Some Analysis of Single Server Queues with Traffic Modeled by Heavy-tailed Distributions

Weldisson Ferreira Ruas
Instituto Nacional de Telecomunicações - Inatel
Santa Rita do Sapucaí - MG - Brazil
weldisson@mtel.inatel.br

José Marcos Câmara Brito
Instituto Nacional de Telecomunicações - Inatel
Santa Rita do Sapucaí - MG - Brazil
brito@inatel.br

Abstract— ***Markovian models are not suitable to characterize traffic in some modern telecommunications networks. Among the new proposed models, those based on heavy-tailed distributions offer lower complexity. However, there are different approaches to traffic modeling using heavy-tailed distributions. In this paper we investigate the influence of these approaches in the performance of an isolated single server queue. The models considered are G/M/1 and G/G/1, with G modeled by Pareto, Lognormal and Weibull distributions.***

I. INTRODUCTION

Traffic in telecommunications networks has evolved from voice traffic to multimedia traffic, including voice, data and video. In this new scenario, the traditional Markov models are not suitable to characterize the traffic in the network.

In 1994, Leland et al [1] demonstrate that Ethernet Local Area Network traffic is statistically self-similar and that none of the traditional traffic models is able to capture this behavior. Since then, several studies were conducted to propose new traffic models to telecommunications networks. These works can be classified in three categories:

a) Based on measurements.
b) Based on fractal models.
c) Based on generic models.

The approach based on generic models is less complex than fractal models [2][3] and is the subject of our work. In this kind of model, the arrival processes is modeled by a heavy-tailed distribution, like Pareto, Lognornal or Weibull distributions, and the service time can be modeled by an exponential distribution (G/M/1 queue), by a heavy-tailed distribution (G/G/1 queue) or can be considered constant (G/D/1 queue).

An important performance parameter of a queuing system is the mean waiting time, computed as a function of the utilization factor of the server. Thus, to define the performance of the system it is necessary to vary the utilization factor of the server. However, we found in the literature three different approaches to obtain this variation [4][5]:

a) Fixing the service time and varying the arrival rate by varying the shape parameters of the heavy-tailed distributions.
b) Fixing the arrival rate and varying the service time. In this case the shape parameters of the heavy-tailed distributions are fixed.
c) Fixing the service time and varying the number of traffic sources (identical and independents). In this case, the shape parameters of the heavy-tailed distributions used in each traffic source are fixed.

In this paper we compare the performance, based on simulations, of single server queues using these three approaches to vary the utilization factor of the server. We consider G/M/1 and G/G/1 systems, with G modeled by the following heavy-tailed distributions: Pareto, Lognormal and Weibull. Also, we introduce the idea of a performance factor, in order to compute the performance of a G/M/1 queue based on results from a M/M/1 queue.

The parameter used to evaluate the performance of the systems is the mean waiting time of each queue, as a function of the utilization factor.

The remaining of this paper is organized as follow: Section II presents some characteristics of the heavy-tailed distributions used in this paper; Section III describes the scenarios used in our simulations; Section IV presents the results for Scenario I; Section V shows the analysis for Scenario II; Section VI presents the results for Scenario III; Section VII presents the performance factor concept and its results; finally, Section VIII presents the conclusions.

II. HEAVY-TAILED DISTRIBUTIONS

Let X a random variable (R.V) with Probability Density Function (PDF) *f(x)* and Cumulative Distribution Function (CDF) F(x). The R.V. X has a heavy-tailed distribution if: [6]

$$P(X > x) \approx L(x)x^{-\alpha}, \quad \alpha > 0, \quad x \to \infty \tag{1}$$

where $L(x)$ is a function which decays slowly, tending to infinity when:

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1, \quad \forall c > 0 \tag{2}$$

Some important heavy-tailed distributions used to traffic modeling in telecommunications networks are Pareto, Lognomal and Weibull distributions. The main characteristics of these distributions are resumed below.

A. Pareto Distribution

Pareto distribution is widely used for traffic modeling in telecommunications networks. This distribution can be represented using one, two or three parameters. Results presented in [7] show that the use of Pareto with two parameters results in a lower mean queuing time, compared with the one parameter distribution. In our work, we opted to use de Pareto Distribution with one parameter.

The Probability Density Function of Pareto distribution with one parameter is given by: [7]

$$f(x) = \frac{\alpha}{(1+x)^{\alpha+1}}, \quad 1 < \alpha < 2, \quad x \ge 0 \qquad (3)$$

The parameter α is the shape parameter of the distribution. If this parameter takes values between one and two, the expected value of the R.V. is finite, its variance is infinity and the process is self-similar. The expected value can be computed by

$$E(x) = \frac{1}{\alpha - 1} \qquad (4)$$

B. Lognormal Distribution

Although Lognormal distribution is mentioned in several works as a heavy-tailed distribution, it does not have infinite variance, which is the main characteristic of a heavy tailed distribution [8][9]. However, as their moments increase very rapidly, it has also been used for traffic modeling.

The Probability Density Function for Lognormal distribution is given by:

$$f(x) = \frac{1}{x\sqrt{2\pi\beta^2}} e^{\left[-\frac{(\ln x - \alpha)^2}{2\beta^2}\right]}, \quad \beta^2 > 0, \quad \alpha \in R, \quad x \in (0, +\infty) \qquad (5)$$

where α and β are the shape parameters of the distribution. The expected value for this distribution is given by:

$$E(x) = e^{\alpha + \beta^2/2} \qquad (6)$$

C. Weibull Distribution

Weibull distribution has also been used to traffic modeling in telecommunications networks [5][10]. The PDF of this distribution is given by:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^{\alpha}}, \quad 0 < \alpha \le 1, \quad \beta > 0, \quad x \in (0, +\infty) \qquad (7)$$

where α and β are the shape parameters of the distribution. To characterize a heavy-tailed distribution, the parameter α must take values between zero and one [11].

The expected value of the Weibull distribution is given by:

$$E(x) = \beta \cdot \Gamma\left(1 + \frac{1}{\alpha}\right) \qquad (8)$$

III. Scenarios for the Simulations

In our simulations, we have used the software Arena 11.0 Professional. This tool does not provide the possibility to generate Pareto distributions directly. Thus, for this distribution we have used the Percentile Transformation Method [12].

Gross et al [13] show that there are some difficulties in simulating queues with Pareto service. To overcome these problems, it is necessary to consider a truncated expected value, obtained from a truncated CDF, for the distribution. This truncated expected value is given by:

$$E_T(x) = \frac{\alpha}{F(T)}\left[\frac{1}{\alpha(1+T)^{\alpha}} - \frac{1}{(\alpha-1)(1+T)^{\alpha-1}} + \frac{1}{\alpha(\alpha-1)}\right]$$
(9)

where *T* is the truncation parameter and *F(T)* is given by:

$$F(T) = 1 - \frac{1}{(1+T)^{\alpha}} \qquad (10)$$

In our first scenario, called Scenario I, we keep the shape parameters of the heavy-tailed distributions fixed and vary the utilization factor varying the service time. Due this, it is necessary to normalize the mean waiting time of the queues in order to compare different results. Thus, in all results presented is this paper, we use the mean waiting time normalized by the service time.

In the second scenario, called Scenario II, we fix the service time and vary the shape parameters of the heavy-tailed distribution, thus varying the input traffic and the utilization factor of the server

Finally, in the third scenario, called Scenario III, we define a basic traffic generator with fixed shape parameters of the heavy-tailed distributions and fix the service time. The variation of the utilization factor is achieved by varying the number of traffic generators in the input of the queue system.

This scenario is illustrated on Figure 1.

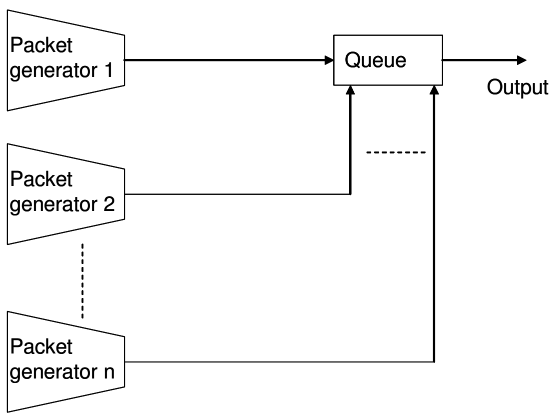

Fig. 1: *Block diagram for Scenario III.*

IV. RESULTS FOR SCENARIO I

At first, let's go to compare the influence of the heavy-tailed distribution on the performance of the system. Figure 2 shows the normalized mean waiting time for Pareto/M/1, Lognormal/M/1 and Weibull/M/1 systems. For Pareto and Weibull distributions, we also present the theoretical results using Transform Approximation Method (TAM) [4][5][14]. The shape parameters are fixed as: Pareto, α = 1.3; Lognormal, α = 1.015 and β = 2; Weibull, α = 0.257 and β = 1.

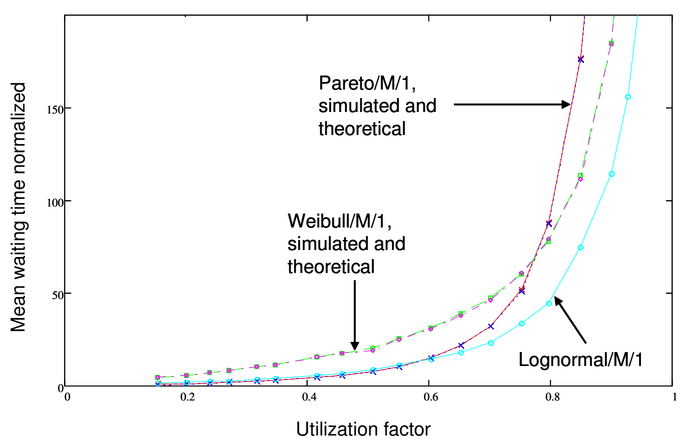

Fig. 2: *Normalized mean waiting time considering G/M/1 models - shape parameter* α*:Pareto, 1.3; Lognormal, 1.015; Weibull, 0.257.*

We can see that, in this comparison, Lognormal distribution results in the best performance. If the utilization factor is less than 0.78, the worst performance is obtained with Pareto distribution; otherwise, Weibull distribution results in the worst performance.

Figure 3 presents the results for a different set of shape parameters: In this case, we change the parameter α to: Pareto, α = 1.7; Lognormal, α = 2.8515; Weibull, α = 0.6515. Now, Lognormal distribution has the best performance and Pareto distribution the worst one, in all range of utilization factor. Comparing figures 2 and 3, we can see that the performance of the queue system depends on the shape parameters of the heavy-tailed distributions.

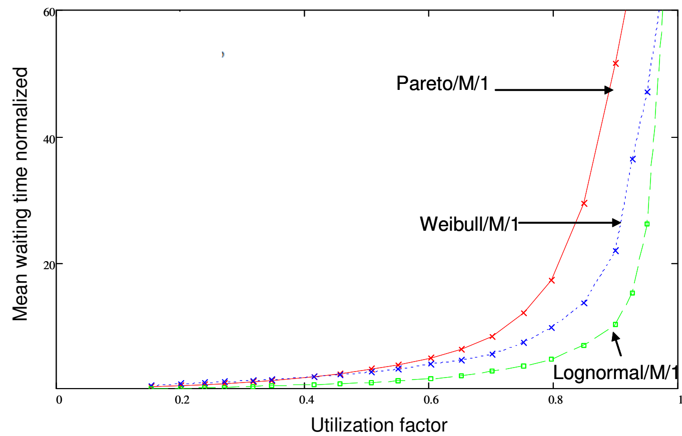

Fig. 3: *Normalized mean waiting time considering G/M/1 models – shape parameter* α*: Pareto, 1.7; Lognormal, 2.8515; Weibull, 0.6515.*

Now, we compare the performance of the queuing system considering G/G/1 model. Figure 4 presents the results for the shape parameters fixed as in Figure 2; and Figure 5 shows the results considering shape parameters fixed as in Figure 3.

Again, Lognormal distribution results in best performance and Pareto, in most of the range of the utilization factor, has the worst performance. Comparing figures 4 and 5, we conclude again that the performance of the queuing system depends on the shape parameters.

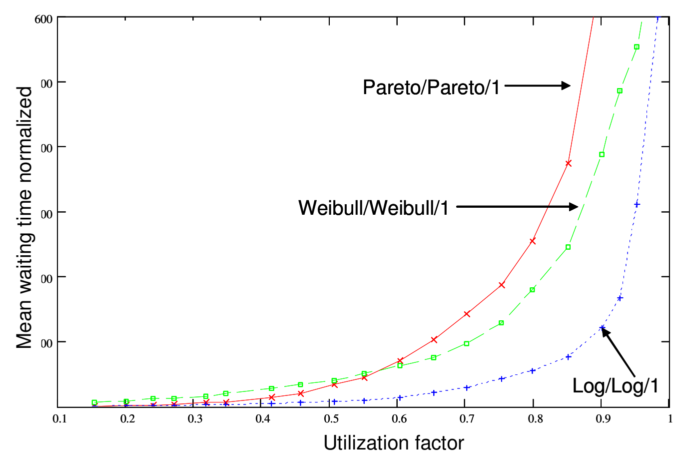

Fig. 4: *Normalized mean waiting time considering G/G/1 models - shape parameter* α*:Pareto, 1.3; Lognormal, 1.015; Weibull, 0.257.*

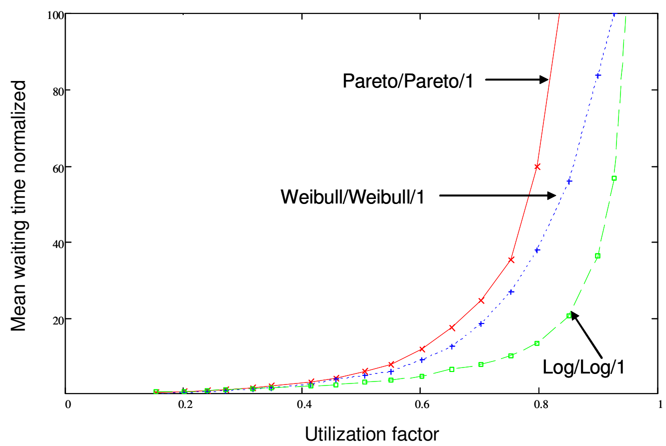

Fig. 5: *Normalized mean waiting time considering G/G/1 models – shape parameter α: Pareto, 1.7; Lognormal, 2.8515; Weibull, 0.6515*

Comparing Figure 2 with Figure 4 and Figure 3 with Figure 5, as expected, we can see that the performance in G/G/1 queues is much worse than the performance in G/M/1 queues.

V. Results for Scenario II

In this scenario we fix the service time and vary the shape parameters of the distributions to obtain the variation of the utilization factor. We use two set of parameters:

Set 1: we fix the service time to 0.333 seconds and vary the shape parameters of the heavy-tailed distribution as follows: Pareto, $0.4596 \leq \alpha \geq 2.9439$; Lognormal, $-1.579 \leq \alpha \geq 0.2773$ and $\beta = 1$; Weibull, $0.1698 \leq \alpha \geq 1.0879$ and $\beta = 0.5$.

Set 2: we fix the service time to 0.1 seconds and vary the shape parameters as follows: Pareto, $1.532 \leq \alpha \geq 9.813$; Lognormal, $-2.7837 \leq \alpha \geq -0.9265$; Weibull, $0.0509 \leq \alpha \geq 0.3263$.

The range of parameter values was chosen to obtain the desired variation in the utilization factor.

The problem with this approach is that the shape parameters assume values outside the range needed to guarantee the self-similarity of the traffic.

At first, figures 6 and 7 presents the results for G/M/1 systems, with different service times. In Figure 6, only to validate the simulation processes, we also present the theoretical results obtained using the method TAM. In Figure 6 we use the parameters as defined in Set 1; while in Figure 7 we use the Set 2.

Based on these figures, we can see the influence of the service time on the performance of the system. In both cases, the best performance was obtained by Lognormal distribution and the worst one by Weibull distribution.

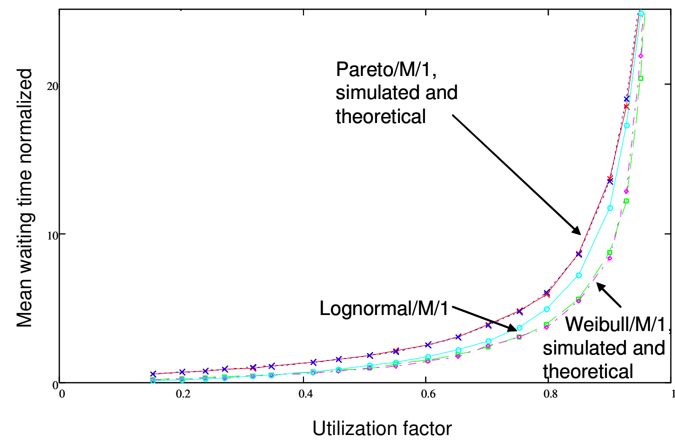

Fig. 6: *Normalized mean waiting time considering G/M/1 models – shape parameter fixed as defined in Set 1.*

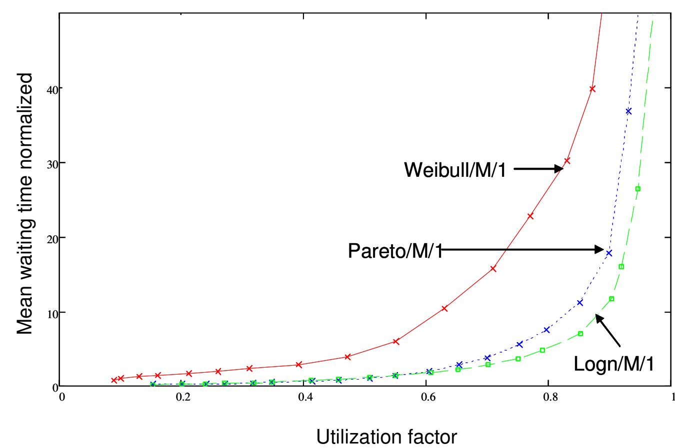

Fig. 7: *Normalized mean waiting time considering G/M/1 models – shape parameter fixed as defined in Set 2.*

Now, we present the results for Scenario II considering G/G/1 models. Figure 8 shows the performance obtained with the parameters defined as in Set 1 and Figure 9 considers the Set 2.

Again, the worst performance was obtained with Weibull distributions. The best performance depends on the set of the shape parameters: for Set 1, Lognormal offer the best performance, while Pareto has the best performance for the Set 2. Thus, we can conclude again that the service time influences the normalized mean waiting time.

Comparing the results obtained for Scenario I (figures 2, 3, 4 and 5) with the results for Scenario II (figures 6, 7, 8 and 9), we can see that the way used to vary the utilization factor of the queue has significant influence on system performance.

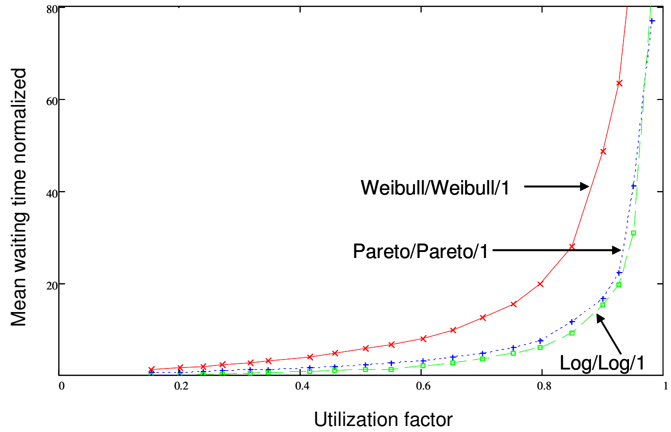

Fig. 8: *Normalized mean waiting time considering G/G/1 models – shape parameter fixed as defined in Set 1.*

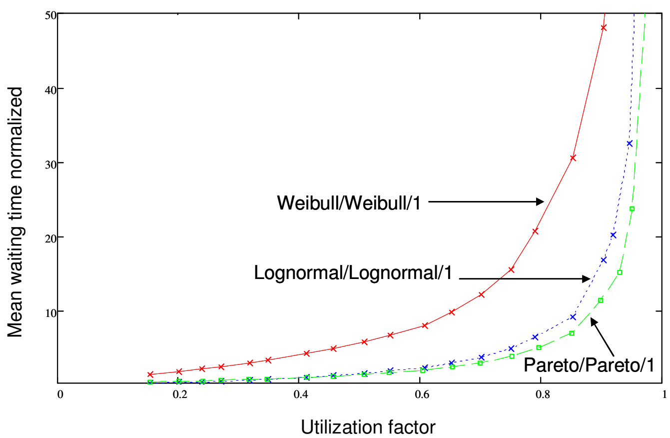

Fig. 9: *Normalized mean waiting time considering G/G/1 models – shape parameter fixed as defined in Set 2.*

VI. RESULTS FOR SCENARIO III

In the scenario analyzed in this section we define a basic traffic generator and vary the number of generators to obtain the variation of the utilization factor (see Figure 1). In this case, the shape parameters of the basic traffic generators are fixed with the same values used in Scenario I. The service time is fixed equal to 6.49 seconds for Pareto distribution and equal to 1 second for Lognormal and Weibull distributions.

Here we are interested in comparing the performance achieved in this scenario with that obtained in Scenario 1. The comparisons are shown on the next figures. Figures 10, 11 and 12 compare the performance for G/M/1 models, with G modeled by Pareto, Lognormal and Weibull, respectively. Figures 13, 14 and 15 compare the performance considering G/G/1 models (Pareto, Lognormal and Weibull, respectively).

Based on figures 10, 11 and 12, we can see that the performance of the Scenario III is better than the performance of Scenario I in most of the range of utilization factor, considering G/M/1 model.

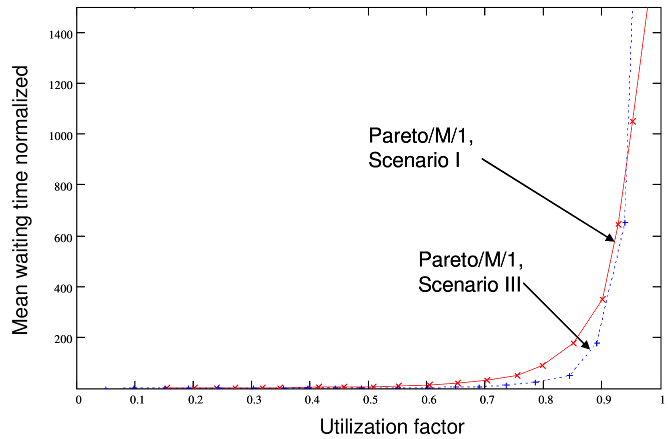

Fig. 10: *Comparing the normalized mean waiting time for Scenario I and Scenario III considering Pareto/M/1 model.*

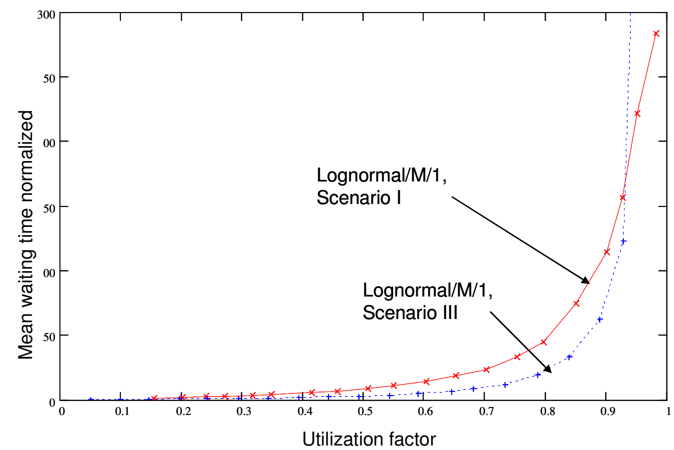

Fig. 11: *Comparing the normalized mean waiting time for Scenario I and Scenario III considering Lognormal/M/1 model.*

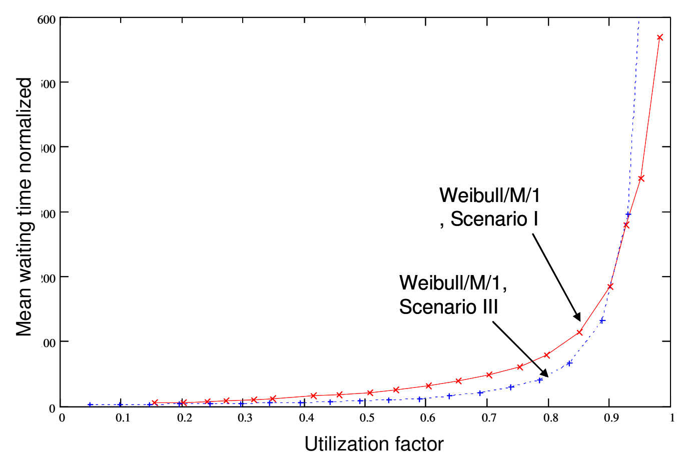

Fig. 12: *Comparing the normalized mean waiting time for Scenario I and Scenario III considering Weibull/M/1 model.*

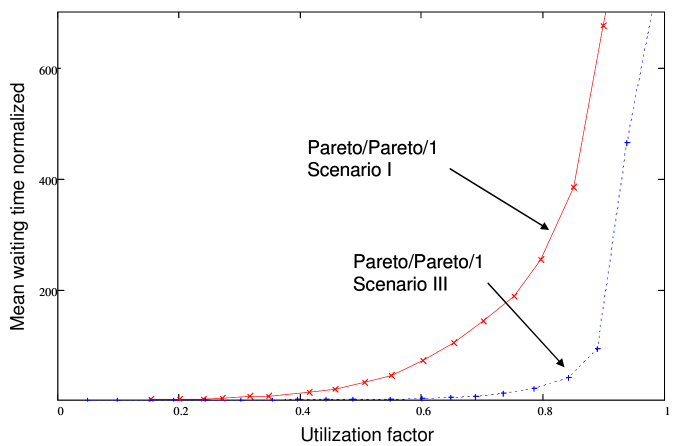

Fig. 13: *Comparing the normalized mean waiting time for Scenario I and Scenario III considering Pareto/Pareto/1 model.*

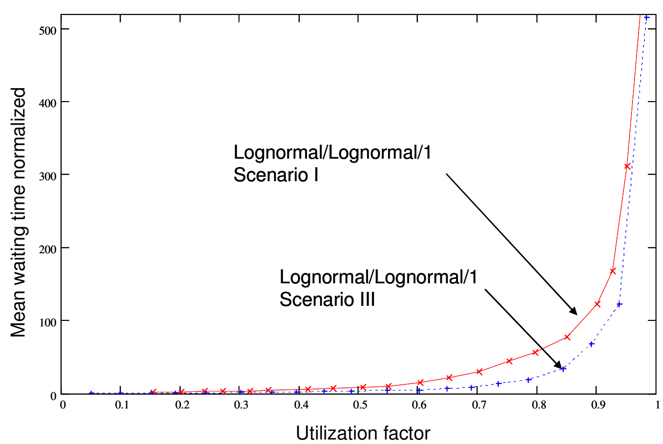

Fig. 14: *Comparing the normalized mean waiting time for Scenario I and Scenario III considering Lognormal/Lognormal/1 model.*

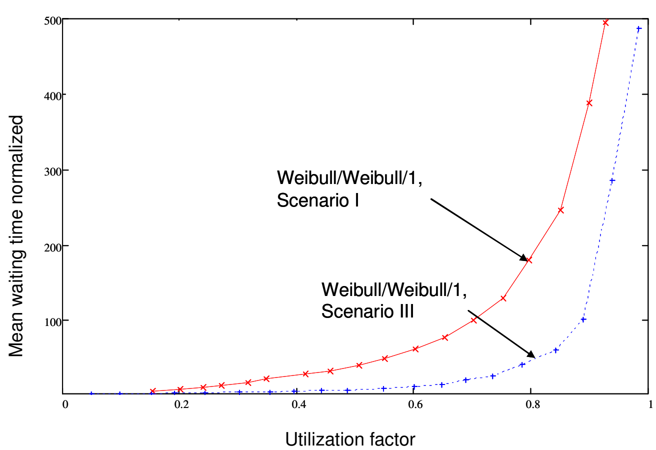

Fig. 15: *Comparing the normalized mean waiting time for Scenario I and Scenario III considering Weibull/Weibull/1 model.*

Based on figures 13, 14 and 15, we can see that, in this case, the results for Scenario III are always better than the results for Scenario I.

VII. Performance Factor

Finally, in this section we investigate the possibility to define a performance factor to compute the mean waiting time for G/M/1 systems from the results obtained for an M/M/1 system. The performance factor (δ) is a number that satisfies the following equality:

$$E\left(t_{wG/M/1}\right) \cong \delta \cdot E\left(t_{wM/M/1}\right) \qquad (11)$$

where $E(t_{wG/M/1})$ and $E(t_{wM/M/1})$ are the mean waiting time for G/M/1 and M/M/1 queues, respectively.

Figures 16 and 17 show the results for Scenario 1 and Pareto/M/1 queue. In Figure 16 the shape parameter is fixed to $\alpha = 1.3$, while in Figure 17 is $\alpha = 1.7$. In Figure 16 we use $\delta = 16$ and the performance factor can be defined only for utilization factor less than 0.75. In Figure 17 we use $\delta = 4.8$ and the performance factor is valid for any value of utilization factor. As we can see, in this case, the performance factor is a function of the shape parameter. Similar conclusions can be obtained for Lognormal and Weibull distributions [15].

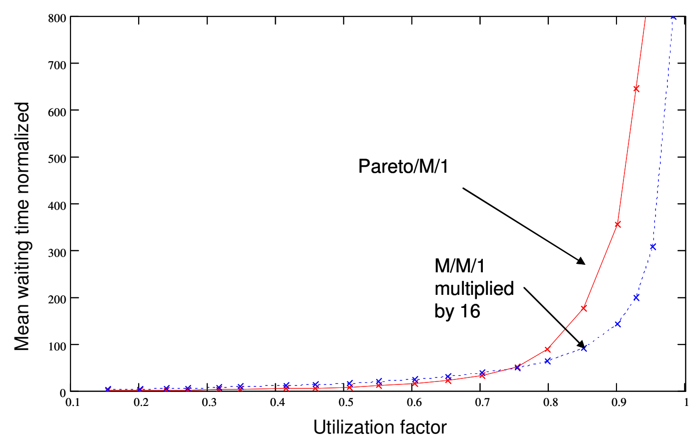

Fig. 16: *Performance factor for Scenario I and Pareto/M/1 queue with $\alpha = 1.3$.*

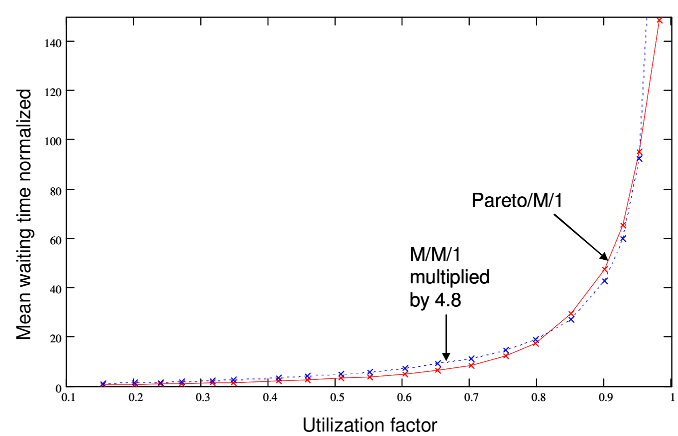

Fig. 17: *Performance factor for Scenario I and Pareto/M/1 queue with $\alpha = 1.7$.*

For Scenario II, we investigated the performance factor as a function of the departure rate (μ) of the queue (the departure rate is the inverse of the service time). We conclude that for μ greater than or equal to 10 packets/second, the performance factor is fixed and equal to 1.2. Figure 18 shows the result for Pareto/M/1 queue. The results for Lognormal and Weibull distributions can be obtained in [15].

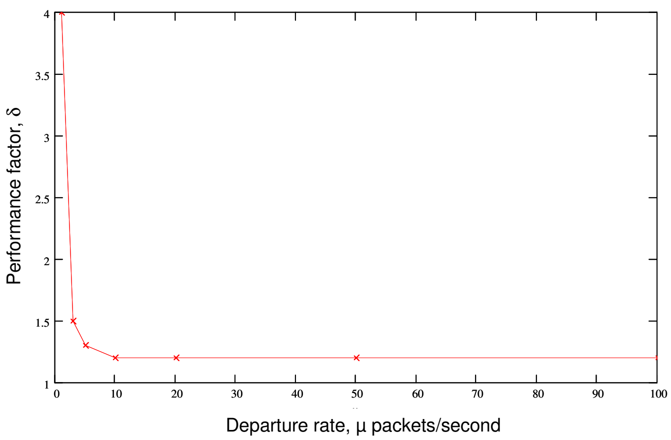

Fig. 17: *Performance factor for Scenario II and Pareto/M/1 queue.*

Finally, Figure 18 shows the performance factor for Scenario III, with the shape parameters fixed as defined on Section VII, considering Pareto/M/1 model. In this case, the performance factor can be defined for utilization factor less than or equal to 0.75 and its value is $\delta = 5$.

Similar results can be obtained for Lognormal and Weibull distributions [15].

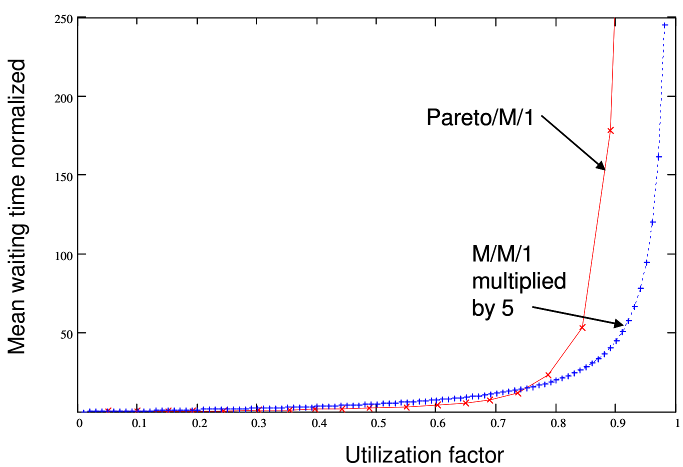

Fig. 18: *Performance factor for Scenario III and Pareto/M/1 queue.*

VIII. CONCLUSIONS

In this paper we analyzed three different approaches to characterize traffic in telecommunications networks using heavy-tailed distributions. We conclude that the normalized mean waiting time in the queue system can vary significantly, depending on the approach used.

We also introduce the performance factor concept, a factor that permits to compute the performance of a G/M/1 queue from results of M/M/1 queue. We show that this factor vary with the approach used to characterize the traffic and with the parameters in each scenario analyzed.

As a future work, we intend to complete the performance factor analysis, trying to establish its value based on a closed equation, as a function of the parameters of the heavy-tailed distribution and the service time.

REFERENCES

[1] Leland W.E. et al, "*On the self-similar nature of Ethernet traffic (Extended version)"* IEEE/ACM Transactions on Networking, Vol. 2, No. 1, Feb 1994, pp. 1-15.

[2] Jiang M., Nikolic M., Hardy S. and Trajkovic L. "*Impact of Self-Similarity on Wireless Data Network Performance.* In IEEE International Conference on Communications, june 2001, pp. 1-5.

[3] Cappe O. and Yang X. "*Long-Range Dependence and Heavy-Tail Modeling for Teletraffic Data*" IEEE Signal Proc. Magazine, Dec 2002, Vol. 19, pp. 14–27.

[4] C. M. Harris and M. J. Fischer *"Internet-Type Queues with Power-Tailed Interarrival Time and Computational Methods for their Analysis"* Published in INFORMS Journal on Computing, pp. July 2000, 261-271.

[5] Fischer M., Gross D., Masi D. M. and Shortle J. F. "*Analyzing the Waiting Time Process in Internet Queueing Systems With the Transform Approximation Method*" The Telecommunications Review, July 2001, pp. 21–32.

[6] Zaliapin V., Kagan Y. and Schoenberg F.*"Approximating the distribution of Pareto sums"* Pure Appl. Geophys., July 2005, pp. 1187–1228.

[7] Shortle J. F., Fischer M. J., Gross D. and Masi D. M. "*One-Parameter Pareto, Two-Parameter Pareto, Three-Parameter Pareto Is there a Modeling Difference?"* The Telecommunications Review May 2005, pp 79-91.

[8] Crovella M. E. and Bestavros A. "*Explaining World Wide Web Traffic Self-Similarity"* Computer Science Department, Boston university, October 1995, pp. 1-19

[9] Beaulieu C. N. *"A Simple Integral Form of Lognormal Characteristic Functions Convenient for Numerical Computation"* Communications Society subject matter experts for publication in the IEEE GLOBECOM proceedings, Dec 2006, PP. 1-4.

[10] Fernandes S., Kamienski C., and Sadok D. "*Accurate and Fast Replication on the Generation of Fractal Network Traffic Using Alternative Probability Models*". Conference on Performance and Control of Next Generation Communication Networks, part of the SPIE International Conference ITCom September 2003, pp. 7-11.

[11] Fischer M., Masi D., Gross D and Shortle J. "*Loss Systems With Heavy-Tailed Arrivals*" The Telecommunications Review, Jun 2004, pp. 1-5.

[12] Papoulis A. and Pillai S. U., *Probability, Random Variables and Stochastic Processes – Fourth Edition.* Mc-Graw Hill, 2002.

[13] Gross, D., Shortle, J.F., Fischer, M.J. and Masi "*Difculties in simulating queues with pareto service*" In Proceedings of the 2002 Winter Simulation Conference, October 2002, pp. 1-9.

[14] Shortle J. F., Fischer M. J., Gross D. and Masi D. "*Using the Transform Approximation Method to Analyze Queues"*Journal of Probability and Statistical Science, Vol 1, No. 1, April 2003, pp. 15-27.

[15] Ruas W. F. "Análise de Desempenho de Redes de Filas com Tráfego Modelado por Distribuição de Cauda Pesada" (in portuguese). Dissertation. Inatel, 2010.