

Фрактальные свойства потоков событий прикладного уровня в информационных системах

Н. Г. Треногин, Д. Е. Соколов

Приводятся результаты оценки статистических характеристик потока событий прикладного уровня в действующей информационной системе оператора связи, позволяющие сделать вывод о наличии фрактальных свойств в данном потоке.

Ключевые слова: фрактальные процессы, самоподобие, моделирование нагрузки.

1. Введение

Ряд исследований последнего времени показывает недостаточность использования классических моделей теории телетрафика для описания нагрузки в современных информационных системах и сетях передачи данных и для оценки их вероятностно-временных характеристик. Традиционные подходы, основанные на марковских моделях, не учитывают такие свойства реальной нагрузки, как выраженная вариативность, коррелированность (потоки с последствием), весовые распределения статистических величин и т.д. Для описания потоков, обладающих указанными свойствами, в настоящее время все чаще используются фрактальные, или самоподобные, модели, переносящие отдельные положения фрактальной геометрии на свойства стохастических процессов.

Адекватность фрактальных моделей для описания потоков нагрузки на уровне сетевых пакетов многократно подтверждена в зарубежных и отечественных работах. При этом причины формирования фрактальных, или самоподобных, свойств сетевого трафика, как правило, рассматриваются в качестве внешних по отношению к рассматриваемому предмету факторов и выносятся за рамки исследования.

В настоящей статье на примере данных действующей информационной системы показывается, что потоки событий не только сетевого, но и прикладного уровня могут обладать фрактальными свойствами, что дает возможность использования единого математического аппарата при комплексном моделировании систем обработки данных на различных уровнях абстракции.

2. Базовые понятия и свойства фрактальных моделей трафика

Наиболее общим определением фрактального (самоподобного) процесса является его неформальное определение как случайного процесса, статистические характеристики которого проявляют свойства масштабирования. Самоподобный процесс существенно не меняет вида при рассмотрении в различных масштабах по шкале времени.

Случайный процесс $X(t)$ является самоподобным с параметром H , если $X(t)$ и $a^{-H}X(at)$ имеют одинаковые конечномерные распределения вероятностей для всех $a > 0$. Параметр

$H \in (0.5; 1)$ называется параметром Хёрста и определяет степень выраженности фрактальных свойств. При $H = 0.5$ такие свойства полностью отсутствуют и процесс является классическим; чем ближе H к 1, тем более ярко проявляются фрактальные свойства. С прикладной точки зрения ценность данного подхода к моделированию заключается в том, что он позволяет описать существенно пульсирующий трафик, моменты поступления пакетов которого группируются в пиковые всплески, причем пульсирующий характер наблюдается при различных степенях усреднения процесса по времени.

Самоподобные процессы, в том числе описывающие явления в сетях передачи данных, обладают рядом свойств, существенно отличающих их от потоков, рассматриваемых в классической теории очередей.

1. Долговременная зависимость.

Самоподобные процессы обладают гиперболически затухающей корреляционной функцией вида

$$R(k) = \frac{\sigma^2}{2} ((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}), \quad (1)$$

или

$$R(k) \approx k^{2H-2} L(t) \quad \text{при } k \rightarrow \infty, \quad (2)$$

где $L(t)$ – медленно меняющаяся функция на бесконечности (т.е. $\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1$ для всех $x > 0$).

Следовательно, корреляционная функция является несуммируемой – ряд, образованный последовательными значениями корреляционной функции, расходится.

Это свойство характеризует практически все самоподобные процессы и отличает их от процессов без долговременной зависимости, у которых корреляционная функция убывает по показательному закону и суммируема. Долговременная зависимость является причиной ярко выраженных пульсаций процесса, однако позволяет говорить о некоторой предсказуемости в небольших пределах времени. С точки зрения теории очередей важным следствием коррелированности потока является неприемлемость оценок параметров очереди, основанных на предположении об одинаковом и независимом распределении интервалов во входящем потоке.

2. Медленно убывающая дисперсия.

При усреднении процесса дисперсия выборочного среднего затухает медленнее, чем величина, обратная размеру выборки, по закону:

$$\sigma^2(X^{(m)}) \propto m^{2H-2}, \quad \text{при } m \rightarrow \infty, \quad (3)$$

в то время как для традиционных стационарных случайных процессов

$$\sigma^2(X^{(m)}) = \frac{1}{m} \sigma^2(X), \quad (4)$$

т.е. уменьшается обратно пропорционально объему выборки.

Свойство медленно убывающей дисперсии говорит о возможности существенных, не сглаживаемых усреднением, «выбросов» в случайном процессе и связывает самоподобие с таким понятием, как распределения с весомами хвостами (РВХ). Важное следствие свойства медленно затухающей дисперсии состоит в том, что в случае классических статистических тестов (например, вычисления доверительных интервалов) общепринятая мера среднеквадратического отклонения σ может не быть достоверной.

С данным свойством связано и нехарактерное поведение индекса разброса для отсчетов процесса (IDC), также называемого фактором Фано:

$$F(T) = \frac{\text{Var}[N(T)]}{E[N(T)]}. \quad (5)$$

Здесь $N(T)$ – число событий исследуемого потока, наступивших в интервале (окне) T .

Для самоподобных процессов логарифм индекса разброса $F(T)$ линейно возрастает:

$$\ln[F(T)-1] = (2H-1) \ln T + y. \quad (6)$$

3. *Весомые распределения (распределения с весовыми хвостами)*

Случайная величина Z имеет распределение с весовым хвостом (PBX), если вероятность

$$P[Z > x] \sim cx^\alpha, \quad x \rightarrow \infty, \quad (7)$$

т.е. хвост распределения затухает по степенному закону. Примером распределения с весовым хвостом является распределение Парето.

При $0 < \alpha < 2$ величина Z обладает бесконечной дисперсией, а при $0 < \alpha < 1$ среднее значение также бесконечно.

Наиболее существенной особенностью случайной величины, обладающей распределением с тяжелым хвостом, является чрезвычайная изменчивость. С вероятностью, которая не является пренебрежимо малой, в выборке может присутствовать некоторое число «очень больших» значений. Такие распределения существенно снижают точность статистических оценок: скажем, конечный объем выборки приводит к заниженной оценке среднего и дисперсии.

Наличие PBX во внешних по отношению к рассматриваемым процессам явлениях является одной из причин возникновения самоподобия в соответствующих стохастических моделях.

Часто при рассмотрении самоподобных процессов говорят о комплексе взаимосвязанных понятий: самоподобии, масштабировании, долговременной зависимости, PBX и степенных законах статистических характеристик. Этот комплекс свойств отличает процессы, называемые самоподобными, от классических случайных процессов, например, пуассоновского.

3. Рассматриваемая информационная система

В настоящей работе проведен анализ потока событий прикладного уровня в действующей информационной системе.

Исследовалась система коммерческого ордеринга (COM – Commercial Order Management), интегрированная с биллинговой системой крупного подразделения оператора связи (порядка 9 млн абонентов).

Элементарным событием для системы данного класса является создание или обработка заявки (ордера, «наряда»). Заявка в общем случае инициирует внесение в асинхронном режиме изменений в абонентскую картотеку биллинга: создание или изменение лицевого счета или его отдельных атрибутов, изменение действующего набора услуг (подключение, отключение, блокировка) и т.д.

Заявки могут порождаться как полностью автоматически при обработке периодических операций, таких, например, как контроль дебиторской задолженности, либо поступать от смежных интегрированных систем: CRM (управление взаимоотношениями с клиентом – Customer Relationship Management), дилерские интерфейсы, интерфейсы абонентского самообслуживания и др. В процессе обработки заявка может проходить несколько этапов, которые могут быть как автоматизированными операциями над данными в смежных системах, так и инициировать внешние по отношению к системам действия специалистов: например, экспертное определение технической возможности или выезд монтажника к абоненту.

В рамках исследования рассматривалась только последовательность моментов наступления элементарных событий вышеописанного рода, т.е. экспериментальные данные являются

полностью обезличенными и не содержат какой-либо коммерческой или абонентской информации.

4. Статистические свойства прикладного потока

Пример среза прикладного потока приведен на рис. 1. Здесь и далее в качестве иллюстрации рассматривается 10-часовой срез процесса в течение рабочего дня (22.08.2017). Процесс в указанный период можно считать условно стационарным, внешние по отношению к предмету исследования аномалии (регламентные работы, аварии) отсутствуют.

Диаграмма на рис. 1 отображает число событий на последовательных 10-секундных интервалах.

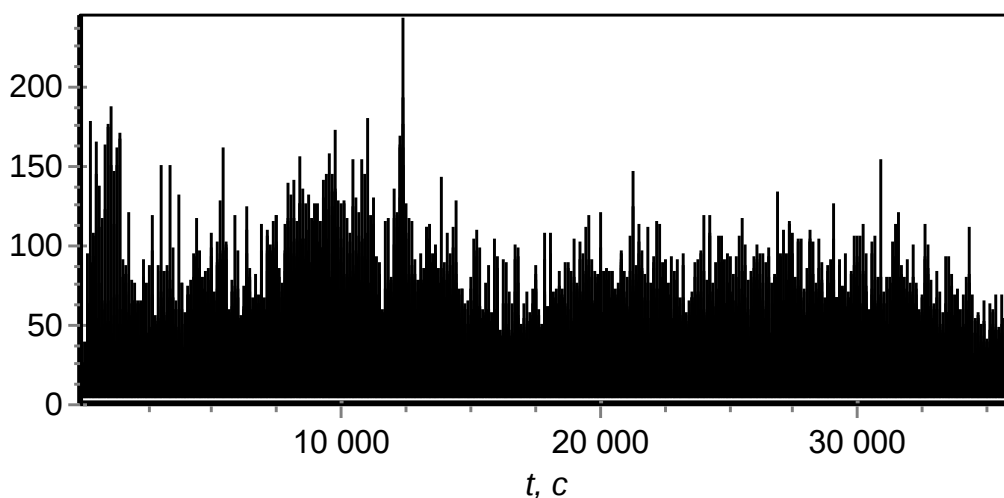


Рис. 1. Срез прикладного потока

Отчетливо заметно наличие существенных всплесков, а также наличие колебаний интенсивности различных масштабов времени, что говорит о наличии внутренней корреляции (последствия) в потоке.

Таблица 1. Основные статистические характеристики среза потока

Характеристика	Значение
Длительность, ч	10
Длительность, с	36000
Число событий	216461
Интенсивность, с^{-1}	6.013

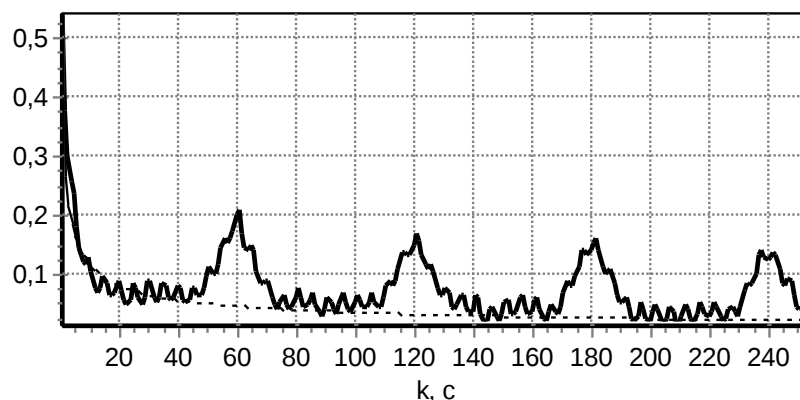


Рис. 2. Автокорреляционная функция (линейный масштаб)

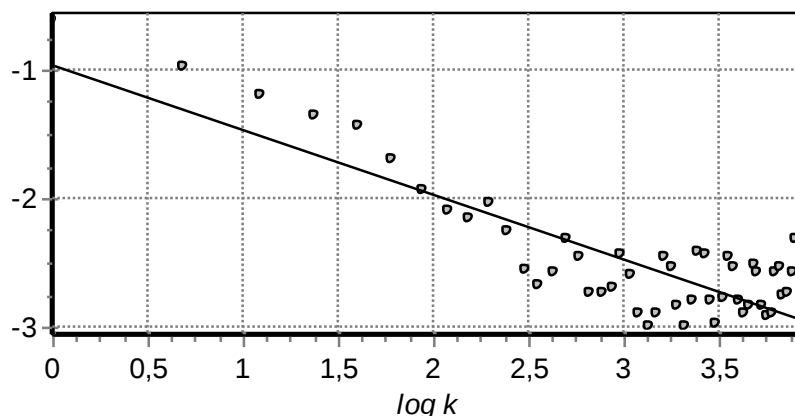


Рис. 3. Автокорреляционная функция (логарифмический масштаб)

Корреляционная структура (рис. 2) показывает наличие как долговременной зависимости в форме гиперболически затухающей кривой АКФ, так и периодической составляющей.

Периодическая составляющая с выраженными пиками для значений, кратных 60 с, вызвана тем фактом, что значительная часть событий в системе порождается автоматическими периодическими процессами обработки, выполняющимися по расписанию – либо ежеминутно, либо в кратные целым минутам заданные моменты времени.

Гиперболическое же затухание АКФ говорит о наличии в рассматриваемом процессе фрактальных свойств. Отсчеты оценок АКФ в логарифмическом масштабе (рис. 3) в целом группируются вокруг наклонной прямой. Для самоподобного процесса в соответствии с (2) угловой коэффициент этой прямой $\beta = 2 - 2H$ напрямую связан с параметром Хёрста. Если путем линейной регрессии оценить этот коэффициент для рассматриваемого случая, получаем $\beta = 0.5047$ и $H = 0.748$, что подтверждает наличие фрактальных свойств. На рис. 2 дополнительно показана кривая для идеального самоподобного процесса с полученным оценочным значением параметра β .

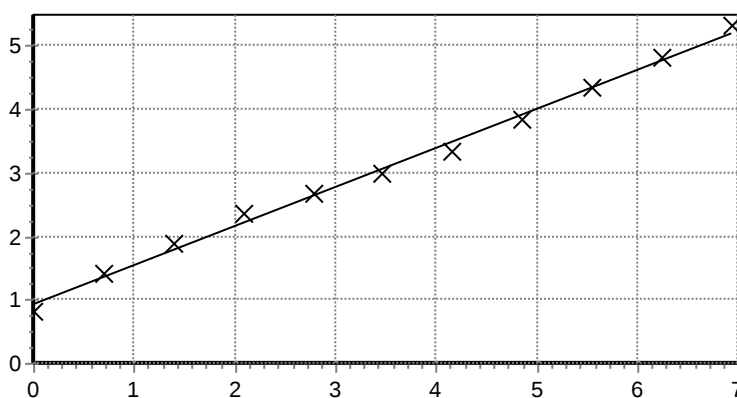


Рис. 4. Логарифм индекса разброса (IDC)

Если рассмотреть график логарифма индекса разброса IDC (рис. 4), то очевидно его линейное возрастание, что характерно для выраженных фрактальных свойств. Угловым коэффициентом соответствующей прямой, оцененный путем линейной регрессии, равен 0.617, что исходя из (6) соответствует параметру Хёрста $H = 0.808$.

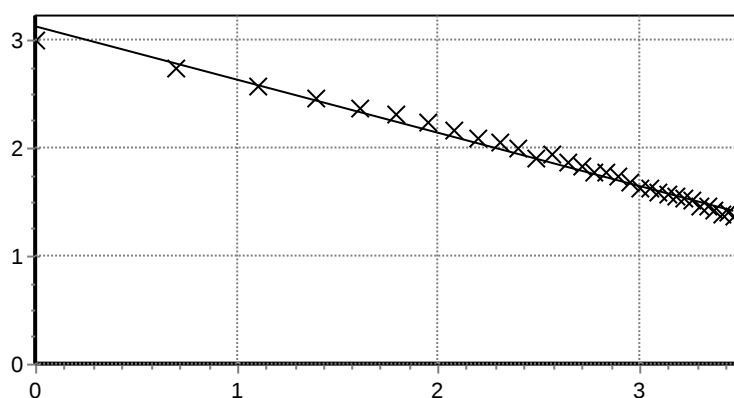


Рис. 5. Убывание дисперсии при усреднении процесса по времени (логарифмический масштаб)

Оценка параметра Хёрста исходя из свойства медленно убывающей дисперсии (рис. 5) на основании (3) дает значение $H = 0.755$, что практически совпадает с оценкой, полученной из анализа автокорреляционной функции.

5. Заключение

Полученные результаты подтверждают наличие фрактальных свойств в потоках событий не только сетевого, но и прикладного уровня, что говорит о том, что самоподобие в распределенных информационных системах не является исключительно сетевым феноменом, порожаемым свойствами сети – протоколов, алгоритмов пакетирования и маршрутизации.

Представляет научный интерес вопрос взаимосвязи статистических характеристик потоков сетевого и прикладного уровня, а также выбор конкретной фрактальной модели, наиболее адекватно описывающей рассматриваемый процесс. Эти направления можно рассматривать как сферу дальнейших исследований.

Литература

1. Кокс Д., Льюис П. Статистический анализ последовательностей событий. Пер. с англ. И. А. Маховой, В. В. Рыкова, ред. Н. П. Бусленко. М.: Мир, 1969. 310 с.
2. Шелухин О. И., Тенякшев А. М., Осин А. В. Фрактальные процессы в телекоммуникациях: монография. М.: Радиотехника, 2003. 480 с.
3. Петров М. Н., Треногин Н. Г. Распределенные информационные системы управления в телекоммуникационной отрасли. Красноярск, 2006. 345 с.
4. Городецкий А. Я., Заборовский В. С. Информатика. Фрактальные процессы в компьютерных сетях: учеб. пособие. СПб: Изд-во СПбГТУ, 2000. 102 с.
5. Leland W. E., Taqqu M. S., Willinger W., Wilson D. W. On the self-similar nature of Ethernet traffic // Computer Communications Review. 1993. V. 23. P. 193–198.
6. Ryu B. K. Fractal Network Traffic: From Understanding to Implications. Ph.D. thesis. Columbia University, 1996. 143 p.
7. Треногин Н. Г., Соколов Д. Е. Фрактальные свойства сетевого трафика в клиент-серверной информационной системе // Вестник НИИ СУВПТ. Сборник научных трудов. 2003. В. 14. С. 163–172.

8. Треногин Н. Г., Соколов Д. Е. Моделирование сетевого трафика в информационных системах на основе фрактального точечного процесса // Вестник университетского комплекса: сб. научных трудов. 2004. В. 2 (16). С. 12–21.
9. Соколов Д. Е., Треногин Н. Г. Фрактальные свойства трафика в действующей двухзвенной системе обработки данных // Современные проблемы информатизации в технике и технологиях: Сб. трудов. Вып. 10. Воронеж: Научная книга, 2005. С. 264.
10. Треногин Н. Г., Петров М. Н., Соколов Д. Е. Свойства фрактального трафика при прохождении системы массового обслуживания с очередью // Вестник СибГАУ. 2017. Т. 18, № 1. С. 105–110.

Статья поступила в редакцию 12.09.2017

Треногин Николай Геннадьевич

к.т.н., доцент кафедры телекоммуникационных сетей и вычислительных средств СибГУТИ, директор департамента развития и эксплуатации информационных систем и платформ макрорегионального филиала «Сибирь» ПАО «Ростелеком», e-mail: nikolay.g.trenogin@sibir.rt.ru.

Соколов Дмитрий Евгеньевич

начальник отдела создания и продвижения биллинговых решений центра компетенций по телекоммуникационным решениям «Старт» макрорегионального филиала «Сибирь» ПАО «Ростелеком», e-mail: dmitrij.sokolov@sibir.rt.ru.

Fractal properties of events flow of application level in information systems

N. Trenogin, D. Sokolov

The results of estimating statistical characteristics of events flow of application level in operating information system of telecommunications operator, allowing to draw a conclusion about the presence of fractal properties in the process are given.

Keywords: fractal processes, self-similarity, load simulation.