Using BIC and AIC for Ethernet traffic model selection. Is it worth?

1st Anderson dos Santos Paschoalon
dept. name of organization (of Aff.)
State University of Campinas
Campinas, Brazil
anderson.paschoalon@gmail.com

2nd Christian Esteve Rothenberg
dept. name of organization (of Aff.)
State University of Campinas
Campinas, Brazil
chesteve@dca.fee.unicamp.br

***Abstract*—In this work, we aim to evaluate how good are the information criteria AIC and BIC inferring which is the best stochastic process to describe Ethernet inter-packet times. Also, we check if there is a practical difference between using AIC or BIC. We use a set of stochastic distributions to represent inter-packet of a traffic trace and calculate AIC and BIC. To test the quality of BIC and AIC guesses, we define a cost function based on the comparison of significant stochastic properties for internet traffic modeling, such as correlation, fractal-level and mean. Then, we compare both results. In this short paper, we present just the results of a public free Skype-application packet capture, but we provide as reference further analyzes on different traffic traces. We conclude that for most cases AIC and BIC can guess right the best fitting according to the standards of Ethernet traffic modeling.**

***Index Terms*—BIC, AIC, stochastic function, inter-packet times, correlation, Hurst exponent, heavy-tailed distribution, fractal-level, burstiness, linear-regression, weibull, pareto, exponential, normal, poison, maximum likelihood, Ethernet traffic, traffic modeling, fractal-level, pcap file, Skype traffic.**

I. Introduction

It is already a well-known fact that the type of traffic used on performing tests matters. Studies show that a realistic Ethernet traffic provides a different and more variable load characteristics on routers [1], even with the same mean bandwidth consumption. It indicates that tests with constant bit rate traffic generator tools are not enough for a complete validation of a new technology. There are many reasons for this behavior, which includes burstiness and packet sizes. A burstier traffic can cause packet more buffer overflows on network [2] [3] [4], degenerating network performance[1]. Also, a burstier and realistic traffic will not just impact on performance, but on measurement as well, since a realistic and burstier traffic impacts on bandwidth measurement accuracy [5] [6]. Knowing that, many traffic generator tools [7] provide a set of pre-defined stochastic models to control the emission packets, controlling packet-trains and inter-pacekt times.

There are many works devoted on studing the nature of the Ethernet traffic [8] [9] [10] [11] [12]. Classic Ethernet models use Poisson related processes, wich represents the probability of events occur in many independent sources with a known average rate, and independently of the last occurrence [8] [13]. However studies made by Leland et al. [8] showed that the Ethernet traffic has a self-similar and fractal nature. Even if they can represent the randomness of an Ethernet traffic, simple Poisson processes can't express traffic "burstiness" in a long-term time scale, such as traffic "spikes" on long-range ripples. These characteristics are an indication of the fractal and self-similar nature of the traffic that usually we express by distributions with infinite variance, called heavy-tailed. Heavy-tail means its distribution is not exponentially bounded [14], such as Weibull, Pareto and Cauchy distribution. Heavy-tailed processes may guarantee self-similarity, but not necessarily will ensure other important features like high correlation between data and same mean packet rate [15].

Therefore, there are many investigations on how to model stochastic functions for different scenarios [16] [3] [4] [17] [18] [3]. However, there are some limitations on this idea of finding a single ideal model. Usually, not the same stochastic distribution will present a proper fitting for all possible kinds of traces [14]. Depending on some variables, such as the capture time, the number of packets or type of traffic, different functions may fit better the available data. On most works the best model representation for an Ethernet traffic is not chosen analytically but based on the researcher own data analyses and purposes [19] [17] [18]. Also, some methods like linear regression may diverge sometimes. Furthermore, it has already been proven that a single model cannot represent arbitrary traffic traces [14].

In this work we propose and evaluate the use of information criteria BIC (Bayesian information criterion) and AIC (Akaike information criterion) as tool for an automatic selection best fitting for inter-packet times of a traffic trace. It is an analytical and deterministic method which spares and avoid human analyzes, is easy to be implemented by software, and don't relies on simulations and generation of random data. We fit a set of stochastic models through different methods and applying BIC and AIC to choose the best.

First, we explain the mathematical meaning of BIC and AIC and state the methods we are going to use to create a set of candidate models for our dataset. Then we define our cross-validation method based on a cost function J, attributing weights from the best to the worst representation

[1]Fuatures such as packet-trains periods and inter-packet times affect traffic burstiness

for each properties using randomly generated data with our stochastic fittings, we can choose the best possible traffic model among these fittings. This cost function depends on important stochastic properties for traffic modelling. Thus we compare the results achieved by AIC/BIC and our cost function. In that way we show that thay are a worth method for inteligent selection of inter-packet times stochastic models, being able to gess models with smaller J values. We also found that for traffic inter-packet times, that the difference between BIC and AIC values is minimal. Thus, selecting one over the other does not seem a key question.

II. Literature Review

A. Modelling Ethernet Traffic

There are plenty of works on the literature which proposes new processes and methodologies for modeling times between packets and packet trains. Fiorini [17] presents a heavy-tailed ON/OFF model, which tries to represent a traffic generated by many sources. The model emulates a multiple source power-tail Markov-Modulated (PT-MMPP) ON/OFF process, where the ON times are power-tail distributed. They achieve analytical performance measurements using Linear Algebra Queueing Theory.

Kreban and Clearwater [19] presents a model for times between job submissions of multiple users over a super computer. They show that the Weibull probability functions are able to express well small and high values of inter-job submission times. They also tested exponential, lognormal and Pareto distributions. Exponential distribution couldn't represent long-range values because it fell off too fast and Pareto was too slow. Lognormal fit well small values, but was poor on larger ones. Kronewitter [18] presents a model of scheduling traffic of many heavy-tail sources. On his work, he uses many Pareto sources to represent the traffic. To estimate the shape parameter α they use linear regression.

B. AIC and BIC

Suppose that we have an statistical model M of some dataset $\boldsymbol{x} = \{x_1, ..., x_n\}$, with n independent and identically distributed observations of a random variable X. This model can be expressed by a probability density function (PDF) $f(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameter of the PDF, $\boldsymbol{\theta} \in \mathbb{R}^k$ (k is the number of parameters). The likelihood function of this model M is given by:

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = f(x_1|\boldsymbol{\theta}) \cdot ... \cdot f(x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \qquad (1)$$

Now, suppose we are trying to estimate the best statistical model, from a set $M_1, ..., M_n$, each one with an estimated vector of parameters $\hat{\boldsymbol{\theta_1}}, ..., \hat{\boldsymbol{\theta_n}}$. AIC and BIC are defined by:

$$AIC = 2k - \ln(L(\hat{\boldsymbol{\theta}}|\boldsymbol{x})) \qquad (2)$$

$$BIC = k\ln(n) - \ln(L(\hat{\boldsymbol{\theta}}|\boldsymbol{x})) \qquad (3)$$

In both cases, the preferred model M_i, is the one with the smaller value of AIC_i or BIC_i.

III. Methodology

We start defining the *pcaps* datasets we are going to use in the rest of this text. We will use for datasets, and for reproduction purposes, three are public available. The first is a lightweight Skype packet capture, found in Wireshark wiki[2], we call it *skype-pcap*. The second is a CAIDA[3] capture, wich we analyze its first second capture[4], and we refer to it as *wan-pcap*. The third we capture in our laboratory LAN, through a period of 24 hours. It was captured firewall gateway between our local and external network. Along with other tests, We intend to verify diurnal behavior on it. That means a high demand of packets during the day and a small in the night. We call it *lan-diurnal-pcap*. Finally, the last is a capture of a busy private network access point to the Internet, available on-line on TCPreplay website [5], we will refer to it *bigflows-pcap*.

Further we collect inter-packet times from the traffic traces. Then, we estimate a set of parameters for stochastic processes, using a set of different methodologies such as linear-regression, maximum likelihood, and direct estimation of parameters. We are modelling:

- Weibull, exponential, Pareto and Cauchy distributions, using linear regression, through the Gradient descendent algorithm. We refer to these exponential and Pareto approximations as Exponential(LR) and Pareto(LR);
- Normal and exponential distribution, using direct estimation the mean and the standard deviation of the dataset for the normal, and the mean for the exponential. We refer for to this exponential approximation as Exponential(Me);
- Pareto distribution, using the maximum likelihood method. We refer to this distribution as Pareto(MLH);

Then, from these parametrized models, we estimate which one best represent our dataset, using AIC and BIC criteria. These results were obtained using Octave language[6], and the scripts are available at [20] for reproduction purposes. Thus, to see if our criterion of parameter selection can find which is the best model according to traffic modeling standards on realism and benchmarking [15], we define a validation methodology. We randomly generated a dataset using our parameterized stochastic processes. Moreover, we compare it with the original and synthetic sample, trough three different metrics, all with a confidence interval of 95%:

- Correlation between the sample data and the estimated model (Pearson's product-moment coefficient);
- Difference between the original and the synthetic Hurst exponent;

[2] Available at https://wiki.wireshark.org/SampleCaptures, named *SkypeIRC.cap*

[3] http://www.caida.org/home/

[4] Available at https://data.caida.org/datasets/passive-2016/equinix-chicago/20160121-130000.UTC, named as *equinix-chicago.dirB.20160121-135641.UTC.anon.pcap.gz*

[5] Available at http://tcpreplay.appneta.com/wiki/captures.html, named `bigFlows.pcap`

[6] https://www.gnu.org/software/octave/

- Difference between the original and the synthetic mean inter-packet time;

The Pearson's product-moment coefficient, or simply correlation coefficient, is an expression of the dependence or association between two datasets. Its value goes from -1 to +1. +1 means a perfect direct linear correlation. -1 indicates perfect inverse linear correlation. 0 means no linear correlation. So, as close the result reach 1, more similar are the inter-packet times to the original values. To estimate it, we use the Octave's function `corr()`.

The Hurst exponent is meter self-similarity and indicates the fractal level of the inter-packet times. As close the result is from the original, more similar is the fractal level of the estimated samples from the original. To estimate this value we use the function `hurst()` from Octave, which uses rescaled range method. Finally, the mean is also relevant, since it will meters if the packet per second rate of the trace and its approximation model are close related.

To quantitatively check if AIC and BIC are good criteria for model selection for inter-packet times, we define a cost function based on the correlation, Hurst exponent and mean. We defined a cost function J, based on the randomly generated data values with our estimated stochastic process. Being Cr the vector of correlations ordered from the greater to the smaller. Let Me and Hr defined by the absolute difference between mean and hurt exponent of the estimated values and the original dataset. Both are ordered from the smaller to the greatest values. Letting $\phi(V, M)$ be an operator who gives the position of a model M in a vector V, we define the cost function J as:

$$J(M) = \phi(Cr, M) + \phi(Me, M) + \phi(Hr, M) \qquad (4)$$

The smaller is the cost J, the best is the model. Then we compare the results achieved by AIC and BIC, and J.

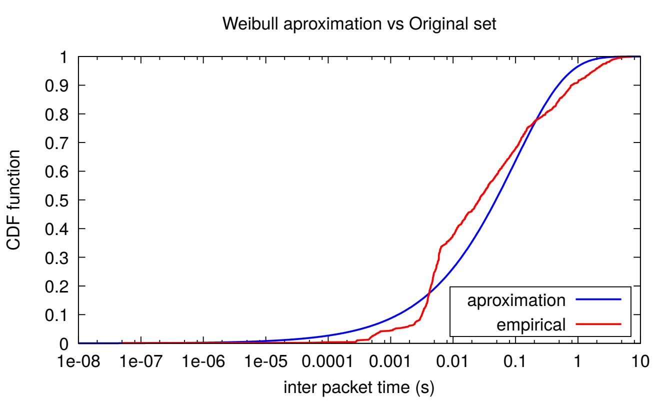


Fig. 1: Cumulative distribution function(CDF) for Weibull fitting and empirical data for inter-packet times of *skype-pcap*

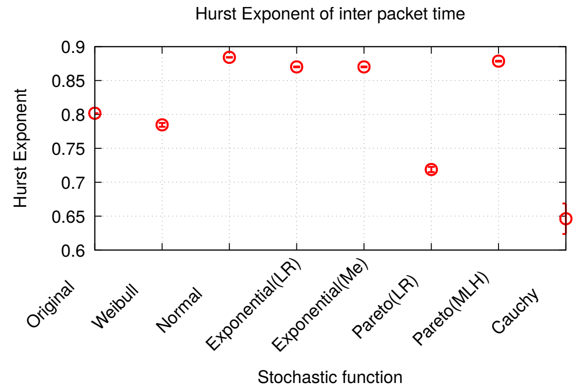


Fig. 2: Hurst exponent calculation of inter-packet times for *skype-pcap* and its approximations

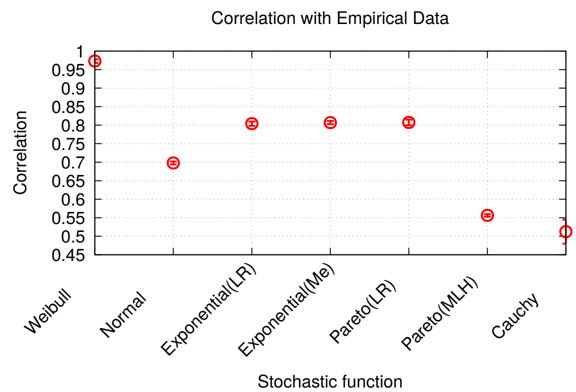


Fig. 3: Correlation between *skype-pcap* inter-packet times and its approximations.

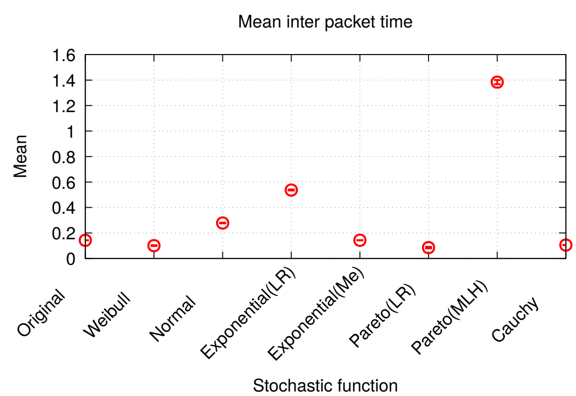


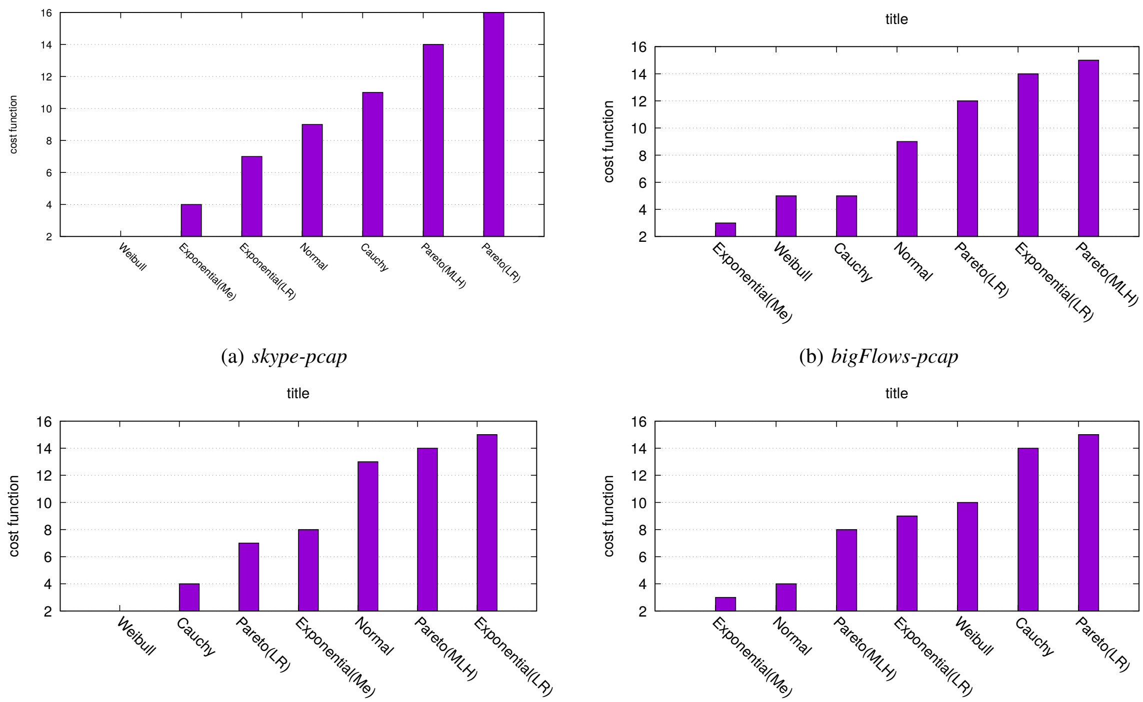
Fig. 4: Mean inter-packet times for *skype-pcap* and its approximations.

IV. Results

In table I we summarize our estimations for AIC, BIC, and the stochastic process estimated parameters for all *pcap* traces. We iserted and aditional column to facilitate the observation of the quality of fitting according BIC and AIC. For visualization, in the figure 1 we present the best fitting chosen by BIC and AIC criteria for the *skype-pcap*. It is on log-scale, which provides a better visualization for small time values. Visually we can see that linear regression with Weibull distribution was

TABLE I: Results of the octave prototype, include BIC and AIC values, para estimated parameters for two of our pcap traces: *skype-pcap* and *bigflows-pcap*. *bigflows-pcap* is much larger, and has a much much smaller mean inter-packet time

	Trace									
Function	Order	AIC	BIC	Parameters		Order	AIC	BIC	Parameters	
	skype-pcap					lan-diurnal-pcap				
Cauchy	7	$1.35e4$	$2.19e4$	$\gamma : 2.59e-4$	$x_0 : 1.05e-1$	5	$-2.85e7$	$-2.85e7$	$\gamma : 9.63e-3$	$x_0 : -3.61e-3$
Exponential(LR)	3	$9.69e1$	$1.02e2$	$\lambda : 1.86$		6	$1.79e6$	$1.79e6$	$\lambda : 8.51e-1$	
Exponential(Me)	2	$-4.26e2$	$-4.28e3$	$\lambda : 7.01$		4	$-3.12e7$	$-3.12e7$	$\lambda : 58.78$	
Normal	5	$2.42e3$	$3.31e3$	$\mu : 1.43e-1$	$\sigma : 5.01e-1$	7	Inf	Inf	$\mu : 1.70e-2$	$\sigma : 8.56e-2$
Pareto(LR)	6	$6.4e3$	$-8.27e3$	$\alpha : 2.52e-1$	$x_m : 5e-8$	3	$-4.60e7$	$-4.60e7$	$\alpha : 2.55e-1$	$x_m : 5e-8$
Pareto(MLH)	4	$3.62e2$	$3.72e2$	$\alpha : 9.21e-2$	$x_m : 5e-8$	2	$-5.03e7$	$-5.03e7$	$\alpha : 1.15e-1$	$x_m : 5e-8$
Weibull	1	$-2.29e3$	$-2.28e3$	$\alpha : 3.20e-1$	$\beta : 1.52e-2$	1	$-5.60e7$	$-5.60e7$	$\alpha : 3.34e-1$	$\beta : 1.83e-3$
	bigFlows-pcap					wan-pcap				
Cauchy	6	$7.14e6$	$7.14e6$	$\gamma : 1.94e0$	$x_0 : -7.25$	7	$5.99e7$	$5.99e7$	$\gamma : 8.28e2$	$x_0 : -4.52e3$
Exponential(LR)	7	$7.33e6$	$7.33e6$	$\lambda : 1.489e-1$		6	$5.68e7$	$5.68e7$	$\lambda : 2.2e-5$	
Exponential(Me)	2	$-1.09e7$	$-1.09e7$	$\lambda : 2.64e3$		1	$-6.58e7$	$-6.58e7$	$\lambda : 6.58e5$	
Normal	5	$-9.35e6$	$-9.35e6$	$\mu : 3.79e-4$	$\sigma : 6.60e-4$	2	$-6.39e7$	$-6.39e7$	$\mu : 2e-6$	$\sigma : 1e-6$
Pareto(LR)	4	$-1.02e7$	$-1.02e7$	$\alpha : 1.489e-1$	$x_m : 5e-8$	4	$-5.31e7$	$-5.31e7$	$\alpha : NaN$	$x_m : 5e-8$
Pareto(MLH)	3	$-1.03e7$	$-1.03e7$	$\alpha : 1.362e-1$	$x_m : 5e-8$	5	$-6.25e7$	$-6.25e7$	$\alpha : 3.39e-1$	$x_m : 5e-8$
Weibull	1	$-1.10e7$	$-1.10e7$	$\alpha : 2.81e-1$	$\beta : 5.54e-4$	3	$-5.46e7$	$-5.46e7$	$\alpha : 7.64e-2$	$\beta : 1e-6$



(a) *skype-pcap*

(b) *bigFlows-pcap*

(c) *lan-diurnal-pcap*

(d) *wan-pcap*

Fig. 5: Cost function for each one of the datasets used in this validation process

able to provide a good approximation for this dataset.

First, let's take a look on table I. On *skype-pcap* and *bigFlows-pcap*, *lan-diurnal-pcap*, the fitting pointed as the best was Weibull, except on the *wan-pcap*, where Exponential(Me) was selected. This can be explained by the fact that this one had the largest bandwidth from all three pcaps, having smaller inter-packet times with a smaller range, and closer to the mean. The difference between BIC and AIC values in all simulations is much smaller than the difference between each distribution (for most of the cases are not larger than 1%). This result indicates that for inter-packet times, using AIC or BIC to pick a model, do not influence the results significantly.

For *skype-pcap*, according to BIC and AIC previsions, Weibull and Exponential (Me and LR) are the best options, and the cost functions gave exact this same order 5. In fact, Weibull pointed as the best stochastic function by BIC and AIC, has half of the penalty imposed by the cost function J. The following models however are not in the same order since some results are flipped. But still, no opposite correspondence can be found. No result found by AIC and BIC were far from the one pointed by J.

Furthermore, for *wan-pcap* and *lan-diurnal-pcap*, the function pointed by AIC/BIC was the same as pointed by J: Exponential(Me) and Weibull respectivally. On *big-Flows-*

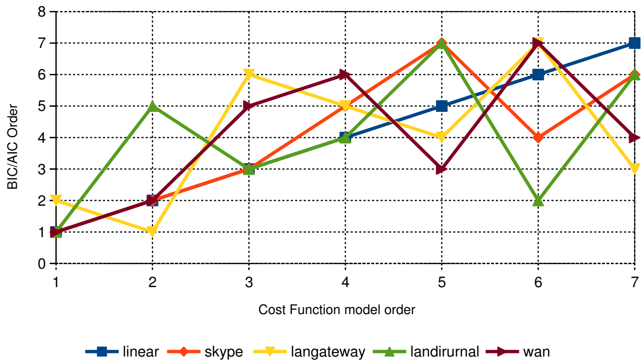


Fig. 6: Comparision of the model selection order for BIC/AIC and the cost function for each pcap.

pcap the first two gesses values are flipped, however we can notice that the difference between the AIC/BIC values were of just 7%, wich indicate a closer level of quality. A more complete chart of the selection order between both metricas are on the figure 6, relative to the J selection order. The linear patern represent the ideal results. In some cases, like the 2nd and 6th

V. Conclusion

In this work, we analyze how BIC and AIC perform being used as analytical selection criteria form stochastic models for Ethernet inter-packet times. Using a cross-validation methodology based on the generation of random data using these models, and pointing a cost function. We saw that both AIC or BIC and the cost function were able to pick the first models in the same order. Therefore, analytically with BIC and AIC, we were able to achieve the same results as pointed by our simulations. Even if AIC and BIC mathematical definitions are unaware of the specific requirements of Ethernet traffic modeling, such as same fractal-level and close packet per second rate, they still can point the best choices according to these constraints.

In this work, we analyze just inter-packet times of a single trace. However at [20] we perform the same methodology on different types of traffic captures, finding similar results. Therefore, we can conclude that BIC and AIC are healthy alternatives for model selection of Ethernet inter-packet times models and we can safely use them. Finally, we must point some advantages of BIC and AIC instead of simulations. Since it is an analytical model, no generation of random data is necessary, being computationally cheaper and easy to code. Also, since we do not use a single stochastic function and parameterization strategy, it is resilient to the fact that some methods like linear-regression over Weibull may diverge sometimes. If it happens, BIC or AIC will discard this guesses, and choose another oneautomatically.

Last but not least, to the best of our knowledge, this is the most comprehensive investigation of the actual quality of BIC and AIC as model selection criteria of for inter-packet times.

References

[1] J. Sommers and P. Barford, "Self-configuring network traffic generation," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '04. New York, NY, USA: ACM, 2004, pp. 68–81. [Online]. Available: http://doi.acm.org/10.1145/1028788.1028798

[2] Y. Cai, Y. Liu, W. Gong, and T. Wolf, "Impact of arrival burstiness on queue length: An infinitesimal perturbation analysis," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, Dec 2009, pp. 7068–7073.

[3] A. J. Field, U. Harder, and P. G. Harrison, "Measurement and modelling of self-similar traffic in computer networks," *IEE Proceedings - Communications*, vol. 151, no. 4, pp. 355–363, Aug 2004.

[4] T. Kushida and Y. Shibata, "Empirical study of inter-arrival packet times and packet losses," in *Proceedings 22nd International Conference on Distributed Computing Systems Workshops*, 2002, pp. 233–238.

[5] G. Bartlett and J. Mirkovic, "Expressing different traffic models using the legotg framework," in *2015 IEEE 35th International Conference on Distributed Computing Systems Workshops*, June 2015, pp. 56–63.

[6] K. V. Vishwanath and A. Vahdat, "Evaluating distributed systems: Does background traffic matter?" in *USENIX 2008 Annual Technical Conference*, ser. ATC'08. Berkeley, CA, USA: USENIX Association, 2008, pp. 227–240. [Online]. Available: http://dl.acm.org.ez88.periodicos.capes.gov.br/citation.cfm?id=1404014.1404031

[7] A. Botta, A. Dainotti, and A. Pescap, "A tool for the generation of realistic network workload for emerging networking scenarios," *Computer Networks*, vol. 56, no. 15, pp. 3531 – 3547, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128612000928

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

[9] F. Ju, J. Yang, and H. Liu, "Analysis of self-similar traffic based on the on/off model," in *2009 International Workshop on Chaos-Fractals Theories and Applications*, Nov 2009, pp. 301–304.

[10] Z. Rongcai and Z. Shuo, "Network traffic generation: A combination of stochastic and self-similar," in *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, vol. 2, March 2010, pp. 171–175.

[11] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similarity through high-variability: statistical analysis of ethernet lan traffic at the source level," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 71–86, Feb 1997.

[12] I. Cevizci, M. Erol, and S. F. Oktug, "Analysis of multi-player online game traffic based on self-similarity," in *Proceedings of 5th ACM SIGCOMM Workshop on Network and System Support for Games*, ser. NetGames '06. New York, NY, USA: ACM, 2006. [Online]. Available: http://doi.acm.org/10.1145/1230040.1230093

[13] F. A. Haight, *Handbook of the Poisson Distribution*. New York: John Wiley & Son, 1967.

[14] N. L. Antoine Varet, "Realistic network traffic profile generation: Theory and practice," *Computer and Information Science*, vol. 7, no. 2, 2014.

[15] S. Molnar, P. Megyesi, and G. Szabo, "How to validate traffic generators?" in *2013 IEEE International Conference on Communications Workshops (ICC)*, June 2013, pp. 1340–1344.

[16] N. M. Markovitch and U. R. Krieger, "Estimation of the renewal function by empirical data-a bayesian approach," in *Universal Multiservice Networks, 2000. ECUMN 2000. 1st European Conference on*, 2000, pp. 293–300.

[17] P. M. Fiorini, "On modeling concurrent heavy-tailed network traffic sources and its impact upon qos," in *1999 IEEE International Conference on Communications (Cat. No. 99CH36311)*, vol. 2, 1999, pp. 716–720 vol.2.

[18] F. D. Kronewitter, "Optimal scheduling of heavy tailed traffic via shape parameter estimation," in *MILCOM 2006 - 2006 IEEE Military Communications conference*, Oct 2006, pp. 1–6.

[19] S. D. Kleban and S. H. Clearwater, "Hierarchical dynamics, interarrival times, and performance," in *Supercomputing, 2003 ACM/IEEE Conference*, Nov 2003, pp. 28–28.

[20] "Projeto mestrado," https://github.com/AndersonPaschoalon/ProjetoMestrado, 1234, [Online; accessed May 30th, 2017].