

УДК 004.41

ИССЛЕДОВАНИЕ МНОГОКАНАЛЬНЫХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С ЭКВИВАЛЕНТНОЙ ПРОИЗВОДИТЕЛЬНОСТЬЮ

В.В. Соснин

В работе с помощью имитационного моделирования анализируются свойства многоканальных систем с эквивалентной суммарной производительностью. Детально исследуется влияние параметров входящего потока заявок на среднее время пребывания заявок в системе.

Ключевые слова: теория очередей, многоканальные системы массового обслуживания, неэкспоненциальные законы распределения, оценка эффективности, имитационное моделирование.

Введение

При проектировании компьютерных систем и сетей широко применяются модели массового обслуживания. При этом моделями отдельных устройств и узлов в общем случае служат системы массового обслуживания (СМО) типа G/G/N (в символике Кендалла [1, с. 107]). В такой системе на обработку к N приборам одинаковой производительности поступает произвольный поток заявок, а время обслуживания в каждом приборе описывается произвольным законом распределения (ЗР). Под производительностью прибора понимается интенсивность обслуживания заявок. При $N > 1$ система называется многоканальной. Зададим для случая $N = 1$ некоторое значение производительности V_1 для обрабатывающего прибора. Тогда класс СМО с $N \geq 1$, в которых производительность каждого из N приборов равна $V_N = V_1/N$, будем называть *системами с эквивалентной производительностью* (СЭП), так как суммарная производительность всех приборов в них остается постоянной: $V_\Sigma = V_N \cdot N$. При этом все СМО из заданного класса СЭП будут одинаково хорошо справляться с нагрузкой, т.е. обеспечат одинаковый уровень загрузки. В литературе [2, с. 82] описываются присущие СЭП свойства:

- 1) с увеличением N возрастает отказоустойчивость в СЭП (при выходе из строя k приборов суммарная производительность снижается в $N/(N - k)$ раз);

- 2) с увеличением N возрастают накладные расходы на обеспечение взаимодействия между обслуживающими приборами;
- 3) с увеличением N среднее время ожидания заявок в очереди уменьшается, так как при большем числе приборов большая часть заявок немедленно поступает на обслуживание;
- 4) с увеличением N среднее время пребывания заявок в СМО увеличивается.

Из пункта 4 этого перечня следует, что для минимизации задержек выгоднее использовать системы с меньшим числом каналов. Это свойство является очень важным при выборе оптимального числа N , так как время пребывания характеризует эффективность работы СМО в целом. Заметим, что это свойство получено и доказано в результате аналитического моделирования СЭП М/М/Н.

Целью данной работы стало исследование свойств СЭП G/G/N. Применение аналитических методов для этого невозможно, так как при этом понадобилось бы точное задание ЗР входящего потока и ЗР времени обслуживания. Поэтому исследование проводилось на имитационных моделях. В результате многочисленных экспериментов были получены новые свойства СЭП, которые противоречат приведенному выше свойству 4. Таким образом, появилась необходимость более детально исследовать зависимость среднего времени пребывания заявок в системе от числа N для СЭП с неэкспоненциальными потоками. Для этого в работе решались следующие задачи:

- 1) определение факторов, влияющих на исследуемую характеристику (среднее время обслуживания заявок);
- 2) выявление характера влияния на нее каждого из этих факторов;
- 3) формулирование рекомендаций по оптимизации выбора числа N при проектировании систем.

Модель СЭП разработана в среде имитационного моделирования GPSS World. В модели предусмотрена возможность варьирования числа приборов и законов распределения времени обслуживания и времени между прибывающими заявками (гамма-распределение, распределение Кокса, экспоненциальное распределение, равномерное распределение).

Постановка задачи

В качестве примера сравним характеристики двух классов СЭП, М/М/Н и G/G/N, с некоторым заданным набором параметров. На практике бывает удобно рассматривать не конкретный ЗР, а ограничиться указанием лишь первых двух моментов этого закона, так как именно эти два момента оказывают определяющее влияние на характеристики. Поэтому в рассматриваемых системах зададим среднее время между приходом заявок $T_{\Pi} = 100$ единиц времени, коэффициент загрузки $K_z = 60\%$, коэффициент вариации времени обслуживания заявок в приборе (KV_{TO}) и коэффициент вариации времени между приходами заявок (KV_{Π}). Среднее время обслуживания заявок будет определяться в зависимости от N по формуле $T_O = K_z \cdot T_{\Pi} \cdot N$. Очевидно, что две рассматриваемые СЭП будут отличаться только коэффициентами KV_{Π} и KV_{TO} .

На рис. 1, а представлена зависимость времени пребывания заявок СЭП М/М/Н (теоретический расчет) от числа приборов N . Принятые обозначения: U – среднее время пребывания заявки в СМО, W – среднее время ожидания обслуживания заявки. На рис. 1, б показана аналогичная зависимость для СЭП G/G/N (по результатам имитационного моделирования). Здесь и далее на графиках с результатами экспериментов приводится 95 % доверительный интервал (по Стьюденту) всех измеренных характеристик.

График на рис. 1, а иллюстрирует классический случай, когда с точки зрения минимизации задержек наиболее выгодным является использование одноканальной СМО, так как время пребывания $U(N)$ монотонно возрастает. Эксперименты показали, что по-

добный характер поведения системы сохраняется при любой загрузке при условии, что $0 \leq KV_{TO} \leq 1$ и $0 \leq KV_{TP} \leq 1$. Однако стоит увеличить оба коэффициента вариации выше этих ограничений, как это правило теряет силу (рис. 1, б): у функции $U(N)$ теперь наблюдается минимум в точке $N = 3$. Этот пример доказывает, что эмпирическое правило о предпочтительности минимизации числа каналов в СЭП нельзя распространять на весь класс систем G/G/N.

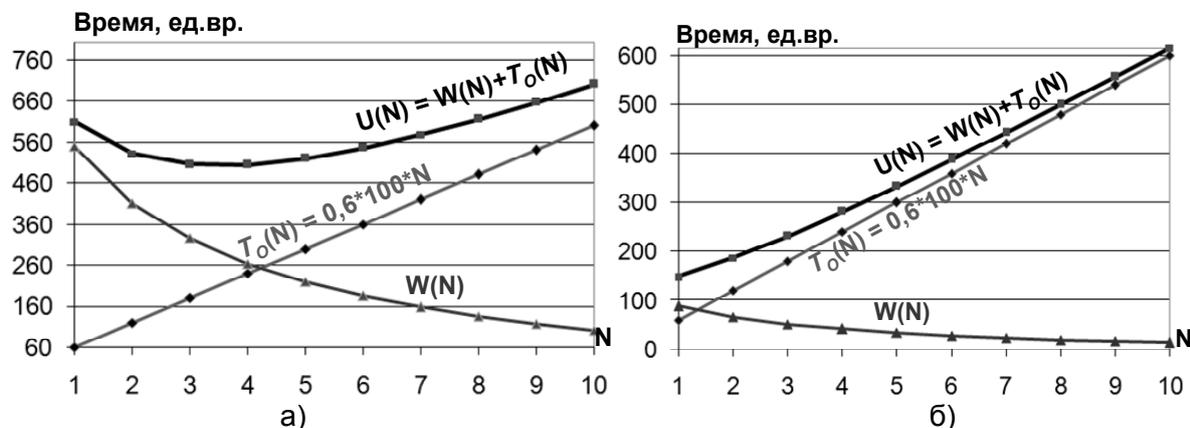


Рис. 1. Зависимость времени пребывания заявок в СЭП от числа приборов:
а) $KV_{TO} = 1, KV_{TP} = 1$; б) $KV_{TO} = 3, KV_{TP} = 1,5$

Таким образом, возникает задача оценки влияния на экстремум функции $U(N)$ следующих параметров:

- 1) коэффициент вариации времен обслуживания заявок в приборе (KV_{TO});
- 2) коэффициент вариации времен между приходами заявок (KV_{TP});
- 3) коэффициент загрузки СМО ($Kз$).

Результаты исследования

В работе исследовалось независимое влияние на U каждого из этих факторов. Ниже приводятся результаты экспериментов, в которых значение исследуемого фактора варьировалось, а значения двух других факторов фиксировались. На рис. 2 приведена зависимость $U(KV_{TO})$ для трех значений N при $T_{П} = 100$ единиц времени, $KV_{TP} = 1$ и $Kз = 60\%$. Для $N=1$ расчет проводился по формуле Поллячека–Хинчина [3, с. 208]. Для $N=2$ и $N=3$ приведен 95%-й доверительный интервал измеренной характеристики. На рисунке $U_{N=i}$ означает среднее время обслуживания заявки при числе обслуживающих приборов, равном i . Обозначим через N_{opt} число приборов, для которого в данном классе СЭП среднее время пребывания заявок является наименьшим. Как видно из рисунка, при $KV_{TO} < 2$ имеем $N_{opt} = 1$, при $2 < KV_{TO} < 3$ имеем $N_{opt} = 2$, при $KV_{TO} > 3$ имеем $N_{opt} = 3$. Очевидно, существует некоторое значение, при превышении которого коэффициентом вариации KV_{TO} оптимальным числом приборов станет 4, 5, 6 и т.д. Наличие этого эффекта подтверждается многочисленными экспериментами.

На рис. 2 зависимость $U_{N=1}(KV_{TO})$, согласно формуле Поллячека–Хинчина, имеет параболический вид: $U_{N=1}(KV_{TO}) = C_1 \cdot KV_{TO}^2 + C_2$, где C_1 и C_2 – константы, не зависящие от KV_{TO} . Расчеты показали, что и две другие зависимости ($U_{N=2}(KV_{TO})$ и $U_{N=3}(KV_{TO})$) имеют с критерием достоверности аппроксимации $R^2 = 0,999$ аналогичный вид, с разницей лишь в значении констант C_1 и C_2 . Знание общего вида этой зависимости позволяет использовать следующий алгоритм анализа систем с неизвестным KV_{TO} :

- 1) на имитационной модели найти несколько пар значений $U_{N=2}$ и KV_{TO} , а также $U_{N=3}$ и KV_{TO} при малых значениях KV_{TO} (так как именно при малых значениях этого коэф-

- фициента вариации получаемые характеристики достаточно точны, т.е. имеют узкий доверительный интервал);
- 2) по найденным парам найти аппроксимирующие функции $U_{N=2}(KV_{TO})$ и $U_{N=3}(KV_{TO})$ – в этом случае аппроксимирующие функции будут определены даже точнее, чем если бы они искались, в том числе, и по известным парам значений с высоким KV_{TO} , приводящим к значительному разбросу в значении получаемых характеристик;
 - 3) решить уравнение $U_{N=2}(KV_{TO}) = U_{N=3}(KV_{TO})$ относительно KV_{TO} ;
 - 4) проделать пункты 1–3 для достаточного числа последующих пар ($U_{N=3}(KV_{TO}) = U_{N=4}(KV_{TO})$, $U_{N=4}(KV_{TO}) = U_{N=5}(KV_{TO})$ и т.д.).

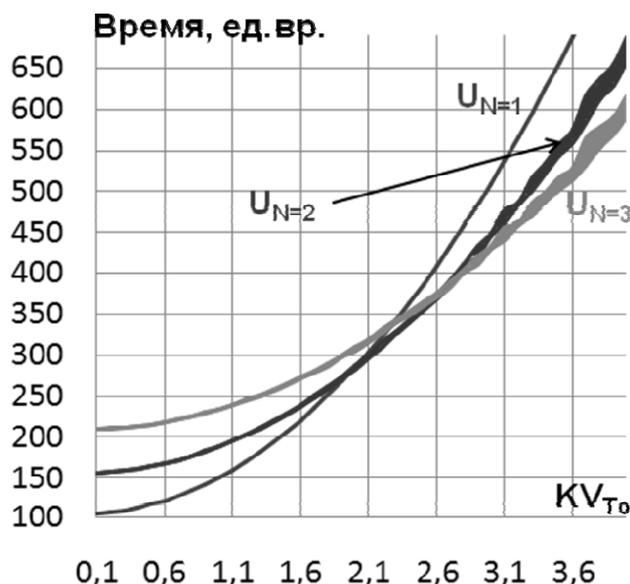


Рис. 2. Зависимость времени пребывания заявок в СЭП от KV_{TO}

В результате будут найдены значения KV_{TO} , являющиеся границами изменения значения N_{opt} . Полученные данные удобно использовать при проектировании рассматриваемого класса систем.

На рис. 3 приведена зависимость $U(KV_{ТП})$ для трех значений N при $T_{П} = 100$ единиц времени, $KV_{TO} = 2$ и $Kz = 60\%$. Каждая из $U_{N=i}$ представлена двумя графиками, задающими нижнюю и верхнюю границу 95%-го доверительного интервала. Очевидно, что ситуация похожа на предыдущую: здесь также наблюдаются точки пересечения, т.е. смена значения N_{opt} . По рисунку можно с уверенностью сказать, что при $KV_{ТП} < 2$ получаем $N_{opt} = 1$, а при $KV_{ТП} > 2$ получаем $N_{opt} > 1$.

При исследовании фактора $KV_{ТП}$ было замечено, что вообще-то наличие точек пересечения зависит от выбранного значения KV_{TO} (он должен быть больше некоторого числа из промежутка 1–2), тогда как при варьировании самого KV_{TO} такой зависимости от других факторов не отмечено.

По аналогии с предыдущим случаем для каждого из $U_{N=i}$ была получена аппроксимирующая функция. Оказалось, что с критерием достоверности аппроксимации $R^2 = 0,999$ для всех трех случаев она имеет вид $U_{N=i} = C_3 \cdot KV_{ТП}^2 + C_4 \cdot KV_{ТП} + C_5$ с разницей лишь в константах C_3 , C_4 , C_5 , которые не зависят от $KV_{ТП}$. Очевидно, что полученная закономерность может быть использована аналогичным образом (можно, например, чисто аналитически узнать значение $KV_{ТП}$, при котором $N_{opt} = 3$, так как определить это по рисунку нельзя из-за перекрытия границ доверительных интервалов $U_{N=2}$ и $U_{N=3}$).

Рассмотрим, наконец, зависимость $U(Kz)$, которая представлена на рис. 3. На этом рисунке $T_{П} = 100$ единиц времени, $KV_{П} = 1$, $KV_{ТО} = 2,5$. Снова видим схожую с предыдущими ситуацию: после некоторого значения варьируемого фактора более выгодным становится использовать СМО с большим числом обслуживающих приборов (на рис. 3 при $Kz > 0,25$ выгоднее использовать 3 прибора, чем 1). Однако и для Kz была замечена зависимость наличия точек смены N_{opt} от выбранного значения $KV_{ТО}$. Таким образом, независимое влияние оказывает только $KV_{ТО}$, а два других фактора нужно рассматривать в непрерывной связи с $KV_{ТО}$.

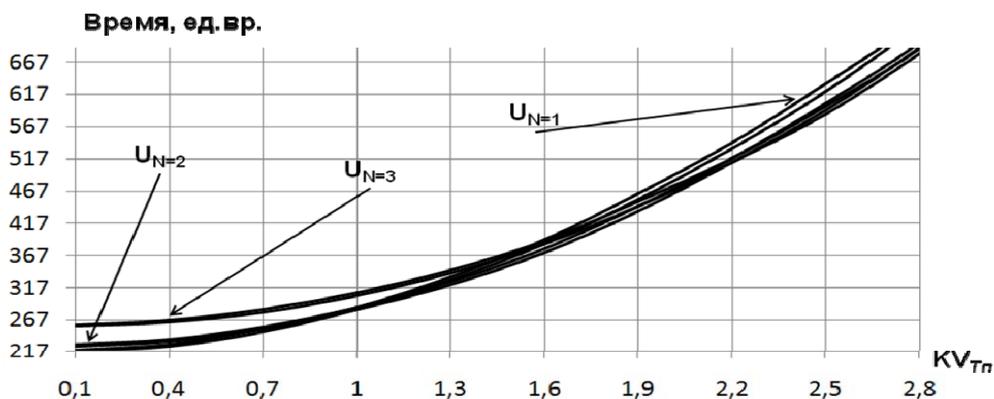


Рис. 3. Зависимость времени пребывания заявок в СЭП от $KV_{П}$

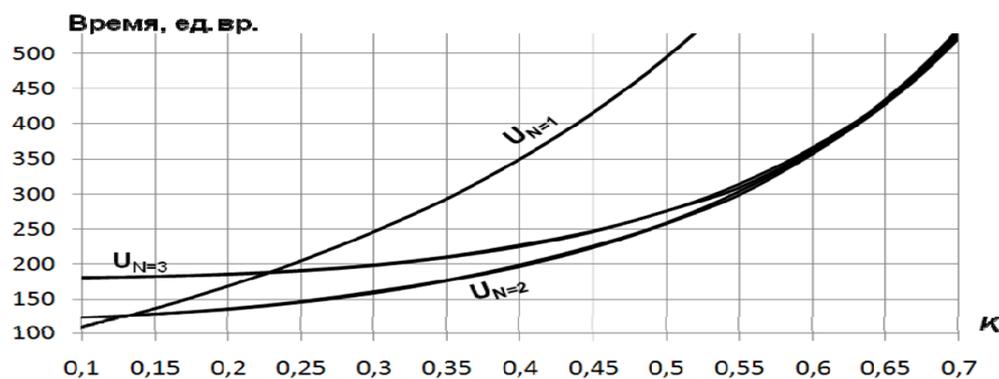


Рис. 3. Зависимость времени пребывания заявок в СЭП от Kz

Аппроксимирующие функции для $U_{N=i}$ с критерием достоверности аппроксимации $R^2=0,999$ для всех трех случаев имеют вид $U_{N=i} = C_6 + C_7/(1-Kz)$ с разницей лишь в константах C_6 и C_7 , которые не зависят от Kz .

Пример

Приведем пример использования полученных результатов на практике. Пусть в некоторой фирме используется выделенный канал Интернет со скоростью 100 Мбит/с на витой паре. Предположим, что в связи с расширением фирмы становится необходимым удвоить указанную производительность канала. Возникает две альтернативы решения задачи:

- 1) проложить дополнительный канал со скоростью 100 Мбит/с на витой паре;
- 2) отказаться от использования существующего канала (сдать его в утиль), проложить новый, основанный на оптоволокне, и уже по нему оплатить подключение на скорости 200 Мбит/с.

Очевидно, что дешевле будет реализация первой альтернативы. Но, согласно бытовавшим представлениям, с точки зрения минимизации задержек выгоднее должен быть второй вариант. Однако рассмотрим конкретную ситуацию: замер параметров трафика в сети Fast Ethernet, состоящей из 500 компьютеров, в течение часа случайно взятого времени показал, что $KV_{\text{ТП}} = 2,1$, $KV_{\text{ТО}} = 1,31$. Время прибытия в этом случае – это время между прибывающими пакетами. А время обработки определяется длиной пакета, так как скорость передачи постоянна. Если провести имитационное моделирование для загрузки 0,05–0,15 (при которой характеристики измеряются очень точно), а затем по полученным данным построить аппроксимирующие функции для числа $N = 1$ и $N = 2$, то для $U_{N=1}$ получим $C_6 = -183$ и $C_7 = 274$, для $U_{N=2}$ получим $C_6 = -80$ и $C_7 = 208$. Если затем составить уравнение $U_{N=1} = U_{N=2}$, то его решением будет $Kз = 0,34$. Получаем, что при $Kз > 0,34$ для минимизации задержек выгоднее использовать два канала, а не один. Иначе говоря, при определенной конфигурации передаваемого потока пакетов становится вполне реальной ситуация, когда даже с точки зрения минимизации задержек выгоднее использовать два низкоскоростных канала вместо одного высокоскоростного.

Выводы

На основании проделанной работы можно сделать следующие выводы.

1. Решение задачи выбора оптимального числа каналов (при сохранении их суммарной производительности) с точки зрения минимизации задержек в М/М/Н нельзя распространять на весь класс G/G/N.
2. Необходимость увеличения числа приборов в СЭП (для минимизации задержек) возрастает:
 - с увеличением загрузки;
 - с увеличением коэффициента вариации времени между приходами заявок;
 - с увеличением коэффициента вариации времени обслуживания заявок.На понятийном уровне такой эффект означает, что большее число каналов (по сравнению с меньшим числом) сглаживает пульсации входящего потока сообщений, давая им быстрее попасть на обслуживание.
3. Независимое влияние на выбор числа каналов оказывает только коэффициент вариации времени обслуживания заявок. Два других фактора увеличивают скорость его воздействия.
4. Полученные аппроксимирующие функции могут быть использованы при выборе оптимального числа каналов в СЭП G/G/N.

Литература

1. Шнепс М. А. Системы распределения информации. Методы расчета. – М.: Связь, 1979. – 344 с.
2. Клейнрок Л. Коммуникационные сети (стохастические потоки и задержки сообщений). – М.: Наука, 1970. – 256 с.
3. Клейнрок Л. Теория массового обслуживания. – М.: Машиностроение, 1979. – 432 с.

Соснин Владимир Валерьевич

– Санкт-Петербургский государственный университет информационных технологий, механики и оптики, аспирант, vsosnin@mail.ru