Journal of Theoretical and Applied Information Technology
15th November 2022. Vol.100. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DECOMPOSITION OF MEASURED DATA ON THE NETWORK SEGMENT BETWEEN LAN AND ISP

AMREEV M[1]., NAUBETOV D.A.[2] , YAKUBOVA M.Z.[3] , MIRZAKULOVA S.A.[4] , SERIKOV T.G.[5]

[1,2,3] Almaty University of Power Engineering and Telecommunication named Gumarbek Daukeev, Almaty, Kazakhstan

[4] Turan University, Almaty, Kazakhstan

[5] Kazakh Agrotechnical University named after S.Seifullina, Nur-Sultan, Kazakhstan

E-mail:[1]m.amreev@aues.kz,[2]daul.naubetov@gmail.com, [3]mubor194@gmail.com, [4]mirzakulova@mail.ru, [5]tansaule_s@mail.ru

ABSTRACT

Functioning packet trunks experience a constant increase in load in wired access networks, which in turn are transformed and becoming more and more optical in accordance with GPON (Gigabit Passive Optical Network) technology. It is established that numerous studies of the measured data on packet network trunks confirm that they are not stationary, and their structure is multicomponent, and there are no studies of the structure of network access traffic. Therefore, this article is devoted to the study of network traffic between LAN and ISP. On the basis of the measured data, scattering diagrams are constructed, a statistical assessment of the correlation of the measured series to the general population of the normal distribution is carried out, a correlogram is constructed and the decomposition of the time series is performed by the SSA (Singular spectrum analysis) method, which has a strict justification within the framework of the theory of dynamical systems. Also, it is concluded that the traffic entering the backbone network also has a complex structure, with its decomposition into the main components of the time series.

Keywords: *LAN, ISP (Internet service provider), SSA (Singular Spectrum Analysis), Time series analysis, OPNET Modeler, Wireshark*

1. INTRODUCTION

The evolution of the development of a unified telecommunications network in the world began to be based on packet IP technology instead of the traditional circuit-switched network, which caused a sharp increase in the volume of data related to information flows from the increase in the use of network technologies by humans. Thanks to convergent solutions, subscribers have access to a variety of services with the same quality, regardless of whether they are in a stationary or mobile environment.

The rapid growth of developing telecommunications services contributes to the evolution of access network technologies and its improvement depending on the development of transport network technologies.

The evolution of the subscriber access network development depends on:

- the basic network (transport network technologies) functioning at that time;
- evolution of the development of switching systems;
- development of technologies of guiding systems;
- trends in the development of subscriber access network technologies;
- the number of growing applications;
- variety of terminal devices.

For broadband signal transmission, a large bandwidth is required so that the signal can carry a larger amount of information. Bandwidth-intensive applications such as video require fast broadband networks capable of delivering content in «real time». As a result, broadband connectivity to telecom operators and service providers is shifting from copper-based metal access (traditional subscriber line) to optical access using optical fibers that enable high-speed data transmission. In addition, demand is growing for Triple Play services, in which one broadband connection provides telephone, broadcast, video services and simultaneously receiving Internet services.

Moreover, optical, digital transmission, VoIP, IPTV, audio, network management, etc. are carried out.

Access networks around the world are constantly evolving and include more and more fiber-optic technologies. Optical fiber already makes up the majority of the underlying networks and will gradually approach the customer until a complete transition to it is achieved. We are talking about access technologies:

- FTTx (Fiber To The x) Internet access with transfer speeds up to 100 Mbit/s;
- GPON/GEPON (Gigabit Passive Optical Network/Gigabit Ethernet PON) – PON with a bandwidth of 10 Gbit/s. GPON is regulated by the G standard.984. The transmission rate is 2488 Mbit/s in the direction of the subscriber and 1244 Mbit/s in the direction from the subscriber, and the symmetric GEPON is regulated by the IEEE 802.3ah standard and has a transmission rate of 1.25 Gbit/s in both directions;
- XG-PON with asymmetric mode 10GPON downlink speed 9.953 Gbit/s, uplink speed 2.488 Gbit/s. With symmetric mode XGS-PON 10G GPON, the linear speeds of incoming and outgoing traffic are 9.953 Gbit/s;
- 40G PON (ITU-T G.989) with time multiplexing, wavelength division (WDM) and 10G GPON wavelength cascading to support bandwidth up to 4 x 9.953 Gbit/s;
- the concept of future optical access systems FOAS (Future Optical Access Systems) using an alternative optical distribution network ODN (Optical Distribution Network) and ensuring the implementation of WDM-PON with wavelength routing (WR-WDM-PON).

Optical access systems of the future should work even faster and have even greater bandwidth, be more economical and provide a high level of usability.

The heterogeneous nature of access networks, which is in constant development, and a wide range of terminal devices with an increasing number of applications using the access network reach the backbone of a multiservice heterogeneous network through the ISP.

The rapid growth in the number of local area networks and ISP networks, the rapid and continuous emergence of new network technologies aimed at improving the speed and reliability of access networks, requires research and improvement of traffic flow between LAN and ISP.

Based on the above, with the expansion of the access network, there is an urgent need to study the traffic structure as described in [1] that most systems, due to their complexity, cannot be modeled with sufficient accuracy and that theoretical studies based on time series analysis are a powerful tool for understanding many phenomena occurring in a complex system related to the timely management of flows, classification and prioritization in the backbone of the network in the best way. It is in packet trunks that traffic has self-similarity properties, the processes under study are considered chaotic, containing deterministic chaos [2, 3]. A multiservice network considers access networks as a vector of user (user group) performance, which consists of several components reflecting traffic properties (voice traffic, text data traffic, video and management network) and it is believed that these components are statistically independent when serving a mixture of packet streams by network devices in a digital economy [4].

In [5] «The Effect of ISP Traffic Shaping on User-Perceived Performances in Broadband Shared Access Networks», Internet provider traffic is considered using models based on user behavior and application/session level metrics that provide quantitative measurements of user-perceived performance for HTTP, FTP and streaming video traffic. The results of this work can provide ISPs with valuable information about the design, deployment and operation of next-generation access networks from the point of view of end users, especially for monitoring peak speeds and spikes. The authors noted that access traffic is pulsating, and in [6], traffic researchers in the BellCore network described, among other properties, the presence of pulsation not only in short periods of time, but also in other time intervals. That is, pulsating processes are already entering the highways.

In [7] «On the Impact of Guest Traffic in Open-Access Domestic Broadband Sharing Schemes», a study of the influence of guest traffic of the upstream channel on the quality of broadband communication both in general and in the policy of package planning for sharing broadband access was conducted. The results confirm the suitability of the hybrid package planning policy for broadband access sharing schemes (home and guest users). Only a few tens of kilobytes per second of upstream guest traffic can be allowed by users without compromising their broadband access quality. This work paves the way for the development of a more complex package planning policy of open access to the internal broadband connection of the data exchange scheme.

In [8] «Improving the Performance of a Network by Managing the Bandwidth Utilization with squidGuard: A Case Study», the LAN of the Mining and Technological University with a download and upload bandwidth of 60 MB is considered, which is connected to the network operating center by a fiber-optic cable and distributed among 3000 users using a Cat 6 Ethernet cable. More in addition, there are wireless connections. The behavior model and trends in the use of network users' bandwidth are investigated using the Squid Analysis report generator with the definition of an effective bandwidth control policy. As a result of installing squidGuard on the firewall server and defining protocols, access policies for effective monitoring, network traffic control, applying priority access to legitimate users, restricting access to low-priority, bandwidth-demanding websites and applications, we have achieved a significant increase in the speed and security of the local network. At the same time, the access network manages the secure access of users to the network and performs segmentation using the technology of switched local networks.

In [9] «Towards Fully-Shared Access: Designing ISP Service Plans Leveraging Excess Bandwidth Allocation», a study of the problems of ISP traffic formation and the corresponding fixed-rate tariff plans in the public access network is conducted. The formation of subscriber traffic based on a bucket of tokens leads to the loss of network resources in public access when there are few active subscribers. Therefore, it is not possible to allocate extra bandwidth in the long term. Based on new bandwidth-related traffic management schemes, a hybrid ISP traffic management scheme is proposed with the gradual introduction of excess bandwidth in collective access.

The article [10] The «Caterpillar» - SSA method for analysis of time series with missing values» is devoted to the problem of applying singular spectral analysis to time series with missing data. A method of filling in the missing data is proposed and applied to time series of finite rank. Conditions for accurate reconstruction of missing data are developed and versions of the algorithm applicable to real time series are presented. The proposed algorithms lead to the extraction of additive components of time series, such as trends and periodic components, while filling in the missing data.

In the article [11] «Singular spectrum analysis and forecasting of failure time series», SSA is used to decompose and predict failure behavior using time series associated with time-to-failure data. Singular Spectrum Analysis (SSA) is a relatively new approach used to model time series without any assumptions about the underlying process. SSA can decompose the original time series into the sum of independent components that represent trend, oscillatory behavior (periodic or quasi-periodic components) and noise. The results are compared with previous approaches and show that SSA is a promising approach for data analysis and prediction of failure time series.

In the article [12] «Improving reconstruction of time-series based in Singular Spectrum Analysis: A segmentation approach», the use case of SSA is aimed at: improving the recovery and separation of components for non-stationary time series; enabling the analysis of large data sets, avoiding the problems of combining many segments; and presenting some advantages of segmentation in terms of better time–frequency characteristics. SSA may be able to decompose a time series into its significant components: trends, fluctuations and noises. However, if the analyzed signal is non-stationary, with an expansion of its spectrum and a change in time, the reliability of recovery is guaranteed only when using many elementary matrices. Consequently, the ability to distinguish dominant structures from time series may be impaired. To work around this problem, a new method called overlap-SSA (ov-SSA) is proposed for segmentation, analysis and recovery of long-term and/or non-stationary signals. The raw time series is divided into smaller, sequential and overlapping segments, and standard SSA procedures are applied to each segment with the resulting series combined.

The paper [13] «Randomized singular spectrum analysis for long time series» proposes a randomized SSA, which is an alternative to SSA for long time series without loss of analysis quality. Singular Spectrum Analysis (SSA) is a relatively new method of time series analysis and is a nonparametric alternative to classical methods. This methodology has proven its effectiveness in the analysis of nonstationary and complex time series, since it is a nonparametric method and does not require classical assumptions about the stationarity or normality of the residuals. Although SSA has proven its advantages over traditional methods, the problems that arise when considering long time series make standard SSA very computationally complex and often unusable. SSA and randomized SSA are compared in terms of model fit and prediction quality, as well as calculation time. This is done using Monte Carlo simulation and real data.

Singular Spectrum Analysis (SSA), since the second half of the 20th century, is a rapidly developing method of time series analysis. Since this can be called Principal Component Analysis (PCA) for time series, SSA will certainly become the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

standard method of time series analysis and signal processing in the future. Moreover, the problems solved by SSA are much broader than those of PCA. In particular, within the framework of SSA, the problems of frequency estimation, forecasting and imputation of missing values can be solved. The idea of SSA came from different scientific communities, such as researchers in the field of time series analysis (Karhunen–Loev decomposition), signal processing (low-rank approximation and frequency estimation) and multidimensional data analysis (PCA). In addition, depending on the field of application, different points of view on the same algorithms, the choice of parameters and the methodology in general are considered. Thus, the purpose of the article [14] «Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing» is to describe and compare different points of view on SSA and its modifications and extensions in order to give people from different scientific communities the opportunity to learn about potentially new aspects of the method.

In [15] «Performance of TCP with Multiple Priority Classes», two classes of user priorities are considered. In case of temporary congestion of the access channel, high priority clients have strict priority over low priority clients, and therefore, low priority clients can use only the bandwidth remaining from the high priority client according to the Diffserv concept. The authors examined empirical data and noted that they have the property of long-term dependence and self-similarity. At the same time, the settings of the parameters of the Kendall M/G/C-PS model for servicing one priority class are considered. The dependencies of the file size on the average download time are constructed for the following distributions: deterministic, exponential and Pareto.

Based on the above review of articles, it becomes clear that the use of Internet services has increased dramatically, and Internet applications have evolved from standard document search functions to advanced multimedia services. There has been a migration from providing services with maximum efficiency to a technology capable of supporting QoS (Quality of Service) guarantees for end users. At the same time, P/P/1 models are increasingly being used (the time of receipt of service requests is distributed according to the Pareto law, and not the M/M/1 model (the time of receipt of service requests is distributed exponentially).

Note that the possibilities of physical modeling in the analysis of networks are very limited. It allows you to solve individual tasks when setting a small number of combinations of the system parameters under study. Indeed, with full–scale modeling of a computer network, it is almost impossible to check its operation for variants using a large number of communication devices - routers, switches, etc. But if the topology is transformed in a certain way, then the possibilities of physical modeling may increase. However, the removal of statistical characteristics from various points of the network is extremely difficult. If it is still possible to programmatically remove interface usage statistics on a workstation, then this may become almost impossible on a switch or in an optical communication line. Again, the analysis of the obtained results is complicated by the complexity of calculations.

Therefore, when analyzing and optimizing networks, in many cases it is preferable to use mathematical modeling. A special class of mathematical models are simulation models. Such models are a computer program that chronologically reproduces the events taking place in a real system step by step. In relation to computer networks, their simulation models reproduce the processes of message generation by applications, divide messages into packets and frames of certain protocols, identify delays associated with the processing of messages, packets and frames inside the operating system, and also allow analyzing the process of a computer gaining access to a shared network environment, etc.

Moreover, when simulating a network, it is not necessary to purchase expensive equipment, since its operation is simulated by programs that accurately reproduce all the main features and parameters of such equipment or the network.

The result of the simulation model is statistical data collected during the model run on the most important characteristics of the network: reaction times and delays, network resource utilization coefficients, packet loss probabilities, etc. All this allows us to look at networks not as a «black box», but from the point of view of information processes occurring in them.

This article uses the software simulation system OPNET Modeler, which is focused on communication networks and allows you to build models without programming. Such software systems themselves generate a network model based on initial data about its topology and protocols used, about the intensity of request flows between network computers, the length of communication lines, about the types of equipment and applications used.

The OPNET Modeler program offers users a graphical environment for creating, executing and analyzing event modeling of communication networks. This convenient software can be used to solve many tasks, for example, to check communication protocols, analyze protocol interactions, optimize and plan the network. It is also possible to use the program to verify the correctness of analytical models and describe protocols. Figure 1 shows the project palette and the workspace where the network will be created, and the palette where the elements that can be used in the network project are displayed.

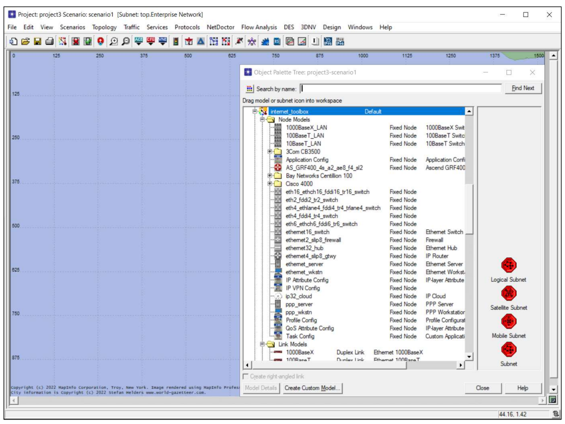

Figure 1: Project palette and workspace

After the end of the simulation, the user gets the following network performance characteristics at his disposal:

- predicted delays between the end and intermediate nodes of the network, channel throughput, utilization rates of segments, buffers and processors;
- traffic peaks and troughs as a function of time, not as averaged values;
- sources of network delays and bottlenecks.

The OPNET Modeler system operates with three types of nodes – processor nodes, router nodes and switches. Nodes can be connected via ports to any type of communication channels, from local area network channels to satellite communication lines. Nodes and channels can be characterized by an average time to failure and an average recovery time to simulate network reliability [16].

It simulates not only the interaction of computers in the network, but also the process of dividing the processor of each computer between its applications. The operation of the application is modeled using several types of commands, including commands for processing data, sending and reading messages, reading and writing data to a file, establishing sessions and suspending the program until messages are received. A so-called command repertoire is set for each application.

Communication channels are modeled by specifying their type, as well as two parameters – bandwidth and the introduced propagation delay. The unit of data transmitted over the channel is a frame. Packets are segmented into frames when transmitted over channels. Each channel is characterized by minimum and maximum frame size, overhead per frame and the intensity of errors in frames. Communication with the global network is simulated using an access channel that has a certain propagation delay and bandwidth.

The workload is created by traffic sources. Each node can be connected to several traffic sources of different types.

Application sources generate applications that are executed by nodes such as processors or routers. The node executes command after command, simulating the operation of applications on the network. Sources can generate complex non-standard applications, as well as simple ones that are mainly engaged in sending and receiving messages over the network.

Call sources generate requests to establish connections in circuit-switched networks (networks with switched virtual connections, ISDN, POTS).

The planned load sources generate data using a time-dependent schedule. At the same time, the source generates data periodically, using a certain law of the distribution of the time interval between data portions. It is possible to model the dependence of the intensity of data generation on the time of day.

Client-server sources allow you to specify not the traffic between clients and the server, but the applications that generate this traffic. These applications work in a client-server model, and a source of this type allows you to simulate the computing load of a computer operating as a server, that is, to take into account the time of computing operations, operations related to disk access, I/O subsystem, etc.

2. INVESTIGATION OF THE STRUCTURE OF TRAFFIC PASSING THROUGH THE ACCESS NETWORK

Figure 2 illustrates the developed network model between LAN and ISP on the desktop of the OPNET Modeler application package.

Journal of Theoretical and Applied Information Technology
15th November 2022. Vol.100. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

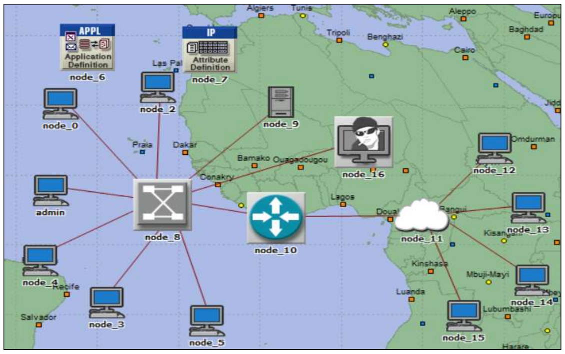

Figure 2: Diagram of the built network between LAN and ISP on the desktop of the OPNET Modeler application package

In recent years, work has been underway to transfer the access network with metal cores to optical access using optical fibers that enable high-speed data transmission. Also, the study of the structure and other parameters of network traffic on the site of the subscriber access network between LAN and ISP is promising. During the experiment, 7655 packets were tracked in just five hours. Figure 3 shows a graph of the intensity of packet receipt. Visually it can be seen that the distribution of the number of packets has an uneven intensity. There are places with destroyed areas where there are few incoming packages.

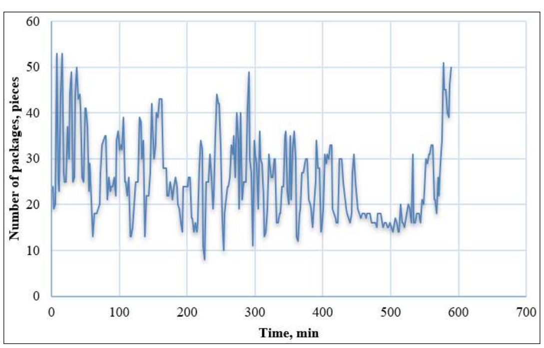

Figure 3: The intensity of packet arrival on the high-speed backbone of the multiservice network

A common model for analyzing access network processes is a time series-based model. A time series, or a series of observations, is a sequence of measurements of a certain quantity performed at arbitrary points in time [17].

In general, a time series is a one-parameter family of random variables $y_t = yt_i, i = 1,\ldots,n$, whose distribution law and numerical characteristics may depend on time *t*. The numerical values of the indicator are called time series levels.

General statistical model of a time series:

$$y_t = f(x_t, a) + \varepsilon_t \tag{1}$$

here y_t – a series of observations that is considered as the sum of the components $f(x_t, a)$;

a – parameter;

ε – white noise.

Next, we examine the obtained time series in the AtteStat add-in in Excel to test the assumption of Gaussian (normality), which means that for any set $t_1,\ldots, t_n$ random variables $X_{t1},\ldots, X_{tn}$ have a normal distribution.

The AtteStat data analysis program offers a set of classical and modern statistical methods. It is made in the form of an add-on to the popular Microsoft Excel spreadsheets running the Microsoft Windows operating system.

The AtteStat includes the following modules:

- cross tabulation;
- data transformations;
- time series analysis and forecasting and others.

The results of the output data of the AtteStat on the statistical evaluation of the measured series for normality are shown in Figure 4.

	A	B	C	D	E	F	G	H	I	J
8	15,7179	53			Modified Kolmogorov criterion					
9	17,6826	27			0,09683	1,3E-05	The hypothesis of normality is rejected			
10	19,6473	25			Modified Smirnov criterion					
11	21,6121	25			0,09683	1E-05	The hypothesis of normality is rejected			
12	23,5768	37			Kramer-Mises criterion					
13	25,5415	30			0,70349	4,1E-07	The hypothesis of normality is rejected			
14	27,5063	44			Anderson-Darling criterion					
15	29,471	49			4,59708	3,8E-07	The hypothesis of normality is rejected			
16	31,4357	25			Shapiro-Francia criterion					
17	33,4005	26			0,94734	1,1E-07	The hypothesis of normality is rejected			
18	35,3652	44			Criterion of the asymmetry coefficient					
19	37,3299	50			0,79944	6,7E-09	The hypothesis of normality is rejected			
20	39,2947	43			Jarque-Bera criterion					
21	41,2594	44			32,1659	1E-07	The hypothesis of normality is rejected			
22	43,2241	41			D'Agostino criterion					
23	45,1889	26			0,27729	0,00274	The hypothesis of normality is rejected			
24	47,1536	25			Fisher's chi-square criterion					
25	49,1183	41			Classes	9				
26	51,0831	41			10,7222	11	19,0485			
27	53,0478	37			16,1667	74	41,4469			
28	55,0125	23			21,6111	66	63,836			
29	56,9773	29			27,0556	55	69,5953			
30	58,942	21			32,5	52	53,7077			
31	60,9067	13			37,9444	16	29,3382			
32	62,8715	18			43,3889	19	11,3442			
33	64,8362	18			48,8333	4	3,10494			
34	66,8009	18			54,2778	3	0,60155			
35	68,7657	19			Criterion statistics, p-value					
36	70,7304	20			53,2085	1,1E-09	The hypothesis of normality is rejected			

Figure 4: Results of processing of the AtteStat of the initial time series by checking it for the normality of the distribution

Based on the above figure, it can be seen that as a result of processing the original time series in Attestat, the modified Kolmogorov criterion, the

Journal of Theoretical and Applied Information Technology
15th November 2022. Vol.100. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

modified Smirnov criterion, and the Kramer-Mises criterion rejected the hypothesis that the distribution of the time series refers to the general population of the normal distribution.

3. DEVELOPMENT AND ANALYSIS OF THE SCATTERING MODEL OF THE STUDIED TIME SERIES

Figure 5 shows a time series model with relationships between the current and previous observations. The previous observation of the time series is called lag. This scattering diagram demonstrates the relationship between observation and lag, that is, it demonstrates the spread of points in the form of an elongated cloud, limited mainly by one quadrant [18].

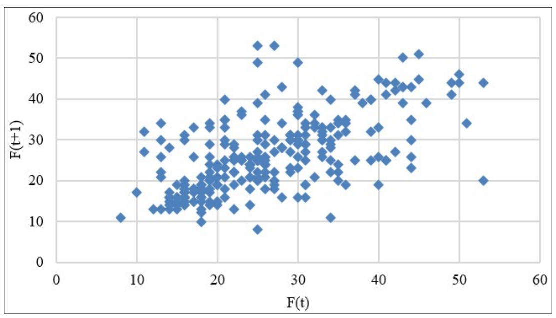

Figure 5: Graph of the dependence of the initial series $f_{i+1} = f(f_i)$

In this scattering diagram, the slope (direction of connection) and width (strength of connection) of the imaginary ellipse are of interest, which reflects the closeness of the linear relationship between the two measured correlation coefficients. This suggests that there is a high positive correlation here. Based on the above, it can be assumed that the series under study is multicomponent.

Moreover, it can be assumed that the series under study is non-stationary. This assumption about the presence of a trend in the series requires the exclusion of a non-stationary component. To exclude a trend, they usually switch to a time series composed of the differences of successive values of the series, that is, to a series of increments of the original time series. The transition to increments makes the time series more stationary.

To transform the initial non-stationary series according to the mathematical expectation of the time series into a stationary one, differentiation is carried out - that is, taking the finite differences of the values of the series (with low frequencies dominating) according to the formula:

$$Y(t) = X(t+1) - X(t) \quad (2)$$

Next, we construct a scattering diagram of the increments of the series under study (Figure 6).

The resulting diagram of a series of increments differs from the diagram of the original series in that the distribution points are distributed in all four quadrants with a relatively high density of points near zero (around zero mathematical expectation). The relative independence of the values of neighboring points is obtained. This diagram of a series of increments confirms the absence of a strong linear correlation between the levels of a series of increments, and if they are, they are relatively weak.

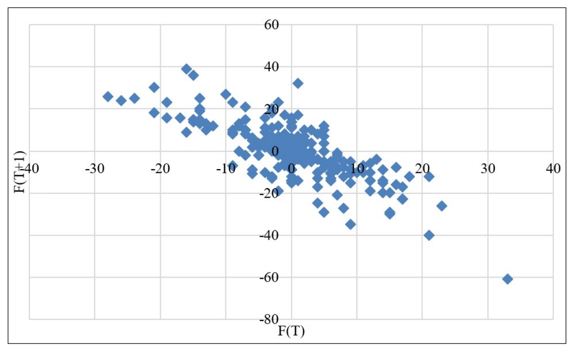

Figure 6: Increment dispersion diagram

Figure 7 shows a scattering diagram of increments that are mixed in random order. At the same time, the resulting scattering diagram differs from the scattering diagram of the increments of the original series in that the points are more evenly distributed in all four quadrants compared to the scattering diagram of increments, where there are more points in the second and fourth quadrants.

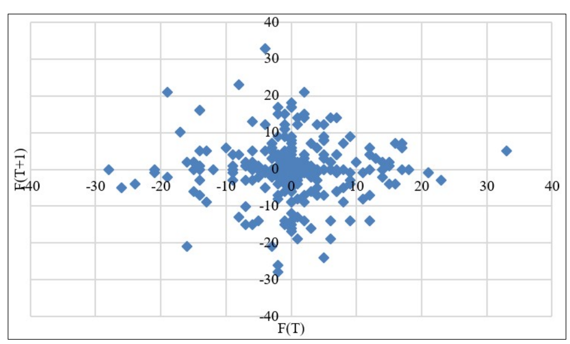

Figure 7. Graph of the dependence of increments mixed in random order $f_{i+1} = f(f_i)$

4. DEVELOPMENT AND ANALYSIS OF A TIME SERIES AUTOCORRELATION FUNCTION MODEL

Correlation shows how strong the dependence of random variables is and how well it is possible to predict the possible values of one indicator, knowing the value of the other [19]. If the time series contains a trend and a cyclical component, then the values of the next level of the series depend on the previous ones. This dependence is the autocorrelation of the series levels. Its quantitative value is measured using the correlation index between the levels of the original time series and the levels of this series shifted by several steps in time.

The correlation coefficient between the series y_t and y_{t-1}.

$$r_{xy} = \frac{\overline{\sum (x_j - \bar{x}) \cdot (y_j - \bar{y})}}{\sqrt{\sum (x_j - \bar{x})^2 \cdot \sum (y_j - \bar{y})^2}} \quad (3)$$

If $x_j = y_{t-1}$, $y_j = y_{t-1}$, then we get:

$$r_1 = \frac{\overline{\sum_{t=2}^{n} (y_{t-1} - \bar{y}_{t-1}) \cdot (y_t - \bar{y}_t)}}{\sqrt{\sum_{t=2}^{n} (y_{t-1} - \bar{y}_{t-1})^2 \cdot \sum_{t=2}^{n} (y_t - \bar{y}_t)^2}} \quad (4)$$

The autocorrelation coefficients of the second and higher orders are determined in the same way. For example, for the second-order autocorrelation coefficient:

$$r_2 = \frac{\overline{\sum_{t=3}^{n} (y_{t-2} - \bar{y}_{t-2}) \cdot (y_t - \bar{y}_t)}}{\sqrt{\sum_{t=2}^{n} (y_{t-2} - \bar{y}_{t-2})^2 \cdot \sum_{t=3}^{n} (y_t - \bar{y}_t)^2}} \quad (5)$$

An autocorrelation function is a sequence containing autocorrelation coefficients of levels of the first, second and other orders. A correlogram is a graphical dependence of the values of autocorrelation coefficients of different orders.

Figure 8 shows a corellogram of the studied time series in AtteStat, on which abbreviations are applied: row 1 is a corellogram; row 2 and row 3 are confidence intervals. The correlogram represents the dependence of the autocorrelation of the sample on the lag.

Using the correlogram, you can determine the highest lag. In our case, the numerical value of the highest level of autocorrelation is 0.6362 with a lag of 1, therefore, the original series contains only a trend. Further, the levels of autocorrelation coefficients are in the corridor of confidence intervals.

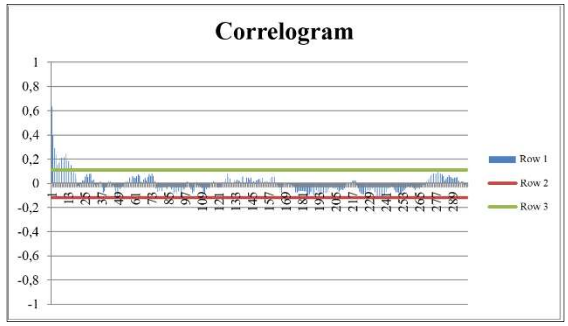

Figure 8: Time series correlogram

5. DECOMPOSITION OF A TIME SERIES BY THE SSA METHOD

The properties of time series are usually analyzed in the time domain, although the same information can be efficiently obtained in the analysis in the frequency domain, that is, using spectral analysis. Any random process has both a time domain and a representation in the frequency domain. To identify some properties, a study in the time domain is suitable and this is more appropriate. For other properties, the frequency domain may be more suitable.

Spectral analysis decomposes the original series into an infinite sum of periodic functions, where each has a different frequency. Spectral methods usually encompass a class of algorithms that represent matrices using linear algebraic methods involving the eigenvalues of the matrix vectors.

Recently, a method considered as part of the field of computational intelligence, such as SSA (Singular Spectrum Analysis) or, in another way, the «Caterpillar» method, has been added to the classical methods of spectral analysis [20]. It is one of the modern tools for analyzing the structural components of a time series. The classical Karhunen-Loeve theorem for the spectral decomposition of time series can be attributed to the sources of the origin of SSA.

The SSA time series decomposition method splits a time series into a set of summable components that are grouped and interpreted as trend, periodicity, and noise. This method emphasizes the separability of the basic components and can easily separate the periodicities that occur at

Journal of Theoretical and Applied Information Technology
15th November 2022. Vol.100. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

different time scales, even in very noisy time series data.

The SSA 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 stationary. It also has only one configurable (and easily interpreted) parameter. The basic algorithm is implemented on the basis of simple linear algebra using a discrete convolution operator.

The SSA method transforms a one-dimensional series of dynamics into a multidimensional one, using a one-parameter shift. Using the properties of the obtained trajectory matrix, the SSA method applies singular value decomposition and reconstructs (approximates) the processed data with a choice of principal components [21]. The SSA method is multifunctional and performs Principal Component Analysis (PCA), cluster analysis and graphical analysis of individual vectors.

This method considers a valid time series X = $X_N = (x_1, ..., x_N)$ of length N. It is assumed that N>2 and the series X is a non-zero series, that is, there is at least one i in the series, which is $f(i) \neq 0$. The main parameter L (1<L<N) is some integer called the window length (should not exceed half the length of the row N), and the embedding procedure creates K=N-L+1 embedding vectors having dimension L [22]:

$$X_i = \left(f_{i-1},...,f_{i+L-2}\right)^T, 1 \le i \le K \qquad (6)$$

Next, the one-dimensional series is transformed into an L-trajectory matrix of the series F, which consists of embedding vectors as columns:

$$X = \left[X_1 : ... : X_K\right] \qquad (7)$$

Trajectory matrix:

$$X = \left(x_{ij}\right)_{i,j=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & . & . & . & f_{K-1} \\ f_1 & f_2 & f_3 & . & . & . & f_K \\ f_2 & f_3 & f_4 & . & . & . & f_{K+1} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ f_{L-1} & f_L & f_{L+1} & . & . & . & f_{N-1} \end{pmatrix} \qquad (8)$$

Since $x_{ij} = f_{i+j-2}$ and the matrix X have the same elements on the diagonals $i + j = const$, this type of matrix is known as the Hankel matrix.

The parameter K represents the number of columns in the trajectory matrix. In this case, columns X are referred to as L-lagging vectors, and rows are referred to as K–lagging vectors. In this case, the analysis of the main components of the covariance matrix is carried out. As a result, the eigenvalues of the matrix are calculated, which are the sample variances of the corresponding principal components, and the square roots of them are the sample standard deviations and the eigenvectors of the matrix. That is, its singular decomposition is performed.

The paper [23] describes the importance of the spectrum of the obtained sample covariance matrix in the modern method of nonlinear spectral analysis of SSA, that is, sets of its eigenvalues. In this case, the presence of a connection between the spectrum of the covariance matrix of the time series and its spectral power is described. It is also noted that the spectrum of eigenvalues of the covariance matrix can be used to estimate the number of signals. The eigenvalues are a characteristic of the contribution of the matrix to the expansion of the series.

This is followed by the procedure of grouping the belonging of the corresponding main components to the trend, or to harmonics, or to noise. To separate the noise, you can use several remarks: the irregular behavior of singular vectors may indicate that they belong to a set generated by a noise component; this may also be evidenced by a slow, almost without jumps, decrease in eigenvalues from a certain number.

In [24-25] it is described that the eigenvalues in the sum give *m* and by the formula it is possible to determine the percentages of the variance of the corresponding principal components:

$$c_i = \frac{\lambda_i}{m} \cdot 100\% \qquad (9)$$

That is, $c_i = \sum_{k=1}^{i} c_k$ is the accumulated percentages that reflect the share of the main component *i*.

To determine *m*, a threshold value is set with c* the accumulated percentage. Next, the decomposition of the series is carried out at some sufficiently large m_0 and takes for *m* such a number of principal components that $c_m \ge c^*$.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

According to the graphs of the dependence of eigenvectors on time, it is determined which main components are attributed to the three components of the time series:

- trend (low-frequency fluctuations, slowly changing or components of a smoothed group);
- periodic components;
- noise (high frequency components).

The original time series is nested in AtteStat for decomposition into additive components (at 5, 10 and 15), shown in Figures 9, 10 and 11.

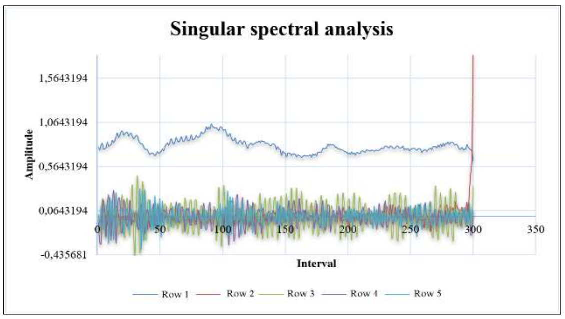

Figure 9: Decomposition into 5 components

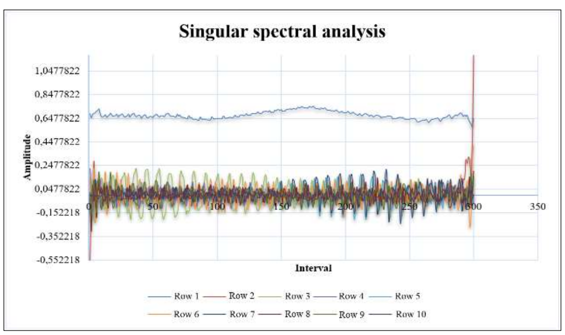

Figure 10: Decomposition into 10 components

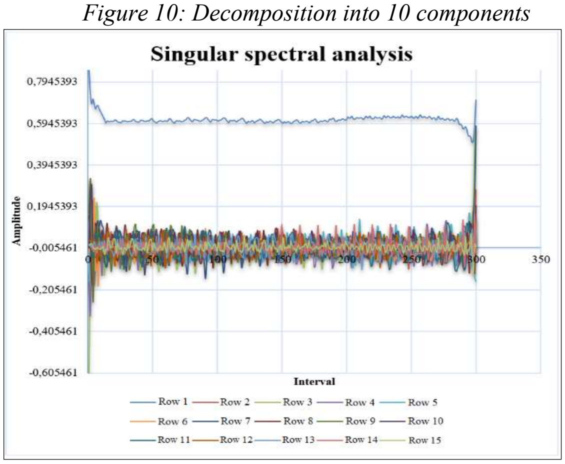

Figure 11: Decomposition into 15 components

6. CHECKING THE NEED FOR SECURITY AND NETWORK ARCHITECTURE

Next, consider the need to ensure network security. Recently, the number of LAN is becoming more and more, and information with a large volume passes through them, therefore, performing the procedure for investigating such a process based on modeling a network experiment is becoming one of the important tasks. Next, consider this problem. To do this, in the built network, we will conduct an experiment on the hacker's impact on the server of this network. To conduct such an experiment, it is necessary to study the Application Capture Manager (ACM) module and the WireShark program. The ACM program is necessary to organize a passive attack on the network server. To do this, select Application Capture Manager (Figure 12).

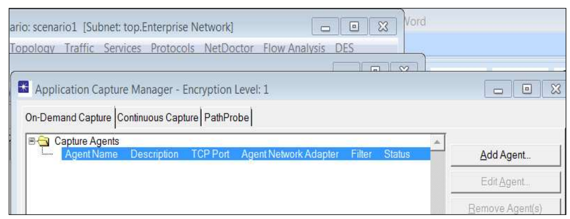

Figure 12: The window of the beginning of the WireShark impact on the network server

Next, by clicking on the Add Agent that allocates the IP address of the computer on which we work (192.168.1.106), we prepare to capture packets transmitted over a network built on OPNET Modeler, and launch WireShark, which is shown in Figure 13. After launching it, we get a window of the graph shown in Figure 14.

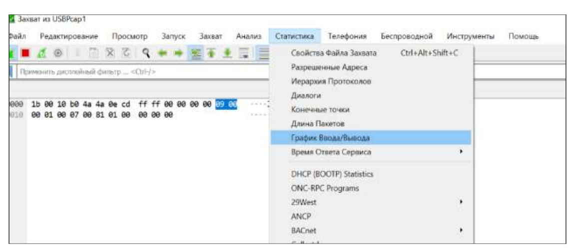

Figure 13: WireShark program window in packet capture mode from the network

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

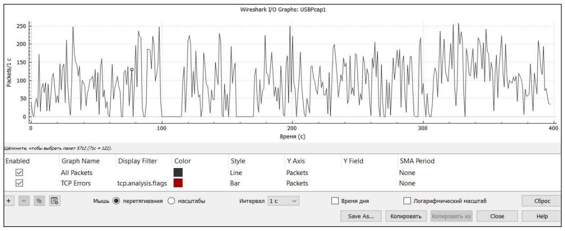

Figure 14: Time-Dependent Packets Captured By Wireshark

Sometimes there are problems checking the correctness of the network construction during its modeling. In this case, the OPNET Modeler module program called NetDoctor is used. Starting the program and getting the results are shown in Figure 15. The figure shows that the network is built correctly as there are no indications of errors.

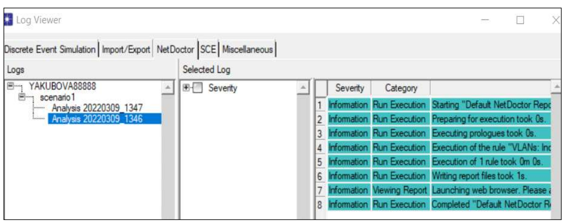

Figure 15: The Window Of The Validation Result Of Checking The Built Network

CONCLUSION

1. A network model between LAN and ISP has been developed on the desktop of the OPNET Modeler application package;
2. A graph of the intensity of network packet arrival between LAN and ISP is constructed, uneven packet arrival is visually determined;
3. According to the results of processing the original time series using the AtteStat add-in in Microsoft Excel, the normality of the distribution is not justified;
4. Various scattering diagrams are constructed, which show the presence of a positive correlation of the levels of the time series;
5. A correlogram of the time series under study is constructed. In our case, the numerical value of the highest autocorrelation level is 0.6362 with a lag of 1, therefore, the original series contains only a trend;
6. Based on the decomposition, it is established that it has both low-frequency (trend), harmonic components and high-frequency components (noise);
7. The results of a passive attack on a simulated network based on Wire Shark show that it is necessary to ensure the security of the gap between LAN and ISP;
8. Using the Net Doctor module in the OPNET Modeler program to check the architecture of the developed network, shows the reliability of the built network.

ACKNOWLEDGMENT

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).

REFERENCES

[1] Loskutov A.Y., Mikhailov A.S., Fundamentals of the theory of complex systems. – M. – Izhevsk: Institute of Computer Research, 2007. – 620.

[2] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," in IEEE/ACM Transactions on Networking, vol. 2, no. 1, pp. 1-15, Feb. 1994, doi: 10.1109/90.282603.

[3] Serikov T., Zhetpisbayeva A., Akhmediyarova A., Mirzakulova S., Kismanova A., Tolegenova, A., Wjcik, W. (2021). City Backbone Network Traffic Forecasting. International Journal of Electronics and Telecommunications, 67 (3), 319–324. doi: https://doi.org/ 10.24425/ijet.2021.135983

[4] Tsitovich, I.I., Stable traffic models of multiservice networks. – St. Petersburg. May 17 – May 19, 2005, pp. 271-273.

[5] Kyeong Soo Kim. The Effect of ISP Traffic Shaping on User-Perceived Performances in Broadband Shared Access Networks. – Preprint submitted to Computer Networks. – 06.2014. – 39 p.

[6] Willinger W., Taqqu M. S. and Erramilli A. “A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks,” Stochastic Networks: Theory and Applications. In Royal Statistical Society Lecture Notes Series. – Oxford University Press, 1996, vol. 4, pp.339 –366.

[7] Sotiris Lenas, Vassilis Tsaoussidis, Srikanth Sundaresan, Arjuna Sathiaseelan and Jon

Crowcroft. On the Impact of Guest Traffic in Open-Access Domestic Broadband Sharing Schemes. –June 2018. https://www.researchgate.net/publication/325643009.

[8] S. Akpah, D. Mireku-Gyimah and F. L. Aryeh. Improving the Performance of a Network by Managing the Bandwidth Utilisation with squidGuard: A Case Study. – Vol. 1, No. 2, March, 2017. 9–17 p. https://www.researchgate.net/publication/316461034.

[9] Serikov T., Zhetpisbayeva A., Mirzakulova S., Zhetpisbayev K., Ibrayeva Z., Soboleva L. et. al. (2021). Application of the NARX neural network for predicting a one-dimensional time series. Eastern-European Journal of Enterprise Technologies, 5 (4 (113)), 12–19. doi: https://doi.org/10.15587/1729-4061.2021.242442

[10] N. Golyandina, E. Osipov, The "Caterpillar"-SSA method for analysis of time series with missing values, Journal of Statistical Planning and Inference, Volume 137, Issue 8, 2007, Pages 2642-2653, ISSN 0378-3758, https://doi.org/10.1016/j.jspi.2006.05.014.

[11] Claudio M. Rocco S, Singular spectrum analysis and forecasting of failure time series, Reliability Engineering & System Safety, Volume 114, 2013, Pages 126-136, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2013.01.007.

[12] M.C.R. Leles, J.P.H. Sansão, L.A. Mozelli, H.N. Guimarães, Improving reconstruction of time-series based in Singular Spectrum Analysis: A segmentation approach, Digital Signal Processing, Volume 77, 2018, Pages 63-76, ISSN 1051-2004, https://doi.org/10.1016/j.dsp.2017.10.025.

[13] P.C. Rodrigues,Pétala G. S. E. Tuy, Rahim Mahmoudvand, Randomized singular spectrum analysis for long time series, Journal of Statistical Computation and Simulation Volume 88, 2018 - Issue 10: 2nd Latin American Conference on Statistical Computing, pp. 1921-1935.

[14] Yakubova, M., Serikov, T. «Development and imitating modeling in the developed network consisting of several knots removed among themselves on NetCracker 4.1» International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices, EDMthis link is disabled, 2016, 2016-August, pp. 210-213

[15] Vranken R., van der Mei R.D., Kooij, R.E., van den Berg J.L. Performance of TCP with Multiple Priority Classes. – 08 July 2014. – p.10.

[16] Yakubova M.Z., Konshin S.V., Berdibaev R.Sh., Manankova O.A., Mukasheva A., Fundamentals of information security, Textbook, Almaty: AUES, 2021, in press.

[17] Vafeiadis, Thanasis & Papanikolaou, Alexandros & Ilioudis, Christos & Charchalakis, Stefanos. (2012). Real-time network data analysis using time series models. Simulation Modelling Practice and Theory. 29. 173-180. 10.1016/j.simpat.2012.07.002.

[18] Mareev V.S. Time series visualization tools in space research. Volume 1. – Research of the science city, 2017. – № 4 (22).

[19] Beate Meffert, Olaf Hochmuth. Signal processing tools – basics, application examples and tasks. The authors of the translation are G.Bezel, V.Baerbach, 2018. – 320 p.

[20] Golyandina, Nina & Zhigljavsky, Anatoly. (2013). Singular Spectrum Analysis for Time Series. 10.1007/978-3-642-34913-3.

[21] D.S. Broomhead, Gregory P. King, Extracting qualitative dynamics from experimental data, Physica D: Nonlinear Phenomena, Volume 20, Issues 2–3, 1986, Pages 217-236, ISSN 0167-2789, https://doi.org/10.1016/0167-2789(86)90031-X.

[22] Matveev, M.G., Semenov M.G., Matveev M.E., Tolokonnikov P.V., Soloviev A.Y. Application of the method of singular spectral analysis to identify signals of electrical activity of the brain: – VSU Bulletin, Series: System Analysis and Information Technology. – 2012. №2.

[23] Zhigljavsky, Anatoly. (2010). Singular Spectrum Analysis for time series: Introduction to this special issue. Statistics and Its Interface. 3. 10.4310/SII.2010.v3.n3.a1.

[24] Fatkieva R. R., Levonevsky D. K. Detection of computer attacks by singular spectral decomposition. – SPIIRAN, – 2013, release 25, p.135–147.

[25] M.Z.Yakubova, T.G.Serikov, A.D.Mekhtiev, A.D.Alkina, Y.G.Neshina. Development and imitating modeling in the developed network consisting of several knots removed among themselves on NetCracker 4.1. IEEE 2016 Dynamics of Systems, Mechanisms and Machines (Dynamics), Omsk, Russia, 2017, doi:10.1109/Dynamics.2016.7819108.