

**АННОТАЦИЯ**  
**диссертационной работы Ибраевой Жанар Базарбековны**  
**на тему: «Разработка моделей анализа сетевого трафика и**  
**прогнозирования», представленной на соискание степени доктора**  
**философии (PhD) по специальности 6D070400 – Вычислительная**  
**техника и программное обеспечение**

**Актуальность темы исследования.** В Казахстане с 2007 года функционирует мультисервисная сеть. Эта сеть нового поколения NGN (Next Generation Network), которая основана на протоколе IP с коммутацией пакетов. Смена сетевой инфраструктуры с технологией TDM (Time Division Multiplexing) с коммутацией каналов на пакетную IP (Internet Protocol) создала современную инфраструктуру сферы ИКТ (инфокоммуникационных технологий) с предоставлением услуг – VoIP (Voice over Internet Protocol), IP VPN (Virtual private network), IPTV (Internet protocol television) и другие. Все увеличивающийся объём пропускаемой информации создаёт для опорной сети передачи данных определенную сложность при её обработке. С другой стороны современное общество требует высоких скоростей передачи обработанной информации.

Современная гетерогенная сеть порождает сетевой трафик со сложной (неоднородной) структурой. Исследование реально измеренных данных показывают, что они обладают не равномерной интенсивностью поступления пакетов на обслуживающие сетевые устройства.

Следовательно, происходит увеличение не только объемов трафика, но и изменение ее структуры, поэтому анализ сетевого трафика по-прежнему остается актуальной задачей. Выявить и количественно выразить компоненты сложной структуры – наличие/отсутствие тенденции, периодичности, случайной компоненты является основной задачей анализа временного ряда.

Рост объемов разнородного трафика в инфокоммуникационных сетях актуализирует вопросы обеспечения качества предоставляемых услуг связи, что в свою очередь требует обращения к моделям прогнозирования.

Современные исследования показали, что анализ и прогнозирование сетевого трафика остается наиважнейшей задачей в управлении трафиком.

Прогнозные данные предоставляют необходимые сведения для решения задачи управления информационными потоками в сети и *позволят на основе управления предотвратить потери пакетов.*

**Целью диссертационной работы** является разработка моделей анализа и прогнозирования реально измеренного сетевого трафика.

Для достижения поставленной цели в диссертации решаются нижеследующие **задачи:**

1. Исследование структуры временного ряда эмпирических данных;
2. Проведение экспериментальных исследований с классическими моделями прогнозирования временных рядов;

3. Разработка моделей прогнозирования сетевого трафика с учетом его разнородности;

**Объект исследования.** Объектом исследования является временной ряд, который отображает совокупность переданных по магистральной сети пакетов протокола MPEG за пять часов за каждую секунду (18000 точек).

**Научная новизна диссертационной работы** определяется следующими полученными результатами:

1. На основе анализа структуры реально измеренного сетевого трафика разработана программа, проверяющая ряд на стационарность;
2. Определены параметры модели ARIMA и доказана адекватность модели ARIMA(0,2,1);
3. Разработана модель прогнозирования сетевого трафика на основе ИНС NARX (Nonlinear AutoRegressive Network with exogenous inputs);
4. Программно реализованы модели нечеткой логики.

**Достоверность** полученных результатов диссертационной работы определяется использованием наиболее точных математически ориентированных инструментов моделирования высокого уровня, таких как Matlab и Python, апробированием результатов диссертационных исследований на научных конференциях, семинарах.

**Теоретическая и практическая ценность.** Теоретическая значимость данной работы заключается в выявлении моделей прогнозирования, которые можно использовать в нестационарных условиях эмпирических данных. Применение моделей прогнозирования дает возможность повысить параметры качества обслуживания анализируемого трафика. Практическая значимость работы заключается в разработке плана управления сетевым трафиком и принятии правильных решений при управлении предложенными моделями прогнозирования. Полученные результаты прошли проверку в ТОО «Алматинский Институт Технологий».

**Структура диссертации** включает введение, 4 раздела, заключение, список использованных источников и приложения. Диссертация изложена на 100 страницах компьютерного набора, включая 41 рисунок и 6 таблиц, 51 формула, список литературы из 113 наименований.

**Во введении** дано обоснование актуальности выбранной темы диссертационной работы. Сформулированы цель, объект, задачи научно-исследовательской работы. Описаны результаты проведенных исследований, показаны их научная новизна и практическая значимость. Приведены данные об апробации основных результатов диссертационной работы.

**В первом разделе** диссертационной работы проведен обзор и анализ сетевого трафика мультисервисной сети, основные особенности сети телекоммуникации в Казахстане, обзор основных научных работ по прогнозированию сетевого трафика. Приведена общая характеристика проблемы и постановка задач исследования. Определена классификация стохастических процессов, параметры определения стационарности и нестационарности временного ряда данных, методы прогнозирования.

Временной ряд называется нестационарным, если характеристики ряда: среднее значение, дисперсия, АКФ зависят от времени. Если среднее значение ряда линейно зависит от времени, то это означает наличие во временном ряду линейной тенденции, если дисперсия временного ряда изменяется во времени или является гетероскедастичной (неоднородной), если АКФ циклически изменяется, то это говорит о том, что временной ряд содержит периодические составляющие. Нестационарный ряд всегда имеет тенденцию, которая обусловлена неслучайными факторами в процессах. Не стационарность в широком смысле означает, что, корреляционная функция ряда при фиксированном сдвиге или первый момент ряда или оба они вместе изменяются во времени. Не стационарность в узком смысле означает изменчивость во времени функции распределения.

Нестационарные ряды обладают следующими особенностями:

- в долгосрочном периоде уровни ряда группируются около разных средних значений;
- дисперсия временного ряда меняется от периода к периоду, то есть зависит от времени;
- автокорреляционная функция убывает очень медленно.

Имеются следующие способы по распознаванию стационарности:

1. представление временного ряда в виде графика и визуальная проверка на наличие тренда;
2. исследование на наличие автокорреляции (описывает линейную тесноту и при этом предполагается стационарность процесса);
3. использование тестов присутствия тренда;

Важно отметить, что проверка стационарности является важным шагом при анализе временных рядов, поскольку многие статистические методы и модели требуют, чтобы временной ряд был стационарным. В случае если ряд нестационарен, необходимо провести соответствующие преобразования данных, например, дифференцирование или сезонное корректирование, чтобы получить стационарный ряд.

Для проверки гипотезы о стационарности ряда на практике используются критерии, которые обеспечивают принятие истинной и отклонение ложной гипотезы с высокой вероятностью.

По первому разделу были сделаны следующие выводы:

1. Временные ряды могут быть стационарными и нестационарными. Для проверки гипотезы о стационарности ряда на практике используются критерии параметрические и непараметрические.
2. Статистические свойства стационарных и нестационарных временных рядов существенно отличаются и для их моделирования должны применяться различные методы.
3. Модель стационарных временных рядов характеризуется постоянными во времени средней, дисперсией и автокорреляцией.
4. Модель нестационарных временных рядов многокомпонентная, содержит трендовую, сезонную и случайную составляющие.
5. Нестационарные ряды разделяются на ряды с детерминированным

трендом TS и со стохастическим трендом DS.

6. Приведение к стационарности осуществляются следующими операциями: выделения тренда, сезонной составляющей или интеграцией, характеризующейся порядком последовательных разниц.
7. Критерий Дики-Фуллера осуществляет проверку на DS-ряд или TS.

**Во втором разделе** проведен анализ числовых характеристик измеренного трафика. Исходные данные получены за пять часов, которые отображают совокупность переданных по магистральной сети пакетов протокола MPEG.

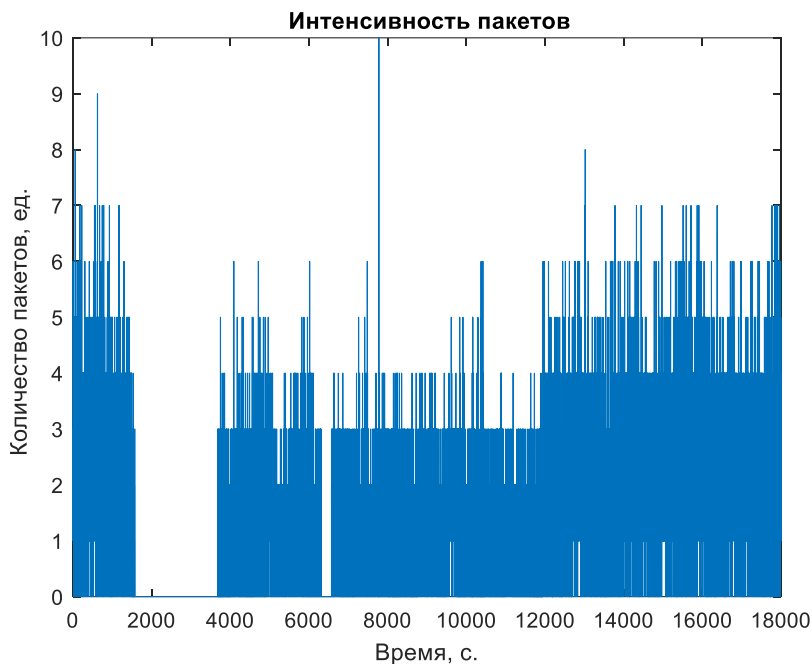


Рис.1 Ряд интенсивности передачи пакетов

Осуществлено статистическое оценивание ряда на нормальность, так как для прогнозирования сетевого трафика это играет важную роль при использовании соответствующих методов. Проверена гипотеза о нормальности распределения исследуемого ряда и отменены следующие критерии: модифицированные критерии Колмогорова и Смирнова; Крамера-Мизеса; Андерсона-Дарлинга; Шапиро-Франсиа; коэффициента асимметрии; эксцесса; Жарка-Бера; Гири и Д`Агостино.

С помощью непараметрических критериев выявлено, что исследуемый ряд имеет тренд. По итогам исследования разработана программа в программной среде Python, проверяющий ряд на стационарность.

В программной среде Matlab исследован исходный ряд на единичный корень тестами Филлипса-Перрона (PP-test), а также тестом тестом Дики-Фуллера.

По второму разделу были сделаны следующие выводы:

1. Характер распределения исследуемого ряда не соответствует условию нормальности распределения.
2. С помощью созданной программы с использованием непараметрических критериев и коррелограммы выявлено, что исследуемый ряд имеет тренд.
3. Исследование методами ADF-test, PP-test и KPSS-test показало наличие единичных корней.

**Третий раздел** посвящен частотно-временному анализу временных рядов. Для исследуемого ряда применена методика анализа сингулярного спектра SSA (Singular Spectrum Analysis). Метод является одним из современных инструментов анализа структурных компонент временного ряда. Метод декомпозиции временных рядов SSA разбивает временной ряд на набор суммируемых компонентов, которые сгруппированы и интерпретируются как тренд, периодичность и шум, и этот метод подчеркивает разделимость базовых компонентов и может легко разделять периодичности, возникающие в разных временных масштабах, даже в очень шумных данных временных рядов.

Далее ряд исследован на гармонические составляющие, он разложился на низкочастотную составляющую (тренд) и на периодические составляющие вместе с высокочастотным шумом. Декомпозиция временного ряда показывает, что временной ряд является не стационарным и он содержит тренд, гармонические составляющие и шум.

**В четвертом разделе** приведено описание особенностей прогноза нестационарных временных рядов.

Среди статистических подходов прогнозирования метод ARIMA (Auto-Regressive Integrated Moving Average) позволяет описывать нестационарные временные ряды. Если в данных есть признаки единичного корня, то можно рассмотреть возможность использования модели ARIMA. Тесты, использованные во второй главе, доказали наличие единичного корня.

Первым шагом при построении моделей является разность данных до тех пор, пока они не станут выглядеть стационарными.

В результате подбора многочисленных параметров модели и визуального анализа получена интегрированная модель ARIMA(0,2,1).

Авто регрессионная интегрированная модель скользящего среднего временного ряда со следующим уравнением:

$$(1 - L)^2 y_t = c + (1 + \Theta_1 L) \varepsilon_t \quad (1)$$

Уравнение описывает ARIMA-модель в форме разностного уравнения, где  $y_t$  — это текущее значение временного ряда,  $L$  - оператор лага (т.е. сдвиг на один период времени),  $c$  - константа,  $\varepsilon_t$  - случайная ошибка в момент времени  $t$ , а  $\Theta_1$  - параметр скользящего среднего первого порядка,  $(1 - L)^2$  означает оператор разности второго порядка, который применяется к значению временного ряда  $y_t$ .

Данное уравнение описывает зависимость текущего значения временного ряда от его прошлых значений, случайных ошибок и параметров модели. Она может быть использована для прогнозирования будущих значений временного ряда на основе его прошлых значений и параметров модели.

Из методов на основе ИИ (Искусственный интеллект) в настоящей работе выбрана модель NARX (Nonlinear AutoRegressive exogenous Network), которая хорошо подходит для моделирования нелинейных систем и алгоритмы нечеткой логики (Fuzzy Logic);

Для осуществления прогнозирования нестационарного временного ряда выбран пакет прикладных программ (ППП) фирмы Math Works Neural Network Toolbox (NNT), который функционирует под управлением ядра системы Matlab.

NARX (Nonlinear Autoregressive with eXogenous inputs) — это рекуррентная нейронная сеть, которая может использоваться для прогнозирования сетевого трафика. Эта сеть относится к классу многослойных перцептронов с обратной связью, которые имеют входные, скрытые и выходные слои.

Для прогнозирования сетевого трафика NARX использует исторические данные о трафике и других связанных переменных, таких как время суток, день недели, праздники и т.д. Эти данные предварительно обрабатываются, нормализуются и затем используются для обучения сети.

Обучение NARX включает в себя настройку весов между узлами входного, скрытого и выходного слоев. В процессе обучения сеть постепенно оптимизирует эти веса, чтобы минимизировать ошибку прогнозирования.

После завершения обучения сеть может быть использована для прогнозирования сетевого трафика на основе новых данных о времени и других связанных переменных.

В модели NARX с внешними входами исходный временной ряд поступая в ИНС, обрабатывается в соответствии с весовыми коэффициентами, и далее снова выходные данные попадают на вход сети, тем самым заменяя собой процедуру обратного распространения ошибки. Это дает возможность учитывать полученные после первичного обучения весовых коэффициентов при повторном обучении сети и это в свою очередь повышает точность ее работы.

В задачах цифровой обработки сигналов широкое распространение получили динамические нейроны, описываемые разностными или дифференциальными уравнениями, поведение которых определяется их предысторией. Одним из простейших динамических нейронов является нейрон Хопфилда. В сети NARX используется модифицированный нейрон Хопфилда, состояние которой определяется более отдаленной предысторией. Динамика модели NARX описывается следующим образом

$$y(n+1) = F(y(n), \dots, y(n-q+1), u(n), \dots, u(n-q+1)) \quad (2)$$

где  $F$  – некоторая нелинейная функция своих аргументов, которая аппроксимируется во время обучения;

$q$  – задержка.

Сеть NARX представляет собой многослойную сеть с прямой передачей сигнала и обратной связью по выходу, выходные данные которой пропущены через вектор задержек во времени. В качестве функции активации используется сигмовидная функция.

Каждый слой сети преобразует входное пространство признаков в некоторое другое пространство, возможно, с иной размерностью. Такое нелинейное преобразование происходит до тех пор, пока классы не оказываются линейно разделимыми нейронами выходного слоя. Все слои ИНС, кроме входного и выходного; придают сети способность моделировать нелинейные явления.

Для обучения нейронной сети использовался алгоритм Левенберга-Маквардта и для оценки производительности нейронной сети используется среднеквадратичная ошибка (MSE).

Возможности нейронных сетей могут усиливаться за счет технологии обработки информации, основанные на нечетких множествах и нечетком выводе. В данной работе исследованы нейронечеткие алгоритмы прогнозирования Чена и Ченга, алгоритмы нечеткой кластеризации.

Метод Чена показал низкие показатели точности прогнозных данных, к тому же числовое значение MSE большое, 9,078161087779954, а метод нечеткого прогнозирования Ченга имеет более высокие показатели, о чем свидетельствует более низкое числовое значение показателя  $MSE=1,2359176594533703$ . Этот результат указывает на то, что метод нечетких временных рядов Ченга достаточно хорош для использования при прогнозировании временных рядов. Различия в этих методах заключаются в этапах формирования нечетких множеств, и в каждой группе нечетких отношений есть веса.

**В заключении** изложены основные результаты и выводы диссертации.

**Апробация работы.** Основные результаты по теме диссертации изложены в 15 работах, 7 статей в журналах, рекомендованных Комитетом по обеспечению качества в сфере образования и науки МН и ВО РК, 1 статья в международном научном издании, входящим в базу данных Scopus и 7 статей в материалах международных конференций.

**Публикации.** Полученные по теме диссертации результаты представлены в следующих публикациях:

1. Serikov T., Zhetpisbayeva A., Mirzakulova S., Zhetpisbayev K., **Ibraeva Zh.**, Tolegenova A., Soboleva L., Zhumazhanov B. (2021). Application of the NARX neural network for predicting a one-dimensional time series. Eastern-European Journal of Enterprise Technologies. Vol. 5 №4 (113) Pp.12-19. Scopus: 56%, WoS: Q2
2. Г.У.Бектемысова, **Ж.Б.Ибраева**, А.Е.Кулакаева, Б.А.Кожаметова. Анализ измеренного сетевого трафика на стационарность. Вестник

- КазАТК, Вестник Казахской Академии Транспорта и Коммуникации им. М.Тынышпаева, ISSN 2790-5802, -№ 3 (122), -2022г. -С.: 302-308
3. G. Bektemyssova, Abdul R., Sh.Mirzakulova, **Zh.Ibraeva**. Time series forecasting by the Arima method. Scientific Journal of Astana IT University, ISSN (P): 2707-9031 ISSN (E): 2707-904X, Volume 11, September 2022, - P.: 14-23
  4. Г.У.Бектемысова, **Ж.Б.Ибраева**. Исследование временного ряда на стационарность. Вестник НИА РК, Вестник Национальной инженерной академии РК, №4(86), 2022, -С.: 20-27
  5. G.U. Bektemyssova, **Zh.B. Ibraeva**, Abd Rakhim Akhmad. Fuzzy model for time series forecasting. Scientific Journal of Astana IT University. ISSN (P): 2707-9031 ISSN (E): 2707-904X VOLUME 13, MARCH 2023, DOI: 10.37943/13EOTU7482, -P.: 93-102
  6. Бектемысова Г.У., **Ибраева Ж.Б.** Возможности применения искусственного интеллекта в строительстве. Вестник КазГАСА. – 2018. – Том 68, выпуск 2. – с. 205-212.
  7. Бектемысова Г.У., **Ибраева Ж.Б.**, Луганская С.П., Миркасилова Т.Ш. инструментов MATLAB для анализа больших данных по энергоэффективности зданий. Вестник КБТУ. – 2019. – Том 16, выпуск 3. – с. 324-328
  8. Аймагамбетова З.Т., **Ибраева Ж.Б.** Аспекты формирования комфортной городской среды, Вестник КазГАСА. – 2021. – Том 81, выпуск 3. с. 15-20
  9. Aimagambetova Z.T, Bektemyssova G.U, **Ibraeva Zh.B.** Buildings energy consumption modeling methods. STCCE-2020 IOP Conf. Series: Materials Science and Engineering (Scopus). 2020. Volume 890, Kazan, Russian Federation doi:10.1088/1757-899X/890/1/012144, p.14-23
  10. **Ибраева Ж.Б.**, Айтжанов Д., Каттабек А. Анализ и оптимизация сетевого трафика. Международный журнал информационных и коммуникационных технологий, Спец выпуск, май, 2022, С.:166-170
  11. **Ибраева Ж.Б.**, Мирзакулова Ш.А. Analysis of a one-dimensional time series for a trend. Материалы международной научной конференции молодых ученых, IMA-2022, Суми-Нур-Султан, апрель, 2022, P.:258-259
  12. **Ибраева Ж.Б.**, Мирзакулова Ш.А. Network traffic analysis using Leybourne-McCabe test. Материалы международной научной конференции молодых ученых, IMA-2022, Суми-Нур-Султан, апрель, 2022, P.:259-261
  13. **Ибраева Ж.Б.**, Миркасилова Т.Ш. Мәліметтерді басқарудың ертеңі мен бүгіні. Сборник материалов международной научно-методической конференции. Современные концепции науки и образования. Алматы, 2017. стр. 56-59



14. Sh.Mirzakulova, **Zh.Ibraeva**. Clustering Time Series Data. Материалы международной научной конференции молодых ученых, IMA-2022, Суми-Нур-Султан, апрель, 2023, Р.:398-399
15. Доскен Б., **Ибраева Ж.Б.** Компьютерное моделирование усилителя сигналов. Материалы международной научной конференции молодых ученых, IMA-2022, Суми-Нур-Султан, апрель, 2023, Р.:400-401

***Авторские свидетельства:***

1. Разработка модели прогнозирования с использованием статистического метода Auto-Regressive Integrated Moving Average. № 32481 от «9» февраля 2023 года.
2. Нечеткая модель прогнозирования временного ряда. № 35224 от «27» апреля 2023 года