

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра дискретной математики и алгоритмики

МАКАРЕВИЧ Евгения Николаевна

АНАЛИЗ СЕТЕВОГО ТРАФИКА С ПОМОЩЬЮ
УСТОЙЧИВОГО РАСПРЕДЕЛЕНИЯ

Магистерская диссертация

специальность 1-31 81 09 «Прикладная математика и информатика»

Научный руководитель
Соболева Татьяна Валентиновна
кандидат физико-математических наук

Допущена к защите
«___» _____ 2021 г.

Заведующий кафедрой дискретной математики и алгоритмики

_____ В.М. Котов
доктор физико-математических наук, профессор

Минск, 2021

ОГЛАВЛЕНИЕ

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ	3
ВВЕДЕНИЕ.....	6
ГЛАВА 1 САМОПОДОБНЫЙ СЕТЕВОЙ ТРАФИК	8
1.1 Понятие сетевого трафика.....	8
1.2 Обзор моделей представления сетевого трафика.....	8
1.3 Самоподобие и фрактальная размерность	9
1.4 Коэффициент Херста и RS-анализ.....	11
1.5 Фрактальные параметры.....	14
ГЛАВА 2 УСТОЙЧИВОЕ РАСПРЕДЕЛЕНИЕ.....	18
2.1 Понятие устойчивого распределения	18
2.2 Метод максимального правдоподобия.....	20
2.3 Обобщенный метод моментов	20
2.4 Итеративный метод Кутровелиса	21
2.5 Метод квантилей МакКаллаха	22
ГЛАВА 3 АНАЛИЗ РЕАЛЬНЫХ ДАННЫХ СЕТЕВОГО ТРАФИКА.....	24
3.1 Постановка задачи	24
3.2 Описание данных	24
3.3 Предобработка данных	26
3.4 Оценка параметров устойчивого распределения	27
3.5 Анализ самоподобия.....	30
3.6 Исследование данных, собранных в течение четырех недель	32
ЗАКЛЮЧЕНИЕ	35
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	36
ПРИЛОЖЕНИЕ А	38
ПРИЛОЖЕНИЕ Б.....	42
ПРИЛОЖЕНИЕ В	43

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Магистерская диссертация, 43 стр., 18 рис., 7 табл., 15 источников, 3 приложения.

СЕТЕВОЙ ТРАФИК, УСТОЧИВОЕ РАСПРЕДЕЛЕНИЕ, САМОПОДОБИЕ, ФРАКТАЛЬНЫЕ ПАРАМЕТРЫ, КОЭФФИЦИЕНТ ХЕРСТА, RS-АНАЛИЗ.

Объект исследования – свойство самоподобия сетевого трафика. В частности, исследуются возможность моделирования сетевого трафика с помощью устойчивого распределения.

Цель работы – изучить основные сведения о проблеме моделирования сетевого трафика; получить информацию о сетевом трафике, позволяющую обнаружить подозрительное поведение сети.

Методы проведения работы – исследование сетевого трафика с помощью методов статистического оценивания параметров вероятностных распределений, оценивание фрактальных показателей сетевого трафика.

Результаты – проведен анализ данных сетевого трафика на свойство самоподобия: подсчитан коэффициент Херста с помощью RS-анализа; произведено оценивание параметров устойчивого распределения различными методами; произведено моделирование сетевого трафика с помощью полученных параметров и произведено оценивание адекватности модели.

Область применения – результаты работы могут быть использованы при дальнейших научных исследованиях сетевого трафика, использоваться в учебном процессе при преподавании учебных дисциплин, связанных со статистическим анализом случайных процессов.

АГУЛЬНАЯ ХАРАКТЕРЫСТИКА ПРАЦЫ

Магістарская дысертацыя, 43 с., 18 мал., 7 табл., 15 крыніц, 3 дадатка.

Ключавыя словы: СЕТКАВЫ ТРАФІК, УСТОЙЛІВАЕ РАЗМЕРКАВАННЕ, САМОПОДОБИЕ, ФРАКТАЛЬНЫЯ ПАРАМЕТРЫ, КАЭФІЦЬЕНТ ХЁРСТА, RS-АНАЛІЗ.

Аб'ект даследавання – самападобіе сеткавага трафіку. У прыватнасці, даследуюцца магчымасць мадэлявання сеткавага трафіку з дапамогай ўстойлівага размеркавання.

Мэта работы – вывучыць асноўныя звесткі аб праблеме мадэлявання сеткавага трафіку; атрымаць інфармацыю аб сеткавым трафіку, якая дазваляе выявіць паводзіны сеткі.

Метады правядзення працы – даследаванне сеткавага трафіку з дапамогай метадаў статыстычнага ацэньвання параметраў імавернасных размеркаванняў, ацэньванне фрактальнай паказчыкаў сеткавага трафіку.

Вынікі – праведзены аналіз дадзеных сеткавага трафіку на самопадобнасць: падлічаны каэфіцыент Херста з дапамогай RS-аналізу; выраблена ацэньванне параметраў ўстойлівага трафіку рознымі метадамі; выраблена мадэляванне сеткавага трафіку з дапамогай атрыманых параметраў і выраблена ацэньванне адэкватнасці мадэлі.

Вобласць прымянення - вынікі працы могуць быць выкарыстаны пры далейшых навуковых даследаваннях сеткавага трафіку, выкарыстоўваюцца ў навучальным працэсе пры выкладанні вучэбных дысцыплін, звязаных са статыстычным аналізам выпадковых працэсаў.

ABSTRACT

Master thesis, 43 p., 18 figures, 7 tables, 15 sources, 3 appendices.

NETWORK TRAFFIC, STABLE DISTRIBUTION, SELF-SIMILARITY, FRACTAL PARAMETERS, HURST COEFFICIENT, RS-ANALYSIS.

Object of research – the self-similarity of network traffic. In particular, the possibility of modeling network traffic using stable distribution is being investigated. The purpose of the work is to study the basic information about the problem of modeling network traffic; Obtain information about network traffic to detect suspicious network behavior.

Methods – research of network traffic using methods of statistical estimation of the parameters of probability distributions, estimation of fractal indicators of network traffic.

Results – the analysis of network traffic data for self-similarity was carried out: the Hurst coefficient was calculated using RS-analysis; the parameters of stable distribution were estimated by various methods; the modeling of network traffic was carried out using the obtained parameters and the adequacy of the model was assessed.

ВВЕДЕНИЕ

С появлением на рынке большого количества устройств и инструментов сбор данных стал как никогда простым. Корпус данных сетевого трафика может генерироваться миллионами или миллиардами записей в секунду [1]. Сетевой трафик является одной из важнейших характеристик компьютерной сети. Также значительно увеличилось количество сетевых маршрутов, что привело к возникновению перегруженности сетевого трафика и последующей задержке в передаче данных, к потере пакетов.

В настоящее время при проектировании сетей используют каналы связи с пакетной коммутацией, что увеличивает эффективность использования каналов, однако снижает надёжность доставки информации. В случае перезагрузки канала, пакет может не поместиться во входной буфер. С увеличением числа передаваемых пакетов растёт и число потерянных пакетов, а также увеличивается время задержек. Постоянное увеличение объёма данных привело к тому, что на данном этапе задача прогнозирования сетевого трафика стала достаточно актуальной: необходимо моделировать сетевой трафик таким образом, чтобы максимально эффективно использовать имеющиеся ресурсы. К тому же аномалии и структурные изменения в поведении сетевого трафика могут свидетельствовать о различных атаках на сеть. В данном случае требуется своевременно обнаружить аномальный пакет и заблокировать возможную атаку.

Кибербезопасность - серьезная проблема для компаний и организаций, которые полагаются на технологии для поддержания своего бизнеса. В любой организации с интенсивным обменом данных из-за сбоя или уязвимости в системе организации могут быть потеряны миллионы или даже миллиарды денежных средств. Таким образом, с появлением новых технологий и инструментов использование и устранение уязвимостей таких систем может быть легко достигнуто с помощью соответствующих ресурсов. Поиск аномалий ставит своей целью обнаружение наличия в трафике изменений, нетипичных для его структуры. Автоматический поиск таких аномалий поможет сетевым

администраторам в качестве дополнительного источника информации, а также инструмента для диагностики поведения трафика и поиска источника проблем.

В данной работе приведены результаты статистических исследований реального сетевого трафика с помощью устойчивых распределений, осуществлено моделирование сетевого трафика с помощью полученных оценок параметров распределения.

ГЛАВА 1 САМОПОДОБНЫЙ СЕТЕВОЙ ТРАФИК

1.1 Понятие сетевого трафика

Сетевым трафиком (англ. traffic - «движение») называется объем информации, передаваемой через компьютерную сеть за определенный промежуток времени. Количество трафика может быть измерено в пакетах или таких единицах измерения как биты, байты, и их производных. Трафик подразделяется на внешний и внутренний, исходящий и входящий. [1]

В данной работе будет использована следующая классификация трафика: нормальный/аномальный трафик. Под аномальным трафиком будет подразумеваться трафик, не характерный для обычной работы сети. Такой трафик может сигнализировать об осуществлении сетевой атаки.

1.2 Обзор моделей представления сетевого трафика

Проектирование устойчивых компьютерных сетей базируется на понимании работы сетевого трафика и его характеристиках. Для этого были разработаны различные модели представления сетевого трафика. С помощью данных моделей могут быть получены оценки, характеризующие реальный трафик и его производительность. Модели полезны при статистическом анализе сетевого трафика и исследовании различных гипотез о его поведении.

Классическая модель

Исходящий трафик от источника представляет собой пуассоновский процесс, а время поступления сообщения и его длина имеют экспоненциальное распределение [2]. В данной модели не учитываются периоды всплесков активности в сети. Свойство независимого приращения данного распределения и обуславливает отсутствие памяти процесса. Это одна из наиболее старых моделей. Подходит для исследования, если у трафика большое количество источников.

Стохастические модели

Стохастическая модель сетевого трафика представляет собой марковский процесс, который обладает кратковременной зависимостью [3]. Марковские модели моделируют сетевой трафик с помощью конечного числа состояний. Обладают марковским свойством: будущее состояние зависит только от текущего состояния сетевого трафика.

Модель цепочки сообщений

В данной модели считается, что пакеты передаются и обрабатываются как единое целое [4]. Трафик представляет собой последовательности пакетов от источника к месту назначения. Решение об обработке последующих сообщений принимается по первому сообщению. Такая модель подходит для исследования, а также категоризации трафика, если у сообщения один пункт назначения.

Самоподобная модель

В первые о самоподобном трафике начали говорить в начале 90-х годов, в связи с увеличением объема исследуемых данных. Статьи [5] указывают на то, что объединенный из нескольких источников трафик становится сильно автокоррелированным с долговременной зависимостью [6]. Такое поведение может быть объяснено тем, что будущее процесса определяется его прошлым поведение, при чем с убывающей степенью влияния этого прошлого на процесс. Совокупность множества источников данных, проявляющих синдром бесконечной дисперсии, в результате дает самоподобный объединенный сетевой трафик. Применение классических моделей к подобному трафика не отражает действительности и приводит к недооценке действительности.

1.3 Самоподобие и фрактальная размерность

Временные ряды, полученные при изучении сетевого трафика, обладают такими свойствами как самоподобие и фрактальная размерность.

В математике под фракталами (лат. fractus — дроблёный, сломанный, разбитый) понимают множества точек в евклидовом пространстве, имеющие дробную метрическую размерность, либо метрическую размерность, отличную от топологической, поэтому их следует отличать от прочих геометрических фигур, ограниченных конечным числом звеньев.

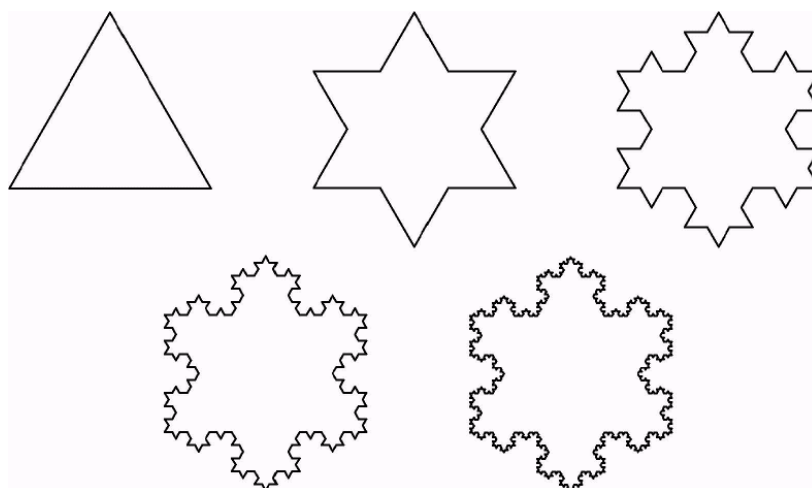


Рисунок 1.1 – Снежинка Коха

Важное свойство, которым обладают почти все фракталы — свойство самоподобия. Фрактал можно разбить на малые части так, что каждая часть окажется просто уменьшенной версией глобального целого (т.е. подобна исходному объекту). Снежинка Коха — пример фрактала, представлена на рисунке 1.1. Природные объекты по своей сути являются фракталами, и их создание обычно определяется уравнениями хаоса. Примеры природных фракталов представлены на рисунке 1.2. Использование теории хаоса для исследования самоподобного сетевого трафика является перспективным направлением и позволяет развить идею самоподобности трафика. Стохастические фрактальные процессы описываются масштабной инвариантностью статистических характеристик второго порядка — свойство неизменности коэффициента корреляции при масштабировании. Как раз с такими стохастическими фракталами имеют дело при изучении характеристик сетевого трафика.

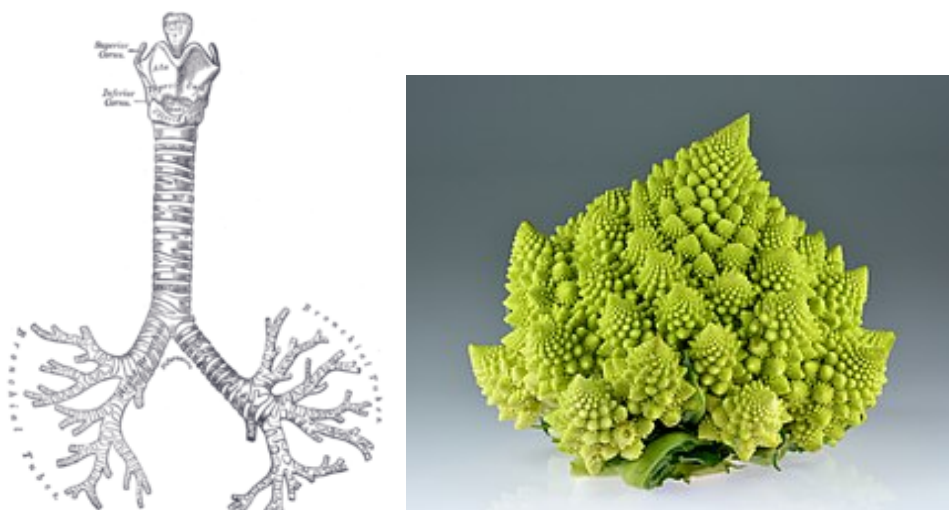


Рисунок 1.2 – Природные фракталы (слева – трахея и бронхи, справа – кочан капусты романеско)

Стохастические фрактальные процессы не обладают четким сходством между частями объекта и глобальным целым, но свойство самоподобия для таких объектов можно показать математически и проиллюстрировать. Далее представлен обзор параметров, по которым мы можем оценить степень самоподобия того или иного временного ряда.

1.4 Коэффициент Херста и RS-анализ

Коэффициент Херста является важнейшим параметром, характеризующим степень самоподобия. Этот параметр был назван в честь Х.Е. Херста – британского гидролога, который посвятил себя изучению реки Нил, а также проблеме хранения воды. Оценка параметра Херста помогает не только решить, является ли процесс самоподобным, но и позволяет применить к процессу ряд методов по прогнозированию фрактальных процессов.

Коэффициент Херста принимает значения $0 < H < 1$.

- При значении коэффициента $0.5 < H < 1$ говорят о *персистентном (поддерживающемся) поведении* процесса, либо о том, что процесс обладает длительной памятью, то есть является самоподобным. Персистентные стохастические процессы обнаруживают четко выраженные тенденции изменения при относительно малом “шуме”.

- В случае $H = 0.5$ говорят о *полностью случайном ряде*, аналогичном смещениям частицы при классическом броуновском движении.
- В случае $0 < H < 0.5$ говорят о *антиперсистентности процесса*. Такой ряд не обладает самоподобием.

Существуют различные методы для оценки параметра H для того, чтобы выявить самоподобие или медленно убывающую зависимость.

Временные методы оценки параметра Херста:

- Метод R/S статистики (В основе R/S анализа лежит формула Альберта Эйнштейна о броуновском движении частиц).
- Метод вариаций
- Метод абсолютного момента
- Метод отношения вариации остатков
- Частотные методы оценки параметра Херста:

Далее в работе будет использован метод R/S статистики.

Показатель Херста, H , определяется в терминах асимптотического поведения масштабированного диапазона как функции отрезка времени временного ряда следующим образом:

$$E \left[\frac{R(n)}{S(n)} \right] = C * n^H, n \rightarrow \infty$$

В данной формуле R – размах накопленных отклонений n значений от среднего значения ряда, S - среднеквадратическое отклонение, n – величина промежутка времени (количество точек в отрезке временного ряда), C – константа.

Данный метод легко применим на практике, обладает достаточным уровнем надежности, но может быть чувствителен к длине временного ряда (чувствительность метода выражается в том, что он плохо оценивает локальный показатель Херста).

размерностью

Алгоритм RS-анализа

1. Дан исходный временной ряд Y_t . Рассчитаем логарифмические отношения

$$L_t = \ln \frac{Y_t}{Y_{t-1}}.$$

2. В полученном ряду L_t последовательно выделим K отрезков длины n .
Обозначим полученные отрезки следующим образом:

$$L_k, \quad k = \overline{1, K}.$$

3. Для каждого из этих отрезков высчитывается среднее значение.

$$E[L_k] = \frac{1}{n} \sum_{i=1}^n L_{k,i}, \quad k = \overline{1, K}.$$

4. Далее рассчитаем размах накопленных отклонений для каждого из отрезков:

$$X_{k,t} = \sum_{i=1}^t (L_{k,i} - E[L_k]), \quad t = \overline{1, n}, \quad k = \overline{1, K}$$

$$R_k = \max_{t=1 \dots n} (X_{k,t}) - \min_{t=1 \dots n} (X_{k,t}), \quad k = \overline{1, K}$$

5. Рассчитаем среднеквадратичное отклонение для каждого отрезка:

$$S_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (L_{k,i} - E[L_k])^2}, \quad k = \overline{1, K}$$

6. Высчитываем значение статистики:

$$\frac{R}{S}(n) = \frac{1}{K} \sum_{k=1}^{K-1} \frac{R_k}{S_k}$$

7. Повторяем алгоритм для следующего значения n .

Получают следующую последовательность точек:

$$H(n) = \frac{\ln(R(n)/S(n))}{\ln(n)},$$

где $n = N, \frac{N}{2}, \frac{N}{4}, \dots, N$ – количество элементов в исходном временном ряду.

Далее решаем задачу регрессии $\ln\left(\frac{R}{S}(n)\right) = H \log(n) + c$ по известным точкам $(H(n), n)$, строим линию тренда и находим аппроксимированное значение показателя Херста [7, 8].

1.5 Фрактальные параметры

Существуют и иные фрактальные параметры, которые позволяют оценивать степень самоподобия временного ряда, а также связанные с ним свойства.

Фрактальная размерность

Мы знаем, что фрактальный объект, с математической точки зрения, обладает дробной размерностью, которую можно вычислить по следующей формуле:

$$D = \frac{\log N}{\log 1/r},$$

где N — количество равных частей, на которое делят объект, а r — коэффициент масштаба.

Фрактальная размерность может рассматриваться как мера неровности изрезанной поверхности объекта $D \in [n, n + 1)$ в n -мерном пространстве, причем более неровные поверхности соответствуют высоким значениям D .

Фрактальная размерность выражается через коэффициент Херста по следующей формуле:

$$D = 2 - H.$$

Фрактальная размерность обобщает понятие евклидовой геометрической размерности и, в отличие от последней, может принимать нецелочисленные значения. С помощью данной размерности можно классифицировать процессы по свойствам хаотичности или сложности.

Корреляционный параметр

Введем понятие автокорреляционной функции. Пусть X_t значение случайного процесса в момент времени t (t может быть как вещественным, если

процесс непрерывный, так и целым, если процесс дискретный). Если X_t имеет среднее значение μ дисперсию σ^2 , то автокорреляция X_t определяется следующим образом:

$$r(t-s) = r(t,s) = \frac{E[(X_t - \mu)(X_s - \mu)]}{\sigma^2}.$$

Случайный процесс X_t обладает *медленно убывающей зависимостью*, если для его автокорреляционной функции выполняется условие:

$$r(k) \sim k^{-\beta} L_1(k) + C_1, \quad k \rightarrow \infty,$$

где $C_1 = \text{const}$, $0 < \beta < 1$ – корреляционный параметр, L_1 — медленно меняющаяся на бесконечности функция $\lim_{t \rightarrow \infty} \frac{L_1(tx)}{L_1(t)} = 1, \forall x$.

Процессы с медленно убывающей зависимостью характеризуются автокорреляционной функцией, которая убывает по степенному закону при увеличении временной задержки.

Связь корреляционного параметра с показателем Херста определяет формула:

$$\beta = 2(1 - H).$$

Спектральный показатель

В частотной области медленно убывающая зависимость приводит к степенному закону поведения спектральной плотности процесса.

Пусть $r(t)$ – автокорреляционная функция процесса X_t , тогда функция $S(f) = \int_{-\infty}^{\infty} r(\tau) e^{-i2\pi f\tau} d\tau$ называется *спектральной плотностью процесса X_t* .

Случайный процесс X_t обладает *медленно убывающей зависимостью*, если для спектральной плотности $S(f)$ выполняется следующее условие:

$$S(f) \sim f^{-b} L_2(f) + C_2,$$

где $f \rightarrow 0$, $0 < b < 4$, L_2 — медленно изменяющаяся в нуле функция, $C_2 = \text{const}$.

Зависимость спектрального показателя от показателя Херста:

$$b = 2H + 1.$$

Фрактальный показатель

Фрактальный показатель α – это мера того, насколько островершинной является функция плотности вероятности. В нормальном распределении $\alpha = 2$. В случае распределений Парето - $0 < \alpha < 2$. Данный показатель характеризует степень тяжести «хвоста» распределения (индекс «хвоста»).

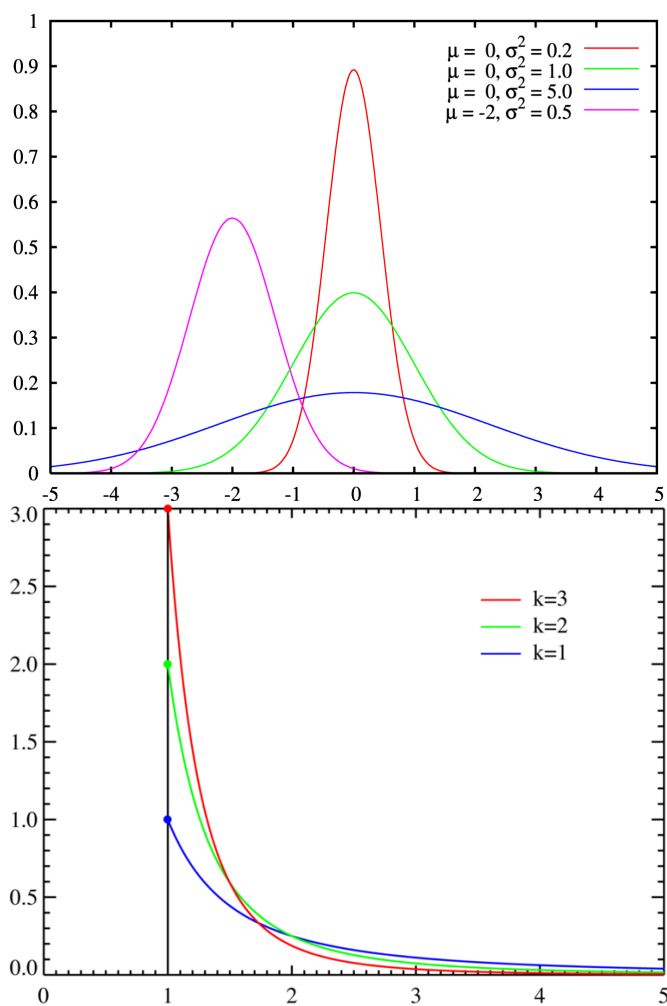


Рисунок 1.3 – Плотности распределений

(сверху – нормальное распределение, снизу –распределение Парето)

Случайная величина ξ имеет распределение с “тяжелым хвостом”, если:

$$P(\xi \geq x) = 1 - F(x) = x^{-\alpha} L_1(x),$$

где $L_1(x)$ — медленно меняющаяся функция на бесконечности, параметр $0 < \alpha < 2$ — индекс «хвоста».

Параметр α связан с параметром Херста H следующим соотношением:

$$\alpha = 3 - 2H.$$

Так как все представленные выше фрактальные параметры могут быть выражены через коэффициент Херста, то в дальнейшей работе будем использовать его.



Рисунок 1.4 – Связь фрактальных параметров друг с другом

ГЛАВА 2 УСТОЙЧИВОЕ РАСПРЕДЕЛЕНИЕ

Многочисленные измерения сетевого трафика показали, что он лучше всего описывается так называемыми распределениями с “тяжелыми хвостами”.

2.1 Понятие устойчивого распределения

Рассмотрим понятие устойчивого распределения. *Устойчивое распределение* — это такое распределение, которое может быть получено как предел по распределению сумм независимых случайных величин [9].

Пусть ξ — случайная величина, заданная на вероятностном пространстве (Ω, F, P) с функцией распределения $F_\xi(t), t \in \mathcal{R}$.

Пусть $\xi_i, i = 1, 2, \dots$ — последовательность независимых одинаково распределенных случайных величин с функциями распределения $F_{\xi_i}(t), t \in \mathcal{R}, i = 1, 2, \dots$, заданных на том же вероятностном пространстве (Ω, F, P) . Причем $F_\xi(t) = F_{\xi_i}(t), t \in \mathcal{R}, i = 1, 2, \dots$.

Случайная величина ξ , ее функция распределения $F_\xi(x), x \in \mathcal{R}$, плотность распределения $p_\xi(x), x \in \mathcal{R}$, характеристическая функция $\varphi_\xi(t), t \in \mathcal{R}$ называются *устойчивыми (устойчивыми в широком смысле)*, если для каждого $m, m = 1, 2, \dots$, существуют такие постоянные $c_m > 0$ и $\gamma_m \in \mathcal{R}$, что для $x \in \mathcal{R}, t \in \mathcal{R}$ выполняются следующие равенства:

$$\begin{aligned}\sum_{i=1}^m \xi_i &= c_m \xi + \gamma_m, \\ F_{\xi_1} * \dots * F_{\xi_m} &= F_\xi \left(\frac{x - \gamma_m}{c_m} \right), \\ p_{\xi_1} * \dots * p_{\xi_m} &= p_\xi \left(\frac{x - \gamma_m}{c_m} \right), \\ \prod_{i=1}^m \varphi_{\xi_i}(t) &= \varphi_\xi(c_m t) e^{i\gamma_m t}.\end{aligned}$$

Случайная величина ξ , ее функция распределения $F_\xi(x), x \in \mathcal{R}$, плотность распределения $p_\xi(x), x \in \mathcal{R}$, характеристическая функция $\varphi_\xi(t), t \in \mathcal{R}$ имеют *строго устойчивое распределение (устойчивое в узком смысле)*, если в приведенных ранее равенствах $\gamma_m = 0$.

Устойчивое распределение с параметрами $(\alpha, \beta, \mu, \sigma)$ однозначно определяется с помощью характеристической функции:

$$\varphi(\alpha, \beta, \mu, \sigma) = \begin{cases} \exp \left[i\mu t - \sigma |t| \left(1 + \frac{2i\beta}{\pi} \operatorname{sgn}(t) \ln |t\sigma| \right) \right], & \alpha = 1 \\ \exp \left[i\mu t - |t\sigma|^\alpha \left(1 + i\beta \tan \frac{\pi\alpha}{2} \operatorname{sgn}(t) (|t\sigma|^{1-\alpha} - 1) \right) \right], & \alpha \neq 1 \end{cases}$$

где $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\mu \in \mathcal{R}$, $\sigma > 0$.

Параметры распределения обладают следующим смыслом:

- α – характеристический параметр (описывает тяжесть хвостов распределения).
- β – параметр асимметрии (описывает скошенность распределения).

При $\beta = 0$ распределение является симметричным.

- μ – параметр положения.
- σ – параметр масштаба.

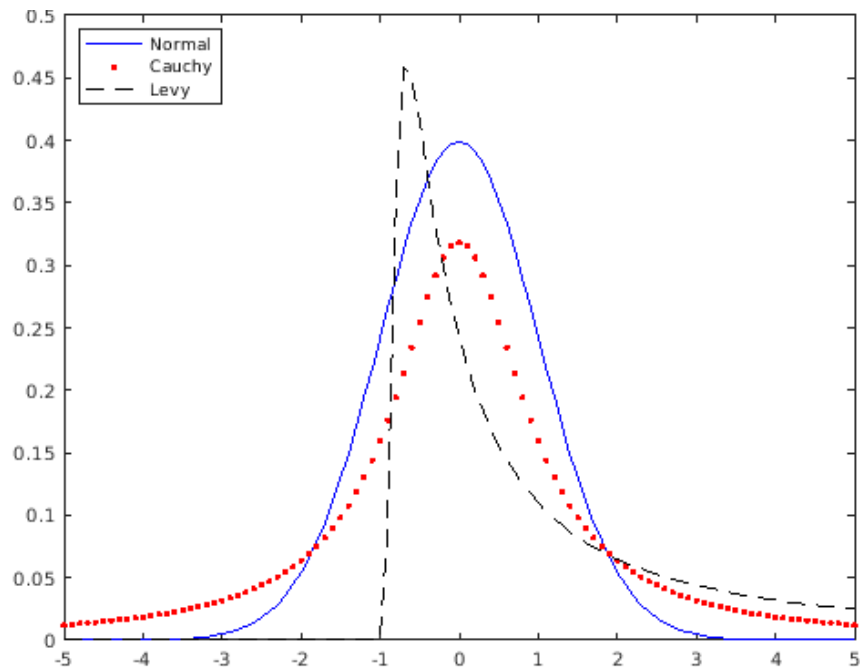


Рисунок 2.1 – Сравнение различных устойчивых распределений

2.2 Метод максимального правдоподобия

Метод максимального правдоподобия – это метод оценивания неизвестных параметров с помощью максимизации функции максимального правдоподобия.

Пусть есть независимая выборка $X = \{x_1, \dots, x_n\}$ из распределения P_θ , где $\theta \in \Theta$ – вектор неизвестных параметров. Пусть $L(x|\theta) = \prod_{i=1}^n f_X(x_i|\theta)$ – функция правдоподобия, где $x \in R$, f_X – плотность распределения выборки.

Точечная оценка

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(x|\theta)$$

называется *оценкой максимального правдоподобия* параметра θ .

Таким образом оценка максимального правдоподобия — это такая оценка, которая максимизирует функцию правдоподобия при фиксированной реализации выборки.

Для упрощения вычислений часто используется логарифмическая функция максимального правдоподобия $l = \ln L(x|\theta) = \sum_{i=1}^n \ln f_X(x_i|\theta)$.

Для устойчивого распределения функция максимального правдоподобия принимает следующий вид:

$$l_n(\alpha, \beta, \mu, \sigma) = \sum_{i=1}^n \ln f(x_i|\alpha, \beta, \mu, \sigma),$$

где $x_i, i = \overline{1, n}$ – исследуемая выборка, f - аппроксимированная с помощью дискретного преобразования Фурье плотность устойчивого распределения.

2.3 Обобщенный метод моментов

Пусть есть выборка независимых одинаково распределенных случайных величин $X = \{x_1, \dots, x_n\}$ некоего устойчивого распределения с параметрами $\theta = (\alpha, \beta, \mu, \sigma)$ с функцией распределения $F(x)$. Тогда $k = 4$ – количество оцениваемых параметров.

Пусть также есть набор моментных функций

$$g_i(x, \theta), i = \overline{1, q}, q \geq k$$

таких, что выполняются моментные условия

$$m_i(\theta) = E[g_i(x, \theta)] = 0.$$

Идея метода заключается в том, чтобы использовать в моментных условиях вместо математических ожиданий их выборочные аналоги:

$$\widehat{m}_i(\theta) = \overline{g_i(x, \theta)},$$

которые согласно закону больших чисел асимптотически сходятся к истинным математическим ожиданиям.

Обобщенным методом моментов называется оценка, которая минимизирует положительно определенную квадратичную форму от выборочных условий на моменты:

$$\hat{\theta} = \arg \min_{\theta} \widehat{m}(\theta)^T W m(\theta),$$

где W – некоторая симметрическая положительно определенная матрица.

В эффективном обобщенном методе моментов в качестве весовой матрицы используется ковариационная матрица моментных функций $W = V_g^{-1}$.

2.4 Итеративный метод Кутровелиса

Пусть есть выборка независимых одинаково распределенных случайных величин $X = \{x_1, \dots, x_n\}$ из некоторого распределения. Тогда определим выборочную характеристическую функцию следующим образом:

$$\hat{\varphi}(t) = \frac{1}{n} \sum_{j=1}^n e^{itx_j}, |\hat{\varphi}(t)| \leq 1.$$

Данная оценка характеристической функции $\varphi(t)$ является состоятельной.

В 1980 году Кутровелис предложил метод регрессионного типа, который начинается с начальной оценки параметров и продолжается, пока не будет удовлетворен какой-либо критерий сходимости.

Метод основан на следующих наблюдениях. Во-первых, из логарифма характеристической функции устойчивого распределения

$$\ln \varphi(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \left\{ 1 - i\beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2} \right\} + i\mu t, & \alpha \neq 1, \\ -\sigma |t| \left\{ 1 - i\beta \operatorname{sign}(t) \frac{2}{\pi} \ln(t) \right\} + i\mu t, & \alpha = 1, \end{cases}$$

можно вывести следующее равенство:

$$\ln(-\ln|\varphi(t)|^2) = \ln(2\sigma^\alpha) + \alpha \ln |t|.$$

Это уравнение зависит только от α и σ , и можно получить оценки $\hat{\alpha}$ и $\hat{\sigma}$ с помощью регрессии $y = \ln(-\ln|\varphi(t)|^2)$, $\omega = \ln|t|$ на модели:

$$y_k = m + \alpha\omega_k + \varepsilon_k,$$

где $m = \ln(2\sigma^\alpha)$, $t_k = \frac{\pi k}{25}$, $k = 1, 2, \dots, K$ и K выбирается из диапазона от 9 до 134 в зависимости от размеров выборки.

Во-вторых, используя равенство

$$\arctan\left(\frac{\Re(\varphi(t))}{\Im(\varphi(t))}\right) = \mu t + \beta\sigma^\alpha \tan\frac{\pi\alpha}{2} \operatorname{sign}(t)|t|^\alpha, \alpha \neq 1,$$

где $\Re(\varphi(t))$, $\Im(\varphi(t))$ – действительная и мнимая части $\varphi(t)$, при фиксированных $\alpha = \hat{\alpha}$ и $\sigma = \hat{\sigma}$, могут быть получены оценки $\hat{\beta}$ и $\hat{\mu}$.

Далее итеративно повторяется решение описанных выше задач регрессии с заданными начальными параметрами равными $\hat{\alpha}$, $\hat{\sigma}$, $\hat{\beta}$ и $\hat{\mu}$, полученными на предыдущем шаге. Алгоритм продолжается до тех пор, пока не будет выполнен заданный критерий сходимости.[15]

2.5 Метод квантилей МакКаллаха

Пусть есть выборка независимых случайных величин $x_1, \dots, x_n$ некоего устойчивого распределения с параметрами $(\alpha, \beta, \mu, \sigma)$ с функцией распределения $F(x)$. Допустим, что выборка отсортирована по возрастанию. Обозначим за x_p - p -квантиль данного распределения, а за $\widehat{x}_p = x_j, j = [p * n + 1]$ – выборочную p -квантиль.

Введем следующие величины, которые зависят только от параметров α и β [10]:

$$\begin{aligned} v_\alpha &= \frac{x_{.95} - x_{.05}}{x_{.75} - x_{.25}} = \varphi_1(\alpha, \beta), \\ v_\beta &= \frac{x_{.95} + x_{.05} - 2x_{.5}}{x_{.95} - x_{.05}} = \varphi_2(\alpha, \beta), \\ v_\sigma &= \frac{x_{.75} - x_{.25}}{\sigma} = \varphi_3(\alpha, \beta), \\ v_\mu &= \frac{\mu - x_{.5}}{\sigma} = \varphi_4(\alpha, \beta). \end{aligned}$$

И их оценки:

$$\widehat{v}_\alpha = \frac{\widehat{x}_{.95} - \widehat{x}_{.05}}{\widehat{x}_{.75} - \widehat{x}_{.25}},$$

$$\widehat{v}_\beta = \frac{\widehat{x}_{.95} + \widehat{x}_{.05} - 2\widehat{x}_{.5}}{\widehat{x}_{.95} - \widehat{x}_{.05}}.$$

Также введем вспомогательную величину ζ и ее оценку $\hat{\zeta}$ для оценки параметра положения μ :

$$\zeta = \begin{cases} \mu + \beta\sigma \tan \frac{\pi\alpha}{2}, & \alpha \neq 1 \\ \mu, & \alpha = 1 \end{cases},$$

$$\hat{\zeta} = \widehat{x}_{.5} + \hat{\sigma}\varphi_5(\hat{\alpha}, \hat{\beta}).$$

Теперь мы можем выразить интересующие нас оценки параметров $\alpha, \beta, \mu, \sigma$ через введенные выше величины:

$$\hat{\alpha} = \psi_1(\widehat{v}_\alpha, \widehat{v}_\beta),$$

$$\hat{\beta} = \psi_2(\widehat{v}_\alpha, \widehat{v}_\beta),$$

$$\hat{\sigma} = \frac{\widehat{x}_{.75} - \widehat{x}_{.25}}{\varphi_3(\hat{\alpha}, \hat{\beta})},$$

$$\hat{\mu} = \hat{\zeta} - \hat{\beta}\hat{\sigma} \tan \frac{\pi\hat{\alpha}}{2}.$$

Значения функций $\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \psi_1, \psi_2$, используемых выше, представлены в таблицах в Приложении А [11].

ГЛАВА 3 АНАЛИЗ РЕАЛЬНЫХ ДАННЫХ СЕТЕВОГО ТРАФИКА

Для последующего исследования были собраны реальные данные о сетевом трафике на предприятии ИООО «Техниксофт».

3.1 Постановка задачи

Как было сказано ранее классическая модель сетевого трафика не отражает реальности: данные реального сетевого трафика при больших объемах обладают свойством самоподобия. Для моделирования такого трафика используется устойчивое распределение.

Задачи:

1. Собрать данные на предприятии, провести их предварительный анализ;
2. Смоделировать полученный сетевой трафик с помощью устойчивого распределения.
3. Провести анализ сетевого трафика на свойство самоподобия.

3.2 Описание данных

В качестве данных для исследования были предоставлены две выборки данных сетевого трафика, собранных за различные периоды времени.

Первая выборка представляет из себя 529 наблюдений за период времени 12 часов: с 08:20 15-го января 2021 до 20:20 этого же дня. Одно наблюдение представляет из себя количество пакетов за минуту.

Вторая выборка представляет из себя 620 наблюдений, собранных за четыре недели: с 17:00 24-го января 2021 до 12:00 19-го февраля 2021. Одно наблюдение представляет из себя количество пакетов, полученных за час.

Данные были собраны с помощью такого инструмента как NetFlow Analyzer и предоставлены в виде набора .csv файлов. Пример собранных данных представлен в Приложении Б. Файлы с данными были предобработаны и

подготовлены для последующего моделирования и исследования с помощью программного кода, написанного на языке Python.

Далее приведено описание первой выборки.

Посмотрим на то, как распределено количество пакетов в зависимости от времени.

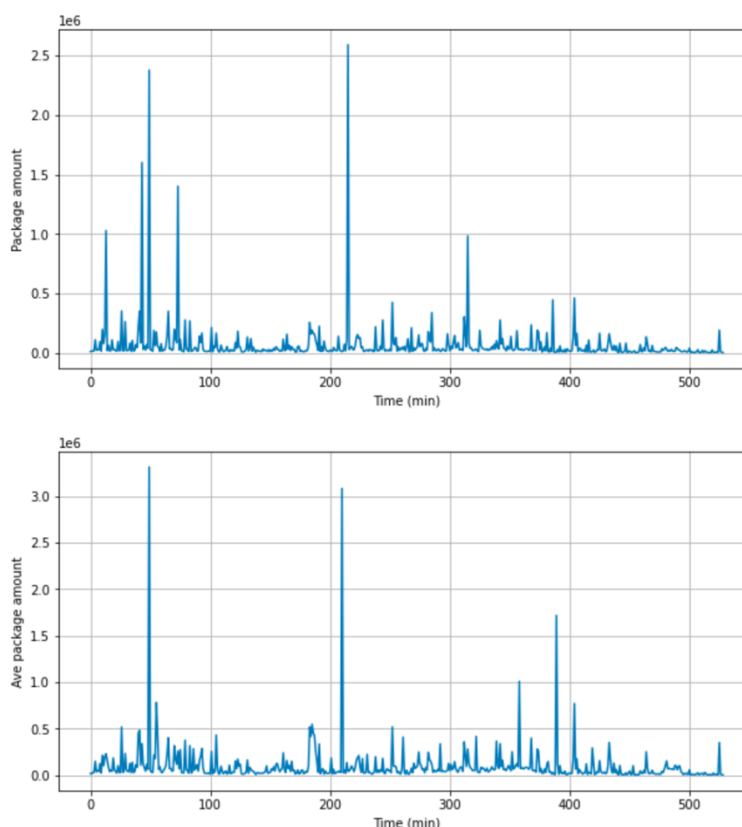


Рисунок 3.1 – График количества пакетов в зависимости от момента времени (сверху – входящих, снизу - исходящих)

Таблица 3.1 – Сводная таблица характеристик для исходных данных по количеству входящих и исходящих пакетов

	Min	25%	50%	Mean	75%	Max
In	3297	12686	25476	69028.1	55640	2592923
Out	3363	23510	47925	99725.7	96182	3311913

Из рисунка 3.1 и сводной таблицы 3.1 видно, что у нас присутствуют скачки активности, которые выходят за обычные значения для нашего трафика.

3.3 Предобработка данных

Будем рассматривать данные по количеству входящих пакетов. Для начала получим следующий временной ряд и рассмотрим его распределение:

$$g_t = \ln(x_t), t = \overline{1, n},$$

где t – момент времени, x_t – значение исходного временного ряда в момент времени t , n – количество наблюдений. График зависимости преобразованной величины и ее гистограмма представлены на рисунках 3.2 и 3.3.

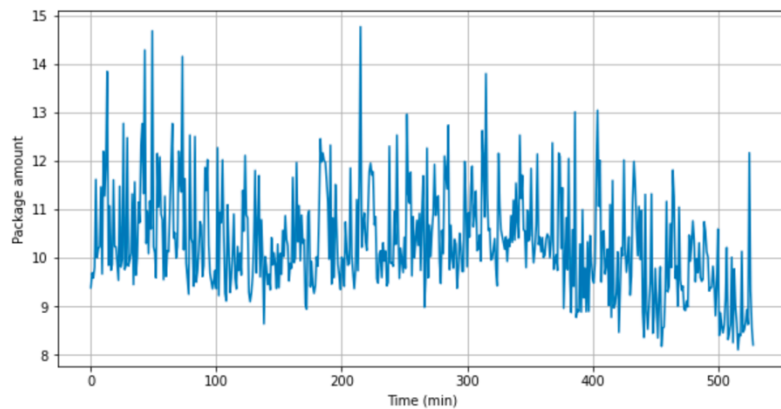


Рисунок 3.2 – График зависимости величины g_t от момента времени t

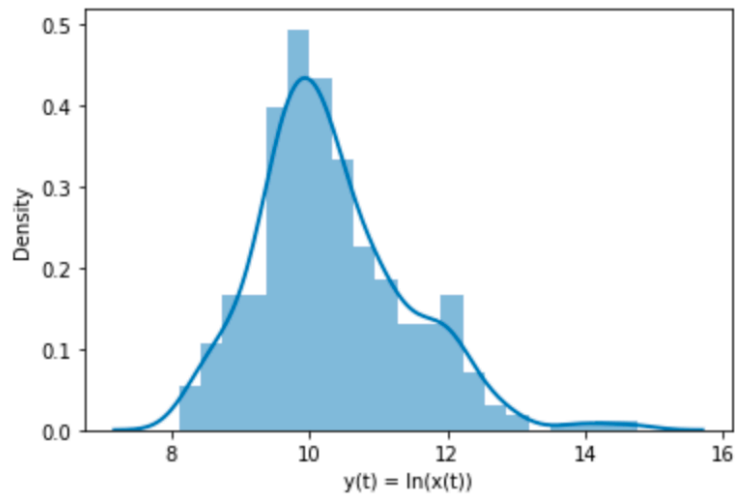


Рисунок 3.3 – Гистограмма распределения величины g_t

Для того, чтобы произвести оценивание параметров устойчивого распределения, проведем еще одно преобразование наших данных:

$$r_t = \ln\left(\frac{x_{t+1}}{x_t}\right), t = \overline{1, n-1},$$

где t – момент времени, x_t – значение исходного временного ряда в момент времени t , n – количество наблюдений.

Физический смысл данного преобразования – приращение логарифма количества пакетов в момент времени t за заданный промежуток времени.

Посмотрим на то, как выглядят данные после описанного выше преобразования.

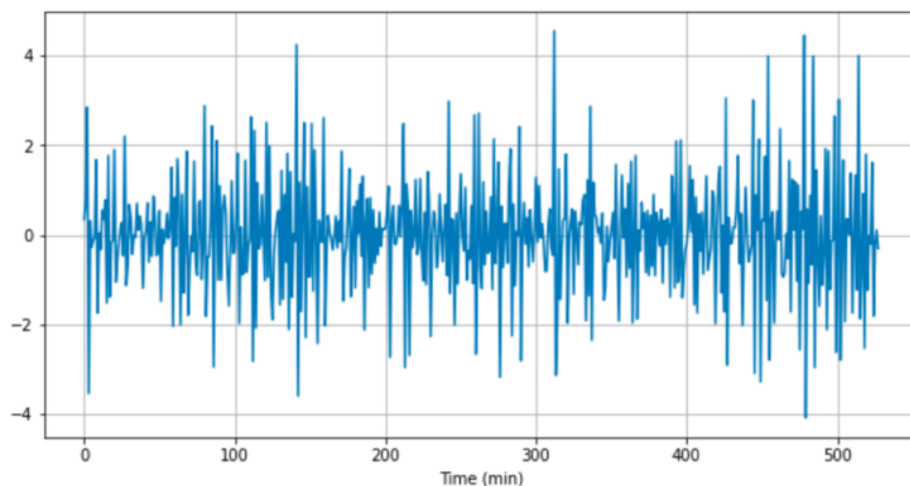


Рисунок 3.4 – График зависимости величины r_t от момента времени t

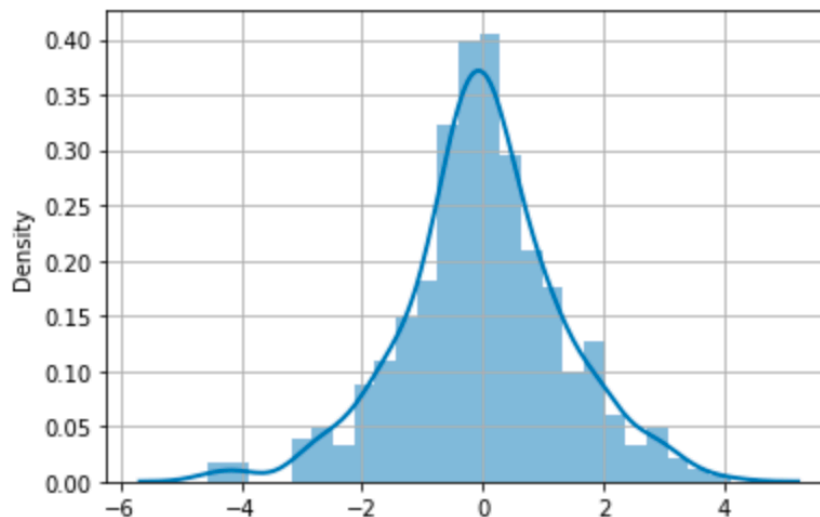


Рисунок 3.5 – Гистограмма распределение величины r_t

3.4 Оценка параметров устойчивого распределения

Для оценки параметров устойчивого распределения воспользуемся пакетом PyLevy [12], представленным в языке Python, а также пакетом StableEstim [13], представленными в языке R.

В таблицах 3.2 и 3.3 представлены результаты оценивания параметров устойчивого распределения для величины g_t и r_t .

Таблица 3.2 – Сводная таблица оценок параметров устойчивого распределения величины g_t

	Alpha	Beta	Mu	Sigma
Метод максимального правдоподобия	1.7487510	0.9990000	10.1120248	0.6761528
Итеративный метод Кутровелиса	1.8030253	0.9990000	10.0607396	0.6694596
Метод квантилей	1.761000	0.853000	10.060740	0.683906
Обобщенный метод МОМЕНТОВ	1.6764908	0.9990000	10.0777521	0.6645146

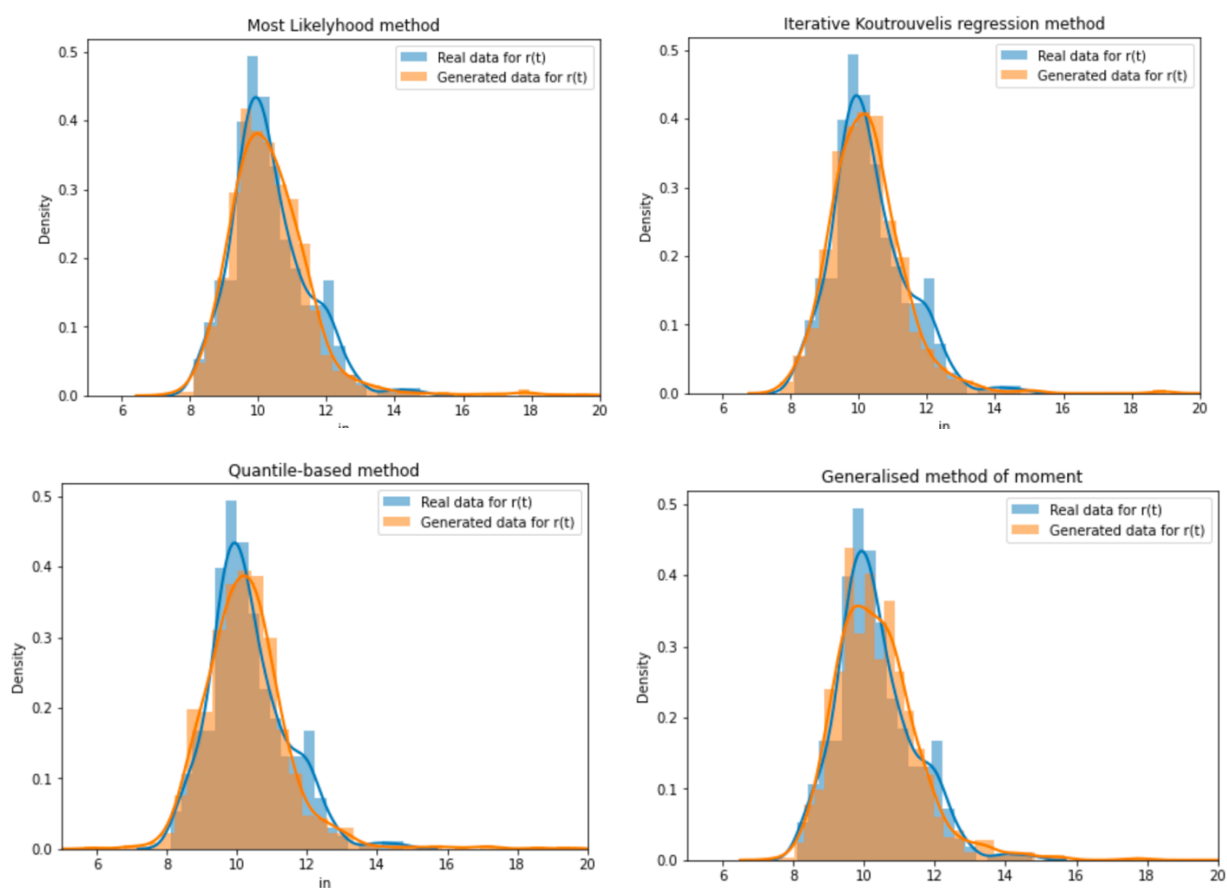


Рисунок 3.6 – Сравнение распределений настоящих и сгенерированных данных для величины g_t

Таблица 3.3 – Сводная таблица оценок параметров устойчивого распределения величины r_t

	Alpha	Beta	Mu	Sigma
PyLevy (метод максимального правдоподобия)	1.87	-0.17	0.03	0.87
StableEstim (метод максимального правдоподобия)	1.8710145	-0.1659061	0.02826065	0.8696862
StableEstim (итеративный метод Кутровелиса)	1.8228606	-0.1300519	0.0184439	0.8187967
StableEstim (метод квантилей)	1.502	-0.078	-0.0045645	0.72803641
StableEstim (обобщенный метод моментов)	1.7109628	0.0453521	-0.00040278	0.83122871

На рисунке 3.7 предоставлено визуальное сравнение распределений реальных данных и данных, полученных с помощью оценок параметров устойчивого распределения. В таблице 3.4 представлены результаты теста Колмогорова-Смирнова для проверки гипотезы о принадлежности выборки величины r_t оцененным законам распределения. Исходя из полученных результатов мы можем определить, что метод квантилей даст лучший результат в оценке параметров. Таким образом, при последующем моделировании сетевого трафика будем использовать оценки, полученные с помощью метода квантилей.

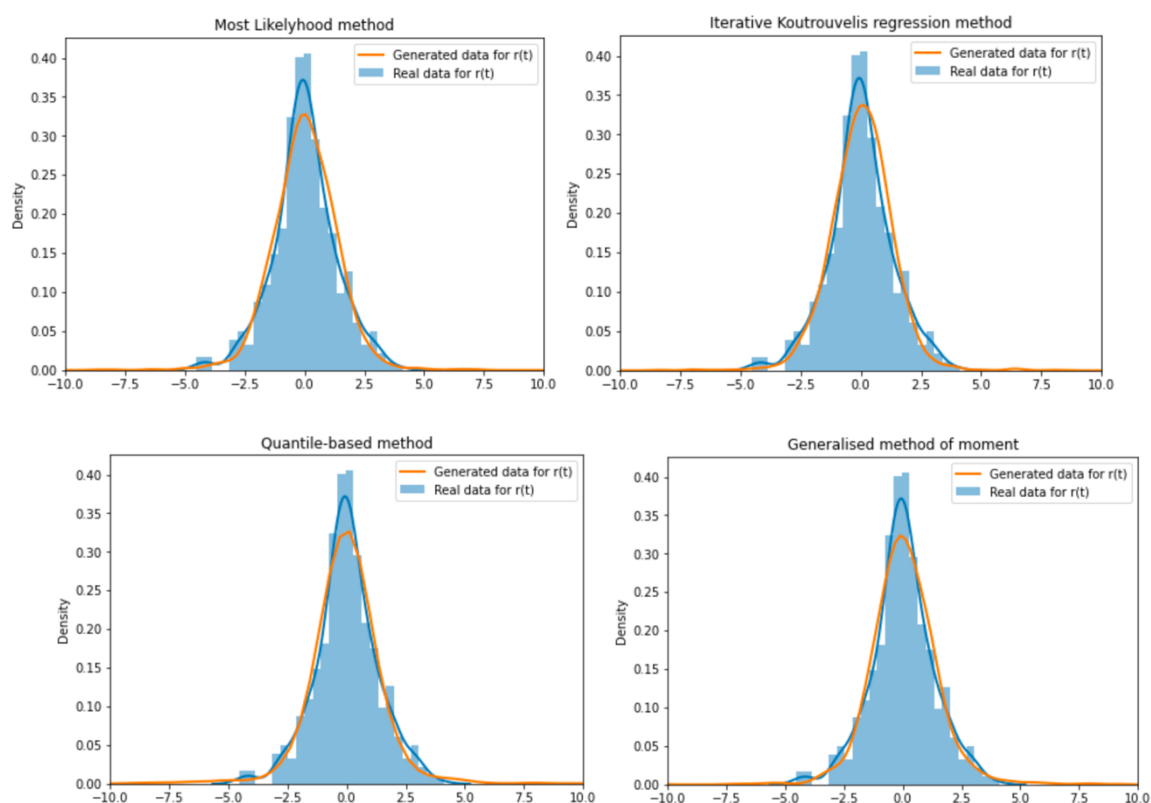


Рисунок 3.7 – Сравнение распределений настоящих и сгенерированных данных для величины r_t

Таблица 3.4 – Сводная таблица результатов теста Колмогорова-Смирнова для сгенерированных данных величины r_t

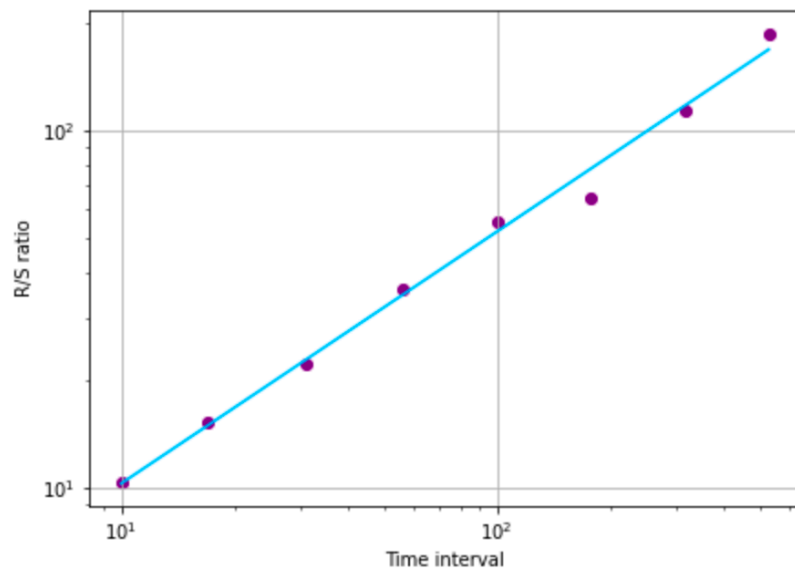
	KS-statistics	p-value
Метод максимального правдоподобия	0.050242	0.233116
Итеративный метод Кутровелиса	0.057242	0.123784
Метод квантилей	0.035833	0.653750
Обобщенный метод моментов	0.049848	0.246193

3.5 Анализ самоподобия

Как было описано ранее, самоподобный объект – это объект, часть которого целиком или приближенно совпадает с самим уменьшенным объектом. Также ранее были приведены характеристики, которые могут быть

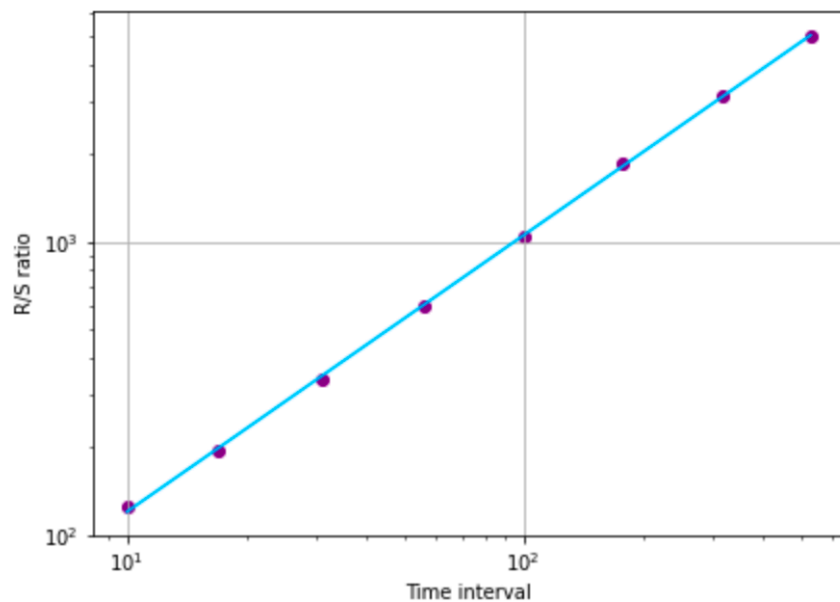
использованы при оценке самоподобности объекта. Все характеристики связаны с показателем Херста, его и используем в дальнейшем анализе.

Для того, чтобы оценить показатель Херста, воспользуемся пакетом «hurst» [14] для языка Python.



$$H=0.7046, \quad c=2.0317$$

Рисунок 3.8 – Результаты R/S анализа исходного временного ряда



$$H=0.9418, \quad c=13.7596$$

Рисунок 3.9 – Результаты R/S анализа для величины g_t

Для обеих величин показатели Херста больше граничного значения 0.5, следовательно, данные временные ряды обладают свойством самоподобия.

3.6 Исследование данных, собранных в течение четырех недель

Как было сказано ранее, вторая выборка представляет из себя 620 наблюдений, собранных в течение четырех недель. Одно наблюдение представляет из себя количество пакетов, полученных за час. Далее будут приведены результаты только для входящих пакетов

Таблица 3.5 – Сводная таблица характеристик для второй выборки

	Min	25%	50%	Mean	75%	Max
In	159439	413472.75	675790	1481865.225	2297970.25	8035464
Out	290544	377574.75	554595	1067225.021	1537639	4336352

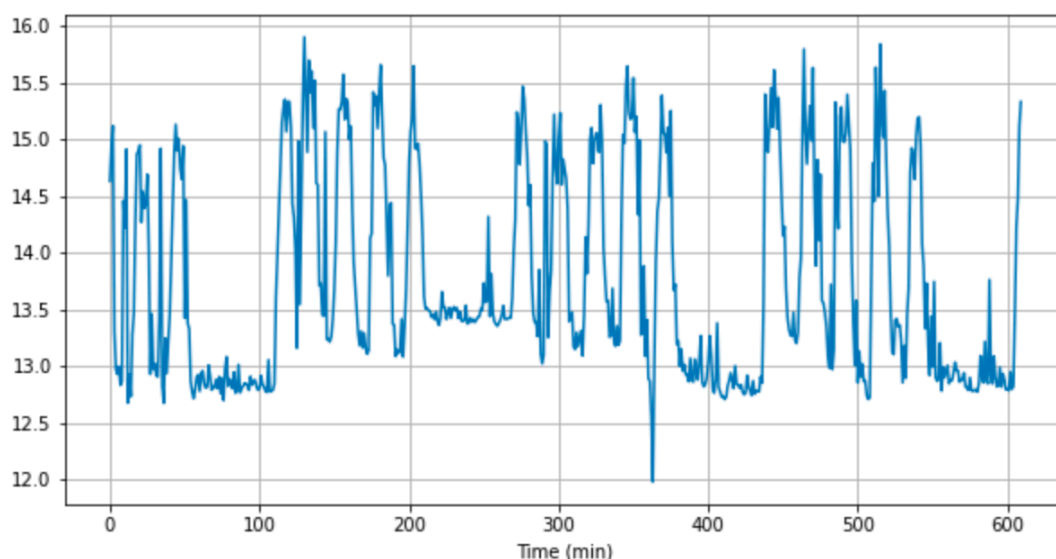


Рисунок 3.10 – График логарифмированного временного ряда

После логарифмирования исходного временного ряда, получим временной ряд, представленный на рисунке 3.10. На рисунке 3.11 представлена гистограмма распределения значений полученного ряда. Таким образом, можем видеть, что такое распределение не может быть смоделировано с помощью устойчивого распределения, так оно является бимодальным. Продолжим анализ временного ряда с помощью введенных ранее приращений.

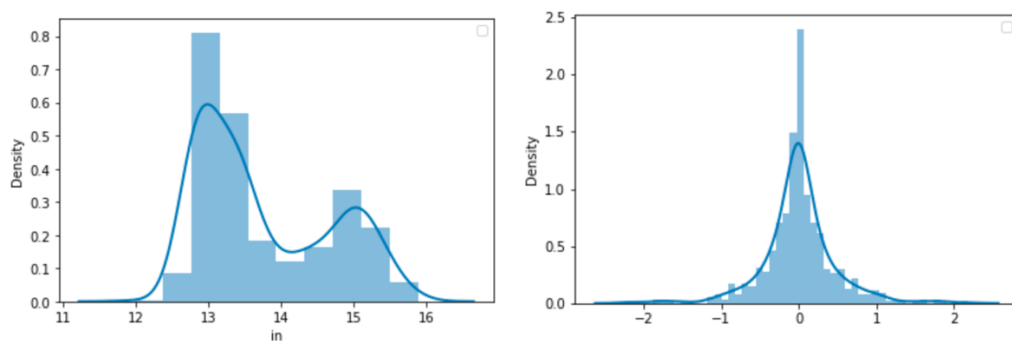


Рисунок 3.11 – Графики преобразованных временных рядов
(слева - g_t , справа - r_t)

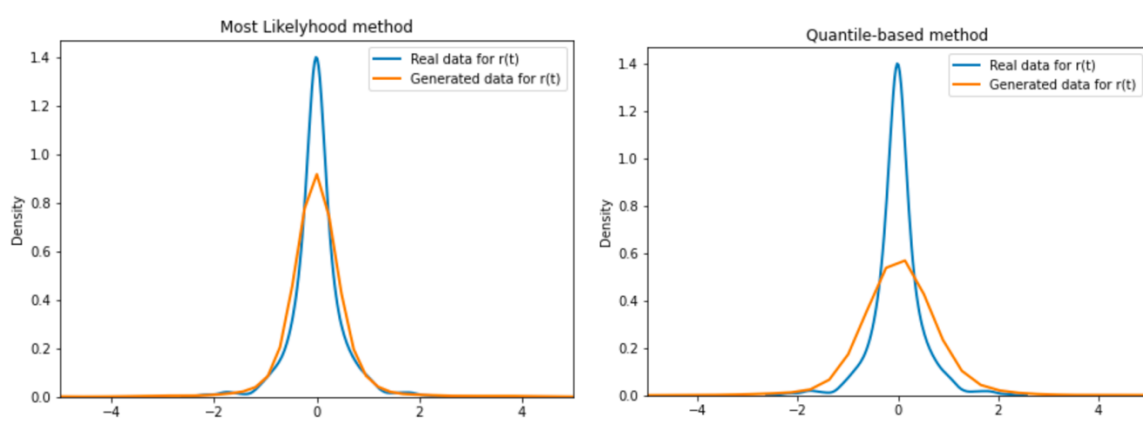


Рисунок 3.12 – Сравнение распределений настоящих и сгенерированных
данных для величины r_t

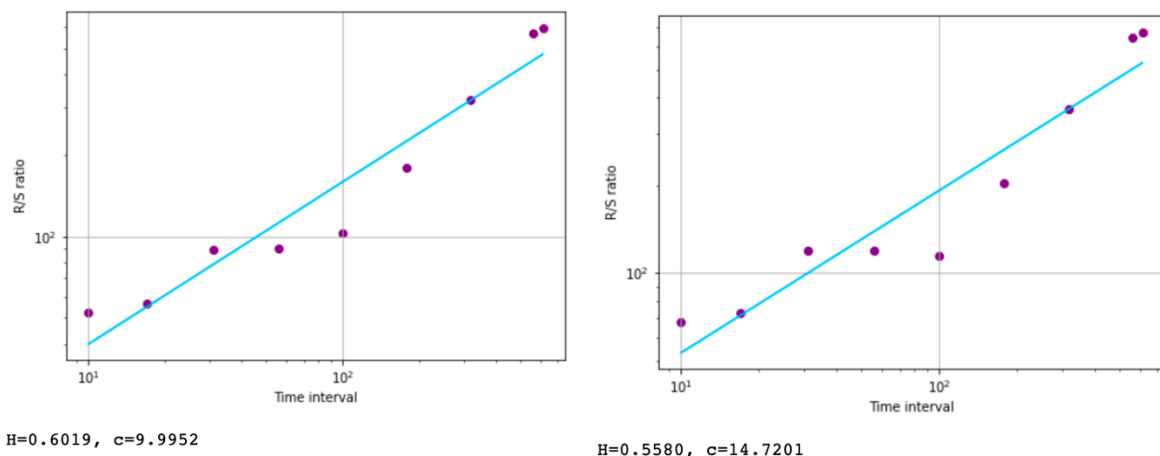
Таблица 3.6 – Сводная таблица оценок параметров устойчивого
распределения величины r_t

	Alpha	Beta	Mu	Sigma
Метод максимального правдоподобия	1.22498	0.03384	-0.00725338	0.180717
Метод квантилей	1.238	0.102	-0.00448875	0.182427

Показатели Херста для исходного и преобразованного временных рядов равны $H = 0.6019$ и $H = 0.558$, что больше граничного значения 0.5, следовательно, наш исходный временной ряд обладают свойством самоподобия.

Таблица 3.7 – Сводная таблица результатов теста Колмогорова-Смирнова для сгенерированных данных

	KS-statistics	p-value
Метод максимального правдоподобия	0.072985	0.033297
Метод квантилей	0.070445	0.018345



Результаты моделирования представлены на рисунке 3.12 и таблица 3.6 и 3.7. Получили, что данные, которые собирались в течение четырех недель плохо моделируются с помощью устойчивого распределения. Это может быть вызвано периодичностью. По причине того, что данные собраны на предприятии, в данных присутствуют пики во время рабочей недели, а во время выходных активность отсутствует. В дальнейшем предлагается увеличить выборку наблюдений: собрать данные за 3 месяца и/или уменьшить промежуток агрегации данных; спрогнозировать временной ряд с помощью SARIMA модели.

ЗАКЛЮЧЕНИЕ

Во время работы над магистерской диссертацией были получены следующие результаты:

- изучены различные методы и принципы формирования данных для анализа сетевого трафика;
- собраны данных сетевого трафика компании ИООО «Техниксофт», произведен предварительный анализ полученных данных, а также их предобработка;
- была проведена оценка параметров устойчивого распределения, осуществлено моделирование сетевого трафика на основе полученных оценок;
- сетевой трафик исследован на свойство самоподобия.

При анализе результатов, полученных при генерации данных с помощью устойчивого распределения, мы выяснили, что наилучшую оценку параметров дает метод квантилей, специально разработанный для оценки параметров устойчивого распределения (реализован и в языке R). Также был проведен R/S анализ исходного временного ряда. Результаты анализа показали, что обе выборки данных обладают свойством самоподобия.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. M. W. Lucas, Network flow analysis, Ch. 1, P. 9–11
2. Heyman D.P., Sobel M.J. Stochastic Models in Operations Research. Vol. 1. Stochastic Processes and Operating Characteristics. New York: McGraw-Hill, 1982
3. Kelly F.P. Networks of Queues with Customers of Different Types of Applied Probability. 1975. Vol. 12, no 3. P. 542-554
4. Jain R. Routhier S.A. Packet Trains – Measurement and a New Model for Computer Network Traffic // IEEE Journal on Selected Areas in Communications. Sep.1986. Vol. 4, no. 6. P. 986-995
5. Willinger W., Taqqu M.S., Sherman R., Wilson D.V. Self-Similarity through High Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level // IEEE/ACM Transcriptions on Networking. Feb. 1997. Vol. 5, no. 1. P. 71-86.
6. Ramakrishnan P. Self-Similar Traffic Models. Technical research report.1999
7. Anis, A.A., Lloyd, E.H. The expected value of the adjusted rescaled Hurst range of independent normal summands. 1976 Biometrika 63: P 283-298.
8. Simmross-Wattenberg F., Anomaly Detection in Network Traffic Based on Statistical Inference and alpha-Stable Modeling // IEEE Transactions on Dependable and Secure Computing Jul. 2001. Vol. 8 no. 4 P.494-509
9. Труш Н. Н., Соболева Т.В. Статистический анализ оценок спектральных плотностей устойчивых случайных процессов – Минск, 2008
10. McCulloch J. H. Simple consistent estimators of stable distribution parameters. Communications in Statistics – 1986 Simulation and Computation. Vol. 15 no. 4 P. 1109-1136.
11. DuMouchel, W. Stable Distributions in Statistical Inference: 1. Symmetric Stable Distributions Compared to Other Symmetric Long-Tailed Distributions – 1973 Journal of the American Statistical Association Vol.68 no.342 P. 469-477
12. PyLevy Module [Электронный ресурс]. — 2020. — Режим доступа: <https://pylevy.readthedocs.io/en/latest/index.html>
13. StableEstim [Электронный ресурс]. — 2016. — Режим доступа: <https://cran.r-project.org/web/packages/StableEstim/index.html>

14. Hurst module [Электронный ресурс]. — 2019. — Режим доступа:
<https://pypi.org/project/hurst/>
15. Koutrouvelis, I.A. Regression-type estimation of the parameters of stable laws. JASA 1980, 75, 918-928.

ПРИЛОЖЕНИЕ А

Таблица 1. Значения функции $\varphi_1(\alpha, \beta)$.

$$v_\alpha = \varphi_1(\alpha, \beta).$$

α	β				
	0.00	0.25	0.50	0.75	1.00
2.00	2.439	2.439	2.439	2.439	2.439
1.90	2.512	2.512	2.513	2.513	2.515
1.80	2.608	2.609	2.610	2.613	2.617
1.70	2.737	2.738	2.739	2.742	2.746
1.60	2.912	2.909	2.904	2.900	2.902
1.50	3.148	3.136	3.112	3.092	3.089
1.40	3.464	3.436	3.378	3.331	3.316
1.30	3.882	3.834	3.720	3.626	3.600
1.20	4.447	4.365	4.171	4.005	3.963
1.10	5.217	5.084	4.778	4.512	4.451
1.00	6.314	6.098	5.624	5.220	5.126
0.90	7.910	7.590	6.861	6.260	6.124
0.80	10.448	9.934	8.779	7.900	7.687
0.70	14.838	13.954	12.042	10.722	10.370
0.60	23.483	21.768	18.332	16.216	15.584
0.50	44.281	40.137	33.002	29.140	27.782

Таблица 2. Значения функции $\varphi_2(\alpha, \beta)$.

$$v_\beta = \varphi_2(\alpha, \beta).$$

α	β				
	0.0	0.25	0.5	0.75	1.0
2.00	0.0	0.0	0.0	0.0	0.0
1.90	0.0	0.018	0.036	0.053	0.071
1.80	0.0	0.039	0.077	0.113	0.148
1.70	0.0	0.063	0.123	0.178	0.228
1.60	0.0	0.089	0.174	0.248	0.309
1.50	0.0	0.118	0.228	0.320	0.390
1.40	0.0	0.148	0.285	0.394	0.469
1.30	0.0	0.177	0.342	0.470	0.546
1.20	0.0	0.206	0.399	0.547	0.621
1.10	0.0	0.236	0.456	0.624	0.693
1.00	0.0	0.268	0.513	0.699	0.762
0.90	0.0	0.303	0.573	0.770	0.825
0.80	0.0	0.341	0.634	0.834	0.881
0.70	0.0	0.387	0.699	0.890	0.927
0.60	0.0	0.441	0.768	0.936	0.962
0.50	0.0	0.510	0.838	0.970	0.985

Таблица 3. Значения функции $\psi_1(v_\alpha, v_\beta)$.

v_α	v_β						
	0.0	0.1	0.2	0.3	0.5	0.7	1.0
2.439	2.0	2.0	2.0	2.0	2.0	2.0	2.0
2.5	1.916	1.924	1.924	1.924	1.924	1.924	1.924
2.6	1.808	1.813	1.829	1.829	1.829	1.829	1.829
2.7	1.729	1.730	1.737	1.745	1.745	1.745	1.745
2.8	1.664	1.663	1.663	1.668	1.676	1.676	1.676
3.0	1.563	1.560	1.553	1.548	1.547	1.547	1.547
3.2	1.484	1.480	1.471	1.460	1.448	1.438	1.438
3.5	1.391	1.386	1.378	1.364	1.337	1.318	1.318
4.0	1.279	1.273	1.266	1.250	1.210	1.184	1.150
5.0	1.128	1.121	1.114	1.101	1.067	1.027	0.973
6.0	1.029	1.021	1.014	1.004	0.974	0.935	0.874
8.0	0.896	0.892	0.887	0.883	0.855	0.823	0.769
10.0	0.818	0.812	0.806	0.801	0.780	0.756	0.691
15.0	0.698	0.695	0.692	0.689	0.676	0.656	0.595
25.0	0.593	0.590	0.588	0.586	0.579	0.563	0.513

Таблица 4. Значения функции $\psi_2(v_\alpha, v_\beta)$.

v_α	v_β						
	0.0	0.1	0.2	0.3	0.5	0.7	1.0
2.439	0.0	2.160	1.0	1.0	1.0	1.0	1.0
2.5	0.0	1.592	3.390	1.0	1.0	1.0	1.0
2.6	0.0	0.759	1.800	1.0	1.0	1.0	1.0
2.7	0.0	0.482	1.048	1.694	1.0	1.0	1.0
2.8	0.0	0.360	0.760	1.232	2.229	1.0	1.0
3.0	0.0	0.253	0.518	0.823	1.575	1.0	1.0
3.2	0.0	0.203	0.410	0.632	1.244	1.906	1.0
3.5	0.0	0.165	0.332	0.499	0.943	1.560	1.0
4.0	0.0	0.136	0.271	0.404	0.689	1.230	2.195
5.0	0.0	0.109	0.216	0.323	0.539	0.827	1.917
6.0	0.0	0.096	0.190	0.284	0.472	0.693	1.759
8.0	0.0	0.082	0.163	0.243	0.412	0.601	1.596
10.0	0.0	0.074	0.147	0.220	0.377	0.546	1.482
15.0	0.0	0.064	0.128	0.191	0.330	0.478	1.362
25.0	0.0	0.056	0.112	0.167	0.285	0.428	1.274

Таблица 5. Значения функции $\varphi_3(\alpha, \beta)$.

$$v_c = \varphi_3(\alpha, \beta).$$

α	β				
	0.0	0.25	0.50	0.75	1.00
2.00	1.908	1.908	1.908	1.908	1.908
1.90	1.914	1.915	1.916	1.918	1.921
1.80	1.921	1.922	1.927	1.936	1.947
1.70	1.927	1.930	1.943	1.961	1.987
1.60	1.933	1.940	1.962	1.997	2.043
1.50	1.939	1.952	1.988	2.045	2.116
1.40	1.946	1.967	2.022	2.106	2.211
1.30	1.955	1.984	2.067	2.188	2.333
1.20	1.965	2.007	2.125	2.294	2.491
1.10	1.980	2.040	2.205	2.435	2.696
1.00	2.000	2.085	2.311	2.624	2.973
0.90	2.040	2.149	2.461	2.886	3.356
0.80	2.098	2.244	2.676	3.265	3.912
0.70	2.189	2.392	3.004	3.844	4.775
0.60	2.337	2.635	3.542	4.808	6.247
0.50	2.588	3.073	4.534	6.636	9.144

Таблица 6. Значения функции $\varphi_4(\alpha, \beta)$.

$$v_d = \varphi_4(\alpha, \beta).$$

α	β				
	0.0	0.25	0.50	0.75	1.00
2.00	0.0	0.0	0.0	0.0	0.0
1.90	0.0	0.023	0.047	0.070	0.094
1.80	0.0	0.051	0.101	0.152	0.202
1.70	0.0	0.084	0.167	0.250	0.331
1.60	0.0	0.126	0.252	0.375	0.495
1.50	0.0	0.184	0.366	0.544	0.717
1.40	0.0	0.269	0.534	0.791	1.041
1.30	0.0	0.407	0.808	1.196	1.573
1.20	0.0	0.679	1.347	1.998	2.631
1.10	0.0	1.483	2.949	4.389	5.806
1.00 ⁺	0.0	∞	∞	∞	∞
1.00	0.0	-0.098	-0.223	-0.383	-0.576
1.00 ⁻	0.0	$-\infty$	$-\infty$	$-\infty$	$-\infty$
0.90	0.0	-1.677	-3.394	-5.159	-6.966
0.80	0.0	-0.865	-1.789	-2.777	-3.820
0.70	0.0	-0.580	-1.243	-1.992	-2.816
0.60	0.0	-0.422	-0.960	-1.613	-2.373
0.50	0.0	-0.311	-0.779	-1.409	-2.198

Таблица 7. Значения функции $\varphi_5(\alpha, \beta)$.

$v_z = \varphi_5(\alpha, \beta)$					
α	β				
	0.0	0.25	0.50	0.75	1.00
2.00	0.0	0.0	0.0	0.0	0.0
1.90	0.0	-0.017	-0.032	-0.049	-0.064
1.80	0.0	-0.030	-0.061	-0.092	-0.123
1.70	0.0	-0.043	-0.088	-0.132	-0.179
1.60	0.0	-0.056	-0.111	-0.170	-0.232
1.50	0.0	-0.066	-0.134	-0.206	-0.283
1.40	0.0	-0.075	-0.154	-0.241	-0.335
1.30	0.0	-0.084	-0.173	-0.276	-0.390
1.20	0.0	-0.090	-0.192	-0.310	-0.447
1.10	0.0	-0.095	-0.208	-0.346	-0.508
1.00	0.0	-0.098	-0.223	-0.383	-0.576
0.90	0.0	-0.099	-0.237	-0.424	-0.652
0.80	0.0	-0.096	-0.250	-0.469	-0.742
0.70	0.0	-0.089	-0.262	-0.520	-0.853
0.60	0.0	-0.078	-0.272	-0.581	-0.997
0.50	0.0	-0.061	-0.279	-0.659	-1.198

ПРИЛОЖЕНИЕ Б

Пример данных собранных на предприятии

Time	IN Traffic	OUT Traffic
Feb-08-21 20:20	3645	3954
Feb-08-21 20:19	5140	4384
Feb-08-21 20:18	11155	12161
Feb-08-21 20:17	192224	351760
Feb-08-21 20:16	5563	5849
Feb-08-21 20:15	7544	5046
Feb-08-21 20:14	5739	11242
Feb-08-21 20:13	4987	4967
Feb-08-21 20:12	4734	22813
Feb-08-21 20:11	25143	3539
Feb-08-21 20:10	4408	3876
Feb-08-21 20:09	4560	7798
Feb-08-21 20:08	3297	3453
Feb-08-21 20:07	5674	4466
Feb-08-21 20:06	8040	12903
Feb-08-21 20:05	17476	29990
Feb-08-21 20:04	3820	5257
Feb-08-21 20:03	22282	20840
Feb-08-21 20:02	5559	13130
Feb-08-21 20:01	4674	17109

ПРИЛОЖЕНИЕ В

Программная реализация алгоритмов

https://github.com/Jeniamakarchik/master_thesis_experiments