

## ПРОБЛЕМЫ И ТЕХНИКА МОДЕЛИРОВАНИЯ ФРАКТАЛЬНЫХ ОЧЕРЕДЕЙ<sup>1</sup> В.Н. Задорожный (Омск), О.И. Кутузов (Санкт-Петербург)

### Введение

Фрактальная природа сетевого трафика, осознанная и изучаемая научным сообществом с 1993 г. [1], привела к необходимости создания новых методов анализа и проектирования сетей связи. Фрактальный трафик описывается новыми формализмами – «самоподобными» случайными процессами, долговременными зависимостями (ДВЗ) и распределениями с тяжелыми хвостами (РТХ) [2]. Создаются новые методы идентификации сетевого трафика; особое внимание уделяется анализу очередей сообщений, поскольку при фрактальном трафике затраты на их буферизацию резко возрастают [1–4].

В статье рассматриваются проблемы имитационного моделирования (ИМ) очередей, обусловленные фрактальной природой трафика. В качестве базовых формализмов используются фрактальные системы с очередями (FQS), определяемые следующими положениями. На вход FQS поступает рекуррентный поток заявок. Интервалы  $\tau_i$  их поступления описываются функцией распределения (ф.р.)  $A(t)$ . Время  $x_i$  обслуживания описывается ф.р.  $B(t)$ . Хотя бы одна из ф.р.  $A(t)$ ,  $B(t)$  имеет бесконечную дисперсию и асимптотически степенной хвост (т.е. задает *фрактальную* случайную величину (с.в.) [5]). Для хранения поступающих заявок имеется буфер размером  $m \leq \infty$ . Заявка, поступившая, тогда когда все каналы и все места в буфере заняты, теряется (получает отказ).

### Задачи исследования фрактальных очередей

Одной из основных задач исследования FQS является определение вероятности  $P$  отказа при заданном размере буфера  $m$  (прямая постановка задачи) или определение наименьшего  $m$ , обеспечивающего заданную малую вероятность  $P$  отказа (обратная постановка). При  $m = \infty$  интерес представляет определение стационарной средней длины очереди  $L$ . Далее рассматриваются FQS с конечными математическими ожиданиями (м.о.)  $\bar{\tau}$ ,  $\bar{x}$  с.в.  $\tau \in A(t)$ ,  $x \in B(t)$  и с коэффициентом загрузки  $\rho = \bar{x}/(n\bar{\tau}) < 1$  ( $n$  – число каналов). Типичные представители FQS – это системы  $M|Pa|n|m$ ,  $Pa|M|n|m$  и  $Pa|Pa|n|m$  (в обозначениях Кендалла), где  $Pa$  – распределение Парето (РП), имеющее ф.р.:

$$F(t) = 1 - (K/t)^\alpha, \quad t \geq K, \quad K > 0, \quad \alpha > 0, \quad (1)$$

далее обозначаемую в виде  $Pa(K; \alpha)$ . Конечное м.о.  $\alpha K/(\alpha - 1)$  РП имеет при  $\alpha > 1$ , бесконечную дисперсию – при  $\alpha \leq 2$ . Поэтому у рассматриваемых FQS ф.р.  $A(t)$  и/или  $B(t)$  есть РП с параметром  $1 < \alpha \leq 2$ . Чем меньше  $\alpha$ , тем «тяжелее» хвост РП. Изменяя при фиксированном  $\alpha$  масштабный параметр  $K$ , можно получать любое требуемое м.о.

Если обе ф.р.  $A(t)$  и  $B(t)$  являются РП, то FQS адекватно отражает все три основные особенности фрактального трафика – его статистическое «самоподобие», наличие характерных взрывных пульсаций нагрузки и ДВЗ объемов поступлений.

### Проблемы имитационного моделирования фрактальных очередей

Основным методом исследования FQS является ИМ. Имеющийся опыт [6–8] позволяет в числе первоочередных проблем моделирования FQS указать следующие три:

- проблему корректной реализации РТХ (в частности, РП) в ИМ,
- проблему длительных переходных процессов (ПП),
- проблему эффективного решения обратной задачи (задачи определения минимального  $m$ , обеспечивающего заданную малую вероятность отказа  $P$ ).

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ (проект 12-07-00149-а).

**Проблема корректной реализации РТХ**

Формула для генерации с.в.  $x \in Pa(K; \alpha)$ , получаемая обращением ф.р. (1), имеет вид  $x = K(1-z)^{-1/\alpha}$  или  $x = Kz^{-1/\alpha}$ , где  $z$  – базовая с.в. (БСВ), равномерно распределенная в промежутке от 0 до 1. Программные датчики БСВ реализуют ее *дискретную версию*  $\hat{z}$  с множеством равновероятных значений  $\{0, \varepsilon, 2\varepsilon, \dots, 1-\varepsilon, 1\}$  (возможно, без нуля и/или единицы), образующим решетку с шагом  $\varepsilon$ . В GPSS датчик Uniform(1,0,1) реализует дискретную с.в. (д.с.в.)  $\hat{z}$  с шагом  $\varepsilon = 10^{-6}$ . В других программах шаг  $\varepsilon$  составляет около  $10^{-15}$ . При реализации РТХ шаг  $10^{-15}$  (и, тем более,  $10^{-6}$ ) часто оказывается слишком большим и приводит к значительным отличиям свойств реализуемой д.с.в.  $\hat{x} = K\hat{z}^{-1/\alpha}$  от свойств непрерывной с.в. (н.с.в.)  $x \in Pa(K; \alpha)$ . Поэтому при ИМ FQS имеет смысл говорить о реализации *дискретного* РП  $Pa(K; \alpha; \varepsilon)$ . В табл. 1 характеристики д.с.в.  $\hat{x}$  сравниваются при различных  $\varepsilon$  с соответствующими характеристиками н.с.в.  $x$ , приведенными в колонках « $\varepsilon \rightarrow 0$ » [6]. Из таблицы видно, что чем тяжелее хвост РП, тем сильнее отличается м.о. реализуемой д.с.в.  $\hat{x} \in Pa(1; \alpha; \varepsilon)$  от м.о. н.с.в.  $x \in Pa(K; \alpha)$ . Очевидно, аналогичные особенности имеют место и при реализации в ИМ других РТХ.

**Таблица 1****Сравнение числовых характеристик д.с.в.  $\hat{x} \in Pa(1; \alpha; \varepsilon)$  и н.с.в.  $x \in Pa(1; \alpha)$** 

$\alpha \backslash \varepsilon$	Математические ожидания				Коэффициенты вариации			
	$10^{-6}$	$10^{-12}$	$10^{-15}$	$\varepsilon \rightarrow 0$	$10^{-6}$	$10^{-12}$	$10^{-15}$	$\varepsilon \rightarrow 0$
<b>1,01</b>	13,415	24,612	29,662	101	2652	$3,9 \cdot 10^4$	$9,8 \cdot 10^5$	$\infty$
<b>1,1</b>	8,0297	10,154	10,549	11	48,3	$1,1 \cdot 10^4$	$1,8 \cdot 10^5$	$\infty$
<b>1,5</b>	2,9755	2,9998	3,0000	3	6,27	63,2	200,0	$\infty$
<b>1,9</b>	2,1089	2,1111	2,1111	2,1111	1,96	3,68	4,67	$\infty$
<b>2</b>	1,9985	1,999985	2	2	1,61	2,46	2,79	$\infty$

В разных задачах отличие дискретного РП от непрерывного проявляется по-разному.

Например, в FQS  $M|Pa|1|_{\infty}$  при  $x \in Pa(K; \alpha)$ ,  $1 < \alpha \leq 2$  стационарная средняя длина  $L(\rho)$  очереди заявок бесконечна при любом коэффициенте загрузки  $\rho > 0$ :

$$L(\rho) = \frac{\lambda^2 x^{(2)}}{2(1-\rho)} = \frac{\rho^2(1+C_x^2)}{2(1-\rho)} = \infty, \quad (2)$$

т.к. второй момент  $x^{(2)}$  времени  $x$  и соответственно его коэффициент вариации  $C_x^2$  бесконечны. Если эту FQS исследовать методом ИМ, то, поскольку вместо н.с.в.  $x \in Pa(K; \alpha)$  реализуется д.с.в.  $\hat{x}$ , например, при  $\varepsilon = 10^{-15}$ ,  $\alpha = 2$  получится (искажаемое погрешностями и незавершенным ПП) приближение к иному результату:

$$L(\rho) = \frac{\rho^2(1+C_{\hat{x}}^2)}{2(1-\rho)} \approx \frac{\rho^2(1+2,79^2)}{2(1-\rho)} \approx \frac{4,39\rho^2}{1-\rho}, \quad 0 < \rho < 1$$

(табл. 1), который принципиально отличается от истинного  $L(\rho) = \infty$ ,  $0 < \rho < 1$ .

В другой FQS – в системе  $Pa|M|1|_{\infty}$  – дискретность цифровых датчиков с.в. влияет на результаты имитационного исследования зависимости  $L(\rho)$  значительно меньше. В этой FQS среднюю длину  $L(\rho)$  очереди можно рассчитать точно по формуле

$$L(\rho) = \frac{\rho\sigma}{1-\sigma}, \quad (3)$$

где  $\sigma$  – единственный в области  $0 \leq \sigma < 1$  корень уравнения  $\sigma = A^*(\mu - \mu\sigma)$ ;  $\mu = 1/\bar{x}$  – интенсивность обслуживания;  $A^*(s) = M(e^{-st}) = \int_0^\infty e^{-st} dA(t)$  – преобразование Лапласа от ф.р.  $A(t)$  интервала  $\tau$  поступления заявок, в данном случае – от ф.р. Парето (1). С учетом (1) уравнение  $\sigma = A^*(\mu - \mu\sigma)$  сводится к решаемому численно уравнению:

$$\sigma = \alpha K^\alpha \mu^\alpha (1 - \sigma)^\alpha \Gamma(-\alpha, K\mu(1 - \sigma)), \quad (4)$$

где  $\Gamma(c, x)$  – неполная гамма-функция:  $\Gamma(c, x) = \int_x^\infty t^{c-1} e^{-t} dt$ . При ИМ системы  $\text{Pa}|M|1|_\infty$  хотя и возникают заметные вычислительные и статистические погрешности, однако характер зависимости  $L$  от  $\rho$  принципиально не искажается. Так, при  $\alpha = 1,1$ ,  $K = 1$  (т.е.  $\bar{\tau} = 11$ ) и  $\rho = 0,1; 0,2; \dots; 0,6$ ;  $\mu = 1/(\rho\bar{\tau})$  решения уравнения (4)  $\sigma \approx 0,259; 0,649; 0,872; 0,964; 0,993; 0,999$  определяют соответствующие  $L(\rho) \approx 0,035; 0,369; 2,04; 10,8; 69,1$  и  $648$  (точные значения округлены). В результате ИМ данной FQS (при  $\varepsilon = 10^{-12}$  и длине прогона 10 млн. заявок) получены имитационные оценки  $\hat{L}(\hat{\rho}) \approx 0,047; 0,51; 2,7; 14,4; 98$  и  $859$ . При этом оценки коэффициента загрузки в отличие от заданных значений  $\rho$  составили  $\hat{\rho} \approx 0,136; 0,275; 0,402; 0,532; 0,673$  и  $0,831$ . Такие погрешности оценок могут быть в ряде задач вполне приемлемыми. Сложнее, однако, путем непосредственного ИМ установить, что в рассмотренной FQS, как следует из (4), при фиксированном  $\mu$  и  $\alpha \rightarrow 1$  (т.е. при  $\rho \rightarrow 0$ ) имеет место  $L(\rho) \rightarrow \infty$ . Кстати, при ИМ м.о. д.с.в.  $\hat{\tau} \in \text{Pa}(1; 1; \varepsilon)$  даже при  $\varepsilon = 10^{-100}$  составит лишь  $230,8357\dots$ , тогда как для н.с.в.  $\tau \in \text{Pa}(1; 1)$   $\bar{\tau} = \infty$ .

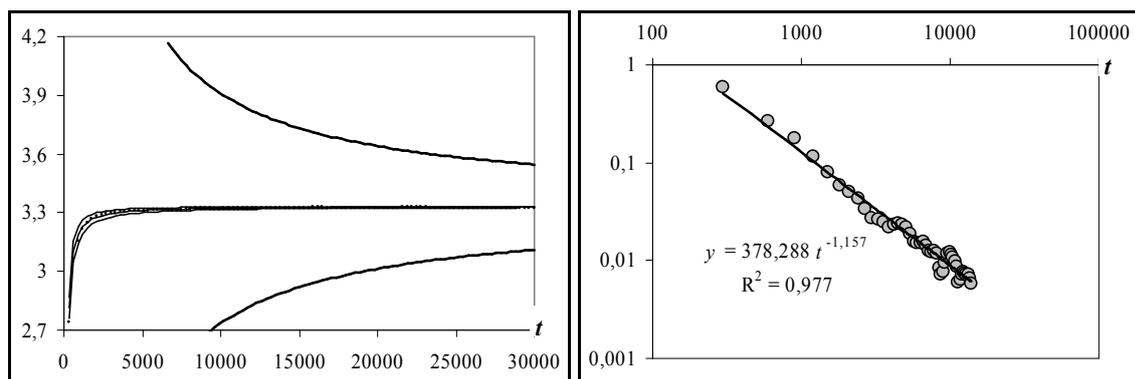
В общем случае для решения проблемы корректной реализации РТХ требуется использовать арифметику с произвольной разрядностью чисел. Для практических целей, по нашим оценкам, в большинстве случаев достаточно 60–100 десятичных значащих цифр. А при использовании машинной арифметики с обычной точностью, когда численные результаты моделирования FQS могут существенно зависеть от шага  $\varepsilon$  дискретизации БСВ, следует этот шаг указывать вместе с результатами моделирования.

### Проблема длительных переходных процессов

Бесконечная дисперсия РП и ДВЗ порождает в рассматриваемых FQS длительные ПП. Математическая структура оценок для средней длины  $L$  очереди и вероятности  $P$  отказа такова, что в среднем они сходятся к своим точным стационарным значениям как степенные функции моделируемого времени  $t$ . Учитывая этот факт, можно находить стационарные значения  $L$  и  $P$ , не дожидаясь завершения ПП. Проиллюстрируем эту идею двумя примерами.

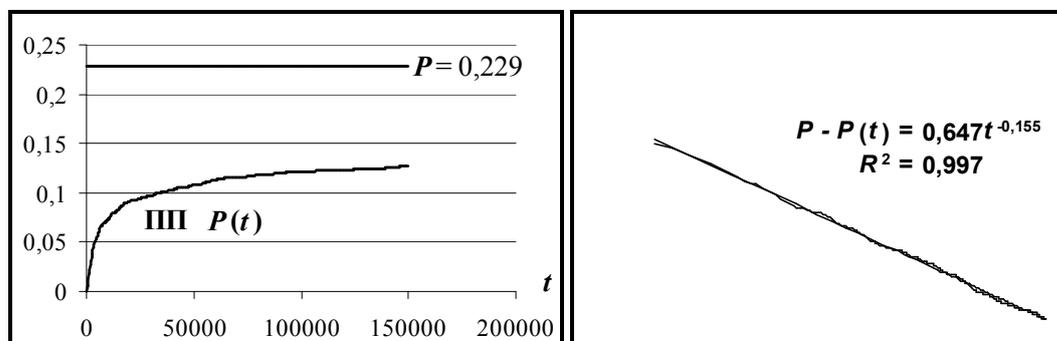
*Пример 1.* Для системы  $\text{Pa}|M|1|_\infty$  ( $K = 1$ ,  $\alpha = 1,5$ ,  $\rho = 0,7$ ,  $\varepsilon = 10^{-12}$ ) на рис. 1 слева показан ПП  $L_{cp}(t)$  с доверительными полосами, полученный на GPSS путем усреднения 10 000 независимых ПП  $L_j(t)$  (т.е. изменений СЧА  $QA_j$  во времени  $t$ ).

В ходе ИМ вместе с оценками  $L_{cp}(t)$  для  $t = 300, 600, \dots, 30\,000$  были рассчитаны оценки  $\sigma(t)$  их средних квадратичных отклонений. Результаты ИМ перенесены в Excel. По графику зависимости  $L_{cp}(t)$  для  $L$  «визуально» определено и записано в отдельную ячейку начальное приближение. К графику зависимости от  $t$  разности  $y = L - L_{cp}(t)$  (см. маркеры на рис. 1 справа) добавлена линия степенного тренда. Далее приближение для  $L$  уточнено подбором такого его значения, при котором достигается максимум коэффициента  $R^2$  детерминации тренда. Наилучший результат  $R^2 \approx 0,977$  получен при  $L \approx 3,331$ . Проверить точность этой оценки, можно расчетом для системы  $\text{Pa}|M|1|_\infty$  ( $K = 1$ ,  $\alpha = 1,5$ ,  $\rho = 0,7$ ) точного решения по формулам (3), (4). Это решение  $L \approx 3,3175$  отличается от найденной для  $L$  оценки лишь на 0,41%.



**Рис. 1. Слева:** аппроксимация для  $L_{cp}(t)$  (центральная линия), доверительная полоса для  $L_{cp}(t)$  (две линии рядом с центральной) и доверительная полоса для усредняемых оценок  $L_j(t)$ .  
**Справа:** график разности  $y = L - L_{cp}(t)$  (маркеры) и линия ее степенного тренда (прямая)

*Пример 2.* На рис. 2 показан ПП оценки для  $P$  при ИМ системы  $Pa|Pa|1|100$  при  $\tau \in Pa(1; 1,1; \varepsilon)$ ,  $x \in Pa(0,5; 1,1; \varepsilon)$ .  $\varepsilon = 10^{-12}$ . Здесь  $\bar{\tau} = 10,154$ ,  $\bar{x} = 5,077$ ,  $\rho = 0,5$ . Слева – сам ПП, усредненный по 1000 независимых реплик, справа – график разности  $P - P(t)$  в логарифмических шкалах (кривая линия) и линия степенного тренда (прямая), в ходе построения которой найдена наилучшая оценка  $P = 0,229$ . Следовательно, усредненный ПП описывается как  $P(t) \approx 0,229 - 0,647t^{-0,155}$ . Из этого уравнения находим, что без аппроксимации, при длине всех 1000 прогонов 300 млн е.в. (т.е. при наблюдении около  $10^{10}$  заявок), оценка  $P(t)$  в среднем составила бы  $0,198$  и примерно на 14% «не дошла» бы до стационарного  $P = 0,229$ .



**Рис. 2. ПП  $P(t)$  оценки для  $P$  при ИМ FQS  $Pa|Pa|1|100$ ,  $\tau \in Pa(1; 1,1; 10^{-12})$ ,  $x \in Pa(0,5; 1,1; 10^{-12})$**

Таким образом, метод степенной аппроксимации ПП позволяет существенно повысить точность и сократить время расчета стационарных характеристик FQS. Вместе с тем этот метод не освобождает от необходимости использования большого запаса качественных случайных чисел. К сожалению, в GPSS такой запас «не дотягивает» и до 1 млрд., т.е. на современных ПК может быть исчерпан за 3–5 минут. Это вместе с потребностью в «длинной» арифметике обуславливает актуальность создания для ИМ FQS специализированного программного обеспечения.

#### Решение обратной задачи

Для решения обратной задачи (определение  $m$  по заданной достаточно малой  $P$ ) нужно преодолеть одновременно проблему оценки в ИМ малых вероятностей (при заданных  $m$ ), и проблему перебора значений  $m$  для отыскания подходящих. Это удается за счет

степенной аппроксимации зависимости  $P(m)$  [8]. В принципе, достаточно решить задачу определения стационарной  $P$  при трех больших  $m$ , т.е. найти три точки  $(m_1, P_1)$ ,  $(m_2, P_2)$ ,  $(m_3, P_3)$ , отложить их на графике с логарифмическими шкалами, и провести через них прямую методом наименьших квадратов. Высокое значение  $R^2$  будет свидетельствовать о хорошей точности аппроксимации, низкое – о плохой.

*Пример 3.* Для системы  $\text{Pa}|\text{Pa}|1|m$  с параметрами  $\tau \in \text{Pa}(1; 1,25; \varepsilon)$ ,  $x \in \text{Pa}(1; 1,5; \varepsilon)$ ,  $\varepsilon = 10^{-12}$ , при  $m = 300, 400$  и  $500$  методом степенной аппроксимации результатов ИМ (использовалось по 10 тысяч независимых реплик) найдены стационарные вероятности отказа  $P = 0,0124, 0,0102$  и  $0,00870$ . Соответствующие три точки  $(m_i, P_i)$  в логарифмических осях «идеально» ложатся на прямую. Полученная по ним степенная аппроксимация  $P = 0,6467m^{-0,6930}$  с показателем  $R^2 = 0,9998$  позволяет решить обратную задачу в общем виде:  $m = (P/0,6467)^{-1/0,6930} = 0,53314 \cdot P^{-1,443}$ . Используя это решение, находим, например, что для достижения вероятности  $P = 10^{-6}$  нужен буфер, рассчитанный на хранение около 243 млн заявок. И это при загрузке  $\rho = 0,6$  (!)

### Решение прикладной задачи с применением предложенных методов

Для снижения вероятности отказа  $P$  можно делать следующее:

- увеличивать размер буфера;
- повышать скорость обслуживания заявок;
- наращивать число каналов.

Применять в FQS буферы большого размера малоэффективно (в *примере 3* размер буфера растет как  $(1/P)^{1,443}$  – быстрее, чем убывает  $P$ ). Повышение скорости каналов – тоже. Так, если в FQS из *примера 3* положить, что  $x \in \text{Pa}(0,1; 1,5; \varepsilon)$ , т.е. ускорить обслуживание на порядок, то для обеспечения вероятности  $P = 10^{-6}$  потребуются (при загрузке  $\rho = 0,06$ ) буфер размером около **67 тыс. мест**. Однако (в отличие от классических СМО) неожиданно высокоэффективным путем становится наращивание числа каналов. Если, например, в этой же FQS вместо ускорения обслуживания увеличить число каналов (не на порядок, а лишь впятеро, то в полученной системе  $\text{Pa}|\text{Pa}|5|m$  уже при  $m = 25$  вероятность отказа практически равна нулю: при прогоне 400 млн заявок не происходит ни одного отказа. Аналогичный эффект наблюдается и при других соотношениях параметра  $\alpha$  в распределениях с.в.  $x, \tau$ , а также в системах  $M|\text{Pa}|n|m$ .

### Выводы

При моделировании фрактальных очередей устоявшиеся приемы ИМ систем с очередями приходится пересматривать и переоценивать заново. Проблемы возникают в связи с особенностями генерации с.в., описываемых РТХ (1), большой длительностью ПП (2) и оценками малых вероятностей  $P$  потерь при размере буфера  $m < \infty$  (3). Корректное применение степенных аппроксимаций, основанных в случае проблемы (2) на асимптотически-степенном характере сходимости оценок, а в случае проблемы (3) – на асимптотически-степенном характере зависимости  $P(m)$ , позволяет эти проблемы решать. При этом необходимо применять большое число независимых прогонов модели.

Проблема (1), требующая применения «длинной» арифметики и качественных датчиков с.в., а также задачи дальнейшего развития техники ИМ фрактальных очередей обуславливают актуальность создания специализированных программных средств ИМ.

### Литература

1. **Stallings, William.** Интернет и телекоммуникации / W. Stallings. [Электронный ресурс]: URL: <http://my.online.ru/it/press/cwm/1997/world.htm> (Дата обращения: 13.03.2010).
2. **Mark E. Crovella, Murad Taqqu and Azer Bestavros,** «Heavy Tailed-Probability distributions in the World Wide Web» 5(6): 835–846, December 1997.

3. Шелухин, О.И. Фрактальные процессы в телекоммуникациях / О.И. Шелухин, А.М. Тенякшев, А.В. Осин: под ред. О.И. Шелухина. – М.: Радиотехника, 2003. – 480 с.
4. Zwart, A. P. (2001) Queueing Systems with Heavy Tails. *Eindhoven University of Technology*. – 227 P.
5. Мандельброт, Б. Фрактальная геометрия природы / Б. Мандельброт. – М.: Институт компьютерных исследований, 2002. – 656 с.
6. Задорожный В.Н., Кутузов О.И. Проблемы генерации случайных величин с фрактальными распределениями / Омский научный вестник. – 2012. – № 3 (113). – С. 20–24.
7. Задорожный В.Н., Кутузов О.И. Методы моделирования очередей в условиях фрактального трафика в сетях с коммутацией пакетов: учеб. пособие. – Омск, ОмГТУ. – 2013. – 104 с.