ISSN 2782-2001
Системы анализа и обработки данных
том 86, № 2, 2022, с. 55–68

http://journals.nstu.ru/vestnik
Analysis and data processing systems
Vol. 86, No. 2, 2022, pp. 55–68

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ

INFORMATION TECHNOLOGIES AND TELECOMMUNICATIONS

УДК 004.738.2

DOI: 10.17212/2782-2001-2022-2-55-68

Сравнительный анализ современных трендов в области моделей трафика сетей передачи данных*

И.Л. РЕВА[a], А.В. ИВАНОВ[b], М.А. МЕДВЕДЕВ[c], И.А. ОГНЕВ[d]

630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

[a] *reva@corp.nstu.ru* [b] *andrej.ivanov@corp.nstu.ru* [c] *m.medvedev@corp.nstu.ru* [d] *i.ognev.2016@corp.nstu.ru*

На сегодняшний день в вопросах обработки и управления сетевым трафиком нет единого подхода, применимого к широкому пулу прикладных задач, который бы позволял решать вопросы по управлению трафиком. Опубликованные работы в этой области направлены на решение узкоспециализированных целей: при применении комплексных решений эти задачи требуют введения множества дополнительных параметров, которые увеличивают вычислительную сложность, или решают только узконаправленные проблемы.

В настоящей статье приводится сравнительный анализ классических моделей сетевого трафика и выявляется возможность практического применения таких моделей в реальных задачах. Подробно рассмотрены классические модели трафика, а именно пуассоновская модель, модели трафика с «тяжелым хвостом», модели на основе цепей Маркова, модели трафика на основе теории фракталов и модели на основе стохастических временных рядов. Также представлено математическое описание каждой модели трафика.

По итогам проведенного сравнительного анализа была проведена оценка применимости математических моделей к реальным проектам. На ее основе были выявлены две основные проблемы: во-первых, отсутствие учета предыдущих результатов обработки сетевого трафика; во-вторых, узконаправленная применимость каждой из моделей с учетом жесткой привязки к предметным областям, что позволяет решать лишь узкий круг задач.

В качестве критериев для оценки моделей сетевого трафика брались за основу следующие показатели: возможность масштабирования анализируемого трафика, возможность учитывать предыдущие данные трафика, вычислительная сложность и отсутствие неких случайных признаков, которые могли бы повлиять на работу модели. При детальном изучении проблемы масштабирования трафика выявлены основные закономерности, зависимости, размерности пакета трафика от времени его обработки.

* *Статья получена 15 января 2022 г.*

Работа выполнена при поддержке ЦК НТИ «Технологии доверенного взаимодействия» в рамках проекта «Технология семантической контент-фильтрация сетевого трафика (нежелательного контента)».

Ключевые слова: сетевой трафик, компьютерные сети, математические модели, пуассоновская модель трафика, модели временных рядов, фрактальное распределение, цепи Маркова, масштабируемость системы, доверенная среда

ВВЕДЕНИЕ

В рамках современного и динамически развивающегося информационного пространства начинает получать всё большую важность проблематика задач в области управления трафиком с точки зрения безопасности и защиты информации [1–3]. Когда начинают выделять конкретные подзадачи для решения определенного пула задач, первоначально встает вопрос о сложности идентификации и классификации трафика.

В рамках задачи по классификации трафика наиболее универсальным методом считается использование математических моделей, а именно их применение для решения нижеизложенных сетевых подзадач:

1) предсказание трафика, полученного в будущем, с целью предварительной оценки ресурсов, необходимых для выявления таких параметров, как необходимая минимальная пропускная способность узла трафика и определение размерности буфера для получения минимальных значений возможных количественных признаков, допустимых в рамках показателей потерь и задержки пакетов трафика;

2) выявление степени влияния алгоритмов управления системным трафиком на количественные и качественные характеристики сети;

3) выявление в рамках задач передачи трафика аномальных процессов во время передачи данных сети, таких как фрактальность трафика, пульсация трафика и т. д.;

4) использование генераторов трафика для имитирования потоков трафиков между объектами в реальной сети и постановка предположения, что выявленные закономерности будут достаточны для классифицирования объектов в реальной сети [4];

5) выявление количественно-качественных признаков источника трафика и на основе полученных признаков выявление идентификационных сигнатур относительно этого же источника трафика [5].

В основе ряда моделей трафика лежат стационарные случайные процессы $X(t)$ с различными законами распределения, с помощью которых воспроизводятся характеристики трафика (количество пакетов, полученных или отправленных в течение определенного промежутка времени; $\{\tau_i\}$, где $i = 1, 2,\ldots$, интервалы между пакетами; длины пакетов $\{l_i\}$, $i = 1, 2,\ldots$, последовательность направлений передачи пакетов $\{\delta_i\}$, $i = 1, 2,\ldots$ [6].

В зависимости от методологии и описания $X(t)$ модели делятся на наиболее часто встречающиеся группы:

1) пуассоновская модель трафика;

2) модели трафика с «тяжелым хвостом»;

3) модели на основе цепей Маркова;

4) модели трафика на основе теории фракталов;

5) модели на основе стохастических временных рядов.

При этом не следует забывать о том, что при увеличении количества обрабатываемых пакетов увеличивается и вычислительная сложность обработки пакетов и, как следствие, всегда присутствует вероятность того, что какие-то решения будут хороши на короткой дистанции, но на длинной – будут недостаточны, где число пакетов стремится к $+\infty$.

1. ПУАССОНОВСКАЯ МОДЕЛЬ ТРАФИКА

Модель Пуассона можно назвать самой старой используемой моделью трафика. Эта традиционная модель трафика возникла, когда количество входящих пакетов начало соответствовать распределению Пуассона, где длина каждого входящего пакета моделируется как экспоненциальное распределение [7, 8]. Один из случаев, когда модель Пуассона подходит лучше всего, – это зависящие от времени пуассоновские процессы, средняя скорость непостоянна. В модели Пуассона $X(t)$ определяет количество входящих пакетов, причем вероятность получения $X(t) = k$ пакетов за интервал времени задается экспоненциальным законом распределения [9]:

$$P\{x(t) = k\} = \frac{(\lambda t)^k e^{(-\lambda t)}}{k!}, \tag{1}$$

где λ – интенсивность поступления пакетов.

При этом вероятность получения ноль-пакетов равна

$$P\{x(t) = 0\} = e^{(-\lambda t)}. \tag{2}$$

Уравнение (1) демонстрирует следующие характеристики: если оценивать интервал времени между двумя пакетами трафика, то можно выявить, что распределение интервала времени между двумя пакетами является не чем иным, как экспоненциальным распределением с параметром λ.

Выгодной особенностью описанной математической модели является простая формулировка, а сумма нескольких независимых пуассоновских процессов составляет новый процесс с суммарной интенсивностью $\lambda = \Sigma_i \lambda_i$ [9].

В рамках работы с моделью необходимо учитывать, что модель не сохраняет предыдущие состояния модели, и вследствие этого не решается проблема описания пульсации трафика. Следует отметить то, что используемая модель распределения трафика крайне редко встречается на практике.

2. МОДЕЛИ ТРАФИКА С «ТЯЖЕЛЫМ ХВОСТОМ»

Эта модель трафика имеет «хвост», который сужается постепенно, а не резко – это подтип распределения с «тяжелым хвостом». В рассматриваемом же выше пуассоновском распределении модель трафика рассматривается как «легкий хвост». Эту модель можно кратко описать как распределение, которое имеет большое количество вхождений далеко от среднего.

В этой модели бесконечно убывающая часть трафика намного длиннее, что дает реальные шансы генерировать большие числа [7, 8]. Это связано с тем, что если будут рассматриваться сети IoT и при попытке отслеживания пакета трафика, сенсорный узел будет передавать данные только при условии того, что пакет находится в зоне слежения, и как только этот пакет покинет пределы зоны информации, о нем не будут поступать данные [10, 11]. На основе этого можно заявить о том, что пакеты данных должны передаваться некими массивами.

Тогда если пытаться рассматривать поведение трафика, то будет просматриваться большая дисперсия квантов времени между моментами поступления данных даже при большом отклонении от среднего. И тогда функция распределения будет иметь слабый тренд убывания «хвоста» [12, 13]:

$$P(Z > x) \approx cx^{(-a)}, \quad x \to \infty, \tag{3}$$

где c – константа, $0 < a < 2$ – параметр формы (индекс «хвоста» распределения), характеризует степень его тяжести.

Как видно из формулы (3), при незначительных колебаниях значений параметров формы распределение имеет «тяжелый хвост». Схожие распределения также имеют логнормальные распределения, Парето и т. д. [12].

3. МОДЕЛИ НА ОСНОВЕ ЦЕПЕЙ МАРКОВА

Марковские модели трафика – это стохастические модели, учитывающие состояние моделей в предыдущих событиях. Такая модель характеризуется как последовательность возможных событий на основе прошлого состояния модели. Эта модель возникает, когда трафик идет пакетами с паузами, как при голосовом общении [7, 8].

С математической точки зрения марковская модель есть не что иное, как набор дискретных случайных величин $\{X_n,\ n \in N\}$ [6, 14], при условии

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, ..., X_0 = x_0) = P(X_{n+1} = x_{n+1} \mid X_n = x_n), \tag{4}$$

где значения случайных величин $\{X_n\}$ образуют пространство состояний с конечным или с четным числом состояний при условии, что при фиксации настоящего прошлое не имеет абсолютно никакого влияния на будущее. При условии конечности пространства состояний (при этом нам известна его размерность) переходная матрица вероятностей будет описываться следующим уравнением [6]:

$$a_{ij}(n) = P(X_{n+1} = x_i \mid X_n = x_j), \tag{5}$$

где размерность равна N, а n-й шаг равен $a(n)$.

Модель Маркова считается гомогенной, если шаг не зависит от элементов матрицы переходных вероятностей, т. е. [6]

$$a_{ij}(n) = a_{ij}, \qquad \forall n \in N. \tag{6}$$

Модель Маркова применяется, как правило, когда у нас задача сводится лишь к определению состояния системы от некоего фактора, к примеру, от состояния пользователя (активен / неактивен), а также при моделировании получения пакетов трафика в течение некоего t_i.

Проблематика в рамках задач по моделированию трафика, использующего марковские процессы, заключается в излишнем «оптимизме», это связано с тем, что модель слабо учитывает нагрузки в сетях, имеющих периодичность протекания процессов.

4. МОДЕЛИ ТРАФИКА НА ОСНОВЕ ТЕОРИИ ФРАКТАЛОВ

На сегодняшний день доказано, что трафик на протяжении некоего временного промежутка меняется. Как следствие, в рамках задач по масштабированию трафика необходимо иметь более мощные инструменты по анализу и производительности компьютерной модели.

Первое упоминание фрактала (самоподобия) ввел в обиход Бенуа Мандельброт [15, 16]. Понятие фрактала использовалось для описания сохранения ряда свойств исследуемых объектов на различных масштабах пространства и времени. Следует отметить, что из-за такого расплывчатого описания, как правило, фракталы определяются в терминах их атрибутов, таких как медленное затухание их дисперсий, гиперболическое хвостовое распределение плотности времени между последовательными приходами, моменты бесконечного порядка или плохо определенная статистика, шум $1/f$, дальнодействующая зависимость, самоподобие и нецелочисленная размерность и т. д. [6].

Когда самоподобные модели трафика были впервые представлены, не существовало эффективных, аналитически поддающихся обработке процессов для создания моделей. Илкка Норрос разработал стохастический процесс для модели хранения с самоподобным входом и постоянной скоростью передачи данных на выходе. Хотя эта первоначальная модель была скорее непрерывной, чем дискретной, она была эффективной, простой и привлекательной [17]. Прежде чем обсуждать стохастический процесс Норроса, сначала определим нормализованное дробное броуновское движение (fBm). fBm – это гауссовский процесс с непрерывным временем, определенный для всех положительных значений времени, со средним значением, равным нулю, и автокорреляцией, определенной для параметра Херста (H) как [18]

$$y(t,s)=\frac{1}{2}(|t|^{2H}+|s|^{2H}-|t-s|^{2H}). \tag{7}$$

Математически процесс Норроса представляется следующим образом [18]:

$$V_{(t)}=\sup_{s\le t}(A(t)-A(s)-C(t-s)),\quad t\in(-\infty,\infty), \tag{8}$$

где $A(t)$ является процессом

$$A(t)=mt+\sqrt{am}\cdot Z(t), \tag{9}$$

где $Z(t)$ представляет собой нормализованную fBm с параметром Херста в интервале $\left(\frac{1}{2}, 1\right]$. Параметры процесса: m – средняя скорость ввода, a – коэффициент дисперсии, H – параметр Херста, C – скорость обслуживания.

5. МОДЕЛИ НА ОСНОВЕ СТОХАСТИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ

В рамках исследования построения моделей трафика были разработаны модели на основе теории временных рядов. Их основная идея заключается в том, что за основу берется некий процесс X_n в моделях сетевого трафика, и он используется для представления интенсивности источника трафика с дискретным временем $n = 1, 2, \ldots$, где интервалы между последовательными отсчетами равны.

Главная идея этой модели заключена в том, что кратковременную зависимость в полной мере раскрывают авторегрессии при представлении процесса X_n, позволяющего использовать их для прогнозирования трафика.

В процессе проработки моделей были выделены модели типов AR (Autoregressive) с авторегрессионным скользящим средним ARMA (Autoregressive moving average) и с авторегрессионным интегрированным скользящим средним ARIMA (Autoregressive integrated moving average) [6].

Упомянутая выше модель AR отличается тем, что здесь от предыдущих значений рядов имеют линейную зависимость текущие значения [6, 19]:

$$X_n = \sum_{i=1}^{p} a_j X_{n-i} + e_n, \tag{10}$$

где p – порядок модели, $\{a_1, a_2, \ldots, a_p\}$ – ее коэффициенты и e_n – белый шум.

При рассмотрении модели уравнения следует, что X_n по сути своей есть не что иное, как изменение для n-го участка временного ряда.

Модель ARMA представляет собой обобщение модели AR и модели скользящего среднего MA (Moving Average). В математическом виде модель MA представляет собой линейную комбинацию двух предыдущих значений шума [6]:

$$X_n = e_n + \sum_{i=1}^{q} \beta_t e_{n-t}, \tag{11}$$

где $\{e_n\}$ – белый шум; $\{\beta_1, \beta_2, \ldots, \beta_q\}$ – коэффициенты скользящего среднего, q – порядок модели.

Модель ARMA (p, q) составляется из комбинации моделей AR и MA следующим образом [6]:

$$X_n = e_n + \sum_{i=1}^{p} a_i X_{n-i} + \sum_{i-1}^{q} \beta_i e_{n-i}. \tag{12}$$

Если в этой модели ввести обратный оператор B такой, что $BX_n = X_{n-1}$, то модель ARMA можно представить в виде [6]

$$\varphi(B)X_n = \theta(B)e_n, \tag{13}$$

где

$$\varphi(B) = 1 - a_1 B - ... - a_p B^p, \tag{14}$$

$$\theta(B) = 1 + \beta_1 B + ... + \beta_q B^q. \tag{15}$$

Модели ARMA из-за наличия скользящего среднего коэффициента могут использоваться в рамках работ, обладающих цикличностью либо сезонным неким характером [6].

Относительно большое распространение она получила в сетях модели данных Peer-to-peer, а также имеет потенциал в системах идентификации сетевых вторжений и атак.

Модель ARIMA является частным случаем ARMA, если принять $Y_n = \nabla^d X_n$ (d порядок разности), и ∇X_n равно

$$\nabla X_n = X_n - X_{n-1} = (1 - B)X_n. \tag{16}$$

Тогда для ARMA (p, q)

$$\varphi(B)Y_w = \theta(B)e_n, \tag{17}$$

$$\varphi(B)(1 - B)^d X_n = \theta(B)e_n. \tag{18}$$

В этом случае X_n является процессом модели ARIMA, так как X_n является интегралом процесса Y_n (модели ARMA) [6].

Как правило, временные ряды имеют некие тренды, то есть постепенное увеличение, либо некие циклические особенности, что и следует из модели ARMA. Для сглаживания циклических изменений используется разность значений временных рядов $(1 - B)^d X_n$. Полученная разность является не чем иным, как стационарным временным рядом [20].

На основании рассматриваемой математической модели можно отметить, что данная модель может применяться в прогнозировании трафика и производительности сети, как это было предложено в работе Г. Рутка «Прогнозирование сетевого трафика с использованием моделей ARIMA и нейронных сетей» [20].

6. СРАВНИТЕЛЬНЫЙ АНАЛИЗ СУЩЕСТВУЮЩИХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ТРАФИКА

Рассмотрев основные тенденции в области построения моделей трафика, можно сделать следующие выводы:

1) каждая из рассмотренных моделей имеет свою узконаправленную тематику в плане применения относительно конкретных предметных областей и очень сильно зависит от той предметной области, в которую она помещается;

2) применение к задачам идентификации, классификации сетевого трафика является затруднительным в связи с тем, что они не учитывают предыдущее поведение данных и ориентированы на трафики с простым принципом распределения, то есть на практике существует большая проблема по аппроксимации поведения трафика относительно его качественно-количественных признаков;

3) модели трафика на основе теории временных рядов выглядят перспективно для моделирования трафиков различного типа, особенно если качественно-количественные признаки зависят от своих прошлых значений, но их проблема в том, что, используя стандартизированные методы вычислений, с увеличением количества признаков увеличивается вычислительная сложность данной модели, и, как следствие этого, она сильно сужает область практического применения, особенно в местах применения систем in-real-time;

4) фрактальные модели работают на выявление долговременных зависимостей, поэтому они достаточно хороши в области выявления ситуаций, отличных от нормы, к примеру аномалий или сетевых атак, но из-за больших ограничений числа пакетов они слабо подходят для идентификации трафика в режиме реального времени;

5) модели с «тяжелым хвостом» позволяют решать проблемы, которые возникают в системах с «легким хвостом» (пуассоновские модели), но из-за отсутствия памяти они могут лишь описывать цикличные изменения;

6) модели на основе цепей Маркова, как правило, используются для решения задач по типу «положительный / отрицательный исход» (пример: распознавание речи, жестов, рукописного текста и т. п.), лишены недостатков относительно памяти и могут описывать различные виды трафиков. Также такие модели не зависят от протоколов, и с расширением количества признаков не так сильно растет вычислительная сложность в сравнении с моделями временных рядов. Но при решении задач по этому принципу используются случайные значения для инициализации параметров модели, и данный аспект требует существенной доработки.

7. ПРОБЛЕМА МАСШТАБИРОВАНИЯ ТРАФИКА

На сегодняшний день тенденции сетевого трафика диктуют новые условия для формирования методологии оценки трафика, при этом учитываются требования к обработке его скорости. Таким образом можно выявить то, что трафик на пути с количеством пакетов N изменяется прямо пропорционально, с увеличением числа пакетов увеличивается и время реакции:

$$K = \frac{N}{V_{\text{обр}}}, \tag{19}$$

где $V_{\text{обр}}$ – скорость обработки 1-го пакета; N – общее число всех пакетов; K – скорость обработки всего отправленного трафика. Общее количество пакетов стремится к $+\infty$.

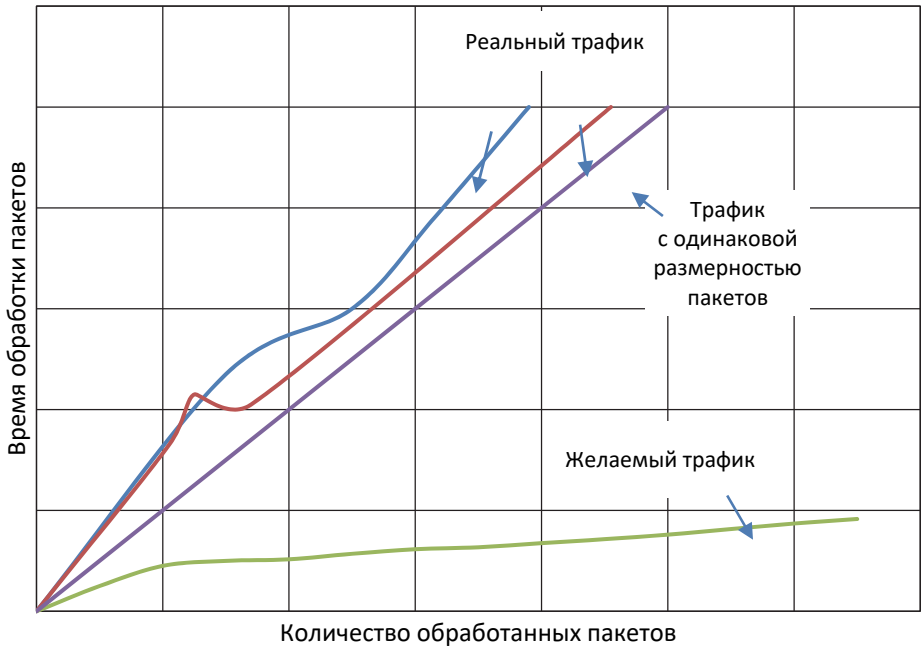
При этом не следует забывать о том, что в реальном трафике пакеты различны и, как следствие, функция будет вести себя нелинейно. При аппроксимировании поведения формулы (19) возникают проблемы классификации пакетов данных, а именно при масштабировании трафика результирующее значение функции (19) должно изменяться незначительно:

$$|v_2 - v_1| \to 0 \text{ на любом участке } t, \tag{20}$$

где t – определенный отрезок времени; $v_1 - v_2$ – разность скоростей, характеризует изменение скорости обработки пакетов.

Визуализируя представленные выше формулы, мы получаем графическое представление на рисунке. Из этого вытекает следующая проблема: при масштабировании системы у нас слишком сильно возрастают временные затраты на обработку пакетов, и как следствие, вырастают затраты на обработку самого трафика в рамках системы. Из-за этого необходимо помимо представленных выше характеристик учитывать проблемы масштабирования трафика при поиске подхода к решению задачи и снижению вычислительной сложности.

В диссертации [8] Д.А. Божалкина и более ранних работах других авторов уже предпринимались попытки классифицировать трафик по размерности пакетов и, как следствие, снизить вычислительную сложность.



Графическое представление обработки трафика

Graphical representation of traffic processing

При попытке построения математической модели необходимо учитывать предыдущие состояния сети для ускорения обработки трафика. Возможно, имеет смысл задуматься о выявлении подобий объектов внутри трафика и, таким образом, снизить временные затраты на обработку пакетов трафика.

ВЫВОДЫ

Исходя из результатов сравнительного анализа моделей сетевого трафика сделан вывод о том, что в настоящее время в данной предметной области существует определенный набор задач и требований, которые необходимо решить. В рамках этой проблематики отсутствует четкий комплекс параметров, по которым можно было бы провести разметку трафика и, как следствие, понимать, какие признаки трафика наиболее важны, а какими можно пренебречь. Современные авторы пытаются решать частные задачи с опорой на существующие решения в области системного анализа.

Большинство вышеописанных моделей неконкурентоспособны при классификации трафика больших объемов, так как в рамках работ математических моделей трафик оценивается целиком. Но при этом не рассматриваются отдельные пакеты трафика и не происходит поиск самоподобия между этими пакетами. Как следствие, мы вынуждены каждый отдельный пакет представлять как самодостаточную единицу трафика. Если бы существующие математические модели данных имели опцию выделения определенного уровня подобия между пакетами трафика, мы смогли бы использовать принцип формирования доверенной среды между пакетами трафика, где при выявлении определенных свойств и совпадений этих свойств между уже обработанными пакетами трафика и новым поступившим пакетом трафика можно было бы заявить о подобии пакетов и не тратить ресурсы на полную классификацию, объявив при этом пакеты подобными, и, как следствие, снизить вычислительную сложность метода обработки трафика. Но мы не можем ускорить обработку пакетов, признав их самоподобными без наличия самих отдельных конкретных признаков, которые мы могли бы учитывать.

Делаем вывод о недостаточной проработке с точки зрения математических моделей существующих решений в области классификации и разметки сетевого трафика.

СПИСОК ЛИТЕРАТУРЫ

1. *Кузьмин В.В.* Модели и процедуры управления трафиком в мультисервисной сети оператора связи: дис. ... канд. техн. наук. – Н. Новгород, 2015. – 189 с.

2. Hidden markov model modeling of SSH brute-force attacks / A. Sperotto, R. Sadre, P. de Boer, A. Pras // Integrated Management of Systems, Services, Processes and People in IT. DSOM 2009. – Venice, Italy, 2009. – P. 164–176.

3. *Velan P., Čermák M., Čeleda P.* A survey of methods for encrypted traffic classification and analysis // International Journal of Network Management. – 2015. – Vol. 25 (5). – P. 355–374.

4. Имитатор сетевого трафика / П.А. Будко, Д.Н. Рыжкова, Ж.О. Карпова, Д.В. Воронина // Техника средств связи. – 2018. – № 2 (142). – С. 86–98.

5. *Ефимов А.Ю.* Использование энтропийных характеристик сетевого трафика для определения его аномальности // Программные продукты и системы. – 2021. – Т. 34, № 1. – С. 83–90.

6. *Джаммул С.М.* Идентификация трафика сетей передачи данных в реальном времени: дис. … канд. техн. наук. – М., 2018. – 143 с.

7. *Manaseer S., Al-Nahar O.M., Hyassat A.S.* Network traffic modeling, case study: the University of Jordan // International Journal of Recent Technology and Engineering (IJRTE). – 2019. – Vol. 7, iss. 5. – P. 13–16.

8. How to accelerate your Internet: a practical guide to bandwidth management and optimisation using open source software / R. Flickenger, M. Belcher, E. Canessa, M. Zennaro. – [S. l.]: INASP/ICTP, 2006. – ISBN 0-9778093-1-5. – 298 p.

9. *Карташевский И.В.* Разработка и исследование методов анализа качества обслуживания коррелированного трафика в телекоммуникационных сетях: дис. … д-ра техн. наук. – Самара, 2020. – 288 с.

10. *Wang Q., Zhang T.* Source traffic modeling in wireless sensor networks for target tracking // PE-WASUN'08: Proceedings of the 5th ACM International Symposium on Performance Evaluation of Wireless Ad-Hoc, Sensor, and Ubiquitous Networks. – Vancouver, British Columbia, Canada, 2008. – P. 96–100.

11. *Vybornova A., Koucheryavy A.* Traffic analysis in target tracking ubiquitous sensor networks // Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN 2014. – Cham: Springer International Publishing, 2014. – P. 389–398. – (Lecture Notes in Computer Science; vol. 8638).

12. *Осовский А.В., Кутузов Д.В., Стукач О.В.* Анализ моделей трафика, создаваемого устройствами интернета вещей // Динамика систем, механизмов и машин. – 2019. – Т. 7, № 4. – С. 220–226.

13. *Кутузов Д.В., Осовский А.В., Стукач О.В.* Модель генерации и обработки трафика IoT параллельными коммутационными системами // Вестник СибГУТИ. – 2019. – № 4. – С. 78–87.

14. *Сарымсаков Т.А.* Основы теории процессов Маркова. – М.: Гостехтеориздат, 1954. – 208 с.

15. *Федер Е.* Фракталы: пер. с англ. – М.: Мир, 1991. – 254 с.

16. *Feder J.* Fractals. – New York: Plenum Press, 1988. – 312 p.

17. *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.

18. *Елагин В.С., Спиркина А.В., Фицов В.В.* Фундаментальные основы моделирования трафика в гетерогенных сетях связи с перспективой канальной идентификации отдельных сервисов и прогнозирования состояния сети // Актуальные проблемы инфотелекоммуникаций в науке и образовании: X Юбилейная Международная научно-техническая и научно-методическая конференция: в 4 т. – СПб., 2021. – Т. 1. – С. 351–356.

19. Characterization of video codecs as autoregressive moving average processes and related queuing system performance / R. Grunenfelder, J.P. Cosmas, S. Manthorpe, A. Odinma-Okafor // IEEE Journal on Selected Areas in Communications. – 1991. – Vol. 9 (3). – P. 284–293.

20. *Rutka G.* Network traffic prediction using ARIMA and neural network models // Electronics and Electrical Engineering. – 2008. – Vol. 4 (48). – P. 47–52.

Рева Иван Леонидович, кандидат технических наук, доцент кафедры защиты информации Новосибирского государственного технического университета. Основное направление научных исследований – информационная безопасность, информационные технологии. Имеет более 46 публикаций. E-mail: reva@corp.nstu.ru

Иванов Андрей Валерьевич, кандидат технических наук, доцент кафедры защиты информации Новосибирского государственного технического университета. Основное направление научных исследований – акустические измерения, техническая защита информации. Имеет более 63 публикаций. E-mail: andrej.ivanov@corp.nstu.ru

Медведев Михаил Александрович, ассистент кафедры защиты информации Новосибирского государственного технического университета. В настоящее время специализируется в области машинного обучения и информационной безопасности. E-mail: M.medvedev@corp.nstu.ru

Огнев Игорь Александрович, ассистент кафедры защиты информации Новосибирского государственного технического университета. В настоящее время специализируется в области анализа сетевого трафика и информационной безопасности. E-mail: i.ognev.2016@corp.nstu.ru

Reva Ivan L., PhD (Eng.), associate professor at the Information Security Department, Novosibirsk State Technical University. The main filed of his research is information security and information technologies. He has more than 46 publications. E-mail: reva@corp.nstu.ru

Ivanov Andrei V., PhD (Eng.), associate professor at the Information Security Department, Novosibirsk State Technical University. The main filed of his research is the acoustic measurements, and technical protection of information. He has more than 63 publications. E-mail: reva@corp.nstu.ru

Medvedev Mikhail A., assistant lecturer at the Information Security Department, Novosibirsk State Technical University. Currently he specializes in the field of machine education and information security. E-mail: M.medvedev@corp.nstu.ru

Ognev Igor A., assistant lecturer at the Information Security Department, Novosibirsk State Technical University. Currently he specializes in the field of network traffic analysis and information security. E-mail: i.ognev.2016@corp.nstu.ru

DOI: 10.17212/2782-2001-2022-2-55-68

Comparative analysis of modern trends in the field of traffic models of data transmission networks[*]

I.L. REVA[a], A.V. Ivanov[b], M.A. MEDVEDEV[c], I.A. OGNEV[d]

Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

[a] *reva@corp.nstu.ru* [b] *andrej.ivanov@corp.nstu.ru* [c] *m.medvedev@corp.nstu.ru*
[d] *i.ognev.2016@corp.nstu.ru*

Abstract

To date, in matters of processing and managing network traffic, there is no single approach applicable to a wide pool of practical and applied tasks that would allow solving traffic management issues. Published works in this area are aimed at solving highly specialized problems: when applying complex solutions, these problems require the introduction of many additional parameters that increase computational complexity or solve only narrowly focused problems.

This article provides a comparative analysis of classical network traffic models and reveals the possibility of practical application of such models in real-life problems. Classical traffic models are considered in detail, namely the Poisson model, heavy-tail traffic models, models based on Markov chains, traffic models based on the fractal theory and models based on stochastic time series. A mathematical description of each traffic model is also presented.

Based on the results of the comparative analysis, the applicability of mathematical models to real projects was assessed. Based on it, two main problems were identified: first, the lack of consideration of the previous results of network traffic processing; secondly, the narrowly focused applicability of each of the models, given the rigid binding to subject areas, which allows solving only a narrow range of problems.

The following indicators were taken as the criteria for evaluating network traffic models: the ability to scale the analyzed traffic, the ability to consider previous traffic data, computational complexity and the absence of some random features that could affect the operation of the model. A detailed study of the problem of traffic scaling revealed the main patterns, dependencies, dimensions of the traffic packet by the time it was processed.

Keywords: network traffic, computer networks, mathematical models, Poisson traffic model, time series models, fractal distribution, Markov chains, system scalability, trusted environment

---

[*] *Received on 15 January 2022.*

The work was supported by the NTI Central Committee "Technologies of trusted interaction" within the framework of the project "Technology of semantic content filtering of network traffic (unwanted content)".

REFERENCES

1. Kuz'min V.V. *Modeli i protsedury upravleniya trafikom v mul'tiservisnoi seti operatora svyazi*. Diss. kand. tekhn. nauk [Models and procedures for traffic management in a multiservice network of a telecom operator. PhD eng. sci. diss.]. Nizhny Novgorod, 2015. 189 p.

2. Sperotto A., Sadre R., Boer P. de, Pras A. Hidden markov model modeling of SSH brute-force attacks. *Integrated Management of Systems, Services, Processes and People in IT. DSOM 2009*, Venice, Italy, 2009, pp. 164–176.

3. Velan P., Čermák M., Čeleda P. A survey of methods for encrypted traffic classification and analysis. *International Journal of Network Management*, 2015, vol. 25 (5), pp. 355–374.

4. Budko P.A., Ryzhkova D.N., Karpova Zh.O., Voronina D.V. Imitator setevogo trafika [Network traffic simulator]. *Tekhnika sredstv svyazi = Means of Communication Equipment*, 2018, no. 2 (142), pp. 86–98.

5. Efimov A.Yu. Ispol'zovanie entropiinykh kharakteristik setevogo trafika dlya opredeleniya ego anomal'nosti [Using the entropy characteristics of network traffic to determine its abnormality]. *Programmnye produkty i sistemy = Software and Systems*, 2021, vol. 34, no. 1, pp. 83–90.

6. Jammul S.M. *Identifikatsiya trafika setei peredachi dannykh v real'nom vremeni*. Diss. kand. tekhn. nauk [Identification of the traffic of data transmission networks in real time. PhD eng. sci. diss.]. Moscow, 2018. 143 p.

7. Manaseer S., Al-Nahar O.M., Hyassat A.S. Network traffic modeling, case study: the University of Jordan. *International Journal of Recent Technology and Engineering (IJRTE)*, 2019, vol. 7, iss. 5, pp. 13–16.

8. Flickenger R., Belcher M., Canessa E., Zennaro M. *How to accelerate your Internet: a practical guide to bandwidth management and optimisation using open source software*. INASP/ICTP, 2006. ISBN 0-9778093-1-5. 298 p.

9. Kartashevskii I.V. *Razrabotka i issledovanie metodov analiza kachestva obsluzhivaniya korrelirovannogo trafika v telekommunikatsionnykh setyakh*. Diss. dokt. tekhn. nauk [Development and research of methods for analyzing the quality of service of correlated traffic in telecommunication networks. Dr. eng. sci. diss.]. Samara, 2020. 288 p.

10. Wang Q., Zhang T. Source traffic modeling in wireless sensor networks for target tracking. *PE-WASUN'08: Proceedings of the 5th ACM International Symposium on Performance Evaluation of Wireless Ad-Hoc, Sensor, and Ubiquitous Networks*, Vancouver, British Columbia, Canada, 2008, pp. 96–100.

11. Vybornova A., Koucheryavy A. Traffic analysis in target tracking ubiquitous sensor networks. *Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN 2014*. Cham, Springer International Publishing, 2014, pp. 389–398.

12. Osovskiy A.V., Kutuzov D.V., Stukach O.V. Analiz modelei trafika, sozdavaemogo ustroistvami interneta veshchei [Analyze traffic patterns generated by IoT devices]. *Dinamika sistem, mekhanizmov i mashin = Dynamics of Systems, Mechanisms and Machines*, 2019, vol. 7, no. 4, pp. 220–226.

13. Kutuzov D.V., Osovskiy A.V., Stukach O.V. Model' generatsii i obrabotki trafika IoT parallel'nymi kommutatsionnymi sistemami [IoT traffic generation and processing model with parallel switching systems]. *Vestnik SibGUTI = Herald of SibSUTIS*, 2019, no. 4, pp. 78–87.

14. Sarymsakov T.A. *Osnovy teorii protsessov Markova* [Fundamentals of the theory of Markov processes]. Moscow, Gostekhteorizdat Publ., 1954. 208 p.

15. Feder J. *Fraktaly* [Fractals]. Moscow, Mir Publ., 1991. 254 p. (In Russian).

16. Feder J. *Fractals*. New York, Plenum Press, 1988. 312 p.

17. Trenogin N.G., Petrov M.N., Sokolov D.E. Svoistva fraktal'nogo trafika pri prokhozhdenii sistemy massovogo obsluzhivaniya s ochered'yu [Properties of fractal traffic on the output of a queuing system]. *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta im. akademika M.F. Reshetneva = Vestnik SibGAU*, 2017, vol. 18, no. 1, pp. 105–110.

18. Elagin V.S., Spirkina A.V., Fitsov V.V. [Fundamental of modeling traffic heterogeneous communication networks, with the prospective channel identification of individual services and forecasting the network state]. *Aktual'nye problemy infotelekommunikatsii v nauke i obrazovanii* [Actual problems of infotelecommunications in science and education]. 10th International Conference on Advanced ICAIT 2021. St. Petersburg State University of Technology. St. Petersburg, 2021, vol. 1, pp. 351–356. (In Russian).

19. Grunenfelder R., Cosmas J.P., Manthorpe S., Odinma-Okafor A. Characterization of video codecs as autoregressive moving average processes and related queuing system performance. *IEEE Journal on Selected Areas in Communications*, 1991, vol. 9 (3), pp. 284–293.

20. Rutka G. Network traffic prediction using ARIMA and neural network models. *Electronics and Electrical Engineering*, 2008, vol. 4 (48), pp. 47–52.

Для цитирования:

Сравнительный анализ современных трендов в области моделей трафика сетей передачи данных / И.Л. Рева, А.В. Иванов, М.А. Медведев, И.А. Огнев // Системы анализа и обработки данных. – 2022. – № 2 (86). – С. 55–68. – DOI: 10.17212/2782-2001-2022-2-55-68.

For citation:

Reva I.L., Ivanov A.V., Medvedev M.A., Ognev I.A. Sravnitel'nyi analiz sovremennykh trendov v oblasti modelei trafika setei peredachi dannykh [Comparative analysis of modern trends in the field of traffic models of data transmission networks]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2022, no. 2 (86), pp. 55–68. DOI: 10.17212/2782-2001-2022-2-55-68.