

ИССЛЕДОВАНИЕ СИСТЕМЫ ОБСЛУЖИВАНИЯ С ОЖИДАНИЕМ
И РЕКУРРЕНТНЫМ ПОТОКОМ ВЫЗОВОВRESEARCH OF A QUEUEING SYSTEM IN DELAY MODE
OF OPERATION AND RECURRENT CALLS FLOW

Аннотация. Исследовано влияние пик-фактора нагрузки на характеристики качества обслуживания в системе с ожиданием и рекуррентным потоком вызовов. Предложена формула расчета средней длительности ожидания задержанных вызовов.

Summary. The influence of traffic peak-factor on a quality of service characteristics in delay-basis system and recurrent calls flow is probed. The calculation formula of average duration of delay for the delayed calls is offered.

При системном подходе к проблеме повышения качества предоставляемых информационных услуг невозможно обойтись без строгих математических методов анализа качества обслуживания вызовов в реальных условиях формирования их потоков. Для систем массового обслуживания (СМО) с потерями и с ожиданием в случае входящего простейшего потока вызовов известны все аналитические выражения для расчета основных характеристик качества обслуживания. Например, для расчета полнодоступной системы при экспоненциальном времени обслуживания используются I и II формулы Эрланга соответственно, а для расчета системы с ожиданием при постоянном времени обслуживания – метод Кроммелина. Практические измерения параметров потоков вызовов показывают, что реальные потоки в телекоммуникационных сетях существенно отличаются от модели простейшего потока тем, что дисперсия интенсивности нагрузки σ^2 может быть от нескольких до десятков раз больше её математического ожидания Λ [1]. При этом такие потоки лучше описываются моделью рекуррентного (пальмовского) потока, являющегося обобщением простейшего. В работах [2, 3] предложен метод оценки вероятности потерь при обслуживании полнодоступной СМО с явными потерями реального потока вызовов. Однако в таких же условиях, но для СМО с ожиданием точных формул расчета не существует. Поэтому, целью данной работы является исследование характеристик качества обслуживания СМО с ожиданием при обслуживании рекуррентного потока вызовов.

Степень отличия рекуррентного потока от простейшего определяет пик-фактор интенсивности нагрузки $S = \sigma^2 / \Lambda$ (для простейшего потока $S = 1$). Учет этого параметра в упомянутом методе [3] позволил получить величину относительной ошибки в расчетах не более $\pm 5\%$. Поскольку на качество обслуживания, хотя и несколько меньшее, чем пик-фактор нагрузки S , влияние оказывает ещё и вид закона распределения длительности обслуживания, то в данном методе он также учитывается, но с помощью поправочных коэффициентов.

Естественно предполагать, что и в системе с ожиданием при обслуживании рекуррентного потока вызовов с большими колебаниями интенсивности нагрузки (больше, чем это предусмотрено пуассоновским распределением), характеристики качества обслуживания будут иными, и будут зависеть как от пик-фактора S , так и от вида закона распределения длительности обслуживания. Поэтому в данной работе СМО исследована для условий, когда пик-фактор S принимает значения в пределах 1...10 (наиболее соответствует реальным потокам), а вид закона распределения длительности обслуживания – регулярный (соответствует технологии с коммутацией пакетов) и экспоненциальный (соответствует технологии с коммутацией каналов в телефонной связи).

Поскольку для рекуррентного потока строгое и точное аналитическое решение задачи получить не удастся, то необходимо использовать моделирование исследуемого процесса. Для этого несколько доработана и модифицирована схема имитационной модели, приведенная в [4]. Дополнительно в ней добавлены комплекты ожидания r , что позволяет исследовать модель СМО типа $GI/G/v/r$. При различных значениях пик-фактора S определяются стационарное распределение числа вызовов в системе (в каналах v и в комплектах ожидания r), вероятность ожидания для поступившего вызова, средняя длина очереди, моменты распределения длительности ожидания и др.

В сериях опытов имитационного моделирования при неограниченном числе комплектов ожидания ($r = \infty$) и показателем времени обслуживания вероятность ожидания для поступившего вызова D_v и средняя длина очереди Q определяются соответственно как:

$$D_v = \sum_{i=0}^{\infty} P_{v+i} \quad \text{и} \quad Q = \sum_{i=1}^{\infty} i \cdot P_{v+i}, \quad (1)$$

где P_i – вероятности состояний системы [5, с. 124]. На рис. 1 показано стационарное распределение числа вызовов в системе для рекуррентного потока, близкого к реальному с $S = 4$ (сплошная линия), и простейшего потока с $S = 1$ (пунктирная линия). Интенсивность нагрузки в обоих случаях составляет 100 вызовов за среднюю длительность обслуживания, а ёмкость системы $v = 105$.



Рисунок 1 – Стационарное распределение состояний системы P_i

Графики демонстрируют, что вероятности состояний системы P_i отличаются весьма существенно, следовательно, будут отличаться и основные характеристики качества обслуживания D_v и Q , рассчитываемые по (1).

Отличие всех характеристик качества обслуживания, получаемых при моделировании обоих потоков сведены в табл. 1, где средние длительности ожидания даны в единицах длительности обслуживания.

Таблица 1

	Вероятность Ожидания, D_v	Ср. длина очереди, Q	Ср. длит. ожид. (всех вызовов), W	Ср. длит. ожид. (задержан. вызовов), t_3
Простейший поток, $S = 1$	0,52073	10,6984	0,10696	0,20526
Реальный поток, $S = 4$	0,67036	33,7203	0,33659	0,48844

Видно, что по сравнению с простейшим потоком вызовов для реального потока с пик-фактором $S = 4$ средняя длительность ожидания любого вызова (ожидающего и не ожидающего) увеличивается и может быть определена в обоих случаях как

$$W = Q / \Lambda. \quad (2)$$

Средняя длительность ожидания только задержанных вызовов также больше, но если при $S = 1$ она может быть определена как $t_3 = 1 / (v - \Lambda)$, то при $S > 1$ точное выражение неизвестно и для постоянного времени обслуживания предлагается рассчитывать так:

$$t_3 \approx \frac{1}{v - \Lambda} \cdot S / k. \quad (3)$$

Для постоянного времени обслуживания при $k = 2$ ошибка аппроксимации результатов имитационного моделирования не превышает 10%. Повышение точности возможно путём уточнения этого коэффициента, весьма сложно зависящего от соотношения v и Λ . Не трудно заметить, что в таком виде (3) при $S = 1$ обращается в исходное выражение для вычисления t_3 , делённому на 2, что соответст-

вует известному результату для одноканальной системы $M/D/1/r=\infty$. Для многоканальной системы это справедливо только при ёмкости системы $\nu = \Lambda$, и с ростом ν коэффициент k нелинейно убывает до 1.

Для экспоненциального времени обслуживания при $k = 2$ ошибка аппроксимации результатов имитационного моделирования может быть больше 10% (при $S = 1$, естественно, $k = 1$). Функциональная зависимость k от ν и Λ несколько иная и для её нахождения требуются дополнительные исследования.

Для сравнения характеристик качества обслуживания в СМО с ожиданием при рекуррентном потоке вызовов с экспоненциальным и постоянным временем обслуживания произведено имитационное моделирование. Результаты моделирования при $\Lambda = 100$ и $\nu = 110$ представлены в табл. 2.

Таблица 2

Поток	Закон длит. обслуж.	Вероятность ожидания, D_ν	Ср. длина очереди, Q	Ср. длит. ожид. (всех заявок), W	Ср. длит. ожид. (задерж. заявок), t_3
S = 1 (простейш.)	эксп.	0,23801	2,38642	0,02386	0,10024
	пост.	0,21396	1,34170	0,01342	0,06270
S = 4 (произвольн.)	эксп.	0,43824	9,7237	0,09723	0,22187
	пост.	0,49997	10,27069	0,10254	0,20509

Результаты моделирования потока с пик-фактором $S = 1$ при экспоненциальном времени обслуживания практически совпадают с расчетными значениями, полученными по второй формуле Эрланга (расхождение $< 1\%$), а при постоянном времени – полученными по методу Кроммелина (системы $M/M/m/r=\infty$ и $M/D/m/r=\infty$ соответственно).

Следует отметить, что вероятность ожидания, рассчитываемая иначе как

$$D_\nu = \frac{C_w}{C} = \frac{W}{t_3} \quad (4)$$

(а это есть доля задержанных вызовов) для простейшего потока независимо от закона распределения длительности обслуживания всегда равна $\sum_{i=\nu}^{\infty} P_i$, т.е. (1) при $\sum_{i=0}^{\infty} P_i = 1$, а для рекуррентного потока с

$S > 1$ – она всегда несколько больше этой величины. Здесь C_w – количество задержанных вызовов, C – общее количество вызовов. Из чего следует, что для рекуррентного потока вызовов с пик-фактором $S > 1$ методика нахождения характеристик качества обслуживания D_ν и Q должна быть отличной от (1).

Характерно, что при $S = 1$ постоянное время обслуживания даёт некоторое увеличение пропускной способности системы, поскольку все параметры на $\approx 10...30\%$ ниже, чем при экспоненциальном времени. И наоборот, эти же параметры при $S > 1$ существенно выше – в несколько раз (см. табл. 2). Очевидно, что фактор «пиковости» интенсивности потока оказывает более существенное влияние, чем закон распределения длительности обслуживания. Справедливость этого утверждения доказывает стационарное распределение числа вызовов в системе, представленное на рис. 2.

На рис. 2 показано стационарное распределение числа вызовов в системе для потока, близкого к реальному с $S = 4$ (штриховая линия), и простейшего потока с $S = 1$ (пунктирная линия – экспоненциальное время, сплошная – постоянное время обслуживания). Очевидно, что если для такого относительно незначительного расхождения стационарных распределений состояния системы при экспоненциальном и постоянном времени обслуживания (пунктирная и сплошная линии) потребовалась разработка специального метода расчета (достаточно сложный в математическом отношении метод Кроммелина), то для рекуррентных потоков с пик-фактором $S > 1$ разработка адекватных методов расчета еще более актуальна.

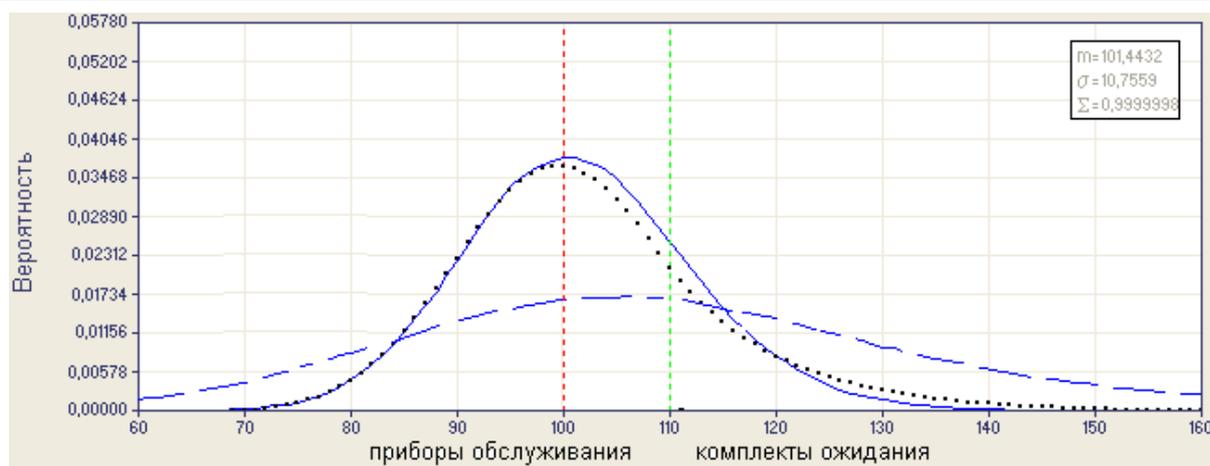


Рисунок 2 – Стационарное распределение числа вызовов в системе

В заключении необходимо подчеркнуть. Для современных мультимедийных сетей с широким диапазоном скоростей передачи (от сотен бит/с до сотен Мбит/с) и нестационарными потоками вызовов, как сказано в [6, с. 281], описание нагрузки кроме средних значений требует дополнительных характеристик. Мультимедийные сети это, как правило, сети с коммутацией пакетов и дисциплиной обслуживания с ожиданием. Здесь источники определённой службы характеризуются максимальной (пиковой) и средней скоростями передачи, их соотношением, т.е. коэффициентом пачкования (*burstness*) и средней длительностью пика нагрузки. Все это может быть интегрально оценено с помощью дисперсии нагрузки или пик-фактора S . Очевидно, что если в таких условиях для расчета условного количества выходных портов пакетного коммутатора, среднего числа задержанных источников (пакеты которых находятся в буфере дольше одного цикла), среднего заполнения входного буфера и пр. использовать II формулу Эрланга [6, с. 285], то с учетом данного исследования ошибка таких расчетов будет достаточно велика.

Литература

1. Ложковский А.Г., Захарченко Н.В., Горохов С.М. Экспериментальная оценка модели потока вызовов на современных телефонных сетях // Наукові праці ОНАЗ ім. О.С. Попова. – №2, 2001. – С. 40-43.
2. Ложковский А.Г. Розрахунок імовірності втрат у разі нормального розподілу кількості зайнятих каналів. – К.: Зв'язок. – №3, 2003. – С. 54-57.
3. Ложковский А.Г. Нова методика оцінювання ймовірності втрат викликів, наближена до реальних умов. – К.: Зв'язок. – №3, 2004. – С. 52-53.
4. Ложковский А.Г. Статистическое моделирование полnodоступного пучка с потерями // Наукові праці ОНАЗ ім. О.С. Попова. – №1, 2003. – С. 75-82.
5. Шнепс М.А. Системы распределения информации. Методы расчета: Справ. пособие. – М.: Связь, 1979. – 344 с., ил.
6. *Современные телекоммуникации. Технологии и экономика* / В.Л. Банкет, О.В. Бондаренко, П.П. Воробийенко и др./ Под общ. ред. Довгого С.А. – М.: Эко-Трендз, 2003. – 320 с.: ил.