УДК 621. 395

Ложковский А.Г.
Ложковський А.Г.
Loshkovskii A.G.

МЕТОД РАСЧЕТА ОДНОКАНАЛЬНЫХ СИСТЕМ С ОЧЕРЕДЬЮ ПРИ ЭКСПОНЕНЦИАЛЬНОЙ ДЛИТЕЛЬНОСТИ ОБСЛУЖИВАНИЯ

МЕТОД РОЗРАХУНКУ ОДНОКАНАЛЬНИХ СИСТЕМ З ЧЕРГОЮ ПРИ ЕКСПОНЕНТНІЙ ТРИВАЛОСТІ ОБСЛУГОВУВАННЯ

CALCULATION METHOD OF SINGLE-CHANNEL SYSTEM WITH QUEUE AT EXPONENTIAL HOLDING TIME

Аннотация. Рассмотрена модель одноканальной системы с очередью при обслуживании произвольного потока заявок с экспоненциальной длительностью обслуживания. Предложен метод расчета основных характеристик качества обслуживания в модели G/M/1.

Анотація. Розглянуто модель одноканальної системи з чергою при обслуговуванні довільного потоку заявок з експонентною тривалістю обслуговування. Запропоновано метод розрахунку основних характеристик якості обслуговування в моделі G/M/1.

Summary. The model of single-channel system with queue at service of arbitrary stream of the requests with exponential duration of service is considered. The calculation method of basic performances of quality of service in G/M/1 model is offered.

Повышающаяся сложность телекоммуникационных систем выдвигает на первый план проблему разработки адекватных математических моделей этих систем с целью получения достоверных оценок их характеристик, реализации задач их оптимизации относительно выбранного критерия качества обслуживания и разработки соответствующих алгоритмов управления ими.

В современных мультисервисных пакетных сетях связи входящие информационные потоки могут иметь постоянную (CBR), переменную (VBR) и смешанную битовую скорость, отчего математическая модель потока может быть самой разной – от простейшей пуассоновской модели до сложнейшей модели фрактальных процессов (самоподобный трафик). Закон распределения промежутка времени между заявками в этих потоках может быть произвольный и поэтому в обобщенной модели резонно исследовать общий (G – *general*) вид распределения случайной величины этого промежутка. Длина пакетов каждой из служб общей для них мультисервисной сети (следовательно, и длительность обслуживания) может быть различной – для одних служб постоянной, а для других – переменной. В таком случае также желательно исследовать общий вид распределения случайной величины длительности обслуживания.

Системы с очередью типа G/G/m относятся к важнейшим моделям, рассматриваемым в теории массового обслуживания. При их исследовании применялись различные методы и получены многочисленные приближенные результаты. Частный случай – система с очередью и одним каналом (m = 1) рассматривается, например, в книге Коэна [1], ориентировочные результаты получены в [2, 3]. Однако до настоящего времени в общем случае не существует достаточно простых и точных, непосредственно применяемых на практике формул, для расчета характеристик качества обслуживания (QoS) в стационарном режиме.

Поскольку во многих случаях одноканальной системы достаточно хорошим приближением является экспоненциальная функция распределения длительности обслуживания заявок, то целью данной статьи является разработка метода расчета основных параметров качества обслуживания заявок в телекоммуникационных системах, представленных моделью G/M/1 (M – *Markov*, экспоненциальное распределение).

Основным математическим аппаратом, позволяющим строить аналитические модели, адекватные распространенным телекоммуникационным системам, является аппарат теории систем массового обслуживания (СМО). При этом для описания функционирования рассматриваемой телекоммуникационной системы используется некоторая СМО с определенной степенью точности, учитывающая ее основные особенности.

Телекоммуникационная система, в которой входящий поток заявок на обслуживание имеет ограничения на доступ к ресурсам системы, также рассматривается как СМО. Если при этом моменты поступления заявок или продолжительность их обслуживания не регламентируются, то при пользовании системой возникают конфликты и образуется очередь. Длина этой очереди зависит от двух характеристик потока заявок – от интенсивности поступления заявок и статистических колебаний этой интенсивности. В случае, когда интенсивность поступления требований превышает пропускную способность системы, система не справляется с потоком этих заявок и начинает расти очередь неограниченной длины. Однако, даже если интенсивность поступления заявок меньше пропускной способности системы, очередь может образоваться из-за статистических колебаний и внезапного накопления заявок (всплеска нагрузки).

1. Основные параметры качества обслуживания в СМО с очередью. Ключевым параметром качества обслуживания (и показателем эффективности функционирования СМО) является *длина очереди* Q, которая определяется количеством заявок, ожидающих обслуживания [4]. Длина очереди зависит от того, когда и сколько заявок поступило в систему, сколько времени затрачено на обслуживание поступивших заявок и т. д. Поскольку длина очереди является случайной величиной, то в качестве показателя длины очереди используется ее математическое ожидание.

С точки зрения обслуживания очень важен и такой показатель как *среднее время ожидания в очереди* t_q, образующееся за счет задержки в очереди. Оно зависит от количества заявок, находящихся в данный момент в очереди, времени окончания обслуживания всех предыдущих заявок и т.д.

Поскольку не все заявки проходят через очередь, а часть из них при наличии свободных каналов системы обслуживаются немедленно, то вводится параметр *среднее время ожидания в системе* W, представляющее собой среднее значение времени ожидания, отнесённое ко всем заявкам – задержанным и не задержанным.

Соотношение задержанных заявок и общего количества поступивших заявок определяет долю ожидающих заявок или *вероятность ожидания* $P_{\text{ож}}$. По этому значению можно найти вероятность того, что обслуживание очередной заявки будет начато. Эта вероятность совпадает с вероятностью того, что в момент прибытия заявки в СМО есть свободный канал.

Степень загруженности обслуживающей системы характеризует такой показатель, как *среднее количество заявок в системе* N, состоящее из заявок, находящихся на обслуживании и в очереди.

Время, проведенное одной заявкой в системе и усредненное по всем заявкам, определяется как с*реднее время нахождения заявки в системе* T. Оно состоит из среднего времени обслуживания и среднего времени ожидания заявок в системе.

Важным параметром является *коэффициент использования системы* ρ, определяемый как отношение интенсивности входящего потока заявок λ к интенсивности обслуживания заявок μ. Для одноканальной системы коэффициент использования равен интенсивности нагрузки и находится в диапазоне $0 \le \rho < 1$ ($\rho < 1$ является условием эргодичности процесса и наличия стационарного распределения вероятностей состояний системы). При этом $\rho = 1 - p_0$, где p_0 – вероятность свободности системы (состояние системы p_0 – занято 0 каналов). Следовательно, ρ – численно совпадает с вероятностью занятости системы $P_{\text{зн}}$ (состояние системы p_1 – занят единственный канал, соответствует доле времени занятости канала). С учетом заявок, находящихся в очереди, в стационарном режиме для системы существует стационарное распределение количества заявок в системе p_k, где k – количество заявок. Это распределение не зависит от момента прибытия заявки в систему.

2. Расчетные формулы. Данные показатели позволяют в конкретных условиях рассчитать все необходимые характеристики протекания процессов в СМО. Тем не менее алгоритм этого расчета существенно зависит от соотношения вероятностей $P_{\text{ож}}$ и $P_{\text{зн}}$. Для одноканальной модели M/G/1 с пуассоновским входящим потоком заявок и произвольным распределением длительности их обслуживания данные вероятности численно равны и, следовательно, $P_{\text{ож}} = \rho$.

Однако, для одноканальной модели G/M/1 такого равенства нет, т.е. по этому параметру модели не инвариантны. В [4, с. 272] показано, что система G/M/1 приводит к геометрическому распределению количества заявок в системе в моменты поступления новых заявок r_k, где k – количество заявок. Распределение p_k отличается от распределения r_k тем, что $p_0 = 1 - P_{\text{зн}}$ (или $p_0 = 1 - \rho$), в то время как $r_0 = 1 - P_{\text{ож}}$. Для системы M/G/1 $p_k = r_k$.

Заявка должна ожидать обслуживания с вероятностью $P_{\text{ож}} = 1 - r_0$. Поэтому при экспоненциальном законе распределения длительности обслуживания безусловное распределение длительности ожидания определится так:

$$W(t) = 1 - P_{\text{ож}} e^{-\mu(1-P_{\text{ож}})t}, \quad \text{при} \quad t \ge 0. \qquad (1)$$

Прямыми вычислениями можно найти среднее время ожидания в системе W и все остальные параметры качества обслуживания. Результаты этих вычислений и известные результаты для моделей M/M/1, M/D/1 и M/G/1 приведены в табл. 1.

Таблица 1 – Основные параметры качества обслуживания

Параметр	**Модель**			
	M/M/1	M/D/1	M/G/1	**G/M/1**
$\boldsymbol{P_{\text{зн}}}$	ρ	ρ	ρ	ρ
$\boldsymbol{P_{\text{ож}}}$	ρ	ρ	ρ	$1-\dfrac{\rho}{N}$
$\boldsymbol{Q}$	$\dfrac{\rho^2}{1-\rho}$	$\dfrac{\rho^2}{2(1-\rho)}$	$\dfrac{\rho^2(1+\nu^2)}{2(1-\rho)}$	$\dfrac{\rho \cdot P_{\text{ож}}}{1-P_{\text{ож}}}$
$\boldsymbol{W}$	$\dfrac{\rho}{1-\rho}$	$\dfrac{\rho}{2(1-\rho)}$	$\dfrac{\rho(1+\nu^2)}{2(1-\rho)}$	$\dfrac{P_{\text{ож}}}{1-P_{\text{ож}}}$
$\boldsymbol{t_q}$	$\dfrac{1}{1-\rho}$	$\dfrac{1}{2(1-\rho)}$	$\dfrac{(1+\nu^2)}{2(1-\rho)}$	$\dfrac{1}{1-P_{\text{ож}}}$
$\boldsymbol{N}$	$\dfrac{\rho}{1-\rho}$	$\rho+\dfrac{\rho^2}{2(1-\rho)}$	$\rho+\dfrac{\rho^2(1+\nu^2)}{2(1-\rho)}$	$\dfrac{\rho}{1-P_{\text{ож}}}$
$\boldsymbol{T}$	$\dfrac{1}{1-\rho}$	$1+\dfrac{\rho}{2(1-\rho)}$	$1+\dfrac{\rho(1+\nu^2)}{2(1-\rho)}$	$\dfrac{1}{1-P_{\text{ож}}}$

Поскольку $\mu = 1$, т. е. средняя длительность обслуживания $T_{\text{обсл}} = 1/\mu$ принимается за условную единицу времени, то W, t_q и T оцениваются в единицах средней длительности обслуживания. Для модели M/G/1 через $v = \sigma_{T\text{обсл}} / T_{\text{обсл}}$ обозначен коэффициент вариации длительности обслуживания. Формула для расчета модели M/G/1 получена Поллачеком-Хинчиным [4].

Таблица 2 – Зависимости между параметрами QoS в системе G/M/1

Параметр	**Модель G/M/1**					G/G/1 и G/G/m
	изв. Q	изв. W	изв. t_q	изв. N		
$\boldsymbol{P_{\text{зн}}}$	ρ	ρ	ρ	ρ		$\dfrac{Q}{P_{\text{ож}} \cdot t_q}$
$\boldsymbol{P_{\text{ож}}}$	$\dfrac{Q}{\rho+Q}$	$\dfrac{W}{1+W}$	$1-\dfrac{1}{t_q}$	$1-\dfrac{\rho}{N}$	$\boxed{\dfrac{Q}{N}}$	$\dfrac{Q}{\rho \cdot t_q}, \ \dfrac{W}{t_q}$
$\boldsymbol{Q}$	–	$\rho \cdot W$	$\rho \cdot (t_q - 1)$	$N-\rho$	$N \cdot P_{\text{ож}}$	
$\boldsymbol{W}$	$\dfrac{Q}{\rho}$	–	$\boxed{t_q - 1}$	$\dfrac{N}{\rho}-1$		$P_{\text{ож}} \cdot t_q$,
$\boldsymbol{t_q}$	$1+\dfrac{Q}{\rho}$	$1+W$	–	$\dfrac{N}{\rho}$		$\dfrac{W}{P_{\text{ож}}}$
$\boldsymbol{N}$	$\rho+Q$	$\rho \cdot (1+W)$	$t_q \rho$	–	$\dfrac{Q}{P_{\text{ож}}}$	
$\boldsymbol{T}$	$1+\dfrac{Q}{\rho}$	$1+W$	t_q	$\dfrac{N}{\rho}$		

Из табл. 2 следует, что при наличии лишь одного известного параметра (например, известен параметр Q, W, t_q или N) все остальные параметры рассчитываются через приведенные в таблице соотношения.

Для модели G/M/1 выявлены и проверены при помощи имитационной модели [5] несколько важных свойств одноканальной системы, справедливых только при экспоненциальной длительности обслуживания (формулы выделены рамкой в табл. 2).

Во-первых: среднее время ожидания в очереди t_q численно совпадает со средним временем нахождения заявки в системе T. Это значит, что среднее время ожидания в системе W меньше среднего времени ожидания в очереди t_q на величину средней длительности обслуживания:

$$W = t_q - 1. \tag{2}$$

Во-вторых: вероятность ожидания можно определить как

$$P_{\text{ож}} = \frac{Q}{N}. \tag{3}$$

Хотя по определению вероятность ожидания $P_{\text{ож}}$ есть отношение количества задержанных заявок к общему количеству поступивших заявок, но как установлено в данном случае, долю ожидающих заявок можно определить и как отношение среднего количества заявок в очереди Q к среднему количеству заявок в системе N. Отсюда с учетом формулы Литтла вытекают несколько важных соотношений между параметрами QoS, справедливыми уже для моделей G/G/1 и G/G/m (выделено фоном в правом столбце табл. 2).

Получившая широкое распространение модель самоподобного трафика не имеет точных и надежных методов расчета, что показано в [6, 7]. В работе [8] предложен новый энтропийный метод расчета части характеристик QoS (нет расчета вероятности ожидания) при обслуживании самоподобного трафика. Новые методы расчета параметров QoS предложены и в [9, 10], но все они предназначены для определенных моделей потоков, имеющих пуассоновский или нормальный закон распределения количества занятых каналов в системе.

В заключении можно отметить, что важность предложенного метода расчета основных характеристик качества обслуживания в модели G/M/1 с бесконечной очередью обосновывается существенным усложнением моделей трафика в современных сетях связи и отсутствием адекватных этому усложнению методов расчета параметров QoS. Предложенный метод является более общим и всеобъемлющим по отношению к различным видам входящих потоков, циркулирующих в телекоммуникационных сетях связи.

Литература

1. *Cohen J.W.* The single server queue / J.W. Cohen // North Holland Publishing Company, Amsterdam/London. – 1969.
2. *Штоян Д., Штоян Х.* Оценки среднего времени ожидания в однолинейных системах массового обслуживания / Д. Штоян, Х. Штоян // Изв. АН СССР. Техническая кибернетика. – 1974. – № 6.
3. *Daley D.J.* Bounds on mean waiting times in GI/G/1 queues, Z. Wahrscheinlichkeitstheorie verw. Geb. – 1969. – №10. – P. 305-317.
4. *Клейнрок Л.* Теория массового обслуживания / Л. Клейнрок: пер. с англ. // Машиностроение. – М.: 1979. – 432 с., ил.
5. *Ложковский А.Г.* Моделирование многоканальной системы обслуживания с организацией очереди / А.Г. Ложковский, Н.С. Салманов, О.В. Вербанов // Восточно-европейский журнал передовых технологий. – Х.: 2007. – №3/6 (27). – С. 72 – 76.
6. *Ложковський А.Г.* Дослідження функціонування телекомунікаційних систем в умовах самоподібного трафіка / А.Г. Ложковський, К.Б. Нікіфоренко // Наукові записки УНДІЗ. – К.: 2009. – №2(10). – С. 60-64.
7. *Ложковский А.Г.* Сравнительный анализ методов расчета характеристик качества обслуживания при самоподобных потоках в сети / А.Г. Ложковский // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Вип. 47. – К., 2008. – С. 187–193.
8. *Ложковский А.Г.* Оценка параметров качества обслуживания самоподобного трафика энтропийным методом / А.Г. Ложковский, Р.А. Ганифаев // Наукові праці ОНАЗ ім. О.С. Попова. – 2008. – № 1. – С. 57–62.
9. *Ложковський А.Г.* Розрахунок якості обслуговування в пакетній мережі при необмеженій довжіні накопичувального буфера / А.Г. Ложковський // Зв'язок. – К.: 2009. – №2. – С. 54–58.
10. *Ложковський А.Г.* Методи аналізу і синтезу систем розподілу інформації в умовах реального трафіка / А.Г. Ложковский // Моделювання та інформаційні технології: зб. наук. пр. ІПМЕ НАН України. – Вип. 48. – К., 2009. – С. 141–153.