# Poisson Packet Traffic Generation Based on Empirical Data

**Andrej KOS**
**Laboratory for Telecommunications**
**University of Ljubljana, Faculty of Electrical Engineering**
**Tržaška 25, 1000 Ljubljana, Slovenia**
andrej.kos@fe.uni-lj.si
http://www.ltfe.org/

**and**

**Janez BEŠTER**
**Laboratory for Telecommunications**
**University of Ljubljana, Faculty of Electrical Engineering**
**Tržaška 25, 1000 Ljubljana, Slovenia**
janez.bester@fe.uni-lj.si
http://www.ltfe.org/

## ABSTRACT

An algorithm for generating equivalent Poisson packet traffic based on empirical traffic data is presented in this paper. Two steps are required in order to produce equivalent Poisson packet traffic. Real traffic trace is analyzed in the first step. In the second step, a new equivalent synthetic Poisson traffic is generated in such a way that the first order statistical parameters remain unchanged. New packet inter-arrival time series are produced in a random manner using negative exponential probability distribution with a known mean. New packet size series are also produced in a random manner. However, due to specified minimum and maximum packet sizes, a truncated exponential probability distribution is applied.

**Keywords:** Packet Traffic Generation, Poisson Model and Fitting onto Empirical Traffic Data.

## 1. INTRODUCTION

Analytical solutions for connection-oriented circuit-switched voice networks, based on Poisson models, have been one of the most successful applications of mathematical techniques in this field [1]. This is possible because voice traffic spans long time scales and has the property of being relatively homogenous and predictable.

Connectionless packet-oriented data traffic, found contemporary packet systems, on the other hand, is much more variable, since a typical multimedia application contains a mix of packets from various sources. Analysis of high-quality traffic measurements confirms the prevalence of long-range dependent features and scaling in traffic processes loading packet switching communications networks [2,3,4,5,6,7,8,9], where aggregation of traffic is really not successful in smoothing out traffic burstiness.

Therefore, the applicability of traffic analysis based on mathematical tractability has been diminishing, as purely mathematical traffic sources cannot capture traffic characteristics in real packet networks to the extent that they would allow a detailed performance evaluation of the network. Consequently, the importance of computer simulations has been growing considerably, which poses different requirements for traffic source models [10,11].

Ideally, a suitable traffic source model should represent real traffic or capture essential characteristics of traffic that have a significant impact on network performance with only a small number of parameters (i.e. the parsimonious model), as well as facilitate the rapid generation of packets [8,9,12].

In contrast to connection-oriented circuit-switched voice networks, there seems to be no so-called "universal laws" as regards homogenous systems, typical users and generic behaviour in connectionless-oriented packet-switched data networks. However, recently, has analysis suggested that in contrast to network edge traffic, highly aggregated core traffic slowly tends back to Poisson [13,14].

Therefore in the field of performance evaluation, packet traffic - the characteristics of which obey Poisson distributions – can still be used as a reference point [15].

This paper is organized as follows:

- the description of Poisson packet traffic generation based on empirical data is presented in Section 2;
- the process of fitting the exponential distribution onto empirical data is shown in Section 3;
- Main implications and conclusions are summarised in Section 4.

## 2. POISSON PACKET TRAFFIC GENERATION BASED ON EMPIRICAL DATA

In packet networks, two random processes are used to define the traffic properties. These are the inter-arrival time series $\{T_i\}$, where $T_i$ denotes the $i$-th inter-arrival time between two successive packet arrivals, and the packet length sequence, $\{L_i\}$, where $L_i$ represents the length of the $i$-th packet [16]. Therefore, in order to produce packet traffic with Poisson-like characteristics based on first order statistical properties of empirical data (traffic), two steps are required. Thus, packet inter-arrival time and packet size series are produced.

The starting point is the capture of real traffic traces. In this instance real traffic traces were collected with a Linux workstation with 100 Mbit/s Ethernet interface [15]. An additional special tool for the post processing of captured real traffic traces has been developed in Perl. Captured real traffic traces were then analysed, and Matlab routines were used for traffic trace analysis. Based on this information on empirical data, the equivalent Poisson packet traffic was generated.

In the first step, the real traffic trace file is analysed. As a result the following parameters are retrieved: number of packets; average throughput; minimal, maximal and average packet size; average packet inter-arrival time, as well as the number of different destinations. In addition, a special filter vector is created for packet destination addresses (i.e. for performance evaluation of packet switches). The purpose of the destination vector is to enable the proper distribution of packets (generated in the second step) between the list of available destination addresses. The filter vector is used in combination with uniform random distribution in order to properly transform the original uniform distribution into a distribution which is in accordance with the distribution of destination addresses in the original real traffic trace.

In the second step, a new equivalent Poisson packet traffic trace is generated in such a way that all the above measured first-order statistical parameters and the number of packets remain unchanged. New packet inter-arrival times are produced in a random manner by using negative exponential probability distribution with a known mean. New packet sizes are also produced in a random manner. However, due to specified minimum and maximum packet size, a truncated exponential probability distribution is applied. Thus, statistical values for minimal, average, and maximal packet size are properly preserved. To select the destination address of a new packet, a uniform probability distribution is applied and then transformed by using the filter vector (generated in the first step).

## 3. FITTING THE TRUNCATED EXPONENTIAL DISTRIBUTION ONTO EMPIRICAL DATA

The lower limit of exponential distribution is zero, whereas there is no upper limit. This is adequate for packet inter-arrival time generation, but not adequate for packet size generation.

### 2.1. Packet size generation

In practice, packet size is always limited. The lower limit is represented by a packet header, which is present in every packet. The upper limit is usually defined by technical restrictions. Therefore, truncated exponential distribution has to be used. Its lower and upper limits are zero and one, respectively. The course of an exponential curve is defined by parameter $a$, while truncated exponential probability distribution is defined as:

$$f\left(x,a\right)=\frac{a\,\mathrm{e}^{a}}{\mathrm{e}^{a}-1}\mathrm{e}^{a\,x}. \tag{1}$$

The truncated exponential probability distribution for different values of parameter $a$ is displayed in Figure 1. The average value of the function for a given value $a$ is

$$\int_0^1 x\,f(x,a)\,\mathrm{d}x=\frac{1}{a}-\frac{1}{\mathrm{e}^{a}-1}. \tag{2}$$

The problem in need of solution is how to generate random values that obey truncated exponential probability distribution for a given parameter $a$, as well as follow the precepts of random number generation.

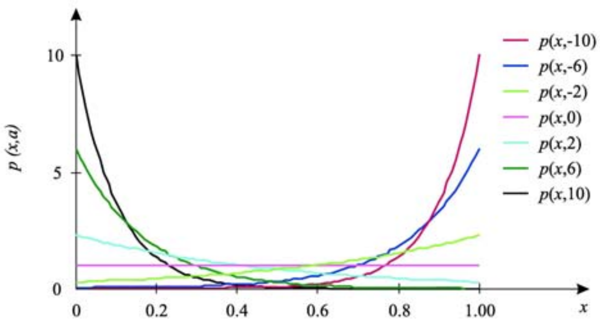


**Figure 1: Truncated exponential probability distribution for different values of parameter *a*.**

Figure 1 reveals that the function $f(x)$ is limited by the values $f(0)$ and $f(1)$. As to which value is greater, depends on the parameter $a$. If a uniform random generator is used for the generation of values $f(x)$, and each $x$ is reversibly calculated from $f(x)$, then $x$ has a truncated exponential probability distribution. Therefore the filter that enables the transformation of uniform random generation to truncated exponential random generation is defined as:

$$x = -\frac{1}{a}\ln\left(f\left(x\right)\frac{e^{a}-1}{a\,e^{a}}\right). \tag{3}$$

Values $f(x)$ are generated with a uniform random generator in the following way:

$$f\left(x\right)=\begin{cases}\text{UniformReal}[f\left(1\right),f\left(0\right)],\ a>0,\\ \text{UniformReal}[f\left(0\right),f\left(1\right)],\ a<0,\\ \text{special case,}\qquad\qquad a=0,\end{cases} \tag{4}$$

where

$$f(0)=\frac{a\,e^{a}}{e^{a}-1} \tag{5}$$

and

$$f(1)=\frac{a}{e^{a}-1}. \tag{6}$$

Defining a new value $y$ (e.g. the size of a newly generated packet), that lies between minimal and maximal value in such a way that the whole series $Y$ obeys truncated exponential probability distribution is achieved in the following manner: first $f(x)$ is defined. The value $x$ is then calculated from $f(x)$; it lies in the interval between 0 and 1. In the final step the transformation of $x$ is performed in order to obtain the desired interval for $y$:

$$y=\min+x^{*}\left(\max-\min\right). \tag{7}$$

In the event that parameter $a$ is 0, the $x$ values are directly generated using a uniform random generator:

$$x=\text{UniformReal}[0,1]. \tag{8}$$

In the execution of the above formulae, it is assumed that the appropriate value $a$ is known a priori. However, it depends on the desired average value (e.g. average packet size) of the sequence $Y$. The value $a$ is determined by iteration. The starting point is set at $a=0$. This corresponds to a special case with uniform probability distribution between minimal and maximal values. In this case the average value of $Y$ lies in the middle, between those two extreme values. In the event that the desired average value of $Y$ lies above or below this mean, iteration is used. For desired average values below the mean, the parameter $a$ is increased. Each time a new average value is calculated by:

$$\text{avrgGen}=\min+(\max-\min)\cdot\left(\frac{1}{a}-\frac{1}{e^{a}-1}\right). \tag{9}$$

Value $a$ is increased until avrgGen is equal or lower than desired average value of $Y$. For desired average values above the mean, the parameter $a$ is decreased in a similar manner until avrgGen is equal or higher than the desired average value of $Y$.

### 2.2. Packet inter-arrival time generation

For packet inter-arrival time generation negative exponential distribution with a known mean is used.

## 5. CONCLUSION

With the rise of packet switching it was thought that modelling of connectionless-oriented packet-switched data traffic differs from conventional connection-oriented circuit-switched voice traffic in so many fundamental ways that the same concepts for traffic models would not be applicable.

However, the latest analyses show that high-speed and highly-aggregated packet traffic in network core tends back to Poisson, which revives the use of classical Poisson concepts for traffic modelling.

In network simulations and optimisation scenarios the application of equivalent Poisson traffic is reasonable, due to the fact that it offers mathematically well-defined and well-known conditions for simulations (at the price of reduced accuracy of traffic description).

An algorithm for the generation of equivalent Poisson packet traffic is presented in this paper. Packet inter-arrival time and packet size series with Poisson-like properties are produced. Empirical data is used as a basis.

To conclude, although Poisson traffic model within packet systems doesn't describe the most important properties of real traffic, in many simulation and network optimisation scenarios the application of equivalent Poisson packet traffic is very reasonable, due to the fact that it offers well-defined and well-known conditions for simulations. However, the best choice (when possible) is an application with a combination of both equivalent Poisson and empirical packet traffic.

## 6. REFERENCES


[1] P. Homan, "Configuration and Control of Queues for Assuring Quality of Telecommunication Services in Packet Switches", Ph. D. Thesis, University of Ljubljana, Faculty of Electrical Engineering, 2001.

[2] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the Self-similar Nature of Ethernet Traffic" (ext. version), **IEEE/ACM Transactions on Networking**, Vol. 2, Iss. 1, 1994, pp. 1-15.

[3] V. Paxson and S. Floyd, "Wide-area Traffic: The Failure of Poisson", **IEEE ACM Transactions on Networking**, Vol. 3, No. 3, 1995, pp. 226-244.

[4] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-range Dependence in Variable-bit-rate Video Traffic", **IEEE Transactions on Communications**, Vol. 43, Feb/Mar/Apr, 1995, pp. 1566-1579.

[5] M. E. Crovella and A. Bestavros, "Self-similarity in World Wide Web Traffic Evidence and Possible Causes", **IEEE ACM Transactions on Networking**, Vol. 5, No. 6, 1997, pp. 835-846.

[6] A. Kos, J. Bešter and P. Homan, "Next Generation Networks and Services in Slovenia", **Second International IFIP-TC6 Networking Conference**, E. Gregori (ed.), Springer Verlag, New York, Pisa, Italy, May 2002, pp. 1147-1152.

[7] R. Žurbi, "Signalling and Control Protocols in Next Generation Networks", Ph. D. Thesis, University of Ljubljana, Faculty of Electrical Engineering, 2001.

[8] P. Homan, J. Bešter, A. Kos and T. Slivnik, "The Impact of Bursty Traffic on FPCF Packet Switch Performance", **Computer Communications,** Vol. 26, Iss. 6, April 2003, pp. 513-521.

[9] A. Kos, "Performance Analysis and Optimisation of Packet Switches Loaded with Telecommunication Traffic", Ph. D. Thesis, University of Ljubljana, Faculty of Electrical Engineering, May 2003.

[10] N. McKeown, A. Mekkittikul, V. Anantharam and J. Walrand, "Achieving 100 % Throughput in an Input-Queued Switch", **IEEE Transactions on Communications**, Vol. 47, No. 8, 1999, pp. 1260-1267.

[11] P. Homan and J. Bešter, "Enabling Quality of Service in Input-Queued Packet Switches", **Proceedings of the International Conference on Trends in Communications EUROCON '01**, Bratislava, 2001, pp. 496-499.

[12] A. Kos, P. Homan, T. Slivnik and J. Bešter, "Performance Evaluation of a Synchronous Bulk Packet Switch under Real Traffic Conditions", **IEICE Transactions on Communications**, Vol. E86-B, No. 5, May 2003, pp. 1612-1624.

[13] N. McKeown and T. E. Anderson, "A Quantitative Comparison of Scheduling Algorithms for Input-Queued Switches", **Computer Networks**, Vol. 30, No. 24, 1998, pp. 2309-2326.

[14] S. T. Chuang, A. Goel, N. McKeown and B. Prabhakar, "Matching Output Queueing with a Combined Input/Output-Queued Switch", **IEEE Journal on Selected Areas in Communications**, Vol. 17, No. 6, 1999, pp. 1030-1038.

[15] "Passive Measurement and Analysis Research Project". **URL:** http://pma.nlanr.net/PMA/

[16] J. B. Gao and I. Rubin, "Multiplicative Multifractal Modeling of Long-Range-Dependent Network Traffic", **International Journal on Communications Systems**, October 2001, Vol. 14, Iss. 8, pp. 783-801.