - a_k}$. Требуемые уровни потерь $\pi_1 = 0,01$; $\pi_2 = 0,001$. Цифрой 1 обозначена доля потерь по времени для 1-го потока, цифрой 2 — для 2-го и 3-го потоков (значения потерь равны в соответствии с (10)). Начальное значение объема ресурса $v = 200$. Последовательное увеличение v , использование процедуры резервирования и проверка условия попадания значений потерь в требуемый коридор $\pi_{t,1} \leq 0,01$; $\pi_{t,2} = \pi_{t,3} \leq 0,001$ приводят к следую-

щим значениям: $v = 374$, $z = 19$. При данном выборе объема ресурса и параметров резервирования выполняются соотношения: $\pi_{t,1} = 0,007929 \leq 0,01$ и $\pi_{t,2} = \pi_{t,3} = 0,000969 \leq 0,001$.

В заключение отметим следующее. Если функция внутренней блокировки равна единице, то осуществляется процедура «жесткого» контроля за допуском заявок. Она дает возможность создать условия для их дифференцированного обслуживания. Однако это происходит за счет некоторого уменьшения коэффициента использования канала. Отрицательные последствия этого эффекта можно ослабить, если выполнить допуск заявок с вероятностью, меньшей единицы. В этом случае можно увеличить коэффициент использования канала за счет некоторого изменения условий дифференцированного обслуживания заявок. Понятно, что в такой постановке выбор значений функции внутренней блокировки зависит от экономических условий обслуживания заявок. Исследование этой проблемы в данной работе не рассматривается.

Теперь обратимся к решению второй из сформулированных в начале раздела задач. Информационные потоки, обладающие свойством эластичности², можно передавать с большей скоростью, используя для этих целей ресурс, оставшийся свободным от обслуживания трафика реального времени. Вернемся к математическому описанию модели мультисервисного узла доступа, изложенному в разделе 3, и внесем в нее изменения, относящиеся к ускоренной передаче данных.

Предположим, что свойством эластичности обладает дополнительный $(n + 1)$ -й поток и его параметры не зависят от номера потока. Заявки на передачу данных (файлов) поступают от конечной группы абонентов. Обозначим через s их число. Заявки от одного абонента поступают через случайное время, имеющее экспоненциальное распределение с параметром ζ . Для простоты предположим, что для обслуживания заявки требуется как минимум один виртуальный канал. Время обслуживания заявки одним каналом имеет экспоненциальное распределение с параметром μ . Таким образом, объем файла имеет экспоненциальное распределение со средним значением $F = u/\mu$, выраженным в битах.

Построим процедуру динамического распределения ресурса при организации пересылки файлов. Допустим, что на обслуживании находятся d заявок на передачу данных и для этого используются $(v - i)$ виртуальных каналов, оставшихся свободными от обслуживания трафика реального времени. Пусть $g = \lfloor \frac{v-i}{d} \rfloor$ — целая часть от деления $(v - i)$ на d . Разделим d заявок на две группы $d = d_1 + d_2$, где $d_1 = v - i - g d$, а $d_2 = (g + 1) d - (v - i)$. Для обслуживания заявок из групп d_1 и d_2 используются соответственно $(g + 1)$ и g виртуальных каналов. При выбранном распределении ресурса все $(v - i)$ свободных каналов занимают на обслуживание d принятых заявок на передачу данных, а время до освобождения первой из d заявок имеет экспоненциальное распределение с параметром $(v - i)\mu$. Понятно, что в рассматриваемых условиях скорость передачи каждого файла возрастает с появлением свободных

² К ним относятся передача файлов и других подобных им данных, допускающая небольшую задержку без потери качества обслуживания.

каналов и, наоборот, уменьшается при их сокращении, не становясь при этом меньше скорости, обеспечиваемой одним виртуальным каналом. В остальном схема функционирования узла не меняется. Сохраним для параметров и характеристик обобщенной модели те же обозначения, что были использованы в разделе 3.

Пусть $i_k(t)$ — число заявок k -го потока на передачу трафика реального времени, обслуживаемых в момент времени t , а $d(t)$ — число заявок на передачу файлов, обслуживаемых в момент времени t . Динамика изменения состояний модели описывается марковским процессом $r(t) = (i_1(t), \dots, i_n(t), d(t))$, определенным на пространстве состояний S и состоящем из векторов (i_1, \dots, i_n, d) с компонентами

$$\begin{aligned} i_1 &= 0, 1, \dots, \min \left\{ s_1, \left\lfloor \frac{v}{b_1} \right\rfloor \right\}; \\ i_n &= 0, 1, \dots, \min \left\{ s_n, \left\lfloor \frac{v - i_1 b_1 - \dots - i_{n-1} b_{n-1}}{b_n} \right\rfloor \right\}; \dots; \\ d &= 0, 1, \dots, v - i_1 b_1 - \dots - i_n b_n. \end{aligned}$$

Значения $p(i_1, \dots, i_n, d)$ стационарных вероятностей $r(t)$ связаны системой уравнений равновесия, имеющей вид

$$\begin{aligned} (13) \quad P(i_1, \dots, i_n, d) &\sum_{k=1}^n \left((s_k - i_k) \zeta_k I(i + d + b_k \leq v) + i_k \mu_k I(i_k > 0) \right) + \\ &+ (s - d) \zeta I(i + d + 1 \leq v) + (v - i) \mu I(d > 0) = \\ &= \sum_{k=1}^n P(i_1, \dots, i_k - 1, \dots, i_n, d) (s_k - i_k + 1) \zeta_k I(i_k > 0) + \\ &+ \sum_{k=1}^n P(i_1, \dots, i_k + 1, \dots, i_n, d) (i_k + 1) \mu_k I(i_k + 1 \leq s_k, i + d + b_k \leq v) + \\ &+ P(i_1, \dots, i_n, d - 1) (s - d + 1) \zeta I(d > 0) + \\ &+ P(i_1, \dots, i_n, d + 1) (v - i) \mu I(d + 1 \leq s, i + d + 1 \leq v), \\ &(i_1, \dots, i_n, d) \in S. \end{aligned}$$

В (13) для состояния $(i_1, \dots, i_n, d) \in S$ значение i определяет величину ресурса, занятого трафиком сервисов реального времени, $i = i_1 b_1 + \dots + i_n b_n$, а $I(\cdot)$ — индикаторная функция события. Найденные значения $P(i_1, \dots, i_n)$ необходимо нормировать.

Характеристики качества обслуживания заявок на передачу трафика реального времени определяются по аналогии с (2). Показатели обслуживания эластичного трафика оценим долей времени недоступности ресурса передачи π_t ; долей потерянных заявок π_c ; средним числом обслуживаемых заявок y ;

средним временем доставки файла t_d ; средним числом виртуальных каналов, используемых для передачи одного файла b_f :

$$(14) \quad \begin{aligned} \pi_t &= \sum_{(i_1, \dots, i_n, d) \in S | i+d+1 > v} p(i_1, \dots, i_n, d); \\ \pi_c &= \frac{\sum_{(i_1, \dots, i_n, d) \in S | i+d+1 > v} p(i_1, \dots, i_n, d)(s-d)}{\sum_{(i_1, \dots, i_n, d) \in S} p(i_1, \dots, i_n, d)(s-d)}; \\ y &= \sum_{(i_1, \dots, i_n, d) \in S} p(i_1, \dots, i_n, d)d; \quad t_d = \frac{y}{(s-y)\zeta(1-\pi_c)}; \\ b_f &= \frac{1}{y} \sum_{(i_1, \dots, i_n, d) \in S | d > 0} p(i_1, \dots, i_n, d)(v-i) = \frac{1}{t_d \mu}. \end{aligned}$$

Для определения объема необходимого ресурса и оценки введенных показателей эффективности использования каналов при обслуживании эластичного трафика необходимо решить систему уравнений (13). Сделать это можно итерационным методом Гаусса–Зейделя. Для оценки характеристик можно также применять приближенные алгоритмы. Построим соответствующую процедуру. Воспользуемся тем, что требуемый объем ресурса соответствует области малых потерь заявок. В этой ситуации совместное обслуживание трафика реального времени и эластичного трафика данных обладает свойствами, упрощающими оценку характеристик. Рассмотрим момент поступления заявки на передачу файла. Обозначим через i число каналов, занятых в этот момент на обслуживание трафика реального времени. Оставшиеся $v-i$ каналов используются для передачи файлов. Применим технику декомпозиции и оценим по отдельности процесс обслуживания трафика реального времени и данных. Пусть $p(i)$ — вероятность занятости i каналов на передачу трафика реального времени в условиях, когда трафик данных не поступает, $i = 0, 1, \dots, v$. Величины $p(i)$ находятся из рекурсивных соотношений (6). Предположим теперь, что в модели рассматривается обслуживание только заявок на передачу файлов. Обозначим через $y(v-i)$ и $t_d(v-i)$ соответственно среднее число заявок на передачу файлов и среднее время передачи файла при числе виртуальных каналов, равном $(v-i)$. Несложно показать, что значения введенных характеристик рассчитываются из выражений:

$$(15) \quad y(v-i) = \sum_{d=0}^{v-i} p_d(d)d; \quad t_d(v-i) = \frac{y(v-i)}{\sum_{d=0}^{v-i-1} p_d(d)(s-d)\zeta},$$

где вероятности $p_d(d)$ находятся из рекурсивных соотношений

$$p_d(d+1)(v-i)\mu = p_d(d)(s-d)\zeta, \quad d = 0, 1, \dots, v-i-1$$

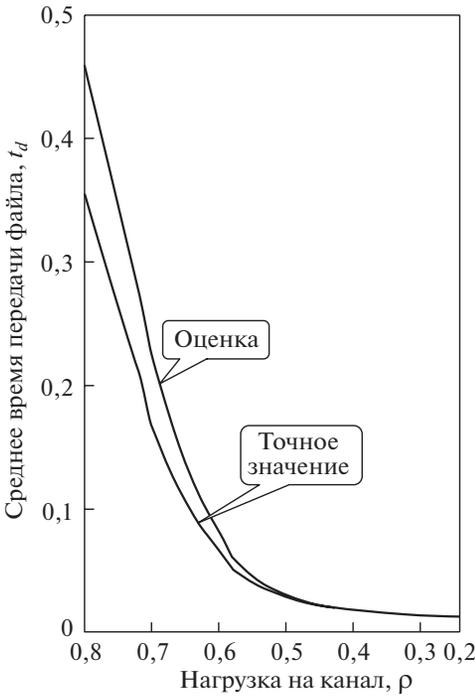


Рис. 7. Точная и приближенная оценки t_d .

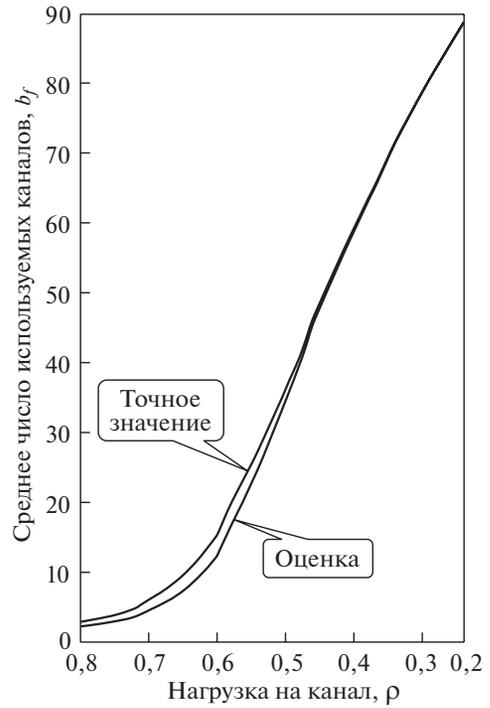


Рис. 8. Точная и приближенная оценки b_f .

с последующей их нормировкой. Обозначим получаемые оценки характеристик передачи эластичного трафика теми же символами, что применялись в исследуемой модели. Для их расчета используются соотношения

$$(16) \quad y = \sum_{i=0}^{v-1} p(i)y(v-i), \quad t_d = \sum_{i=0}^{v-1} p(i)t_d(v-i), \quad b_f = \frac{1}{t_d\mu}.$$

Оценим погрешность приближенного алгоритма расчета характеристик передачи эластичного трафика. Рассмотрим процесс совместного обслуживания трафика реального времени и эластичных данных для модели узла доступа со следующими фиксированными значениями входных параметров: $v = 100$; $n = 2$; $b_1 = 5$; $\mu_1 = 1$; $b_2 = 10$; $\mu_2 = 1$. Величины v , b_1 и b_2 выражены в виртуальных каналах. Примем, что $a_1 = v\rho/3b_1$, $a_2 = v\rho/3b_2$, $s_k = \lfloor a_k \rfloor + 10$, $k = 1, 2$. Здесь ρ — потенциальная нагрузка одного в.к., а $\zeta_k = \frac{a_k}{s_k - a_k}$. Параметры эластичного трафика $a = v\rho/3$, $s = \lfloor a \rfloor + 10$, $\zeta = \frac{a}{s-a}$, $\mu = 1$. Точные и приближенные значения t_d и b_f в зависимости от величины ρ приведены соответственно на рис. 7 и рис. 8. Точные величины характеристик находились в результате решения системы уравнений (13) итерационным методом Гаусса–Зейделя, а приближенные — с использованием выражений (16).

Анализ погрешности вычислений позволяет сделать вывод о хорошей точности расчетов особенно в области малых потерь, характерных для решения задач планирования требуемого объема ресурса. Сформулируем последова-

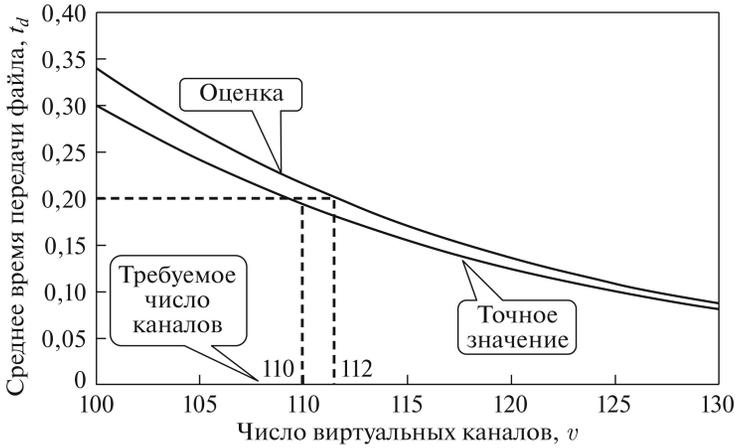


Рис. 9. Оценка требуемого объема ресурса при передаче эластичного трафика.

тельность действий, которую можно использовать для решения этой задачи. Предположим, что заданы параметры трафика реального времени n , s_k , ζ_k , b_k , μ_k и эластичных данных s , ζ , μ . Требуется определить число каналов v , обеспечивающее выполнение ограничений на долю потерянных заявок

$$(17) \quad \max(\pi_1, \dots, \pi_n, \pi_c) < \pi$$

и среднее время передачи файла

$$(18) \quad t_d \leq t.$$

Если вести передачу эластичного трафика на условиях трафика реального времени, то это приведет к увеличению доли потерянных заявок и даст возможность получить обоснованную оценку числа каналов v_1 в смысле выполнения неравенства (17). Для решения этой задачи используется либо рекурсия (6), либо алгоритм, рассмотренный в разделе 6. Далее при последовательном увеличении v_1 определяется величина ресурса v , обеспечивающая выполнение неравенства (18). Приведем пример реализации сформулированного алгоритма.

Рассмотрим потоки трафика со следующими значениями входных параметров: $n = 2$; $b_1 = 3$; $b_2 = 5$, $\mu_1 = 1$, $\mu_2 = 1$. Величины v , b_1 и b_2 выражены в виртуальных каналах. Примем, что $a_1 = 75\rho/3b_1$, $a_2 = 75\rho/3b_2$, $s_k = \lfloor a_k \rfloor + 10$, $k = 1, 2$. Здесь ρ — потенциальная нагрузка одного в.к., $\zeta_k = \frac{a_k}{s_k - a_k}$. Параметры эластичного трафика $a = 75\rho/3$, $s = \lfloor a \rfloor + 10$, $\zeta = \frac{a}{s - a}$, $\mu = 1$. Контрольные значения параметров $\pi = 0,03$, $t = 0,2$. После выполнения 1-го этапа получаем $v_1 = 99$. Результаты реализации 2-го этапа представлены на рис. 9, где показана зависимость t_d от v . Приведены данные приближенного вычисления t_d с использованием (16) и точные величины, полученные с помощью решения (13). Для $t = 0,2$ приближенный расчет дает ответ $v = 112$, а точный $v = 110$. Результаты вычислений отличаются хорошей точностью, приемлемой для практических приложений.

В заключение отметим следующее. Выбор экспоненциального распределения для описания времени обслуживания заявок упрощает моделирование мультисервисного узла доступа. Если положить равными нулю трафик реального времени или эластичный трафик, то расчетные формулы и алгоритмы перестанут зависеть от функции распределения времени обслуживания заявок [10, 21]. Это свойство дает основание ожидать слабую зависимость полученных результатов от выбора соответствующей функции и для модели совместного обслуживания заявок. Отмеченное свойство нуждается в численном анализе средствами имитационного моделирования и здесь не исследуется.

8. Заключение

Построена и исследована математическая модель распределения ресурса передачи информации мультисервисного узла доступа. В модели рассматривается произвольное число потоков мультимедийного трафика, которые различаются интенсивностью поступления заявок, величиной ресурса, используемого для обслуживания одной заявки, и временем занятия ресурса. Интервалы времени между поступлением заявок имеют экспоненциальное распределение с параметром, зависящим от числа заявок рассматриваемого потока, находящихся на обслуживании. Построен рекурсивный алгоритм оценки характеристик. Установлены соотношения между интегральными и потоковыми характеристиками качества обслуживания заявок, которые упрощают выбор метрики при решении задач планирования ресурса узла доступа. Построен эффективный алгоритм оценки объема ресурса, требуемого для обслуживания заданных потоков трафика с необходимым качеством. Эффективность расчетной процедуры достигается в результате организации рекурсии по объему ресурса и использовании нормированных значений вероятностей состояний. Рассмотрено решение задачи оценки необходимого объема ресурса для модели мультисервисного узла, допускающего использование механизмов резервирования ресурса и его динамического распределения при обслуживании эластичного трафика. Приведены численные примеры, иллюстрирующие особенности реализации построенных расчетных процедур. Построенную модель и методы ее расчета можно обобщить на случаи, когда поступление заявок носит групповой характер.

СПИСОК ЛИТЕРАТУРЫ

1. 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.
2. Kaufman J.S. Blocking in a shared resource environment // *IEEE Transact. Communicat.* 1981. V. 29. No. 1. P. 1474–1481.
3. Roberts J.W. A service system with heterogenous user requirements application to multi-service telecommunications systems / *Performance of Data Communication Systems and their Applications*. Pujolle G. (ed.). North Holland, 1981. P. 423–431.
4. Delbrouck L.E.N. On the Steady-State Distribution in a Service Facility Carrying Mixtures of Traffic with Different Peakedness Factor and Capacity Requirements // *IEEE Transactions on Communications*. 1983. V. COM-31. P. 1209–1211.

5. Broadband network traffic. Performance evaluation and design of broadband multi-service networks. Final report of action COST 242 / James Roberts (ed). Lecture notes in computer sciences. Springer, 1996.
6. *Ross K.W.* Multiservice loss models for broadband telecommunication networks. London: Springer, 1995.
7. *Iversen V.B.* Teletraffic Engineering and Network Planning. Technical University of Denmark, May. 2010.
8. *Степанов С.Н.* Теория телетрафика: концепции, модели, приложения. М.: Горячая линия – Телеком, 2015.
9. *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.
10. *Bonald T., Virtamo J.* A recursive formula for multirate systems with elastic traffic // *IEEE Communicat. Lett.* 2005. V. 9. No. 8. P. 753–755.
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. *Iversen V.B., Stepanov S.N.* The optimal dimensioning of multi-service links // *Proc. COST-285 Mid-term Symposium.* Munchen, September 8–10, 2005. Chapter 7 (pp. 151–178) in A. Nejat Ince & Ercan Topuz (editors): Modeling and simulation tools for emerging telecommunication networks. Springer, 2006.
13. *Berezner S.A., Krzesinski A.E.* An Efficient Stable Recursion to Compute Multiservice Blocking Probabilities // *Performance Evaluation.* 2001. V. 43. P. 151–164.
14. *Begishev V., Petrov V., Samuylov 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.
15. *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. Scie., Springer, Cham.* 2015. V. 9012. P. 473–483.
16. *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.
17. *Степанов С.Н.* Модель совместного обслуживания трафика сервисов реального времени и данных. I // *АиТ.* 2011. № 4. С. 121–132.
Stepanov S.N. Model of Joint Servicing of Real-Time Service Traffic and Data Traffic. I // *Autom. Remote Control.* 2011. V. 72. No. 4. P. 787–797.

18. *Степанов С.Н.* Модель совместного обслуживания трафика сервисов реального времени и данных. II // *АиТ.* 2011. № 5. С. 139–147.
Stepanov S.N. Model of Joint Servicing of Real-Time Service Traffic and Data Traffic. II // *Autom. Remote Control.* 2011. V. 72. No. 5. P. 1028–1035.
19. *Evans J., Filsfils C.* Deploying IP and MPLS QOS for Multiservice Networks. Theory and Practice / *The Morgan Kaufmann Series in Networking.* Elsevier Ltd, 2007.
20. *Kelly F.P.* Notes on effective bandwidths. In: F. Kelly, S. Zachary and I. Ziedinis (Eds.), *Stochastic Networks: Theory and Applications Telecommunications Networks.* Volume 4 of Royal Statistical Society Lecture Notes Series, Oxford. 1996. P. 141–168. Oxford University Press.
21. *Bonald T., Feuillet M.* Network Performance Analysis. United Kingdom, London: Wiley, 2011.
22. *Bonald T., Comte C.* The multi-source model for dimensioning data networks // *Computer Networks*, 2016 (10.1016/j.comnet.2016.03.019. hal-01314992).
23. *Баширин Г.П.* Лекции по математической теории телетрафика: Учеб. пособие. Изд. 3-е исправ. и доп. М.: Изд-во РУДН, 2009.

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

Поступила в редакцию 17.03.2000

После доработки 15.05.2020

Принята к публикации 09.07.2020