Некоторые модели многопроцессорных систем обслуживания с тяжелыми хвостами *

Е. В. Морозов, А. С. Румянцев

Учреждение Российской академии наук
Институт прикладных математических исследований
Карельского научного центра РАН

Работа посвящена изучению свойств процесса загрузки в многопроцессорных системах обслуживания с распределениями, обладающими тяжелым хвостом. Такие распределения могут приводить к появлению долговременной зависимости, что существенно затрудняет построение надежных статистических оценок. Построена модель вычислительного кластера, являющаяся модификацией классической многосерверной системы обслуживания. Приведены результаты статистических экспериментов.

1. Введение

Теоретический интерес последних двух десятилетий к моделям систем обслуживания, описываемых распределениями с так называемыми тяжелыми хвостами, обусловлен обнаружением этого эффекта в экспериментах по изучению компьютерных систем и сетей (см., напр., [4]). Более медленное, чем экспоненциальное, убывание хвостовой вероятности $\overline{F}(x) := P(X > x)$ для заданной случайной величины X (*«тяжелый хвост распределения»*), т.е.

$$\lim_{x\to\infty} e^{-ax}\overline{F}(x) = \infty, \quad \forall a > 0 \tag{1}$$

ограничило ранее повсеместное использование экспоненциальных распределений для моделирования процессов в современных компьютерных системах. Эмпирически подтверждено присутствие тяжелых хвостов в распределениях размеров файлов на жестких дисках, времен вычисления задач на процессоре, времен передачи файлов между сервером и клиентом в высокоскоростных сетях и т.д. [4]. Наиболее важными для аналитического исследования таких процессов подклассами распределений с тяжелыми хвостами являются *субэкспоненциальные распределения*, которые определяются как

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} = \lim_{x\to\infty} \frac{\overline{F} * \overline{F}(x)}{\overline{F}(x)} = 2, \quad x \geqslant 0, \tag{2}$$

где $\overline{F} * \overline{F}(x) = \int_0^\infty \overline{F}(x-y)\overline{F}(y)\,dy$ означает свертку распределений (для неотрицательных с.в.), а также *правильно меняющиеся распределения*, характеризуемые соотношением

$$\lim_{x\to\infty} \frac{\overline{F}(xt)}{\overline{F}(x)} = t^{-\alpha}, \quad \alpha \geqslant 0. \tag{3}$$

(При $\alpha = 0$ функция F называется медленно меняющейся). Наиболее часто используемым является распределение Парето, имеющее (стандартный) вид

$$\overline{F}(x) = x^{-\alpha},\ x \geqslant 1, \quad \overline{F}(x) = 1,\ x \leqslant 1. \tag{4}$$

Отметим, что хороший обзор субэкспоненциальных распределений представлен в работе [7], свойства регулярно меняющихся функций подробно обсуждаются в [12], а работа [8] посвящена применению распределений с тяжелыми хвостами к сетевым моделям и системам обслуживания.

*Работа поддержана РФФИ, проект 10-07-00017.

Однако, обнаружение тяжелых хвостов в компьютерных системах поставило ряд проблем, недооценивание которых приводит к серьезным последствиям как для компаний-производителей сетевого оборудования, так и для потребителей. (Частично эти проблемы обсуждаются в [16].) Одна из проблем состоит в том, что многие алгоритмы управления сетевыми устройствами и процессами в вычислительных системах строились с учетом предположения об экспоненциальности соответствующих распределений [9].

Присутствие тяжелого хвоста на уровне одного клиента системы или сети (состоящей из нескольких обслуживающих устройств) в результате суперпозиции большого числа независимых источников приводит к тому, что сетевой процесс $\{X(t), t > 0\}$ ведет себя не регулярно [10]. Точнее говоря, имеет место следующее равенство по распределению

$$C^{-H}\{X(Ct), t > 0\} =_d \{X(t), t > 0\}, \quad H \in (0.5, 1]$$

для шкалированного процесса $C^{-H}\{X(Ct), t > 0\}$. Такой процесс называется *самоподобным*, а постоянная H называется параметром Херста.

В таком процессе, в частности, возникают длительные периоды пребывания процесса $\{X(t)\}$ выше или ниже своего среднего значения, которые можно трактовать или как отсутствие стационарности, или как *долговременную зависимость* (долгую память) («эффект Иосифа» [18]). Аналитически долгая память выражается в расходимости ряда автокорреляций процесса, что (в случае дискретного времени) выражается как

$$\lim_{t\to\infty} \sum_{k=1}^{t} corr(X_0, X_k) = \infty. \tag{5}$$

Расходимость ряда (5) приводит к сложностям, связанным с надежностью статистических оценок параметров системы, выражающих качество обслуживания (QoS), например задержки в системе [1]. В этой связи отметим классическую работу [14], посвященную самоподобию и долговременной зависимости сетевого трафика.

Практический интерес для исследователя часто представляет вероятность большого уклонения (например, вероятность переполнения большого буфера, вероятность большого времени ожидания в системе и т.д.), а также моментные свойства сетевых процессов, в частности, моментный индекс, т.е. максимальный конечный момент. Особенно важным является величина дисперсии длины (передачи) сообщения. Для пояснения заметим, что дисперсия длины сообщений, варьирующихся в современных сетях от коротких электронных сообщений до видеоконференций, часто при моделировании может быть принята равной бесконечности. Заметим, что свойства корреляции изучались при исследовании эффекта долгой памяти [5], а зависимость моментных свойств сетевых процессов от дисциплины обслуживания рассматривается в [3]. Однако такого рода результаты известны для достаточно узкого класса односерверных систем обслуживания, и в гораздо меньшей степени для многосерверных систем. Особенно сложен анализ в случае, когда заявки в систему обслуживания поступают в виде пачек/групп (см., напр., работы [13, 15, 24]), или в виде регенеративного процесса [11]. Вычислительные кластеры и Грид представляют в этом смысле особую сложность, т.к. обладают вышеперечисленными свойствами, такими как многопроцессорность, сложные дисциплины обслуживания потока заявок, возможность поступления групп, большие времена обслуживания на процессорах (тяжелые хвосты).

2. Модели многосерверных систем обслуживания

В этом разделе дан обзор ряда важных известных результатов для многосерверных систем обслуживания, в т.ч. с бесконечным числом серверов. Эти результаты в основном опираются на классические результаты для односерверных систем, которые также приводятся для сравнения.

2.1. Система G/G/s

Пусть в систему обслуживания, состоящую из идентичных s серверов (процессоров), в моменты $\{t_i, i \geqslant 1\}$, образующие процесс восстановления, поступают заявки с независимыми, одинаково распределенными временами обслуживания $\{S_i\}$. Обозначим интервалы между приходами заявок $T_i := t_{i+1} - t_i$, $i \geq 1$. Введем функции распределения интервала $A(x) = P(T \leqslant x)$, и времени обслуживания $B(x) = P(S \leqslant x)$ (индекс опущен, когда рассматривается типичный представитель последовательности). Заявка ожидает время $D_i \geq 0$, в очереди, формируемой в порядке поступления (FIFO), пока не освободится наименее загруженный сервер, на который она поступает на обслуживание. Можно определить вектор загрузки в системе W_i, состоящий из оставшегося времени обслуживания на каждом процессоре в момент прихода заявки i (вектор загрузки Кифера-Вольфовица [20]):

$$W_{i+1} = R\left(W_i(1) + S_i - T_i, W_i(2) - T_i, \ldots, W_i(s) - T_i\right)^+, \quad i \geqslant 1, \tag{6}$$

компоненты которого представляют незавершенные нагрузки на серверах, отсортированные в возрастающем порядке, $W_i(1) \leqslant \ldots \leqslant W_i(s)$. (Оператор $R(\cdot)$ располагает компоненты вектора в возрастающем порядке, $(x)^+ = \max(0, x)$.) Таким образом, время ожидания i-й заявки является первой компонентой вектора загрузки, $D_i = W_i(1)$. При условии $\rho := ES/ET < s$ имеет место слабая сходимость вектора загрузки к стационарному значению, $W_i \Longrightarrow W := (W(1), \ldots, W(s))$. Обозначим $U_i = W_i(2) - W_i(1)$ — время, оставшееся на втором наименее загруженном сервере в момент поступления заявка i на обслуживание. Обозначим $P_i = \min(U_i, S_i)$, тогда время ожидания D_i удовлетворяет рекурсии [20]:

$$D_{i+1} = (D_i + P_i - T_i)^+, \quad i \geqslant 1. \tag{7}$$

Заметим, что при $s = 1$ формула (7) определяет классическую рекурсию Линдли

$$D_{i+1} = (D_i + S_i - T_i)^+, \quad i \geqslant 1.$$

2.1.1. Моменты времени ожидания

Значительное развитие классических результатов о конечности моментов стационарного времени ожидания D в системе G/G/s (с дисциплиной FIFO) достигнуто в работе [23]. Именно, пусть для некоторого натурального k выполнено условие $k < \rho < k + 1 \leqslant s$. Тогда

$$E\left(S^{1+\frac{\alpha}{s-k}}\right) < \infty \quad \text{влечет} \quad E(D^\alpha) < \infty. \tag{8}$$

При дополнительных ограничениях на распределение B это условие является также необходимым.

Таким образом, ввиду (8) имеет место зависимость моментных свойств процесса загрузки от коэффициента загрузки ρ. Эта зависимость имеет ступенчатый характер, в отличие от классического результата для односерверной системы, где, при условии $ET < \infty$,

$$ES^{\alpha+1} < \infty \quad \text{тогда и только тогда, когда} \quad ED^\alpha < \infty. \tag{9}$$

Частный случай (8) при загрузке $\rho < 1$ в системе G/G/s приведен в работе [19]. Именно, для любого $2 \leqslant j \leqslant s$ в системе G/G/j имеет место следующее условие конечности момента порядка j/s времени ожидания $D(j)$:

$$ES^{1+s^{-1}} < \infty \quad \text{влечет } E\left[D(j)\right]^{j/s} < \infty. \tag{10}$$

В частности, для $j = s$

$$ES^{(s+1)/s} < \infty \quad \text{влечет} \quad ED < \infty. \tag{11}$$

В предельном случае при $s \to \infty$ получается классическое условие $ES < \infty$ конечности среднего времени пребывания в системе G/G/∞. В работе [21] приведена также верхняя граница для стационарной средней задержки в системе G/G/s.

Одним из вариантов практического применения формулы (8) может быть обоснование выбора между s «медленными» серверами (работающими со скоростью $1/s$) и одним «быстрым» (имеющим скорость 1). Действительно, если выбрать $s > 1/(1-\rho)$, то при одном и том же входном потоке задержки в s-серверной системе будут иметь конечный момент более высокого порядка, чем в односерверной. Действительно, обозначив $\beta = 1 + \alpha/(s-k)$, из (8) получим

$$ES^\beta < \infty \quad \text{влечет } E\left(D^{(s-k)(\beta-1)}\right) < \infty. \tag{12}$$

Таким образом, для улучшения моментных свойств задержки целесообразнее использовать несколько медленных серверов/процессоров, чем один быстрый [23].

Рассмотрим численный пример. Пусть система с процессором частотой 3 ГГц обслуживает процессы, времена выполнения которых имеют тяжелый хвост, $\overline{B}(x) = x^{-1.5}$, $x \geq 1$ (распределение Парето). Пусть загрузка системы составляет $\rho = 0.3$. Согласно (9), стационарное время ожидания имеет лишь конечный момент $ED^{0.5} < \infty$, т.е. бесконечное среднее. Заменим систему на двухпроцессорную, где каждый процессор имеет частоту 1.5 ГГц. Тогда загрузка системы возрастет до 0.6, и согласно (12), $ED^{2\cdot 0.5} = ED < \infty$, т.е. средняя задержка конечна. (Заметим, что поскольку $\rho < 1$ то здесь $k = 0$.) Если же заменить систему на восьмипроцессорную, то $2 < \rho < 3$ и $k = 2$, т.е. $ED^3 < \infty$ и стационарная задержка имеет конечный третий момент.

2.1.2. Моменты компонент вектора загрузки

Еще более неожиданные моментные свойства компонент вектора загрузки в многосерверной системе $(W(1), \ldots, W(s))$ получены в работе [22]. Так, если коэффициент загрузки $\rho = ES/ET < s$, то для всех компонент с индексами $i \leqslant \lceil \rho \rceil$ (наименьшее целое большее ρ) имеют место *одни и те же* достаточные условия конечности моментов порядка $\alpha \geqslant 1$:

$$ES^{(s-\lfloor\rho\rfloor+\alpha)/(s-\lfloor\rho\rfloor)} < \infty \quad \text{влечет} \quad E\left[W(i)\right]^\alpha < \infty. \tag{13}$$

Однако для компонент с индексами $i > \lceil \rho \rceil$ моментные свойства зависят от индекса i:

$$ES^{(s-i+\alpha)/(s-i)} < \infty \quad \text{влечет} \quad E\left[W(i)\right]^\alpha < \infty. \tag{14}$$

При дополнительных ограничениях на распределение $B(x)$ эти условия являются также необходимыми [22]. Таким образом, имеет место зависимость условий конечности моментов компонент вектора W от порядка компонент в векторе.

2.1.3. Двухсерверная система: случай тяжелого хвоста

Классический результат (9) имеет интуитивное объяснение. Определяющим параметром для времени ожидания в системе является оставшееся время обслуживания S_I клиента i в момент прихода клиента $i+1$. Известно, что стационарное незавершенное время обслуживания (при условии, что система занята) имеет распределение

$$\overline{B}_I(x) = \frac{1}{ES}\int_x^\infty \overline{B}(y)\,dy, \;\; x \geq 0. \tag{15}$$

При этом для выполнения $ES_I^{k-1} < \infty$ при некотором $k > 1$ требуется $ES^k < \infty$, т. е. моментные свойства S_I хуже, чем у S.

В работе [6] приведены асимптотические результаты для хвоста распределения задержки D в двухсерверной системе G/G/2 при условии, что незавершенное время обслуживания

имеет субэкспоненциальное распределение. Так, в случае «максимальной устойчивости», $\rho := ES/ET < 1$

$$C_1 \leqslant \liminf_{x\to\infty} \frac{P(D>x)}{(\overline{B}_I(x))^2} \leqslant \limsup_{x\to\infty} \frac{P(D>x)}{(\overline{B}_I(x))^2} \leqslant C_2, \tag{16}$$

где константы C_1, C_2 зависят от ET, ES. В частности, в случае правильно меняющейся функции $\overline{B}(x)$ имеет место асимптотика

$$P(D>x) \sim C(\overline{B}_I(x))^2, \quad x\to\infty, \tag{17}$$

где постоянная C зависит от ES, ET [6].

В случае «минимальной устойчивости», $1 < \rho < 2$, для правильно меняющейся $\overline{B}(x)$ имеет место асимптотика

$$P(D>x) \sim C\overline{B}_I(\frac{\rho}{\rho-1}x), \quad x\to\infty, \tag{18}$$

где постоянная C зависит от ES, ET.

2.2. Дисциплины распределения задач по серверам

В работе [9] рассматривается многосерверная система обслуживания, состоящая из s одинаковых вычислителей, каждый из которых имеет собственную очередь (практическим примером такой системы является Грид). Пуассоновский входной поток заявок обслуживается диспетчером, который, зная время обслуживания заявки S, назначает ей вычислитель. Предметом эмпирического исследования в [9] является оптимальная дисциплина работы диспетчера с точки зрения среднего времени ожидания EW и *замедления*, определяемого как EW/ES. При этом время обслуживание заявки моделируется при помощи усеченного распределения Парето

$$dB(x) = \frac{\alpha k^\alpha}{1-(k/p)^\alpha} x^{-\alpha-1}, \quad k \leqslant x \leqslant p < \infty. \tag{19}$$

Время обслуживания с таким распределением может принимать большие, но все же конечные значения. В работе рассмотрены три типа известных дисциплин обслуживания: назначение вычислителя случайным образом; циклическое назначение (Round-Robin); динамическое назначение (заявка отправляется на вычислитель с минимальной незавершенной работой). В работе [9] также предлагается новая дисциплина, SITA-E (Size Interval Task Assignment with Equal Load), которая позволяет сбалансировать нагрузку на все вычислители в среднем. Интервал возможных размеров задач $[k, p]$ делится на участки $k = x_0 < x_1 < \ldots < x_s = p$ так, чтобы в среднем загрузка на этих участках была одинакова:

$$\int_{x_0}^{x_1} x\, dB(x) = \cdots = \int_{x_{s-1}}^{x_s} x\, dB(x) = \frac{ES}{s}. \tag{20}$$

В результате численного моделирования в работе было показано, что наиболее эффективной дисциплиной работы диспетчера в случае конечной дисперсии для тяжелохвостого распределения времени обслуживания ($\alpha \approx 2$) динамическое распределение наиболее эффективно. Тогда замедление в большинстве случаев меньше 1, а среднее время ожидания на 1-2 порядка меньше, чем у других дисциплин. Однако, при $\alpha \approx 1$ динамическое распределение теряет свои преимущества из-за неограниченного роста дисперсии времен обслуживания. В то же время дисциплина SITA-E равномерно хорошо себя ведет с точки зрения обеих метрик (замедления и среднего времени ожидания) на всем интервале $\alpha \in (1, 2)$.

3. Модель вычислительного кластера

Вычислительный кластер — система, состоящая из нескольких узлов, объединенных быстрой коммуникационной сетью, предоставляющая пользователю вычислительные ресурсы. Для построения модели кластера будем рассматривать следующие упрощения:

1. поток заявок представляет собой процесс восстановления (нет пачек, времена между приходами независимы);
2. разделяемым ресурсом является количество процессоров и время вычисления; времена вычисления заявок и требуемое каждой заявке число процессоров независимы;
3. заявки поступают в бесконечную очередь и обслуживаются в порядке прихода (FIFO);
4. заявка поступает на обслуживание только в момент освобождения требуемого ей количества процессоров.

3.1. Теоретические результаты

В дополнение к обозначениям секции 2.1, пусть N_i — требуемое число процессоров для заявки i, на каждом из которых она будет выполняться в течение S_i единиц времени. При этом очевидно, $1 \leqslant N_i \leqslant s$. Тогда можно предложить следующую модификацию рекурсии Кифера-Вольфовица (6) для данной модели:

$$W_{i+1} = R\left(W_i(N_i) + S_i - T_i, \ldots, W_i(N_i) + S_i - T_i, W_i(N_{i+1}) - T_i, \ldots, W_i(s) - T_i\right)^+, \ i \geq 1. \tag{21}$$

Действительно, поскольку i-я заявка должна дождаться освобождения всех требуемых ей процессоров, то ей придется ждать освобождения наиболее загруженного процессора с номером N_i (т.е. обнуления соответствующей компоненты вектора W_i). Поскольку все сервера с меньшими номерами будут простаивать в ожидании наиболее загруженного, не принимая другие заявки, то их загрузку также можно считать равной $W_i(N_i)$. Отметим, что на практике такая неэффективная дисциплина используется редко. Примером более эффективного распоряжения ресурсами является алгоритм Backfill, позволяющий заполнять промежутки простоя отдельных процессоров заявками без существенного нарушения приоритетов и очередности.

Заметим, что при $N_i = s$ все компоненты вектора загрузки можно считать равными, и поэтому заявка с номером $i + 1$ начнет обслуживание в полностью пустой системе (все процессоры освобождаются с уходом заявки i). Такая ситуация при некоторых дополнительных предположениях может порождать *регенерации* в системе, что открывает возможности применения регенеративного моделирования для надежного оценивания стационарных характеристик системы (см., напр., [17]).

Суммарная незавершенная загрузка системы в данном случае в момент прихода заявки i составит

$$N_i \cdot (W_i(N_i) + S_i - T_i)^+ + \sum_{k=N_i+1}^{s} (W_i(k) - T_i)^+. \tag{22}$$

Поэтому условие стационарности должно включать в себя ограничения на распределение задач в системе. Проведенные предварительные эксперименты позволяют предполагать, что в случае, когда значения N_i распределены в интервале $[1, s]$, условие устойчивости имеет вид $\rho = ES/ET < 1$ (максимальная устойчивость), в отличие от классического условия $\rho < s$ для системы G/G/s. В частности, если каждая заявка будет требовать ровно s процессоров, система вырождается в классическую G/G/1, условием устойчивости которой является $\rho < 1$.

Отметим, что задержка в рассматриваемой системе $D_i := W_i(1)$ ограничена снизу значением задержки $D_i^{(low)} := W_i^{(low)}(1)$ в классической системе G/G/s с тем же входным

потоком с интервалами $\{T_i\}$ и временами обслуживания $\{S_i\}$. Действительно, из рекурсии (21) следует, что, в предположении индукции $D_i \geqslant D_i^{(low)}$

$$D_{i+1} = W_{i+1}(1) = (W_i(N_i) + S_i - T_i)^+ \geqslant (W_i(1) + S_i - T_i)^+ =$$
$$= (D_i + S_i - T_i)^+ \geqslant (D_i^{(low)} + S_i - T_i) \geqslant (D_i^{(low)} + P_i - T_i)^+ = D_{i+1}^{(low)}.$$

С другой стороны, величина D_i в рассматриваемой системе ограничена сверху задержкой $D_i^{(up)}$ для системы, в которой $N_i = s, i \geqslant 1$. Действительно,

$$D_{i+1} = W_{i+1}(1) = (W_i(N_i) + S_i - T_i)^+ \leqslant$$
$$\leqslant (W_i(s) + S_i - T_i)^+ = (D_i^{(up)} + S_i - T_i)^+ = D_{i+1}^{(up)}.$$

Последняя (доминирующая) система представляет систему G/G/1, с теми же интервалами между приходами T_i и временами обслуживания S_i, что и в исходной системе. Таким образом, для любых реализаций $\{T_i, S_i\}$ имеет место

$$D_i^{(low)} \leqslant D_i \leqslant D_i^{(up)}. \tag{23}$$

Неравенства (23) для моментов порядка α соответствующих стационарных величин (полученных в пределе при $i \to \infty$) принимают вид

$$E(D^{(low)})^\alpha \leqslant ED^\alpha \leqslant E(D^{(up)})^\alpha. \tag{24}$$

Однако в «нижней» и «верхней» системах обслуживания (при условии $ES/ET < 1$) имеют место одинаковые достаточные условия конечности моментов времени ожидания. Следовательно условие $ES^{\alpha+1} < \infty$ является достаточным для конечности момента ED^α в предложенной модели. Таким образом, доказана

Теорема. В предложенной модели кластера (в предположении $\rho = ES/ET < 1$) условие $ES^{\alpha+1} < \infty$ является достаточным для конечности момента порядка α времени ожидания в системе, т.е. $ED^\alpha < \infty$.

3.2. Численный эксперимент

Моделирование предложенного типа систем обслуживания проводилось на вычислительном кластере ЦКП КарНЦ РАН [2]. В качестве модельного примера рассматривался сам кластер.

За период времени с 01.03.2010 г. по 01.12.2010 г. на кластере было посчитано 3682 задачи, суммарное время вычисления составило 1072804.5 минут. Астрономическое время в минутах между указанными датами составляет 396000 минут. Таким образом, среднее время вычисления задачи составило $ES \approx 291$ минуту, а среднее время между приходами составило $ET \approx 107$ минут. Кластер имеет 80 вычислительных ядер, сгруппированных по 8 штук на каждый вычислительный узел. В процессе вычислений разрешено занимать все узлы, но не менее одного узла. Очевидно, достаточное условие устойчивости такой системы не выполнено, так как $ES/ET > 1$. Действительно, эксперименты показали, что в системе наблюдается неограниченный рост значений задержек. График зависимости величины задержки от номера задачи (для пуассоновского входного потока и экспоненциального времени обслуживания) приведен на рисунке 1 (одна реализация процесса).

Предположим, что заявкам разрешено занимать не более 4-х узлов одновременно. Очевидно, в такой системе задержки будут меньшими, чем в исходной, где разрешено занимать все узлы (см. рис. 2). Кроме того, имеет место сходимость суммы ряда автокорреляций к конечной величине, что говорит об отсутствии долговременной зависимости. График частичных сумм автокорреляций представлен на рисунке 3.

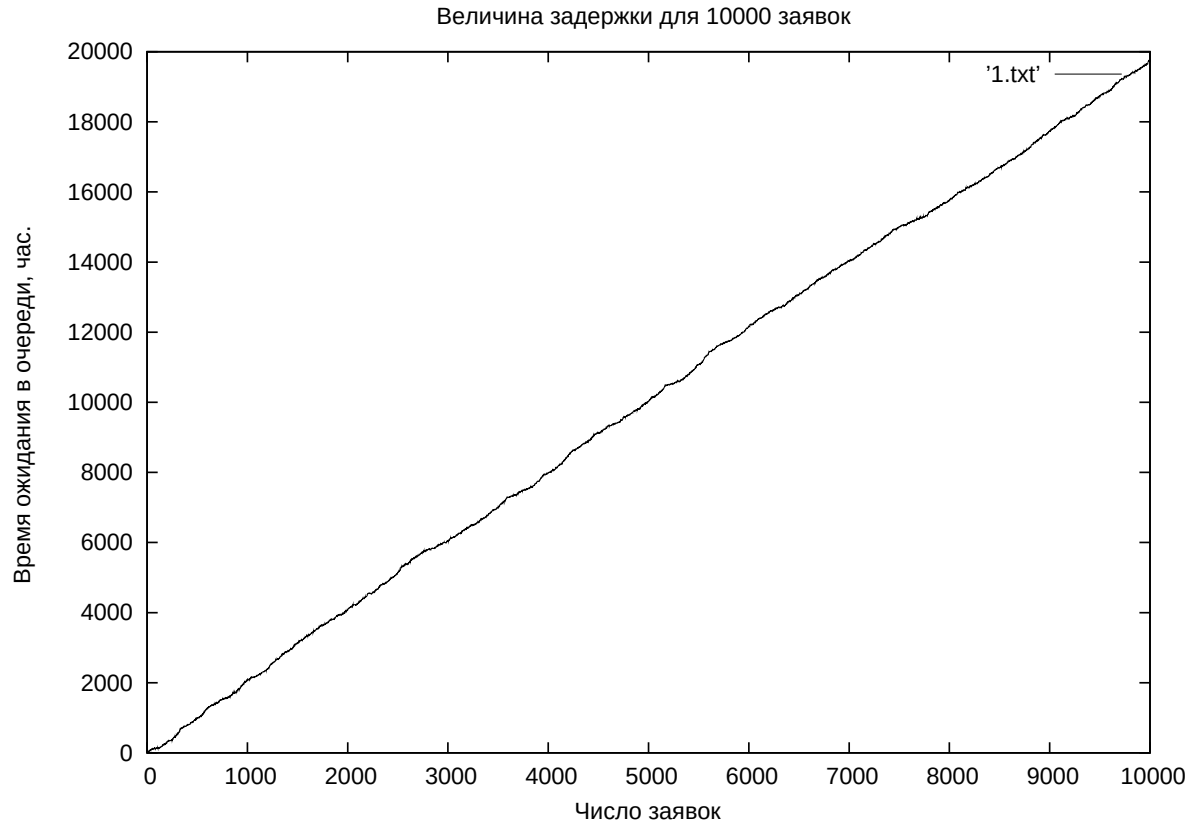

Рис. 1. Задержки в модели 10-узловой системы

Пусть теперь время обслуживания S имеет тяжелый хвост. Промоделируем этот случай при помощи распределения Парето с параметром $\alpha = 2.2$. Входной поток предполагается пуассоновским, с параметром $\lambda = 1$. В этом случае средняя загрузка равна [1] $\lambda/(\alpha-1) < 1$, и поэтому система должна обладать стационарным режимом. В то же время Рис. 4 показывает рост суммы автокорреляций, что говорит о возможном присутствии долговременной зависимости у процесса времени ожидания в очереди.

Однако, времена ожидания не уходят на бесконечность, хотя наблюдается тенденция длительных периодов нахождения времени ожидания выше или ниже своего среднего, что также косвенно указывает на долговременную зависимость (рис. 5).

Необходимо отметить, что модельные результаты отличаются от реального наблюдаемого поведения кластера, прежде всего потому, что на кластере ЦКП КарНЦ РАН предельный размер очереди существенно ограничен. Именно, каждому пользователю разрешается запускать не более 4 задач одновременно. Кроме того, используется алгоритм Backfill для выбора заданий из очереди, а также имеется ряд других ограничений, не отраженных в модели. Поэтому предложенная модель может рассматриваться лишь как предварительный вариант модели, описывающей функционирование кластера.

Для лучшей адаптации модели к реальным кластерам следует рассмотреть модели с конечным буфером для заявок, с возможностью появления пачек заявок, с более сложными дисциплинами выбора заявок на обслуживание (например, алгоритм Backfill и некоторые другие алгоритмы планировщиков). Аналитическое исследование в этих случаях существенно усложнится. В целом, точность предложенной модели и возможные направления ее развития должны быть установлены с помощью более обширных экспериментов.

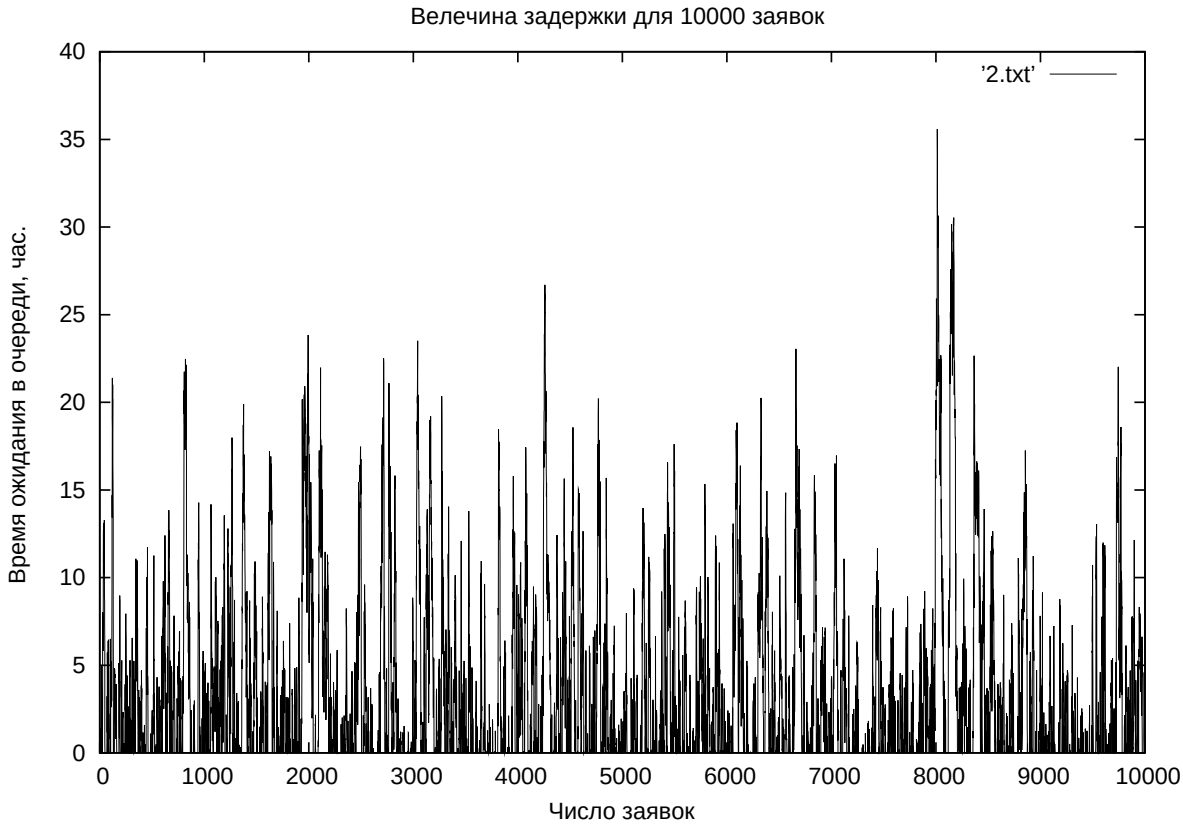

Рис. 2. Задержки в модели 10-узловой системы с правом занимать не более 4 узлов

4. Заключение

В работе рассмотрены некоторые важные результаты, касающиеся моментных свойств процесса обслуживания в многосерверных системах обслуживания, включая случай, когда время обслуживания имеет тяжелый хвост. Эти результаты могут применяться для оценивания качества обслуживания вычислительных кластеров и Грид. Предложена новая модель вычислительного кластера на основе модифицированной рекурсии Кифера-Вольфовица для многосерверной системы $GI/G/s$. Для этой модели получены условия конечности соответствующих моментов стационарного времени ожидания (задержки) в очереди. Приведены результаты вычислительных экспериментов по моделированию системы, в том числе для времени обслуживания с тяжелым хвостом.

Литература

1. Морозов Е.В., Румянцев А.С. Регенерация и корреляционные свойства стационарной задержки в одноканальной очереди. // Распределенные компьютерные и коммуникационные сети (DCCN-10). М., 2010. С. 58–67.

2. Центр высокопроизводительной обработки данных ЦКП КарНЦ РАН. URL: http://cluster.krc.karelia.ru (дата обращения: 13.12.2010 г.).

3. Borst S., Boxma O., Núñez-Queija R. Heavy Tails:The Effect of the Service Discipline. // Proceedings of Performance TOOLS 2002. 2002.

4. Crovella M., Bestavos A. Self-similarity in World Wide Web traffic: evidence and possible causes. // IEEE/ACM Transactions on Networking. 2002. Vol. 5. N. 6. P. 835–845.

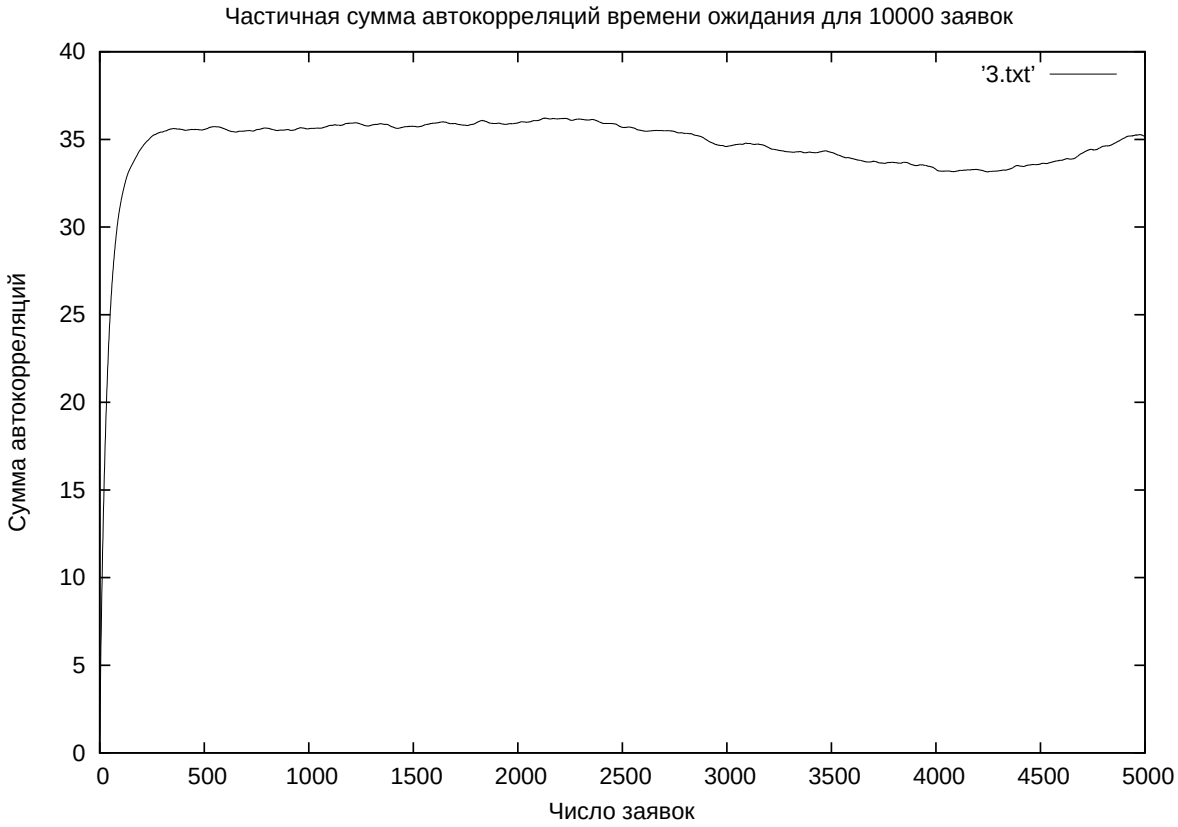

Рис. 3. Автокорреляции в 10-узловой системе с правом занимать не более 4 узлов

5. Daley D. The serial correlation coefficients of waiting times in a stationary single server queue. // Journal of Australian Mathematical Society. 1969. N. 8 P. 683–699.

6. Foss S., Korshunov D. Heavy Tails in Multi-Server Queue. // Queueing Systems. 2006. Vol. 52. P. 31-48.

7. Goldie C., Klüppelberg C. Subexponential Distributions // A practical guide to heavy tails. 1998. P. 435–459.

8. Greiner M., Jobmann M., Klüppelberg C. Telecommunication traffic, queueing models and subexponential distributions. // Queueing Systems. 1999. Vol. 33. P. 125–152.

9. Harchol-Balter M. The Effect of Heavy-Tailed Job Size Distributions on Computer System Design. // Proceedings of ASA-IMS Conference on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics, Washington, DC. June 1999.

10. Heath D., Resnick S., Samorodnitsky G. Heavy Tails and Long Range Dependence in ON/OFF Processes and Associated Fluid Models. // Mathematics of Operations Research 1998. Vol. 23. P. 145–165

11. Huang T., Sigman K. Steady-state Asymptotics for Tandem, Split-Match and Other Feedforward Queues with Heavy-Tailed Service. // Queueing Systems. 1999. Vol. 33. P. 233–259.

12. Jessen A., Mikosch T. Regularly Varying Functions. // Publications de L'Institut Mathematique. Nouvelle serie, tome 80 (94) 2006. P. 171–192

13. Keilson J., Seidmann A. M/G/∞ with Batch Arrivals. September 1987.

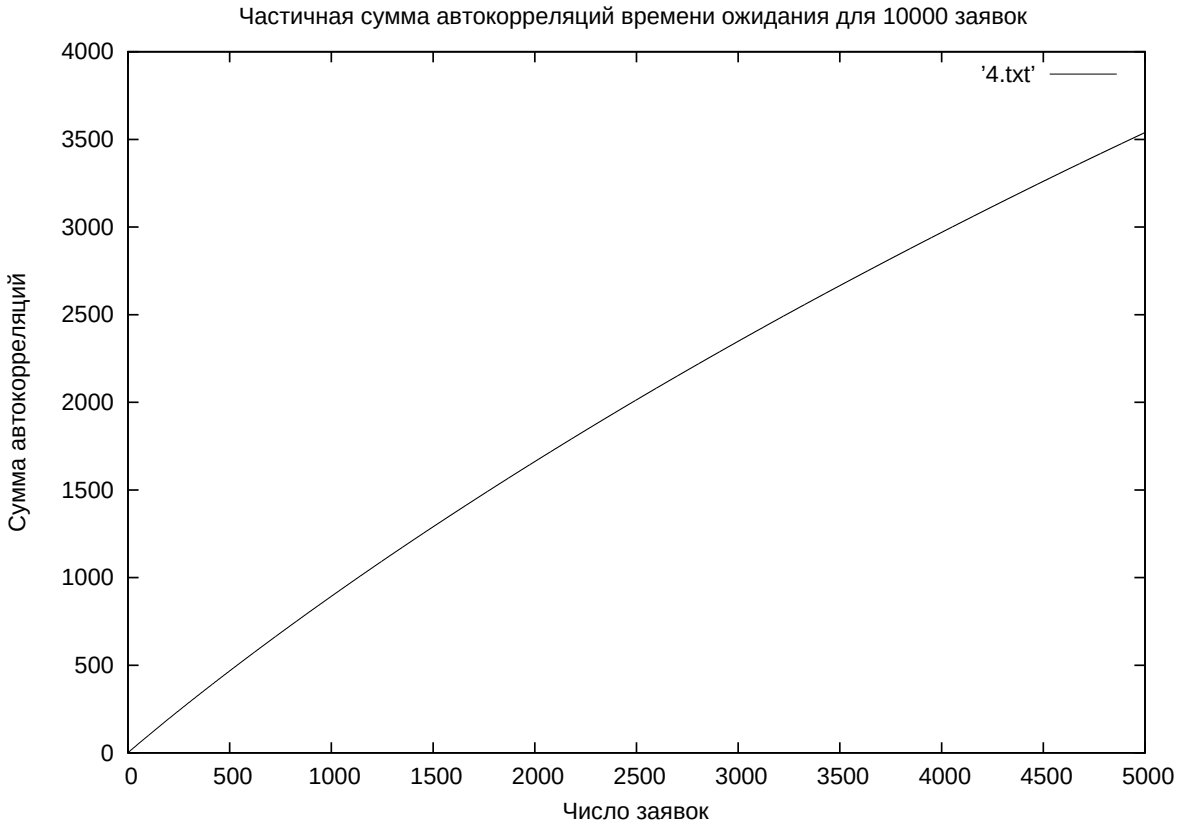

Рис. 4. Автокорреляции в 10-узловой системе, случай тяжелых хвостов

14. Leland W., Taqqu M., Willinger W., Wilson D. On the Self-Similar Nature of Ethernet Traffic // Journal IEEE/ACM Transactions on Networking (TON). February 1994. Vol. 2, N. 1.

15. Liu, L. Batch arrival infinite server queues with constant service times. // Proceedings of the First Symposium on Queueing Theory in China September 1993. P. 47–53

16. Morozov E., Pagano M., Rumyantsev A. Heavy-tailed Distributions with Applications to Broadband Communication Systems. // Proceedings of AMICT'2007. 2008. Vol. 9. P. 157–174.

17. Morozov E., Rumyantsev A. Moment properties of queueing systems and networks // Proceedings of ICUMT'2010 (Ультрасовременные телекоммуникации и системы управления). 2010.

18. Samorodnitsky G. Long Range Dependence. // Foundations and Trends in Stochastic Systems. 2006. Vol. 1. P. 163–257.

19. Scheller-Wolf A. Further delay moment results for FIFO multiserver queues. // Queueing Systems. 2000. Vol. 34 P. 387–400.

20. Scheller-Wolf A., Sigman K. Delay Moments for FIFO GI/GI/s queues. // Queueing Systems. 1997. Vol. 25. P. 77–95.

21. Scheller-Wolf A., Sigman K. New bounds for expected delay in FIFO GI/GI/c queues. // Queueing Systems. 1997. Vol. 26. P. 169–186.

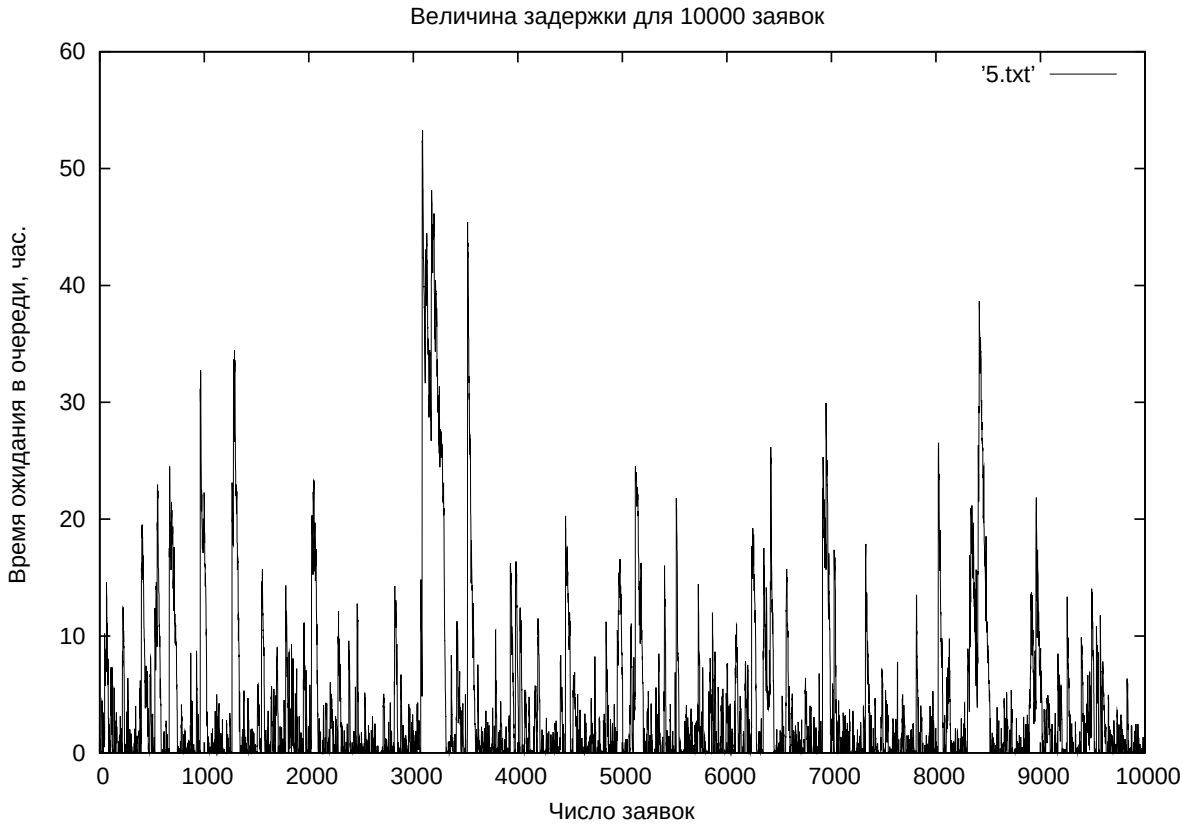

Рис. 5. Задержки в модели 10-узловой системы, случай тяжелых хвостов

22. Scheller-Wolf A., Vesilo R. Sink or Swim Together: Necessary and Sufficient Conditions for Finite Moments of Workload Components in FIFO Multiserver Queues. March 2008.

23. Scheller-Wolf A., Vesilo R. Structural interpretation and derivation of necessary and sufficient conditions for delay moments in FIFO multiserver queues. // Queueing Systems. 2006. Vol. 54. P. 221–232

24. Wolff R. On Finite Delay-Moment Conditions in Queues. // Operations Research. September-October 1991. Vol. 39. N. 5. P. 771–775.