

УДК 519.872.1

ОЦЕНИВАНИЕ ВЕРОЯТНОСТИ БЛОКИРОВКИ В СИСТЕМЕ С ПОВТОРНЫМИ ВЫЗОВАМИ И ПОСТОЯННОЙ СКОРОСТЬЮ ВОЗВРАЩЕНИЯ ЗАЯВОК С ОРБИТЫ

Е. В. Морозов, Р. С. Некрасова

*Институт прикладных математических исследований
Карельского научного центра РАН*

В статье исследуется вероятность блокировки заявки в системе с повторными вызовами и постоянной интенсивностью повторных заявок, идущих с орбиты на сервер. Изучаются теоретические свойства, а также свойства различных статистических оценок этой вероятности. Анализ проводится с помощью регенеративного метода. Исследован эффект уменьшения дисперсии оценки. Результаты имитационного моделирования хорошо согласуются с известными аналитическими результатами, а также с соответствующими результатами для системы с потерями.

Ключевые слова: система с повторными вызовами, вероятность блокировки, зона стационарности/нестационарности, (квази-)регенерации, уменьшение дисперсии.

E. V. Morozov, R. S. Nekrasova. ESTIMATION OF BLOCKING PROBABILITY IN A RETRIAL QUEUING SYSTEM WITH CONSTANT RETRIAL RATE

The paper investigates the blocking probability in a retrial system with constant retrial rate. Both theoretical properties and the properties of various estimates of the probability are studied. The analysis is based on the regenerative method. The variance reduction effect is studied. Simulation results are consistent with known analytical results and with corresponding results for a lossy system.

Key words: retrial queuing system, blocking probability, stability/instability region, (quasi-) regenerations, variance reduction.

ВВЕДЕНИЕ

Данная статья посвящена применению регенеративного метода для доверительного оценивания стационарной вероятности блокировки (ухода на орбиту) заявки в односерверной системе с повторными вызовами без буфера для ожидания и с постоянной интенсивностью возвращения заявок с орбиты на сервер. Показано, что данный метод применим как в

области стационарности, так и в области нестационарности системы, когда величина орбиты неограниченно растет. В выводе соотношений для вероятности блокировки существенно использован метод доказательства обобщенной формулы Литгла для классических систем обслуживания с буфером. Рассмотрены различные оценки вероятности

блокировки. Для уменьшения дисперсии используется выпуклая комбинация *прямой оценки* (отношения числа отказов к общему числу приходов) и оценки доли времени занятости сервера. Приведены результаты численного моделирования для ряда систем.

ОПИСАНИЕ СИСТЕМЫ И УСЛОВИЕ СТАЦИОНАРНОСТИ

Рассмотрим односерверную систему с повторными вызовами типа $M/G/1/1$ (без буфера) с входным пуассоновским потоком с интенсивностью λ и произвольным (типичным) временем обслуживания S со средним $ES := 1/\mu$. Заявки, поступающие в систему, когда сервер занят, уходят на орбиту бесконечно большого объема, а затем вновь пытаются попасть на обслуживающее устройство. В отличие от классических систем, поток повторных заявок (попыток) является пуассоновским с параметром μ_0 , *не зависящим от величины (непустой) орбиты*. Такая модель успешно применена при моделировании телефонных линий [Fayolle, 1986], протоколов множественного доступа АЛОНА [Choi et al., 1993], протокола TCP с короткими сообщениями (short TCP transfers) [Avrachenkov, Yechiali, 2008]. В частности, рассматриваемая в данной работе модель односерверной системы без буфера описывает протокол множественного доступа АЛОНА, когда при наличии на орбите n заявок каждая поступает (повторно) в систему с интенсивностью μ_0/n . Как хорошо известно, тогда суммарный поток является пуассоновским с параметром μ_0 .

Таким образом, суммарный поток заявок, поступающих на сервер, складывается из двух (вообще говоря, зависимых) потоков: λ -потока первичных заявок (λ -заявок) и потока вторичных заявок с интенсивностью $\tilde{\mu}_0 \leq \mu_0$. (Величина $\tilde{\mu}_0$ строго определяется ниже см. (21).) Для удобства далее обозначим исходную систему через Σ .

Рассмотрим также следующую вспомогательную односерверную систему $\hat{\Sigma}$ с потерями, с нулевым буфером и без орбиты. Система $\hat{\Sigma}$ имеет тот же входной λ -поток, то же распределение времени обслуживания, что и исходная система Σ , но, в отличие от Σ , имеет еще один (независимый) пуассоновский входной поток заявок с параметром μ_0 . (Назовем их μ_0 -заявки.) Таким образом, если в момент прихода (любой) новой заявки сервер занят, то она *покидает систему и не оказывает влияния на ее будущее состояние*. Для удобства можно считать, что такие заявки уходят на виртуальную орбиту, описываемую как систе-

ма вида $\cdot/M/1$ с интенсивностью обслуживания μ_0 , и покидают систему навсегда после обслуживания.

Можно ожидать, что сервер в системе $\hat{\Sigma}$ более загружен, чем в исходной системе, так как интенсивность потока повторных заявок $\tilde{\mu}_0 \leq \mu_0$. Более того, можно ожидать, что с ростом величины орбиты $\tilde{\mu}_0 \uparrow \mu_0$, и поэтому распределение состояний сервера (занят или пуст) в системе Σ должно сближаться с соответствующим распределением в системе $\hat{\Sigma}$. Следовательно, поток потерь в системе $\hat{\Sigma}$ (или поток заявок, идущих на виртуальную орбиту) должен быть более интенсивным, чем поток заблокированных заявок, идущих на орбиту в исходной системе. Отметим, что источником неустойчивости (нестационарности) исходной системы (с конечным буфером) может быть только неограниченно растущая орбита. Тогда условие устойчивости виртуальной орбиты должно гарантировать устойчивость орбиты в (менее загруженной) системе Σ .

Для пояснения рассмотрим орбиту как односерверную систему с входным потоком заявок, получивших отказ. Интуитивно понятно, что верхняя граница интенсивности входного потока на сервер равна $\lambda + \mu_0$ (что соответствует постоянно непустой орбите). Можно ожидать, что поток с большей интенсивностью приводит к большему числу потерь, и таким образом, более интенсивному потоку заявок на орбиту. Этот вывод опирается на свойство монотонности систем с потерями относительно интервалов входного потока, которое кратко рассматривается ниже. Суммарный поток заявок на сервер в системе $\hat{\Sigma}$ (в отличие от исходной системы Σ) представляет собой суперпозицию *независимых пуассоновских потоков* с интенсивностями μ_0 и λ . В исходной системе поток повторных заявок прерывается, когда начинается *период простоя* орбиты. Этот период заканчивается при поступлении очередной повторной заявки.

Обозначим через P_{loss} стационарную вероятность потери в системе $\hat{\Sigma}$. (Точное определение P_{loss} приведено ниже.) Из вышесказанного следует ожидать, что условие

$$(\lambda + \mu_0)P_{loss} < \mu_0 \quad (1)$$

является достаточным условием устойчивости орбиты в системе Σ . Действительно, этот результат доказан в работе [Avrachenkov, Morozov, 2010].

Замечание 1. На самом деле данный результат доказан в [Avrachenkov, Morozov, 2010] для

систем с повторными вызовами более общего вида, имеющих несколько серверов, ненулевой буфер и входной λ -поток восстановления. Более того, условие (1) является *критерием стационарности*, по крайней мере, для систем с пуассоновским λ -поток, в том числе для рассматриваемой в данной статье системы.

Обсудим кратко свойство монотонности системы с потерями, установленное в [Sonderman, 1979, 1981]. (Этот результат верен в соответствующей форме и для многосерверной системы с ненулевым буфером.)

Обозначим через t_n момент прихода n -й заявки, B_n – n -й фактический момент прихода (когда заявка поступает на обслуживание), через S_n – время n -го обслуживания, а через D_n – n -й момент завершения обслуживания, $n \geq 1$. Пусть n -я заявка (пришедшая в момент t_n) попала в систему, и это есть m -й фактический вход в систему во временном интервале $[0, t_n]$. Обозначим этот момент через $t(m)$, таким образом, введенные величины связаны как $t_n = t(m) := B_m$, $n \geq m \geq 1$. Напомним, что ν_n – число заявок в системе перед n -м приходом. Имеют место следующие соотношения (в предположении, что система изначально пуста):

$$\begin{aligned} B_{n+1} &= \min(t_i : t_i \geq t(n), \nu_i = 0), \quad B_0 = 0, \\ D_1 &= B_1 + S_1, \\ D_{n+1} &= (\max(D_n, B_{n+1}) + S_{n+1}), \quad n \geq 1. \end{aligned} \quad (2)$$

В очевидных обозначениях рассмотрим две системы с потерями Σ_1 и Σ_2 , у которых интервалы между приходами (стохастически) равны, $\tau_n^{(1)} =_{st} \tau_n^{(2)}$, а времена обслуживания упорядочены как $S_n^{(1)} \leq S_n^{(2)}$, $n \geq 1$ (с вероятностью 1 или стохастически). В [Sonderman, 1979] по индукции доказано, что $D_n^{(1)} \leq D_n^{(2)}$, $B_n^{(1)} \leq B_n^{(2)}$, $\nu_n^{(1)} \leq \nu_n^{(2)}$, $n \geq 1$. Этот результат непосредственно переносится на случай непрерывного времени (на число событий в любом интервале $[0, t]$). Также отсюда легко следует, что число потерь в обеих системах в интервале $[0, t]$ связано как $R^{(1)}(t) \leq R^{(2)}(t)$, $t \geq 0$. Именно этот результат является ключевым при доказательстве условия стационарности в работе [Avrachenkov, Morozov, 2010]. (Аналогично, в [Sonderman, 1979] доказано, что при $\tau_n^{(1)} \geq \tau_n^{(2)}$ и $S_n^{(1)} = S_n^{(2)}$ также $\nu_n^{(1)} \leq \nu_n^{(2)}$, $n \geq 1$.) Заметим, что для экспоненциальных распределений времен обслуживания упорядоченность $S_n^{(1)} \leq S_n^{(2)}$ сводится к неравенству $\mu_1 \geq \mu_2$ для параметров этих распределений.

Чтобы применить результаты о монотонности к анализу вероятности блокировки в исходной системе с повторными вызовами и *без потерь*, заметим, что различие между системами Σ и $\hat{\Sigma}$ в этом анализе проявляется лишь на *периодах простоя* орбиты в исходной системе. Напомним, что λ -потоки у обеих систем совпадают. Чтобы синхронизировать поток с орбиты и пуассоновский (с параметром μ_0) в системе $\hat{\Sigma}$, мы продолжаем отсчитывать *фиктивные* интервалы выходного потока с орбиты и на периодах простоя (эти интервалы завершаются приходом фиктивных заявок). В моменты завершения каждого периода простоя орбиты (в системе Σ) разыгрываются заново незавершенные интервалы в пуассоновских μ_0 -потоках в обеих системах Σ и $\hat{\Sigma}$. По свойству потери памяти экспоненциального распределения эта процедура не меняет распределения потоков. Кроме того, фиктивным заявкам назначается *нулевое время обслуживания*. В результате моменты (суммарного) входного потока в обеих системах оказываются полностью синхронизованными, т. е. (в очевидных обозначениях) $\tau_n = \hat{\tau}_n$, в то время как времена обслуживания упорядочены как $S_n \leq \hat{S}_n$, $n \geq 1$. Остается применить свойство монотонности системы с потерями, приведенное выше.

Для некоторых систем с потерями известны аналитические формулы для вероятности потери, в частности, для обобщенной c -серверной системы Эрланга $M/G/c/c$ (без буфера)

$$P_{loss} = \frac{\left((\lambda + \mu_0)/\mu\right)^c}{c!} \left[\sum_{n=0}^c \frac{\left((\lambda + \mu_0)/\mu\right)^n}{n!} \right]^{-1}. \quad (3)$$

Это позволяет записать условие (1) для системы $M/G/1/1$ в виде

$$\frac{\lambda + \mu_0}{\lambda + \mu_0 + \mu} < \frac{\mu_0}{\lambda + \mu_0}, \quad (4)$$

что эквивалентно

$$\frac{\lambda^2}{\mu_0} + \lambda < \mu. \quad (5)$$

Полагая $\lambda = 1$, последнее условие можно записать в форме

$$\frac{1}{\mu_0} + 1 < \mu, \quad (6)$$

удобной для численного исследования области стационарности по двум параметрам μ_0 , μ , характеризующим скорости ухода заявок с орбиты и обслуживания на сервере, соответственно. Напомним, что проведенный анализ

верен для произвольного времени обслуживания, в частности, имеющего (стандартное) распределение Парето интервалов $P(S > x) = x^{-\alpha}$, $\alpha > 1$, $x \geq x_0 = 1$. Тогда $ES = \frac{\alpha}{\alpha-1} := 1/\mu$ и условие (5) принимает вид

$$\lambda \left(\frac{\lambda}{\mu_0} + 1 \right) < 1 - 1/\alpha. \quad (7)$$

(В последней формуле нельзя положить $\lambda = 1$, что связано с выбором дополнительного параметра $x_0 = 1$.)

РЕГЕНЕРАТИВНАЯ СТРУКТУРА СИСТЕМЫ С ПОВТОРНЫМИ ВЫЗОВАМИ

В данном разделе описана регенеративная структура процессов в исходной системе с повторными вызовами.

Пусть $N(t)$ – число заявок на орбите (величина орбиты) в момент времени t , а $\nu(t)$ – состояние сервера (0 или 1) в момент t . Пусть $\{t_n\}$ – моменты прихода λ -заявок, $N(t_n^-) = N_n$, $\nu(t_n^-) = \nu_n$, $n \geq 1$. Введем общее число заявок в системе $X = \{X(t) = N(t) + \nu(t), t \geq 0\}$, и положим $X(t_n^-) = X_n := N_n + \nu_n$, $n \geq 1$. Заметим, что система $\hat{\Sigma}$ регенерирует в моменты, когда λ -заявки встречают пустой сервер (здесь используется свойство потери памяти входного потока μ_0 -заявок). Исходная же система Σ регенерирует в моменты, когда заявка поступает в полностью пустую систему (пустой сервер и пустая орбита). Таким образом, моменты регенерации основного процесса $\{X_n\}$ в дискретном времени (полученного из исходного процесса X путем наблюдения лишь в моменты прихода λ -заявок) определяются следующим образом:

$$\beta_{n+1} = \inf_k (k > \beta_n : X_k = 0), \quad n \geq 0 \quad (\beta_0 = 0). \quad (8)$$

Процесс регенераций $\{\beta_n\}$ (и основной процесс X) называется положительно возвратным, если $E\beta < \infty$ [Morozov, 2004]. (Эта терминология, ставшая уже общепризнанной, заимствована из теории цепей Маркова, возвратных по Харрису, у которых положительная возвратность также означает конечность средней длины цикла регенерации. По этому поводу (см. [Wolf, 1989; Morozov, 2004]).

Обозначим через $R(t)$ – общее число неудачных попыток заявок попасть на сервер в интервале $[0, t]$ (т. е. это поток, идущий на орбиту), а через $A(t)$ – общее число попыток попасть на сервер в этом интервале (порожденное первичными λ -заявками и повторными μ_0 -заявками). Рассмотрим вначале стационарный режим. Иными словами, пусть условие (1)

выполнено, тогда процесс X – положительно возвратен. Обозначим через R и A типичное число орбитальных заявок и (типичное) общее число попыток на цикле регенерации, соответственно. Пусть $I_n = 1$, если n -я попытка неудачна (иначе, $I_n = 0$). Поскольку, как легко увидеть, $P(\tau > S) > 0$, то длина цикла регенерации β – непериодическая случайная величина. Поэтому существует предел по распределению $I_n \Rightarrow I$. Более того, ввиду равномерной интегрируемости индикаторов, имеет место также сходимость в среднем

$$P(I_n = 1) \rightarrow P_{orb}, \quad n \rightarrow \infty, \quad (9)$$

где $P_{orb} := P(I = 1)$ есть стационарная вероятность того, что вновь поступающая заявка идет на орбиту. Рассмотрим выборочные средние

$$B_N(t) := \frac{R(t)}{A(t)}, \quad P_{orb}(n) := \frac{\sum_{k=1}^n I_k}{n}. \quad (10)$$

Из теории регенерации для этих оценок с вероятностью 1 следует сходимость

$$\lim_{t \rightarrow \infty} B_N(t) = \lim_{n \rightarrow \infty} P_{orb}(n) = \frac{ER}{EA} = P_{orb}. \quad (11)$$

Так как исходная система менее загружена, чем система $\hat{\Sigma}$, то можно ожидать, что в области стационарности исходной системы

$$P_{orb}(n) \rightarrow P_{orb} < P_{loss}, \quad \text{при } n \rightarrow \infty. \quad (12)$$

С другой стороны, в области нестационарности исходной системы Σ можно ожидать неограниченного роста величины ее орбиты. В этом случае поток повторных заявок (идущих с орбиты на сервер) через некоторое время t_0 склеивается с пуассоновским потоком с параметром μ_0 . Иными словами, после момента t_0 поток повторных заявок стохастически эквивалентен пуассоновскому потоку. (Такое склеивание называется *каплингом* [Asmussen, 2002].) Поэтому можно ожидать, что $P_{orb}(n) \rightarrow P_{loss}$. Но при условии нестационарности процесс X не является положительно возвратным (т. е. $E\beta = \infty$), и применение регенеративного метода для получения предела P_{orb} , вообще говоря, становится проблематичным. Однако в данном случае для оценки P_{orb} можно использовать другой тип регенераций, а именно, *квази-регенерации*. Квази-регенерации исходной (нестационарной) системы определяются как моменты, когда поступающая в систему первичная заявка встречает пустой сервер (в то время как орбита, вообще говоря, может быть не пустой).

Более точно, положим $\alpha_0 = 0$ и определим рекурсивно моменты квази-регенераций следующим образом:

$$\alpha_{n+1} = \inf_k (k > \alpha_n : \nu_k = 0), n \geq 0. \quad (13)$$

Возможность использования данных моментов в нестационарном режиме для оценки P_{orb} рассмотрена в [Avrachenkov, Morozov, 2010; Avrachenkov et al., 2011]. Важно заметить, что в области нестационарности классические моменты регенерации через некоторое время после начала работы системы перестают происходить. Точнее это можно сформулировать следующим образом. Пусть выполнено условие нестационарности

$$(\lambda + \mu_0)P_{loss} > \mu_0. \quad (14)$$

(Напомним, что для рассматриваемой системы с пуассоновским входным потоком условие (1) является критерием стационарности.) Тогда, после конечного (с вероятностью 1) момента времени t_0 , орбита никогда не опустошается, и, с точки зрения оценки вероятности P_{loss} , исходная система ведет себя как система $\hat{\Sigma}$, на вход которой поступают два независимых потока с параметрами λ и μ_0 . Другими словами, квази-регенерации становятся классическими регенерациями, но для процесса $\{\nu_n\}$, рассматриваемого изолированно, а оценка $P_{orb}(n)$ сходится к P_{loss} с вероятностью 1. Заметим, что определенная проблема, связанная с частотой квази- и классических регенераций, может проявиться на границе области устойчивости (определяемой равенством в (1) и (14)) или вблизи нее. Можно ожидать, что чем больше разность между левой и правой частями неравенства (14) (т. е., чем *глубже* параметры λ , μ , μ_0 находятся *внутри области нестационарности*), тем быстрее (и устойчивее) должна быть сходимость $P_{orb}(n) \rightarrow P_{loss}$. Совершенно другую ситуацию следует ожидать в области стационарности системы. Именно, чем *глубже внутри области стационарности* находятся параметры λ , μ , μ_0 (т. е. чем больше разность правой и левой частей неравенства (1)), тем меньше должно быть значение предела $P_{orb}(n) \rightarrow P_{orb}$ по сравнению с P_{loss} .

Эти соображения подтверждаются приводимыми ниже результатами численного оценивания вероятности P_{orb} методом регенеративного (или квази-регенеративного) моделирования. Заметим, что детальное изложение регенеративного метода моделирования можно найти в [Bratley, Fox, 1987; Law, Kelton, 1991; Glynn, Iglehart, 1993].

ВЕРОЯТНОСТЬ БЛОКИРОВКИ: ПРИМЕНЕНИЕ ФОРМУЛЫ ЛИТТЛА

Обобщенная формула Литтла является ключевым элементом анализа при выводе соотношений для вероятности блокировки P_{orb} . Поэтому, следуя [Whitt, 1981], ниже приведена как сама обобщенная формула Литтла, так и пояснения к ее выводу. Кроме того, рассмотрены и исследованы две оценки вероятности блокировки: введенная ранее прямая (классическая) оценка $B_N(t)$ и (определенная ниже) альтернативная (непрямая) оценка $B_I(t)$, предложенная в [Spricant, Whitt, 1999].

Рассмотрим s -серверную систему с потерями общего вида $GI/G/s/k+s$ (здесь k -размер буфера для ожидания). Пусть $U(t)$ – число поступивших в систему заявок, а $D(t)$ – общее число обслуженных заявок, в интервале $[0, t]$. Пусть W_k – время, проведенное k -й заявкой в системе (для систем без буфера эта величина равна времени обслуживания S_k), и $\nu(t)$ – число заявок в системе в момент t (в случае одного сервера $\nu(t) = 0$ или 1). Обобщенная формула Литтла содержится в следующем утверждении из [Spricant, Whitt, 1999].

Теорема 1. Пусть с вероятностью 1 существуют пределы

$$t^{-1}U(t) \rightarrow \lambda, \quad t^{-1}D(t) \rightarrow \lambda, \quad t \rightarrow \infty, \quad (15)$$

где $0 < \lambda < \infty$. Пусть также

$$k^{-1} \sum_{j=1}^k W_j \rightarrow W, \quad k \rightarrow \infty. \quad (16)$$

Тогда с вероятностью 1

$$\frac{1}{t} \int_0^t \nu(s) ds \rightarrow L \quad \text{при } t \rightarrow \infty, \quad (17)$$

и кроме того,

$$L = \lambda W. \quad (18)$$

Для пояснения заметим, что интеграл в (17) представляет собой суммарное время, которое провели в системе $D(t)$ заявок (уже обслужившихся к моменту t) плюс время, проведенное в системе к моменту t остальными $A(t) - D(t)$ заявками. Также в доказательстве используются следующие неравенства

$$\sum_{j=1}^{D(t)} W_j \leq \int_0^t \nu(s) ds \leq \sum_{j=1}^{U(t)} W_j, \quad (19)$$

откуда утверждение теоремы следует после деления на t и перехода к пределу при $t \rightarrow \infty$.

Вернемся к системе Σ с повторными вызовами и без буфера, описанной в разделе 1. Рассмотрим обслуживающее устройство как изолированную систему, к которой и применим Теорему 1. Входной поток можно разделить на *эффективный* поток заявок, поступающих на обслуживание, и поток отказов, идущий на орбиту. Таким образом, эффективный поток есть число поступлений $U(t)$ на сервер в интервале $[0, t]$, $t \geq 0$. Так как все заявки эффективного потока получают обслуживание, то с вероятностью 1

$$\lim_{t \rightarrow \infty} \frac{U(t)}{t} = \lim_{t \rightarrow \infty} \frac{D(t)}{t} := \lambda_e. \quad (20)$$

(Как доказано в Теореме 2 ниже, данный предел, т. е. интенсивность λ_e , действительно существует). Таким образом, условия Теоремы 1 выполнены, причем время, проведенное заявкой в системе (на сервере), является ее временем обслуживания S . Заметим также, что приращения на циклах регенерации процесса накопления $\{R(t)\}$ в стационарном режиме образуют положительно возвратный регенерирующий процесс. Поэтому с вероятностью 1 существует упоминавшийся выше предел

$$\lim_{t \rightarrow \infty} \frac{R(t)}{t} := \tilde{\mu}_0, \quad (21)$$

который является интенсивностью потока неудачных попыток (т. е. потока, идущего на орбиту) и, одновременно, потока попыток (удачных и неудачных) попасть с орбиты на сервер. (Анализ процессов накопления можно найти в [Smith, 1955].)

Теперь сформулируем и докажем следующую теорему, обозначив $\tilde{\rho} = (\lambda + \mu_0)ES$.

Теорема 2. *В рассматриваемой системе Σ с повторными вызовами предел*

$$\lim_{t \rightarrow \infty} \frac{D(t)}{t} = \lambda_e \quad (22)$$

всегда существует, и выполнено следующее равенство:

$$\lambda_e ES = P_{busy}, \quad (23)$$

где P_b – стационарная вероятность занятости сервера. При этом

а) в стационарном режиме, т. е. при выполнении условия (1)

$$\lambda_e = \lambda, \quad P_{orb} = \frac{\tilde{\mu}_0}{\lambda + \tilde{\mu}_0}; \quad (24)$$

б) в нестационарном режиме, т. е. при выполнении условия (14)

$$\lambda_e = (\lambda + \mu_0)(1 - P_{orb})$$

и

$$P_{orb} = 1 - \frac{P_b}{\tilde{\rho}}. \quad (25)$$

Доказательство. Напомним, что $A(t)$ – общее число обращений к серверу (включая повторные попытки орбитальных заявок), $U(t)$ – число *успешных* попыток, а $R(t)$ – число уходов на орбиту (т. е. число неудачных попыток), все – в интервале $[0, t]$. Очевидно, что $A(t) = U(t) + R(t)$. Определим исходную последовательность (независимых, одинаково распределенных) времен обслуживания заявок $J = \{S_1, S_2, \dots\}$, из которой реальные времена обслуживания выбираются не в порядке поступления λ -заявок, а в порядке успешных попыток попасть на сервер. Таким образом, если j -я попытка оказывается не успешной, т. е. соответствующая ей заявка уходит на орбиту (возможно, не в первый раз), то величина S_j оказывается фиктивным временем обслуживания и выбрасывается из последовательности J . Это правило назначения приводит к тому, что последовательность реальных времен обслуживания, вообще говоря, не совпадает с исходной последовательностью J , что объясняет появление стохастических неравенств ниже (вместо неравенств с вероятностью 1). Пусть $\nu(t)$ – число заявок на сервере в момент t ($\nu(t) = 0$ или 1). Независимо от режима работы системы выполняются следующие неравенства:

$$\sum_{i=1}^{D(t)} S_i \leq_{st} \int_0^t \nu(u) du \leq \sum_{j=1}^{A(t)} I_j S_j, \quad (26)$$

где индикатор $I_j = 1$, если j -я попытка успешна, $j \geq 1$. (Отметим, неравенства (19) также можно трактовать как стохастические.) Заметим, что оба неравенства переходят в равенство, если $\nu(t) = 0$. Иначе значение суммы слева меньше значения интеграла на уже *прошедшую* часть времени обслуживания заявки, находящейся на сервере в момент t , а сумма справа больше интеграла на *оставшееся* в момент t время обслуживания этой заявки. Поскольку в системе лишь один сервер, то процесс $\nu(t)$ фактически является индикатором занятости, т. е. $\nu(t) = I(\nu(t) > 0)$. Следовательно, предполагая, что существует слабый предел $\nu(t) \Rightarrow \nu$, получим

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \nu(u) du = E\nu = P(\nu > 0) := P_{busy}. \quad (27)$$

Далее доказательство проводится отдельно для стационарного и нестационарного режимов.

Стационарный режим. Пусть выполнено условие (1). Перепишем (26), усилив правое неравенство,

$$\sum_{i=1}^{D(t)} S_i \leq_{st} \int_0^t \nu(u) du \leq \sum_{i=1}^{A(t)} S_i. \quad (28)$$

Ввиду стационарности все λ -заявки, поступающие в систему, в конце концов получают обслуживание (сразу или после возвращения с орбиты). Таким образом, с вероятностью 1,

$$A(t) - D(t) \leq N(t) + 1 = o(t). \quad (29)$$

(Условие $N(t) = o(t)$ следует из положительной возвратности процесса величины орбиты $\{N(t)\}$, см. [Smith, 1955; Morozov, 2004]). Отсюда следует что

$$\lim_{t \rightarrow \infty} \frac{D(t)}{t} = \lim_{t \rightarrow \infty} \frac{U(t)}{t} \rightarrow \lambda, \quad t \rightarrow \infty, \quad (30)$$

и $\lambda_e = \lambda$. Поскольку с вероятностью 1

$$\frac{1}{t} \sum_{i=1}^{D(t)} S_i = \frac{D(t)}{t} \frac{\sum_{i=1}^{D(t)} S_i}{D(t)} \rightarrow \lambda ES,$$

$$\frac{1}{t} \sum_{i=1}^{U(t)} S_i \rightarrow \lambda ES, \quad t \rightarrow \infty, \quad (31)$$

то первое утверждение Теоремы 2 доказано. Далее, очевидно, что

$$\sum_{i=1}^{U(t)} S_i \stackrel{st}{=} \sum_{j=1}^{U(t)+R(t)} I_j S_j. \quad (32)$$

(Напомним, что реализованные интервалы обслуживания S_j при $I_j = 1$ в (32) справа, вообще говоря, не совпадают с S_i слева.) Кроме того, $S_j I_j \Rightarrow S I_\infty$, при $j \rightarrow \infty$, где S – типичное время обслуживания и $I_j \Rightarrow I_\infty$. Заметим, что событие $I_\infty = 1$ означает поступление заявки на сервер в стационарном режиме. Поэтому $P(I_\infty = 1) = 1 - P(I = 1) = 1 - P_{orb}$ (см. (9)). Поскольку случайные величины S_j и I_j независимы для каждого j , то получаем $E(S I_\infty) = ES(1 - P_{orb})$. Таким образом, учитывая также (21), можно было бы ожидать, что

$$\begin{aligned} \frac{1}{t} \sum_{j=1}^{U(t)+R(t)} I_j S_j &\rightarrow (\lambda + \tilde{\mu}_0) E(S I_\infty) \\ &= (\lambda + \tilde{\mu}_0) ES(1 - P_{orb}), \quad t \rightarrow \infty. \end{aligned} \quad (33)$$

Однако случайные величины $\{S_j I_j, j \geq 1\}$ зависимы. Поэтому для корректного применения усиленного закона больших чисел будем использовать другое представление суммы $\sum_{j=1}^{U(t)+R(t)} I_j S_j$, опираясь на равенство (32). Именно, запишем (для $t > 0$)

$$\begin{aligned} &\frac{\sum_{i=1}^{U(t)} S_i}{t} \\ &= \frac{\sum_{i=1}^{U(t)} S_i}{U(t)} \cdot \frac{U(t)}{R(t) + U(t)} \cdot \frac{U(t) + R(t)}{t}. \end{aligned} \quad (34)$$

Очевидно, что

$$\begin{aligned} \frac{U(t) + R(t)}{t} &\rightarrow \lambda + \tilde{\mu}_0, \\ \frac{U(t)}{U(t) + R(t)} &\rightarrow 1 - P_{orb}, \quad t \rightarrow \infty, \end{aligned} \quad (35)$$

Отсюда, а также из (32) следует утверждение (33). Приравняв пределы (31) и (33), получаем

$$\lambda ES = (\lambda + \tilde{\mu}_0) ES(1 - P_{orb}), \quad (36)$$

откуда следует формула (24) для вероятности ухода на орбиту в стационарном режиме.

Нестационарный режим. Пусть выполняется условие (14). Так как тогда величина орбиты $N(t) \rightarrow \infty$ с вероятностью 1, то $R(t)/t \rightarrow \mu_0$, и легко видеть, что теперь соотношение (33) выполнено с заменой $\tilde{\mu}_0$ на μ_0 . Далее, как и выше, получим

$$P_{busy} = (\lambda + \mu_0) ES(1 - P_{orb}). \quad (37)$$

Отсюда и из (26) следует $\lambda_e = (\lambda + \mu_0)(1 - P_{orb})$ и выражение (25). Теорема доказана. \square

Заметим, что в случае пуассоновского λ -потока в нестационарном режиме суммарный входной поток в систему также будет пуассоновским с параметром $\lambda + \mu_0$. Тогда, опираясь на свойство PASTA (по которому $P_{orb} = P_{busy}$), для стационарной вероятности занятости из ((37)) получаем хорошо известное выражение

$$P_{busy} = \frac{\tilde{\rho}}{\tilde{\rho} + 1}. \quad (38)$$

(Свойство PASTA рассматривается, например, в [Asmussen, 2002]).

АЛЬТЕРНАТИВНЫЕ ОЦЕНКИ ВЕРОЯТНОСТИ БЛОКИРОВКИ. МЕТОД УМЕНЬШЕНИЯ ДИСПЕРСИИ

Этот раздел в значительной мере опирается на важную работу [Spricant, Whitt, 1999]. Напомним, что *прямая оценка* вероятности блокировки имеет следующий вид:

$$B_N(t) = \frac{R(t)}{A(t)}. \quad (39)$$

Поскольку $B_N(t) \rightarrow P_{orb}$ при $t \rightarrow \infty$, то эта оценка состоятельная. Однако она представляет собой отношение двух величин, а такие оценки, вообще говоря, являются смещенными, т. е. $EB_N(t) \neq P_{orb}$. Но если процессы $\{R(t)\}$ и $\{A(t)\}$ имеют стационарные приращения, то, очевидно, $ER(t)/EA(t) = P_{orb}$ для любого t .

Близкая к прямой *простая оценка*, предложенная в [Spricant, Whitt, 1999], опирается на асимптотику процессов восстановления и принимает для рассматриваемой здесь системы вид

$$B_S(t) = \frac{R(t)}{(\lambda + \tilde{\mu}_0)t}. \quad (40)$$

(В нестационарном режиме $\tilde{\mu}_0$ заменяется на μ_0 .) Оценка $B_S(t)$ может быть предпочтительнее прямой оценки в системе с потерями, где параметр μ_0 задан. Однако присутствие параметра $\tilde{\mu}_0$, требующего оценивания (наряду с $R(t)$), лишает ее данного преимущества в системе с повторными вызовами.

Гораздо более интересная *непрямая оценка* вероятности P_{orb} , предложенная в [Spricant, Whitt, 1999], имеет вид

$$B_I(t) = 1 - \frac{P_{busy}(t)}{\tilde{\rho}}, \quad (41)$$

где

$$P_{busy}(t) := \frac{1}{t} \int_0^t \nu(u) du, \quad (42)$$

есть доля времени занятости сервера в интервале $[0, t]$.

В работе [Spricant, Whitt, 1999] показано, что оценка (41) является более (менее) эффективной нежели оценка $B_N(t)$ в случае большой (малой) нагрузки. (Большая эффективность означает меньшую дисперсию.) Это приводит к рассмотрению следующей *комбинированной оценки*

$$B_C(t) = pB_N(t) + (1-p)B_I(t), \quad (43)$$

где параметр p выбирается таким образом, чтобы минимизировать дисперсию. Идея комбинирования основана на том, что с ростом

$P_{busy}(t)$ оценка $B_I(t)$ убывает (это видно из (41)), в то время как оценка $B_N(t)$ должна расти, так как увеличение доли времени занятости сервера в системе с потерями увеличивает потери. Таким образом, оценки $B_I(t)$ и $B_N(t)$ должны быть отрицательно коррелированы.

Для системы с повторными вызовами было бы естественно использовать аналог этой оценки,

$$\tilde{B}_I(t) = 1 - \frac{P_{busy}(t)}{(\lambda + \mu_0(t))ES}, \quad (44)$$

где параметр μ_0 заменен на оцениваемый параметр $\mu_0(t) = R(t)/t$ – локальную интенсивность потока попыток, идущих с орбиты (и на орбиту) в интервале $[0, t]$. Заметим, что в стационарном режиме (см. (24))

$$\tilde{B}_I(t) \rightarrow 1 - \frac{\rho}{\tilde{\rho}} = \frac{\tilde{\mu}_0}{\lambda + \tilde{\mu}_0} = P_{orb}, \quad t \rightarrow \infty. \quad (45)$$

Преимущество комбинированной оценки (43) в случае большой нагрузки в системе с потерями достигается за счет непрямой оценки (41). В рассматриваемой модели без потерь большая нагрузка, как правило, соответствует нестационарному режиму, в котором, как показано в разделе 5, комбинированная оценка действительно является более эффективной.

Ниже приведено краткое обоснование этого утверждения, следуя [Spricant, Whitt, 1999]. Для анализа эффективности оценок удобно представить отношение их дисперсий в виде

$$\frac{DB_C(t)}{DB_N(t)} = \frac{DB_C(t)}{DB_I(t)} \frac{DB_I(t)}{DB_N(t)}. \quad (46)$$

В [Spricant, Whitt, 1999] показано, что отношение дисперсий $DB_C(t)/DB_I(t)$ зависит от отношения $r^2 := DB_I(t)/DB_N(t)$ и корреляции $\rho := Corr(B_I(t), B_N(t))$. (Для удобства мы иногда опускаем зависимость от t .) Можно показать, что

$$\begin{aligned} V(p) &:= DB_C(t) \\ &= DB_I(t) \left(\frac{p^2}{r^2} + (1-p)^2 \right. \\ &\quad \left. + 2p(1-p) \frac{\rho}{r} \right). \end{aligned} \quad (47)$$

Несложно также показать, что вторая производная $V''(p) > 0$ при всех p , а значит минимум достигается при $p = p^*$ таком, что $V'(p^*) = 0$, причем

$$p^* = \frac{r(r-\rho)}{1+r^2-2r\rho}. \quad (48)$$

При этом $0 < p^* < 1$, так как мы предполагаем неположительную корреляцию, $\rho \leq 0$. Окончательно минимальная дисперсия имеет вид:

$$V(p^*) = \frac{DB_I(t)(1 - \rho^2)}{1 + r^2 - 2r\rho}. \quad (49)$$

Поэтому коэффициент уменьшения дисперсии

$$R(p) := \frac{V(p)}{DB_I(t)} \quad (50)$$

принимает при $p = p^*$ значение

$$R(p^*) = \frac{1 - \rho^2}{1 + r^2 - 2r\rho}. \quad (51)$$

При $r \leq 1$ получим следующую нижнюю оценку выигрыша за счет использования комбинированной оценки

$$R(p^*) \geq \frac{1 - \rho^2}{4}, \quad (52)$$

а при равных дисперсиях $DB_I = DB_N$, т. е. при $r = 1$, получим $R(p^*) = (1 + \rho)/2$. Эта величина принимает большие значения лишь при $\rho \approx -1$. В этом случае нижняя граница (52) может быть аппроксимирована как

$$\frac{1 - \rho^2}{4} \approx \frac{1 + \rho}{2}. \quad (53)$$

Таким образом, при равных дисперсиях прямой и не прямой оценок и сильной отрицательной корреляции получаем $R(p^*) \approx 0$, т. е. очень значительное уменьшение дисперсии. При малом r (т. е. при $DB_I \ll DB_N$) из (51) следует

$$\lim_{r \rightarrow 0} R(p^*) = 1 - \rho^2, \quad (54)$$

независимо от знака ρ , что дает большой выигрыш при $|\rho| \approx 1$. Предел (54) отличается от нижней границы для r в (52), достигаемой при $r = 1$ и $\rho \approx -1$, в 4 раза. Согласно [Spricant, Whitt, 1999], в случае малых значений r уменьшение дисперсии в (46) можно приближенно выразить через произведение $1 - \rho^2$ и r . Поэтому, как было сказано выше, комбинированная оценка особенно эффективна, когда $\rho \approx -1$.

Как показано в следующем разделе, приведенные выше результаты для системы с потерями хорошо согласуются с результатами численного моделирования исходной системы с повторными вызовами без потерь. Это вновь подтверждает определенную близость двух данных систем, что уже было показано при анализа стационарности.

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ СИСТЕМЫ ТИПА $M/G/1$

В данном разделе представлены результаты оценивания вероятности блокировки P_{orb} в исходной системе Σ типа $M/G/1/1$ с повторными вызовами с использованием регенеративного метода. Имитационное моделирование проведено для 20 000 заявок. Заметим, что оценки построены на основе соответствующего числа регенераций (или квази-регенераций), полученных в результате процедуры моделирования. Однако для удобства, зависимость оценок от числа заявок (или регенераций) опущена. Положим $\lambda = 1$ и обозначим через

$$\Gamma = \mu - 1 - \frac{1}{\mu_0}, \quad (55)$$

разность правой и левой частей в условии (6). Эта величина характеризует меру стационарности системы, поскольку $\Gamma > 0$ ($\Gamma < 0$) в области стационарности (нестационарности).

Зоны стационарности/нестационарности данной системы представлены на рис. 1, где граница зон определяется уравнением $\Gamma = 0$.

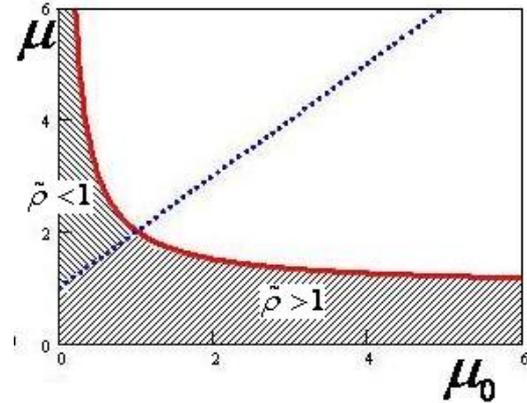


Рис. 1. Зоны стационарности/нестационарности системы типа $M/M/1/1$ при $\lambda = 1$

Обсудим результаты, представленные в табл. 1. В области стационарности (т. е. при $\Gamma > 0$), значение прямой оценки $B_N \equiv B_N(t)$ отличается от значения вероятности потери P_{loss} , получаемого по формуле Эрланга. При этом, чем больше значение Γ (т. е. чем «глубже» в области стационарности выбраны параметры μ, μ_0), тем больше значение разности $P_{loss} - B_N > 0$. Как уже отмечалось в разделе 1, это вызвано убыванием интенсивности потока заявок, идущих на орбиту, что, в свою очередь, вызывает убывание вероятности

Таблица 1. Оценивание вероятности P_{orb} в системе $M/M/1/1$ при $\lambda = 1$

№	μ_0	$\tilde{\mu}_0$	μ	$\tilde{\rho}$	Γ	P_{loss}	B_N	B_I	B_C	r^2	ρ	$\frac{DB_C}{DB_N}$
1	10,0	0,11	15,0	0,07	13,90	0,423	0,100	0,007	-0,035	1,562	0,907	1,467
2	5,0	0,44	5,0	0,29	3,80	0,545	0,296	0,081	0,112	0,879	0,913	0,876
3	17,0	1,36	4,0	0,59	2,94	0,818	0,595	0,254	0,054	0,531	0,908	0,449
4	2,0	1,99	1,0	2,99	-0,50	0,750	0,748	0,751	0,750	0,071	-0,556	0,036
5	6,0	5,91	0,3	23,04	-0,87	0,959	0,958	0,959	0,959	0,001	-0,702	0,000
6	0,1	0,10	2,0	0,55	-9,00	0,355	0,355	0,354	0,354	3,309	-0,358	1,702
7	10,0	5,36	1,5	4,24	0,40	0,880	0,840	0,781	0,735	0,325	0,932	0,163

P_{orb} и возрастание разности $P_{loss} - P_{orb}$. Отметим, что прямая и непрямо́е оценки в стационарном режиме положительно коррелируют, $\rho > 0$. Кроме того, заметим, что в стационарном режиме загрузка может быть низкой даже в случае достаточно больших значений μ_0 , поскольку реальная интенсивность $\tilde{\mu}_0$ может быть достаточно мала (см. табл. 1, строка 3). Отметим, что в «глубине» области стационарности $r^2 > 1$, т. е. дисперсия непрямо́е оценки больше дисперсии прямо́е оценки (см. табл. 1, строка 1). Поэтому для оценивания P_{orb} следует использовать лишь прямо́е оценку B_N . (Это полностью соответствует случаю малой нагрузки в системе с потерями [Spricant, Whitt, 1999]). В нестационарном случае прямо́е и непрямо́е оценки отрицательно коррелируют, $\rho < 0$. При этом максимальное снижение дисперсии достигается при больших значениях $\tilde{\rho}$, т. е. при большой загрузке, что вновь соответствует результатам работы [Spricant, Whitt, 1999] для системы с потерями. Заметим, что при малом r^2 и $\rho \approx -1$, дисперсия комбинированной оценки снижается практически до нуля (табл. 1, строка 5), что соответствует выводу из формулы (54). Однако сильная нестационарность (т. е. большое отрицательное значение Γ) не гарантирует большую загрузку системы (табл. 1, строка 6). В частности, при выполнении двойного неравенства

$$1 + \mu_0 < \mu < 1 + \frac{1}{\mu_0}, \quad (56)$$

система нестационарна при малой загрузке, т. е. при $\tilde{\rho} < 1$. В этом случае дисперсия непрямо́е оценки превышает дисперсию прямо́е оценки B_N (строка 6) и комбинирование не имеет смысла.

Таким образом, в системе типа $M/M/1/1$ комбинированная оценка B_C эффективна (и близка к прямо́е) только при большом значении $\tilde{\rho}$ в нестационарном режиме. В противном случае более эффективна прямо́е оценка B_N .

Области значений параметров μ и μ_0 , определяющие малую/большую загрузку, представлены на рис. 2.

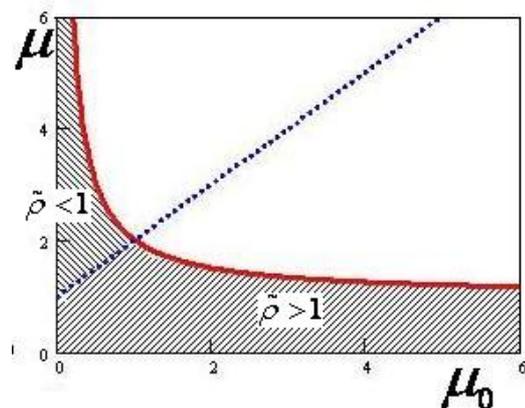


Рис. 2. Области малой ($\tilde{\rho} < 1$) и большой ($\tilde{\rho} > 1$) загрузки в зоне нестационарности системы типа $M/M/1/1$ при $\lambda = 1$

Обратимся к анализу результатов оценивания в системе типа $M/Pareto/1/1$, представленных в табл. 2. В данном случае мера стационарности равна

$$\Gamma = 1 - \lambda - \frac{1}{\alpha} - \frac{\lambda^2}{\mu_0}. \quad (57)$$

Эта величина (при фиксированном λ) увеличивается с ростом μ_0 и α . Коэффициент загрузки в данной системе равен

$$\tilde{\rho} = (\lambda + \tilde{\mu}_0) \frac{\alpha}{\alpha - 1} \quad (58)$$

и возрастает (при фиксированном λ) с ростом $\tilde{\mu}_0$.

Как видно из табл. 2 (см. табл. 2, строки 1–3), изменения параметра μ_0 (в области больших значений) мало влияют на величину $\tilde{\mu}_0$. (Это связано с тем, что в глубине области стационарности орбита имеет большой запас мощности равный $\mu_0 - \tilde{\mu}_0$.) Таким обра-

Таблица 2. Оценивание вероятности P_{orb} в системе $M/Pareto/1/1$ при $\lambda = 0,5$

№	μ_0	$\tilde{\mu}_0$	α	$\tilde{\rho}$	Γ	P_{loss}	B_N	B_I	B_C	r^2	ρ	$\frac{DB_C}{DB_N}$
1	10,00	3,01	10	3,90	0,37	0,921	0,856	0,770	0,695	0,302	0,953	0,108
2	5,00	2,22	5	3,40	0,25	0,873	0,814	0,735	0,663	0,320	0,953	0,122
3	2,00	1,66	3	3,23	0,04	0,789	0,770	0,713	0,657	0,332	0,958	0,119
4	1,00	1,00	3	2,25	-0,08	0,692	0,694	0,694	0,694	0,099	-0,085	0,085
5	0,05	0,05	3	0,83	-4,83	0,452	0,452	0,452	0,452	0,938	-0,058	0,456
6	0,05	0,05	10	0,61	-4,60	0,379	0,380	0,378	0,379	1,892	-0,008	1,228

зом, значение коэффициента загрузки $\tilde{\rho}$ остается стабильным и достаточно большим. Это, в свою очередь, влечет эффективность оценки B_C (см. последний столбец), что вновь согласуется с результатами [Spricant, Whitt, 1999]. По аналогии с системой $M/M/1/1$, можно получить следующее двойное неравенство, соответствующее зоне нестационарности при малой загрузке ($\tilde{\rho} < 1$),

$$1 - \frac{\lambda}{\mu_0}(\lambda + \mu_0) < \frac{1}{\alpha} < 1 - (\lambda + \mu_0). \quad (59)$$

Следовательно, в зоне нестационарности система имеет малую загрузку только при $\lambda < \mu_0$. В этом случае эффективность оценки B_C уменьшается (см. строки 4-5), что вновь согласуется с результатами для системы с потерями.

Области значения параметров α и μ_0 , определяющие малую/большую загрузку в зоне нестационарности системы $M/Pareto/1/1$, представлены на рис. 3.

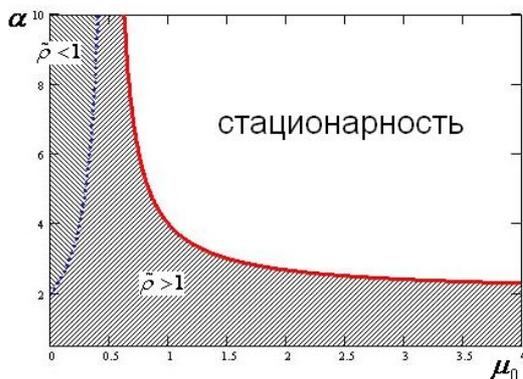


Рис. 3. Области малой и большой загрузки в зоне нестационарности системы типа $M/Pareto/1/1$ при $\lambda = 0,5$

Заметим, что при определенной конфигурации параметров система находится в зоне стационарности при большой загрузке ($\tilde{\rho} > 1$) (табл. 1, строка 7; табл. 2 строки 1–3). В этом

случае оценка B_C обладает меньшей дисперсией, однако ее значение систематически меньше значения прямой оценки B_N . Это говорит о том, что в зоне стационарности использовать комбинированную оценку нежелательно. Причина такой несогласованности оценок пока остается неясной и требует дальнейшего исследования. Скорее всего, это связано с различием между исходной системой с повторными вызовами и системой с потерями, для которой изначально были построены данные оценки.

Подчеркнем, что в обоих исследованных системах малая загрузка сочетается с нестационарностью, и это отражает их существенное отличие от системы с потерями.

ЗАКЛЮЧЕНИЕ

В статье рассмотрена проблема оценивания вероятности блокировки в системе с повторными вызовами и постоянной скоростью возвращения заявок с орбиты на сервер. Приведено условие стационарности такой системы. Обсуждается обобщенная формула Литтла, на основе которой найдены соотношения для вероятности блокировки как в стационарном, так и в нестационарном режимах. Рассмотрены две основные альтернативные статистические оценки вероятности блокировки, имеющие отрицательную корреляцию в нестационарном режиме. Это позволяет построить комбинированную оценку, имеющую существенно меньшую дисперсию. Проведено статистическое оценивание вероятности блокировки в стационарном режиме с использованием классической регенерации, а также в нестационарном режиме с использованием newline квази-регенерации. Представленные численные расчеты хорошо согласуются с теоретическими условиями стационарности, а также с результатами работы [Spricant, Whitt, 1999] (для системы с потерями) об эффективности комбинированной оценки при большой загрузке.

Работа выполнена при поддержке РФФИ, проект 10-07-00017.

ЛИТЕРАТУРА

Asmussen S. Applied Probability and Queues // Springer, 2002. 440 p.

Avrachenkov K., Goricheva R. S., Morozov E. V. Verification of stability region of a retrial queueing system by regenerative method // Proceedings of the International Conference «Modern Probabilistic Methods for Analysis and Optimization of Information and Telecommunication Networks». Minsk, 2011. P. 22–28.

Avrachenkov K., Morozov E. V. Stability analysis of GI/G/c/K Retrial Queue with Constant Retrial Rate // INRIA (Sophia Antipolis), Research Report. 2010. N. 7335.

Avrachenkov K., Yechiali U. Retrial networks with finite buffers and their application to Internet data traffic // Probability in the Engineering and Informational Sciences. 2008. Vol. 22. P. 519–536.

Bratley P., Fox B. L. A guide to simulation // New York: Springer-Verlag, 1987. 424 p.

Choi B. D., Rhee K. H., Park K. K. The M/G/1 retrial queue with retrial rate control policy // Probability in the Engineering and Informational Sciences, 1993. Vol. 7. P. 29–46.

Fayolle G. A simple telephone exchange with delayed feedback / In O. J. Boxma, J. W. Cohen and H. C. Tijms (ed.) // Teletraffic Analysis and

Computer Performance Evaluation. 1986. Vol. 7. P. 245–253.

Glynn P. W., Iglehart D. L. Conditions for the applicability of the regenerative method // Management Science. 1993. Vol. 39. P. 1108–1111.

Law A. M., Kelton W. D. Simulation modeling and analysis. New York: McGraw-Hill, 2000. 560 p.

Morozov E. V. Weak regeneration in modeling of queueing processes // Queueing Systems. 2004. Vol. 46. P. 295–315.

Smith W. L. Regenerative stochastic processes // Proc. Roy. Soc. 1955. Ser. A 232. P. 6–31.

Sonderman D. Comparing multi-server queues with finite waiting rooms, I: Same number of servers // Advances in Applied Probability. 1979. Vol. 11. P. 439–447.

Sonderman D. Comparing multi-server queues with finite waiting rooms, II: Different number of servers // Advances in Applied Probability. 1981. Vol. 11. P. 448–455.

Spricant R., Whitt W. Variance reduction in simulations of loss models // Operation research. 1999. Vol. 47, N. 4. P. 509–523.

Whitt W. Comparing counting processes and queues // Advances in Applied Probability. 1981. Vol. 13. P. 207–220.

Wolff R. W. Stochastic Modeling and the Theory of Queues // Prentice-Hall, 1989. 560 p.

СВЕДЕНИЯ ОБ АВТОРАХ:

Морозов Евсей Викторович

ведущий научный сотрудник, д. ф.-м. н.
Институт прикладных математических исследований
КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика
Карелия, Россия, 185910
эл. почта: emorozov@karelia.ru
тел.: (8142) 763370

Morozov, Evsey

Institute of Applied Mathematical Research, Karelian
Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,
Russia
e-mail: emorozov@karelia.ru
tel.: (8142) 763370

Некрасова Руслана Сергеевна

аспирантка
Институт прикладных математических исследований
КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика
Карелия, Россия, 185910
эл. почта: rulsana.nekrasova@mail.ru
тел.: (8142) 763370

Nekrasova, Ruslana

Institute of Applied Mathematical Research, Karelian
Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,
Russia
e-mail: rulsana.nekrasova@mail.ru
tel.: (8142) 763370