Non Gaussian and long memory statistical characterisations for Internet traffic with anomalies

Antoine Scherrer, Nicolas Larrieu, Philippe Owezarski, Pierre Borgnat, Patrice Abry

Laboratoire de l'Informatique du Parallélisme

École Normale Supérieure de Lyon
Unité Mixte de Recherche CNRS-INRIA-ENS LYON-UCBL n° 5668

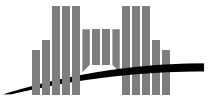
Non Gaussian and Long Memory Statistical Characterisations for Internet Traffic with Anomalies

Antoine Scherrer (LIP),
Nicolas Larrieu (LAAS),
Philippe Owezarski (LAAS)
Pierre Borgnat (Physics Lab),
Patrice Abry (Physics Lab)

September 2005

Research Report N° 2005-35

École Normale Supérieure de Lyon
46 Allée d'Italie, 69364 Lyon Cedex 07, France
Téléphone : +33(0)4.72.72.80.37
Télécopieur : +33(0)4.72.72.80.80
Adresse électronique : `lip@ens-lyon.fr`



Non Gaussian and Long Memory Statistical Characterisations for Internet Traffic with Anomalies

Antoine Scherrer (LIP), Nicolas Larrieu (LAAS), Philippe Owezarski (LAAS)
Pierre Borgnat (Physics Lab), Patrice Abry (Physics Lab)

September 2005

Abstract

The Internet aims at providing a wide range of services for a large variety of applications. Hence, it is highly sensitive to traffic anomalies (e.g., failures, flash crowds,...) as well as to DoS attacks, which are likely to significantly reduce the Quality of Service level. Current intrusion detection systems, specially those based on anomaly detection, are not providing efficient nor satisfactory solutions for DoS attack tracking. This is mainly due to difficulties in distinguishing between strong but legitimate traffic variations and DoS attack induced changes. The goal of this work is to compare relevant statistical characteristics of regular traffic to those of traffic presenting anomalies. To do so, we introduce a non Gaussian long memory model and develop estimators for the corresponding parameters. First, we show that this model relevantly describes Internet traffic for a wide range of aggregation levels, using both a large set of data taken from public reference repositories (Bellcore, LBL, Auckland, UNC, CAIDA) and data collected by ourselves. Second, we show that the proposed model also describes meaningfully traffic with anomalies such as flash crowd and DoS attacks which we generated and collected. We show that the behaviors of the parameters of the model enables us to discriminate between regular and anomalous traffic, and between flash crowds and DoS attacks. We also derive analytically procedures to numerically synthesize realizations of stochastic processes with prescribed non Gaussian marginals and long range dependent covariances. This enables us to validate the relevance and accuracy of our characterization procedures. Finally, we describe various applications based on the proposed model.

Keywords: Intrusion detection, DoS attack, Flash Crowd, Non Gaussian Long Range Dependent Process.

Résumé

Internet fourni de multiples services de communication de qualités variées à un très grand nombre d'applications. C'est pourquoi il est très sensible aux anomalies du trafic (dysfonctionnement, flash crowds,...) ainsi qu'aux attaques DoS (déni de service) qui réduisent sensiblement la qualité de service fournie par le réseau. Les mécanismes actuels de détection d'intrusion, notamment ceux basés sur la recherche d'anomalies, ne parviennent pas à lutter efficacement contre les DoS car il est très difficile de différencier les variations légitimes de trafic des variations dues à des attaques. Nous présentons ici une étude statistique de ces différentes variations (légitime et illégitime), dans le but d'aider les détecteurs d'intrusion à être plus efficace. Pour cela, nous introduisons un modèle stochastique de trafic basé sur des processus à longue mémoire non gaussien. Nous montrons que ce modèle parvient à capturer les caractéristiques statistiques d'ordre 1 et 2 du traffic de manière satisfaisante sur des traces de trafic classiques (Bellcore, LBL, Auckland, UNC, CAIDA), ainsi que sur des traces de trafic contenant des anomalies. Nous montrons aussi que les variations des paramètres de ce modèle permettent de différencier les flash crowds et les attaques DoS. Nous présentons enfin une méthode pour générer des traces de trafic synthétiques en accord avec le modèle, ce qui nous permet de valider les performances de nos outils d'analyse.

Mots-clés: Détection d'intrusion, Attaques DoS, Flash Crowd, Processus à longue mémoire non gaussiens.

1 Introduction

The Internet is on the way of becoming the universal communication network for all kinds of information, ranging from the simple transfer of binary computer data to the real time transmission of voice, video, or interactive information. Simultaneously, the Internet is evolving from a single best effort service to a multi-service network, a major consequence being that it becomes highly sensitive to attacks, specially denial of services (DoS) and distributed DoS ones (DDoS). Indeed, DoS attacks are responsible for large changes in traffic characteristics which may in turn significantly reduce the quality of service (QoS) level perceived by all users of the network. This hence breaks the service level agreement (SLA) at the Internet service provider (ISP) fault. DoS attacks can then induce financial losses for ISPs.

Reacting against DoS attacks is a difficult task and current intrusion detection systems (IDS), specially those based on anomaly detection, often fail in detecting them efficiently. This is first because DoS attacks can take a wide variety of multiple forms so that proposing a common definition is in itself a complex issue. Second, it is commonly observed that Internet traffic under normal conditions presents *per se* or *naturally* large fluctuations and variations in its throughput at all scales [34], often described in terms of long memory [13], self-similarity [35], multifractality [15]. This significantly impairs the detection of anomalies. Third, Internet traffic may exhibit strong, possibly sudden but legitimate variations (flash crowds, for instance) that may be hard to distinguish from illegitimate ones.
For those reasons, anomaly detection based IDS are often suffering from high false alarm rates, a major shortcomings for their actual use. The current evolution of the Internet traffic, allowing a larger variety of traffic and diversity of communication should result in an increase of difficulties in efficient IDS design.

This work, lead in the framework of the METROSEC project, mainly aims at analyzing the impact of anomalies on traffic statistical characteristics as well as at determining discriminative profile signatures for traffic containing legitimate (e.g., flash crowds) and illegitimate (e.g., DDoS attacks) anomalies. At end, it is expected that these findings could serve for improving networking schemes to react against traffic anomalies and particularly malicious ones.

To do so, we propose the use of non Gaussian long memory stochastic processes to model Internet traffic. We show experimentally that this model proves versatile enough to provide us with a relevant statistical description for a large variety of regular traffic as well as for traffic with anomalies, be they legitimate or not.
We also show that the evolution with respect to the aggregation level of the fitted parameters for the proposed model enables us to distinguish between traffic with and without anomalies and to classify them.

• **Outline.** Section 2 briefly reviews traffic modeling with a detection of anomalies perspective. Section 3 describes the analyzed data which consist both of standard reference traces taken from publicly available traffic data repositories and of data collected by ourselves. In particular, it explains how DDoS attacks and flash crowds were conducted by ourselves in a controlled, reproducible and chosen manner as well as how the corresponding traffic was collected by ourselves. Section 4 introduces the proposed model and describes the practical analysis procedures. Section 5 presents with details the statistical characterizations and results obtained on traffic with and without anomalies and proposes some discussions and interpretations. Section 6 explains an exact method to generate numerical time series whose marginals and covariances follow that of the observed traffic, and it validates the analysis and modeling procedures used in the previous section. Section 7 presents applications of the proposed modeling and analysis: traffic with prescribed statistics generator, traffic prediction and anomaly or attack detection.

2 Traffic Modeling : A Brief Overview

2.1 Traffic without anomalies

• **Statistical modeling and network performance.** Computer network traffic consists of packets arrival processes. It has long been recognized that those arrival processes depart from Poisson or renewal processes, see for instance [37], insofar as the inter-arrival delays are not independent. This can be modeled using, either non stationary Point processes [24] or stationary Markov modulated Point processes [3].

Therefore, a general description is in terms of marked point processes $\{(t_l, A_l), l = 0, 1, 2, ...\}$ where

the t_l denotes the arrival time stamp of the l-th packet and A_l some attributes of the packet (such as its payload, its application/source/destination ports,...). However, given the huge number of packets involved in any computer network traffic, this would result in huge data sets.
This is why it is most often preferred to work on byte or packet aggregated count processes, denoted $W_\Delta(k)$ and $X_\Delta(k)$. They consists of the number of bytes (resp., packets) that lives within the k-th window of size $\Delta > 0$, i.e., whose time stamps lie between $k\Delta \leq t_l < (k+1)\Delta$. It has also been proposed to model at the connexion or flow level instead of the packet one. In that case, one groups IP packets with identical standard 5-tuples into connections or given any other definition for more generic flows, and one models the flow arrival process or, for example, the fluctuation of the number of active flows [4]. In the present work, we remain at the packet level and mainly concentrate on the modeling of $X_\Delta(k)$. When modeling X_Δ, one mostly uses stationary processes and it is commonly accepted that marginal distributions and auto-covariance functions are the two major factors that affect the performance of the network and hence need to be modeled in priority. Reviews on traffic models can be found in, e.g., [31, 35].

• **Marginal Distributions.** Because $X_\Delta(k)$ is designed from a packet arrival process, it has been proposed to model its marginal via a Poisson distribution [14]. As $X_\Delta(k)$ is by definition a positive random variable, other works proposed to describe its marginal with common positive laws such as (one-sided) exponential, log-normal, Weibull or gamma distributions [31]. In many cases for highly aggregated data, Gaussian distributions are also used as relevant approximations.

• **Covariance Function.** After the seminal work reported in [26], it has been commonly accepted that computer network traffic is characterized by a long memory or long range dependence property (cf. [7, 40]). Long range dependence (LRD) is usually defined through the fact that the power spectral density $f_{X_\Delta}(\nu)$ of the process behave at the origin as:

$$f_{X_\Delta}(\nu) \sim C|\nu|^{-2d},\ |\nu| \to 0, \quad \text{with } 0 < d < 0.5. \tag{1}$$

Let us recall that the power spectral density is the Fourier transform of the auto-covariance. Long range dependence in computer network traffic is a central property as it is likely to be responsible for decreases of the QoS as well as of the performance of the network (see e.g., [43]). Taking it into account precisely is a necessary condition to perform accurate and relevant network design (buffer size,...) and performance predictions (delay as a function of utility,...). It is hence crucial that long memory be incorporated in description models. This rules out the use of processes such as Poisson or Markov processes as well as some of their declinations, Markov Modulated Poisson Processes for instance [36]. Consequently, canonical long range dependent processes such as fractional Brownian motion, fractional Gaussian noise [33] or Fractionally Integrated Auto-Regressive Moving Average (FARIMA) have ben widely used to describe and/or analyze Internet times series (see [35] and the references therein). It is also interesting to note that long memory can be incorporated directly into point processes using cluster point process models, yielding fruitful description of the packet arrival processes as pointed out in [18].

• **Higher Orders.** Multifractal processes are used to model scaling properties that are not fully captured by the second statistical order and involve higher ones. The relevance and interest of multifractal processes for the modeling and analysis of the Internet has been discussed in details in numerous papers and is still an open issue (see [11, 15, 19, 41, 42, 52]). It turns out that the *joint* modeling of the first and second order statistics is in itself already a difficult task. We will show in next sections that it also provides us with rich enough model to detect and characterize anomalies in traffic. Therefore, in the present work, we concentrate only on the joint modeling of the first and second statistic orders by means of non Gaussian long range dependent processes.

• **Aggregation Level.** A recurrent issue in traffic modeling lies in the choice of the relevant aggregation level Δ. This is an involved question whose answer mixes up the characteristics of the data themselves, the goal of the modeling as well as technical issues such as real time, buffer size, computational cost constraints. Facing this difficulty of choosing a priori Δ, it is of great interest to have at disposal a statistical model that may be relevant for a large range of values of Δ, a feature we seek in the model proposed below.

Data	Date	T (s)	Network(Link)	PKT	IAT	Repository
PAUG	1989-08-29(11:25)	2620	LAN(10BaseT)	1	2.6	ita.ee.lbl.gov/index.html
LBL-TCP-3	1994-01-20(14:10)	7200	WAN(10BaseT)	1.7	4	ita.ee.lbl.gov/index.html
AUCK-IV	2001-04-02(13:00)	10800	WAN(OC3)	9	1.2	wand.cs.waikato.ac.nz/wand/wits
CAIDA	2002-08-14(10:00)	600	Backbone(OC48)	65	0.01	www.caida.org/analysis/workload/oc48/
UNC	2003-04-06(16:00)	3600	WAN(10BaseT)	4.6	0.8	www-dirt.cs.unc.edu/ts/
METROSEC-ref1	2004-12-09(18:30)	5000	LAN(10BaseT)	3.9	1.5	www.laas.fr/METROSEC/
METROSEC-ref2	2004-12-10(02:00)	9000	LAN(10BaseT)	2.1	4.3	www.laas.fr/METROSEC/
METROSEC-DDoS	2004-12-09(20:00)	9000	LAN(10BaseT)	6.9	1.3	www.laas.fr/METROSEC/
METROSEC-FC	2005-04-14(14:30)	1800	LAN(10BaseT)	3.7	0.48	www.laas.fr/METROSEC/

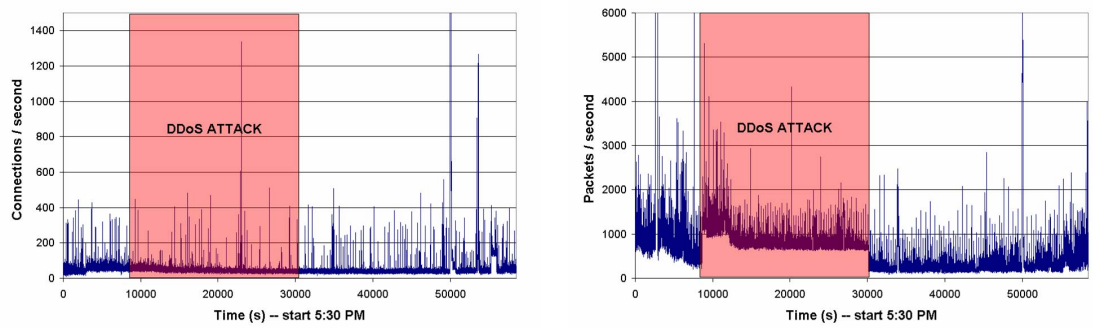
Table 1: **Data Description.** General parameters of the studied traces. T is the duration is the trace, in second. PKT is the number of packets in the trace, given in million. IAT is the mean inter-arrival time, in ms.

2.2 Anomaly Detection

IDS based on anomalies detection generally do not rely on the use of tools involving rich statistical models. They are mainly based on monitoring simple traffic parameters such as its throughput or packet rate, and most IDS make use of specific packet sequences known as attack signatures [38]. Essentially, alarms are raised whenever a threshold is reached [8, 17, 21, 45], yielding a significant number of false positives [32]. Therefore, they often remain unsatisfactory as they cannot discriminate between legitimate traffic variations and attacks.

Recently, progresses in traffic modeling obtained in various Internet traffic monitoring projects, significantly renewed IDS design strategies. Though still at early stages of developments, interesting results making use of statistical characterizations were published. For instance, Ye proposed in [50] a Markov model for the temporal behavior of the traffic, that raises alarms whenever the traffic significantly deviates from the model. Other authors [22, 51] have shown that DDoS attacks increase correlations in traffic, and this could provide a robust detection technique. Based on inter-correlation between traffic across different links, Lakhina *et al.* proposed a method for detecting network wide anomalies in traffic matrices [25]. Hussain and co-authors used the spectral density to find out signatures for various attacks [20]. Similarly, spectral estimation was used for comparing traffic with and without attacks [9]. While the spectral density exhibits peaks around the Round Trip Time values for normal traffic, such peaks no longer appear under attacks. This fact can then be used in IDS design. Finally, Li and Lee used the wavelet technique developed in [47] to compute a so-called energy distribution; it was observed that this energy distribution presents peaks under attacks that do not exist for regular traffic [27]. The work in [5] exploits the multiresolution nature of the wavelet decomposition to track and detect traffic anomalies in a so-called medium range of scales.

A collection of promising works have hence already been published in the field of DDoS attacks [5, 23]. The present work intends to contribute to that area of research by comparing statistical characterizations obtained first on regular traffic, second on traffic presenting a variety of anomalies including legitimate ones. A goal at end is to be able to discriminate between legitimate and illegitimate changes in traffic.



(a) (b)

Figure 1: **DDoS Attack.** (a) Packets and (b) connections per second time series.

3 Data and Experiments

3.1 Traffic without anomalies

The model and analysis proposed hereafter are first illustrated on traffic presenting a priori no anomaly, described in details in Table 1. We use both standard data, gathered from most of the major available Internet traces repositories (Bellcore, LBL, UNC, Auckland Univ, Univ North Carolina, CAIDA), and time series collected by ourselves within the METROSEC research project. Therefore, we cover a significant variety of traffic, networks (Local Area Network, Wide Area Network,... and edge networks, core networks,...) and links, collected over the last 16 years (from 1989 to 2005). On each repository, a large number of traces are available, we focussed here on one (or a few ones) that are representative of a collection of others. **PAUG** corresponds to one of the celebrated early Bellcore Ethernet LAN traces, over which long range dependence was first evidenced [26]. **LBL-TCP-3** is provided by the Lawrence Berkeley Laboratory and was collected at LAN gateways. Multifractal models were validated for the first time in computer network traffic on these data [15, 46]. **AUCK-IV** constitute high precisions TCP/IP traces gathered at the Internet access point of the University of Auckland over a non saturated link. We also processed the **CAIDA** time series, another high time-stamp precision, collected over a large backbone, kindly made available by CAIDA from their MFN network. **UNC** corresponds, to our knowledge, to the most recent publicly available data. The **METROSEC** data were collected late 2004 and early 2005, on the RENATER[1] network using DAG systems [10] deployed in the framework of two French research projects: METROPOLIS and METROSEC.

3.2 Traffic (or traces) with anomalies

Analyzing standard RENATER traffic, the only illegitimate activity we found is port scanning, but no DDoS attack. Therefore, we decided to perform both legitimate (Flash Crowds) and illegitimate (DDoS attacks) anomalies in an accurate and completely controlled manner. For each case, several experiments were run on RENATER network.

• **DDoS Attack.** The DDoS attack studied here consists in a distributed UDP flooding attack. It has been generated from 5 different sites: IUT Mont de Marsan, LIAFA Paris, ENS Lyon, ESSI Nice, France and the university of Coimbra, Portugal, against the target site, LAAS, Toulouse, France. LAAS is connected to RENATER thanks to a 100 Mbps Ethernet link which has not been saturated during the attack. A traffic trace is captured on the LAAS access link. The attack started at 8 pm on December 9th, 2004 and lasted more than 5.30 hours. The basic traffic characteristics are depicted in Figures 2.2(a) and 2.2(b) which respectively shows the number of flows and packets on the LAAS access link. While the former remains quite stable, the later presents a significant increase (the packet rate is multiplied by almost 3 during the attack).

• **Flash Crowd (FC).** To compare the impact on the traffic characteristics of DDoS attacks to that of legitimate traffic changes, we created flash crowds on a web server. To make them more realistic, i.e., humanly random, we chose not to use an automatic program, but to ask many French academics to browse the LAAS website (http://www.laas.fr). Results will be presented on the flash crowd performed on April 14th, 2005 which lasted 30 minutes and gathered more than a one hundred participants. Fig. 3.2(a) shows the number of requests (HTTP GET requests) received by the LAAS web server, distinguishing between inner and outer requests. One can clearly see that many people started browsing the LAAS web server at 2.30pm (important increase of the number of hits), but also that many of them did not participate for the whole 30 minutes. The Figures 3.2(b) and 3.2(c) show respectively the number of flows and the packet rate on the LAAS access link. As expected, both plots present an increase of respectively the average number of flows and the average packet rate during the flash crowd. However, one also notices an important increase of the number of flows and packet rate after the end of the flash crowd experiment. Fig. 3.2(c) also shows an increase of the average packet rate before the flash crowd experiment. To understand that increase, we analyzed the different components of the traffic using the QoSMOS Traffic Designer tool [39] (cf. Fig. 3.2(d)). It reveals that the increase recorded around 2 pm (i.e., before our experiment) is due to people inside LAAS browsing the web right after lunch. Such a pattern has been observed systematically on all traces collected on the LAAS access link since then.

[1] RENATER is the French network for education and research that interconnects academics and some industrial partners.

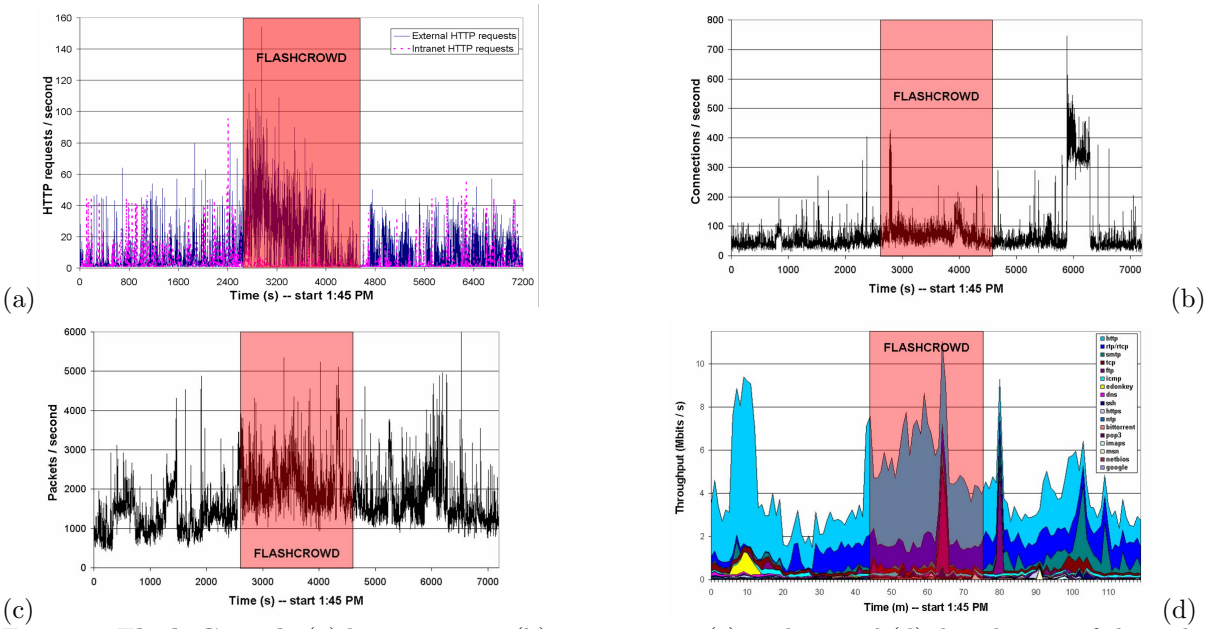


Figure 2: **Flash Crowd.** (a) http requests, (b) connections, (c) packets and (d) distribution of throughputs per second time series. Fig. (d) follows a top-down approach: the application on top generates the larger traffic.

The second peak, after the experiment, appears to be due to SMTP traffic. It can be explained in two ways. First, many researchers at LAAS use webmail. Because the server has been significantly slowed down during the flash crowd experiment, they had stopped sending e-mails until the web server restarted to work with satisfactory performance. Second, the grey listing mechanism (used for spam reduction) delays some e-mails, and sends them all at scheduled door opening. The nearest one occurred at 3.15 pm, just after the end of the flash crowd.

4 Non Gaussian Long Range Dependent Processes

4.1 The Gamma farima model

In the present work, we propose to model the $\{X_\Delta(k), k \in \mathbb{Z}\}$ time series for each aggregation level Δ. Modeling the $\{W_\Delta(k), k \in \mathbb{Z}\}$ time series gives equivalent results and we restrict here the discussion for the sake of clarity. The model is a non Gaussian, long range dependent, stationary process: the Gamma (marginal) farima (covariance) process. Stationarity is assumed for the model because it is convenient from a theoretical point of view, and we empirically check during the analysis that the hypothesis is valid.

• **First order Statistics (Marginals): Gamma distribution.** Because of the point process nature of traffic, Poisson and or exponential distributions are expected at small aggregation levels Δ, while Gaussian laws are at larger Δs. None of them can satisfactorily model traffic marginals for a wide range of (small and large) Δs. From our empirical studies, we found out that a Gamma distribution $\Gamma_{\alpha,\beta}$ capture best the marginals of the X_Δ.
A $\Gamma_{\alpha,\beta}$ distribution is defined for positive random variables X with

$$\Gamma_{\alpha,\beta}(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \tag{2}$$

where $\Gamma(u)$ is the standard Gamma function (see e.g., [14]), depends on two parameters: the shape α and the scale β. It has mean $\mu = \alpha\beta$ and variance $\sigma^2 = \alpha\beta^2$. Note that the inverse of the shape parameter, $1/\alpha$, acts as an indicator of the distance from a Gaussian law. For instance, skewness and kurtosis (relative third and fourth moments) behave respectively as $2/\sqrt{\alpha}$ and $3 + 6/\alpha$.

• **Second order Statistics (covariance): Long Range Dependence.** As explained in Section 2, because long memory is a crucial property for computer traffic, stochastic processes that intrinsically incorporate it are suitable for modeling. However, because of the many different network mechanisms and various source characteristics, short term dependencies are also present and superimposed to this long memory property. Therefore using the FARIMA process [7] is natural in that it allows both short and long range dependencies.
A farima(P, d, Q) model is defined via two polynomials of order P and Q and a fractional integration $\mathbf{D}^{-d}$, of order $-1/2 < d < 1/2$, as:

$$X_\Delta(k) = \sum_{p=1}^{P} \phi_p X_\Delta(k-p) + \mathbf{D}^{-d}(\epsilon(k) - \sum_{q=1}^{Q} \theta_q \epsilon(k-q)),$$

where the $\epsilon(l)$ are independent, identically distributed random variables, commonly referred to as innovations, with zero mean and variance σ_ϵ^2. For fractional d, the fractional integrator is read via a formal power series expansion: $\mathbf{D}^{-d} = \sum_{i=0}^{\infty} b_i(-d)B^i$, where B is the backward operator, $B\epsilon(i) = \epsilon(i-1)$, and $b_i(-d) = \Gamma(i+d)/\Gamma(d)\Gamma(i+1)$. The Fourier spectrum for this process reads:

$$f_X(\nu) = \sigma_\epsilon^2 |1 - e^{-i2\pi\nu}|^{-2d} \frac{|1 - \sum_{q=1}^{Q} \theta_q e^{-iq2\pi\nu}|^2}{|1 - \sum_{p=1}^{P} \phi_p e^{-ip2\pi\nu}|^2}, \tag{3}$$

for $-1/2 < \nu < 1/2$. This shows that for $d \in (0, 1/2)$, this process is long range dependent.
In this case, the ARMA(P, Q) contribution and the fractional integration of order d account respectively for short and long range correlations, in an independent and versatile way. The shape polynomials P and Q can be used to fit the spectrum at high frequencies or, equivalently, fine scales, while d measures the "strength" of the long memory.

• **Comments.** For the analysis and the illustrations reported in the present work, we will restrict ourselves to the use of farima processes with polynomials P and Q of degree at most 1, hereafter labeled farima(ϕ, d, θ).
Then, the $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ) processes involve only 5 parameters that need to be adjusted from the data. As such they are parsimonious models, a much desired property as far as robust, practical, efficient real time on-the-fly network monitoring is concerned.
Note however that the specifications of the first and second order statistical properties do not fully characterize the process, because this model is not Gaussian. Room for further design to adjust other properties of the traffic remains available in the framework of the model, but this task, difficult to achieve, is not needed here for the properties of the traffic we intend to capture.

4.2 Analysis

• **Stationarity of data.** For each aggregation level Δ independently, the analysis of the X_Δ is conducted. Because for theoretical modeling stationarity of X_Δ is assumed, we first perform empirically a consistency check of analysis and estimation results obtained from adjacent non overlapping sub-blocks. Then, we analyze only data sets for which stationarity is a reasonable hypothesis. This approach is very close in spirit to the ones developed in [44, 48]. Then we estimate the parameters of the model for each chosen Δ.

• **Gamma parameter estimation.** Instead of the usual moment based technique, $\hat{\beta} = \hat{\sigma^2}/\hat{\mu}$, $\hat{\alpha} = \hat{\mu}/\hat{\beta}$ where $\hat{\mu}$ and $\hat{\sigma^2}$ consist of the standard sample mean and variance estimators, we use maximum likelihood based estimates for the parameters α and β [16]. The joint distribution of n i.i.d. $\Gamma_{\alpha,\beta}$ variables can be obtained as a product of n terms as in Eq. 2. Derivation of this product with respect to α and β yields the estimates. It is important to note that the term ML standardly attributed to that method is here abusively used. Obviously, in our case, the $X_\Delta(k)$ are strongly dependent and hence do not satisfy the i.i.d. assumption. Section 6is devoted to show empirically from numerical simulations that this estimation procedure nevertheless provides us with very accurate estimates even when applied to processes with long range dependence.

• **Farima parameter estimation.** It is well known that the estimation of the long memory parameter is a difficult statistical task that received a considerable amount of works and attention (see e.g. [12] for

a thorough and up-to-date review), and hence, so will the joint estimation of both long and short range parameters of the farima(ϕ, d, θ) process. Full maximum likelihood estimation based on the analytical form of the spectrum recalled in Eq. 3 are possible but computationally heavy.
In this work, we develop a two step estimation procedure: the wavelet based methodology developed in [1, 2, 47] enables us to estimate first the long range dependence parameter d; second, the parameters of the residual ARMA part are estimated.

Let us first recall the wavelet framework (see [30] for a thorough introduction). Let ψ_0 denote respectively the mother wavelet. Let $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$ denote its dilated and translated templates and $d_X(j,k) = \langle \psi_{j,k}, X_0 \rangle$ the corresponding *wavelet* coefficients. The mother-wavelet ψ_0 is further characterized by an integer $N \geq 1$, the number of vanishing moments that plays a key role in the theoretical and practical analysis of long memory.
For any second order stationary process X, its spectrum $f_X(\nu)$ can be related to its wavelet coefficients through:

$$\mathbb{E}d_X(j,k)^2 = \int f_X(\nu) 2^j |\Psi_0(2^j\nu)|^2 d\nu, \tag{4}$$

where Ψ_0 stands for the Fourier transform of ψ_0 and $\mathbb{E}$ for the expected value. When X is a long range dependent process, with parameter d, Eq. 1 implies that:

$$\mathbb{E}d_X(j,k)^2 \sim C2^{j(2d+1)}, \quad \text{if } 2^j \to +\infty. \tag{5}$$

Moreover, it has been proven that the $\{d_X(j,k), k \in \mathbb{Z}\}$ form short range dependent sequences as soon as $N > d + 1/2$. This means that they no longer suffer from statistical difficulties implied by the long memory property. In particular, the time averages $S_j = 1/n_j \sum_{k=1}^{n_j} |d_X(j,k)|^2$ can then be used as relevant, efficient and robust estimators for $\mathbb{E}d_X(j,k)^2$. Together with Eq. 4 above, this leads to the following estimation procedure: a weighted linear regression of $\log_2 S_j$ against $\log_2 2^j = j$, performed in the limit of the coarsest scales, provides us with an estimate of $2d + 1$, hence of d. The plots $\log_2 S_j$ versus $\log_2 2^j = j$ are commonly referred to as logscale diagrams (LD). The possibility of varying N brings robustness to these analysis and estimation procedures. The full definition as well as the performance of this estimation procedure are detailed in [1, 2, 47].

From this wavelet based estimate $\hat{d}_W$ of d, we perform a fractional derivation of order $\hat{d}_W$ of X_Δ. This removes the long memory from the process so that only the ARMA component is left. A standard iterative procedure (based on a Gauss-Newton algorithm) [28] is then applied to estimate the ARMA parameters.
Obviously, the major weakness of this two step estimation procedure lies in the fact that be d poorly estimated, then so will the ARMA parameters. However, the estimation performance of the procedure are studied numerically in Section 6 using synthetic $\Gamma_{\alpha,\beta}$ farima(ϕ, d, θ) process.

5 Results and discussions

5.1 Traffic without anomalies

The $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ) analysis procedures described above are applied to traffic time series for various levels of aggregation independently. We present here detailed results for the **AUCK-IV** series and for the **Metrosec-ref1** series only. Similar results are obtained for the other series mentioned in Table 1, they are not reported here by lack of room (they are available on request).

• **Marginals.** For these two series, Figs. 3 and 5, left columns (model fit superimposed to empirical histogram), respectively illustrate the relevance of the $\Gamma_{\alpha,\beta}$ fits of the marginal statistics of X_Δ for a wide range of aggregation levels: 1ms $\leq \Delta \leq$ 10 s. The adequacy of these fits has been characterized by means of χ^2 and Kolmogorov-Smirnow goodness-of-fit tests (not reported here). Gamma distributions show usually a better adequacy compared to those obtained from exponential, log-normal and χ^2 laws. For some of the analyzed time series and some aggregation levels, one of these laws may better adjust the data. However, the Gamma distributions are never significantly outperformed, and if a particular distribution performs better than Gamma for a given Δ, this does not hold over a wide range of Δs. As opposed to this, the adequacy of the Gamma laws remains very satisfactory over wide ranges of Δs, and we advocate that this gives a scale-evolving characterization of marginals of the traffic. To some extent,

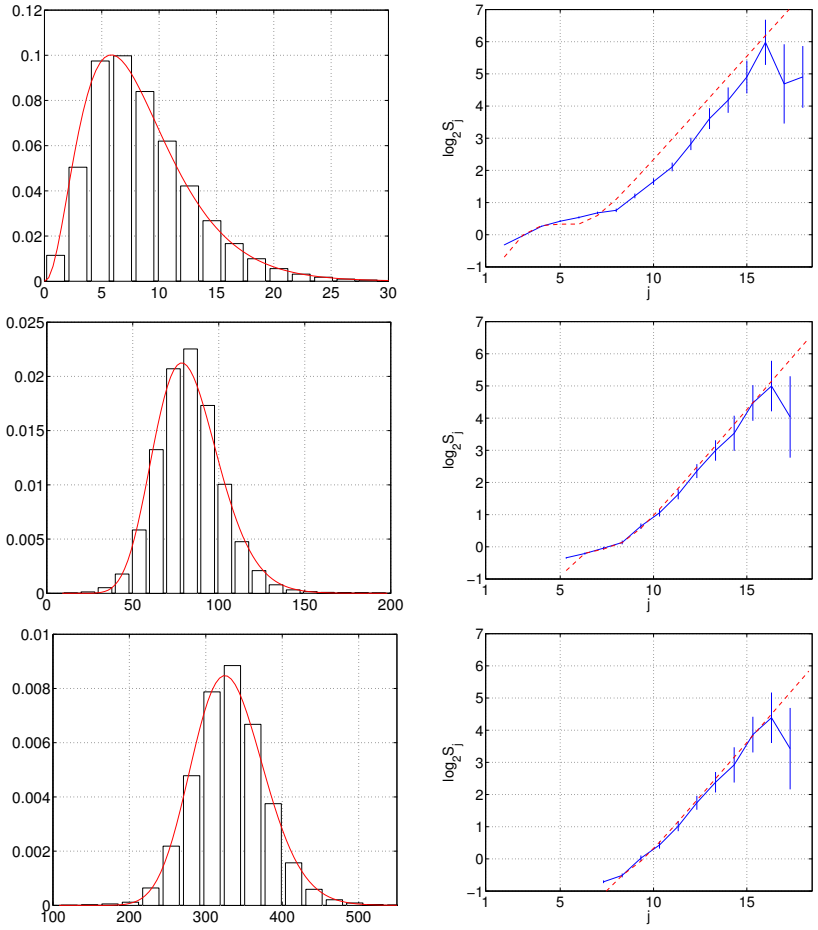


Figure 3: **AUCK-IV.** $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ) fits for the marginals (left column) and covariances (right column) for $\Delta = 10, 100, 400$ ms (top to bottom); $j = 1$ corresponds to 10 ms.

$\Gamma_{\alpha,\beta}$ laws, by variation of their shape and scale parameters, offer a continuous and smooth evolution from pure exponential to Gaussian laws.
Together, these facts are very much in favor of the use of Gamma laws to model computer traffic marginals and is related to the fact that Gamma laws form a family that is stable under addition: for any two X_i and $i = 1, 2$ independent random variables $\Gamma_{\alpha_i,\beta}$, their sum $X = X_1 + X_2$ follows a $\Gamma_{\alpha_1+\alpha_2,\beta}$ law. Under aggregation one has $X_{2\Delta}(k) = X_\Delta(2k) + X_\Delta(2k+1)$. Using the stability under addition argument, assuming independence, one would expect that α increases linearly with Δ while β remains constant. Figs. 4 and 6, top row, show the evolution of $\hat{\alpha}$ and $\hat{\beta}$ as a function of $\log_2 \Delta$. For all time series, significant departures from these ideal behaviors are observed. A careful analysis shows that $\hat{\alpha}(\Delta)$ does not increase at small Δ, then grows roughly like $\log_2 \Delta$ for larger Δ, whereas $\hat{\beta}(\Delta)$ behavior is close to a power-law increase. These facts constitute clear evidences for the existence of dependencies in the data. Moreover note that $\Delta \simeq 1$ s, corresponds to the onset of long memory (as will be discussed below). That tells us that the evolutions of α and β with Δ shown here accommodate mainly for short range dependencies.
The joint variations of α and β with respect to the aggregation level Δ provide us with a relevant feature of nominal traffic. We will take advantage of that to characterize and classify traffic with anomalies.

• **Covariances.** For the two reference series, Figs. 3 and 5, right columns, respectively, compare the logscale diagrams computed from data to fits of model logscale diagrams. These latter are computed numerically from the combination of the analytical Eqs. 3 (where the estimated $\hat{d}_W$, $\hat{\theta}$ and $\hat{\phi}$ are plugged-in) and 4 [2]. These plots illustrate the relevance of the farima(ϕ, d, θ) fits of the covariances of X_Δ.

[2] Procedure developed with D. Veitch, cf. [49].

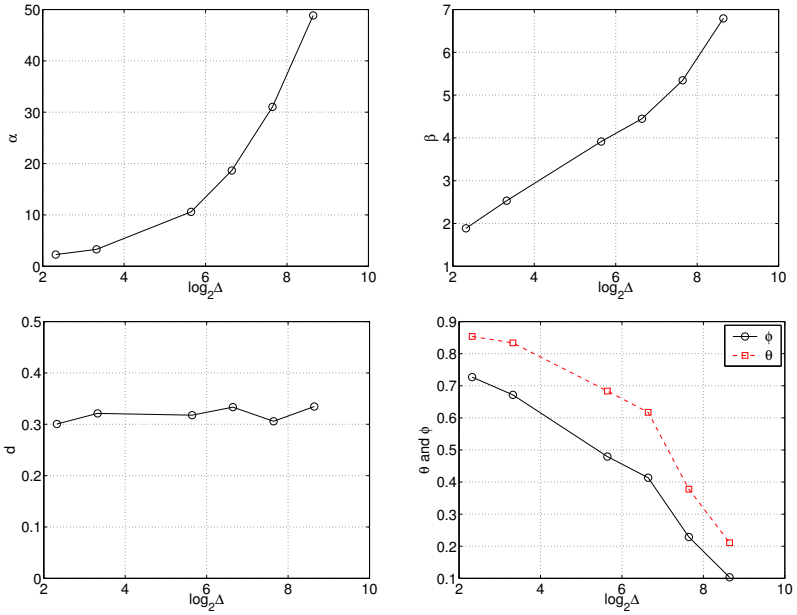


Figure 4: **AUCK-IV.** Estimated $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ) parameters as a function of $\log_2 \Delta$ (with Δ in ms).

One can notice that the logscale diagrams are almost obtained, as Δ increases, as coarser versions of the logscale diagram at finest Δ, shifted toward coarser scales. This can be easily understood as aggregating data mainly consists in smoothing out details at fine scales but leaving coarse scales unaffected. As expected, aggregation does not cancel the long memory characteristic and does not alter it. This can be checked in Figs. 4 and 6, bottom left, where the $\hat{d}_W$ remain notably independent of Δ and this again underlines that long range dependence captures a long-time feature of the traffic that has no inner time-scale.
On the other side, the short-time correlations are cancelled out when the aggregation level increases. This can be seen in Figs. 4 and 6, bottom right, that $\hat{\phi}$ and $\hat{\theta}$ significantly decrease as Δ increases. One expects that they would be null if the aggregation level Δ becomes larger than the characteristics time scales of the short range dependencies. The covariance theoretically converges to that of a fractional Gaussian noise that turns out to be practically extremely close to that of a farima$(0, d, 0)$. For all the reference time series studied here, the time scale where the long memory is dominant (measured as the approximate aggregation level Δ at which the short memory part of the model has vanished) corresponds to 600 ms $\leq \Delta \leq$ 2 s . To finish with, let us note that for some time-series (CAIDA), higher ARMA model proved necessary to model the covariance.

• **Conclusions.** As a partial conclusion, let us put the emphasis on the fact that for a wide range of different traffic collected on various networks, the proposed $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ) model reproduces accurately the marginals and both the short range and long range correlations of traffic time series. The fact that the proposed model is versatile enough to work equally well for a wide range of aggregation levels is a key feature, for two reasons. First, as mentioned in Section 2.1, choosing Δ a priori may be uneasy so that using a process that offers an evolutive modeling with Δ is of high interest. Second, the values of the parameters of the models obviously vary, possibly significantly from one traffic to another. But, these are not the values themselves that are of interest but the curves of evolution of these parameters with respect to Δ that provide us with the relevant statistical feature to analyze traffic.

5.2 Traffic with anomalies

• **DDoS Attack.** Fig. 7, left plot, presents the log scale diagrams for 1 hour block of data during the DDoS Attack (Δ = 1ms) compared to those of 1h long regular traffic times series, recorded a couple of hours before and after the attack. These plots tell us first that a farima(ϕ, d, θ) fits the traffic under DDoS attack equally satisfactorily. Other plots not presented here show that this is true for a wide range

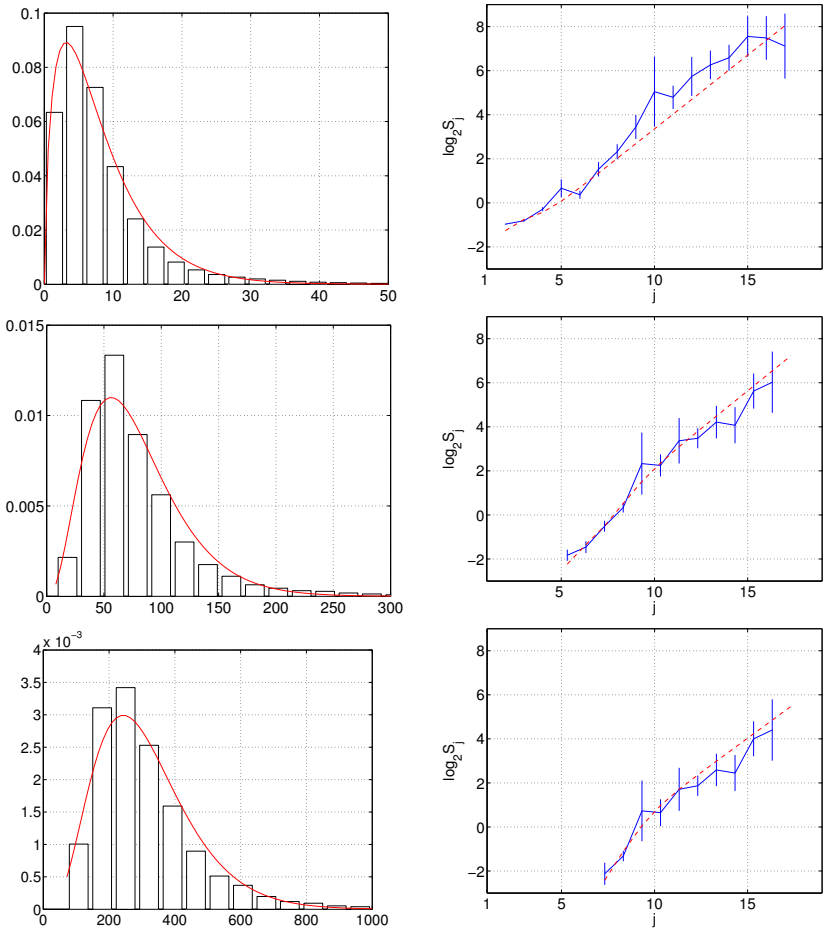


Figure 5: **METROSEC-ref1.** $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ) fits for the marginals (left column) and covariances (right column) for $\Delta = 10, 100, 400$ ms (top to bottom). $j = 1$ corresponds to 10 ms.

of aggregation levels.
Moreover, for the behaviors of the logscale diagrams at scales larger than 500 ms ($j = 9$ in Fig. 7, left plot), no discrepancies can be detected between before/after and during the attack. In particular, the long memory parameter $\hat{d}_W$ remains astonishingly constant. This tell us that long memory is not only not created by the attack, but also totally insensitive to its occurrence. The only change that can be noticed on the logscale diagram changes is a relative increase of the short-time component (at scales j from 4 to 7) after the attack: this is due to the fact that the traffic series after attack was recorded at night, with a lower traffic load. The logscale diagram was shifted upwards to show that the long memory parameter $\hat{d}_W$ (given by the slope) does not change, even when the load is smaller, and consequently the fine scale part has been raised. We cannot detect the anomaly from the logscale diagram.

Fig. 8, left column, illustrates, in two plots, that $\Gamma_{\alpha,\beta}$ distributions adequately fit the marginals of the traffic under DDoS attack. Fig. 9, left column, illustrate the compared evolutions of the estimated $\hat{\alpha}$ and $\hat{\beta}$ with respect to Δ for traffic during (red curves with crosses) and before (black squares) and after (blue circles) the DDoS event. We represent the experimental average of estimation 15 minute-long non-overlapping blocks of data, superimposed with extremal values taken by the estimate during each period. Windows before and after the anomaly correspond to relevant nominal behaviors for regular traffic. One sees that the functions of $\hat{\alpha}(\Delta)$ and $\hat{\beta}(\Delta)$ corresponding to the DDoS attack depart dramatically from regular behaviors. Let us put the emphasis on the fact that the values of the parameters may vary notably from one block to the other even within the DDoS event but that the evolutions in Δ remain comparable and define a path different from what happens with normal traffic.
The attack causes an immediate and sharp increase of α starting from the finest Δ levels whereas under normal circumstances, α remains constant or with only small variations up to $\Delta \simeq 20$ ms. The evolution

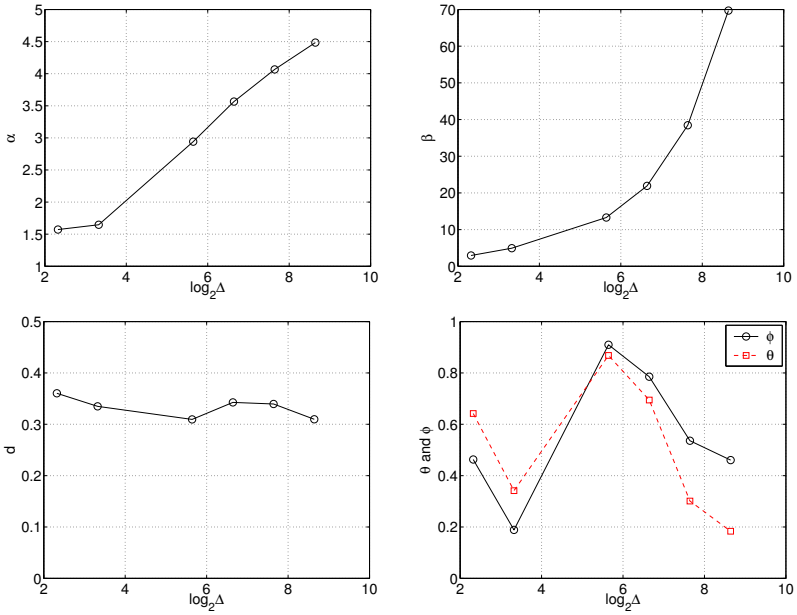


Figure 6: **METROSEC-ref1.** Estimated parameters of $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ), as a function of $\log_2 \Delta$ (Δ in ms).

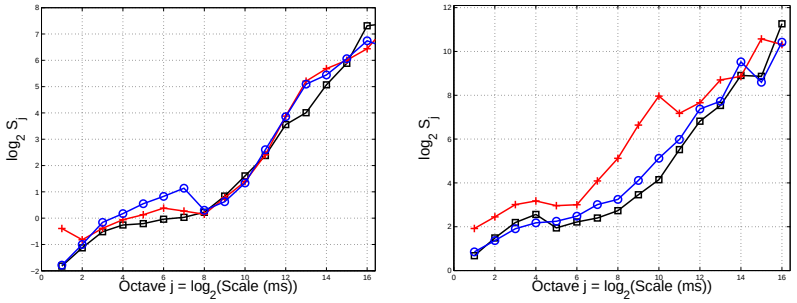


Figure 7: **Logscale Diagrams.** For the DDoS (left) and for the Flash Crowd (right). For both events, the curves are given during the anomaly (red crosses on curve), and before (black squares) or after (blue circles) the anomaly as references for normal traffic.

is inverse for β: it is decreasing from $\Delta \simeq 1$ ms to $\Delta \simeq 30$ ms during the DDoS attack, whereas it should increases smoothly and regularly with Δ under normal traffic.

These evolutions can receive several interpretations, in terms of occurrence of *0 packet event* and of *Gaussianization* effect. First, because during the DDoS attack a large number of packets are emitted at the highest possible rate, a major consequence lies in the fact the possibility that one observes regularly *0 packet* within a window of size Δ goes extremely fast to 0 as soon as Δ reaches 1 ms. More precisely, we observe that the marginals of traffic under DDoS attacks are strictly non zero only above a threshold that depends on Δ. This is a major discrepancy with marginals observed on regular traffic that smoothly go to 0 when $X_\Delta \to 0$ (compare Figs. 3 or 5 to Fig. 8). This effect precisely impacts the values taken by the shape parameter α with respect to Δ, implying that α will grow with Δ slowly for regular traffic and much faster for DoS attack one.
Second, as mentioned in Section 4.1 above, $1/\alpha$ controls the departure of $\Gamma_{\alpha,\beta}$ distributions from Gaussian ones. Under aggregation α tends to always increase. However, DDoS attacks are responsible for a significant acceleration in the *Gaussianization* effect. This constitutes a major statistical feature that differentiate traffic under DDoS attacks from regular one. To finish with, let us note that this effect involves time scales in the traffic ranging from 1 ms to 0.5 s and that the ARMA part of the covariance modeling (plots not reproduced here) is hardly seeing it while the LRD part of the covariance modeling

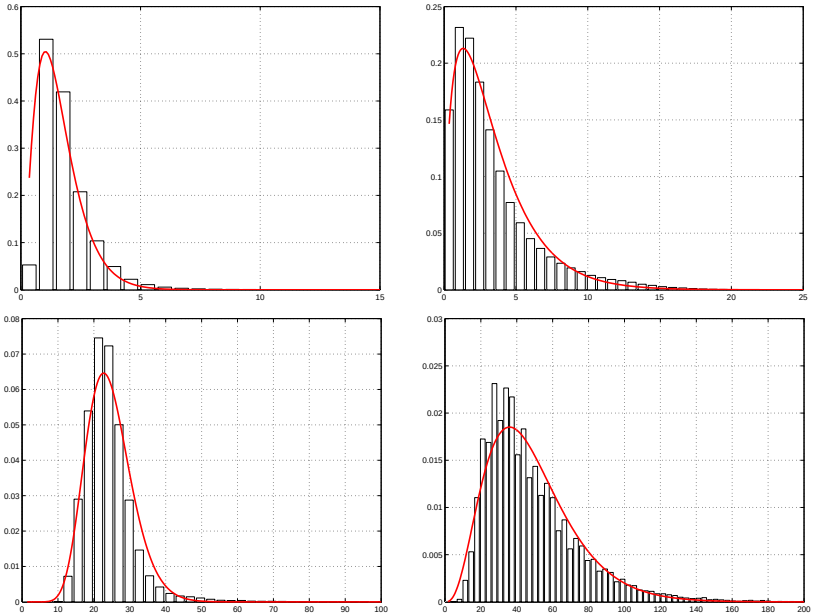


Figure 8: **Marginals.** For the DDoS Attack (left) and for the Flash Crowd (right), empirical histograms of X_Δ and their $\Gamma_{\alpha,\beta}$ fits of the marginals, for $\Delta = 2$ms (top) and $\Delta = 32$ms (bottom).

is totally blind to it.

• **Flash Crowd.** Fig. 8, right column, illustrates that $\Gamma_{\alpha,\beta}$ distributions adequately fit the marginals of the traffic under flash crowd for a wide range of aggregation levels (from 1ms to 1s). Fig. 8, right column, illustrate the compared evolutions of the $\hat{\alpha}(\Delta)$ and $\hat{\beta}(\Delta)$ curves for traffic during (red curves) and before (black) and after (blue) the flash crowd event. Each curve corresponds to 15 min-long non-overlapping blocks of data. The shapes of $\hat{\alpha}(\Delta)$ and $\hat{\beta}(\Delta)$ curves observed during the event do not depart significantly from those recorded under normal circumstances. Some departures are found out for the largest Δs (from 0.5 to 1 s), and this will prove consistent with the findings obtained from the log scale diagrams. This difference in $\hat{\alpha}(\Delta)$ observed under DDoS attack and flash crowd is consistent with the fact that the flash crowd does not involve any mechanisms that tend to forbid the *0 packet per window* event as the DDoS attack does.
Fig. 7, right plot, presents the logscale diagrams of two 15 min long block of data during the flash crowd ($\Delta = 1$ ms) compared to those of 15 min long blocks of normal traffic times series, recorded a few minutes before and after the flash crowd. On this plot, one sees that the the logscale diagrams undergo a significant change under flash crowds. At octaves $j = 8$ to $j = 10$, i.e., for scales of times ranging from 250 ms to 1 s, a strong peak of energy grows (such a peak was never observed on traffic under regular circumstances). Obviously, farima(ϕ, d, θ) fits (not shown here) will fail to reproduce simultaneously the short range dependences, the long range dependences and this energy peaks. Goodness-of-fit tests between data and fitted models yield rejection providing us with a relevant tool for detecting this traffic anomaly. Let us note moreover that the LRD parameter d when estimated for octaves coarser than those corresponding to the evidenced energy peak does not notably depart from that estimated value before or after the flash crowd. This tells that long memory is not caused by the flash crowd and neither is it affected by it. At most, the energy peak act as a masking effect in a subrange of time scales, from 250 ms to 2 s.

• **Comments and conclusions.** Before concluding this section, let us mention that the estimation of the mean and variance of X_Δ as functions of Δ would not enable us to discriminate between traffic with or without anomalies, nor between DDoS attack and Flash Crowd traffics. All the $\hat{\mu}(\Delta)$ and $\hat{\sigma^2}(\Delta)$ curves exhibit identical forms (plots not reproduced here).
The farima form of the covariance fails to accommodate the energy peak due to the flash crowd at specific times scales. The $\Gamma_{\alpha,\beta}$ distributions reproduce the marginals of all types of traffic with and without anomalies, be they legitimate or not. Therefore the monitoring of the goodness-of-fit test outputs for

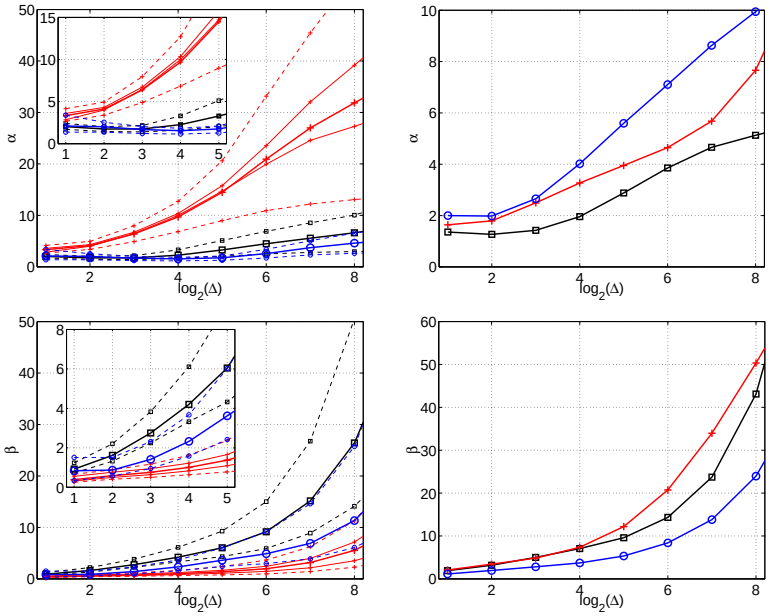


Figure 9: **Estimated $\Gamma_{\alpha,\beta}$ parameters.** Estimation of $\hat{\alpha}$ (top) and $\hat{\beta}$ (bottom) as a function of $\log_2 \Delta$ for the DDoS Attack (left) and for the Flash Crowd (right). In both cases, the curves are given during the anomaly (red crosses on curve), and before (black squares) or after (blue circles) the event as references for normal traffic. For the **DDoS**, the mean evolution (thick line) of the parameters on various 15 min data blocks is drawn, superimposed with the extremal values taken during each period (dashed lines); for the sake of example, two typical evolutions over one block during the DDoS are shown (in thin lines) on the graph. A zoom for the small scales is shown as an inside-plot. For the **FC** event, of smaller duration, one estimation on a 15 min. window is reported for each period (before, during and after the event).

the $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ) model, together with that of the evolutions of the estimated parameters as functions of Δ, for Δ ranging from 1 ms to 10 s, enables us to detect between traffic with and without anomalies as well as to classify them into legitimate and illegitimate ones. This is further discussed in Section 7.3.

6 Validation

6.1 Numerical synthesis

In this section, we want to produce numerically sample paths of processes with $\Gamma_{\alpha,\beta}$ marginals and farima(θ, d, ϕ) covariance and with prescribed length. One objective is the validation of the analysis procedure, especially the estimation performance.

• **Principles.** Our construction relies on three steps.
i) A $\Gamma_{\alpha,\beta}$ random variable X can be obtained as

$$X = \sum_{i=1}^{i=2\alpha} Y_i^2, \tag{6}$$

where the Y_is are zero-mean independent identically distributed Gaussian random variables.
ii) We can relate analytically the covariance of the process $X(k)$, $\gamma_X(l) = \sigma_X^2 \rho_X(l)$, to that of the $Y_i(k)$, $\gamma_Y(l) = \sigma_{Y_i}^2 \rho_Y(l)$. This computation is derived below. It develops calculations proposed in [29], and extends them to the Gamma case.
iii) We synthesize 2α zero mean Gaussian processes Y_i, with prescribed covariance $\gamma_Y(\tau) = \sigma_Y^2 \rho_Y(\tau)$, using the so-called circulant embedded matrix method (see, e.g., [12] for a review).
Obviously, the procedure we propose here works only for integer αs. Its extension to non integer α is under study.

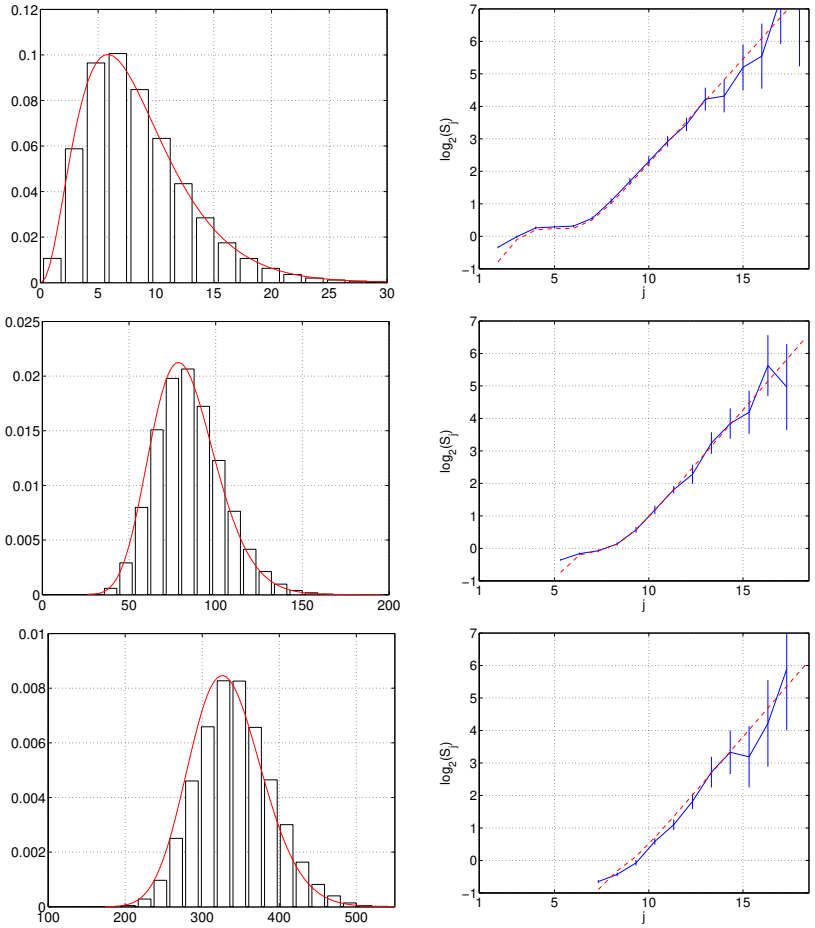


Figure 10: **Synthetic data.** $\Gamma_{\alpha,\beta}$ - farima(ϕ, d, θ) fits for the marginals (left column) and covariances (right column) of synthetic data for $\Delta = 10, 100, 400$ ms. Parameters correspond to those of **AUCK-IV**.

• **Derivation of the key result.** First, one can easily obtain that $\mathbb{E}X = \alpha\beta = 2\alpha\sigma_Y^2$ and $\sigma_X^2 = \alpha\beta^2 = 4\alpha\sigma_Y^4$, hence, $\sigma_Y^2 = \beta/2$. Second, from the canonical decomposition $Y(k+l) = \rho_Y(l)Y(k) + Z(k,l)$, one can show that $Z(k,l)$ is a Gaussian random variable, with $\mathbb{E}Z(k,l) = 0$, $\mathbb{E}Z^2(k,l) = \sigma_Y^2(1-\rho_Y^2(l))$ and $\mathbb{E}Y(k)Z(k,l) = 0$. From these results, one can derive that $\mathbb{E}Y^2(k)Y^2(k+l) = \sigma_Y^4(1+2\rho_Y^2(l))$. Combining those finding with the fact that the Y_i are i.i.d. zero-mean Gaussian processes, one obtains that

$$\rho_X = \rho_Y^2 \text{ or } \gamma_X = 4\alpha\gamma_Y^2. \tag{7}$$

The equivalent procedure for processes with other distributions such as log normal, exponential, chi-squared, Weibull uniform,... are under analysis.

6.2 Analysis Procedures

Now, we generate a large number of realizations of $\Gamma_{\alpha,\beta}$ farima(θ, d, ϕ) processes with different lengths n, and for different sets of α, β, θ, ϕ and d. In particular, we study with care the impact of the strength of the long memory by varying $d = 0.1, 0.15, 0.2, ..., 0.4$, and 0.45. To each realization, we apply the analysis procedures described in Section 4. Averaging results over several realizations enable us to evaluate numerically the performance of our analysis (goodness-of-fit and estimation) procedures.
One can compare Fig. 10 with Fig. 3: the parameters were chosen so that they correspond to those measured on the **AUCK-IV** time series. It illustrates that for different Δs, the $\Gamma_{\alpha,\beta}$ - farima(θ, d, ϕ) model matches the marginals (left column) and covariances (right column) of both real data and the synthetically produced X_Δ time series. Results for the goodness-of-fit tests (not reported here) are also extremely satisfactory.

Fig. 11 presents the relative biases of the estimated parameters as a function of the $\log_{10}$ of the observation duration. One notices that, for all parameters and for all durations, these biases remain low, within a few percent, even when d is large (close to 0.5). Variance plots, not shown here, present a standard $1/n$ decrease, a very satisfactory result. Together, these facts show that the wavelet based estimate of d is satisfactory (as it was already thoroughly studied in [47]), but also that the other parameters are correctly estimated, as they would from data with short range dependencies only. This shows a posteriori that the removal of the long memory performed by the fractional derivation of order $\hat{d}_W$ is actually efficient.
Therefore, the numerical simulations reported here validate the relevance of the analysis procedures used to analyze actual Internet time series in the previous section.

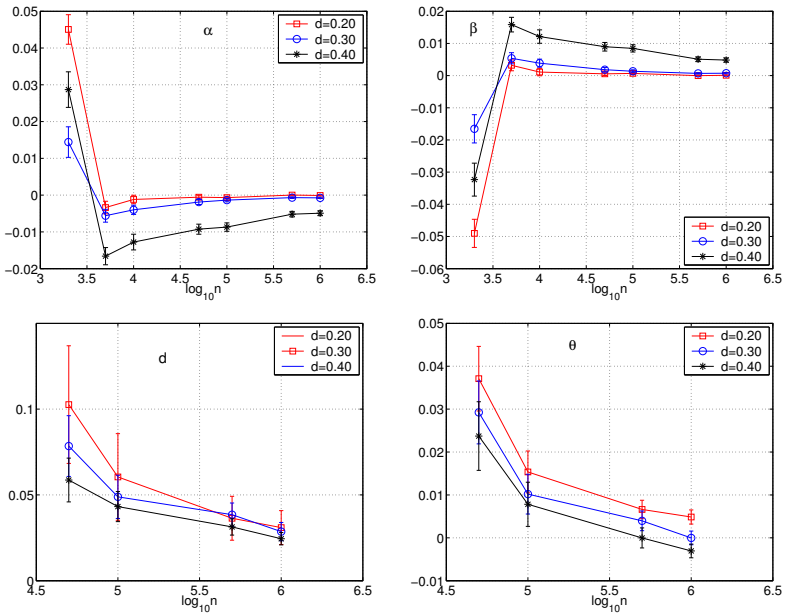


Figure 11: **Performance of estimation procedures.** Relative biases for $\hat{\alpha}$, $\hat{\beta}$, $\hat{d}$ and $\hat{\phi}$. $\hat{\phi}$ is not shown but is comparable to $\hat{\theta}$. Curves are plotted for 3 values of d only for clarity. n is the number of points. Performance were obtained by averaging 500 realizations.

7 Applications

7.1 Traffic generators

Provided that the analytical derivation that connect the covariance functions of a Gaussian process and some non gaussian process, the synthesis procedure described in Section 6.1 can be extended to other types of marginals (log-normal, exponential, chi-squared,...) and covariances (fractional Gaussian noise (fGn), kinked fGn,...). This is under current investigation. Other forms of statistical dependencies may as well be incorporated, including higher order statistics.
Such synthesis procedures constitute traffic generators for non Gaussian long range dependent traffic. Such generators may be used to feed simulation platforms aiming at estimating QoS and performance. The impact of deviations of given parameters from those observed on traffic can hence be studied efficiently and systematically.

7.2 Traffic prediction

The relevance of the proposed $\Gamma_{\alpha,\beta}$ - farima(P,d,Q) model enable to think of using them for traffic evolution prediction. Indeed, within this frame, when the traffic $X_\Delta(t)$ is observed for $t \in [0,T]$, the control over the first and second statistical orders of the model enables to derive a procedure to predict

the samples $X_\Delta(T+\tau)$ for $\tau > 0$ within a window time that needs to be studied, that will depend on the orders of the ARMA(P,Q) part of the model as well as on the strength of the long memory (i.e., the value of the d parameter). In that respect, the use of larger orders for P and Q as long as they are relevant may prove beneficial. This is under study.

7.3 Detection

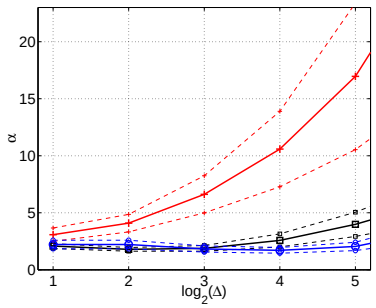


Figure 12: **DDoS Attack: $\hat{\alpha}$ averaged over 1 minute windows.** $\hat{\alpha}$,(thick lines), maximum deviation (dashed lines) as a function of $\log_2 \Delta$, during (red crosses), before (black squares) or after (blue circles) the anomaly.

A major perspective of this work is the real time detection of anomalies in traffic. For the sake of a good statistical characterization, analysis in Section 5.1 were performed over 1h long blocks of data for the reference times series, and 15 min long blocks of data for the DDoS attack and flash crowd. This allowed for a precise statistical description of the data but obviously corresponds to too long observation durations to match detection goals. We are currently developing a joint monitoring of the evolution of α, β, θ and ϕ, as well as the global shapes of the marginals and of the logscale diagrams based of much shorter observation duration.

Fig. 12 reports the estimation of $\hat{\alpha}(\Delta)$ over one minute windows. The evolutions of the mean $\hat{\alpha}$ reproduce the features underlined for 15 minutes windows (see Fig. 8), during the attack or for reference traffic (before and after the attack). One sees clearly that the shape of $\hat{\alpha}(\Delta)$ undergoes a drastic change at every aggregation level during the DDoS attack. This change could be quantified by means of (Kullback or Bhattacharyya-type) distances or divergences [6]. Moreover the analysis is made over blocks of 1 minutes, so that shortly delayed detection can be achieved. This could serve as a basic ingredient for the design of an IDS.

8 Conclusion

In the present work, we introduced a non Gaussian long range dependent process, the $\Gamma_{\alpha,\beta}$ - farima(P,d,Q), to model (the first and second order statistics) of computer network traffic. We also describe estimation procedures for the corresponding parameters. We showed on a large variety of standard reference traffic time series that it constitutes a relevant versatile model, and this for a very large range of aggregation levels Δ. Morevoer, its parameters are smoothly evolving with Δ hence providing us with a useful statistical characterization of regular traffic. We also showed that discrepancies from these reference behaviors with respect to Δ enabled us to distinguish between traffic with and without anomalies and to further discriminate between legitimate (flash crowds) and illegitimate (DDoS attacks) ones.

This work can be further developed along numerous directions. First, we intend to further explore the zoo of regular and anomalous traffics by analyzing very recent data as well as by creating a larger variety of anomalies (more complex and more diffuse DDoS attacks, larger flash crowds, network failures,...). To this end, an experimental platform is being developed within the METROSEC project. We want to explore both the ability of our model to characterize meaningfully this wider set of anomalies and the need to further enrich it and increase its complexity. Second, making use of the found statistical characterizations, we intend to come up in a short future with a detection scheme that is able to automatically identify changes in the traffic statistical characterizations and classify them as legitimate or illegitimate anomalies.

It should operate as an IDS, and be based on short-time time window observation duration.
Ultimately, the METROSEC project aims at developing network based (protocols, architectures,...) strategies to improve the robustness of the network against attacks. This increased insensitivity should help maintaining the targeted level of QoS. The present work constitute a first step toward that global goal.

9 Acknowledgments

The authors acknowledge the help of CRI ENSLyon, and L. Gallon (IUT Mont-de-Marsan, France) and L. Bernaille (LIP6, Paris) for their help in conducting data collection and DDoS attack. They also gratefully acknowledge all the people who freely accepted to take part into the scheduled flash crowd event analyzed here. Finally, they gratefully acknowledge colleagues from the major internet traces repositories (Bellcore, LBL, UNC, Auckland Univ, Univ North Carolina, CAIDA) for making their data available to us. They specially thank S. Marron and F. Hernandez-Campos and C. Park from UNC, USA, and D. Veitch and N. Hohn from CubinLab, University of Melbourne, Australia for having performed the pre-formatting of some of the time series used here. This work has been made possible thanks to the financial support the French MNRT ACI *Sécurité et Informatique* 2004 grant, within the METROSEC project.

References

[1] P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch. Wavelets for the analysis, estimation and synthesis of scaling data. In K. Park and W. Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*. Wiley, 2000.

[2] P. Abry and D. Veitch. Wavelet analysis of long-range dependent traffic. *IEEE Trans. on Info. Theory*, 44(1):2–15, January 1998.

[3] A. Andersen and B. Nielsen. A Markovian approach for modelling packet traffic with long range dependence. *IEEE journal on Selected Areas in Communications*, 5(16):719–732, 1998.

[4] C. Barakat, P. Thiran, G. Iannaccone, C. Diot, and P. Owezarski. A flow-based model for internet backbone traffic. In *ACM/SIGCOMM Internet Measurement Workshop*, pages 35–47, New York, NY, USA, 2002. ACM Press.

[5] P. Barford, J. Kline, D. Plonka, and A. Ron. A signal analysis of network traffic anomalies. In *ACM/SIGCOMM Internet Measurement Workshop*, Marseille, France, November 2002.

[6] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18:349–369, 1989.

[7] J. Beran. *Statistics for Long-memory processes*. Chapman & Hall, New York, 1994.

[8] J. Brutlag. Aberrant behavior detection in time series for network monitoring. In *USENIX System Administration Conference*, New Orleans, December 2000.

[9] C-M. Cheng, H.T. Kung, and K-S. Tan. Use of spectral analysis in defense against DoS attacks. In *IEEE Globecom*, Taipei, Taiwan, 2002.

[10] J. Cleary, S. Donnelly, I. Graham, A. McGregor, and M. Pearson. Design principles for accurate passive measurement. In *Passive and Active Measurements*, Hamilton, New Zealand, April 2000.

[11] N. Desaulniers-Soucy and A. Iuoras. Traffic modeling with universal multifractals. In *IEEE Globecom*, 1999.

[12] P. Doukhan, G. Oppenheim, and M.S. Taqqu. *Long-Range Dependence: Theory and Applications*. Birkhäuser, Boston, 2003.

[13] A. Erramilli, O. Narayan, and W. Willinger. Experimental queueing analysis with long-range dependent packet traffic. *ACM/IEEE transactions on Networking*, 4(2):209–223, 1996.

[14] M. Evans, N. Hastings, and B. Peacock. *Statistical Distributions*. Wiley (Interscience Division), June 2000.

[15] A. Feldmann, A.C. Gilbert, and W. Willinger. Data networks as cascades: Investigating the multifractal nature of internet wan traffic. In *ACM/SIGCOMM conference on Applications, technologies, architectures, and protocols for computer communication*, 1998.

[16] G.J. Hahn and S.S. Shapiro. *Statistical Models in Engineering*, page 88. Wiley (Interscience Division), June 1994.

[17] J. Hochberg, K. Jackson, C. Stallings, J.F. McClary, D. DuBois, and J. Ford. NADIR: an automated system for detecting network intrusion and misuse. *Journal of Computer Secururity*, 12(3):235–248, 1993.

[18] N. Hohn, D. Veitch, and P. Abry. Cluster processes, a natural language for network traffic. *IEEE Transactions on Signal Processing Special Issue on Signal Processing in Networking*, 8(51):2229–2244, October 2003.

[19] N. Hohn, D. Veitch, and P. Abry. Multifractality in tcp/ip traffic: the case against. *Computer Networks Journal*, to appear, 2005.

[20] A. Hussain, J. Heidemann, and C. Papadopoulos. A framework for classifying denial of service attacks. In *SIGCOMM*, Karlsruhe, Germany, 2003.

[21] Javits and Valdes. The SRI IDES statistical anomaly detector. *ESORICS*, May 1991.

[22] S. Jin and D. Yeung. A covariance analysis model for DDoS attack detection. In *IEEE International Conference on Communications*, Paris, France, June 2004.

[23] J. Jung, B. Krishnamurthy, and M. Rabinovich. Flash Crowds and Denial of Service Attacks: Characterization and Implications for CDNs and Web Sites. In *International WWW Conference*, Honolulu, HI, May 2002.

[24] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido. A non stationary Poisson view of the internet traffic. In *INFOCOMM*, 2004.

[25] A. Lakhina, M. Crovella, and C. Diot. Diagnosing network-wide traffic anomalies. In *SIGCOMM*, August 2004.

[26] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic (extended version). *ACM/IEEE transactions on Networking*, 2(1):1–15, February 1994.

[27] L. Li and G. Lee. DDoS attack detection and wavelets. In *International Conference on computer communications and networks*, August 2003.

[28] L. Ljung. *System identification: theory for the user*, chapter 10.2. PTR Prentice Hall, 1999.

[29] S.B. Lowen, S.S. Cash, M. Poo, and M.C. Teich. Quantal neurotransmitter secretion rate exhibits fractal behavior. *The journal of Neuroscience*, 17(15):5666–5677, August 1997.

[30] S. Mallat. *A Wavelet tour of signal processing*. Academic Press, 1999.

[31] Benjamin Melamed. An overview of TES processes and modeling methodology. In *Performance/SIGMETRICS Tutorials*, pages 359–393, 1993.

[32] D. Moore, G.M. Voelker, and S. Savage. Inferring internet denial-of-service activity. In *Usenix Security Symposium*, 2001.

[33] I. Norros. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE journal on Selected Areas in Communications*, 13(6), 1995.

[34] K. Park, G. Kim, and M. Crovella. On the relationship between file sizes, transport protocols, and self-similar network traffic. In *International Conference on Network Protocols*, page 171, Washington, DC, USA, 1996. IEEE Computer Society.

[35] K. Park and W. Willinger. Self-similar network traffic: An overview. In Kihong Park and Walter Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*, pages 1–38. Wiley (Interscience Division), 2000.

[36] S. Paulo, V. Rui, and P. António. Multiscale fitting procedure using Markov Modulated Poisson Processes. *Telecommunication Systems*, 23 (1/2):123–148, June 2003.

[37] V. Paxon and S. Floyd. Wide-area traffic: The failure of Poisson modeling. *ACM/IEEE transactions on Networking*, 3(3):226–244, June 1995.

[38] V. Paxson. Bro: a system for detecting network intruders in real-time. *Computer Networks Journal*, 31(23–24):2435–2463, 1999.

[39] QoSMOS Traffic Designer. http://www.qosmos.net.

[40] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman&Hall, 1994.

[41] S. Sarvotham, R. Riedi, and R. Baraniuk. Connection-level analysis and modeling of network traffic. Technical report, ECE Dept., Rice Univ., 2001.

[42] M. Taqqu, V. Teverosky, and W. Willinger. Is network traffic self-similar or multifractal ? *Fractals*, 5(1):63–73, 1997.

[43] B. Tsybakov and N.D. Georganas. Self similar processes in communications networks. *IEEE Trans. on Info. Theory*, 44(5):1713–1725, 1998.

[44] S. Uhlig, O. Bonaventure, and C. Rapier. 3D-LD: a graphical wavelet-based method for analyzing scaling processes. In *ITC Specialist Seminar*, pages 329–336, Würzburg, Germany, 2003.

[45] H.S. Vaccaro and G.E. Liepins. Detection of anomalous computer session activity. In *IEEE Symposium on Security and Privacy*, pages 280–289, Oakland, California, May 1989.

[46] J. Lévy Véhel and R. H. Riedi. *in Fractals in Engineering'97,* J. Lévy Véhel and E. Lutton and C. Tricot, editors, chapter Fractional Brownian motion and data traffic modeling: The other end of the spectrum. Springer, 1997.

[47] D. Veitch and P. Abry. A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Trans. on Info. Theory special issue on "Multiscale Statistical Signal Analysis and its Applications"*, 45(3):878–897, April 1999.

[48] D. Veitch and P. Abry. A statistical test for the time constancy of scaling exponents. *IEEE Transactions on Signal Processing*, 49(10):2325–2334, October 2001.

[49] D. Veitch, P. Abry, and M. S. Taqqu. On the automatic selection of the onset of scaling. *Fractals*, 11(4):377–390, 2003.

[50] N. Ye. A Markov chain model of temporal behavior for anomaly detection. In *Workshop on Information Assurance and Security*, West Point, NY, June 2000.

[51] J. Yuan and K. Mills. DDoS attack detection and wavelets. Technical report, National Institute of Standards and Technology, 2004.

[52] Z. Zhang, V. Ribeiro, S. Moon, and C. Diot. Small time scaling behavior of internet backbone traffic: an emprirical study. *INFOCOMM*, March 2003.