Торайғыров университетінің хабаршысы
ҒЫЛЫМИ ЖУРНАЛЫ

НАУЧНЫЙ ЖУРНАЛ
Вестник Торайгыров университета

Торайғыров университетінің ХАБАРШЫСЫ

Энергетикалық сериясы

1997 жылдан бастап шығады

ВЕСТНИК Торайгыров университета

Энергетическая серия

Издается с 1997 года

ISSN 2710-3420

№ 3 (2022)

ПАВЛОДАР

НАУЧНЫЙ ЖУРНАЛ
Вестник Торайгыров университета

Энергетическая серия
выходит 4 раза в год

СВИДЕТЕЛЬСТВО
о постановке на переучет периодического печатного издания,
информационного агентства и сетевого издания
№ 14310-Ж
выдано
Министерство информации и общественного развития
Республики Казахстан

Тематическая направленность
публикация материалов в области электроэнергетики,
электротехнологии, автоматизации, автоматизированных и
информационных систем, электромеханики и теплоэнергетики

Подписной индекс – 76136

https://doi.org/10.48081/AFHU6838

SRSTI 50.39.15

https://doi.org/10.48081/UVJS2247

D. A. Naubetov[1], M. Z. Yakubova[2], S. A. Mirzakulova[3], T. G. Serikov[4]

[1,2]Almaty University of Power Engineering
and Telecommunication named Gumarbek Daukeev,
Republic of Kazakhstan, Almaty;
[3]Turan University, Department of Radio Engineering,
Republic of Kazakhstan, Almaty;
[4]Kazakh Agrotechnical University named after S.Seifullin,
Republic of Kazakhstan, Nur-Sultan

NONPARAMETRIC CRITERIA FOR ASSESSING THE TREND OF NETWORK ACCESS TRAFFIC

The article uses nonparametric criteria to assess the trend of network access traffic. Practice shows that most real processes do not satisfy the properties of stationarity. They are characterized by a trend, systematic changes in variance, the presence of periodic fluctuations and the presence of changing interdependencies between the levels of the time series. A review of some foreign articles devoted to this topic has been carried out. Also, the Wireshark sniffer program is used, the task of which is to capture packets passing through the network. Empirical data were obtained on the access network segment between LAN and ISP. In the article, using nonparametric tests (the series criterion, the «ascending and descending» series criterion, the Cox-Stewart sign criterion and the «Caterpillar» method), the hypothesis of the presence of a trend in the measured time series on the access network was tested. The decomposition of the time series into its components is performed.

Keywords: access network, time series, trend, stationarity, nonparametric criteria.

Introduction

The functioning modern network architecture NGN/IMS (Next Generation Network/IP Multimedia Subsystem) promotes the development of access network technology based on optical infrastructure, in order to provide a wide range of multimedia services to users and numerous enterprises subscribed to VoIP service

through an ISP (Internet Service Provider) service provider with a variety of converged communication solutions such as IP-PBX Asterisk – an indispensable application for corporate office.

The rapid growth in the number of local networks and networks of Internet service providers, the rapid and continuous emergence of new network technologies aimed at improving the quality of service (QoS) of real-time applications indicates the relevance of the study of network traffic between LAN and ISP.

Review of some foreign articles on this topic:

In [1] «Fundamentals of the Theory of Complex Systems», the authors describe that most systems, due to their complexity, cannot be modeled with sufficient accuracy and that this can be done using another approach based on observing their behavior. An observable is a sequence of values of some variable recorded continuously or at certain intervals (time series). The presence of only a time series instead of a complete solution of the equations limits knowledge about the system under study, but at the same time, the identification problem in the analysis of the measured data provides answers to questions about the parameters of the system that gave rise to this time series – embedding dimension, correlation dimension, entropy, and others. It is also described that theoretical research based on time series analysis is a powerful tool for understanding many phenomena occurring in the system.

In the work [2] «Review of modern models and methods of time series analysis of the dynamics of processes in social, economic and sociotechnical systems» provides general information about time series and the tasks of their analysis, as well as modern methods of time series analysis. It is noted that the processes occurring in complex systems cannot be considered completely random, since they tend to self-organize and, moreover, are influenced by the memory of previous states. The dynamics of processes in social, economic and sociotechnical systems are described by non-stationary time series, and attempts to represent them in the form of quasi-stationary sections can lead to an incorrect description of the observed processes and incorrect construction of control forecasts. When describing the evolution of time series, an important point is the definition of the so-called “breakdown” points or, in another way, «trend change points». Most traditional methods and models for detecting the effect of the «trend change point» are based on the assumption that the probabilistic characteristics of the output signal of a serviceable technical object do not change over time, i.e. the measured stochastic signal is stationary. The authors of this paper state that it is necessary to search for new methods for analyzing the dynamics of complex systems or new approaches to describing non-stationary time series, especially if self-organization of such systems and the presence of memory of previous states is possible, for example, models and methods based on various types or combinations of machine

learning algorithms, such as neural networks, fuzzy logic, the method of regression support vectors and others.

In the work [3] «Building an experimental stand for studying the problem of time series classification in order to improve the quality of forecasting» it is described that time series are one of the most important objects in the field of machine learning. The time series contains a lot of information. Identification of the structure of the time series is necessary in order to build a mathematical model of the phenomenon described by the time series.

The study of the appearance of a diverse type and volume of traffic leads to the appearance of non-stationary traffic, unlike the previous network based on channel switching and with flows called the simplest, which had the properties of stationarity, ordinariness and with no aftereffect. They belonged to the general population of the normal distribution.

To recognize the non-stationarity of the time series, various methods can be used: visual analysis of the graphical representation of the time series for the presence of a trend and a periodic component, the method of averages, nonparametric tests, etc.

Materials and methods

Empirical data on the access network segment between LAN and ISP were obtained using the Wireshark sniffer program, which plays the role of capturing packets passing through the network.

In just five hours, 7655 packages were tracked. Figure 1 shows a sample from some general population of volume , or in another way a one-dimensional time series that characterizes the intensity of the number of data packets for every three minutes of time. At the same time, it is visually visible that the distribution of the number of packets has an uneven intensity, there are places with discharged areas where there are few incoming packets.

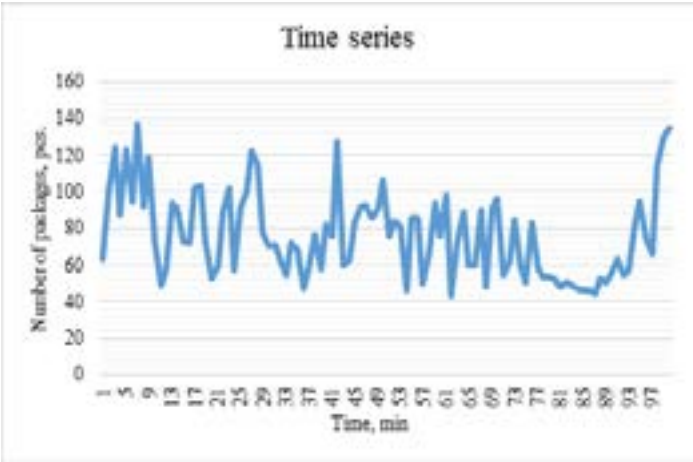

Figure 1 – Time series packet intensity on the access network

To find out whether the data is dependent or independent, there are parametric and nonparametric tests.

Parametric tests are those that make assumptions about the distribution parameters of the general population from which the sample is taken. It is often assumed that these data have a normal distribution.

Nonparametric tests are statistical conclusions that do not require any assumptions about the parameters of the studied population [4]. In this case, it is not the measured values themselves that are processed, but its rank (position within the sample) or frequencies.

To check the independence and stationarity of a number of observations, we will use nonparametric methods:

- series criterion based on the median of the sample;
- criterion of «ascending and descending» series;
- the iconic criterion of the Cox-Stewart trend.

Results and discussion

The series criterion allows you to check whether the order of occurrence of two values of a variable is random. A series is a sequence of similar observations. If there are either too many series or too few in the sample, then this sample is not random [5].

To identify the random order of occurrence of values or the fact of the presence or absence of a non-random (i.e., t-dependent) component as a whole, or to find out in another way whether there is a trend in the original sample at all, we will use the series criterion based on the median of the sample.

After constructing a variation series from the original time series, the median value of equal to 73 is determined.

Then the procedures of the series criterion are performed – accounting as a plus if and minus if . In this case, the members of the sample equal to in the sequence of pros and cons obtained in this way are not taken into account.

The resulting sequence of pros and cons is characterized by the total number of series and the length of the longest series .

By «series» is meant a sequence of consecutive pluses or consecutive minuses. The essence of testing the null hypothesis of randomly selecting values in the sample is that in the absence of systematic influence, the alternation of «+» and «-» should also occur randomly, therefore the number of series should be large enough, and the length of the longest series should not be large.

The values of and represent a two-dimensional random variable, which is a statistical criterion for testing the hypothesis.

The necessary calculations were given:

If one of the inequalities turns out to be violated, then the hypothesis of stochastic independence of the initial data is rejected with an error probability between :

$$v(n) > \left[\frac{1}{2}(n + 1 - 1{,}96\sqrt{n-1}\right] \quad (1)$$

$$\tau(n) < [3{,}3\, log_{10}(n+1)] \quad (2)$$

$$v(n) > \left[\frac{1}{2}(100 + 1 - 1{,}96\sqrt{100-1}\right] = 40{,}75$$

$$\tau(n) < [3{,}3\, log_{10}(100+1)] = 6{,}614$$

The hypothesis of randomness is rejected, since both inequalities are violated. Since u(n)=34 is less than 40.75 and τ(n)=17 is greater than 6.614.

This means there is a dependence in the observation, that is, evidence in favor of the presence of a trend.

The criterion of “ascending and descending” series is a more powerful criterion compared to the criterion of series based on the median.

The criterion of “ascending and descending” series, according to which the trend is determined by the following algorithm [6]:

$$\tau(n) < [3{,}3\, log_{10}(100+1)] = 6{,}614 \quad (3)$$

This criterion captures the bias of the estimation of mathematical expectation of a monotonous and periodic nature. If the subsequent observation is equal to the previous one, then only one observation is taken into account. Next, the number of series ν(n) is calculated [7].

As a result, the number of series was ν(n) =57 and the longest series was

$$l_{max}(n) = 4.$$

If at least one of the following inequalities is violated, then the hypothesis of the absence of a trend is rejected with a confidence probability of 0.95:

$$\begin{cases} \nu(n) > \left[\frac{1}{3}(2n-1) - u_t\sqrt{\frac{16n-29}{90}}\right] \\ l_{max}(n) < t_{kp} \end{cases} \quad (4)$$

where – quantile of the normal level distribution (1–α)/2;

$$t_{kp} = 6, \text{ by } 26 \le n < 153 \ (n = 100)$$

$$\begin{cases} \nu(n) > \left[\frac{1}{3}(200-1) - 1{,}96\sqrt{\frac{1600-29}{90}}\right] = 58{,}14 \\ l_{max}(n) < 6 \end{cases}$$

The numbers $\nu(n)$ and $l_{max}(n)$ must be rounded down to the nearest integer.

If at least one of the inequalities is violated, it is assumed that there is a trend (trend).

If both conditions are met, respectively, there is no trend (trend).

As a result, we get:

$$\begin{cases} 57 > 58{,}14 \\ 4 < 6 \end{cases}$$

Checking the fulfillment of the conditions shows that one of the conditions is not fulfilled, so the hypothesis of the presence of a trend is accepted.

The method of the «Iconic Cox-Stewart trend criterion» was developed in 1955. This test requires dividing the studied series consisting of n values into three equal groups of 34 values each (n/3) in our case [8].

At the same time, we have the first x^1_i second x^2_i and third x^3_i groups, but the first and third groups (x^1_i, x^3_i) will be considered with the difference of values

(x^3_i, x^1_i). Ultimately, based on these data, a «plus» sign is fixed with an increasing trend (s^+), a «minus» sign with a decreasing trend (s^-), zero with no trend (s0),. so $s^+ + s^- + s^0 = n/3$.

$$\hat{z} = \frac{\left|S - \frac{n}{6}\right| - 0{,}5}{\sqrt{\frac{n}{12}}} = 6{,}65. \quad (5)$$

where $S = \max\{s^+, s^-\}$, approximately distributed normally.

To calculate the value of ž, we use the data obtained during the study: $s^+ = 9$, $s^- = 25$, $s^0 = 0$, $S = \max\{9,25\} = 25$.

$$\hat{z} = \frac{\left|25 - \frac{34}{6}\right| - 0{,}5}{\sqrt{\frac{34}{12}}} = 6{,}65. \quad (6)$$

If $\hat{z} = 6{,}65 > z_{0,05} = 1{,}96$ ($\alpha = 5\ \%$), then the null hypothesis is canceled, the time series has a trend.

Three nonparametric criteria confirmed the presence of a trend in the studied time series.

Singular spectral analysis (the «Caterpillar» method) was developed in Russia, and SSA (Singular Spectrum Analysis) was developed in the UK and the USA. This method of studying the structure of time series combines the advantages of the principal component method, Fourier analysis, the classical Karunen-Loev theorem and regression analysis. It transforms a one-dimensional series into a multidimensional one using a one-parameter shift procedure in order to decompose the time series into simple components: slow trends, seasonal and other periodic or oscillatory components, as well as noise components [9].

The «Caterpillar» method is nonparametric and does not require the one-dimensional series under study to belong to the corresponding distribution laws, preliminary stabilization of the series and does not depend on whether the series under study is stationary or not, it also has only one configurable parameter and the basic algorithm is implemented on the basis of simple linear algebra using a discrete convolution operator [10].

The original time series is embedded in the AtteStat program for decomposition into additive components in the amount of 15 and its petal diagram is shown in Figure 2.

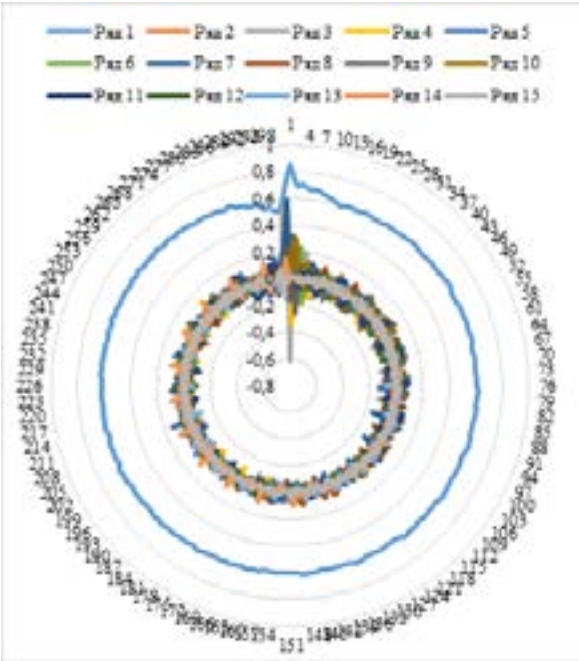

Figure 2 – Decompositions of the original time series into additive components

The above graph shows an explicit decomposition of a complex series into two parts:

- the first part with one low-frequency component of the time series is a trend (low-frequency oscillation, slowly changing);

- the second part is a set of medium and high frequencies of the components of the time series, which are the periodic components of the series and noise (high-frequency components).

Information about financing

This article was prepared as part of the government contract as requested by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan on the subject formulated as «Development of a method for improving the security of a telecommunications network based on IP-PBX Asterisk» (project No. AP14871745).

Conclusions

The actual measured data has a complex structure. Nonparametric tests (the series criterion, the «ascending and descending» series criterion and the Cox-

Stewart sign trend criterion) confirmed the presence of a trend in the original time series. The Caterpillar method graphically displayed the trend as a low-frequency component of a complex time series structure.

REFERENCES

1 **Лоскутов, А. Ю., Михайлов, А.** С. Основы теории сложных систем [Текст]. – М. – Ижевск: Институт компьютерных исследований, 2007. – С. 620.

2 **Андрианова, Е. Г., Головин, С. А., Зыков, С. В., Лесько, С. А., Чукалина, Е. Р.** Обзор современных моделей и методов анализа временных рядов динамики процессов в социальных, экономических и социотехнических системах. Russian Technological Journal. 2020; 8 (4)– С. 7 – 45.

3 Построение экспериментального стенда для исследования задачи классификации временных рядов с целью повышения качества прогнозирования [Электронный ресурс] – http://www.machinelearning.ru/wiki/index.php?title=%D0%A3%D1%87%D0%B0%D1%81%D1%82%D0%BD%D0%B8%D0%BA:Stsupkov, (дата обращения 05. 08. 2022).

4 Непараметрические тесты [Электронный ресурс], – https://www.statmethods.ru/statistics-metody/neparametricheskie-testy/, (дата обращения 05.08.2022).

5 **Градов, В. М., Овечкин, Г. В., Овечкин, П. В., Рудаков, И. В.** Компьютерное моделирование [Текст]: учебник. – М. : КУРС : ИНФРА-М, 2020. – 264 с. – ISBN 978–5-906818–79–9.

6 **Кремер, Н. Ш.** Теория вероятностей и математическая статистика [Текст]: Учебник для ВУЗов. 2-е изд., перераб. и доп. – М.: ЮНИТИ – ДАНА, 2004. – 573 с. –. ISBN 5-238-00573–3.

7 **Тимофеев, В. С.** Методы и алгоритмы анализа нечисловых данных [Текст]: учеб. пособие / В. С. Тимофеев, А. Ю. Тимофеева. – Новосибирск: Изд-во НГТУ, 2021. – 83 с. – 100 экз. – ISBN 978-5–7782–4483–2.

8 Ledvinka, Ondrej. Is the Cox-Stuart test for trend really insensitive to autocorrelation? // Conference: FRIEND-Water 2014. – 2014. – P. 68–69.

9 **T. Fujiwara, Shilpika, N. Sakamoto, J. Nonaka, K. Yamamoto and K. – L.** Ma, «A Visual Analytics Framework for Reviewing Multivariate Time-Series Data with Dimensionality Reduction,» in IEEE Transactions on Visualization and Computer Graphics, – Vol. 27, no. 2, P. 1601-1611, Feb. 2021, doi: 10.1109/TVCG.2020.3028889.

10 **Голяндина, Н. Э.** Метод «Гусеница» – SSA: анализ временных рядов [Текст]: Учеб.пособие. – СПб., 2004. – С. 76.

REFERENCES

1 **Loskutov, A. YU., Mihajlov, A. S.** Osnovy teorii slozhnyh sistem [Fundamentals of the theory of complex systems]. – M. – Izhevsk: Institut komp'yuternyh issledovanij, 2007. – S. 620.

2 **Andrianova, E. G., Golovin, S. A., Zykov, S. V., Les'ko, S. A., Chukalina, E. R.** Obzor sovremennyh modelej i metodov analiza vremennyh ryadov dinamiki processov v social'nyh, ekonomicheskih i sociotekhnicheskih sistemah [Review of modern models and methods for analyzing time series of process dynamics in social, economic and sociotechnical systems]. Russian Technological Journal. 2020; 8(4): S. 7 – 45.

3 Postroenie eksperimental'nogo stenda dlya issledovaniya zadachi klassifikacii vremennyh ryadov s cel'yu povysheniya kachestva prognozirovaniya [Elektronnyj resurs], – http://www.machinelearning.ru/wiki/index.php?title=%D0%A3%D1%87%D0%B0%D1%81%D1%82%D0%BD%D0%B8%D0%BA:Stsupkov, (data obrashcheniya 05.08.2022).

4 Neparametricheskie testy [Elektronnyj resurs], - https://www.statmethods.ru/statistics-metody/neparametricheskie-testy/, (data obrashcheniya 05.08.2022).

5 **Gradov, V. M., Ovechkin, G. V., Ovechkin, P. V., Rudakov, I. V.** Komp'yuternoe modelirovanie [Computer simulation]: uchebnik. – M: KURS: INFRA–M, 2020. – 264 s. – ISBN 978–5-906818–79–9.

6 **Kremer, N. S.** Teoriya veroyatnostej i matematicheskaya statistika [Probability theory and mathematical statistics]: Uchebnik dlya VUZov. 2-e izd., pererab. i dop. – M. : YUNITI-. DANA, 2004. – 573 s. –. ISBN 5-238-00573-3.

7 **Timofeev, V. S.** Metody i algoritmy analiza nechislovyh dannyh [Methods and algorithms for analyzing non-numeric data]: ucheb. posobie / V. S. Timofeev, A. YU. Timofeeva. – Novosibirsk: Izd-vo NGTU, 2021. – 83 s. – 100 ekz. – ISBN 978-5-7782-4483-2.

8 **Ledvinka, Ondrej.** Is the Cox-Stuart test for trend really insensitive to autocorrelation? // Conference: FRIEND-Water 2014. – 2014. – P. 68–69.

9 **T. Fujiwara, Shilpika, N. Sakamoto, J. Nonaka, K. Yamamoto and K. -L. Ma**, «A Visual Analytics Framework for Reviewing Multivariate Time-Series Data with Dimensionality Reduction,» in IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 2, pp. 1601-1611, Feb. 2021, doi: 10.1109/TVCG.2020.3028889.

10 **Golyandina, N. E.** Metod «Gusenica»-SSA: analiz vremennyh ryadov [Caterpillar-SSA method: Time series analysis]: Ucheb.posobie. – SPb., 2004. – C. 76.

Material received on 15.09.22.

**Д. А. Наубетов[1], М. З. Якубова[2], Ш. А. Мирзакулова[3], Т. Ғ. Сериков[4]*
[1,2]Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті, Қазақстан Республикасы, Алматы қ;
[3]Тұран университеті, Қазақстан Республикасы, Алматы қ.;
[4]С. Сейфуллин атындағы Қазақ агротехникалық университеті, Қазақстан Республикасы, Нұр-Сұлтан қ.
Материал баспаға 15.09.22 түсті.

ЖЕЛІЛІК ҚАТЫНАУ ТРАФИГІН БАҒАЛАУДЫҢ ПАРАМЕТРЛІК ЕМЕС КРИТЕРИЙЛЕРІ

Мақалада желіге кіру трафигінің тенденциясын бағалау үшін параметрлік емес критерийлер қолданылады. Тәжірибе көрсеткендей, көптеген нақты процестер стационарлық қасиеттерді қанағаттандырмайды. Олар трендпен, дисперсияның жүйелі өзгеруімен, мерзімді ауытқулардың болуымен және уақыт қатарларының деңгейлері арасында өзгеретін өзара тәуелділіктердің болуымен сипатталады. Осы тақырыпқа арналған кейбір шетелдік мақалаларға шолу жасалды. Wireshark сниффер-багдарламасы қолданылды, оның міндеті – желі арқылы өтетін пакеттерді ұстау. LAN және ISP арасындағы қол жеткізу желісінің аймағында эмпирикалық мәліметтер алынды. Мақалада параметрлік емес тестілердің көмегімен (серия критерийі, «жоғары және төмен» сериялары критерийі, Кокс-Стюарттың белгі критерийі және «Жұлдызқұрт» әдісі) қол жеткізу желісінде өлшенген уақыт қатарында трендтің болуы туралы гипотезаны тексеру жүзеге асырылды. Уақыт қатарының оның компоненттеріне ыдырауы орындалды.

Кілтті сөздер: қол жеткізу желісі, уақыт сериясы, тренд, стационарлық, параметрлік емес өлшемдер.

**Д. А. Наубетов[1], М. З. Якубова[2], Ш. А. Мирзакулова[3], Т. Ғ. Сериков[4]*
[1,2]Алматинский университет энергетики и связи имени Гумарбека Даукеева, Республика Казахстан, г. Алматы;
[3]Университет Туран, Республика Казахстан, г. Алматы;
[4]Казахский агротехнический университет имени С. Сейфуллина, Республика Казахстан, г. Нур-Султан
Материал поступил в редакцию 15.09.22.

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ ОЦЕНКИ ТЕНДЕНЦИИ СЕТЕВОГО ТРАФИКА ДОСТУПА

В статье используются непараметрические критерии для оценки тенденции трафика доступа к сети. Практика показывает, что большинство реальных процессов не удовлетворяет свойствам стационарности. Они характеризуются трендом, систематическими изменениями дисперсии, наличием периодических флуктуаций и наличием изменяющихся взаимозависимостей между уровнями временного ряда. Выполнен обзор некоторых иностранных статей, посвященных данной теме. Также, использована программа-сниффер Wireshark задача которой является захват пакетов, проходящих через сеть. Получены эмпирические данные на участке сети доступа между LAN и ISP. В статье с помощью непараметрических тестов (критерий серий, критерий «восходящих и нисходящих» серий, знаковый критерий Кокса-Стюарта и метода «Гусеница») осуществлена проверка гипотезы о наличии тренда в измеренном временном ряде на сети доступа. Выполнено разложение временного ряда на его составляющие.

Ключевые слова: сеть доступа, временной ряд, тренд, стационарность, непараметрические критерии.

Теруге 15.09.2022 ж. жіберілді. Басуға 30.09.2022 ж. қол қойылды.
Электронды баспа
3 Mb RAM
Шартты баспа табағы 19,8. Таралымы 300 дана. Бағасы келісім бойынша.
Компьютерде беттеген: А. К. Мыржикова
Корректор: А. Р. Омарова
Тапсырыс № 3989

Сдано в набор 15.09.2022 г. Подписано в печать 30.09.2022 г.
Электронное издание
3 Mb RAM
Усл. печ. л. 19,8. Тираж 300 экз. Цена договорная.
Компьютерная верстка: А. К. Мыржикова
Корректор: А. Р. Омарова
Заказ № 3989

«Toraighyrov University» баспасынан басылып шығарылған
Торайғыров университеті
140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы
Торайғыров университеті
140008, Павлодар қ., Ломов к., 64, 137 каб.
67-36-69
E-mail: kereku@tou.edu.kz
www.vestnik-energy.tou.edu.kz