

УДК 004.9

DOI 10.52167/1609-1817-2022-122-3-302-308

Г.У.Бектемысова[✉], **Ж.Б. Ибраева, А.Е.Кулакаева, Б.А.Кожаметова**
Международный университет информационных технологий, Алматы, Казахстан
E-mail: g.bektemisova@gmail.com

АНАЛИЗ ИЗМЕРЕННОГО СЕТЕВОГО ТРАФИКА НА СТАЦИОНАРНОСТЬ

Аннотация. С развитием сети Интернета вещей (Internet of Things, IoT) объем данных, в том числе объем сетевого трафика увеличивается все больше и больше. Число подключаемых устройств к сети IoT увеличивается с каждым днем. Следовательно, актуальность задач управления сетью, в том числе на основе прогнозируемых будущих данных очевидна. Нахождение и оценка компонентов структуры ряда, таких как наличие или отсутствие тренда, периодичности, случайной составляющей, которые необходимы для принятия правильного решения, является основной задачей анализа временных рядов. Анализ сетевого трафика - важный шаг в развитии успешных схем управления перегрузкой и определения нормальных и вредоносных пакетов.

В данной статье рассматривается временной ряд с реальными данными, снятые на магистрали в городе Алматы. Для анализа временного ряда на стационарность применены методы исследования на наличие автокорреляции, а также визуальная оценка графиков одномерного ряда и ковариационной матрицы, которая разбивает одномерный ряд на многомерный. Использован тест Льюнга-Бокса на наличие автокорреляции. Результаты исследования получены в среде численно-математического моделирования Matlab и все они указывают на не стационарность ряда.

Ключевые слова. Временной ряд, сетевой трафик, стационарный процесс, нестационарный процесс, автокорреляция, прогнозирование.

Введение.

На сегодняшний день наблюдается активное внедрение технологии IoT (Internet of Things, IoT), находящий свое применение практически во всех сферах человеческой деятельности. Трафик, передаваемый устройствами в сети IoT, обладает разнородным свойством, ввиду различного применения технологии. При этом модели, построенные на основе данных, характеризующего один объект за ряд последовательных моментов времени (временные ряды) обладают свойством не стационарности, что означает, что их структура многокомпонентная.

Временные ряды можно исследовать с разной целью: чтобы выявить характерные особенности ряда, выяснить механизм, лежащий в основе процесса, либо очистить временной ряд от выбросов или предсказать будущее на основании существующих данных [1].

Нахождение и оценка компонентов структуры ряда, таких как наличие или отсутствие тренда, периодичности, случайной составляющей, которые необходимы для принятия правильного решения, является основной задачей анализа временных рядов [2].

При анализе временного ряда применяются различные математико-статистические методы, предназначенные для выявления структуры временных рядов. В свою очередь исследование структуры временного ряда требуется для того, чтобы построить математическую модель того явления, которое является источником анализируемого ряда. Прогнозирование временного ряда помогает для принятия эффективных решений задач [3,4].

Материалы и методы.

В данной статье рассматривается временной ряд с реальными данными, которые были сняты на магистрали в городе Алматы.

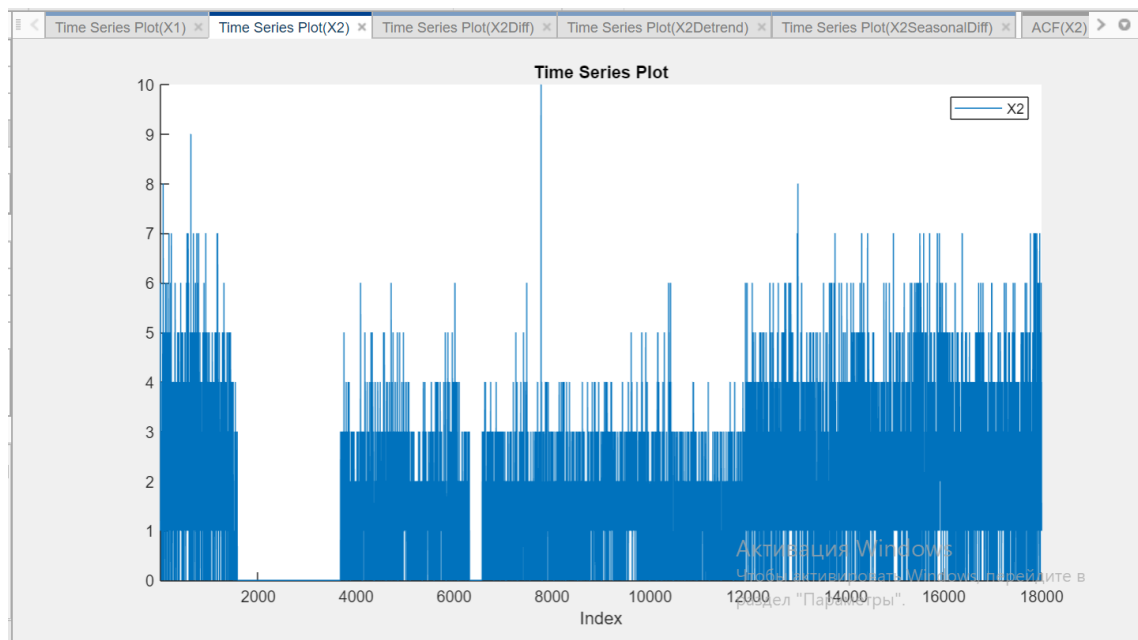


Рисунок 1 - Исследуемый ряд интенсивности передачи пакетов

Измеренный ряд показывает совокупность переданных по магистральной сети пакетов за каждую секунду. Количество точек – 18000. График измеренных данных показан на рисунке 1, по вертикали отображены количество пакетов данных, поступивших за 5 часов, по горизонтали - время (в секундах).

Полученный график (рис.1), отображающий интенсивность передачи данных показывает характерные особенности ряда, позволяет допустить наличие тренда или периодичности, либо случайной составляющей.

К характеристикам случайного процесса во временном ряду относятся:

- математическое ожидание;
- дисперсия;
- ковариационная функция.

Математическим ожиданием случайного процесса называется траектория математических ожиданий, составляющих этот процесс случайных величин:

$$m(t) = MX(t). \quad (1)$$

Математическое ожидание $m(t)$ случайного процесса $X(t)$ представляет собой некоторую неслучайную «среднюю функцию», около которой варьируются реализации случайного процесса [5].

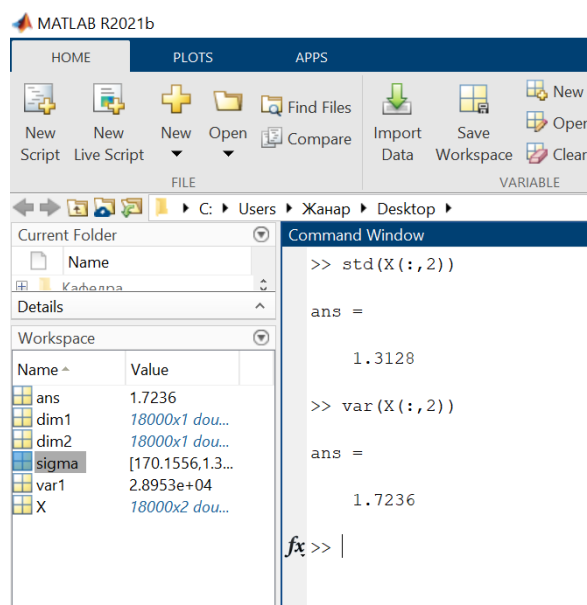


Рисунок 2 - Значения среднего квадратичного отклонения и дисперсии

Дисперсией случайного процесса называется траектория дисперсий, составляющих этот процесс случайных величин:

$$\sigma^2(t) = DX(t) \quad (2)$$

Дисперсия случайного процесса $\sigma^2(t)$ представляет собой неслучайную неотрицательную функцию, характеризующую степень разброса реализаций случайного процесса $X(t)$ около его математического ожидания $m(t)$. Ее значение должно быть положительной, как мы и получили (рисунок 2), она равна = 1.7236.

Дисперсия и стандартное отклонение обычно используются для описания одномерных данных, для многомерных данных в статистике используется ковариация [6-8].

Ковариационной функцией (рисунок 3) случайного процесса называется функция двух переменных, значения которой представляют собой коэффициенты ковариации сечений процесса в соответствующие моменты времени:

$$B(t, s) = cov[X(t), X(s)] = M\{[X(t) - m(t)] \cdot [X(s) - m(s)]\} \quad (3)$$

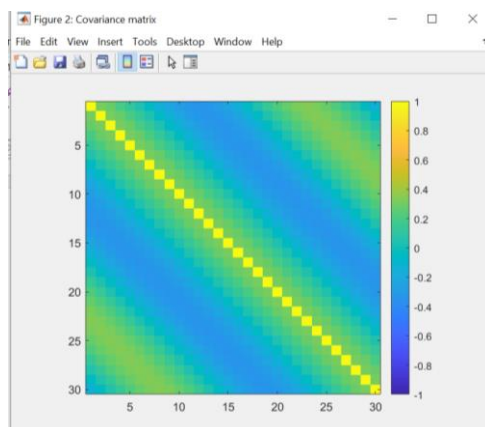


Рисунок 3 - Ковариационная матрица, полученная в Matlab

Результаты.

Анализируя график ковариационной матрицы, можно увидеть разброс уровней, следовательно исследуемый ряд сложный, в ряду допускается наличие тренда.

Временной ряд – ряд одной и той же переменной, полученных в результате измерений. Временной ряд представляет собой: $y_1, \dots, y_T, \dots, y_R$ значения признака (уровень ряда, количество пакетов), измеренные через постоянные временные интервалы. При этом его можно также назвать интервальным, так как она сформирована накоплением переменной в течение некоторого периода времени.

При моделировании временной ряд рассматривают как случайный процесс (стохастический), как статистическое явление, развивающееся во времени согласно законам теории вероятностей.

Стохастические процессы классифицируются:

- стационарные;
- нестационарные.

Стационарные и нестационарные временные ряды отличаются наличием/отсутствием факторов, формирующих уровни ряда [9].

При исследовании временных рядов важно знать стационарный или нестационарный ряд, так как они обладают разными статистическими характеристиками и соответственно оцениваются разными способами.

Свойства стационарных временных рядов позволяют сделать предположение о характере временных рядов на основе графика временного ряда (рис.1) и автокорреляционной функции (АКФ).

Стационарность временного ряда означает отсутствие:

- тренда;
- сезонной составляющей;
- интервенции;
- гетероскедастичности;
- любых систематически изменяющихся взаимозависимостей между составляющими временного ряда.

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:

- представление временного ряда в виде графика и визуальная проверка на наличие тренда;
- исследование на наличие автокорреляции (описывает линейную тесноту и при этом предполагается стационарность процесса);
- использование тестов присутствия детерминистского тренда;
- использование тестов наличия стохастического тренда (например, тесты на единичный корень).

Функция выборочной автокорреляции (ACF) и функция частичной автокорреляции (PACF) являются полезными качественными инструментами для оценки наличия автокорреляции при отдельных задержках. Q-тест Льюнга-Бокса — это более количественный способ проверки автокорреляции при множественных задержках.

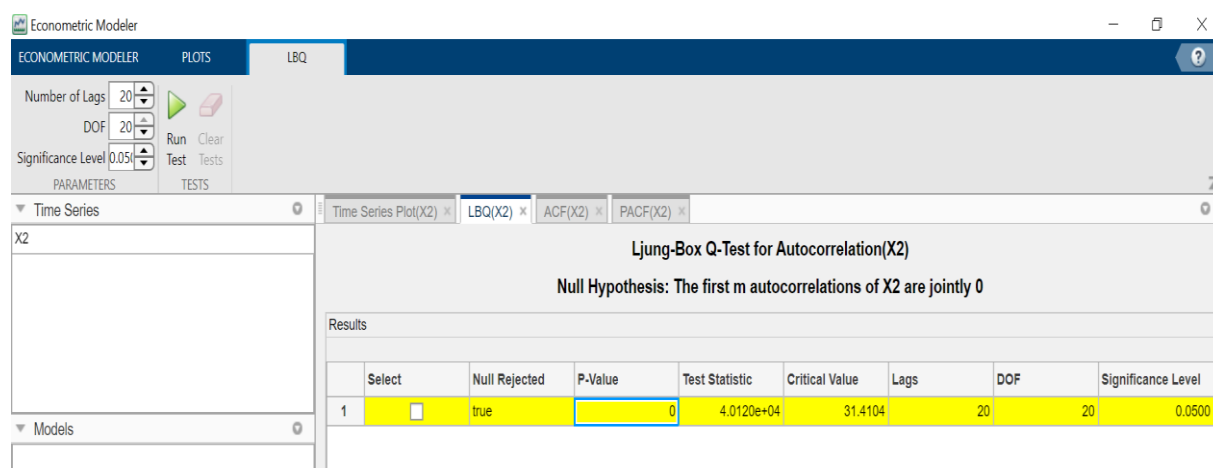


Рисунок 4 - Результаты теста Льюнга-Бокса

Нулевая гипотеза для этого теста состоит в том, что первые m автокорреляций вместе равны нулю (рисунок 4).

$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$.

Q-критерий Льюнга-Бокса можно использовать также для оценки автокорреляции в любом ряду с постоянным средним значением. Сюда входят остаточные ряды, которые можно проверить на автокорреляцию во время диагностических проверок модели.

Обсуждение.

По результатам исследования временного ряда получено представление ряда в виде графика, он позволяет допустить наличие тренда или периодичности, либо случайной составляющей. Также график ковариационной матрицы показывает разброс уровней, что допускает наличие тренда в ряду. Исследование на наличие автокорреляции показало, что первые m автокорреляций вместе равны нулю.

Получены такие наблюдения по тесту Льюнга-Бокса: Аргумент *Lags* влияет на мощность теста. Если L слишком мало, то тест не обнаруживает автокорреляции высокого порядка. Если L слишком велико, то тест теряет силу, когда значительная корреляция при одном лаге размывается незначительными корреляциями при других лагах.

Заклучение.

Первым шагом при анализе временного ряда является выявление закономерностей, лежащих в основе данных с течением времени. В данной статье ряд исследован с помощью определения характеристик случайного процесса с использованием пакета Matlab. Анализ и оценка данных временного ряда на стационарность проведены как графически, так и с использованием теста Льюнг-Бокс.

Целесообразно продолжить исследование одномерного временного ряда с помощью тестов присутствия детерминистского тренда и тестов наличия стохастического тренда (например, тесты на единичный корень).

ЛИТЕРАТУРА

[1] О.А. Мишулина. Статистический анализ и обработка временных рядов. - М.: «МИФИ», 2004. -180 с.

[2] Serikov T., Zhetpisbayeva A., Mirzakulova S., Zhetpisbayev K., Ibraeva Z., Soboleva L., Tolegenova A., & Zhumazhanov B. Application of the NARX neural network for predicting

a one-dimensional time series //Eastern-European Journal of Enterprise Technologies. – 2021. – Vol. 5. – №4 (113). - P. 12–19. DOI: 10.15587/1729-4061.2021.242442

[3] Власюк А. А., Орлов Ю. Н. Точность идентификации выборочных распределений временных рядов в зависимости от типа распределения, нормы и длины выборки // Препринты ИПМ им. М.В. Келдыша. - 2015. - № 17. - С.25

[4] Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. - М.: «Статистика», 1973. –106 с.

[5] В.П. Дьяконов, В.В. Круглов. Matlab 6.5 SP1/7/7 SP1/7 SP2 Simulink 5/6. Инструменты искусственного интеллекта и биоинформатики. Серия «Библиотека Профессионала». – М.: СОЛОН-ПРЕСС, 2020. – 454 с.: ил

[6] Кобзарь А.И. Прикладная математическая статистика. - М.: Физматлит, 2006. – 816 с.

[7] Nielsen A. Practical time series analysis: Prediction with statistics and machine learning. – USA.: «O'Reilly Media», 2019. – 504 p.

[8] Beard E., Marsden J., Brown J., Tombor I., Stapleton J., Michie S., & West, R. Understanding and using time series analyses in addiction research // Society for the Study of Addiction. – 2019. – №114 (10). – P.1866-1884 DOI:10.1111/add.14643

[9] F. Xu et al., Big Data Driven Mobile Traffic Understanding and Forecasting: A Time Series Approach, // IEEE Transactions on Services Computing. – 2016. - Vol. 9. - № 5. - P. 796-805, DOI: 10.1109/TSC.2016.2599878.

Гульнара Бектемысова, т.ғ.к., қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан, g.bektemisova@gmail.com

Жанар Ибраева, магистр, аға оқытушы, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан, zh.ibraeva@iitu.edu.kz, zhanarbazar@gmail.com

Айгуль Кулакаева, магистр, аға оқытушы, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан, aigul_k.pochta@mail.ru

Бағдат Кожаметова, магистр, аға оқытушы, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан, bagdatka_0794@mail.ru

ӨЛШЕНГЕН ЖЕЛІЛІК ТРАФИКТЫ СТАЦИОНАРЛЫҚҚА ТАЛДАУ

Аңдатпа. Заттар Интернеті (Internet of Things, IoT) желісінің дамуымен деректер көлемі, соның ішінде желілік трафик көлемі артып келеді. IoT желісіне қосылатын құрылғылардың саны күн сайын артып келеді. Сондықтан желіні басқару міндеттерінің өзектілігі, оның ішінде болашақ болжамды мәліметтер негізінде айқын көрінеді. Трендтің болуы немесе болмауы, периодтылығы, кездейсоқ құраушылар сияқты қатар құрылымының компоненттерін табу және бағалау дұрыс шешім қабылдау үшін қажет және уақыттық қатарларды талдаудың негізгі міндеті болып табылады. Желілік трафикті талдау – жүктемені басқарудың сәтті схемаларын дамытудағы және қалыпты және зиянды пакеттерді анықтаудағы маңызды қадам.

Бұл мақалада Алматы қаласындағы магистральдан алынған нақты деректері бар уақыттық қатарлар қарастырылады. Стационарлық уақыттық қатарларды талдау үшін автокорреляцияның бар-жоғын зерттеу әдістері, сондай-ақ бір өлшемді қатардың және бір өлшемді қатарды көп өлшемдіге бөлетін коварианттық матрицаның графиктерін визуалды бағалау қолданылды. Автокорреляцияға арналған Лjung-Бокс сынағы қолданылды. Зерттеу нәтижелері Matlab сандық модельдеу ортасында алынды және олардың барлығы қатардың стационарлы еместігін көрсетеді.

Түйінді сөздер. уақыттық қатар, желілік трафик, стационарлы процесс, стационарлық емес процесс, автокорреляция, болжау.

Bektemyssova Gulnara, c.t.s, associate professor, International University of Information Technology, Almaty, Kazakhstan, g.bektemisova@gmail.com

Ibraeva Zhanar, master's degree, senior lecturer, International Information Technology University, Almaty, Kazakhstan, zh.ibraeva@iitu.edu.kz, zhanarbazar@gmail.com

Kulakayeva Aigul, master's degree, senior lecturer, International Information Technology University, Almaty, Kazakhstan, aigul_k.pochta@mail.ru

Kozhakhmetova Bagdat, master's degree, senior lecturer, International Information Technology University, Almaty, Kazakhstan, bagdatka_0794@mail.ru

ANALYSIS OF THE MEASURED NETWORK TRAFFIC ON STATIONARY

Abstract. With the development of the Internet of Things (IoT), the volume of data, including the volume of network traffic, is increasing more and more. The number of connected devices to the IoT network is increasing every day. Therefore, the relevance of network management tasks, including tasks based on projected future data, is obvious. Finding and evaluating the components of the series structure, such as the presence or absence of a trend, periodicity, random component, which are necessary for making the right decision, is the main task of time series analysis. Network traffic analysis is an important step in the development of successful congestion control schemes and the identification of normal and malicious packets.

This article discusses a time series with real data taken on the backbone in Almaty. Research methods for the presence of autocorrelation, as well as a visual assessment of the graphs of a one-dimensional series and a covariance matrix that splits a one-dimensional series into a multivariate one, are used to analyze the time series for stationarity. The Ljung-Box test was used for the presence of autocorrelation. The results of the study were obtained in the Matlab numerical simulation environment, and all of them indicate the non-stationarity of the series.

Keywords. Time series, network traffic, stationary process, non-stationary process, autocorrelation, forecasting.
