

СВОЙСТВА БЕСПРИОРИТЕТНОЙ ДИСЦИПЛИНЫ ОБСЛУЖИВАНИЯ В СИСТЕМАХ ВИДА GI/G/1

В. В. Соснин (Санкт-Петербург)

В теории массового обслуживания (ТМО) важное место занимает правило FIFO, называемое также беспriorитетной дисциплиной обслуживания (ДОБП). Традиционно считается [1], что при ДОБП классы заявок разных классов получают одинаковое качество обслуживания, если показателем качества обслуживания считать среднее время ожидания заявки в очереди. В работе [2] показано, что в СМО M/G/1 ДОБП, действительно, не даёт какому-либо классу преимущества в уровне качества обслуживания, т.е. если времена ожидания в очереди заявок k классов суть случайные величины W_1, W_2, \dots, W_k , то их математические ожидания равны: $M[W_1] = M[W_2] = \dots = M[W_k]$.

Целью автора было проверить, обладает ли этим свойством весь класс систем GI/G/1 с ДОБП. При этом в качестве второго показателя качества обслуживания было решено использовать вариацию времени ожидания в очереди (иначе называемую «джиттер задержки»). Эту величину принято использовать для оценки качества обслуживания в компьютерных сетях [3]. Вариация случайной величины вычисляется как разница между квантилью порядка 0.999 и минимальным значением случайной величины. Будем вариацию случайной величины A обозначать как $J[A]$. Исследования проводились для СМО, в которых время обслуживания всех классов имеет одинаковый закон распределения. Примером реальной системы, для моделирования которой можно использовать этот класс СМО, является выходной порт АТМ-коммутатора. В качестве заявок в такой системе рассматриваются АТМ-ячейки, время обработки которых не зависит от типа потока, к которому они принадлежат, так как все ячейки имеют одинаковый размер.

Поставленная задача решалась с помощью имитационного моделирования. Рассмотрим в качестве иллюстрации проблемы пример результатов исследования СМО GI/G/1 с двумя классами заявок, которые создают загрузки $\rho_{BK} = 0.3$ и $\rho_{HK} = 0.03$. Аббревиатура НК означает низконагружающий класс, BK – высоконагружающий класс. Времена обслуживания заявок BK и НК – случайные величины B_{BK} и B_{HK} – такие, что $M[B_{BK}] = M[B_{HK}] = 10$ единиц времени. Времена между приходом заявок BK и НК – случайные величины A_{BK} и A_{HK} . Для моделирования величин A_{BK} , A_{HK} , B_{BK} и B_{HK} используется Гамма-распределение. Каждая из величин имеет фиксированное значение мат. ожидания, а коэффициент вариации изменяется от 0 до 3 с шагом 0.1 так, что $v[A_{BK}] = v[A_{HK}] = v[B_{BK}] = v[B_{HK}] = KV$. Измеряются времена ожидания в очереди заявок каждого из классов (W_{BK} и W_{HK}), а также относительное отличие их средних значений: $\varepsilon = \frac{M[W_{BK}] - M[W_{HK}]}{M[W_{HK}]} \cdot 100\%$.

Результаты имитационного моделирования этой системы приведены на рис. 1, на котором длина вертикальных чёрточек на графиках равна величине 99%-го доверительного интервала измеренных в соответствующей точке величин (по Стьюденту). На рис. 1, а видно, что средние задержки НК и BK заметно различаются за пределами доверительного интервала. На рис. 1, б видно, что среднее время ожидания заявок BK может быть либо на 90% меньше или на 20% больше среднего времени ожидания заявок НК.

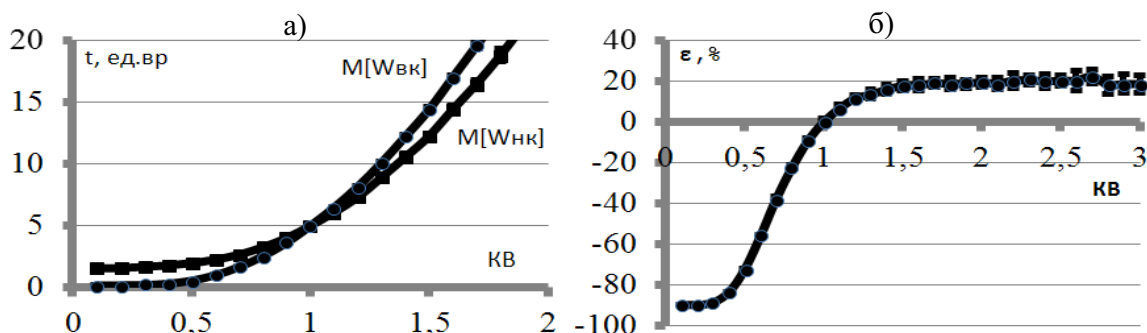


Рис. 1. Среднее время ожидания в очереди заявок НК и ВК:
 а – абсолютные значения; б – процентное отличие

Аналогичный характер имеет соотношение вариации задержек с тем отличием, что относительная разница значений несколько меньше. Приведённый пример позволяет однозначно утверждать, что "свойство бесприоритетности" ДООП в СМО M/G/1 и не может быть распространено на весь класс систем GI/G/1. Однако для инженерной практики одного этого факта мало. Важно уметь численно оценить преимущество в качестве обслуживания того или иного класса заявок для любой СМО, по возможности не прибегая каждый раз к ресурсоёмкому имитационному моделированию. Для решения этой проблемы было проведено большое количество имитационных экспериментов на СМО GI/G/1, чтобы на основе анализа их результатов выявить общие свойства и закономерности, присущие ДООП. Например, проводилась серия экспериментов по факторному плану 2^k для $\nu[A_{VK}], \nu[A_{HK}], \nu[B_{VK}], \nu[B_{HK}]$ при следующих сочетаниях высокого и низкого уровней факторов: 0.2/0.5, 0.2/1.0, 0.2/1.5, 1.0/1.5, 1.5/2.0, – и при нагрузках ВК и НК, изменяющихся от 0.001 до 0.500 с шагом 0.01. Для удобства сравнения W_{VK} и W_{HK} анализировались величины $\delta w = \frac{2 \cdot (M[W_{VK}] - M[W_{HK}])}{M[W_{VK}] + M[W_{HK}]} \cdot 100\%$ и

$$\delta j = \frac{2 \cdot (J[W_{VK}] - J[W_{HK}])}{J[W_{VK}] + J[W_{HK}]} \cdot 100\%,$$

которые имеют ограниченную область допустимых значений $[-200\%; 200\%]$ и дают симметричную оценку сравниваемых величин. Например, при $\delta w = 200\%$ заявки ВК имеют наибольшее относительное преимущество (т.е. ожидают в очереди меньшее время) по отношению к заявкам НК; при $\delta w = -200\%$ ситуация диаметрально противоположная; а при $\delta w = 0\%$ заявки обоих классов получают одинаковый уровень качества обслуживания. Результаты экспериментов по факторным планам позволили сделать вывод, что δw и δj тем больше по модулю, чем:

- больше отношение $\frac{\rho_{VK}}{\rho_{HK}}$;
- меньше сумма $\rho = \rho_{VK} + \rho_{HK}$;
- меньше дисперсия B_{VK} и B_{HK} .

Удалось выявить, что $\delta w \in [C_1; 200)$ и $\delta j \in [C_2; 200)$, где C_1 и C_2 существенно больше 200%, т.е. при некотором сочетании факторов заявки ВК могут получить неограниченно большое преимущество в обслуживании, а для заявок НК возможный выигрыш в качестве обслуживания ограничен некоторым пределом (значения C_1 C_2). Проиллюстрируем этот эффект с помощью рис. 2, на котором приведены результаты имитационных экспериментов при $\nu[B_{VK}] = \nu[B_{HK}] = 0.5$, $\rho_{VK} = 0.3$ и $\rho_{HK} = 0.03$. Координаты ячеек представленных матриц соответствуют использованному в эксперименте

сочетанию $v[A_{BK}]$, $v[A_{HK}]$, а в самих ячейках записаны экспериментально полученные величины δw и δj (соответственно на рис. 2, а и б). Доверительный интервал не указывается, так как его 99%-ое значение составляет не более 1% от приводимых величин. Для упрощения анализа результатов значения δw и δj показаны на фоне, интенсивность затемнения которого пропорциональна абсолютному значению величин.

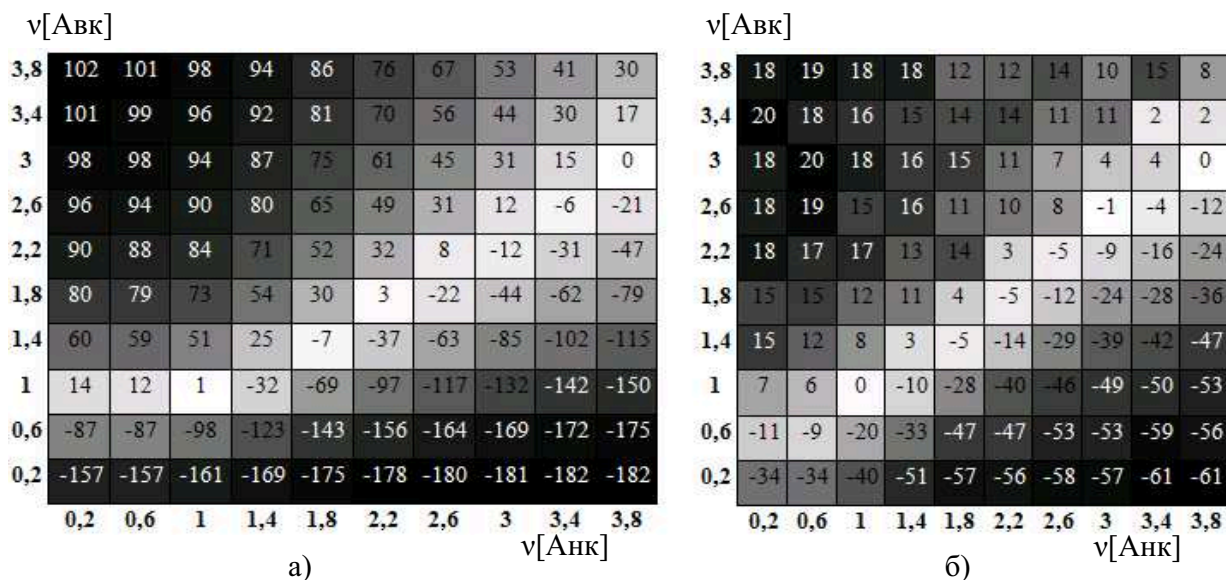


Рис. 2. Взвешенное отклонение времени ожидания НК и ВК:
а – по средним значениям; б – по вариации (джиттеру)

Очевидно, что существует "светлая область" сочетаний значений $v[A_{BK}]$ и $v[A_{HK}]$, при которых заявки НК и ВК получают почти одинаковый уровень качества обслуживания. Но существуют также и две другие "тёмные области", в которых тот или иной класс получает преимущество. Эти результаты можно использовать в оценочных инженерных расчётах, если сформулировать их следующим образом:

– при $v[A_{BK}] > v[A_{BK}] + 1$ преимущество в качестве обслуживания получит ВК;

– при $v[A_{BK}] < v[A_{BK}] - 1$ преимущество в качестве обслуживания получит НК.

Дальнейший анализ экспериментальных данных показал, что заявки НК получают тем лучшее качество обслуживания, чем ближе параметры системы к следующим предельным значениям: $v[A_{HK}] \rightarrow 0$, $v[A_{BK}] \rightarrow \infty$, $v[B_{BK}] \rightarrow 0$, $v[B_{HK}] \rightarrow 0$,

$\frac{\rho_{BK}}{\rho_{HK}} \rightarrow \infty$. Указанное свойство проверялось при увеличении $v[A_{BK}]$ до 5. Дальнейшее

увеличение $v[A_{BK}]$ затруднено, так как для получения результатов с приемлемым доверительным интервалом требуются существенные вычислительные мощности. Примем, что полученное свойство верно для всей области значений $v[A_{BK}]$. Получим на основании этого допущения нижнюю оценку для значения C_1 и C_2 . На рис. 3, а и б штрихами на оси времени отмечены моменты прихода заявок НК и ВК. На рис. 3, в показана диаграмма занятости прибора заявками ВК в отсутствии заявок НК, вертикальными рисками обозначены моменты времени поступления заявок НК, а чёрные прямо-

угольники обозначают периоды занятости прибора $T_3(i)$, где $i=1,2,\dots,\infty$. Остальные промежутки назовём периодами незанятости $T_H(i)$. При $\frac{\rho_{BK}}{\rho_{HK}} \rightarrow \infty$ получим, что

$$\frac{\sum_{i=1}^{\infty} T_3(i)}{\sum_{i=1}^{\infty} (T_3(i) + T_H(i))} = \rho_{BK} \approx \rho. \tag{1}$$

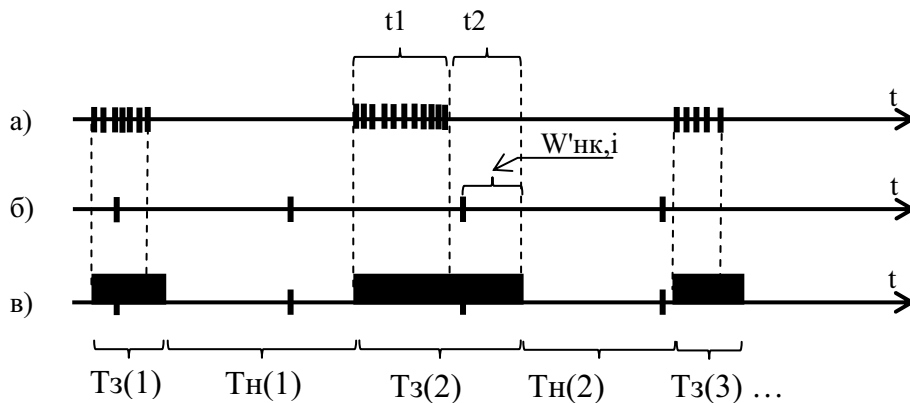


Рис. 3. Временная диаграмма:

a – поступления заявок ВК; **б** – поступления заявок НК; **в** – занятости прибора

Рассмотрим отдельно взятый период $T_3(i)$ на примере $T_3(2)$. Пусть в этот период пришла группа из $N_{BK}(i) = N_{BK}(2)$ заявок ВК. При $\nu[A_{BK}] \rightarrow \infty$ получим, что $N_{BK}(2)$ растет, а время между появлением заявок внутри группы уменьшается. При $\nu[B_{BK}] \rightarrow 0$ длительность $T_3(2)$ пропорциональна $N_{BK}(i)$ и зависит от загрузки, но не зависит от $\nu[A_{BK}]$, т.е. при $\nu[A_{BK}] \rightarrow \infty$ и $\nu[B_{BK}] \rightarrow 0$ получим, что $\frac{t1}{t2} \rightarrow 0$. Тогда времена ожидания $N_{BK}(2)$ заявок ВК, пришедших в период $T_3(2)$, равны

(в порядке поступления): $0, \frac{T_3(2)}{N_{BK}(2)}, \frac{2 \cdot T_3(2)}{N_{BK}(2)}, \dots, \frac{(N_{BK}(2)-1) \cdot T_3(2)}{N_{BK}(2)}$, а их мат.ожидание

равно:
$$\frac{\sum_{k=1}^{N_{BK}(2)} \frac{(k-1) \cdot T_3(2)}{N_{BK}(2)}}{N_{BK}(2)} = \frac{T_3(2) \cdot \sum_{k=1}^{N_{BK}(2)} (k-1)}{N_{BK}^2(2)} = \frac{T_3(2) \cdot (N_{BK}(2)-1)}{2 \cdot N_{BK}(2)}$$

всех значений i , получим:
$$M[W_{BK}] = \frac{M[T_3(i)] \cdot (M[N_{BK}(i)] - 1)}{2 \cdot M[N_{BK}(i)]}$$

Рассмотрим случайную величину $W'_{HK,i}$ – время ожидания заявок НК, которые придут на обслуживание в течение $T_3(i)$. При $\frac{\rho_{BK}}{\rho_{HK}} \rightarrow \infty$ и $\nu[A_{HK}] \rightarrow 0$ они ждут обслуживания только заявок ВК. Все заявки НК, поступившие в СМО в течение $T_3(i)$, попадут в период времени $t2$, поскольку $\frac{t1}{t2} \rightarrow 0$. Следовательно, они будут ждать освобождения прибора до конца данного $T_3(i)$ как раз в течение времени $W'_{HK,i}$, сред-

нее значение которого $M[W'_{HK,i}] = \frac{T_3(i)}{2}$. За время $T_3(i)$ в СМО в среднем поступит

$N_{HK}(i) = \frac{T_3(i)}{M[A_{HK}]}$ заявок. Тогда суммарное время ожидания всех заявок НК, поступивших в СМО в течение Q периодов занятости, равно

$S(W'_{HK}) = \sum_{i=1}^Q (N_{HK}(i) \cdot M[W'_{HK,i}]) = \sum_{i=1}^Q \left(\frac{T_3^2(i)}{2 \cdot M[A_{HK}]} \right)$. В периоды загрузки по-

ступит $N'_{HK} = \sum_{i=1}^Q N_{HK}(i)$ заявок НК, тогда среднее время ожидания в очереди равно:

$$M[W'_{HK}] = \lim_{Q \rightarrow \infty} \frac{S(W'_{HK})}{N'_{HK}} = \lim_{Q \rightarrow \infty} \frac{\sum_{i=1}^Q \left(\frac{T_3^2(i)}{2 \cdot M[A_{HK}]} \right)}{\sum_{i=1}^Q \frac{T_3(i)}{M[A_{HK}]}} = \lim_{Q \rightarrow \infty} \frac{\sum_{i=1}^Q \frac{T_3^2(i)}{Q}}{2 \cdot \sum_{i=1}^Q \frac{T_3(i)}{Q}} =$$

$$= \frac{\alpha_2[T_3(i)]}{2 \cdot M[T_3(i)]} = \frac{M^2[T_3(i)] \cdot (v^2[T_3(i)] + 1)}{2 \cdot M[T_3(i)]} = \frac{M[T_3(i)]}{2} \cdot (v^2[T_3(i)] + 1). \quad (2)$$

Пусть $W''_{HK}(i)$ – случайная величина, равная времени ожидания в очереди тех заявок НК, которые пришли в СМО в течение $T_n(i)$. В соответствии с формулой (1) доля таких заявок среди всех заявок НК составляет $(1 - \rho)$. Тогда среднее время ожидания W_{HK} заявок НК, пришедших в произвольный момент времени, равно $M[W_{HK}] = \rho \cdot M[W'_{HK}] + (1 - \rho) \cdot M[W''_{HK}]$. Очевидно, что $M''[W_{HK}] \rightarrow 0$, следовательно, $M[W_{HK}] \rightarrow \rho \cdot M[W'_{HK}] = \frac{\rho \cdot M[T_3(i)]}{2} \cdot (v^2[T_3(i)] + 1)$, а с учётом форму-

лы (2).
$$\frac{M[W_{HK}]}{M[W_{BK}]} \geq \frac{\frac{\rho \cdot M[T_3(i)]}{2} \cdot (v^2[T_3(i)] + 1)}{\frac{M[T_3(i)] \cdot (M[N_{BK}(i)] - 1)}{2 \cdot M[N_{BK}(i)]}} = \frac{\rho \cdot (v^2[T_3(i)] + 1) \cdot M[N_{BK}(i)]}{M[N_{BK}(i)] - 1}.$$

Учитывая, что $\frac{M[N_{BK}(i)]}{M[N_{BK}(i)] - 1} > 1$ и $v^2[T_3(i)] \geq 0$, получим $\frac{M[W_{HK}]}{M[W_{BK}]} \geq \rho$. Этот ре-

зультат полностью согласуется с данными имитационных экспериментов. Аналогичное выражение оказывается верным и для вариации задержек. Окончательно полученный результат формулируется так:

$$M[W_{HK}] > \rho \cdot M[W_{BK}] \text{ и } J[W_{HK}] > \rho \cdot J[W_{BK}]. \quad (3)$$

Приведём пример инженерного применения полученных выражений. Рассмотрим процессы, протекающие на отдельно взятом выходном порту некоторого магистрального маршрутизатора. При этом в качестве потока заявок ВК рассматривается сетевой трафик, который проходит через рассматриваемый порт. Современные маршрутизаторы позволяют в реальном времени получить данные о текущих характеристиках передачи пакетов, поэтому будем считать информацию о средних задержках пакетов потока ВК заведомо доступной и достоверной. В качестве потока заявок НК рассматри-

вается сетевой трафик низкой интенсивности, который администратор сети планирует дополнительно пустить через рассматриваемый порт. Тогда полученные неравенства (3) позволяют оценить максимально возможное преимущество в качестве обслуживания, которое могут получить заявки НК. Например, пусть заявки ВК создают 60%-ю нагрузку, а их среднее время ожидания в очереди равно 200 мс. Пусть требуется запустить в эту СМО ещё один поток заявок, нагружающий её на 3%. Тогда формула (3) позволяет утверждать, что время ожидания в очереди заявок нового потока не может быть менее 123 мс. Если бы к качеству обслуживания заявок НК применялись требования ITU-T Y.1541 для трафика класса 0 (средняя задержка не должна превышать 100 мс), то полученная оценка позволила бы системному администратору отвергнуть саму возможность добавления этого потока данных в компьютерную сеть ещё на этапе планирования, не проводя натурных экспериментов.

Заключение

1. С помощью имитационных экспериментов показано, что при беспriorитетной дисциплине FIFO заявки разных типов могут иметь разные средние значения времени ожидания в очереди, а также разные вариации этого времени.

2. Получена численная оценка нижней границы для показателей ожидания низконагружающих потоков заявок. Эта оценка имеет практическое значение, позволяя решать некоторые задачи анализа и синтеза беспriorитетных систем массового обслуживания с неоднородным потоком заявок.

Литература

1. **Bolch G., Greiner S., Meer H., Triverdi K.** Queueing networks and Markov chains: modeling and performance evaluation with computer science applications. New York: John Wiley & Sons, 1998.
2. **Клейнрок Л.** Вычислительные системы с очередями. М.: Мир, 1979. 600 с.
3. ITU-T recommendation Y.1541. Amendment 1: An example showing how to calculate IPDV across multiple sections. Geneva, 2006.