ПРИБЛИЖЕННЫЙ МЕТОД АНАЛИЗА ОДНОЛИНЕЙНЫХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С ПОСТОЯННЫМ ВРЕМЕНЕМ ОБРАБОТКИ ЗАЯВОК

А. Н. Соколов
Санкт-Петербургский государственный университет телекоммуникаций им. М. А. Бонч-Бруевича, 191186, Санкт-Петербург, Россия

УДК 621.394.343

Предложен метод оценки характеристик однолинейных систем массового обслуживания, в которых время обработки заявок может считаться постоянной величиной. Метод основан на дискретизации функции распределения длительности интервалов между поступлениями заявок. Предложен критерий выбора интервала дискретизации, учитывающий допустимую ошибку оценки исследуемых характеристик системы массового обслуживания.

Ключевые слова: задержка, аппроксимация, дискретизация, функция распределения, система массового обслуживания.

In clause, the evaluation method of the queuing system characteristics with determined service time is proposed. Method is based on the discretization of interarrival distribution function. Criterion for the choice of discretization interval is introduced. This criterion takes into account permissible error of the investigated values.

Key words: delay, approximation, discretization, distribution function, queuing system.

Введение. Ряд компонентов современных телекоммуникационных сетей можно исследовать с помощью модели, которая представляет собой однолинейную систему массового обслуживания (СМО) с постоянным временем обработки заявок. В классификации Кендалла [1] такая система обозначается $G/D/1$. Символ G указывает на произвольный характер функции распределения (ФР) длительности интервалов между моментами поступления заявок $A(t)$. Символ D используется в тех случаях, когда длительность обработки (в теории телетрафика используется также термин “обслуживание”) заявок может считаться постоянной величиной. Такая гипотеза приемлема для многих компонентов пакетных сетей. В этих сетях IP-пакет, представляющий собой заявку в терминах теории телетрафика, обрабатывается за время, которое можно считать неизменной величиной. Цифра 1 в третьей позиции классификации Кендалла указывает на тот факт, что количество обслуживающих устройств равно единице.

Анализ модели $G/D/1$ существенно упрощается, если функцию $A(t)$ можно представить как экспоненциальное распределение. Однако такое допущение нельзя считать приемлемым для большинства компонентов сети, построенной с использованием IP-технологии. В [2] для систем вида $G_s/D/1$ приведены выражения для расчета кумулянтов, позволяющих провести анализ вероятностно-временных характеристик, которые определяют качество обслуживания трафика. Нижний индекс s (первая буква в слове “stepped”) означает, что функция $A(t)$ является ступенчатой. Такой вид распределения $A(t)$ представляет интерес, когда для описания входящего потока заявок используются результаты измерений.

При исследовании некоторых СМО (например, проектируемых) невозможно провести измерения для получения функции $A(t)$, но можно ввести обоснованную гипотезу о ее характере. В таких случаях функция $A(t)$ описывается одним из известных законов распределения случайных величин [3]. В данной работе предлагается приближенный метод анализа СМО $G/D/1$ на основе результатов, полученных в [2]. В основу метода положена дискретизация функции $A(t)$ с некоторым периодом τ, что позволяет перейти от исследуемой системы к модели вида $G_s/D/1$.

Для описания ступенчатой функции $A(t)$ целесообразно использовать преобразование Лапласа — Стилтьеса $\alpha(s)$. Если в точке $i\tau$ определено приращение исследуемой функции X_i, то преобразование $\alpha(s)$ имеет вид [4]

$$\alpha(s) = \sum_{i=0}^{\infty} X_i \, \mathrm{e}^{-i\tau s}.$$

На практике верхний предел суммирования ограничен значением индекса i, для которого в процессе измерений было зафиксировано последнее приращение функции $A(t)$. Значение τ обычно очень мало. Это позволяет выбрать некоторое целое значение g, позволяющее считать время обслуживания заявок равным произведению $g\tau$. Тогда преобразование Лапласа – Стилтьеса ФР длительности обслуживания заявок $\beta(s)$ определяется следующим соотношением [4]:

$$\beta(s) = \mathrm{e}^{-g\tau s}. \tag{1}$$

Очевидно, что при замене распределения $A(t)$ ступенчатой функцией точность результатов исследования СМО будет зависеть от величины τ. С этой точки зрения рассматриваемая задача подобна выбору периода дискретизации аналогового сигнала. Для определения требований к величине τ целесообразно исследовать несколько моделей с функцией $A(t)$ различного вида.

1. Три аппроксимации функции $A(t)$ для экспоненциального распределения. В теории телетрафика нередко используется модель с пуассоновским входящим потоком заявок [1], для которого

$$A(t) = 1 - \mathrm{e}^{-\lambda t}. \tag{2}$$

Величину λ, называемую интенсивностью потока заявок, можно определить по результатам измерений длительности интервалов между моментами поступления заявок t_i. Если проведено N измерений, в результате которых получены величины t_i $(i = \overline{1,\ N})$, то среднее время между поступлениями заявок $A^{(1)}$ и величина λ рассчитываются следующим образом [1]:

$$A^{(1)} = \frac{1}{N}\sum_{i=1}^{N} t_i, \quad \lambda = \frac{1}{A^{(1)}}. \tag{3}$$

На рис. 1 показаны распределение $A(t)$ и две ступенчатые функции $A_1(t)$ и $A_2(t)$, позволяющие оценить верхнюю и нижнюю границы ряда параметров для систем с пуассоновским входящим потоком. В точке $k\tau$ приращения функций $A_1(t)$ и $A_2(t)$ определяются величинами P_k и Q_k соответственно. Заметим, что $P_0 = 0$.

Величины P_k и Q_k для ступенчатых функций $A_1(t)$ и $A_2(t)$ могут быть вычислены по формулам

$$P_k = (\mathrm{e}^{\lambda\tau} - 1)\,\mathrm{e}^{-k\lambda\tau}, \quad Q_k = (\mathrm{e}^{\lambda\tau} - 1)\,\mathrm{e}^{-(k+1)\lambda\tau}.$$

Преобразования этих двух ступенчатых функций $\alpha_1(s)$ и $\alpha_2(s)$ связаны простым соотношением

$$\alpha_2(s) = \alpha_1(s)\,\mathrm{e}^{\tau s}.$$

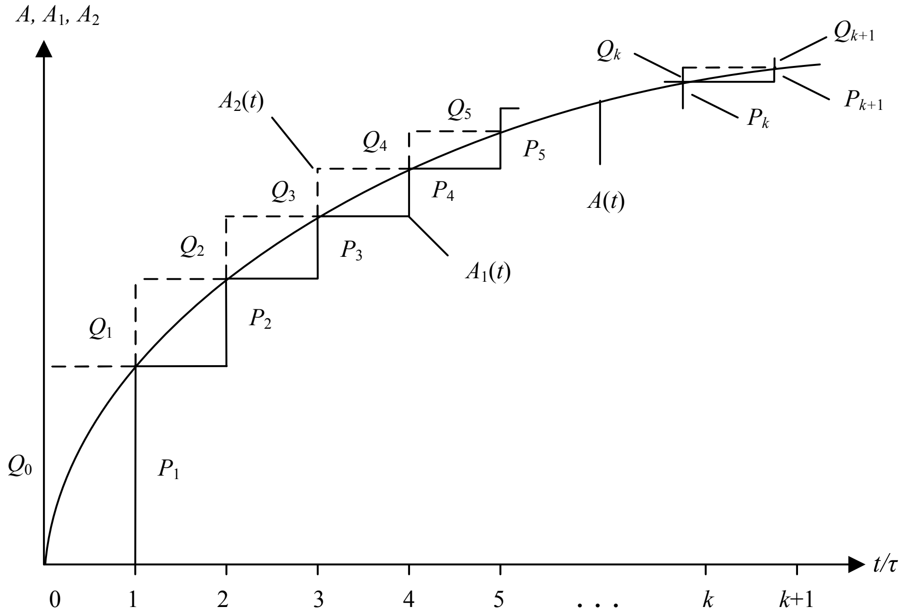

Рис. 1. Две ступенчатые функции $A_1(t),\ A_2(t)$, заменяющие распределение $A(t)$

Целесообразно ввести также функцию $A_3(t)$, которая получается из ступенчатых функций $A_1(t)$ или $A_2(t)$ смещением их по оси абсцисс на $0{,}5\tau$ влево или вправо соответственно. Функции, подобные $A_3(t)$, получаются при выборе аппроксимирующей зависимости методом наименьших квадратов [5]. Преобразование Лапласа – Стилтьеса третьей функции $\alpha_3(s)$ удобно выразить через формулы для изображений $\alpha_1(s)$ и $\alpha_2(s)$:

$$\alpha_3(s) = \alpha_1(s)\,\mathrm{e}^{0{,}5\tau s} = \alpha_2(s)\,\mathrm{e}^{-0{,}5\tau s}.$$

После подстановки выражений для расчета приращений P_k и Q_k функции $\alpha_1(s)$, $\alpha_2(s)$ и $\alpha_3(s)$ определяются как суммы членов геометрической прогрессии [5]:

$$\alpha_1(s) = \frac{\left(1-\mathrm{e}^{-\lambda\tau}\right)\mathrm{e}^{-s\tau}}{1-\mathrm{e}^{-(\lambda+s)\tau}}, \quad \alpha_2(s) = \frac{1-\mathrm{e}^{-\lambda\tau}}{1-\mathrm{e}^{-(\lambda+s)\tau}}, \quad \alpha_3(s) = \frac{\left(1-\mathrm{e}^{-\lambda\tau}\right)\mathrm{e}^{-0{,}5\tau s}}{1-\mathrm{e}^{-(\lambda+s)\tau}}. \tag{4}$$

По правилу нахождения моментов случайной величины r-го порядка из формул для изображений ФР можно определить средние значения длительности интервалов между поступлениями заявок $A_1^{(1)}$, $A_2^{(1)}$ и $A_3^{(1)}$:

$$A_1^{(1)} = \frac{\tau}{1-\mathrm{e}^{-\lambda\tau}}, \quad A_2^{(1)} = \frac{\tau\mathrm{e}^{-\lambda\tau}}{1-\mathrm{e}^{-\lambda\tau}}, \quad A_3^{(1)} = \frac{\tau\left(1+\mathrm{e}^{-\lambda\tau}\right)}{2\left(1-\mathrm{e}^{-\lambda\tau}\right)}. \tag{5}$$

Это же правило позволяет определить второй момент, необходимый для расчета дисперсии. Как и следовало ожидать, для всех трех ступенчатых функций значения дисперсии (σ_1^2, σ_2^2 и σ_3^2) идентичны, поэтому нижний индекс можно опустить, используя стандартное обозначение σ^2:

$$\sigma^2 = \frac{\tau^2\mathrm{e}^{-\lambda\tau}}{\left(1-\mathrm{e}^{-\lambda\tau}\right)^2} = \frac{\tau^2}{4\mathrm{sh}\,(0{,}5\lambda\tau)}. \tag{6}$$

Для оценки квантилей некоторых ФР используется коэффициент асимметрии Sk [6]. Следуя правилам расчета этого показателя [3] для трех исследуемых распределений, можно убедиться, что он один и тот же:

$$Sk = \mathrm{e}^{0{,}5\lambda\tau} + \mathrm{e}^{-0{,}5\lambda\tau}.$$

Выражения (5), (6) позволяют вычислить коэффициенты вариации исследуемой случайной величины C_1, C_2 и C_3:

$$C_1 = \mathrm{e}^{-0{,}5\lambda\tau}, \quad C_2 = \mathrm{e}^{0{,}5\lambda\tau}, \quad C_3 = \frac{2}{\mathrm{e}^{0{,}5\lambda\tau} + \mathrm{e}^{-0{,}5\lambda\tau}}.$$

При $\tau \to 0$ все три коэффициента вариации стремятся к единице. Можно показать, что при $\tau \to 0$ $A_1^{(1)}$, $A_2^{(1)}$ и $A_3^{(1)}$ стремятся к λ^{-1}, σ^2 — к λ^{-2}, C_1, C_2 и C_3 — к единице, а Sk — к двум. Значения λ^{-1}, λ^{-2}, 1 и 2 свойственны соответственно математическому ожиданию, дисперсии, коэффициентам вариации и асимметрии случайной величины, которая подчиняется экспоненциальному закону распределения.

Для значений τ, отличных от нуля, существует относительная ошибка в оценке моментов случайной величины, обусловленная аппроксимацией распределения $A(t)$ ступенчатой функцией. Для функций $A_1(t)$, $A_2(t)$ и $A_3(t)$ эти ошибки обозначаются следующим образом: $\delta_1(j)$, $\delta_2(j)$ и $\delta_3(j)$. Переменная j идентифицирует ту характеристику случайной величины, для которой рассчитывается относительная ошибка.

Нетрудно найти ошибки $\delta_1(C)$, $\delta_2(C)$ и $\delta_3(C)$ для коэффициентов вариации, так как точное значение этой характеристики для рассматриваемого распределения $A(t)$ равно единице. Кроме того, несложно показать, что минимальной ошибкой является величина $\delta_3(C)$. Формулы для расчета ошибок $\delta_1(C)$, $\delta_2(C)$ и $\delta_3(C)$ имеют следующий вид:

$$\delta_1(C) = 1 - \mathrm{e}^{-0{,}5\lambda\tau}, \quad \delta_2(C) = 1 - \mathrm{e}^{0{,}5\lambda\tau}, \quad \delta_3(C) = 1 - \frac{2}{\mathrm{e}^{0{,}5\lambda\tau} + \mathrm{e}^{-0{,}5\lambda\tau}}.$$

Для достаточно малых значений τ величина относительной ошибки оценивается слагаемым $0{,}5\lambda\tau$. Для доказательства этого утверждения функции $\delta_1(C)$, $\delta_2(C)$ и $\delta_3(C)$ следует разложить в ряд Маклорена [5]. Если задан уровень допустимой относительной ошибки для оценки коэффициента вариации $\delta(C)$ и известен параметр λ, то для выбора величины τ следует использовать неравенство

$$\tau \leq \frac{2\delta(C)}{\lambda}. \tag{7}$$

Определение требований к величине τ с учетом допустимой относительной ошибки в оценке средних значений исследуемой случайной величины связано с громоздкими преобразованиями, после выполнения которых несложно получить неравенство вида (7) с одним отличием. Вместо сомножителя $\delta(C)$ в нем содержится величина допустимой относительной ошибки для оценки среднего значения длительности интервала между поступлениями соседних заявок $\delta(A)$.

Коэффициент вариации представляет собой нормированную дисперсию. Это означает, что при использовании аппроксимаций вида (4) для расчета двух основных характеристик исследуемой случайной величины (среднего значения и дисперсии) период τ выбирается из следующего неравенства:

$$\tau \leq \frac{2\delta}{\lambda}. \tag{8}$$

В данном случае множитель δ целесообразно рассматривать как допустимый уровень относительной ошибки при оценке среднего значения и дисперсии случайной величины, распределение которой подчиняется экспоненциальному закону. Несложно убедиться, что условие (8) позволяет оценивать с заданной точностью и коэффициент асимметрии.

2. Аппроксимация функции $A(t)$ для закона равномерной плотности. Исследование ошибок для функции $A(t)$, заданной на ограниченном интервале времени, можно провести на примере равномерного закона распределения длительности интервалов между моментами поступления заявок [3]. На рис. 2 показано распределение $A(t)$ на нормированном к τ интервале времени от 0 до $n-1$. В точках $k\tau$ ступенчатая функция $A_3(t)$ имеет одинаковые приращения, равные n^{-1}.

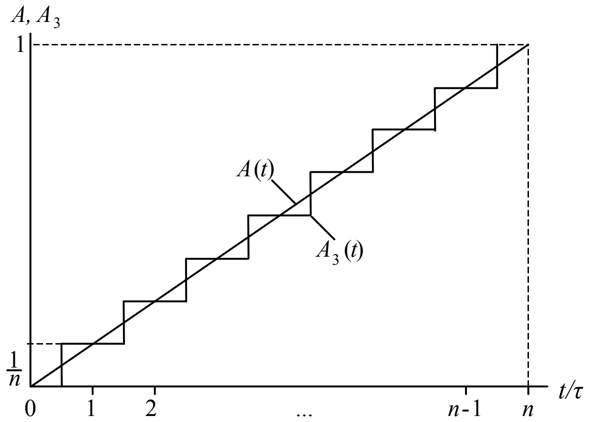

Рис. 2. Равномерное распределение и его аппроксимация

Для функции $A(t)$ среднее значение длительности интервалов между моментами поступления заявок $A^{(1)}$, дисперсия σ^2, коэффициент вариации C случайной величины определяются по известным формулам [3]

$$A^{(1)} = \frac{n\tau}{2}, \quad \sigma^2 = \frac{n^2\tau^2}{12}, \quad C = \frac{1}{\sqrt{3}}.$$

В качестве ступенчатой функции $A(t)$ целесообразно использовать распределение, которое соответствует третьему виду функции, введенному в п. 1. Для ступенчатой функции вида $A_3(t)$ исследуемые характеристики (с нижним индексом 3) вычисляются по формулам [3]

$$A_3^{(1)} = \frac{(n-1)\tau}{2}, \quad \sigma_3^{(2)} = \frac{(n^2-1)\tau^2}{12}, \quad C_3 = \frac{\sqrt{n^2-1}}{\sqrt{3}(n-1)}.$$

Коэффициенты асимметрии для обоих распределений равны нулю, поэтому соответствующие ошибки далее не рассматриваются. При оценке среднего значения исследуемой случайной величины $\delta(A)$ относительная ошибка определяется следующим образом:

$$\delta(A) = \frac{A^{(1)} - A_3^{(1)}}{A^{(1)}} = 0{,}5\lambda\tau.$$

Это означает, что величина τ должна выбираться из неравенства (8), полученного для другого закона распределения исследуемой случайной величины. Можно показать, что именно условие (8) должно выполняться при выборе τ для равномерного закона распределения $A(t)$.

3. Аппроксимация функции $A(t)$ для гиперэкспоненциального распределения. Третий пример связан с распределениями, для которых коэффициент вариации длительности интервалов между моментами поступления заявок превышает единицу. На профессиональном сленге они называются распределениями с “тяжелыми хвостами” [7]. Одним из удачных (с точки зрения простоты последующего анализа) примеров такого распределения следует считать гиперэкспоненциальное. Для гиперэкспоненциального распределения второго порядка [3] с параметром формы p преобразование Лапласа – Стилтьеса ФР длительности интервалов между поступлениями заявок $\alpha(s)$ можно представить в виде

$$\alpha(s) = p\frac{2p\lambda}{s+2p\lambda} + (1-p)\frac{2(1-p)\lambda}{s+2(1-p)\lambda}. \tag{9}$$

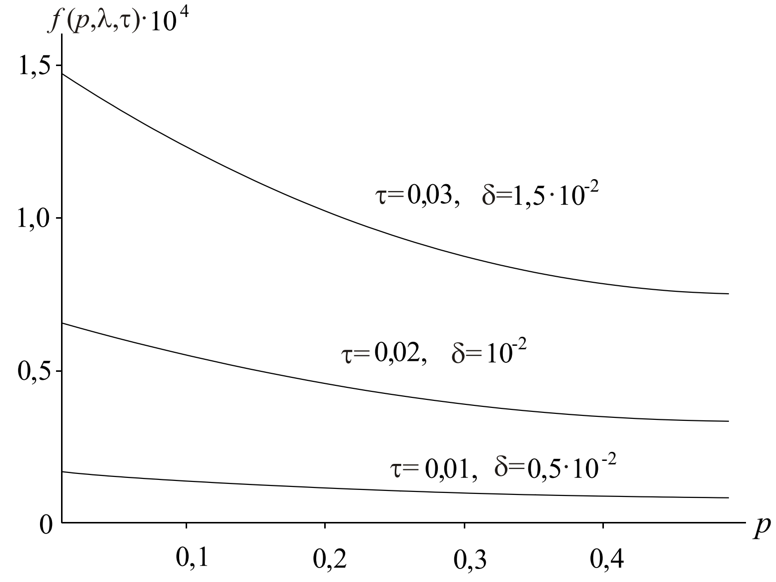

Рис. 3. Зависимость относительной ошибки $f(p, \lambda, \tau)$ от параметра формы p

Математическое ожидание времени между моментами поступления заявок $A^{(1)}$ связано с величиной интенсивности потока заявок λ соотношением (3). При замене гиперэкспоненциального распределения дискретным с отсчетами, взятыми с периодом τ, целесообразно использовать функцию вида $A_3(t)$. Для этой функции среднее значение времени между моментами поступления заявок $A_3^{(1)}$ может быть получено в виде

$$A_3^{(1)} = \frac{\tau p\left(1 + \mathrm{e}^{-2p\lambda\tau}\right)}{2\left(1 - \mathrm{e}^{-2p\lambda\tau}\right)} + \frac{\tau(1-p)\left(1 + \mathrm{e}^{-2(1-p)\lambda\tau}\right)}{2\left(1 - \mathrm{e}^{-2(1-p)\lambda\tau}\right)}.$$

Относительную ошибку оценки $A^{(1)}$ целесообразно исследовать как функцию $f(p, \lambda, \tau)$, которая зависит от трех параметров p, λ и τ:

$$f(p,\lambda,\tau) = 1 - \lambda\tau \frac{p\left(1 + \mathrm{e}^{-2p\lambda\tau}\right)\left(1 - \mathrm{e}^{-2(1-p)\lambda\tau}\right) + (1-p)\left(1 + \mathrm{e}^{-2(1-p)\lambda\tau}\right)\left(1 - \mathrm{e}^{-2p\lambda\tau}\right)}{2\left(1 - \mathrm{e}^{-2p\lambda\tau}\right)\left(1 - \mathrm{e}^{-2(1-p)\lambda\tau}\right)}.$$

На рис. 3 приведены графики функций $f(p, \lambda, \tau)$, построенные для двух значений τ в диапазоне значений параметра формы от 0,01 до 0,49 при $\lambda = 1$. Выбор такого диапазона обусловлен тем, что распределение вида (9) определено при условии $0 < p < 0{,}5$.

Для каждой кривой указаны также значения относительной ошибки, вычисленные на основе соотношения (8): $\delta = 0{,}5\lambda\tau$. Очевидно, что $f(p, \lambda, \tau) << \delta$. Данный факт объясняется характером изменений функции $A(t)$. Распределения с "тяжелыми хвостами" с ростом t приближаются к единице не столь быстро, как, например, экспоненциальное. Поэтому отсчеты, взятые с периодом τ, вычисленным по формуле (8), более точно представляют ФР. Следовательно, соотношение (8) приемлемо для распределений с "тяжелыми хвостами". Можно показать, что данный вывод справедлив также для оценок других параметров ФР длительности интервалов между моментами поступления заявок. Соответствующие формулы очень громоздки и в данной работе не приводятся.

Исследование СМО посредством дискретизации ФР длительности интервалов между моментами поступления заявок, для которой значение коэффициента вариации превышает (иногда весьма существенно) единицу, позволяет анализировать процессы функционирования мультисервисных сетей. Именно для сетей такого рода характерны распределения с “тяжелыми хвостами” [7].

4. Оценка ошибок при исследовании параметров времени задержки заявок. Сформулированное условие для выбора величины τ приемлемо для получения численных характеристик функции $A(t)$. Необходимо определить подобное условие для моментов времени задержки. С практической точки зрения представляют интерес среднее время задержки $S^{(1)}$ и дисперсия этой случайной величины σ_S^2. Вместо дисперсии можно использовать также коэффициент вариации C_S.

Если функция $A(t)$ описывается законом (2), то рассматриваемой моделью служит система вида $M/D/1$. Для этой системы известны выражения, позволяющие рассчитать $S^{(1)}$, σ_S^2 и C_S (см., например, [1]). Для компактной записи этих выражений следует ввести параметр нагрузки ρ, который для исследуемой модели $M/D/1$ с учетом соотношения (1) определяется как отношение интенсивности входящего потока заявок λ к интенсивности их обработки μ:

$$\rho = \frac{\lambda}{\mu}.$$

Тогда формулы для вычисления характеристик $S^{(1)}$, σ_S^2 и C_S можно представить в следующем виде:

$$S^{(1)} = \frac{2-\rho}{2\,(1-\rho)} g\tau, \qquad \sigma_S^2 = \frac{\rho(4-\rho)}{12(1-\rho)^2}(g\tau)^2, \quad C_S = \frac{\sqrt{\rho(4-\rho)}}{\sqrt{3}(2-\rho)}. \tag{10}$$

Ступенчатые функции, используемые для аппроксимации распределения $A(t)$ оценки $S^{(1)}$, σ_S^2 и C_S, приведены в [2]. Эти оценки вместе с результатами, полученными в расчетах по формулам (10), позволяют определить максимальное значение τ, для которого относительная ошибка в оценке исследуемых характеристик времени задержки заявок не превышает установленного порога δ.

Проведенные вычисления показали, что величина τ зависит от нагрузки ρ. Для задач проектирования систем массового обслуживания представляет интерес диапазон значений нагрузки от 0,1 до 0,7. Именно для этих значений ρ целесообразно оценить величину относительной ошибки, возникающей при расчетах параметров времени задержки заявок. В число исследуемых параметров этой случайной величины целесообразно включить среднее значение и коэффициент вариации. Эти два параметра часто используются для оценки показателей качества обслуживания в системах массового обслуживания.

В формулы (10) входит величина нагрузки ρ, а неравенство (8) позволяет определить интенсивность потока заявок λ. Для дальнейшего анализа целесообразно предположить, что время обслуживания равно единице. Тогда значения ρ и λ равны.

Результаты вычисления ошибок, возникающих при оценке двух исследуемых параметров времени задержки заявок, приведены в таблице. Величина τ определялась из соотношения (8) при условии $\delta = 2{,}5\,\%$.

Данные, приведенные в таблице, свидетельствуют о том, что в выбранном диапазоне значений нагрузки ошибки оценки двух исследуемых параметров времени задержки заявок не превышают порога $\delta = 2{,}5\,\%$, который использован для дискретизации функции $A(t)$. Это означает, что условие (8) можно считать достаточным для приближенного анализа параметров времени задержки заявок в однолинейных СМО с постоянным временем обслуживания заявок.

Относительные ошибки δ в оценке параметров времени задержки заявок

Параметр времени задержки заявок	$\rho = 0{,}1$	$\rho = 0{,}3$	$\rho = 0{,}5$	$\rho = 0{,}7$
Среднее значение	0,010	0,002	0,001	0,011
Коэффициент вариации	0,009	0,006	0,003	0,015

Заключение. Замена непрерывного закона распределения длительности интервалов между моментами поступления заявок дискретной функцией позволяет исследовать основные параметры времени задержки в СМО при условии, что время обслуживания постоянно. Ошибка в расчетах исследуемых параметров времени задержки, возникающая вследствие замены закона распределения, зависит от периода дискретизации τ. Проведенные исследования показали, что для известного значения интенсивности входящего потока λ и заданного уровня относительной ошибки δ достаточно брать отсчеты распределения длительности интервалов между моментами поступления заявок с периодом $\tau \leq 2\delta\lambda^{-1}$.

Список литературы

1. Клейнрок Л. Теория массового обслуживания. М.: Машиностроение, 1979.

2. Соколов А. Н. Метод оценки задержки IP-пакетов в узле коммутации // Науч.-техн. ведомости СПбГПУ. 2009. № 4. С. 37–40.

3. Вадзинский Р. Н. Справочник по вероятностным распределениям. СПб.: Наука, 2001.

4. Диткин В. А. Интегральные преобразования и операционное исчисление / В. А. Диткин, А. П. Прудников. М.: Наука, 1974.

5. Корн Т. Справочник по математике для научных работников и инженеров / Т. Корн, Г. Корн. М.: Наука, 1984.

6. ITU-T. Network performance objectives for IP-based services. Recommendation Y.1541. Geneva: ITU, 2006.

7. Шелухин О. И. Моделирование информационных систем / О. И. Шелухин, А. М. Тенякшев, А. В. Осин. М.: Радиотехника, 2005.

Соколов Андрей Николаевич — асп. Санкт-Петербургского государственного университета телекоммуникаций им. М. А. Бонч-Бруевича;
тел. (905) 221-28-93; e-mail: a.n.sokolov@hotmail.com

Дата поступления — 31.05.10