

Г. У. БЕКТЕМЫСОВА*, Ж. Б. ИБРАЕВА

*Международный университет информационных технологий,
г. Алматы, Казахстан*

ИССЛЕДОВАНИЕ ВРЕМЕННОГО РЯДА НА СТАЦИОНАРНОСТЬ

Реальный сетевой трафик интенсивности пакетов MPEG имеет неравномерную скорость поступления пакетов на обслуживающие сетевые устройства. С развитием концепции Интернета вещей (Internet of Things, IoT) актуальность задач управления сетью растёт с каждым днём. Исследование позволяет получить содержательные ответы на интересующие исследователя вопросы. Применение различных тестов (критериев) при исследовании ряда является актуальной потому, что они позволяют выявить их структуру.

С увеличением разнообразия сетевых приложений, новых протоколов передачи данных в поведении трафика стали проявляться свойства и особенности, которые связаны с его нестационарностью. В этой статье рассматривается временной ряд с реальными данными, снятые на магистрали города Алматы.

Для исследования временного ряда на стационарность применены критерии Колмогорова – Смирнова, Шапиро-Уилка, Дарбина, Дэвида-Хартли-Пирсона, Андерсона-Дарлингга, данные QQ plot, Variance Ratio test for Random Walk в среде численно-математического моделирования Matlab. Получены оценки исследования ряда по распределениям Пирсона и Пуассона. Также определены числовые характеристики формы распределений: асимметрия и эксцесс (куртозис).

В статье осуществлены упрощенные проверки измеренного временного ряда.

Ключевые слова: *временной ряд, сетевой трафик, анализ данных, критерии проверки стационарности, закон нормального распределения.*

Введение. Функционирующая гетерогенная сеть в г. Алматы сочетает в себе как проводные, так и беспроводные технологии, а на протяжении последних лет продолжает свое развитие в соответствии с концепцией IoT. Различные виды коммуникационных услуг и различные конфигурации сетей порождают существенно различающиеся виды трафика.

Все больше поведение сетевого трафика описывается теорией детерминированного хаоса. Следовательно, реальный трафик современной сети имеет сложную (неоднородную) структуру, имеет неравномерную интенсивность поступления пакетов на обслуживающие сетевые устройства. Многочисленные исследования реально измеренных данных подтверждают, что они являются нестационарными, и их структура многокомпонентная.

Сегодня существует более тысячи статистических тестов или критериев, которые применяются для определения отношения исследуемого процесса к тому или иному классу.

Долгое время считалось, что поведение пакетов в сети адекватно описывается экспоненциальными распределениями, например, Пуассоновским (интервал времени между пакетами, длина пачки пакетов и др.) [1]. Такое допущение является верным для сетей небольшого размера и позволяло использовать классические методы теории

* E-mail корреспондирующего автора: g.bektemisova@gmail.com

массового обслуживания. При этом расчет описывал среднюю задержку, среднюю длину очередей и другие параметры. Но с ростом размера сетей и появлением новых протоколов передачи данных появились характеристики в трафике, указывающие на нестационарность. В связи с этим, вопросы анализа и моделирования нестационарных временных рядов является актуальной задачей исследования во всех сферах жизнедеятельности.

Методы и материалы. В данной работе проведен анализ измеренного одномерного ряда (интенсивность пакетов протокола MPEG). 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, поступивших за 5 часов, по горизонтали – время (в секундах).

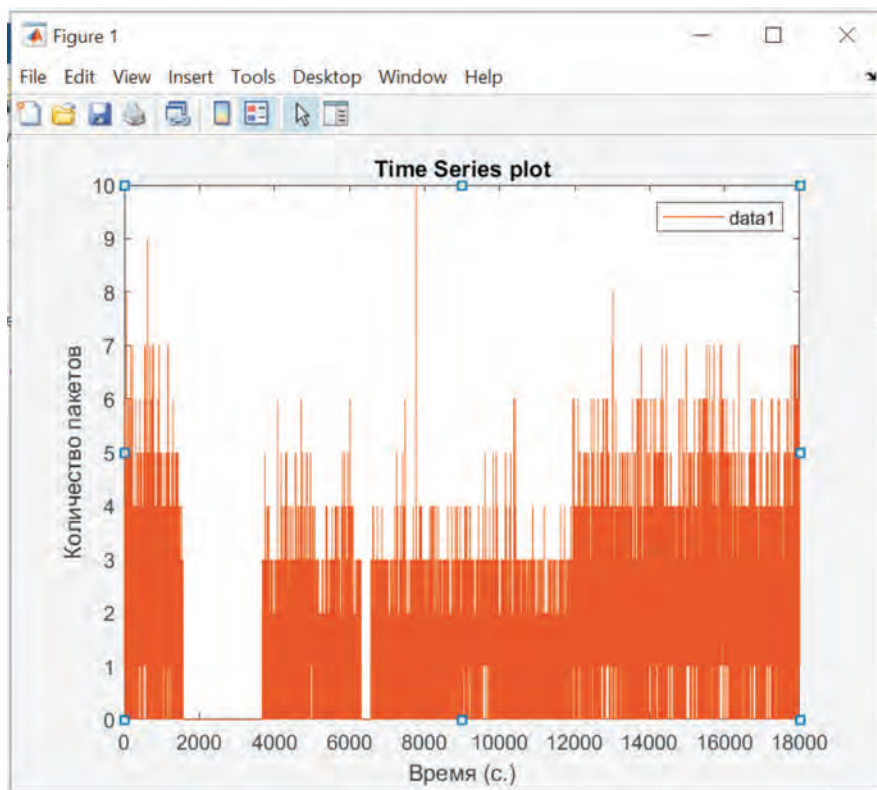


Рисунок 1 – Временной ряд

Визуальная оценка графика полученных данных временного ряда показывает, что он имеет неравномерную интенсивность, разброс уровня данных и др.

Осуществим проверку исходного временного ряда на соответствие нормальному закону распределения с помощью QQ plot теста. Смысл этой проверки сводится к тому, чтобы сравнить идеальное нормальное распределение с фактическим распределением исследуемого ряда (рисунок 2).

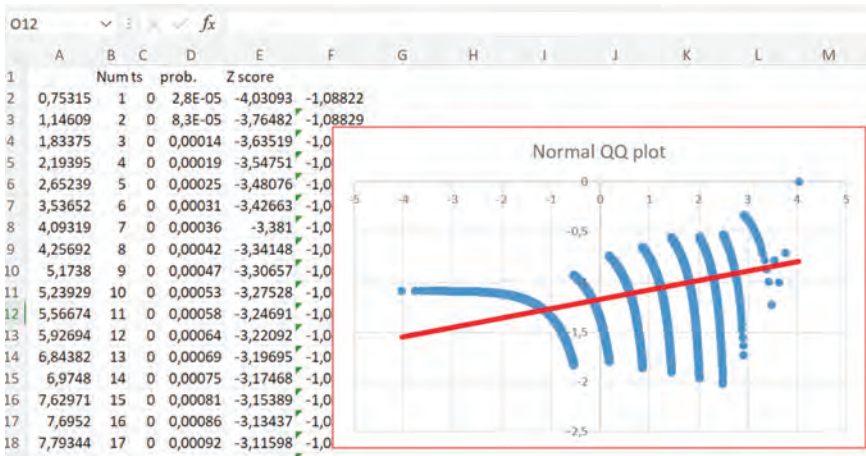


Рисунок 2 – Проверка нормальности распределения данных с помощью QQ plot-теста

Из рисунка 2 можно увидеть, что распределение исследуемого ряда не похоже на нормальный закон распределения, так как форма гауссовского распределения не похожа на кривую колокола.

Рассмотрим распределение исследуемого ряда по закону Пуассона. Распределение Пуассона с параметром λ определяется следующим образом:

$$P_m = \frac{\lambda^m}{m!} e^{-\lambda}, \text{ где } \lambda = np$$

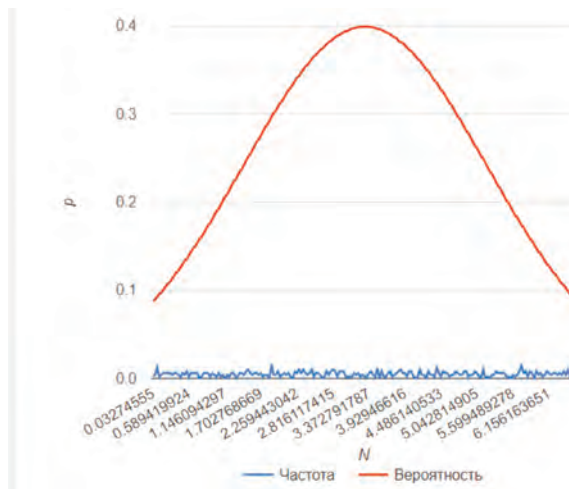


Рисунок 3 – Полигон эмпирических частот и вероятность для распределения Пуассона

Результаты выходных данных на рисунках 3,4 свидетельствуют о том, что исходный ряд интенсивности пакетов протокола MPEG не распределен по закону Пуассона и не относится к генеральной совокупности нормального закона распределения [2-5].

Выводы:

Каждое значение ряда отличается от среднего значения 3.313 в среднем на 1.886.

Гипотеза о том, что случайная величина X подчинена нормальному закону распределения, отвергается (по критерию согласия Пирсона).

Наблюдаемое значение статистики Пирсона попадает в критическую область: $K_{набл} > K_{кр}$, поэтому есть основания отвергать основную гипотезу. Данные выборки распределены **не по закону Пуассона**.

Рисунок 4 – Выходные данные программы

Целесообразно проверить исходный ряд на нормальность с помощью критериев, таких как критерий Колмогорова – Смирнова, Шапиро-Уилка, Дарбина, Дэвида-Хартли-Пирсона, Андерсона-Дарлингга и др. На рисунке 5 показаны результаты исследуемого ряда по проверке на нормальность распределения, используя ряд вышеуказанных критериев.

	A	B	C	D	E	F	G	H	I
1	0,032746	1		Проверка нормальности распределения					
2	0,065491	2		Выдача обычно включает:					
3	0,098237	6		Статистика, Р-значение двустороннее, вывод					
4	0,130982	1		Выбранное пороговое значение					
5	0,163728	2		0,05					
6	0,196473	3		Численность выборки					
7	0,229219	3		18000					
8	0,261964	3		Модифицированный критерий Колмогорова					
9	0,29471	3		0,200817	0	Гипотеза о нормальности отклоняется			
10	0,327456	2		Модифицированный критерий Смирнова					
11	0,360201	3		0,200817	0	Гипотеза о нормальности отклоняется			
12	0,392947	3		Критерий Крамера-Мизеса					
13	0,425692	2		115,6663	0	Гипотеза о нормальности отклоняется			
14	0,458438	1		Критерий Андерсона-Дарлингга					
15	0,491183	3		722,6446	0	Гипотеза о нормальности отклоняется			
16	0,523929	2		Критерий Шапиро-Уилка					
17	0,556674	4		Много данных. Расчет не выполнен.					
18	0,58942	1		Критерий Шапиро-Франсиа					
19	0,622165	3		0,876696	0,367667	Гипотеза о нормальности не отклоняется			
20	0,654911	3		Критерий коэффициента асимметрии					
21	0,687657	3		0,886606	0	Гипотеза о нормальности отклоняется			
22	0,720402	3		Критерий эксцесса					
23	0,753148	0		0,693005	0	Гипотеза о нормальности отклоняется			
24	0,785893	1		Критерий Жарка-Бера					
25	0,818639	3		2717,461	0	Гипотеза о нормальности отклоняется			

Рисунок 5 – Выходные данные проверки временного ряда на нормальность

Полученные данные временного ряда показывают его несоответствие нормальному закону распределения.

К числовым характеристикам формы частотных распределений относятся: асимметрия и эксцесс.

Асимметрия Ax показывает степень отклонения распределения от симметричного, который характерен для нормальности распределения, рассчитывается по формуле:

$$Ax = \frac{1}{(n-1)Sx^3} \sum_{i=1}^n x_i - Mx$$

и принимает значения от -3 до +3. При $Ax = 0$ распределение симметрично, при $Ax < 0$ – левосторонняя асимметрия, при $Ax > 0$ – распределение правосторонней асимметрии.

Эксцесс Ex (Куртозис) показывает степень острровершинности кривой распределения, рассчитывается по формуле:

$$Ex = \frac{1}{(n-1)Sx^4} \sum_{i=1}^n (x_i - Mx)^4 - 3$$

и принимает значения от -3 до +3. При $Ex = 0$ распределение нормальное, при $Ex < 0$ – плосковершинное, при $Ex > 0$ – острровершинное распределение.

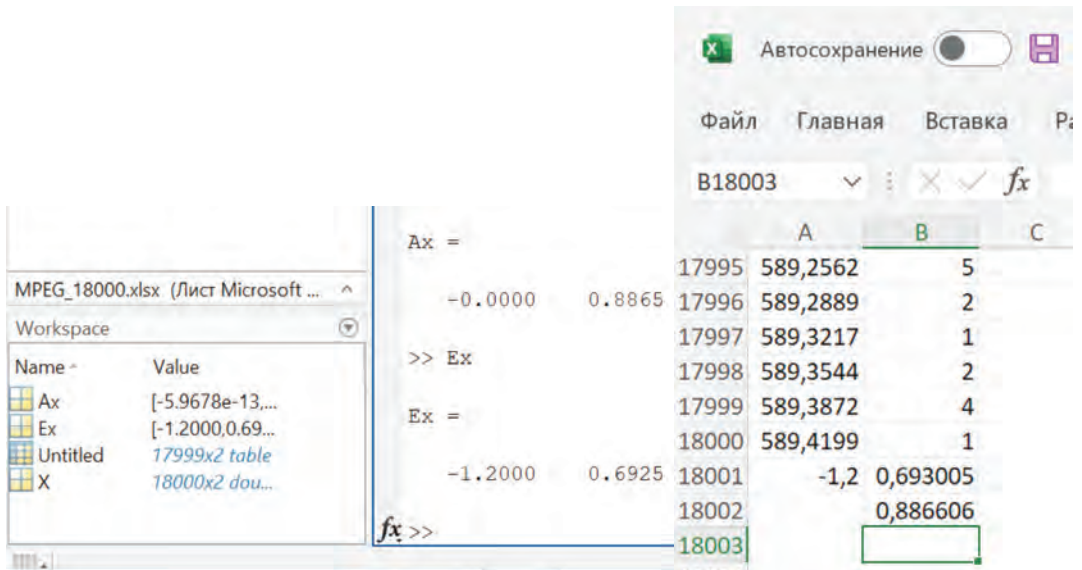


Рисунок 6 – Выходные данные асимметрии и куртозиса

Критерии коэффициентов асимметрии и эксцесса, полученные с использованием программ Attestat (рис.5) Matlab (рис.6), Excel (рис.6), совпадают. $Ax = 0,88 > 0$ – распределение правосторонней асимметрии. Куртозис $Ex = 0,69 > 0$. Оба значения указывают на несоответствие исследуемого ряда нормальному закону распределения.

Исследуем ряд на случайность [6-7] с помощью VRatio test в Matlab (рис.7). Тест отношения дисперсии Variance Ratio test for Random Walk оценивает нулевую гипотезу о том, что одномерный временной ряд является случайным блужданием.

Нулевая модель $Y(t) = c + y(t-1) + e(t)$, где c — константа дрейфа (случайный процесс), $e(t)$ — некоррелированные инновации с нулевым средним значением.

Variance Ratio Test for Random Walk(Y)									
Null Hypothesis: Y is a random walk									
Results									
Select	Null Rejected	P-Value	Test Statistic	Critical Value	Ratio	IID Innovations	Period	Significance Level	
1	<input type="checkbox"/>	true	0	-44.6162	1.9600	0.4887	false	2	0.0500

Рисунок 7 – Выходные данные проверки временного ряда на случайность

По полученным данным обнаружено, что исследуемый временной ряд является случайным блужданием.

Результаты и их обсуждение. В данной работе исследован одномерный временной ряд данных с целью дальнейшего использования этих данных для прогнозирования сетевого трафика. Так как закон распределения исследуемого временного ряда отличается от нормального, целесообразно дальнейшее исследование временного одномерного ряда на существование тенденции (тренда).

Заключение. Принадлежность наблюдаемых данных нормальному закону является необходимой предпосылкой для корректного применения большинства классических методов математической статистики [8-10]. Для проверки стационарности ряда используются параметрические и непараметрические тесты. Применение параметрических тестов возможно только в случае нормального распределения данных. Так как исследование временного ряда в данной статье показало, что он не подчиняется нормальному закону распределения и ряд нестационарный, дальнейшее исследование планируется проводить с помощью непараметрических тестов.

ЛИТЕРАТУРА

1 Федоров С. Л. Моделирование нестационарных временных рядов и построение оператора эволюции их выборочных распределений непараметрическими методами. Диссертация, 2017 – Сургутский государственный университет. – Сургут: 2017. – 108 с.

2 Мефферт Б. Инструменты обработки сигналов – основы, примеры применения и задачи [Текст]: Беате Мефферт, Олаф Хохмут – Авторы перевода Г.Безель, В.Баербах, 2018. 320с.

3 Вершинина М.С. Анализ предположений о стационарности некоторых временных рядов [Текст]: М.С. Вершинина. Сборник трудов Всероссийской конференции по математике с международным участием «МАК-2018» / АлтГУ и др. – Барнаул: Изд-во АлтГУ, 2018. – С. 172–176.

4 Д. С. Шингисов, В. Ю. Гойхман, А. Р. Лаврова, Ш. Ж. Сеилов, Е. Ш. Журсинбек. Об устойчивости распределений характеристик трафика WhatsApp. Вестник НИА РК. 2021 №4(82)

5 Мараев, В.С. Инструменты визуализации временных рядов в космических исследованиях. Том 1. [Текст]: Мараев В.С. – Исследования наукограда, 2017. – № 4 (22).

6 Brockwell P.J., Davis R.A. Introduction to Time Series and Forecasting. 3rd ed. – Springer, 2016. – 428 p.

7 Lo, A. W., and A. C. MacKinlay. A Non-Random Walk Down Wall St. Princeton, NJ: Princeton University Press, 2001.

8 Кобзарь А. И. Прикладная математическая статистика.– М.: Физматлит, 2006. – 816 с.

9 <https://www.machinelearningmastery.ru/time-series-data-stationary-python/>

10 Капитонова О.В. Прогнозирование социально-экономических процессов: Учебно-методическое пособие. [Текст]: О.В. Капитонова – Нижний Новгород: Нижегородский университет, 2016. – 74 с.

REFERENCES

1 Fedorov S. L. Modelirovanie nestacionarnykh vremennykh ryadov i postroenie operatora evolyucii ih vyborochnykh raspredelenij neparametricheskimi metodami. Dissertaciya, 2017 – Surgutskij gosudarstvennyj universitet. – Surgut: 2017. – 108 s.

2 Meffert B. Instrumenty obrabotki signalov – osnovy, primery primeneniya i zadachi [Tekst]: Beate Meffert, Olaf Hohmut – Avtory perevoda G.Bezeľ, V.Baerbah, 2018. 320s.

3 Vershinina M.S. Analiz predpolozhenij o stacionarnosti nekotorykh vremennykh ryadov [Tekst]: M.S. Vershinina. Sbornik trudov Vserossijskoj konferencii po matematike s mezhdunarodnym uchastiem “MAK-2018” / AltGU i dr. – Barnaul: Izd-vo AltGU, 2018. – S. 172–176.

4 D. S. SHingisov, V. YU. Gojhma, A. R. Lavrova, SH. ZH. Seilov, E. SH. ZHursinbek. Ob ustojchivosti raspredelenij harakteristik trafika WhatsApp. Vestnik NIA RK. 2021 №4(82)

5 Maraev, V.S. Instrumenty vizualizacii vremennykh ryadov v kosmicheskikh issledovaniyah. Tom 1. [Tekst]: Maraev V.S. – Issledovaniya naukograda, 2017. – № 4 (22).

6 Brockwell P.J., Davis R.A. Introduction to Time Series and Forecasting. 3rd ed. – Springer, 2016. – 428 p.

7 Lo, A. W., and A. C. MacKinlay. A Non-Random Walk Down Wall St. Princeton, NJ: Princeton University Press, 2001.

8 Kobzar’ A. I. Prikladnaya matematicheskaya statistika.– М.: Fizmatlit, 2006. – 816 с.

9 <https://www.machinelearningmastery.ru/time-series-data-stationary-python/>

10 Kapitonova O.V. Prognozirovaniye social’no-ekonomicheskikh processov: Uchebno-metodicheskoe posobie. [Tekst]: O.V. Kapitonova – Nizhnij Novgorod: Nizhegorodskij universitet, 2016. – 74 s.

Г. У. БЕКТЕМЫСОВА, Ж. Б. ИБРАЕВА

Халықаралық ақпараттық технологиялар университеті, Алматы қ., Қазақстан

УАҚЫТТЫҚ ҚАТАРДЫ СТАЦИОНАРЛЫҚҚА ЗЕРТТЕУ

Желілік трафиктің нақты MPEG пакетінің қарқындылығы желілік құрылғыларға қызмет көрсету кезінде пакеттердің келу жылдамдығы біркелкі емес. Заттардың интернеті (Internet of Things, IoT) концепциясының дамуымен желіні басқару міндеттерінің өзектілігі күн сайын артып келеді. Зерттеу зерттеушіні қызықтыратын сұрақтарға мазмұнды жауап алуға мүмкіндік береді. Қатарларды зерттеуде әртүрлі тесттерді (критерийлерді) қолдану өзекті болып табылады, өйткені олар қатардың құрылымын анықтауға мүмкіндік береді.

Желілік қолданбалардың әртүрлілігінің көбеюімен және деректерді берудің жаңа протоколдарының пайда болуымен трафик тәртібінде оның стационарлық еместігімен байланысты

қасиеттер мен мүмкіндіктер пайда бола бастады. Бұл мақалада Алматы қаласының тас жолында түсірілген нақты деректерден тұратын уақыттық қатар қарастырылады.

Уақыттық қатарды стационарлыққа зерттеу үшін Колмогоров-Смирнов, Шапиро-Уилк, Дарбин, Дэвид-Хартли-Пирсон, Андерсон-Дарлинг критерийлері, QQ сызбасы және де Matlab сандық модельдеу ортасында кездейсоқтықты анықтайтын Variance Ratio тесті қолданылды. Пирсон және Пуассон үлестірімдері бойынша қатарларды зерттеу үшін бағалар алынды. Таралу формасының сандық сипаттамалары да анықталды: асимметрия және эксцесс (куртозис).

Мақалада өлшенген уақыт қатарының оңайлатылған тексерулері жүргізілді.

Түйін сөздер: уақыттық қатар, желілік трафик, мәліметтерді талдау, стационарлық сынақ критерийлері, қалыпты таралу заңы.

G. U. BEKTEMYSOVA, J. B. IBRAEVA

International University of Information Technologies, Almaty, Kazakhstan

RESEARCH OF A TIME SERIES FOR STATIONARITY

Real MPEG packet intensity network traffic has a non-uniform rate of packet arrival at serving network devices. With the development of the concept of the Internet of Things (IoT), the relevance of network management tasks is growing every day. The study allows you to get meaningful answers to the questions of interest to the researcher. The use of various tests (criteria) in the study of a series is relevant because they allow us to reveal their structure.

With the increase in the variety of network applications, new data transfer protocols, properties and features that are associated with its non-stationarity began to appear in the behavior of traffic. This article considers a time series with real data taken on the highway of the city of Almaty.

To study the time series for stationarity, the Kolmogorov-Smirnov, Shapiro-Wilk, Darbin, David-Hartley-Pearson, Anderson-Darling criteria, QQ plot, Variance Ratio test for Random Walk data in the Matlab numerical simulation environment were applied. Estimates are obtained for studying the series in terms of Pearson and Poisson distributions. Numerical characteristics of the form of distributions are also determined: asymmetry and kurtosis.

The article carried out simplified checks of the measured time series.

Key words: time series, network traffic, data analysis, stationarity test criteria, normal distribution law.