

ИССЛЕДОВАНИЕ И ВЫБОР МАТЕМАТИЧЕСКОЙ МОДЕЛИ СЕТЕВОГО ТРАФИКА

А.В. Черниговский

chernigovsky.alex@gmail.com

М.В. Кривов

vmk@angtu.ru

А.Л. Истомин

a.l.istomin@mail.ru

Ангарский государственный технический университет,
Ангарск, Иркутская обл., Российская Федерация

Аннотация

Проведенное исследование посвящено изучению различных видов сетевого трафика в целях получения математического описания не только отдельного вида трафика, но и общего трафика сети. Охарактеризованы основные виды данных, передаваемых в ходе работы сети, выполнено их сравнение с наиболее распространенными математическими моделями — Пуассона, Парето, Вейбулла, экспоненциального и логнормального распределения. Выявлено, что независимо от вида трафика распределение объема передаваемых пакетов данных имеет «длинный хвост» и хорошо описывается моделью логнормального распределения. Проведенная оценка автокорреляционной функции показала, что практически все данные характеризуются медленно убывающей зависимостью. Это свидетельствует об их самоподобности. Данный вывод также подтвержден путем расчета параметра Херста. Установлено, что степень самоподобия зависит не только от типа передаваемых данных, но и от их соотношения в общем трафике сети. Для сравнения математического описания трафика выбраны такие модели, как классическое и фрактальное броуновское движение, а также модели *AR*, *MA*, *ARMA* и *ARIMA*. Расчеты показали, что среди данных моделей наиболее точное математическое описание трафика сети может быть получено с использованием модели фрактального броуновского движения

Ключевые слова

Сетевой трафик, самоподобие, параметр Херста, фрактальное броуновское движение, модель ARMA, модель ARIMA

Поступила 09.01.2020

Принята 06.04.2020

© Автор(ы), 2020

Введение. С развитием сферы информационных технологий появилась потребность в организации передачи большого объема различных данных. В свою очередь, это привело к развитию сетевых технологий. Однако

с расширением возможностей сетей передачи данных нагрузка на сети также возросла, вследствие чего возникла необходимость разработки программных и аппаратных средств для контроля и регулирования сетевой активности [1].

Одной из основных проблем для решения указанной задачи стало отсутствие универсальной математической модели, способной адекватно описать структуру сетевого трафика. В основном это связано с увеличением объемов передаваемых данных, а также с применением различных механизмов их передачи (различные протоколы, способы маршрутизации и т. д.) [2]. Большинство математических моделей, применяемых в настоящее время, может использоваться для описания отдельных видов передаваемых данных, а общей теории трафика до сих пор не существует.

В рамках теории массового обслуживания передаваемые пакеты данных рассматриваются как случайные события, не связанные друг с другом. Наиболее распространенной является классическая модель Пуассона, в которой поток данных рассматривается как независимая случайная величина, экспоненциально зависящая от времени сеанса t :

$$f(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

где $f(t)$ — плотность распределения; λ — средняя интенсивность потока за сеанс; $n = 0, 1, 2, \dots$ — число потоков трафика (событий) за период времени t .

Эта модель является достаточно простой и может применяться для анализа трафика, содержащего незначительный объем данных [3]. Однако в связи с расширением возможностей сетей передачи данных, а также с необходимостью передачи данных различных видов (аудио, видео и др.) выявлено, что модель Пуассона не подходит для их описания ввиду низкой адекватности получаемых результатов [4, 5].

Попытки модификации распределения Пуассона привели к появлению нескольких математических моделей, в рамках которых могли быть описаны отклонения в структуре трафика от стандартной модели Пуассона, например:

– ON/OFF-модель, согласно которой трафик представляет собой прерывающийся пуассоновский процесс [6]. Отличительная особенность — учет периодов неактивного состояния системы (когда пакеты данных не передаются), что позволяет достаточно хорошо описывать процесс коммутации пакетов;

– модель Пуассона, модулированная марковским процессом, которая позволяет учитывать не только неактивное состояние системы, но и число активных на данный момент пользователей [7, 8]. В этом случае поток данных от одного пользователя представляет собой прерывающийся пуассоновский процесс, а результирующий поток рассматривается как марковский процесс.

Особенностью перечисленных моделей является то, что пакеты передаваемых данных рассматриваются в них как случайные события, независимые друг от друга, т. е. размер каждого последующего пакета данных не зависит от размера предыдущего.

Тезис о взаимной независимости поступающих пакетов данных опровергнут в [9], где сформулированы основы теории самоподобия. Согласно этой теории, при работе сети имеет место взаимозависимость не только между отдельными пакетами данных, но и между целыми фрагментами трафика. В результате был создан принципиально новый класс моделей трафика — самоподобные модели.

Отличительная особенность самоподобных процессов — наличие у их распределений «тяжелых хвостов» и медленно убывающей зависимости автокорреляционной функции (АКФ). Такие распределения достаточно хорошо описываются моделями Парето и Вейбулла [10]. Для них характерно наличие степенной зависимости функции плотности от скорости передачи данных, однако вид функции для них несколько различается:

– для модели Парето $f(t) = \beta \alpha^\beta t^{-\beta-1}$;

– для распределения Вейбулла $f(t) = \alpha \beta^{-\alpha} t^{\alpha-1} e^{-(t/\beta)^\alpha}$, где α, β — коэффициенты уравнений.

Одной из наиболее распространенных математических моделей для самоподобных процессов является классическое броуновское движение (БД), а также его аналог — модель фрактального броуновского движения (ФБД). В этом случае поток событий рассматривается как случайная величина, которая зависит не только от времени, но и от величины предыдущего потока:

$$X_t = X_{t-1} + \varepsilon_t.$$

Здесь X_t — значение случайной величины в момент времени t ; ε_t — белый шум.

Плотность распределения приращений такой величины подчиняется закону Гаусса [11], для которого математическое ожидание равно нулю, а дисперсия —

$$\sigma^2 (X_2 - X_1) = C (t_2 - t_1)^{2H},$$

где X_1, X_2 — значения случайной величины в моменты времени t_1 и t_2 ; C — некоторая положительная константа; H — параметр Херста.

Параметр Херста позволяет оценивать степень самоподобия трафика. При $0,5 \leq H \leq 1,0$ процесс является строго самоподобным и описывается моделью ФБД, при $H = 0,5$ — моделью классического БД, а при $0 \leq H < 0,5$ имеет стохастический характер [12].

Еще одна группа моделей, с использованием которых возможно моделирование самоподобных процессов, это так называемые модели временных рядов. Наиболее распространенными из них являются модели авторегрессии и скользящего среднего.

Модель авторегрессии *AR* предполагает, что текущее значение функции определяется линейной зависимостью от нескольких предыдущих значений этой функции, а погрешность — как белый шум:

$$X_t = C + \sum_{i=1}^p \alpha_i X_{t-i} + \varepsilon_t,$$

где p — порядок модели *AR*; α_i — коэффициенты уравнения авторегрессии; X_{t-i} — лаговый оператор.

В свою очередь, модель скользящего среднего *MA* предполагает, что значение функции в данный момент времени колеблется около некоторого среднего значения, а величина отклонения зависит от значения функции в предыдущие периоды времени:

$$X_t = C + \varepsilon_t + \sum_{i=1}^q \beta_i X_{t-i}.$$

Здесь q — порядок модели *MA*; β_i — коэффициенты уравнения скользящего среднего.

На базе моделей *AR* и *MA* сформированы два вида моделей:

1) модель *ARMA*, в которой предыдущие значения функции влияют не только на текущее значение функции, но и на его отклонение:

$$X_t = C + \sum_{i=1}^p \alpha_i X_{t-i} + \varepsilon_t + \sum_{i=1}^q \beta_i \varepsilon_{t-i},$$

где ε_{t-i} — лаговый оператор ошибки;

2) модель *ARIMA* (модель Бокса — Дженкинса, или модель интегрированной авторегрессии — скользящего среднего), которая позволяет работать с зависимостями, имеющими тренд:

$$X_t = C + \sum_{i=1}^p \alpha_i \Delta^d X_{t-i} + \varepsilon_t + \sum_{i=1}^q \beta_i \Delta^d \varepsilon_{t-i},$$

где d — порядок модели *ARIMA*, характеризующий степень интегрирования; Δ^d — оператор взятия конечной разности порядка d .

Несмотря на то что модели временных рядов получили довольно широкое распространение при изучении трафика, задача получения математического описания трафика все еще остается достаточно сложной. В том числе это связано с тем, что:

- сетевой трафик, как правило, сочетает в себе различные виды данных, которые описываются разными моделями;
- в зависимости от типа передаваемых данных степень самоподобия трафика может варьироваться [13].

В связи с этим актуальной задачей является анализ сетевого трафика, а также подбор математической модели, способной адекватно описывать различные виды трафика и их сочетания.

Цель работы — статистический анализ различных видов сетевого трафика, а также получение их математического описания с тем, чтобы адекватно описать не только передачу отдельных видов данных, но и трафик с комбинированной структурой.

Материалы и методы решения задач. Для выполнения исследования использовались данные, полученные на базе сети Ангарского государственного технического университета. Для анализа собраны данные по различным видам трафика:

- 1) потоковое видео (протокол *UDP*);
- 2) *YouTube*-трафик (протокол *QUIC*);
- 3) трафик смешанного типа, включающий в себя одновременную передачу данных *YouTube* и потокового видео;
- 4) *torrent*-трафик (протоколы *GVSP*, *UDP*, *BitTorrent*);
- 5) комбинированный трафик, характеризующийся сочетанием различных видов трафика.

Сбор данных осуществлялся как с отдельных компьютеров, так и со всей сети в целом.

Поступающий с центральных коммутаторов сетевой трафик дублировался на отдельный порт, откуда осуществлялась передача данных на компьютер с работающим анализатором трафика *WireShark*. Далее проводилась группировка полученной информации либо по видам трафика, либо по адресам отдельных ПК. Полученные таким образом данные обрабатывались в пакете *MATLAB* согласно методике, приведенной в [14].

Процедура обработки сводилась к следующему. Для получения данных, более однородных по времени, проводилась процедура их группировки путем вычисления среднего значения объема пакетов (Байт), передаваемых за определенный период времени, с: 0,1; 0,5; 1,0; 5,0. Далее на основе сгруппированных данных рассчитывались относительные частоты распределения скорости передачи данных (Байт/с). Затем осуществлялось сравнение полученных функций плотности распределений со стандартными моделями. В качестве моделей сравнения выбраны: классическая модель Пуассона, модели логнормального и экспоненциального распределения, а также модели Парето и Вейбулла, характеризующие распределение с «тяжелым хвостом».

Проверка на самоподобность трафика включала в себя расчет АКФ, определение типа зависимости между членами ряда данных (медленно или быстро убывающая зависимость) [14], а также нахождение параметра Херста по R/S -методу [15].

Автокорреляционная функция определялась путем нахождения коэффициента корреляции между исходным рядом данных и рядом, в котором все элементы смещены на один шаг. Затем методом наименьших квадратов находились коэффициенты для экспоненциальной и степенной зависимости, к которым АКФ наиболее близка.

Если АКФ может быть описана экспоненциальной функцией, то имеет место быстро убывающая зависимость (БУЗ). Если функция носит степенной характер, то речь идет о медленно убывающей зависимости (МУЗ), которая характерна для самоподобных процессов. Оценка типа зависимости проводилась с помощью показателя LRD , который рассчитывался как отношение дисперсий разностей фактического значения АКФ и расчетного, полученного для степенной ($\sigma_{МУЗ}$) и экспоненциальной ($\sigma_{БУЗ}$) моделей.

Задача подбора модели для описания сетевого трафика заключалась в сравнении эмпирически полученных данных с моделями временных рядов. В качестве таких моделей выбраны модели БД, ФБД, а также AR , MA , $ARMA$ и $ARIMA$. Для каждой модели подобраны параметры уравнений, на основе которых далее получены расчетные значения скоростей передачи данных. Сравнение фактических и расчетных значений скорости осуществлялось с помощью коэффициента детерминации. Процедура повторялась для различных порядков моделей (p и q), после чего рассматривалась зависимость коэффициента детерминации от соответствующего порядка модели.

Результаты. Были проанализированы данные, собранные для различных видов трафика — мультимедийного, пирингового и потокового (видео-

конференции, вебинары). Данные для анализа трафика получены как для отдельных ПК, так и для сети в целом.

Ввиду неравномерности появления запросов в сети по времени проводилась группировка полученных данных с различными временными интервалами, с: 0,1; 0,5; 1,0; 5,0. Для сгруппированных данных рассчитывалась средняя скорость передачи данных. Далее выполнялась оценка плотности распределения полученных значений скорости (рис. 1).

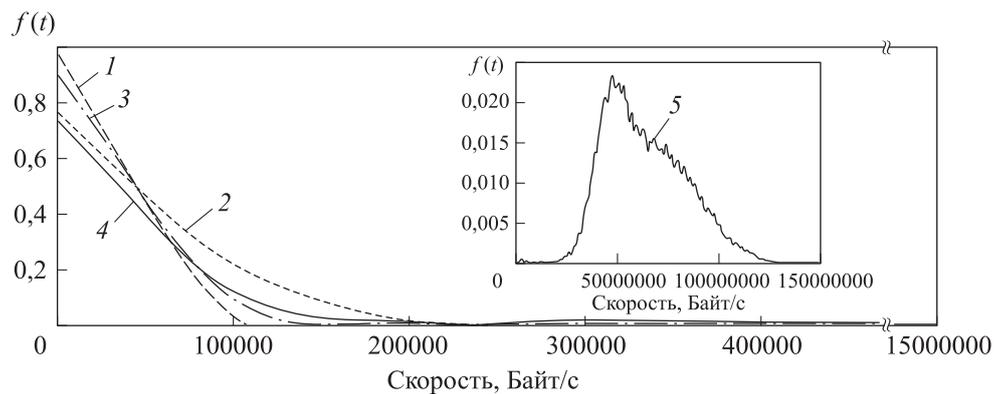


Рис. 1. Плотность распределения скоростей передачи данных (временной интервал 0,1 с) для *YouTube*-трафика (1), потокового видео (2), трафика смешанного типа (потоковое видео + *YouTube*-трафик) (3), комбинированного (4) и *torrent*-трафика (5)

Вид полученных распределений позволяет предположить, что они могут иметь «тяжелый хвост».

Далее на основании экспериментальных данных определялись параметры изучаемых моделей для получения уравнения функции плотности распределения. Для каждой модели рассчитывались коэффициенты детерминации. Полученные результаты приведены в табл. 1.

Таблица 1

Сравнение математических моделей для различных видов сетевого трафика (временной интервал 0,1 с)

Вид передаваемых данных	Коэффициент детерминации R^2 для распределения					LRD	H
	Пуассона	экспоненциального	логнормального	Парето	Вейбулла		
Потоковое видео (UDP)	0,880	0,271	0,985	0,732	0,553	0,5301	0,7528

Вид передаваемых данных	Коэффициент детерминации R^2 для распределения					LRD	H
	Пуассона	экспоненциального	логнормального	Парето	Вейбулла		
<i>YouTube (QUIC)</i>	–	0,212	0,974	0,908	0,825	0,9569	0,2891
Смешанный (потокоевое видео + <i>YouTube</i>)	–	0,258	0,9401	0,908	0,917	0,8997	0,4586
<i>Torrent (GVSP, UDP, BitTorrent)</i>	0,012	–	0,957	0,0070	0,038	0,553	0,8057
Комбинированный	0,857	0,489	0,980	0,958	0,920	2,173	0,8251

Для проверки трафика на самоподобность рассчитывалась АКФ, которая отражает зависимость текущих значений скорости от ее предыдущих (рис. 2). При этом зависимость между передаваемыми пакетами данных наблюдается для всех изученных видов трафика. Однако для *YouTube*-трафика и смешанного трафика эта зависимость выражена слабее.

Следующим шагом стала оценка типа зависимости между членами ряда данных. Для этого рассчитывался показатель LRD , значение которого должно быть менее 1 для МУЗ и более 1 для БУЗ.

Рассчитанное для различных видов трафика значение показателя LRD (см. табл. 1) свидетельствует о том, что для большинства типов данных АКФ характеризуется наличием МУЗ. Для комбинированного трафика наблюдается колебательный характер АКФ (период колебаний составляет 8 с), что несколько снижает точность расчета.

Проверка самоподобности трафика также оценивалась с помощью параметра Херста H [15]. Результаты расчета параметра Херста R/S -методом приведены в табл. 1.

Далее выполнялось сравнение экспериментальных данных с моделями временных рядов: БД, ФБД, AR , MA , $ARMA$, $ARIMA$. Для каждой модели подобраны соответствующие порядки (p и q), при которых она показывала наилучшую надежность. Зависимости коэффициента детерминации от порядка модели приведены на рис. 3.

Обсуждение полученных результатов. При сравнении полученных функций плотности для различных видов трафика с классической

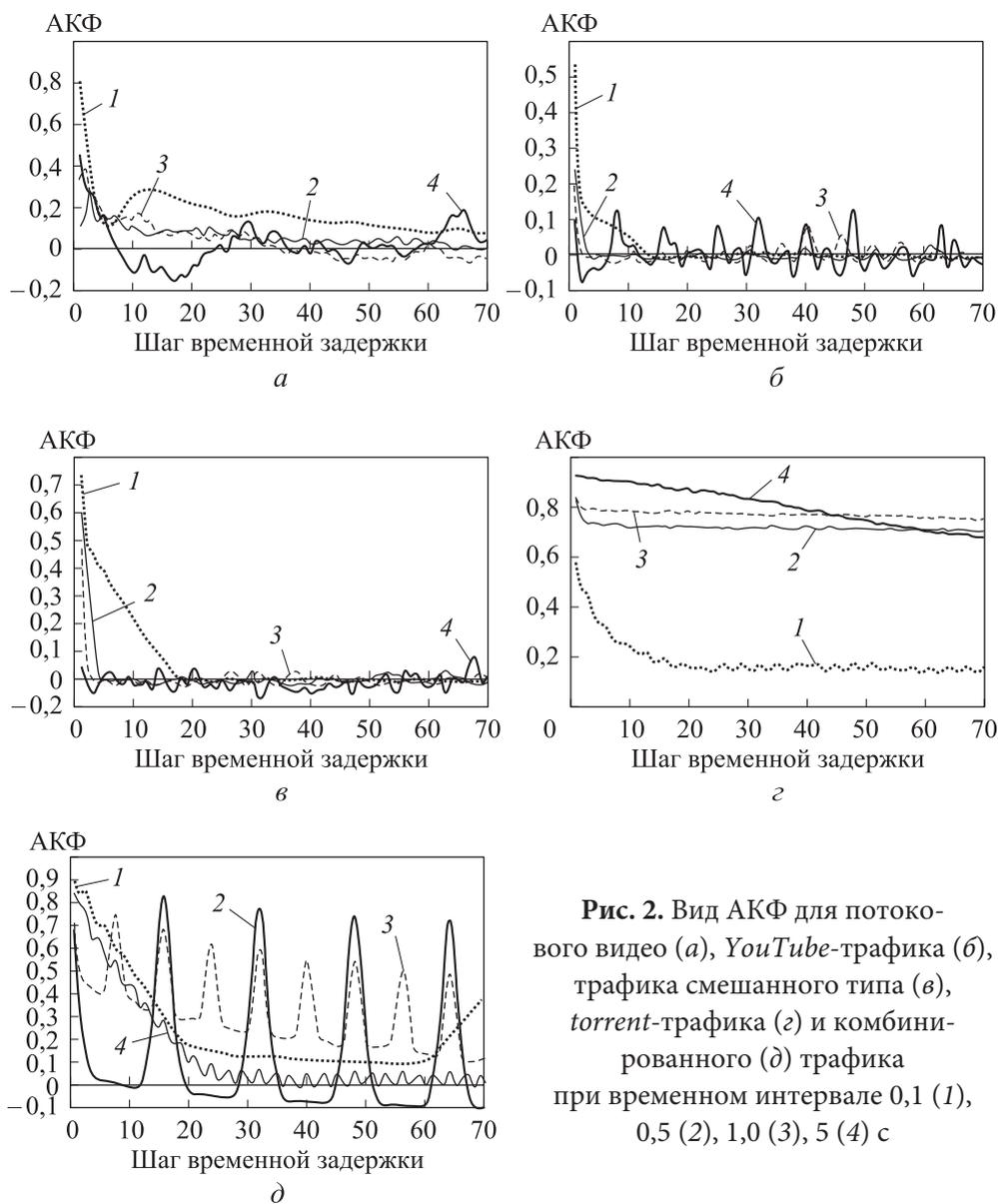


Рис. 2. Вид АКФ для потокового видео (*a*), YouTube-трафика (*б*), трафика смешанного типа (*в*), torrent-трафика (*г*) и комбинированного (*д*) трафика при временном интервале 0,1 (1), 0,5 (2), 1,0 (3), 5 (4) с

моделью Пуассона (см. табл. 1) выявлено, что адекватность полученной модели снижается с увеличением времени работы сети. Это связано с тем, что вид функции плотности распределения для модели Пуассона существенно зависит от таких параметров, как время сеанса связи, интервал между пакетами, а также от типа передаваемых данных.

Ввиду того, что распределение Пуассона является практически симметричным относительно среднего значения скорости процесса, обозначаемого как λ , а значения рассчитываемого параметра изменяются в пределах $\pm \lambda$,

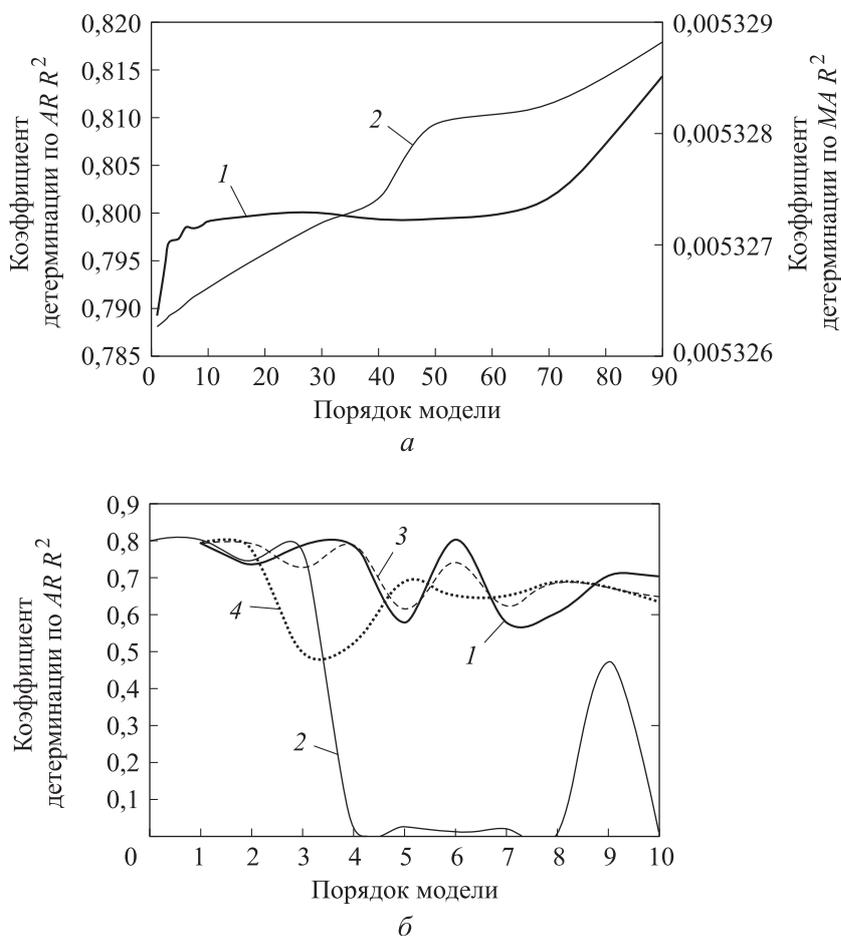


Рис. 3. Зависимости коэффициента детерминации от порядка модели:
 a — $AR(p, 0, 0)$ (1), $MA(0, 0, q)$ (2); b — $ARMA(p, 0, 1)$ (1) и $(6, 0, q)$ (2),
 $ARIMA(p, 1, 1)$ (3) и $(10, 1, q)$ (4)

с его помощью невозможно описать группировку данных в области низких значений скорости, а также наличие «длинного хвоста». Это не позволяет использовать модель Пуассона для моделирования сетевого трафика системы.

Модели распределения с «тяжелыми хвостами» показали достаточно высокую адекватность математического описания только при малом временном интервале (0,1 с), а для больших временных интервалов (1...5 с) такие модели оказались неприменимы. Это может быть обусловлено тем, что эти модели являются монотонно убывающими, а экспериментально полученные данные характеризуются наличием «возмущений», и чем больше временной интервал, тем сильнее выражены эти «возмущения». Аналогичный вывод может быть сделан для экспоненциального распределения.

Коэффициент детерминации, рассчитанный для функции плотности логнормального распределения (см. табл. 1), свидетельствует о достаточно высокой степени достоверности полученного математического описания, а также позволяет применять данную модель для описания различных видов сетевого трафика.

Анализируя вид АКФ, полученных для различных видов трафика (см. рис. 1), можно отметить, что *YouTube*- и смешанный трафики характеризуются кратковременной зависимостью: значение АКФ для них приближается к нулю менее чем после 15 шагов временной задержки. При этом для остальных видов трафика (потокное видео, *torrent*-трафик, комбинированный) зависимость носит более долговременный характер.

При сравнении АКФ потокное видео, *YouTube*- и смешанного трафиков можно отметить, что значения АКФ смешанного трафика несколько выше, чем *YouTube*-трафика, но меньше, чем у потокное видео.

Особенность АКФ комбинированного трафика — наличие выраженной циклической структуры с периодом колебаний 8 с. Снижение интенсивности колебаний при увеличении временного интервала связано с усреднением значений скорости в процессе процедуры агрегирования.

Данные результаты позволяют сделать вывод о наличии автокорреляции для всех типов передаваемых данных, а также о ее зависимости от соотношения видов трафика, входящих в состав общего потока данных.

Результаты, полученные при оценке показателей *LRD* и *H* (см. табл. 1), свидетельствуют о том, что для потокное видео, *torrent*- и комбинированного трафика характерно наличие фрактальности между членами ряда, а для *YouTube*-трафика степень самоподобия довольно низка. Это подтверждает сделанные ранее выводы о его несамоподобности. Интересно, что значение параметра Херста для смешанного трафика находится в интервале значений между результатами видео и *YouTube*-трафика. На основании этого можно предположить следующее: на степень самоподобия влияет не только тип передаваемых данных, но и их соотношение в общем объеме данных.

При изучении моделей временных рядов выявлено следующее. Для модели *AR* (рис. 3, а) в целом точность возрастает при увеличении числа элементов ряда. Модель *MA* также характеризуется увеличением сходимости результатов при увеличении числа шагов, однако коэффициент детерминации в таком случае не превышает 0,0055.

Сравнение экспериментальных данных с рассчитанными по модели *ARMA* показало, что увеличение шага по любому из параметров (*p*, *q*) приводит к снижению точности (рис. 3, б).

Интегрирование исходного ряда данных и расчет модели *ARIMA* привели к незначительному снижению сходимости результатов (см. рис. 3, б). Это можно объяснить отсутствием какого-либо тренда (возрастающего или убывающего) в исходном временном ряде и, как следствие, снижением надежности результатов подобного усреднения.

Для сравнения различных моделей прогнозирования трафика выбраны те параметры уравнений, при которых данная модель показала наилучшую сходимость результатов (табл. 2). Наилучшая сходимость результатов наблюдается для модели *AR*. Однако для использования такой модели в режиме реального времени требуются большие вычислительные мощности, связанные со сложностью уравнения модели. В свою очередь, модель ФБД характеризуется меньшими затратами вычислительной мощности при незначительном снижении точности (менее 1 %). Это делает ее более привлекательной не только для моделирования трафика, но и для внедрения в различные системы его регулирования. Таким образом, путем оценки полученных результатов можно сделать вывод о том, что для прогнозирования сетевого трафика наиболее подходящей является модель ФБД.

Таблица 2

Сравнение моделей прогнозирования трафика
(для комбинированного трафика, временной интервал 0,1 с)

Модель	p	d	q	R^2
<i>AR</i>	90	0	0	0,8144
<i>MA</i>	0	0	90	0,00533
<i>ARMA</i>	6	0	1	0,7973
<i>ARIMA</i>	10	1	1	0,7972
БД	1	0	0	0,7895
ФБД	1	0,8251	0	0,8067

Заключение. Проведенные исследования показали, что выбор математической модели сетевого трафика зависит от вида передаваемых данных, которые присутствуют в его составе, а также от их соотношения. В большинстве случаев распределение данных подчиняется логнормальному закону.

Для всех исследованных видов трафика распределение данных имеет «тяжелый хвост» и характеризуется медленно убывающей зависимостью. Это, с учетом рассчитанных значений параметра Херста, свидетельствует о самоподобии изученных видов трафика (кроме *YouTube*), а также о зависимости степени самоподобия от соотношения данных различных видов.

Результаты анализа моделей, применяемых для прогнозирования трафика, показали следующее: для описания сетевого комбинированного трафика может быть применена модель ФБД.

Полученные результаты можно использовать для математического моделирования сетевого трафика различных видов, в том числе комбинированного, а также для разработки систем его регулирования, включая системы на основе нейросетевых технологий.

ЛИТЕРАТУРА

- [1] Черниговский А.В., Кривов М.В. Анализ методов распределения сетевого трафика между пользователями сети. *Вестник Ангарского государственного технического университета*, 2018, т. 1, № 12, с. 168–173.
DOI: <https://doi.org/10.36629/2686-777X-2018-1-12-168-173>
- [2] Chen Th.M. Network traffic modeling. In: *The Handbook of Computer Networks*. Vol. 3. Wiley, 2007, pp. 326–339.
- [3] Ihler A., Hutchins J., Smyth P. Learning to detect events with Markov-modulated Poisson processes. *ACM Trans. Knowl. Discov. Data*, 2007, vol. 1, no. 3, art. 13.
DOI: <https://doi.org/10.1145/1297332.1297337>
- [4] Paxson V., Floyd S. Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Trans. Netw.*, 1995, vol. 3, iss. 3, pp. 226–244. DOI: <https://doi.org/10.1109/90.392383>
- [5] Сидорова О.И. Математические модели трафика в современных телекоммуникационных системах. Автореф. дис. ... канд. физ.-мат. наук. Тверь, ТвГУ, 2009.
- [6] Сидорова О.И. ON/OFF-модель трафика с неоднородными источниками в режиме «быстрого роста числа соединений». *Вестник ТвГУ. Серия: Прикладная математика*, 2015, № 2, с. 75–94.
- [7] Heffes H., Lucantoni D. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE J. Sel. Areas Commun.*, 1986, vol. 4, iss. 6, pp. 856–868.
DOI: <https://doi.org/10.1109/JSAC.1986.1146393>
- [8] Singh L.N., Dattatreya G.R. A novel approach to parameter estimation in Markov-modulated Poisson processes. *Proc. IEEE (ETC)*, 2004, pp. 1–6.
- [9] Leland W.E., Taqqu M.S., Wilinger W., et al. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Netw.*, 1994, vol. 2, iss. 1, pp. 1–15.
DOI: <https://doi.org/10.1109/90.282603>
- [10] Adas A. Traffic models in broadband networks. *IEEE Commun. Mag.*, 1997, vol. 35, iss. 7, pp. 82–89. DOI: <https://doi.org/10.1109/35.601746>
- [11] Петров В.В. То, что вы хотели знать о самоподобном телетрафике, но стеснялись спросить. *pi.314159.ru: веб-сайт*.
URL: <http://pi.314159.ru/petroff1.pdf> (дата обращения: 03.01.2020).
- [12] Едемская Е.Н., Бельков Д.В. Исследование сетевого трафика с помощью функции Херста. *Информатика и кибернетика*, 2015, № 2, с. 39–46.

[13] Федорова М.Л., Леденева Т.М. Об исследовании свойства самоподобия трафика мультисервисной сети. *Вестник ВГУ. Серия: Системный анализ и информационные технологии*, 2010, № 1, с. 46–54.

[14] Петров В.В. Статистический анализ сетевого трафика. *pi.314159.ru: веб-сайт*. URL: <http://pi.314159.ru/petroff2.pdf> (дата обращения: 03.01.2020).

[15] Воробьев О.В., Можаяев А.А., Осколков А.П. Анализ моделей описания трафика. *Наука і техніка Повітряних Сил Збройних Сил України*, 2014, № 2 (15), с. 170–172.

Черниговский Александр Валерьевич — аспирант кафедры «Вычислительные машины, комплексы, системы и сети» факультета технической кибернетики Ангарского государственного технического университета (Российская Федерация, 665835, Иркутская обл., Ангарск, ул. Чайковского, д. 60).

Кривов Максим Викторович — канд. техн. наук, заведующий кафедрой «Вычислительные машины, комплексы, системы и сети» факультета технической кибернетики Ангарского государственного технического университета (Российская Федерация, 665835, Иркутская обл., Ангарск, ул. Чайковского, д. 60).

Истомин Андрей Леонидович — д-р техн. наук, профессор кафедры «Вычислительные машины, комплексы, системы и сети» факультета технической кибернетики Ангарского государственного технического университета (Российская Федерация, 665835, Иркутская обл., Ангарск, ул. Чайковского, д. 60).

Просьба ссылаться на эту статью следующим образом:

Черниговский А.В., Кривов М.В., Истомин А.Л. Исследование и выбор математической модели сетевого трафика. *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение*, 2020, №3 (132), с. 84–99.

DOI: <https://doi.org/10.18698/0236-3933-2020-3-84-99>

INVESTIGATING NETWORK TRAFFIC AND SELECTING A MATCHING MATHEMATICAL MODEL

A.V. Chernigovskiy

chernigovsky.alex@gmail.com

M.V. Krivov

vmk@angtu.ru

A.L. Istomin

a.l.istomin@mail.ru

Angarsk State Technical University, Angarsk, Irkutsk Region, Russian Federation

Abstract

The investigation aimed to study various network traffic types so as to derive a mathematical description not only for a specific type of traffic, but also for the aggregated network traffic. We characterized the main types of data transmitted during network operation and

Keywords

Network traffic, self-similarity, Hurst parameter, fractal Brownian motion, ARMA model, ARIMA model

compared the results with the most common mathematical models, that is, Poisson, Pareto, Weibull, exponential and lognormal distributions. We established that regardless of traffic type the volume distribution of data packets transmitted has a “long tail” and is well described by the lognormal distribution model. We evaluated the autocorrelation function, which showed that a long-range dependence characterises virtually all data, which indicates their self-similarity. We also confirmed this conclusion by calculating the Hurst exponent. At the same time, we determined that the degree of self-similarity depends not only on the type of data transmitted, but also on the data ratio in the aggregated network traffic. We selected the following models so as to compare the mathematical descriptions of traffic: classical and fractal Brownian motion, and the AR, MA, ARMA and ARIMA models. The results showed that the fractal Brownian motion model provides the most accurate mathematical description of network traffic

Received 09.01.2020

Accepted 06.04.2020

© Author(s), 2020

REFERENCES

- [1] Chernigovskiy A.V., Krivov M.V. Methods of network traffic distribution between users. *Bulletin of the Angarsk State Technical University*, 2018, vol. 1, iss. 12, pp. 168–173 (in Russ.). DOI: <https://doi.org/10.36629/2686-777X-2018-1-12-168-173>
- [2] Chen Th.M. Network traffic modeling. In: *The Handbook of Computer Networks*. Vol. 3. Wiley, 2007, pp. 326–339.
- [3] Ihler A., Hutchins J., Smyth P. Learning to detect events with Markov-modulated Poisson processes. *ACM Trans. Knowl. Discov. Data*, 2007, vol. 1, no. 3, art. 13. DOI: <https://doi.org/10.1145/1297332.1297337>
- [4] Paxson V., Floyd S. Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Trans. Netw.*, 1995, vol. 3, iss. 3, pp. 226–244. DOI: <https://doi.org/10.1109/90.392383>
- [5] Sidorova O.I. Matematicheskie modeli trafika v sovremennykh telekommunikatsionnykh sistemakh. Avtoref. dis. kand. fiz.-mat. nauk [Mathematical traffic models in modern telecommunication systems. Abs. Cand. Sc. Phys.-Math. Diss.]. Tver, TSU Publ., 2009.
- [6] Sidorova O.I. ON/OFF-model' trafika s neodnorodnymi istochnikami v rezhime “bystrogo rosta chisla soedineniy” [ON/OFF model with heterogenous sources under fast growth condition]. *Vestnik TvGU. Seriya: Prikladnaya matematika* [Herald of Tver State University. Series: Applied Mathematics], 2015, no. 2, pp. 75–94 (in Russ.).
- [7] Heffes H., Lucantoni D. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE J. Sel. Areas Commun.*, 1986, vol. 4, iss. 6, pp. 856–868. DOI: <https://doi.org/10.1109/JSAC.1986.1146393>

- [8] Singh L.N., Dattatreya G.R. A novel approach to parameter estimation in Markov-modulated Poisson processes. *Proc. IEEE (ETC)*, 2004, pp. 1–6.
- [9] Leland W.E., Taqqu M.S., Wilinger W., et al. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Netw.*, 1994, vol. 2, iss. 1, pp. 1–15.
DOI: <https://doi.org/10.1109/90.282603>
- [10] Adas A. Traffic models in broadband networks. *IEEE Commun. Mag.*, 1997, vol. 35, iss. 7, pp. 82–89. DOI: <https://doi.org/10.1109/35.601746>
- [11] Petrov V.V. To, chto vy khoteli znat' o samopodobnom teletrafike, no stesnyalis' sprositi' [What you wanted to know about self-similar network traffic, but were embarrassed to ask]. *pi.314159.ru: website* (in Russ.).
Available at: <http://pi.314159.ru/petroff1.pdf> (accessed: 03.01.2020).
- [12] Edemskaya E.N., Belkov D.V. Network traffic research by the Hurst function. *Informatika i kibernetika* [Informatics and Cybernetics], 2015, no. 2, pp. 39–46 (in Russ.).
- [13] Fedorova M.L., Ledeneva T.M. About research of property of self-similarity of the traffic of a multiservice network. *Vestnik VGU. Seriya: Sistemnyy analiz i informatsionnye tekhnologii* [Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies], 2010, no. 1, pp. 46–54 (in Russ.).
- [14] Petrov V.V. Statisticheskiy analiz setevogo trafika [Statistical analysis of network traffic]. *pi.314159.ru: website* (in Russ.). Available at: <http://pi.314159.ru/petroff2.pdf> (accessed: 03.01.2020).
- [15] Vorobyev O.V., Mozhaev A.A., Oskolkov A.P. Analysis of traffic description models. *Nauka i tekhnika Povitryanikh Sil Zbroynikh Sil Ukraini* [Science and Technology of the Air Force of Ukraine], 2014, no. 2 (15), pp. 170–172 (in Russ.).

Chernigovskiy A.V. — Post-Graduate Student, Department of Computers, Computer Installations, Systems and Networks, Faculty of Engineering Cybernetics, Angarsk State Technical University (Chaykovskogo ul. 60, Angarsk, Irkutsk Region, 665835 Russian Federation).

Krivov M.V. — Cand. Sc. (Eng.), Head of Department of Computers, Computer Installations, Systems and Networks, Faculty of Engineering Cybernetics, Angarsk State Technical University (Chaykovskogo ul. 60, Angarsk, Irkutsk Region, 665835 Russian Federation).

Istomin A.L. — Dr. Sc. (Eng.), Professor, Department of Computers, Computer Installations, Systems and Networks, Faculty of Engineering Cybernetics, Angarsk State Technical University (Chaykovskogo ul. 60, Angarsk, Irkutsk Region, 665835 Russian Federation).

Please cite this article in English as:

Chernigovskiy A.V., Krivov M.V., Istomin A.L. Investigating network traffic and selecting a matching mathematical model. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering*, 2020, no. 3 (132), pp. 84–99 (in Russ.).
DOI: <https://doi.org/10.18698/0236-3933-2020-3-84-99>