Self-similar Traffic Research Experiment*

Tatiana Tatarnikova[1][0000-0002-6419-0072] and Ekaterina Poymanova[2][0000-0001-9318-6454]

[1]Russian State Hydrometeorological University, ul. Voronezhskaya, 79, 192007 St. Petersburg, Russia

[2]Saint-Petersburg State University of Aerospace Instrumentation, Bolshaya Morskaya str. 67, 190000 St. Petersburg, Russia

tm-tatarn@yandex.ru, e.d.poymanova@gmail.ru,

Abstract. The article discusses the procedure of researching network traffic with a self-similar structure. It is a mixture of voice traffic, data and multimedia. Self-similar traffic covers very different time scales. Considered the characteristic properties of self-similar traffic both in geometric and statistical terms. Self-similar traffic is represented by the Pareto distribution. Received the characteristics of queuing systems using simulation. Results presented in this article make it possible to substantiate the prospective requirements for network node equipment. Obtained the evaluations quality of service for self-similar traffic in terms of buffer length and time delay in the network node. An experiment proving the operability of the ARIMA(p,d,q) model in the study of self-similar traffic was conducted

Keywords: Network Traffic, Self-similarity, Long-time Dependence, Distribution with Heavy «Tails», Slow-damping Dispersion, Hurst Index, Simulation, Buffer Storage Volume, forecasting, model of autoregressive integrated moving average

1 Introduction

Packet network traffic is the integration of voice, data and multimedia. Such traffic covers very different time scales - from microseconds till seconds and even minutes [1]. Mixture of data streams different on content and properties generates to so named self-similar traffic.

Self-similar traffic at any time scale is longtime dependence – availability of pulsations - activity periods, divided to less active periods [2].

In classical models of information streams, such as Poisson stream, Erlang, gamma-distribution and other pulsations are strongly smoothed on large time scales, which makes the property of long-time dependence is missed. As a result, the classical models are not allowed to appreciate the volumes of calculation resources of systems when servicing pulsating traffic [3].

According to this fact, the present day task is simulation of network nodes with self-similar traffic at the entrance. The aim of this modification is to reconsider characteristics received in classical models of information streams.

In addition, the self-similarity allows one to compose forecast models in different time scales, which allows a long-term forecast of incoming traffic.

2 Properties of self-similar traffic

Self-similarity can be characterized as geometrically and so statistically.

Self-similarity as geometric concept underlines that fact that process keeps the structure on different time scales.

In fig. 1-4 can be seen daily data of different traffic types from 08.20.2018. The data are given by mobile operator MTS in St. Petersburg. They demonstrate the persistence traffic structure over time.

Self-similarity as a statistical understanding is characterized by such properties:

- slowly damped dispersion;
- a long-time dependence;
- availability of distribution with heavy "tails" of time intervals between two consistent events [2, 3].

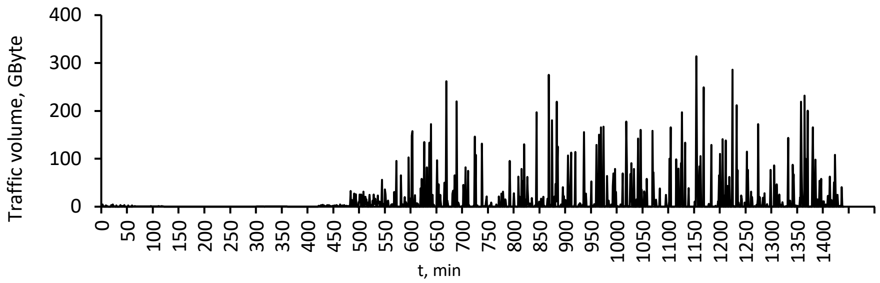

Fig. 1. 2G traffic during 1440 min

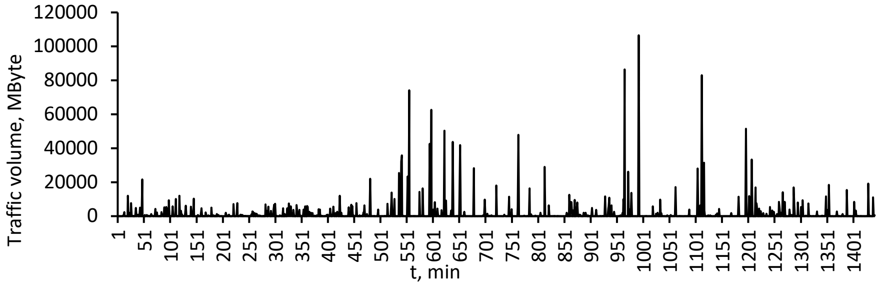

Fig. 2. CS Voice traffic during 1440 min

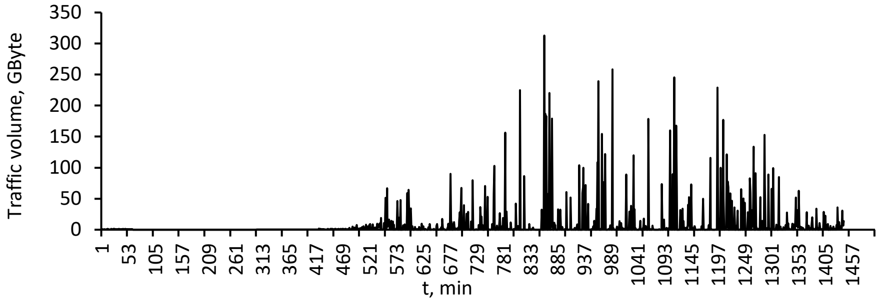

Fig. 3. HSDPA traffic during 1440 min

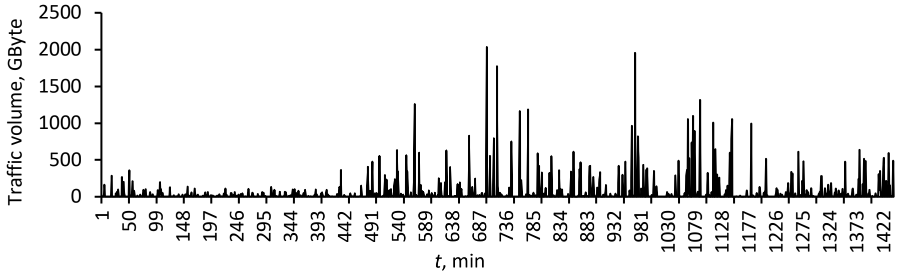

Fig. 4. Mix initial traffic during 1440 min

The property slowly damped dispersion is that the dispersion of the sampling average dampens slower than quantity reciprocal of sampling size

$$D\left(X^{(n)}(t)\right) = \sigma^2 n^{2H-2},\ n \to \infty, \tag{1}$$

where σ^2 – dispersion of process $X(t)$;
n – the sampling size;
H – Hurst index.

For meaning $H > 0.5$ the direction of process dynamics most likely will not change;
for $H < 0.5$ prognoses that process will change direction;
for $H = 0.5$, we have uncertainty – Brownian moving.

Availability of a long-time dependence shows that the self-similar process has hyperbolically damped correlation function

$$R(k) \cong k^{(2H-2)} A(k),\ \forall k \geq 1,\ k \to \infty, \tag{2}$$

$A(k)$ – changing function on infinity, for which

$$\lim_{k\to\infty} \frac{A(kx)}{A(k)} = 1 \ \text{ for all x> 0} \tag{3}$$

Property of availability of distribution with heavy "tail" treat to random variable X

$$P(X > x) \sim cx^{-\alpha},\ x \to \infty, \tag{4}$$

where 0 <α <2 – parameter of distribution form, the smaller meaning α, the heavier the "tail" of the distribution;
c – some positive constant.

For α≤2, distribution has endless dispersion; for α≤1 distribution has an endless mean.

Adequate description of self-similar traffic is given by probability distributions with heavy "tails", in particular, by the Pareto distribution [3, 4].

Pareto distribution is done by following function

$$F(t) = 1 - \left(\frac{K}{t}\right)^{\alpha},\ t \geq K, \tag{5}$$

where α – form parameter (simply parameter);
K – a border parameter, specifies minimum meaning of random variable x, plays a role of a scale coefficient.

In further, the Pareto distribution will be denoted as $P(\alpha, K)$ or simply P.

Mathematical expectation $M(x)$ of a random variable x distributed on Pareto is determined by the expression $M(x)=\alpha K/(\alpha-1)$.

3 Analysis of queuing systems *M*|*M*|1 and *P*|*M*|1

As known bufferisation is the main strategy provide resources. Research are concentrated around statistical characteristics of the queues [5,6]. It is obvious, that for self-similar traffic needs buffers by more size, then predicted classical queue analysis [7]. Let us show it by example of estimating the characteristics of queuing systems (QS) type $M|M|1$ and $P|M|1$. With this purpose were worked out simulation models QS in program AnyLogic.
Experiment on simulation model was carried out in follow limits: buffer size $L = \infty$; average processing time of application in service node $T = 0.02$ s; service node load factor varied $\rho \in$ [0,2; one]; $K = 0.01$; $\alpha \in$ [1.1; 2].

Evaluate difference in characteristics QS type $M|M|1$ and $P|M|1$ in form of relations:

- average queues lengths $\overline{L}_P$ and $\overline{L}_E$ with Pareto and exponential distribution of time intervals between two arrivals of traffic packets, accordingly;
- maximum queues lengths $\hat{L}_P$ and $\hat{L}_E$ with Pareto and exponential distribution of time intervals between two arrivals of traffic packets, accordingly;
- average waiting time T_w and stay time T_s packets into QS type $M|M|1$ and $P|M|1$ for different load.

Since different QS type $M|M|1$ and $P|M|1$ are compared, that besides using one and the same generator of random numbers and one the same scale coefficient, needs to use one the same intensity arrivals of traffic packets [8-10].

For QS type $P|M|1$, the intensity may be done through parameter α. If suppose that the mathematical expectation of exponential distribution tends to mathematical expectation of Pareto distribution, so $\frac{1}{\lambda}=\frac{\alpha K}{\alpha-1}$, so then $\alpha=(1-K\lambda)^{-1}$. Thus, if the intensity λ = 25 packets / s, K =0.01, then for QS type $P|M|1$ with the same intensity, parameter α= 1.33, that is not against property of distribution with heavy tails.

The results of statistical characteristics queuing systems type $M|M|1$ and $P|M|1$ are done in Table. 1. The number of experiments are 5104. Average queue length come to whole rounded on the right.

Table 1. The statistical characteristics results of queues

α	$\bar{L}_P/\bar{L}_E$	$\hat{L}_P/\hat{L}_E$
1,1	1	3,70
1,2	1	3,75
1,3	2	3,71
1,4	2	3,86
1,5	3	3,89
1,6	4	3,82
1,7	5	4,00
1,8	10	4,21
1,9	21	9,23
2,0	216	29,93

Analysis of results from the table. 1 allows to make following conclusions.

Difference in required buffer length for QS type $M|M|1$ and $P|M|1$ is obvious, starting for α = 1.1. For $\alpha\in$ [1,1; 1.7], this difference is stable when comparing the maximum queue lengths and this difference rise for α > 1.7. At comparison of average queue lengths, the buffer length for $P|M|1$ begins show the double rise already at α> 1.2.

The rise of queue is influenced not so much by the distribution of time intervals, as by correlation structure of process.

4 Traffic research experiment using the ARIMA(*p*,*d*,*q*) model

A feature of the self-similarity is also the ability to predict the amount of data in the network, which helps prevent data loss associated with a denial of service.

Based on [11–13], it was concluded that the autoregressive integrated moving average model (ARIMA) can be used to predict self-similar traffic.

An experiment was conducted, which task was to show the operability of the ARIMA(p,d,q) model in the study of traffic. For the experiment, there was taken the data of LTE traffic for six days received from MTS (Fig.5).

The purpose of the experiment: to build an ARIMA forecast model based of traffic data for 5 days for the sixth day and compare it with traffic data for this period.

It can be seen from the graph in Figure 5 that there is seasonality with a period of 24 hours, the graph fluctuates around a certain value, and it is also seen that there is a number of peak values. An autocorrelation analysis of the existing time series was carried out and an autoregression function was constructed with the number of steps equal to 48 (Fig.6).

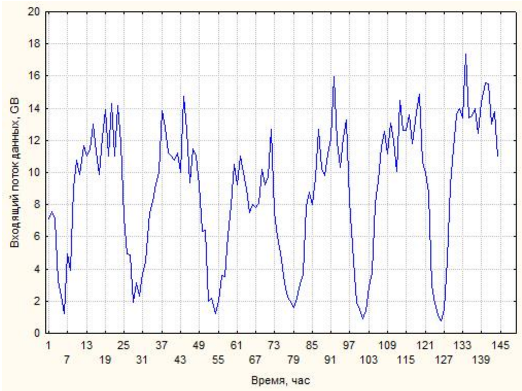

Fig. 5. MTS LTE traffic graph

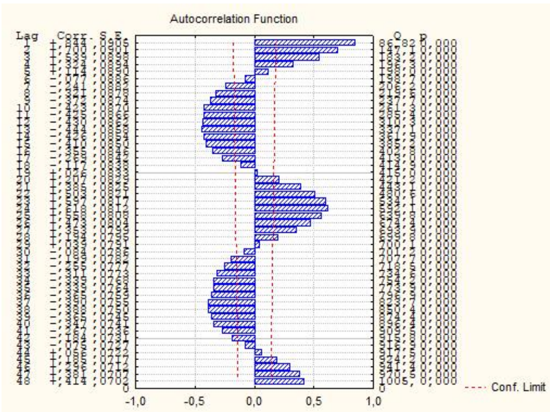

Fig. 6. The initial time series autocorrelation function

On the graph there was a peak of the first value of the time lag. This peak was removed by taking a difference of the order of 1 (*D*-1) and the autocorrelation function for the transformed time series and the partial autocorrelation function were constructed (Fig.7). The parameters *p*, *d*, *q* were determined.

Since the sequential difference operation was applied once, d=1. The parameter p was chosen from the partial autocorrelation model p=1 (the first significant value of the series of functions). The parameter q was chosen similarly from the autocorrelation model and q=1.

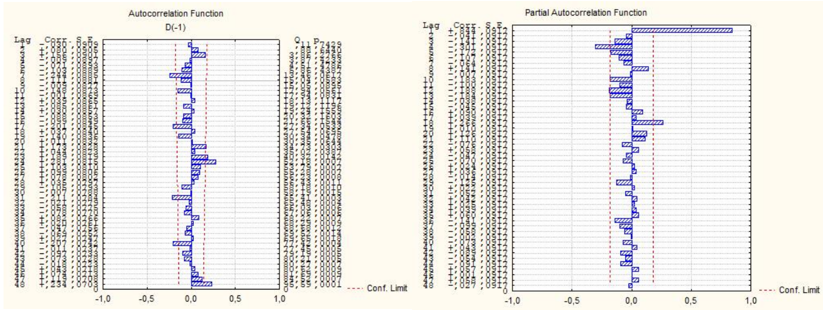

Fig. 7. Autocorrelation and partial autocorrelation functions

As a result, the ARIMA(1,1,1) model was built and a forecast for one period was obtained (Fig. 8).

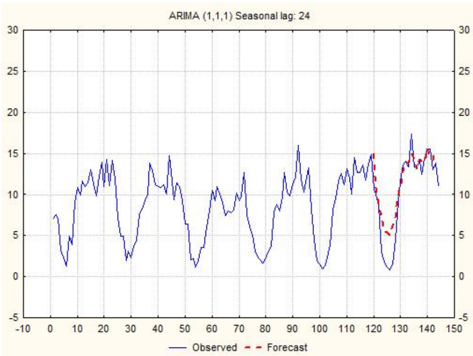

Fig. 8. Data forecast model

Since, due to the properties of self-similarity, the structure of the time series is preserved in different time scales, the constructed model is also suitable for other periods of time (month, year, etc.).

This experiment is an example for constructing an ARIMA model and illustrates the order of forecasting using this model. A feature of the study of the time series using the ARIMA model is a mandatory expert assessment of the obtained autocorrelation models. On the one hand, this is a drawback, because forecasting is a labor-intensive process and it requires the participation of a specialist, on the other hand, expert assessment allows you to more accurately build a forecast model.

5 Conclusion

The rise in the share of multi-service traffic in networks actualizes the problem of meeting customer requirements for the quality of network services provided.

It was conducted the simulation experiment for QS type $M|M|1$ and $P|M|1$. The obtained evaluations quality of service for self-similar traffic in terms of the buffer length and the time delay in the network node.

The results presented in the article make it possible to substantiate the prospective requirements for network node equipment.

An experiment was conducted proving the operability of the ARIMA(p,d,q) model in the study of self-similar traffic.

References

1. Bogatyrev, V.A., Bogatyrev, S. V., Parshutina, Bogatyrev A. V.: Model and Interaction Efficiency of Computer Nodes Based on Transfer Reservation at Multipath Routing. In: 2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), pp. 1-4 (2019). doi: 10.1109/WECONF.2019.8840647
2. Tatarnikova, T.M. Statistical methods for studying network traffic. In Informatsionno-Upravliaiushchie Sistemy, vol. 96, no.5, pp. 35-43 (2018). doi: 10.31799/1684-8853-2018-5-35-43
3. Tanenbaum, A., Wetherall, D.: Computer Networks. 5th ed. Prentice Hall (2010).
4. Kutuzov O. I., Tatarnikova T. M. Model of a self-similar traffic generator and evaluation of buffer storage for classical and fractal queuing system. In Moscow Workshop on Electronic and Networking Technologies, MWENT 2018 - Proceedings 1, pp. 1-3 (2018).
5. Leland, W.E., Taqqu, M.S., Willinger, W. and Wilson, D.V. On The Self-Similar Nature of Ethernet Traffic. In Proc. ACM SIGCOMM'93, pp. 183-193. San-Fransisco (1993).
6. Zwart, A. P.: Queueing Systems with Heavy Tails. Eindhoven University of Technology Publ. (2001).
7. Poymanova, E.D., Tatarnikova, T. M. Models and Methods for Studying Network Traffic. In 2018 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), pp. 1-5 (2018). doi: 10.1109 / WECONF.2018.8604470
8. Bogatyrev, V.A. Protocols for dynamic distribution of requests through a bus with variable logic ring for reception authority transfer. In Automatic Control and Computer Sciences, vol. 33, no. 3, pp. 57-63 (1999).
9. Tatarnikova, T., Kolbanev, M. Statement of a task corporate information networks interface centers structural synthesis. In IEEE EUROCON 2009, pp. 1883-1887. St. Petersburg (2009).
10. Bogatyrev, V. A., Vinokurova M. S. Control and Safety of Operation of Duplicated Computer Systems. In Communications in Computer and Information Science, IET – 2017, vol. 700, pp. 331-342 (2017).
11. Будько М. Ю., Будько М. Б., Гирик А. В. Применение авторегрессионного интегрированного скользящего среднего в алгоритмах управления перегрузками протоколов передачи потоковых данных // Научно-технический вестник информационных технологий, механики и оптики. 2007. №39. С. 319-323
12. Крюков Ю.А., Чернягин Д.В. ARIMA – модель прогнозирования значений трафика // Информационные технологии и вычислительные системы. 2011/2, с. 41-49
13. Соловьев А.Ю. О задаче прогнозирования самоподобных сетевых процессов // III Международная научная конференция «Современные проблемы информатизации в системах моделирования, программирования и телекоммуникациях». URL: http://econf.rae.ru/article/4745