

Comparison of Approximation Methods for the Estimation of Distributions in the Analysis of the G/G/1 Queue

O. Karaulova

The Department of Multi-Service Network and Information Security
Povolzhskiy State University of Telecommunications and Informatics (PSUTI)
Samara, 443010, Russia
E-mail: olya4369@yandex.ru

N. Kireeva

The Department of Multi-Service Network and Information Security
Povolzhskiy State University of Telecommunications and Informatics (PSUTI)
Samara, 443010, Russia
E-mail: zeppelinSN@yandex.ru

L. Chupakhina

The Department of Multi-Service Network and Information Security
Povolzhskiy State University of Telecommunications and Informatics (PSUTI)
Samara, 443010, Russia
E-mail: garip4ik555@mail.ru

A. Gazizulina

Center of Monitoring Science and Education
Peter the Great St. Petersburg Polytechnic University
St. Petersburg, 195251, Russia
Corresponding author: albinagazizulina@gmail.com

(Received November 25, 2018; Accepted May 11, 2019)

Abstract

The analysis of Weibull and Pareto distribution functions in the approximation of the density of a sum of damping functions, PMRQ approximation and approximation of service distribution in the peak mode of the flow. There was found one of the characteristics of the network, the average waiting time of packets in the queue after application of spectral method of the Lindley's integral equation. Finally, PMRQ approximation of the average waiting time in TES+/G/1, MMRP, was compared with the analytic values of the same value in the spectral solution of the Lindley's integral equation obtained by simulation with real traffic. Traffic is captured using the Wireshark protocol analyzer program. The time distributions were obtained in the EasyFit data analysis program, designed for quick statistical data analysis and decision making. Methods of approximation of network traffic allows us to estimate the average packet latency using statistical data analysis, which will improve the quality of service and predict the behavior of traffic.

Keywords- Approximation, PMRQ, PMRS, Damping function, Lindley's equation.

1. Introduction

Analytical study and calculation of traffic parameters, according to its statistical parameters, help to get closer to the definition of network capacity, the optimal buffer size for telecommunications equipment. Both analytical and simulation models can be used to solve problems involving random events. In this case, analytical models are preferred to simulation models for the following reasons: they do not require a large number of tests to get a good estimation of the value of the objective

function for each individual solution; it is possible to obtain an optimal solution (Chupahina, 2013).

The main part of the traditional queue theory was devoted to queues, the time between arrival and service time of which are independent update processes. The assumption that the service time of the packet flow is mutually independent is often justified in practice, and the corresponding assumption about the independence of the packet flow from the intervals between arrivals is often unjustified – therefore, it is necessary to consider self-similarity of these processes while modeling $G/G/1$ queue. Typical examples are packet flows in emerging high-speed integrated telecommunication networks. The relevance and practical importance of studying queues with dependent arrival processes is obvious. It should be concluded that exact solutions (analytical or numerical) are not usually available for the $G/G/1$ queue, since the researchers resort to Monte Carlo simulation or use classical Markov theory and its variations for parametric analysis of the average length of the queue (Heyman, 1976; Blatov et al., 2013).

It is known that correlations in the arriving event flow, assuming a queue per server, significantly affect the average wait time compared to the corresponding event flow coming again to the same server. However, if we use such models like $G/G/1$, for which real or the most possible to the reality approximations of the distribution of time intervals between packets and long-term service for queues with correlated flows (Altiok and Melamed, 2001) are created, it becomes possible to find a solution to the unknown average characteristics of traffic. The study mainly uses distributions with a «heavy» Pareto and Weibull tail to build the process of the incoming packet flow and determine the average waiting time of the packet in the queue, considering its speed, coefficient of variation and other characteristics. A study on the distribution of time between packets described by distributions with a «heavy» Weibull tail, Pareto function showed that it can be used for analytical determination of traffic parameters (Kartashevskii et al., 2015; Akhin et al., 2015; Kartashevskii et al., 2016; Klochkov and Gazizulina, 2016; Skripal and Itsykson, 2017). The application of heavy tail distributions is best approximated to the distribution functions describing the statistical characteristics of the traffic. Determining the Laplace transform of the laws of distributions, which are subject to time intervals between packets and duration of packets and using a spectral method solution to Lindley's integral equation, it is possible to find a function of waiting time of packets in the queue. The definition of this network characteristic allows to improve its quality of service when configuring the network. Therefore, it allows you to track the changeable nature of the traffic, considering the properties, so that the network parameters may be estimated in the most accurate way (Heyman, 1976).

2. Analysis of Waiting Time in $G/G/1$ Queue Using the $PMRQ$ Approximation

A number of studies confirmed the theoretical calculations, demonstrating the effect of autocorrelations in the arrival process (or service process) on the statistics of the queues in the associated queue, as compared to their counterparts updates. More generally, if data on statistical parameters are available, the methods considered and analysed in this study may be used to obtain the average waiting time. There was performed the analysis of the results obtained in (Jagerman et al., 2004) for this study. The idea of this paper is to compare the results of the $PMRQ$ approximation and $PMRS$ while comparing the $G/G/1$ queue which approximate $GI/G/1$ and approximation by the sum of damped exponents (Livny et al., 1993; Patuwo et al., 1993) with the same parameters of the network simulated and approximate to real values.

The proposed approximation scheme presented in (Jagerman et al., 1997), which maps $G/G/1$ queue approximating $GI/G/1$ $PMRQ$ (Peakedness Matched Renewal Queue) uses the

approximation of the flux distribution services in peak mode, which is PMRS (Peakedness Matched Renewal Stream). The approximate arrival process of PMRS preserves the peak of the original arrival process and its speed of arrival; in addition, the square coefficient of variation of the constructed arrival process of PMRS is equal to the variance index of the original arrival process. The merit of the PMRQ approximation is that it is easily solvable, unlike the original G/G/1 queue. To do this, (Jagerman et al., 2004) uses the peak sharpness measure function (Melamed, 1993), and the lag - autocorrelation coefficient in the original stream.

The approach of the *PMRQ* approximation method to find the average waiting time in the *TES+G/I* scheme (Transform Expand Sample) is better compared to the reference values obtained through simulation, where a simple version of the processes of arrival *TES+* served as the arrival process associated with autocorrelation (Melamed, 1991; Jagerman and Melamed, 1992). Consideration of universal classes of autocorrelation (vane) stochastic processes called *TES+* (Melamed, 1991; Jagerman, 1992) subsequently serve as arrival processes in the *G/G/I* queue, the resulting *TES+G/I* is a check of the efficiency of the mean waiting time of the approximation proposed in (Jagerman et al., 2004). When considering this model, both the marginal distribution and the correlation function are selected. The purpose of the *PMRQ* approximation and PMRS is to build the model corresponding to three requirements simultaneously: the marginal distribution should match its original (histogram); the basic correlation model should be close to the originals to an acceptable delay; sample generated by the model should "resemble" with the experimental time series.

Interestingly, the *PMRS* approach is possible in the following way. It can be shown (Melamed, 1993) that the peak function of any renewable traffic flow has a representation of (1)

$$Z_R(s) = \frac{1}{1-\tilde{a}_R(s)} - \frac{\lambda_R}{s}, S \geq 0 \quad (1)$$

where $\tilde{a}_R(S)$ – Laplace transform of the density of the time interval between time and λ_R . From equation (1) is obvious that, for the overall movement of the process *X*, the corresponding *PMRS* approximating the update process *P* is completely determined by the Laplace transform from its time-interval density between time (2)

$$\tilde{a}_R(S) = 1 - \frac{1}{Z_X(s) + \frac{\lambda_X}{s}}, S \geq 0 \quad (2)$$

The equation (2) actually implies that $\lambda_R = \lambda_X$.

The justification for the *PMRS* approximation is obvious: in general, *R* is analytically simpler than *X* is a fact, which will be used later in the approximate analysis of the *G/G/I* queue. More specifically, the *PMRQ* approximations of the *G/G/I* queue with the General Process *X* is simply corresponding to the *GI/G/I* queue with an update of the proposed traffic process *R*, where *R* – *PMRS* approximations.

Therefore, traffic is considered as a pulsating process of time (Shanthikumar and Buzacott, 1980, Downey, 2005), having descriptions and properties of TES processes (Melamed, 1991; Jagerman, 1992). TES processes are a universal class of stationary time series that allow any limit distribution, a wide variety of autocorrelation functions (monotony, oscillation, alternation, etc.), and a wide

range of behavior, including directional and non-directional processes. Processes TES may be designed to fit any empirical (marginal) distribution, and at the same time approximate the empirical autocorrelation functions of different functional forms. It is possible to determine TES processes with different autocorrelation functions, thereby controlling (estimating) their degree.

Solving the statistics of the equilibrium waiting time in the $GI/G/1$ queue usually requires a complex calculation procedure. To find the time-out function $W(\gamma)$ it is necessary to use the inverse transformation of the Laplace function $W^*(S)$ by integrating, considering the Delta function properties (e.g. Lindley's integral equation for the special case).

$$W(\gamma) = \int_0^\gamma w(\tau) d\tau \quad (3)$$

To estimate the parameters of the approximate average waiting time as an alternative to the calculation of the average waiting time, we consider the approximation of large traffic. From (Jagerman et al., 2004) when $GI/G/1$ has a queue with high traffic, namely $\rho \rightarrow 1$ approximation of the average waiting time (4)

$$\bar{W} \approx \frac{E[Y](c_x^2 + c_y^2)}{2} * \frac{\rho}{1-\rho} \quad (4)$$

where c_x, c_y – the coefficients of variation of the time receipt intervals X and service time Y , respectively (Blatov et al., 2013).

In (Melamed, 1993) the efficiency of the proposed approximation is investigated by comparison with modeling. The accuracy of the approximation W is compared in two ways: the scale averaging method and the approximation W , three types of parameters in the system $TES+/G/1$ were changed. Frequencies of receipt with values are considered $\lambda = 0,6$ and $\lambda = 0,8$. These frequencies correspond to the average system load and heavy load respectively, because the average service value has always been one. Parameter of kidney A_E must be distinguished by the range of values by the proper selection of the parameters L and R . The values L and R have a strong effect on the result of kidney failure, while λ has a secondary effect. As a result, four medium-term service time distributions are considered – Exponential, Erlang, Deterministic, $MCE2$ and $MMPP$ (Markov-modulated Poisson processes) (Melamed, 1993), where results for the average waiting time of $SMO TES+/G/1$ are obtained.

3. Approximation as a Density of Sum of Damping Functions

According to the scheme of organization of communication presented in Figure 1, users receive and send data through the switch with a single uplink to the global network. The bandwidth from the subscribers has changed with the software switch, and the port is an edge for the switch, have the configuration of mirroring, allowing you to capture all traffic going to the network. The uplink to the aggregation node was understated in speed, namely the transfer rate was limited to 60 and 80 Mbps.

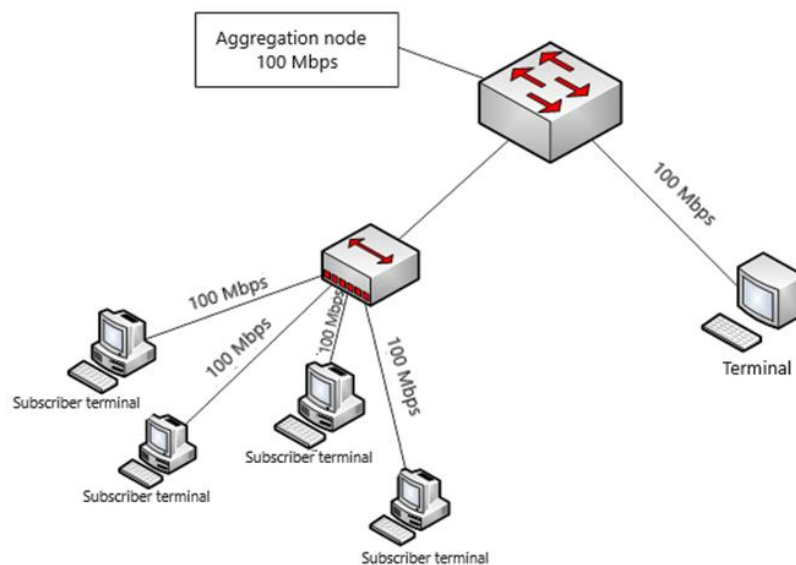


Figure 1. The scheme of organization of communication

Measurements on the user's PC were carried out using the Wireshark protocol analyzer program, this program captures all the packets entering the network interface of the PC. The program allows to get statistical data in the form of text files. In the process of measuring traffic, there were obtained time sequences that contain the required statistics for the study: the time intervals between packets in seconds, the length of packets in bytes, in addition, we need to study the two – dimensional sequence – "the moment of receipt of the package (seconds) – the length of the package (bytes)". The resulting real-time traffic implementation is shown in Figures 2 and 3.

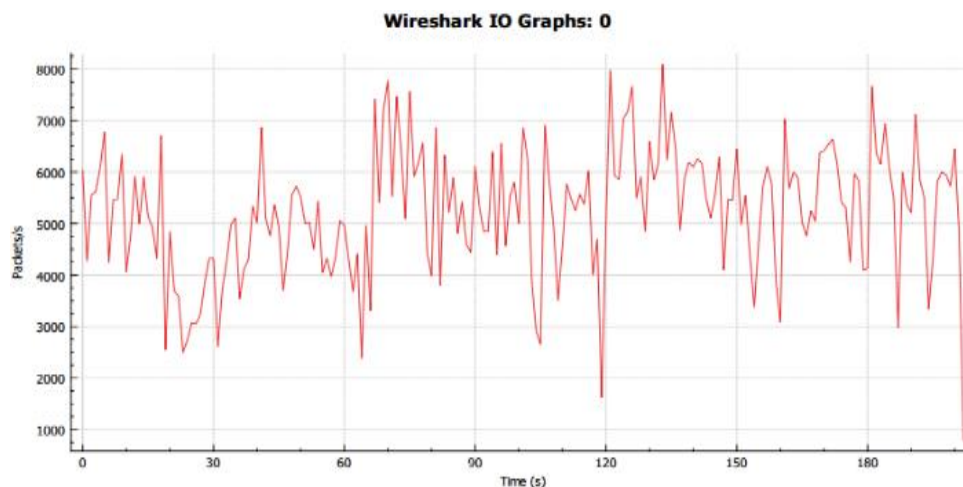


Figure 2. Traffic intensity (packet / sec) at $\lambda = 0.6$

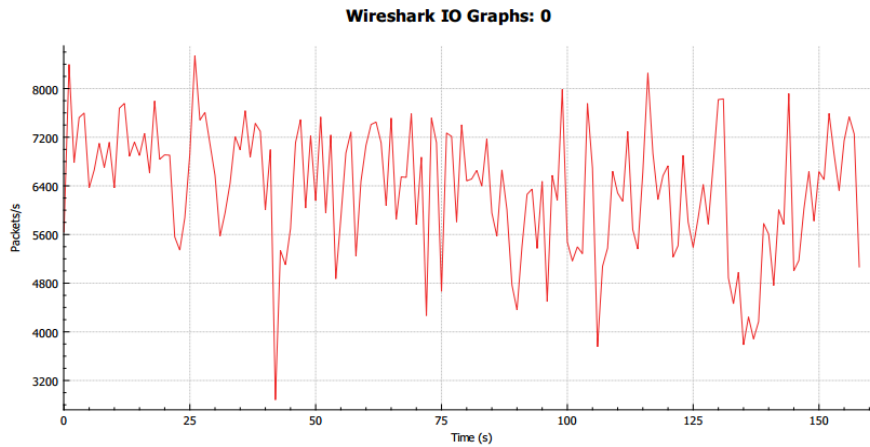


Figure 3. Traffic intensity (packet / sec) at $\lambda = 0.8$

Based on available traffic studies for loads with values $\lambda = 0.6$ and $\lambda = 0.8$ and the resulting distributions according to their histogram, there were considered cases of $G/G/1$ queue system: system types $P/W/1$ and $W/P/1$, where the symbols P and W mean, respectively, Pareto and Weibull distributions (Chupahina, 2013; Itsykson, 2017). The Pareto and Weibull distribution, respectively, have the form (5) and (6):

$$f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}} \quad (5)$$

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right) \quad (6)$$

where α – shape parameter; β – scale parameter.

The Lindley's equation has the following form:

$$F(x) = \int_0^\infty H(x-y)dF(y) \quad (7)$$

where $F(*)$ the distribution function of the waiting time of requirements in the queue;

$H(*)$ – kernel linking an arbitrary probability distribution function of time intervals between receipts of neighboring requirements $A(\tau)$ and an arbitrary function of the distribution of the duration of service requirements $B(\tau)$.

Note that the Lindley's equation itself is derived under the assumption of independence of the elements of the sequence of time intervals between applications and time intervals of applications processing.

It is known that for the $G/G/1$ queue, a spectral method for solving Lindley's integral equation is possible if for probability $\alpha(\tau)$ and $b(\tau)$, the respective distributions $A(\tau)$ and $B(\tau)$ are used for the approximation as a sum of exponentially decaying functions.

Herewith $\alpha(\tau)$ and $b(\tau)$ know

$$\alpha(\tau) = \sum_{k=1}^n \xi_k e^{-\eta_k(\tau)} \quad (8)$$

$$b(\tau) = \sum_{k=1}^l \varphi_k e^{-\theta_k(\tau)} \quad (9)$$

where $n = 5$; $l = 5$.

The average waiting time of the packet in the queue for the above examples is found according to the known property of the characteristic function

$$t_{cp} = \left. \frac{dF(s)}{ds} \right|_{s=0} \quad (10)$$

with an average load of $\lambda = 0.6$, the statistical parameters obtained histograms correspond to the $W/P/I$ system for the time intervals between the packets (W): $\alpha_1=0.13$, $\beta_1=0.000225$ and for the length of the packet (P) – parameters: $\alpha = 0.39$, $\beta = 54$.

Based on the simulation results, the parameters of the stream generated by the sending node were checked against the specified parameters using the EasyFit program. The following distributions were obtained for the time intervals between packets and packet durations Figures 4, 5.

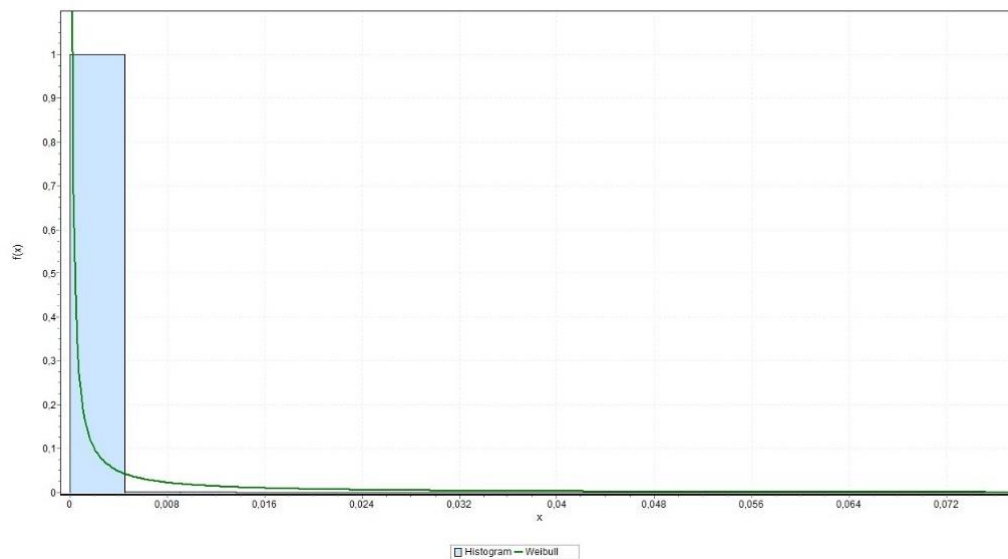


Figure 4. Weibull-distribution of time intervals between packages, the value of the parameters: $\alpha_1=0.13$, $\beta_1=0.000225$

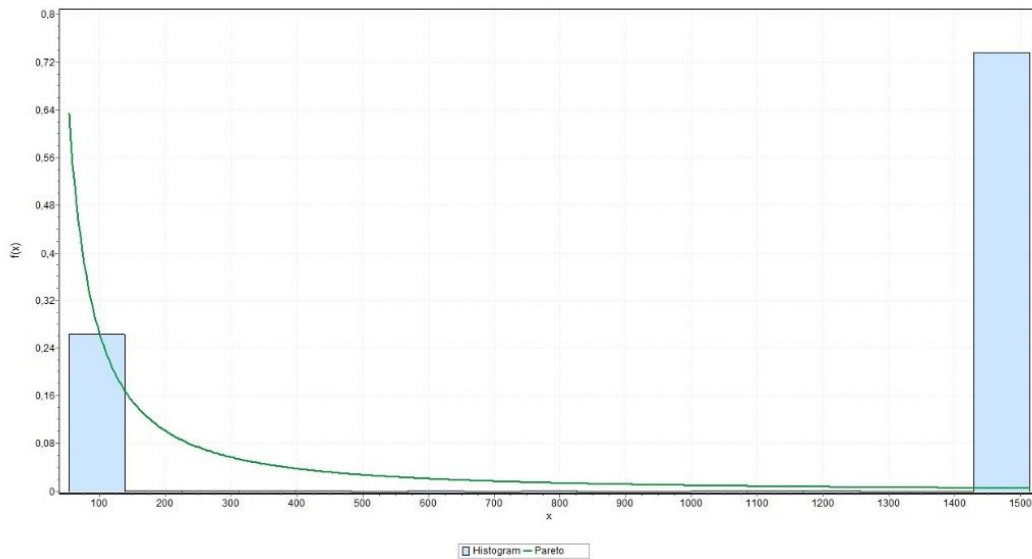


Figure 5. Pareto distribution of the lengths of the packets, the value of the parameters: $\alpha = 0.39, \beta = 54$

After the calculation according to the above algorithm, the average waiting time of the packet was equal to $t_{cp} = 42.6$.

At large load $\lambda = 0.8$ the statistical parameters of obtained histograms correspond to the W/P/1 system for the time intervals between the packets (W) parameters have been set $\alpha_1 = 0.78, \beta_1 = 1500$ and for the length of the packet (P) – parameters: $\alpha = 0.41, \beta = 62$ (Figures 6, 7).

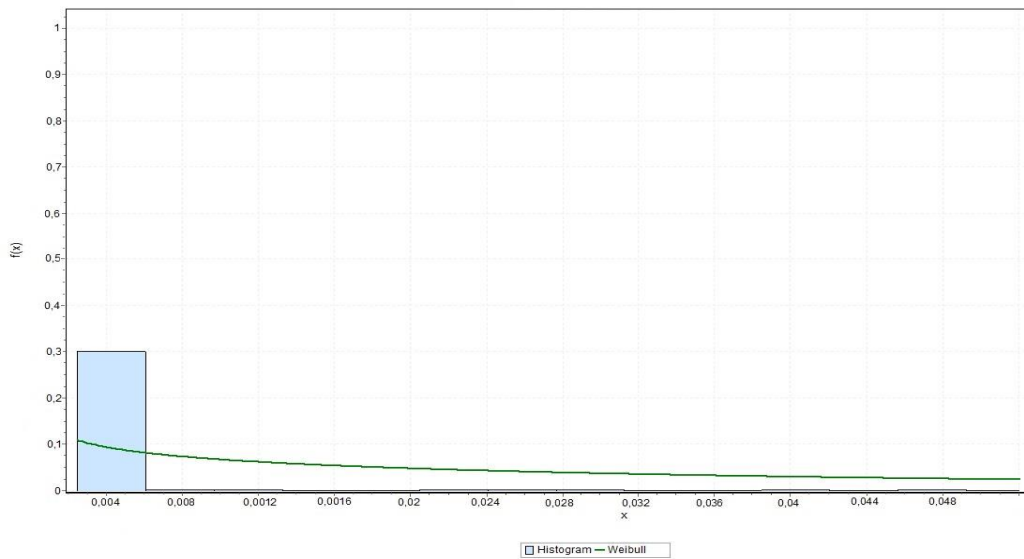


Figure 6. Weibull-distribution of time intervals between packages, the value of the parameters: $\alpha_1 = 0.78, \beta_1 = 1500$

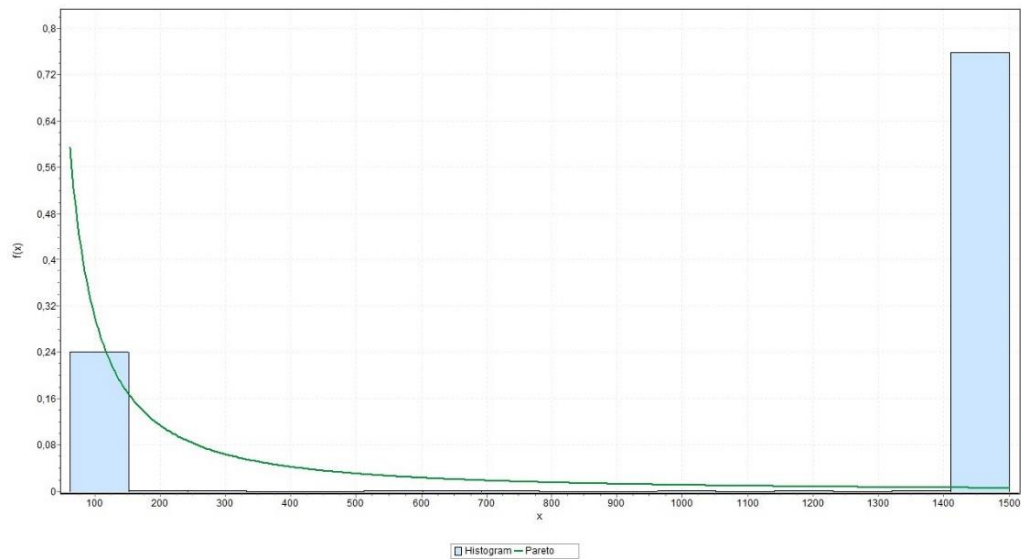


Figure 7. Pareto distribution of the lengths of the packets, the value of the parameters: $\alpha=0.41$, $\beta=62$

After the calculation, the average waiting time of the packet is equal to $t_{cp}=50.662$.

4. Conclusion

This article presents the methods of determining the average waiting time in the $G/G/1$ queue: scale averaging method and approximation \bar{W} , as well as an approximation in the form of a sum of damping functions to find the solution of Lindley's integral equation. In this case, the TCP module is selected as the transport layer module, and the intensity of the packets is equal to $\lambda = 0.6$ and $\lambda = 0.8$.

The error of solving the Lindley's equation by the spectral method in comparison with the results of section 11 of (Melamed, 1993) is approximately: $\lambda = 0.6 - 4\%$, and with parameter $\lambda = 0.8 - 14\%$, which is perfectly acceptable. This error may be associated with the idealization of the initial (real) distributions of time parameters of the considered models, with the assumption of the independence of the elements of the considered sequences of time intervals and with certain accuracy characteristic of any program used in the removal and processing of network traffic. Traffic capture is carried out using the Wireshark protocol analyzer program. Time distributions were obtained in the EasyFit data analysis program, designed for fast statistical data analysis and decision making.

The comparison of methods of approximating network traffic allows to estimate the average waiting time of the packet using statistical data analysis, which will make it possible to improve the quality of service and predict the behavior of traffic.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgement

The research carried out with the financial support of the grant from the Program Competitiveness Enhancement of Peter the Great St. Petersburg Polytechnic University, Project 5-100-2020.

References

- Akhin, M., Kolton, S., & Itsykson, V. (2015). Random model sampling: making craig interpolation work when it should not. *Automatic Control and Computer Sciences*, 49(7), 413-419.
- Altiok, T., & Melamed, B. (2001). The case for modeling correlation in manufacturing systems. *IIE Transactions*, 33(9), 779-791.
- Blatov, I.A., Kartashevskii, V.G., Kireeva, N.V., & Chupakhina, L.R. (2013). The method of approximating an arbitrary density distribution of the sum of exponents. *Vestnik VGU*, 2, 53–57.
- Chupahina, L.R. (2013). Analysis of the characteristics of queuing systems in the transmission nonpoissonian traffic approximation method of distribution functions. *Samara, Povolzhskiy State University of Telecommunications and Informatics* Vol. 1.
- Downey, A.B. (2005). Lognormal and Pareto distributions in the Internet. *Computer Communications*, 28(7), 790-801.
- Heyman, D.P. (1976). *Queueing systems: theory by Leonard Kleinrock*. John Wiley & Sons. Inc., New York. *Networks*, 6(2), Vol. 1., 189-190.
- Itsykson, V.M. (2017). Formalism and language tools for specification of the semantics of software libraries. *Automatic Control and Computer Sciences*, 51(7), 531-538.
- Jagerman, D., Melamed, B., & Willinger, W. (1996, March). Stochastic modeling of traffic processes. In *Frontiers in queueing: Models, Methods and Problems*, CRC Press.
- Jagerman, D.L., & Melamed, B. (1992). The transition and autocorrelation structure of TES processes: Part II: Special Cases. *Communications in Statistics. Stochastic Models*, 8(3), 499-527.
- Jagerman, D.L., Balcioğlu, B., Altiok, T., & Melamed, B. (2004). Mean waiting time approximations in the G/G/1 queue. *Queueing Systems*, 46(3-4), 481-506.
- Jagerman, D.L., Melamed, B. & Willinger, W. (1997). Stochastic modeling of traffic processes. In *Frontiers in Queueing: Models and Applications in Science and Engineering* (pp.271–320). CRC Press.
- Kartashevskii, V.G., Kireeva, N.V., Buranova, M.A., & Chupakhina, L.R. (2015). Simulation and analysis of generalized queue system with arbitrary distributions of system parameters. *Infokommunikacionnye Tehnologii*, 3, 252-258.
- Kartashevskiy, V., Kireeva, N., Buranova, M., & Chupakhina, L. (2016, October). Approximation of distributions in the problems of the analysis of self-similar traffic. In *Third International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T), 2016*, (pp. 105-108), IEEE.
- Klochkov, Y., & Gazizulina, A. (2016). Application of the method of performance evaluation of the production process design using associative design. *Key Engineering Materials*, 684, 448-452.
- Livny, M., Melamed, B., & Tsiolis, A.K. (1993). The impact of autocorrelation on queuing systems. *Management Science*, 39(3), 322-339.

- Melamed, B. (1991). TES: A class of methods for generating auto correlated uniform variates. *ORSA Journal on Computing*, 3(4), 317-329.
- Melamed, B. (1993). An overview of TES processes and modeling methodology. In *Performance Evaluation of Computer and Communication Systems* (pp. 359-393). Springer, Berlin, Heidelberg.
- Patuwo, B.E., Disney, R.L., & McNickle, D.C. (1993). The effect of correlated arrivals on queues. *IIE Transactions*, 25(3), 105-110.
- Shanthikumar, J.G., & Buzacott, J.A. (1980). On the approximations to the single server queue. *International Journal of Production Research*, 18(6), 761-773.
- Skripal, B., & Itsykson, V. (2017, April). Aspect-oriented extension for the Kotlin programming language. In *CEUR Workshop Proceedings, CEUR*, pp. 1864.



Original content of this work is copyright © International Journal of Mathematical, Engineering and Management Sciences. All rights reserved. Except of uses under a Creative Commons Attribution 4.0 International (CC BY 4.0) license at <https://creativecommons.org/licenses/by/4.0/>