Modeling Randomness in Network Traffic

Markus Laner
Vienna University of Technology, Austria
mlaner@nt.tuwien.ac.at

Philipp Svoboda
Vienna University of Technology, Austria
psvoboda@nt.tuwien.ac.at

Markus Rupp
Vienna University of Technology, Austria
mrupp@nt.tuwien.ac.at

ABSTRACT

A continuous challenge in the field of network traffic modeling is to map recorded traffic onto parameters of random processes, in order to enable simulations of the respective traffic. A key element thereof is a convenient model which is simple, yet, captures the most relevant statistics.

This work aims to find such a model which, more precisely, enables the generation of multiple random processes with arbitrary but jointly characterized distributions, auto-correlation functions and cross-correlations. Hence, we present the definition of a novel class of models, the derivation of a respective closed-form analytical representation and its application on real network traffic.

Our modeling approach comprises: (i) generating statistical dependent Gaussian random processes, (ii) introducing auto-correlation to each process with a linear filter and, (iii) transforming them sample-wise by real-valued polynomial functions in order to shape their distributions. This particular structure allows to split the parameter fitting problem into three independent parts, each of which solvable by standard methods. Therefore, it is simple and straightforward to fit the model to measurement data.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Stochastic Processes, Random Number Generation

Keywords

Gaussian Process, Polynomial Transformation

1. INTRODUCTION

Accurate characterization of network source traffic (e.g., Video streams, VoIP, online gaming) is required for various applications, such as, traffic classification or service quality assessment. For this characterization random processes provide a useful framework, by statistically describing representative characteristics of the traffic. Probability Density Functions (PDFs), Auto-Correlation Functions (ACFs) and cross-correlations are among the most convenient measures for random processes and most often used in practice. Nevertheless, from the multitude of models proposed for network traffic, only a few are able to **jointly** characterize those measures with **arbitrary** accuracy. Examples are:

SIGMETRICS'12, June 11–15, 2012, London, England, UK.
ACM 978-1-4503-1097-0/12/06.

(i) Markov Models and respective generalizations [1], for which the fitting of PDF, ACF and cross-correlations has to be performed jointly, leading to prohibitive complexity if a high quality-of-fit is aimed.

(ii) TES Models, based on uniformly distributed random processes [3]. They rely on a separation of the fitting problem into multiple independent problems, hence, the complexity only grows linearly with the requested quality.

(iii) Modified Gaussian Models, based on the sample-wise transformations of Gaussian random processes [4], which also rely on a separation of the fitting problem.

The latter two are rather equivalent in terms of fitting complexity. Further, both suffer from the fact that there is no closed form analytic expression for the ACF.

In this work we present a specific class of Modified Gaussian Models, for which we are able to derive a closed form expression for the ACF. This strongly simplifies the fitting procedure, yielding a clear advantage over TES Models.

2. THE MODEL

The functional principle of our proposed model is depicted in Fig. 1. It consists of four main units/blocks:

[1] The generation of a real-valued Independent and Identically Distributed (I.I.D.) Gaussian random process $W[n]$ with zero mean and unit variance.

[2] A weighted addition of an arbitrary number I of such processes $W_i[n]$, introducing cross-correlation between processes, without altering the distribution and I.I.D. property of the process: $X[n] = \sum_{i=1}^{I} w_i \cdot W[n]$, with the condition that $\sum_{i=1}^{I}(w_i)^2 = 1$.

[3] An Linear Time Invariant (LTI) filter $h[m]$, which introduces auto-correlation $r_{YY}[m]$ to the process but leaves the distribution unchanged: $Y[n] = h[m] \star X[n]$, subject to the condition that $\sum_{m=-\infty}^{\infty}(h[m])^2 = 1$, where $\star$ denotes the convolution operation.

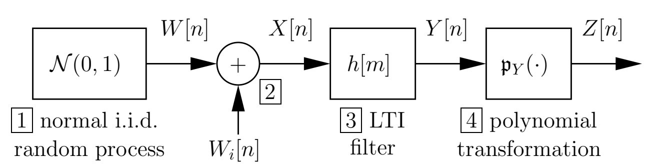

Figure 1: Proposed model for the generation of a random process with arbitrary but jointly defined PDF, ACF and cross-correlation to other processes.

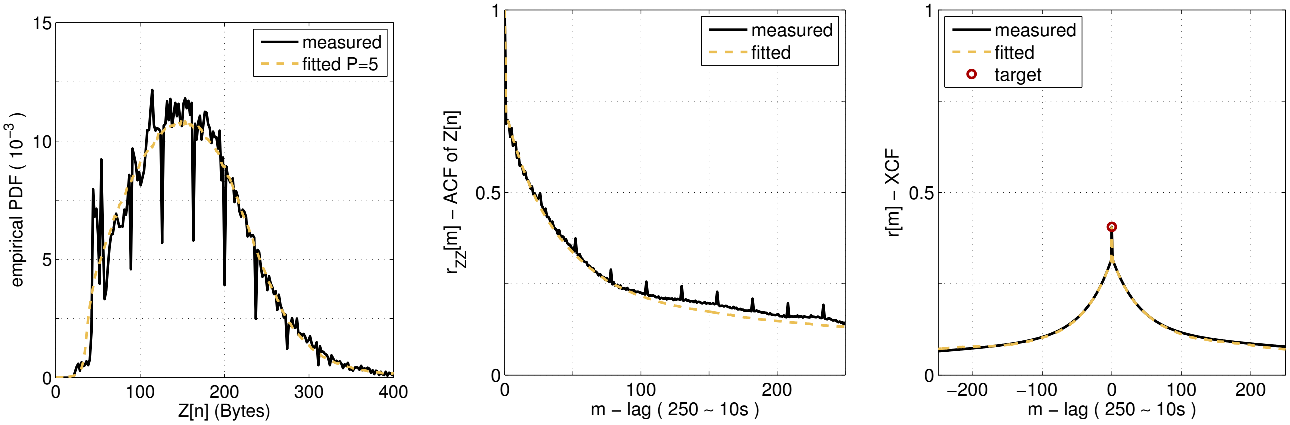

Figure 2: Data traffic of the online game *OpenArena* [5] and respective model-fit. Time series of the downlink packet size of two different players are analyzed. (left): PDF of Player 1 with fitted PDF resulting from a 5th order polynomial. (center): ACF of Player 1 with fitted ARMA(2,2) process. (right): XCF between both players. Note, that only $r_{Z_1 Z_2}[0]$ (red circle) was subject to the fitting procedure.

4 A polynomial function $\mathfrak{p}_Y(\cdot)$, transforming each sample (hence, the PDF): $Z[n] = \mathfrak{p}_Y(Y[n]) = \sum_{p=0}^{P} a_p \cdot (Y[n])^p$.

3. ANALYTICAL RESULTS

Due to the choice of a Gaussian process (closed on addition) and the conditions implied in Sec. 2, the problem of fitting this model to measurement data can be separated into the tree independent problems of fitting the distribution, the auto-correlation and the cross-correlation.

Nevertheless, the fitted parameters of Block 4 influence the fitting of Bl. 3 and, similarly, Bl. 2 is influenced by Bl. 3 and Bl. 4. This problem can be reduced to the question of *How is the auto-correlation function influenced by the polynomial*, which we are able to answer by the Theorem given below.

Note that it gives a closed form expression of the ACF with only P terms, which is crucial for efficient fitting algorithms. To the best of our knowledge, this is the first general model allowing for the exact inference of the auto-correlation function in closed form.

THEOREM: *Let $Y[n]$ denote a Gaussian random process with zero mean, unit variance and auto–correlation function $r_{YY}[m]$ and $Z[n]$ the random process obtained by the transformation of $Y[n]$ by a polynomial $\mathfrak{p}_Y(\cdot)$ according to $Z[n] = \mathfrak{p}_Y(Y[n]) = \sum_{p=0}^{P} a_p \cdot (Y[n])^p$. Then the auto–correlation function of the random process $Z[n]$ equals*

$$r_{ZZ}[m] = \mathfrak{p}_a(r_{YY}[m]) = \sum_{\rho=1}^{P} \alpha_\rho \cdot \left(r_{YY}[m]\right)^\rho,$$

where $\mathfrak{p}_a(\cdot)$ denotes a polynomial with coefficients α_ρ which, for $\rho = 1, \cdots, P$, are calculated to

$$\alpha_\rho = \frac{1}{\sigma_Z^2} \rho! \left(\sum_{p=0}^{P} a_p \cdot \binom{p}{\rho} \cdot (p-\rho-1)!! \cdot I_e(p-\rho) \right)^2,$$

where σ_Z^2 denotes the variance of $Z[n]$, $\binom{p}{\rho}$ the binomial coefficient, $(p-\rho-1)!!$ the double factorial operator and, $I_e(p-\rho)$ the indicator function for parity (1 if even, 0 if odd).

4. FITTING TO MEAUSREMENT DATA

Due to the splitting the fitting problem into three independent standard fitting procedures, the whole process is straightforward and profits form the variety of available literature and software support. More precisely, the remaining problems are: polynomial curve fitting, ARMA modeling and the Cholesky decomposition. A detailed description of the fitting procedure as well as a software implementation can be found at [2]. As a proof of concept Fig. 2 shows the model fitted to measured online gaming traffic.

5. CONCLUSIONS

This study presents a novel class of models, which allows for the convenient generation of random processes with arbitrary but jointly defined distributions, auto-correlation functions and cross-correlations.

Compared to other models our approach allows for strongly simplified parameter fitting, due to the availability of closed form expressions for all parameters, which is unique in this field.

By fitting the model to measurements, the number of parameters can be kept small, which is vital for dissemination and reproducibility. Further, the generation of random samples according to specified model parameters shows very low complexity.

6. REFERENCES

[1] A. Dainotti et al. Internet traffic modeling by means of Hidden Markov Models. *Elsevier Computer Networks*, 52:2645–2662, 2008.

[2] Institute of Telecommunicaions, Vienna University of Technology. http://www.nt.tuwien.ac.at/downloads.

[3] B. Melamed. An Overview of TES Processes and Modeling Methodology. In L. Donatiello, editor, *Performance Evaluation of Computer and Communication Systems*, chapter 1. Springer, 1993.

[4] V. Ogorodnikov, S. Prigarin, and A. Rodionov. Quasi-Gaussian Model of Network Traffic . *Springer Automation and Remote Control*, 71(3):117–130, 2010.

[5] OpenArena. http://openarena.ws.