Automated Selection of Inter-Packet Time Models through Information Criteria

Anderson dos Santos Paschoalon and Christian Esteve Rothenberg
School of Electrical and Computer Engineering (FEEC)
University of Campinas (UNICAMP), Campinas, Sao Paulo, Brazil
Email: anderson.paschoalon,chesteve@dca.fee.unicamp.br

***Abstract*—A well-known problem of network traffic representation over time is that there is no "one-fits-all" model. The selection of the "best" model is traditionally made in a time-consuming and ad-hoc manner by human experts. In this work, we evaluate the feasibility of using Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) as tools for automated selection of the best-fit stochastic process for inter-packet times. We propose and validate a methodology based on Information Criteria, resulting in an automated and accurate approach for such traffic modelling tasks.**

***Index Terms*—BIC, AIC, stochastic function, inter-packet times, Hurst exponent.**

I. Introduction

Traffic identification [1] and generator tools [2] [3] rely on a set of pre-defined stochastic models to set the packet classification/generation rules by configuring packet bursts and inter-packet times. Studies show that realistic network traffic provides different and more variable load characteristics on routers [4], even for the same average bandwidth. Bursty traffic can cause more packet buffer overflows on a given network [5], resulting in higher network performance degradation than under constant-rate traffic [4].

Many efforts have been devoted to understanding the traffic nature, which has been proved to be self-similar and fractal [6] [7]. Classical network traffic models based on Poisson related processes cannot express well this type of scenarios. Therefore, research has been devoted to processes with high-variability [8]. For example, the use of heavy-tailed stochastic processes, such as Weibull, Pareto, and Cauchy, have non-exponentially bounded distributions [3] and can guarantee self-similarity via Joseph and Noah effects [8]. However, they do not necessarily ensure correlation on other quality measures between the model and the actual traffic, such as the average packet rate [9]. There are works that advocate for the use of Cauchy [5], Weibull [10], Bivariate gamma [11], and Moravian-related process [12], just to cite some.

While there is an extensive amount of study-cases on network traffic modeling, there is a gap of suitable generic methods for automating the choice of the "best" model. Specific models valid for some research studies do not guarantee that the same model will apply for new cases. Investigations point to the opposite direction: a change in the scenario can change the best model as well [5] [10]. Since no "one-fits-all" model is viable, the *status quo* of traffic modeling is to be done on an *ad-hoc* manner by human specialists [13]. Another option would be to simulate all outputs a given set of random processes and choose the model that best fits the data. However, this task turns into a research project itself, involving definition of metrics, random-data generation, cross-validation methods, repetitions to guarantee high confidence intervals, and so on. Therefore, such an approach is not practical if that is not the primary research target.

In this work, we propose and evaluate the use of the Information Criteria (IC), more specifically BIC (Bayesian Information Criterion) and AIC (Akaike Information Criterion) [14], as suitable methods for automated model selection for network traffic inter-packet times. Being analytic and deterministic methods which spare model designer humans in the loop, they are also simple to implement and do not rely on hypothesis testing. In addition, We define a cross-validation method based on a cost function J, which acts as an aggregator of traditional and key metrics used for validation of stochastic models and traffic samples. J assigns weights from the best to the worst representation for each property of each trace model by using randomly generated data with our stochastic fittings. Through this process, we choose the best-fitted traffic model under evaluation. Afterward, we compare the results achieved by AIC/BIC and our cost function. Given the aforementioned approach, we show that AIC/BIC methods provide an accurate stochastic process selection strategy for inter-packet times models. Some marginal caveats include limiting our work to independent, and identical distributed random variables, since they are commonly used to describe network traffic [5] and are widely supported in traffic generators [2]. Information criteria on more complex models such as Markov-chain and envelope processes [15] have been left for future work.

II. A Primer on BIC and AIC

Let M represent a statistical model 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 the PDF's parameters, $\boldsymbol{\theta} \in \mathbb{R}^k$ (k is the $\boldsymbol{\theta}$'s dimension). The likelihood function of this model M is given by [14]:

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

The goal is to estimate the best statistical model, from the set $\{M_1, ..., M_n\}$, where each one has an estimated vector of parameters $\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_n$. AIC and BIC are defined by:

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

$$BIC = k\ln(n) - 2\ln(L(\hat{\boldsymbol{\theta}}|\boldsymbol{x})) \quad (3)$$

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

III. Methodology

We used four packet captures (*pcaps*) to extract inter-packet times we used in this work, where three of them are publicly available. The first one is a Skype packet capture[1], which we name *skype-pcap*. The second one is a CAIDA capture[2], from which we use its first capture second[3], referred to as *wan-pcap*. The third one is a capture of a busy private network Internet access point[4], which is referred as *lan-gateway-pcap*. Finally, we capture the last traffic trace at our INTRIG/UNICAMP laboratory LAN through a period of one hour on a firewall gateway. We call it *lan-firewall-pcap*. All the developed scripts and data sets are publically available [16], for reproducibility purposes. The results were obtained using the Octave tool.[5] We retrieved inter-packet times from the traffic traces and divided them into two equally sized datasets. to avoid data *over-fitting*, we use odd-indexed elements as *training dataset*, and even-indexed as *cross-validation dataset*. We then apply the *training dataset* on several techniques for model estimation:

- Weibull, exponential, Pareto and Cauchy distributions: We use linear regression through the Gradient descendent algorithm. Also, we refer to these exponential and Pareto approximations as Exponential (LR) and Pareto (LR);
- Normal and exponential distributions: We approximate the mean and variance by the average and standard deviation on the normal, and the rate by the inverse of the average on the exponential, we refer as Exponential (Me);
- Pareto distribution: We use the maximum likelihood method, which we refer to as Pareto (MLH);

Given the seven models (*hypothesis*) above, we then compute a quality ranking to evaluate AIC and BIC using the *cross-validation* dataset. To validate the information criteria effectiveness, we develop a weight system based on traditional methodologies for model quality verification and synthetic traffic validation [9] [6]. First, we randomly generate datasets following each stochastic processes hypothesis resulting in the synthetic inter-packet times, which are then compared with the *cross-validation* dataset based on the following metrics:

- The Pearson's product-moment coefficient between the sample data and the estimated model. The closer to one, the better;
- Hurst exponent estimation, via range re-scaling. The closer to the *cross-validation* Hurst value, the better;
- Average inter-packet time. The closer to the cross-validation dataset average, the better.

We choose these metrics according to traffic standards on realism and benchmarking [9]. The "Pearsons product-moment coefficient" is a measure of the correlation[6] between datasets. The Hurst exponent is a measure of self-similarity [6][7] and indicates the fractal level of the distribution of inter-packet times within a trace. Finally, a trace's average inter-packet time is inversely proportional to its packet rate. The closer the model's average inter-packet is to the original, the closer will also be its packet rate and throughput [9]. We consolidate all these metrics in a best-effort weight system, we call cost function J. Let Cr be the array of correlations between the randomly generated data and the *cross-validation dataset*, sorted from the better (greater) to the worst (smaller). Let Me and Hr be defined as vectors of absolute difference of the mean and Hurst exponent between the synthetic and the *cross-validation dataset*. These vectors are sorted: the lower the differences, the better the model hypothesis represents the same cross-validation measured metric (throughput and fractal-level). Letting $\phi(V, M)$ be an operator giving the position (starting from 0) of a model M in a vector V, we define the cost function J as:

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

To illustrate an example application, suppose a model m_1 with the best correlation, second and third smaller values of Hr and Me, respectively, would result in: $J(m_1) = 0+1+2 = 3$. Therefore, the smaller J, the better the model to represent a wide range of different metrics, since it consolidates many widely adopted metrics [9] in a single value or *ranking*. The estimation of these values was repeated 30 times, with a confidence interval of 95%, small enough to not interfere with the results. If the information criteria and J returns related results, this is interpreted as a strong indication of the reliability and robustness of AIC and BIC.

IV. Results

Table I summarizes the estimates obtained for AIC, BIC, and the stochastic process estimated parameters for all *pcap* traces. Each model order is graphically presented in Figure 1. For all *pcap* experiments, we verify that the difference between BIC and AIC for a given function is always smaller than its value among different distributions. As shown in the table I, AIC and BIC criteria always pointed to the same model ordering. Table II presents the percentage difference between the obtained values. We verify that their values tend to converge when the dataset increases.

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

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

[3]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*

[4]Available at http://tcpreplay.appneta.com/wiki/captures.html named *bigFlows.pcap*

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

[6]Octave's function `corr()`

[7]Octave's function `hurst()`, which uses the re-scaled range method.

TABLE I: Experimental results, including the estimated parameters and the BIC and AIC values of the four pcap traces.

Function	Trace							
	AIC		Parameters		AIC	BIC	Parameters	
	skype-pcap				lan-firewall-pcap			
Cauchy	$6.94E+03$	$6.95E+03$	$\gamma: 1.71E-04$	$x_0: 1.88E-01$	$-2.29E+05$	$-2.29E+05$	$\gamma: 1.93E-02$	$x_0: -4.97E-02$
Exponential(LR)	$-4.70E+01$	$-4.28E+01$	$\lambda: 1.79E+00$		$-2.22E+06$	$-2.22E+06$	$\lambda: 4.05E-01$	
Exponential(Me)	$-2.16E+02$	$-2.12E+02$	$\lambda: 3.45E+00$		$3.63E+05$	$3.63E+05$	$\lambda: 1.13E+02$	
Normal	$1.21E+03$	$1.22E+03$	$\mu: 2.90E-01$	$\sigma: 6.95E-01$	$-1.48E+06$	$-1.48E+06$	$\mu: 8.85E-03$	$\sigma: 3.49E-02$
Pareto(LR)	$3.38E+03$	$3.39E+03$	$\alpha: 4.28E-01$	$x_m: 5.00E-08$	Inf[1]	Inf[1]	$\alpha: 2.51E-01$	$x_m: 5.00E-08$
Pareto(MLH)	$1.88E+02$	$1.97E+02$	$\alpha: 7.48E-02$	$x_m: 5.00E-08$	$-1.80E+06$	$-1.80E+06$	$\alpha: 1.15E-01$	$x_m: 5.00E-08$
Weibull	$-1.15E+03$	$-1.14E+03$		$\beta: 9.68E-02$	$-1.97E+06$	$-1.97E+06$	$\alpha: 3.46E-01$	$\beta: 1.79E-03$
	lan-gateway-pcap				wan-pcap			
Cauchy	$3.65E+06$	$3.65E+06$	$\gamma: 1.95$	$x_0: -4.45E+03$	$2.99E+07$	$2.99E+07$	$\gamma: 8.17E+02$	$x_0: -4.45E+03$
Exponential(LR)	$3.67E+06$	$3.67E+06$	$\lambda: 9.75E-03$		$2.84E+07$	$2.84E+07$	$\lambda: 2.20E-05$	
Exponential(Me)	$-5.44E+06$	$-5.44E+06$	$\lambda: 2.64E+03$		$-3.29E+07$	$-3.29E+07$	$\lambda: 6.58E+05$	
Normal	$-4.67E+06$	$-4.67E+06$	$\mu: 3.79E-04$	$\sigma: 1.00E-06$	$-3.19E+07$	$-3.19E+07$	$\mu: 2.00E-06$	$\sigma: 1.00E-06$
Pareto(LR)	$-5.13E+06$	$-5.13E+06$	$\alpha: 1.49E-01$	$x_m: 5.00E-08$	$4.51E+07$	$4.51E+07$	$\alpha: 4.00E-14$[2]	$x_m: 5.00E-08$
Pareto(MLH)	$-5.13E+06$	$-5.13E+06$	$\alpha: 1.36E-01$	$x_m: 5.00E-08$	$-3.13E+07$	$-3.13E+07$	$\alpha: 3.39E-01$	$x_m: 5.00E-08$
Weibull	$-5.50E+06$	$-5.50E+06$	$\alpha: 2.81E-01$	$\beta: 1.00E-06$	$-2.73E+07$	$-2.73E+07$	$\alpha: 7.64E-02$	$\beta: 1.00E-06$

[1] The computation of the likelihood function has exceeded the computational precision used, so it was the highest AIC and BIC for this trace.
[2] The linear regression did not converge to a valid value, so we used a small value instead to perform the computations.

TABLE II: Relative difference(%) between AIC and BIC.

	skype-pcap	lan-gateway -pcap	wan-pcap	lan-firewall -pcap
Weibull	$7.47E-01$	$3.96E-04$	$8.86E-05$	$9.21E-04$
Normal	$7.04E-01$	$4.66E-04$	$7.58E-05$	NaN
Exponential(LR)	$9.54E+00$	$2.97E-04$	$4.26E-05$	$2.81E-03$
Exponential(Me)	$2.00E+00$	$2.00E-04$	$3.68E-05$	$6.90E-04$
Pareto(LR)	$2.53E-01$	$4.25E-04$	$5.36E-05$	$1.13E-03$
Pareto(MLH)	$4.45E+00$	$4.25E-04$	$7.74E-05$	$1.04E-03$
Cauchy	$1.23E-01$	$5.97E-04$	$8.08E-05$	$8.90E-03$

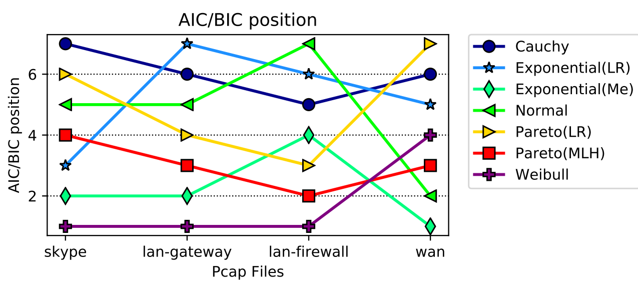

Fig. 1: Comparison of the quality order of each model given by AIC and BIC

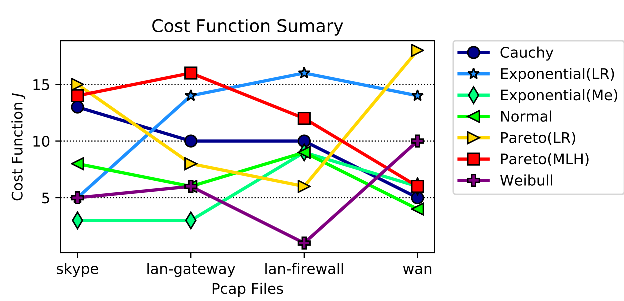

Fig. 2: Cost function for each one of the datasets used in this validation process.

Figure 2 illustrates the cost function values for all the models on each *pcap* file. For example, for *skype-pcap*, BIC and AIC point that Weibull and Exponential (Me) are the best representation for the traffic trace. The cost function used for cross-validation points both as best options, along with Exponential (LR). To simplify the visualization and comparison of the differences between the rankings given by both methodologies, Figure 3 presents a chart with the relative differences from the order of each model. Taking as a reference the position of each model given by J, we sorted them from the better to the worst (0 to 6, on the x-axis), and measured the position distance with the ones given by the information criteria. Since the worst case for this value is 6 (opposite correspondence), we draw a line on the average: the expected value in the case no positive or negative correspondence existed between both information criteria and J. Using the ϕ operator, as defined before, we can calculate the ranking delta, as explained, for the *i-th* model by:

$$\delta(m_i) = \phi(Jv, m_i) - \phi(IC, m_i) \quad (5)$$

where Jv and IC are the ordered pairs vectors on models and cost functions/information criteria, from the best to the worst, respectively. We can observe that in most cases, the information criteria and the cost function choose the best models in a similar order. A hypothesis ranked as good by one tends to be ranked also as good by the other. For the 28 possible study cases, 19 (68%) resulted in the same ranking or at most one position difference. In addition, AIC/BIC tend to prioritize most of the heavy-tailed processes, such as Weibull and Pareto (except of Cauchy). This is a useful feature when the scaling and long-range characteristics of the traffic have to be prioritized by the selected model.

Finally, we observe AIC and BIC presenting a bias in favor of Pareto (MLH). Even though it was never ranked as the best model, it was always better positioned by AIC and BIC than by J. We explain this result by the fact that AIC and BIC calculation uses the model likelihood, which Pareto (MLH) maximizes. This effect is clear on the *lan-firewall-pcap*. Figure 4 presents results from the cross-validation dataset, where we can observe the best fitting pointed by both methods (Weibull), and the second-best indicated by J (Pareto (LR))

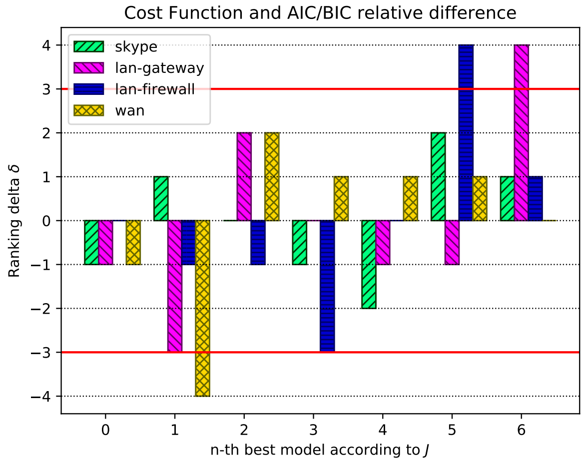

Fig. 3: Comparison of the model selection order for BIC/AIC and the cost function J for each *pcap* traffic trace.

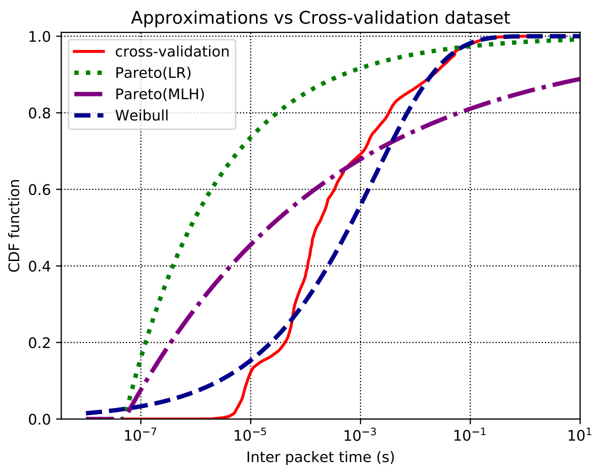

Fig. 4: Inter-packet times CDF function and stochastic models for *firewall-pacap*.

and by AIC/BIC (Pareto (MLH)). Even though Pareto (MLH) presents a good performance representing small values, about 10% of the inter-packet times are higher than 10 seconds, a prohibitive high value that overall turns Pareto (LR) into a better performing option.

V. Conclusion

This work presents and evaluates a method based on BIC and AIC for automated selection criteria of the best stochastic process to model network traffic in terms of inter-packet times. Through a cross-validation methodology based on random data generation following the selected models and cost function measurements, we observe that the proposed methodology is able to accurately pick the first models in the same order, in support of the feasibility and automation benefits of using Information Criteria as reliable model selectors for network network. We conclude that BIC and AIC are suitable alternatives to derive realistic network traffic models that could be used for diverse scenarios to add useful and efficiently add realism to experiments based on synthetic traffic generation or network traffic identification. One identified caveat is the use of the Maximum Likelihood method, which can over-prioritized some models over more performing ones. As future work, we will investigate the use of different and more complex stochastic processes such as Markovian-related and Envelope processes, beyond the scope of this article.

References

[1] M. Jaber, R. G. Cascella, and C. Barakat, “Can we trust the inter-packet time for traffic classification?” in *2011 IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–5.

[2] A. Botta, A. Dainotti, and A. Pescape, “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

[3] A. Varet and N. Larrieu, “Realistic network traffic profile generation: Theory and practice,” *Computer and Information Science*, vol. 7, no. 2, 2014.

[4] 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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] A. Bhattacharjee and S. Nandi, “Bivariate gamma distribution: A plausible solution for joint distribution of packet arrival and their sizes,” in *2010 13th International Conference on Computer and Information Technology (ICCIT)*, Dec 2010, pp. 125–130.

[12] 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.

[13] P. Tune, M. Roughan, and K. Cho, “A comparison of information criteria for traffic model selection,” in *2016 10th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Dec 2016, pp. 1–10.

[14] A. Chakrabarti and J. K. Ghosh, “Aic, bic and recent advances in model selection,” in *Philosophy of Statistics*, ser. Handbook of the Philosophy of Science, P. S. Bandyopadhyay and M. R. Forster, Eds. Amsterdam: North-Holland, 2011, vol. 7, pp. 583 – 605. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780444518620500186

[15] C. A. Melo and N. L. da Fonseca, “Envelope process and computation of the equivalent bandwidth of multifractal flows,” *Computer Networks*, vol. 48, no. 3, pp. 351 – 375, 2005, long Range Dependent Traffic. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128604003305

[16] “Github code and data set repository,” https://github.com/AndersonPaschoalon/aic-bic-paper, 2018, [Online].