

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

Anderson dos Santos Paschoalon , Christian Esteve Rothenberg

Departamento de Engenharia de Computação e Automação Industrial (DCA)

Faculdade de Engenharia Elétrica e de Computação (FEEC)

Universidade Estadual de Campinas (Unicamp)

Caixa Postal 6101, 13083-970 – Campinas, SP, Brasil

{apaschoalon, chesteve}@dca.fee.unicamp.br

**Abstract –** In this work, we aim to evaluate how good are the information criterion 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. To do so, 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 source as reference further analyzes on different traces. We conclude that for most of the cases AIC and BIC can guess right the best fitting, according to the standards of Ethernet traffic modeling.

**Keywords –** 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

## 1. Introduction

There are many works devoted to studying the nature of the Ethernet traffic[1]. Classic Ethernet models used Poisson related processes. Initially, it makes sense since a Poisson-related process represents the probability of events occur in many independent sources with a known average rate, and independently of the last occurrence [1] [2]. But studies made by Leland et al.[1] 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[3]. Examples of heavy-tailed functions are Weibull, Pareto, Cauchy. But heavy-tailed processes may guarantee self-similarity, but not necessarily they will ensure other important features like a high correlation between data and same mean packet rate.

Many investigations were made on the literature about the nature of the internet traffic[1][4][5][6][7], and many others on the modeling of stochastic functions for specific scenarios[8][9][10][11][12][9]. But there are some limitations on this idea of finding a single model. Usually, not the same stochastic distribution will present a proper fitting for all possible kinds of traces[3]. 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 of the works the best model representation for an Ethernet traffic is not chosen analytically but based on the researcher own data analyses and specific purposes[13][11][12]. Also, some methods like linear regression may diverge sometimes. And it has already been demonstrated that a single model cannot represent arbitrary traffic traces[3].

In this work test the use of information criteria BIC(Bayesian information criterion) and AIC(Akaike information criterion) as tool for choosing the best fitting for inter-packet times of a traffic trace. It is an analytical 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. On this article, we analyze the results of inter-packet time fitting for one public available trace we call as *skype-pcap*<sup>1</sup>.

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 for each

<sup>1</sup>It is a lightweight Skype capture, available at <https://wiki.wireshark.org/SampleCaptures>, named *SkypeIRC.cap*

properties using randomly generated data with our stochastic fittings, we can choose the best possible traffic model among these fittings. We compare the results achieved by AIC/BIC and our cost function. We show that BIC and AIC are good at guessing the model with the lesser  $J$  values. Also, we found that for traffic inter-packet times, that the difference between BIC and AIC values is minimal. So choosing one over the other do not seem to be a key question.

## 2. AIC and BIC

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

$$L(\theta|\mathbf{x}) = f(x_1|\theta) \cdot \dots \cdot f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) \quad (1)$$

Now, suppose we are trying to estimate the best statistical model, from a set  $M_1, \dots, M_n$ , each one whit an estimated vector of parameters  $\hat{\theta}_1, \dots, \hat{\theta}_n$ . *AIC* and *BIC* are defined by:

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

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

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

## 3. Methodology

We collect inter-packet times from the traffic capture we call *skype-pcap*. Then, we estimate a set of parameters for stochastic processes, using a set of different methodologies, including linear-regression, maximum likelihood, and direct estimation. We are modeling:

- Weibull distribution, using linear regression, through the Gradient descent algorithm;
- Normal distribution, using direct estimation the mean and the standard deviation of the dataset;
- Exponential distribution, using linear regression, through the Gradient descent algorithm. We refer to this distribution as Exponential(LR);

- Exponential distribution, using a direct estimation of the dataset mean. We refer to this distribution as Exponential(Me);
- Pareto distribution, using linear regression, through the Gradient descent algorithm. We refer to this distribution as Pareto(LR);
- Pareto distribution, using the maximum likelihood method. We refer to this distribution as Pareto(MLH);
- Cauchy distribution, using linear regression, through the Gradient descent algorithm;

Then, from these parametrized models, we estimate which best represent our dataset, using AIC and BIC criterion. These results were obtained using Octave language<sup>2</sup>, and the scripts are available at [14] for reproduction purposes. 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 generate randomly generated a same sized dataset compared to the original, using our stochastic processes fittings. Then 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;
- Difference between the original and the synthetic mean inter-packet time;

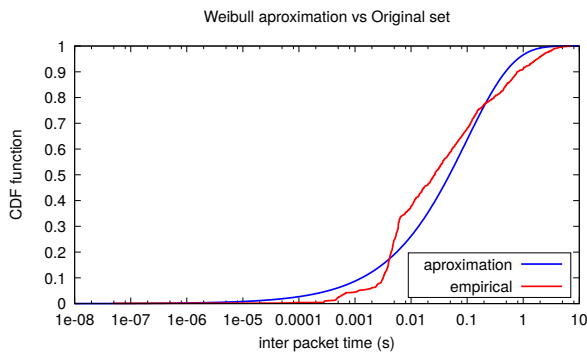
The Pearson's product-moment coefficient, or simply correlation coefficient, is an expression of the linear dependence or association between two datasets. To estimate it, we use the Octave's function `corr()`. The Hurst exponent is meter self-similarity and indicates the fractal level of the inter-packet times. 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 rate of the approximation and the original trace are close to each other.

To quantitatively check if AIC and BIC are suitable criteria for model selection for inter-packet

<sup>2</sup> <https://www.gnu.org/software/octave/>

**Table 1. Results of our modeling and simulation methodology for the traffic trace *skype-pcap*. The stochastic processes are ordered from the worst to the best fitting, according to AIC and BIC.**

Function	AIC	BIC	Parameters	
Weibull	-2293.8	-2283.8	$\alpha : 0.522$	$\beta : 0.097$
Exponential(Me)	-426.13	-421.1	$\lambda : 3.319$	
Exponential(LR)	96.9	101.8	$\lambda : 1.505$	
Pareto(MLH)	361.9	371.8	$\alpha : 0.0747$	$x_m : 5e - 8$
Normal	2423.8	2433.8	$\mu : 0.301$	$\sigma : 0.749$
Pareto(LR)	6411.0	6421.08	$\alpha : 0.413$	$x_m : 5e - 8$
Cauchy	13464.6	13474.5	$\gamma : 0.000275$	$x_0 : 0.219$



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

times, we define a cost function based on the correlation, Hurst exponent and mean. Being  $Cr$  the vector of correlations of the models ordered from the greater to the smaller. Let's define the vectors  $Me$  and  $Hr$  as the absolute difference (modulus of the difference) between mean and hurt exponent of the estimated models and the original dataset. We order both 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:

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

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

## 4. Results

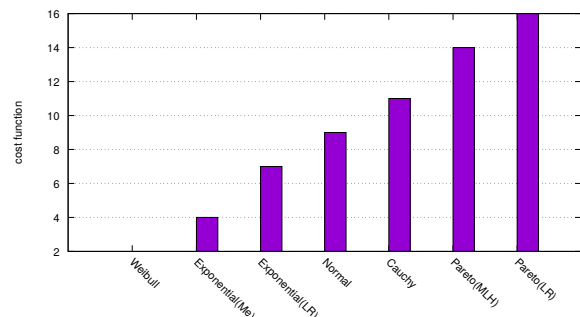
In the table 1 we summarize our estimations for AIC, BIC, and the stochastic process estimated parameters. We organize the values in crescent order of quality, according to the selection criteria (the smaller, the better). In the figure 1 we present the best fitting chosen both by BIC and AIC criteria. 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 able to provide a good approximation for this dataset.

The difference between BIC and AIC values in all simulations are much smaller than the difference between the distributions. This result indicates that for inter-packet times, using AIC or BIC to pick a model, do not influence the results significantly. According to BIC and AIC previsions, Weibull and Exponential (Me and LR) are the best options. Our cost functions points exact this order 2. 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 model is not in the same order since some results are flipped. But still, no opposite correspondence is found. No result found by AIC and BIC were far from the one pointed by  $J$ .

## 5. 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



**Figure 2. Cost function  $J$  of each stochastic process for *skype-pcap*.**

generation of random data with 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.

Here in this work, we analyze just inter-packet times of a single trace. But at [14] 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 be safely 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 liner-regression over Weibull may diverge sometimes. If it happens, BIC or AIC will discard this guesses, and choose another automatically.

## References

- [1] 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*, 2(1):1–15, Feb 1994.
- [2] Frank A. Haight. *Handbook of the Poisson Distribution*. John Wiley & Son, New York, 1967.
- [3] Nicolas Larrieu Antoine Varet. Realistic network traffic profile generation: Theory and practice. *Computer and Information Science*, 7(2), 2014.
- [4] 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*, pages 301–304, Nov 2009.
- [5] Zhao Rongcai and Zhang Shuo. Network traffic generation: A combination of stochastic and self-similar. In *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, volume 2, pages 171–175, March 2010.
- [6] 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*, 5(1):71–86, Feb 1997.
- [7] Ibrahim Cevizci, Melike Erol, and Sema 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, NetGames '06*, New York, NY, USA, 2006. ACM.
- [8] 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*, pages 293–300, 2000.
- [9] A. J. Field, U. Harder, and P. G. Harrison. Measurement and modelling of self-similar traffic in computer networks. *IEE Proceedings - Communications*, 151(4):355–363, Aug 2004.
- [10] 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*, pages 233–238, 2002.
- [11] 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)*, volume 2, pages 716–720 vol.2, 1999.
- [12] F. D. Kronewitter. Optimal scheduling of heavy tailed traffic via shape parameter estimation. In *MILCOM 2006 - 2006 IEEE Military Communications conference*, pages 1–6, Oct 2006.
- [13] S. D. Kleban and S. H. Clearwater. Hierarchical dynamics, interarrival times, and performance. In *Supercomputing, 2003 ACM/IEEE Conference*, pages 28–28, Nov 2003.
- [14] Projeto mestrado. <https://github.com/AndersonPaschoalon/ProjetoMestrado>, 1234. [Online; accessed May 30th, 2017].
- [15] Sandor Molnar, Peter Megyesi, and Geza Szabo. How to validate traffic generators? In *2013 IEEE International Conference on Communications Workshops (ICC)*, pages 1340–1344, June 2013.