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

A. Scherrer, *Student Member, IEEE*, N. Larrieu, P. Owezarski, P. Borgnat, *Member, IEEE*, P. Abry, *Member, IEEE*

A. Scherrer is with the LIP, ENS Lyon, UMR-CNRS.

N. Larrieu and P. Owezarski are with the LAAS, UPR-CNRS.

P. Borgnat and P. Abry are with the Physics Lab., ENS Lyon, UMR-CNRS

Abstract

The goals of the present contribution are twofold. First, we propose the use of a non-Gaussian long-range dependent process to model Internet traffic aggregated time series. We give the definitions and intuition behind the use of this model. We detail numerical procedures that can be used to synthesize artificial traffic following exactly the model prescription. We propose also original and practically effective procedures to estimate the corresponding parameters from empirical data. We show that this empirical model describes relevantly a large variety of Internet traffic, including both regular traffic obtained from public reference repositories and traffic containing legitimate (flash crowd) or illegitimate (DDoS attack) anomalies. We observe that the proposed model fits accurately the data for a wide range of aggregation levels. The model provides us with a meaningful multi-resolution (i.e., aggregation level dependent) statistics to characterize the traffic: the evolution of the estimated parameters with respect to the aggregation level. It opens the track to the second goal of the paper: anomaly detection. We propose the use of a quadratic distance computed on this statistics to detect the occurrences of DDoS attack and study the statistical performance of these detection procedures. Traffic with anomalies were produced and collected by ourselves so as to create a controlled and reproducible database allowing for a relevant assessment of the statistical performance of the proposed (modeling and detection) procedures.

Key Words: Traffic statistical modeling, DoS attack, Flash Crowd, Non-Gaussian Long-Range Dependent Process

I. Motivation

Internet is becoming the universal communication network, conveying 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, Internet is evolving from a single best effort service to a multi-service network, a major consequence being that it becomes highly exposed to attacks, especially to denial of services (DoS) and distributed DoS (DDoS) attacks. 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 may result in the breaking of the service level agreement at the Internet Service Provider fault, potentially inducing major financial losses for them.

Detecting and reacting against DoS attacks is therefore a major issue that is continuously receiving numerous research efforts. However, this is also a difficult task and current intrusion detection systems (IDS), especially those based on anomaly detection from profile, often fail in detecting DDoS attacks efficiently. This can be explained via different lines of arguments. First, DDoS attacks can take a large variety of 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 [1], often described in terms of scaling [2], long memory [3], self-similarity [4], multifractality [5]. Such properties significantly impair anomaly detection procedures by decreasing their statistical performance. Third, Internet traffic may exhibit strong, possibly sudden, however legitimate, variations (flash crowds for instance, such as the notorious Slashdot effect) that may be hard to distinguish from illegitimate ones. Fourth, profile-based IDS generally do not rely on the use of rich enough statistical models that correctly account for traffic large variability. 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 [6]. Alarms are raised whenever a threshold is reached [7], [8], [9], [10], often yielding a significant number of false positives [11], a major shortcoming for their actual use. The current evolution of Internet traffic, allowing for a larger variety of traffic and diversity of communication, results in an increase of difficulties to design efficient IDS.

Recently, various Internet traffic monitoring projects have obtained important improvements in traffic modeling. Mostly, they have better taken into account the large variability and scaling properties mentioned above via the use of richer statistics of the traffic such as correlation functions or spectra. This has significantly renewed IDS design strategies. For instance, Ye rely on a Markov modeling of the traffic time behavior [12]. Other authors have shown that DDoS attacks increase correlations in traffic and indicated that a robust detection technique can be based on this observation [13], [14]. Making use of traffic inter-correlation across different links, Lakhina *et al.* have proposed a method for detecting network wide anomalies using traffic matrices [15]. Hussain and co-authors have defined spectral density signatures for attacks [16]. Similarly, spectral estimation has been used for comparing traffic with and without attacks [17]. While spectral densities exhibit peaks around the Round Trip Time values for regular traffic, such peaks tend to disappear under attacks, and this observation can then be used for IDS design. Finally, Li and Lee has used the wavelet technique developed in [18] 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 [19]. Works in [20], [21] exploit the multiresolution nature

TABLE I: **Regular traffic.** Description for the studied traces containing no anomaly. For each trace, T denotes the time duration (in second), # Pkts (10^6) the number of packets (in million) and IAT the mean inter-arrival time (in ms).

table

Data	Date(Start Time)	T (s)	Network(Link)	# Pkts	IAT (ms)	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-ref3	2006-03-20(11:00)	3600	LAN(10BaseT)	2.8	3.7	www.laas.fr/METROSEC/
METROSEC-ref4	2006-03-21(15:00)	3600	LAN(10BaseT)	2.9	3.9	www.laas.fr/METROSEC/

of wavelet decompositions to track and detect traffic anomalies in a so-called medium range of scales.

The present contribution, conducted in the framework of the METROSEC (Metrology for Security) project (see http://www.laas.fr/METROSEC), pursues along this research line. It is organized around two major goals: Internet traffic statistical modeling and attack detection. Mainly, it aims at analyzing the impact of anomalies on the (parameters) of the statistical modeling as well as at determining discriminative profile signatures for traffic containing legitimate (e.g., flash crowds) and illegitimate (e.g., DDoS attacks) anomalies. First, a long-range dependent non-Gaussian stochastic process is introduced and argued for. Its definition and properties, together with numerical synthesis and parameter estimation procedures, are fully worked out in Section III. Section IV shows that this model describes accurately and relevantly both a wide variety of Internet traffic time series (available from major public international trace repositories) and traffic containing anomalies (generated by ourselves), be they legitimate or not. The originality of the proposed statistical modeling lies in its *multiresolution* nature (several aggregation levels Δs are jointly analyzed). It provides us with robust statistics (the evolution of the model parameters with respect to Δ), taking into account jointly the marginal distributions and the correlation structure of the aggregated traffic, thus opening the track to the second goal of the paper: anomaly detections.

The detection procedure proposed here is based on identifying changes in the model parameter evolutions and hence ruptures in the statistical modeling. Therefore, it is generic and robust as it do not depend on any specific anomaly or attack production mechanism. The detection procedure consists of computing *quadratic distances* between the statistics estimated from a sliding observation time window and those obtained from an a priori chosen reference window. Then, distances are thresholded to yield detections. A key issue in validating anomaly detection procedure lies in the assessment of its statistical performance. As it is difficult to have at disposal traces containing a labeled and documented set of attacks that could be used to benchmark detection procedures, we have chosen to perform a collection of DDoS attacks and flash crowd anomalies, whose characteristics and parameters can be modified in a *controlled* and *reproducible* manner. Both regular data and data containing labeled anomalies are described in section II, together with the operating modes used to perform various DDoS attacks and flash crowd. From this database, we can evaluate the statistical performance (detection vs. false alarm probabilities) and reliability of the proposed detection procedures. Though artificial or simplistic this approach may seem, we see this reference database production methodology as a mandatory step for reliable development and validation of attack detection method. Detection procedures as well as their statistical performance are detailed in Section V. Both regular data and data containing labeled anomalies are described in section II, together with the operating modes used to perform various DDoS attacks and flash crowd. Section VI concludes with further developments under investigation.

II. Data and Experiments

A. *Traffic without anomalies*

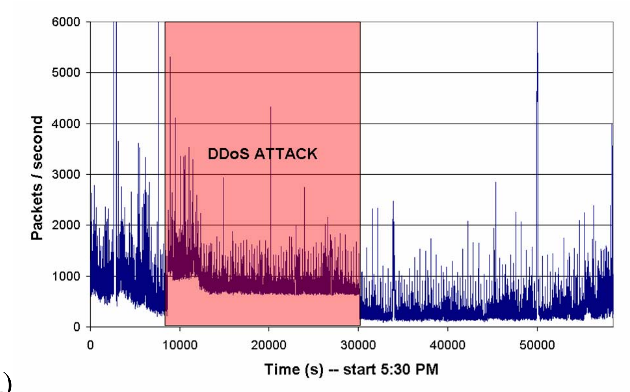
The model and analysis proposed hereafter are first illustrated on regular traffic (i.e. traffic presenting a priori no anomaly) fully described in Table I. We use both standard data, gathered from major available Internet traces repositories and time series collected by us within the METROSEC research project. Therefore, we cover a significant variety of traffic, networks (Local Area Network, Wide Area Network, etc. and edge networks, core networks, etc.) and links, collected over the last 17 years (from 1989 to 2006). For each repository, a large number of traces are available, we have focused here on a few ones that are representative of a collection of others. **PAUG** corresponds to one of the celebrated Bellcore Ethernet LAN traces, over which long-range dependence was first evidenced [22]. **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 [5], [23]. **AUCK-IV** constitute high precision TCP/IP traces gathered at the Internet access point of the University of Auckland over a non saturated link and made available by WAND. We also have processed one of 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 data collected at the University of North Carolina in 2003. The METROSEC data were collected from late 2004 to early 2006, on the RENATER[1] network using DAG systems [24] deployed in the framework of the METROPOLIS and METROSEC French research projects.

B. Traffic (or traces) with anomalies

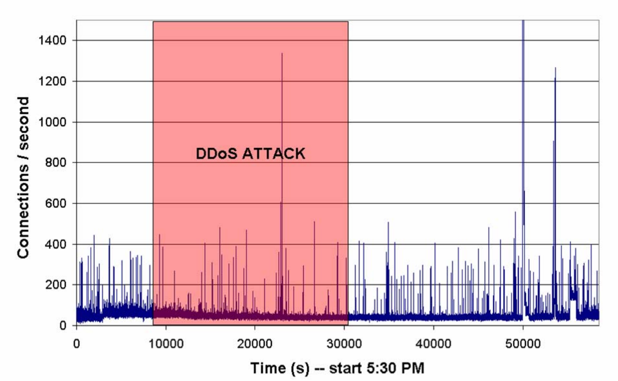
To assess the relevance and performance of our data modeling and anomaly detection procedures, we need to have at disposal a set of traffic containing labeled and documented anomalies. Because no such repository of reliable anomalies exists, we have created a database of legitimate (flash crowds) and illegitimate (DDoS attacks) anomalies, produced in a reproducible, accurate and controlled manner. This section details the anomaly production methodology and characteristics.

1) DDoS attack: **Experimental setting.** We performed UDP flooding DDoS attacks using either IPERF [25] or Trinoo [26] (on computers with Linux distribution) to generate UDP flows with different throughputs. Compared to IPERF, Trinoo uses a “daemon” installed on each attacking site (4 French research laboratories located in Mont-de-Marsan, Lyon, Nice and Paris), and enabled us to create more complex and realistic attacks. The single computer target was located at LAAS in Toulouse. The traffic related to these attacks was transported via the French national network for education and research (RENATER). DDoS attacks were performed so as to be able to reproduce and modify their characteristics (duration, DoS flow intensity, packets

[1]RENATER is the French network for education and research that interconnects academics and some industrial partners, see http://www.renater.fr/ for topology and further informations.



(a)



(b)

Fig. 1: **DDoS Attack.** Time series corresponding to the numbers per second of Packets (a) and connections (b).
figure

length and sending rate). In each case, traffic was collected by ourselves (for a duration of 60 or 90 minutes, the attack mostly occurring during the second-third) before, during and after the DDoS, so that regular traffic can be analyzed before and after each attack. The contribution of the attacks to the global throughput of the monitored link is highly variable, depending on the attack parameters and ranging from a major impact on global traffic profile (IV, V and X) to attacks that are completely hidden in the global traffic (I, II). Attacks X and tT were for instance designed to be clearly dominant in the traffic, whereas attack tM was to be almost hidden. Attack parameters and characteristics are fully detailed in Table II.

DDoS attack traffic characteristics. The LAAS is connected to RENATER with an 100 Mbps Ethernet link that has not been overflowed during attacks. Therefore, most of the conducted attacks have remained low in traffic volume so that they cannot be easily detected via simple statistics such as sample mean or variance estimates. Anyway, the goal of the detection procedures proposed in Section V is to detect anomalies even and mostly when their intensity level remain low, i.e., before they have a negative impact on the network QoS. Experiments were conducted to emulate this situation.

The plots illustrating the modeling of the DDoS were obtained from the reference DDoS, labeled R in Table II. For instance, Fig. 1 shows respectively the numbers of packets and flows on the LAAS access link. While the former remains quite stable, the latter presents a significant increase (the packet rate is multiplied by almost 3 during the attack). But this change remains in the range of the natural fluctuations of internet traffic. However, note that all the modeling (see Section IV) and detection procedures (see Section V) described here were applied to each and every time series reported in Tables I and II and that is has been checked that satisfactory and consistent results were obtained.

2) Flash Crowd: **Experimental setting.** We created anomalies that are considered as legitimate, under the guise of flash crowds (FC) on a web server. Our goal was to generate realistic FCs. This is why we chose not to use automatic programs or robots, but to involve human volunteers. To do so, we have asked to a large number of people (mostly French academics but not only) to browse the LAAS website (http://www.laas.fr). The LAAS website contains a large variety of files of all sizes, from simple html pages to movies, big reports, high definition pictures (of nano devices,...), movies (of autonomous robots,...), etc. There are every indications of heavy-tailed file sizes on this website as it is largely expected. Participants were instructed to browse the web server on their own, as they would do in the real world when visiting a website publishing a new set of information they would be interested in. Precise starting and stopping times were given. FC lasted 30 min or so. A detailed analysis of the IP addresses present in the LAAS incoming traffic enabled us to identify academic laboratories and to find out that more than a hundred persons participated. Information sent to us by participants enable us to say that at least 50 more persons took part in the FC from their individual and personal ISPs.

FC traffic characteristics. Illustrations are presented on FC-1 (cf. Table II). Fig. 2(a) shows the number of (HTTP GET) requests received by the LAAS web server, distinguishing between inner and outer requests. One can clearly see that most people started browsing the LAAS web server precisely when instructed (important increase of the number of hits), but also that some of them did not participate for the whole 30 minutes. Fig. 2 shows respectively the numbers of

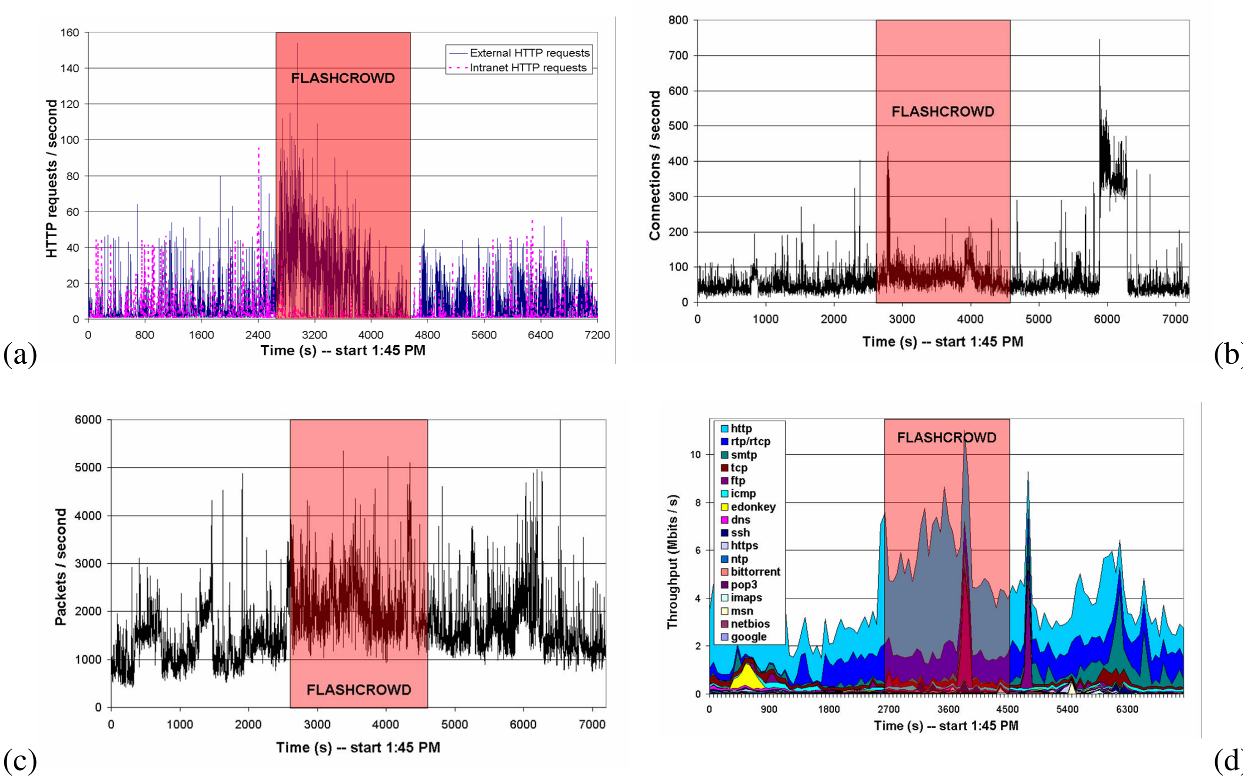


Fig. 2: **Flash Crowd.** (a) http requests, (b) connections, (c) packets and (d) distribution of throughputs per second time series. Fig. (d) follows a top-down approach: the application on top generates the larger traffic.

figure

flows and the packet rate on the LAAS access link. As expected, both plots show an increase in the average number of flows and in the average packet rate during the flash crowd.

Fig. 2(c) also shows increases in the average packet rate (20 min) before and (15 min) after the flash crowd experiment. To understand those increases, we have analyzed the different components of the traffic using the QoSMOS Traffic Designer tool [27] (cf. Fig. 2(d)). It has revealed that the increase occurring round 2 p.m., (before the FC) is caused by 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. The second peak, after the experiment, appears to be due to SMTP traffic. It can be given two explanations. First, many researchers at LAAS use web-mail. Because the server was significantly slowed down during the flash crowd experiment, they had to stop 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 took place at 3.15 p.m., just after the end of the flash crowd.

Note that the FC are used as examples of increase of traffic that are not attacks, and the peaks we just commented are other occurrences of legitimate increase of the traffic. We will see in Section V that they would no be detected as attacks or anomalies by the detection procedure we propose based on the statistical model developed in the very next section.

III. Non-Gaussian Long-Range Dependent Processes

A. *The Gamma arfima model*

Point process vs aggregated traffic. Computer network traffic consists of IP packets arrival processes. Thus, a general description can be formulated 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,...). It has long been observed that such arrival processes differ from simple standard Poisson or renewal processes, see for instance [28]. The inter-arrivals are not independent but display intricate correlation structures. It could be modeled using either non stationary Point processes [29] or stationary Markov modulated Point processes [30]. However, given the huge number of packets involved in any computer network traffic, these models would result in huge data sets. Therefore one often prefers 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 $k\Delta \leq t_l < (k+1)\Delta$. Various traffic models (including stationary or multifractal processes [31]) for X_Δ and $W_\Delta$have been proposed in e.g., [5], [32], [33], [34], [35], [36], [37] and a review of traffic models can be found in, e.g., [4], [38]. However, it is commonly accepted that the marginal distributions and auto-covariance functions are the two major factors that affect the performance of the network and hence that need to be accounted for in priority. Thus mainly concentrate here on the joint modeling of the marginal distributions and covariance functions of $X_\Delta(k)$ (modeling $W_\Delta(k)$ gives equivalent results not reported here for the sake of clarity).

Non-Gaussian marginals: Gamma distributions. By definition, $X_\Delta(k)$ is a positive random variable (RV). Hence, various works propose to describe the marginals of aggregated traffic with classical positive RV distributions such as (one-sided) exponential, log-normal, Weibull or gamma distributions [38]. Because of the packet arrival nature of the traffic, Poisson and exponential distributions are expected at small aggregation levels Δ for the marginals of $X_\Delta(k)$, while for data aggregated at larger Δs, Gaussian laws are relevant approximations, as suggested by a central limit argument. However none of them can satisfactorily model traffic marginals for a wide range of (small and large) Δs. A recurrent issue in traffic modeling lies in the choice of the relevant aggregation level Δ. This is an enduring question which answer involves together 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. Therefore, it would be of great interest to have at disposal a statistical model that may be relevant for a large range of values of Δ. In the present work, we choose to use Gamma distributions, $\Gamma_{\alpha,\beta}$, to model aggregated traffic because i) they naturally offer a smooth and continuous evolution from exponential to Gaussian laws, ii) the empirical studies reported here suggest that they are able to capture best the marginals of X_Δ over a wide range of Δs.

A $\Gamma_{\alpha,\beta}$ distribution is defined for positive RV as:

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

where $\Gamma(u)$ is the standard Gamma function (see e.g., [39]). It has mean $\mu = \alpha\beta$ and variance $\sigma^2 = \alpha\beta^2$. $\Gamma_{\alpha,\beta}$ laws are stable under multiplication and addition. If X is $\Gamma_{\alpha,\beta}$, then λX is $\Gamma_{\alpha,\lambda\beta}$, showing that β mostly acts as a multiplicative, or scaling, factor. For any two X_i and $i = 1, 2$ independent RVs $\Gamma_{\alpha_i,\beta}$, their sum $X = X_1 + X_2$ follows a $\Gamma_{\alpha_1+\alpha_2,\beta}$ law. Therefore, 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 $3 + 6/\alpha$. Hence, α is referred to as the shape parameter, controlling the (smooth and continuous) evolution from exponential to Gaussian distributions.

Long-range vs. short-range dependencies: ARFIMA covariance. After the seminal work reported in [22], it has been commonly accepted that computer network traffic is characterized by a long memory or long-range dependence property (cf. [40], [41]) (LRD). It is usually defined by the behavior at the origin of the power spectral density $f_{X_\Delta}(\nu)$:

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

LRD constitutes a central property in traffic modeling as it is likely to be responsible for the decrease of both the QoS and the performance of the network (see e.g., [42]). Incorporating it precisely into description models is therefore a crucial issue. It would allow to perform accurate and relevant network design (buffer size,...) and performance predictions (delay as a function of utility,...). LRD rules out the use of processes such as Poisson, Markov or Auto-Regressive Moving Average (ARMA) processes as well as other declinations such as Markov Modulated Poisson Processes for instance [43]. Instead, canonical long-range dependent processes such as fractional Brownian motion, fractional Gaussian noise [44] or Fractionally Integrated processes have been widely used to describe and/or analyze Internet times series (see [4] 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 a fruitful description of the packet arrival processes as pointed out in [45]. 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 (this has been explored for VBR video traffic, see for instance [46]). Therefore, we use the covariance function (or spectrum) of the Auto-Regressive Fractionally Integrated Moving Average (ARFIMA, or arfima hereafter) process [40], a natural choice as it allows to account for both short and long-range dependencies.

The covariance function of an arfima(P, d, Q) process, X, is fully defined via two polynomials of order P and Q, a fractional integration $\mathbf{D}^{-d}$, of order $-1/2 < d < 1/2$, and a power multiplicative constant σ^2. Its power spectral density, or spectrum (Fourier transform of the covariance function), takes the following analytical form:

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

for $-1/2 < \nu < 1/2$. It is evident that in the limit $|\nu| \to 0$, $f_X(\nu) \sim \sigma_\epsilon^2 |\nu|^{-2d}$ and (comparing to Eq. 2) that, for $d \in (0, 1/2)$, X is long-range dependent process. Hence, the parameter d accounts for the long-range dependence property and measures its “strength”. Conversely, the polynomials P and Q (i.e., the ARMA(P, Q) contribution to the arfima process) can be used to fit the spectrum at high frequencies or, equally, the covariance function at fine scales, in an independent and versatile way. Hence, they model the short range correlations.

Comments. To model aggregated Internet time series, we therefore propose to use a stochastic stationary non-Gaussian long-range dependent process: the Gamma (marginal) arfima (covari-

ance) process. This models benefits of a number of qualities and calls for a number of comments.
i) The specifications of the first and second order statistical properties described above 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. This difficult task to achieve is under investigation, with respect to detection purposes.
ii) The Gamma-arfima model is fully prescribed by a small number of parameters. For the analysis and the illustrations reported in the present work, we restrict ourselves to the use of arfima processes with polynomials P and Q of degree 1, hereafter labeled arfima(ϕ, d, θ) (ϕ and θ are the sole coefficient of the normalized polynomials P and Q). Then, the $\Gamma_{\alpha,\beta}$ - arfima(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 issues are concerned.
iii) As it is reported in the next section, the model fits Internet data for a large range of aggregation levels Δ. Therefore, it contains a form of covariance with respect to changes in the chosen resolution of analysis.
iv) Most of all, the proposed model not only accurately models traffic both with and without anomalies. Compared to other models, it proves useful to design anomaly detection procedures and to perform classification.

B. *Numerical synthesis*

Principles. The goal of this section is to present an original procedure that enables us to synthesize numerically sample paths (of any length) of stochastic processes with prescribed $\Gamma_{\alpha,\beta}$ marginals and arfima(θ, d, ϕ) covariance. Our construction consists of a three step procedure, stemming from ideas in [47], [48] and extending them to the Gamma case.
i) X, a $\Gamma_{\alpha,\beta}$ RV, can be obtained as $X = \sum_{i=1}^{i=2\alpha} Y_i^2$, where the Y_i's are zero-mean independent identically distributed Gaussian random variables, with variance σ_Y^2.
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 $Y_i(k)$, $\gamma_Y(l) = \sigma_{Y_i}^2 \rho_Y(l)$. The computation is derived below.
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., [49] for a review).

Obviously, the procedure we propose here works only for integer α. An efficient approximation for non integer α can be obtained.

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 $\mathbb{E}Y(k)Z(k,l) = 0$. One can derive from these results that $\mathbb{E}Y^2(k)Y^2(k+l) = \sigma_Y^4(1+2\rho_Y^2(l))$. Combining those findings with the fact that the Y_i are i.i.d. zero-mean Gaussian processes, tedious calculations not reported here lead to the following original and analytical result:

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

Traffic generators. The synthesis procedure described above can be extended to other types of marginals (log-normal, exponential, chi-squared, etc) and covariances (fractional Gaussian noise (fGn), kinked fGn, etc). Preliminary results are available in [50]. Other forms of statistical dependencies may as well be incorporated, including higher order statistics. Such synthesis procedures have been used for the validation of the analysis procedure, especially the estimation performance. Also, they constitute traffic generators for non-Gaussian long-range dependent traffic, that can be used for instance to feed simulation platforms aiming at estimating QoS and network performance.

C. Practical parameter estimation

The section details the practical estimation procedures for the gamma-arfima parameters used on actual data. While the estimation of α and β makes use of standard procedures, that of the arfima parameters is based on a original combination of techniques used for long-range and short-range correlations independently.

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 [51]. The joint distribution of n i.i.d. $\Gamma_{\alpha,\beta}$ variables can be obtained as a product of n terms as in Eq. 1. 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. It has been checked empirically from numerical simulations that this estimation procedure provides us with very accurate estimates even when applied to processes with long-range dependence [50].

Arfima parameter estimation. It is well known that the estimation of the long memory parameter is a difficult statistical task that has received a considerable amount of works and attention (see e.g. [49] for an up-to-date review), and so has the joint estimation of both long and short range parameters of the arfima(ϕ, d, θ) process. Full maximum likelihood estimation based on the analytical form of the spectrum recalled in Eq. 3 is possible but computationally heavy. Here we develop an original two step practically effective estimation procedure.

First the long-range dependence parameter d is estimated using a standard wavelet-based methodology [52]. Let $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$ denote an orthonormal wavelet basis, designed from the mother wavelet ψ_0 and $d_X(j,k) = \langle \psi_{j,k}, X_0 \rangle$ the corresponding *wavelet* coefficients. For any second order stationary process X, its spectrum $f_X(\nu)$ can be related to its wavelet coefficients through [18], [53]:

$$\mathbb{E}d_X(j,k)^2 = \int f_X(\nu) 2^j |\Psi_0(2^j \nu)|^2 d\nu, \tag{5}$$

where Ψ_0 stands for the Fourier transform of ψ_0 and $\mathbb{E}$ for the mathematical expectation. When X is a long-range dependent process, with parameter d, Eq. 2 implies that $\mathbb{E}d_X(j,k)^2 \sim C 2^{2jd}$, if $2^j \to +\infty$. It has been proven that 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. 5 above, this property 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 d. The plots $\log_2 S_j$ versus $\log_2 2^j = j$ are commonly referred to as logscale diagrams (LD). The full definition as well as the performance of this estimation procedure are detailed in [18], [53], [54].

Second, from this wavelet based estimate $\hat{d}$, we perform a fractional derivation of order $\hat{d}$ of X_Δ. It 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) [55] is then applied to estimate the ARMA parameters. Obviously, the major weakness of this two steps estimation procedure lies in the fact that be d poorly estimated, so would the ARMA parameters. However, the estimation performance of the procedure are studied numerically in [50] using synthetic $\Gamma_{\alpha,\beta}$

arfima(ϕ, d, θ) process.

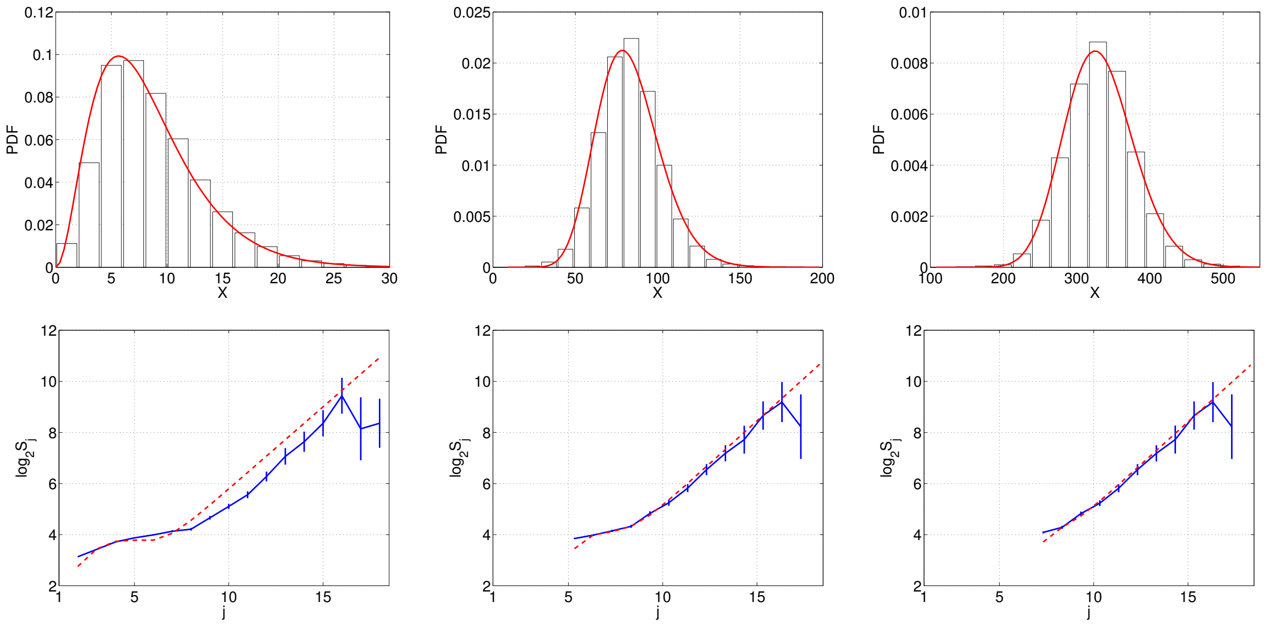


Fig. 3: **AUCK-IV.** $\Gamma_{\alpha,\beta}$ - arfima(ϕ, d, θ) fits for the marginals (top row) and covariances (bottom) for $\Delta = 10, 100, 400$ms (left to right); $j = 1$ corresponds to 10ms.
figure

IV. Traffic Modeling

The $\Gamma_{\alpha,\beta}$ - arfima(ϕ, d, θ) analysis procedures are applied, independently for different levels of aggregation, to the various traffic time series described in Section II, containing or not anomalies. For the theoretical modeling of X_Δ, stationarity is assumed. We first check the consistency of the results obtained for 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 [56], [57]. Then we estimate the parameters of the model for each chosen Δ. Results are analyzed and interpreted.

A. Regular Traffic

We give here detailed results for the **AUCK-IV** series and for the **Metrosec-ref1** series only. Similar results are obtained for the other series (cf. Table I), they are not reported here and are available on request.

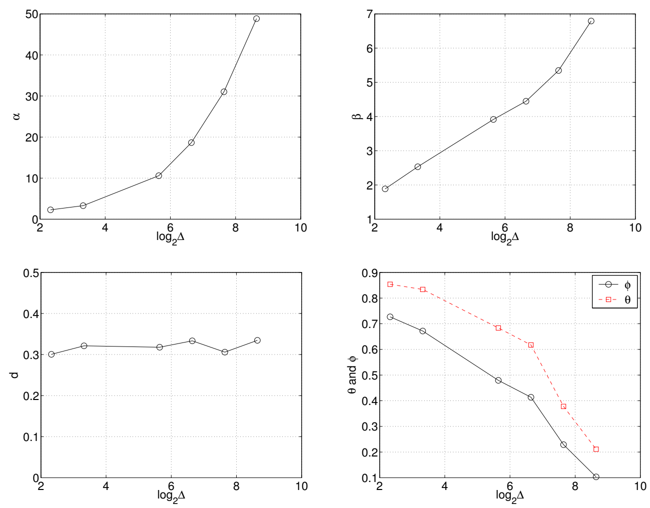


Fig. 4: **AUCK-IV.** Estimated $\Gamma_{\alpha,\beta}$ - arfima(ϕ, d, θ) parameters as a function of $\log_2 \Delta$ (with Δ in ms).
figure

Covariances. Figs. 3 and 5, bottom row, compare, for the two chosen time series respectively, the empirical LDs against their best fits obtained with the arfima covariance model. The latter are computed numerically from the combination of Eqs. 3 and 5 with the estimated $\hat{d}$, $\hat{\theta}$ and $\hat{\phi}$. This numerical procedure has been developed in collaboration with D. Veitch, cf. [58]. The LD plots illustrate the relevance of the arfima(ϕ, d, θ) fits of the covariances of X_Δ. As Δ increases, one can notice that the LDs almost correspond to coarser versions of the LD obtained at finer Δs, shifted toward upper scales. This is easily understood: aggregating data mainly consists of smoothing out details at fine scales though leaving coarse scales unaffected. The plots also show clearly that the onset of long memory occurs around $j = 10$, i.e., around 1s, hence providing us with a characteristic time scale separating short from long correlation time scales. As may have been expected, aggregation does not cancel the long memory and does not alter it. It can be checked in Figs. 4 and 6, bottom left, where the $\hat{d}$ remain remarkably independent of Δ. This underlines that long-range dependence captures a long-time feature of the traffic that has no inner time-scale. The situation is very different for the short-time correlations that are cancelled out when the aggregation level increases, see the Figs. 4 and 6, bottom right: $\hat{\phi}$ and $\hat{\theta}$ significantly decrease as Δ increases. One expects that they would be null (or identical) when the aggregation

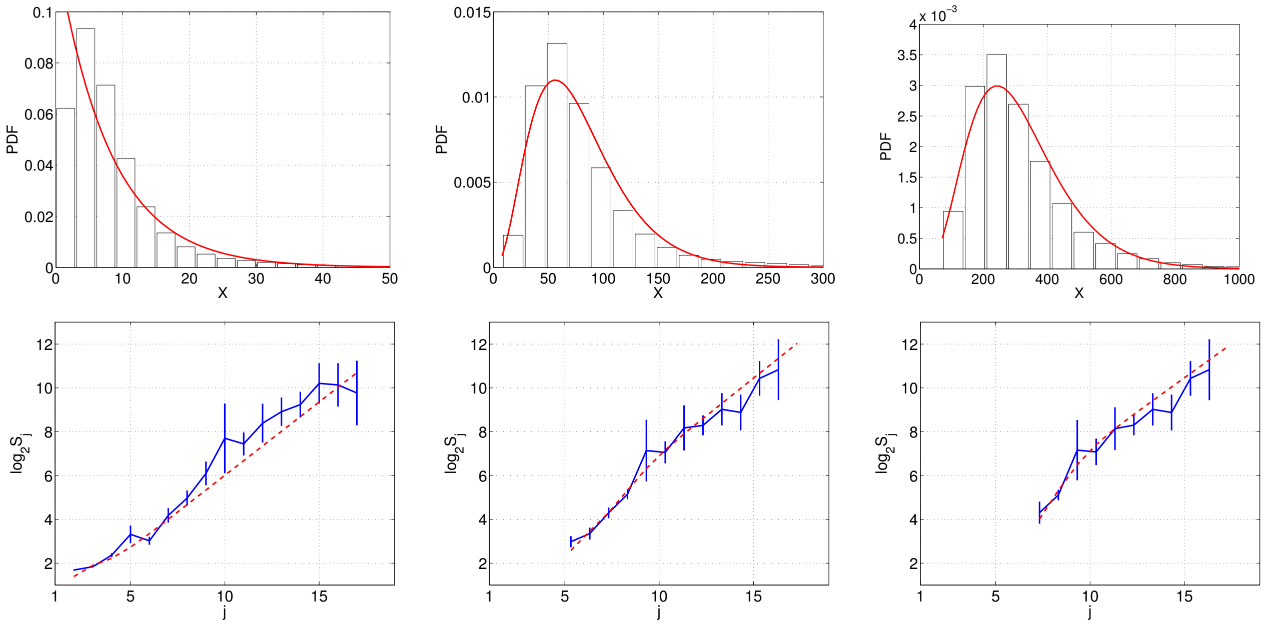


Fig. 5: **METROSEC-ref1.** $\Gamma_{\alpha,\beta}$ - arfima(ϕ, d, θ) fits for the marginals (top row) and covariances (bottom) for $\Delta = 10, 100, 400$ms (left to right). $j = 1$ corresponds to 10ms.

figure

level Δ becomes larger than the critical 1s time scale. Indeed, under aggregation, the covariance theoretically converges to that of a fractional Gaussian noise that turns out to be practically extremely close to that of an arfima$(0, d, 0)$ [40].

To finish with, let us note that for some time-series (CAIDA), higher order for the ARMA part of the arfima model proved necessary to model the covariance.

Marginals. Figs. 3 and 5, top rows, show empirical histograms, obtained from the chosen time series, together with the $\Gamma_{\alpha,\beta}$ fits. They illustrate the relevance of the $\Gamma_{\alpha,\beta}$ distributions to model the marginals of X_Δ, for a wide range of aggregation levels: 1ms $\leq \Delta \leq$ 10s. The adequacy of the fits has been characterized by means of χ^2 goodness-of-fit tests. 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 the other 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, it does not hold over a wide range of Δs. Conversely, the adequacy of the Gamma laws remains very satisfactory

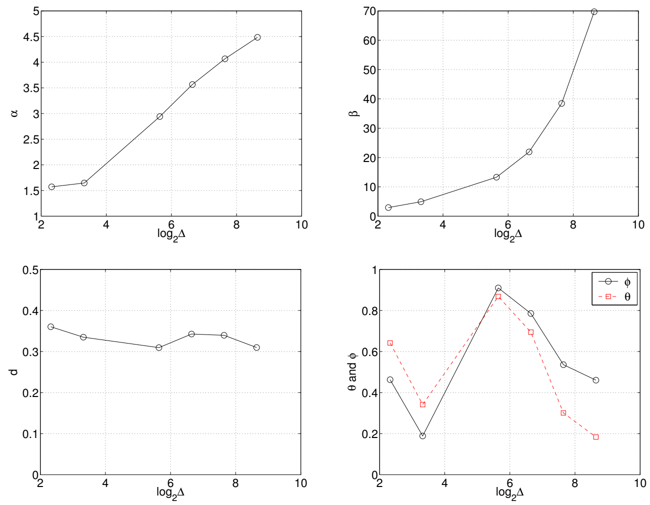


Fig. 6: **METROSEC-ref1.** Estimated parameters of $\Gamma_{\alpha,\beta}$ - arfima(ϕ, d, θ), as a function of $\log_2 \Delta$ (Δ in ms).
figure

over wide ranges of Δs, hence, they provides us with a scale-evolving characterization of the marginals of the traffic. $\Gamma_{\alpha,\beta}$ laws, by variation of their shape and scale parameters, offer a continuous and smooth evolution from pure exponential to Gaussian laws. These empirical facts are very much in favor of the use of Gamma laws to model computer traffic marginals, as is, from a theoretical point of view, their stability under addition property. Indeed, aggregation implies $X_{2\Delta}(k) = X_\Delta(2k)+X_\Delta(2k+1)$. Using stability under addition and 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 $\log_2 \Delta$, noted $\hat{\alpha}_\Delta$ and $\hat{\beta}_\Delta$. Significant departures from these behaviours under I.i.d. hypothesis are observed. The 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 of the existence of dependencies in the data and tell us the evolutions of α and β with Δ accommodate mainly for short range dependencies of X_Δ.

Synthetic time series. Using the synthesis method described in Section III-B, we produce numerical sample paths of the $\Gamma_{\alpha,\beta}$ - arfima(θ, d, ϕ) for different Δs. The parameters have been chosen so that they correspond to those measured on the **AUCK-IV** time series. Comparing Fig.

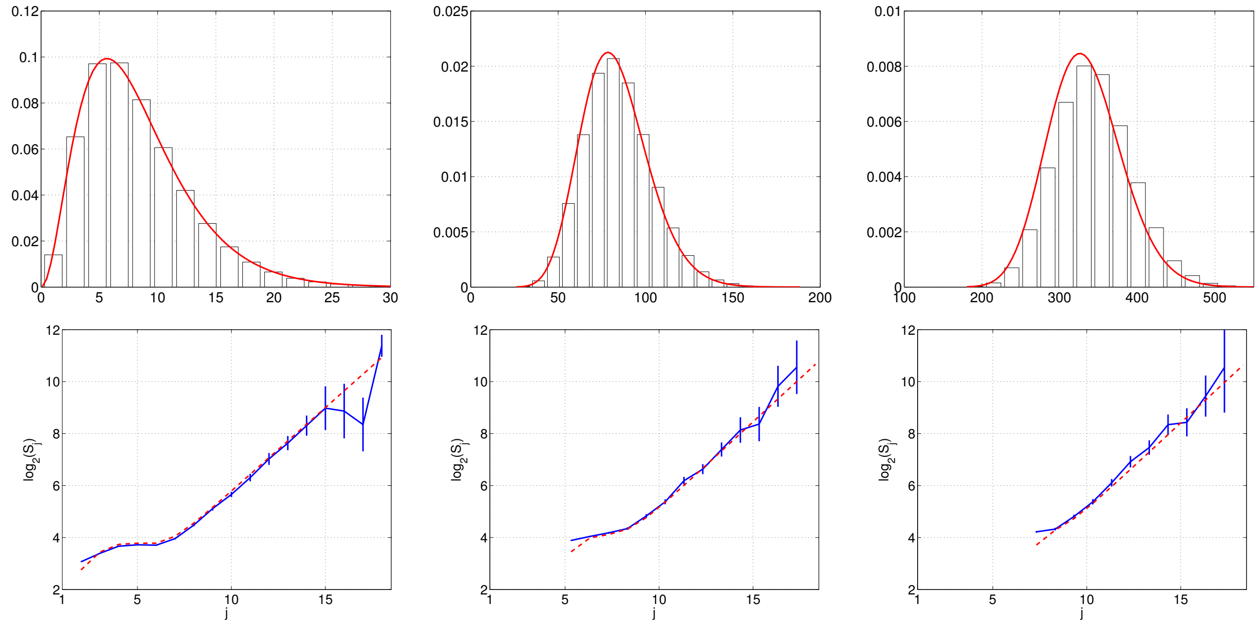


Fig. 7: **Synthetic data.** $\Gamma_{\alpha,\beta}$ - arfima(ϕ, d, θ) fits for the marginals (top row) and covariances (bottom) of synthetic data for $\Delta = 10, 100, 400$ms (left to right). The parameters correspond to those of **AUCK-IV**.

figure

7 with Fig. 3, the plots illustrate that the marginals (top row) and covariances (bottom row) of the synthetic time-series match those of the data.

B. Traffic with anomalies

1) DDoS Attack: **Covariance.** Fig. 8, left plot, presents the LDs for 1 hour block of data during the DDoS Attack ($\Delta = 1$ms) compared to those of 1h long regular traffic times series, recorded a couple of hours before and after the attack. The LD plots tell us first that an arfima(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 of aggregation levels.

Moreover, for the behaviors of the LDs at scales larger than 1s ($j = 10$ in Fig. 8, left plot), no discrepancies can be detected between before/after and during the attack. In particular, the long memory parameter $\hat{d}$ remains astonishingly constant. It tell us that long memory is not created by the attack, and also totally insensitive to its occurrence. The only change that can be noticed

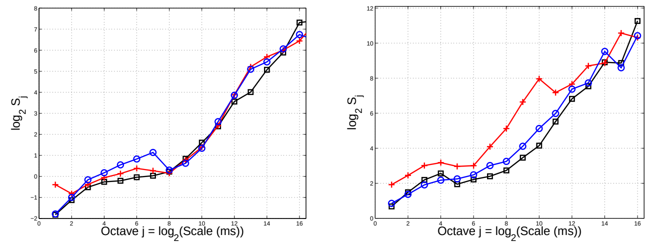


Fig. 8: **Logscale Diagrams.** For the DDoS (left) and for the Flash Crowd (right). For both events, the curves are given during the anomaly (crosses), and before (squares) or after (circles) the anomaly, taken as references for normal traffic.
figure

on the LDs consists of a relative increase of the short-time component (at scales j from 4 to 7) after the attack. The reason is that the traffic series after attack was recorded at night, with a lower traffic load. The LD was shifted upwards to show that the long memory parameter $\hat{d}$ (given by the slope) does not change, even when the load is smaller. Hence, DDoS attack cannot detect the from the LDs.

Marginals. Fig. 9, left column, illustrates, in two plots, that $\Gamma_{\alpha,\beta}$ distributions adequately fit the marginals of the traffic under DDoS attack. Fig. 10, left column, compares the evolutions of the estimated $\hat{\alpha}$ and $\hat{\beta}$ with respect to Δ for traffic during and before/after the DDoS event. Estimations are performed over 15 minute-long non-overlapping blocks of data. One sees that the functions of $\hat{\alpha}_\Delta$ and $\hat{\beta}_\Delta$ observed during the DDoS attack differ significantly from those corresponding to a regular traffic. The attack causes an immediate and sharp increase of α starting from the finest Δs, whereas under normal circumstances, α remains constant or with only small variations up to $\Delta \simeq 20$ms. The evolution is the inverse for β: it is decreasing from $\Delta \simeq 1$ms to $\Delta \simeq 30$ms during the DDoS attack, whereas it increases smoothly and regularly with Δ under normal traffic. These evolutions can receive several interpretations. First, because during the DDoS attack a large number of packets are emitted at the highest possible rate, the probability to observe *0 packet* within a window of size Δ decreases extremely fast to 0 even for small Δs, a major discrepancy compared to marginals observed on regular traffic that smoothly go to 0 when $X_\Delta \to 0$ (compare Figs. 3 or 5 to Fig. 9). It affects the shape of the marginals and hence the value of the α parameter, implying that α grows slowly with Δ for regular traffic

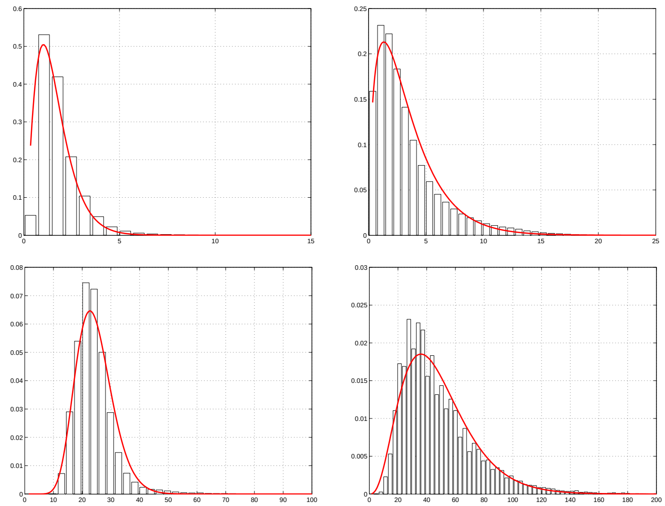


Fig. 9: **Marginals.** For the DDoS Attack (left) and for the Flash Crowd (right), empirical histograms of X_Δ and their $\Gamma_{\alpha,\beta}$ fits of the marginals, for $\Delta = 2$ms (top) and $\Delta = 32$ms (bottom).

figure

and much faster under DDoS attack. Second, the accelerated increase of α with respect to Δ under DDoS attack indicates that the marginal distributions of the traffic under attack tend to Gaussian laws much faster than under regular circumstances. Both properties constitute major statistical features that differentiate traffic under DDoS attacks from regular one.

2) Flash Crowd: **Marginals.** Fig. 9, 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. 10, right column, illustrates the $\hat{\alpha}(\Delta)$ and $\hat{\beta}(\Delta)$ curves observed during the event do not depart significantly from those recorded under normal circumstances. It is consistent with the fact that the flash crowd does not involve any mechanism that forbids the *0 packet per window* event as do DDoS attacks. Therefore, $\hat{\alpha}(\Delta)$ does not enable to detect the FC.

Covariance. Fig. 8, right plot, shows the LDs for two 15 min long blocks of data during the flash crowd ($\Delta = 1$ms) compared to those of 15 min long blocks of regular traffic, recorded before and after. On this plot, one sees that the LDs undergo a significant change during the flash crowd. From octaves $j = 8$ to $j = 10$, i.e., for scales of time ranging from 250ms to 1s, a strong peak of energy grows. Such a peak is never observed on traffic under regular circumstances and can

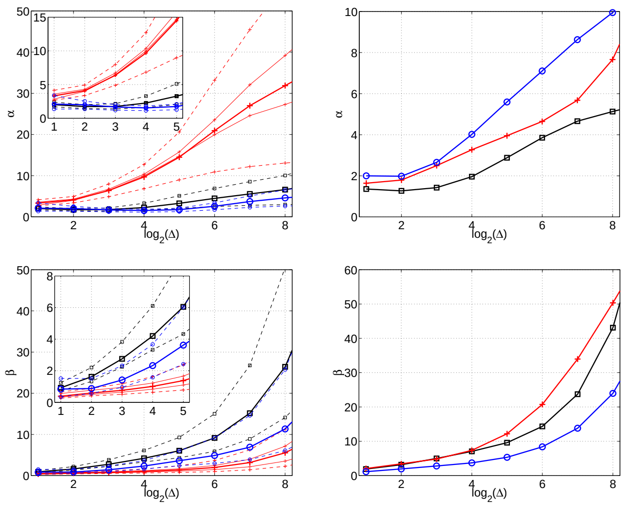


Fig. 10: **Estimated $\Gamma_{\alpha,\beta}$ parameters.** Estimation of $\hat{\alpha}$ (top) and $\hat{\beta}$ (bottom) as a function of $\log_2 \Delta$ (Δ in ms) for the DDoS Attack (left) and for the Flash Crowd (right). In both cases, the curves are given during the anomaly (crosses), and before (squares) or after (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).
figure

therefore be used to detect and characterize FC. Obviously, arfima(ϕ, d, θ) fits (not shown here) will fail to reproduce simultaneously the short range dependences, the long-range dependences and the energy peak. Goodness-of-fit tests between data and fitted models yield rejection also providing us with a relevant tool for designing a flash crowd detection procedure. Moreover the LRD parameter d, when estimated from octaves j coarser than those corresponding to the energy peak, does not notably depart from the one estimated before/after the FC. It tells that long memory is neither caused by the flash crowd nor modified by its occurrence. At most, the energy peak act as a masking effect in a subrange of time scales.

C. Discussions and conclusions

Let us summarize and comment our empirical findings. First, we have shown that the $\Gamma_{\alpha,\beta}$ - arfima(ϕ, d, θ) model reproduces accurately the marginals and both the short range and long-range correlations of traffic time series. It holds for a wide range of different regular traffic collected on various networks, as well as for traffic containing legitimate and illegitimate anomalies such as DDoS attacks and flash crowds. Second, the fact that the proposed model is versatile enough to work equally well for a wide range of aggregation levels is a key feature. This offers an alternative answer to the enduring issue in traffic modeling about the choice of the relevant aggregation level Δ. Hence choosing Δ a priori is uneasy. Therefore using a process that offers an evolutive modeling with Δ is of high interest. Moreover, the values of the parameters of the models obviously vary, possibly significantly from one traffic to another. But, one is not interested in the values themselves, but rather in the evolution of these parameters with respect to Δ. The detection procedures, detailed in the next section, specifically take advantage of the relevance of this multiresolution statistical description of the traffic.

V. DDoS Attack Detection

A. Distance-based detection procedure

As reviewed in the first section, real-time detection of anomalies in the traffic is a major issue of the internet of today. Anomaly detection is roughly divided between profile-based methods and signature-based procedures (or other methods relying on the analysis the attack mechanisms [59] or on application-dependent analysis [60]). Previous sections have shown that, even if one remains at the packet level and works with aggregated time-series, a joint analysis of the statistical profile at various time-scales of the series is sensitive to changes caused by anomalies in the traffic. From these properties, we propose a detection procedure that exploits the multiresolution nature of our statistical modeling.

Because the developed analysis is not based only on simple statistics (mean, variance), we were able to empirically discriminate between legitimate (FC) and illegitimate (DDoS) changes in traffic. In this section, we are dealing with the detection of illegitimate anomalies because we have a wider database of DDoS attacks than of FC; hence, FC are used as a benchmark to test the behaviors of the detection procedures in presence of a natural and long-lasting increase of the traffic (however, the small number of experiments do not allow us yet to assess statistical method to detect them specifically).

The detection scheme is as follows. The time series under analysis are split into adjacent non overlapping time windows of length T, starting at time lT and labeled by l. Independently for each time window l, one computes a *distance* between a statistical characteristic measured on window l and the same characteristic measured on an a priori chosen reference window. In a second step, one thresholds this distance to detect unexpectedly large deviations and hence anomalous traffic behaviors.

There exists a variety of distances that could be used (cf. e.g., [61] for an exhaustive review). For instance, one could use a generic non-parametric distance, such as the Kullback divergence for the marginal distribution or the log- spectral deviation for the spectrum (covariances). However, it would not explicitly take advantage of the relevance of the multiresolution model proposed in Section III. Therefore, we base the detection on a distance computed from the parameters of the model, especially $\hat{\alpha}_{\Delta}(l)$ and $\hat{\beta}_{\Delta}(l)$ for a collection a dyadic scales Δ, going from a fine scale $2^1\Delta_0$ to a large scale $2^J\Delta_0$. A simple, yet robust, Mean Quadratic Distances (MQD) is defined as (Δ_0 is left out for the ease of notation):

$$D_{\alpha}(l) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\alpha}_{2^j}(l) - \hat{\alpha}_{2^j}(ref)\right)^2, \quad (6)$$

$$D_{\beta}(l) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\beta}_{2^j}(l) - \hat{\beta}_{2^j}(ref)\right)^2. \quad (7)$$

After the computation of the distance, one a priori choice is left: the threshold level (values under the threshold are deemed as normal traffic, and values above are considered as anomalies). In the present work, a collection of threshold values is systematically explored so as to derive the statistical performance for the detection procedure.

B. Results and Statistical Performance

1) Results: In the results detailed below, the reference window consists of T_{Ref} minutes of traffic before the occurrence of the attacks and therefore assumed to be regular traffic. We used both $T_{Ref} = 1$ min and $T_{Ref} = 10$ min for comparison and we set $\Delta_0 = 1$ms and $J = 10$ in agreement with the results reported in Section IV.

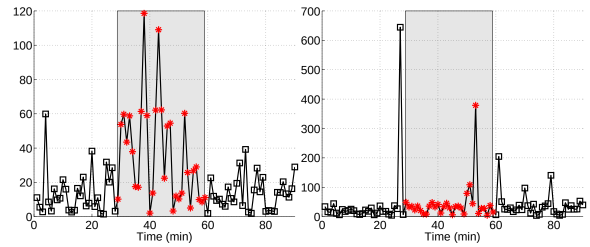


Fig. 11: **MQD for traffic containing a DDoS Attack.** $D_\alpha(l)$ (left) and $D_\beta(l)$ (right), computed on non overlapping 1 min time windows for run Iperf-III. Time windows containing the attack are those in the grey area.

figure

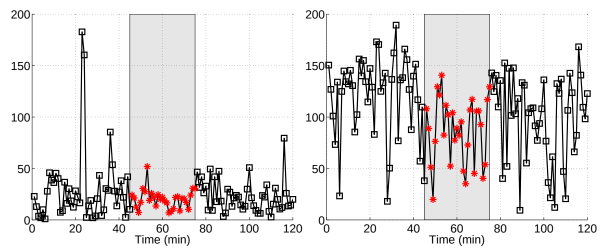


Fig. 12: **MQD for traffic containing a legitimate FC anomaly.** $D_\alpha(l)$ (left) and $D_\beta(l)$ (right), computed on non overlapping 1 min time windows for experiment FC-1. Time windows corresponding to the FC are those in the grey area.

figure

MQDs are depicted in Fig. 11 for traffic containing an attack (here the Iperf-III run). On the left, one sees that $D_\alpha(l)$ takes large values within time windows l containing the attack. It clearly confirms that the occurrence of the anomaly significantly alters the dependency of α with respect to Δ, hence increases the MQD. Conversely, one observes that $D_\beta(l)$ remains mostly stable and is not significantly shifted by the occurrence of the attack. Remember that β is a scale parameter mostly sensitive to the intensity of the traffic. The attacks do not correspond to traffic increase with unchanged correlation structure, but rather to significant dynamical and statistical changes. The large values observed in the $D_\beta(l)$ plot correspond to time windows that do not satisfy the χ^2 goodness-of-fit test because they contain both regular and under attack traffics, yielding aberrant estimates. Note that the large values occur at the start and stop times of the attack.

Conversely, for the Flash Crowd experiments, the traffic is not seen as a clear-cut anomaly. MQDs are plotted in Fig. 12 for FC-1 and we see no particular increase of these distances during the FC. This is because the statistical characterization of the FC by means of α_Δ and β_Δ is insensitive to this kind of variation of traffic, that is mainly a small increase of the traffic but with exactly the same variability, as argued in Section IV-B. Hence, the multiscale characterization through α_Δ and β_Δ is mostly unchanged and the distances $D_\alpha(l)$ and $D_\beta(l)$ to the reference traffic profile are not significantly different during the FC from those without anomaly.

2) Experimental Statistical Performance: The statistical performance of detection procedure are usually assessed in terms of Receiver Operational Characteristics (or ROC curves), consisting of the correct detection probability P_D vs. false alarm probability P_F. Therefore, one plots the curves $P_D = f(\lambda)$ vs. $P_F = g(\lambda)$, parametrized by the threshold value λ. They are obtained empirically from our database as follows. Because we know which time window contains the attack, and which does not, we are able to calculate for each detection level λ both P_D and P_F; the probability of detection P_D is the ratio of the number of windows containing the attack whose distance is above the threshold to the total number of windows with anomaly; the probability of false alarm P_F is the ratio of the number of windows containing no attack whose distance is yet above the threshold to the total number of windows without anomaly.

Plots P_D vs. P_F and $P_D = f(\lambda)$ and $P_F = g(\lambda)$ are shown on Fig. 13, on the example of the Iperf-III run, as an illustration. The ideal set point (all attacks would be detected and no false alarm raised) is the left upper corner. The worst case is the diagonal, when the results do not significantly differ from those obtained at random. Fig. 13 clearly shows the efficiency of the proposed detection procedure. ROC curves were calculated for each and every trace containing anomalies, listed in Table II. All plots can obviously not be displayed, instead, we report in Table III P_D for two a priori chosen levels of false alarm, set respectively to 10% and 20%. They are obtained by reading on ROC curves P_D for the chosen P_F level.

3) Discussion: The results are satisfactory. In the worst cases, detection is possible with better chance than at random. For illegitimate anomalies with high impact on the traffic (for instance the runs IPerf-B, X or Trinoo-T, cf. Table II), the detection probability is really high. For attacks with very low intensity (such as IPerf-I, II, III or Trinoo-M) and hence little impact on traffic volume profiles, detection rates, even if low at first sight, are encouraging as most traditional IDS based on simple mean and variance statistics would totally miss them. Moreover, let us mention that the use of the mean $\hat{\mu}_\Delta$ or variance $\hat{\sigma^2}_\Delta$of X_Δ as functions of Δ yields curves that exhibit identical forms with or without anomaly (plots not reproduced here). Hence, sample mean and variance estimates are blind to anomalies.

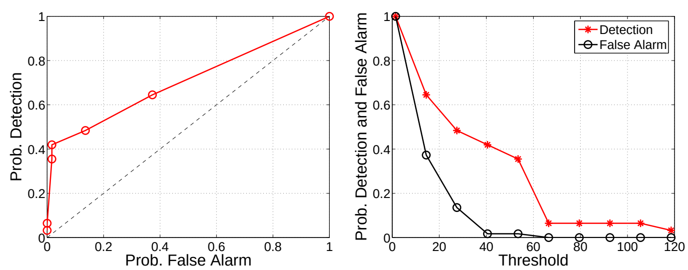


Fig. 13: **Statistical performance.** Left, Detection probability P_D vs. False Alarm probability P_F, $P_D = f(P_F)$, right, $P_D = f(\lambda)$ and $P_F = g(\lambda)$ for $D_\alpha(l)$.
figure

The performance reported here uses only statistics over one minute. Obviously better performance are expected from the use of a detection scheme over several minutes, combining the score functions over a past of a few windows, because, during anomalies, the probability of crossing the threshold on 2 successive window is much larger than it is for a single window. This trade-off between increased detection performance and increase of the delay in the alert time (a few minutes instead of 1), needs to be further explored.

An interesting feature of the detection method based on the multiscale marginal modeling ($\Gamma_{\alpha_\Delta,\beta_\Delta}$ for $\Delta = 2^j\Delta_0$ with $j = 1, ..., J$) lies in its being able to differentiate between legitimate anomalies and illegitimate ones. For instance, the Flash Crowd anomalies, which consists of a regular increase of traffic, are not detected as attacks: the assigned detection probabilities are close to the false alarm rate – this is not detection, this is false alarm! Further along the way, it gives the capability to classify between classes of anomalies, when combining this with other characteristics of the traffic (such as the arfima parameters that are not yet used in the proposed attack detection scheme). Another perspective is to combine this detector that uses only the profile of the traffic with methods based on signature in a full-fledge IDS. It is a accepted fact that to achieve good efficiency, one has to use jointly both approaches.

VI. CONCLUSIONS AND FUTURE WORKS

In the present work, we have proposed a non-Gaussian long-range dependent process, the $\Gamma_{\alpha,\beta}$ - arfima(P, d, Q) process, to model the first and second order statistics of aggregated computer network traffic time series. We have fully described operational parameter estimation procedures and we have defined original numerical synthesis procedures. We have shown from a large variety of standard reference traffic time series that the $\Gamma_{\alpha,\beta}$ - arfima(P, d, Q) process constitutes a relevant versatile model, and this for a very large range of aggregation levels Δ. Moreover, its parameters are smoothly evolving with Δ hence providing us with a useful statistical characterization of regular traffic. We have also shown 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. A detection procedure using this model, and yielding satisfactory results, has been defined.

This work can be further developed along numerous directions. First, the numerical synthesis procedures can be used for traffic generation, performance assessment, and on line traffic samples prediction, $X_\Delta(T+\tau)$ for $\tau > 0$. 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. Second, thanks to the experimental platform being developed within the METROSEC project, we intend to further explore the zoo of anomalies. Third, we are working on extending the proposed detection scheme to the use of other statistical distances (Kullback divergence, etc) that should help in identifying changes in the traffic statistical characterizations and classify them as legitimate or illegitimate. Our ultimate goal is to develop network based (protocols, architectures,...) strategies to improve the robustness of the network against attacks, and thus to help maintaining the targeted level of QoS.

VII. ACKNOWLEDGMENTS

The authors acknowledge the help of CRI ENSLyon, and numerous colleagues from the METROSEC project, for their help in conducting data collection and in performing anomalies. They gratefully acknowledge all the people who freely accepted to take part into the scheduled flash crowd events analyzed here. They also gratefully acknowledge colleagues from the major internet traces repositories for making their data available to us or having performed the pre-formatting of some of the time series used here: S. Marron, F. Hernandez-Campos and C.

Park from UNC (USA); D. Veitch and N. Hohn from CubinLab (Australia). This work has been supported by the French MNRT ACI *Sécurité et Informatique* 2004 grant, within the METROSEC project. The authors also thank anonymous reviewers for their fruitful comments.

References

[1] 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*, Washington, DC, USA, 1996, p. 171, IEEE Computer Society.

[2] P. Abry, Barananiuk, P. Flandrin, R. Riedi, and D. Veitch, "Multiscale network traffic analysis, modeling, and inference using wavelets, multifractals, and cascades," *IEEE Signal Processing Magazine*, vol. 3, no. 19, pp. 28–46, May 2002.

[3] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent packet traffic," *ACM/IEEE transactions on Networking*, vol. 4, no. 2, pp. 209–223, 1996.

[4] K. Park and W. Willinger, "Self-similar network traffic: An overview," in *Self-Similar Network Traffic and Performance Evaluation*, Kihong Park and Walter Willinger, Eds., pp. 1–38. Wiley (Interscience Division), 2000.

[5] A. Feldmann, A.C. Gilbert, and W. Willinger, "Data networks as cascades: Investigating the multifractal nature of internet WAN traffic," in *SIGCOMM*, 1998.

[6] V. Paxson, "Bro: a system for detecting network intruders in real-time," *Computer Networks Journal*, vol. 31, no. 23–24, pp. 2435–2463, 1999.

[7] J. Brutlag, "Aberrant behavior detection in time series for network monitoring," in *USENIX System Administration Conference*, New Orleans, Dec. 2000.

[8] 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 Security*, vol. 12, no. 3, pp. 235–248, 1993.

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

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

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

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

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

[14] J. Yuan and K. Mills, "DDoS attack detection and wavelets," Tech. Rep., National Institute of Standards and Technology, 2004.

[15] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *SIGCOMM*, Aug. 2004.

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

[17] 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.

[18] 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"*, vol. 45, no. 3, pp. 878–897, Apr. 1999.

[19] L. Li and G. Lee, "DDoS attack detection and wavelets," in *International Conference on computer communications and networks*, Aug. 2003.

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

[21] 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.

[22] 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*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[23] 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.

[24] 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, Apr. 2000.

[25] "IPERF – The TCP/UDP Bandwidth Measurement Tool," http://dast.nlanr.net/Projects/Iperf/.

[26] "TRINOO – Distributed network DoS tool," http://staff.washington.edu/dittrich/misc/trinoo.analysis.

[27] "QoSMOS Traffic Designer," http://www.qosmos.net.

[28] V. Paxon and S. Floyd, "Wide-area traffic: The failure of Poisson modeling," *ACM/IEEE transactions on Networking*, vol. 3, no. 3, pp. 226–244, June 1995.

[29] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido, "A non stationary Poisson view of the internet traffic," in *INFOCOM*, 2004.

[30] A. Andersen and B. Nielsen, "A Markovian approach for modelling packet traffic with long range dependence," *IEEE journal on Selected Areas in Communications*, vol. 5, no. 16, pp. 719–732, 1998.

[31] M. Crouse, R. Riedi, V. Ribeiro, and R. Baraniuk, "A Multifractal Wavelet Model for Positive Processes," in *IEEE-SP International Symposium on Time-frequency and Time-scale Analysis*, October 1998, Submitted.

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

[33] N. Hohn, D. Veitch, and P. Abry, "Multifractality in TCP/IP traffic: the case against," *Computer Networks Journal*, to appear, 2005.

[34] S. Sarvotham, R. Riedi, and R. Baraniuk, "Connection-level analysis and modeling of network traffic," Tech. Rep., ECE Dept., Rice Univ., 2001.

[35] M. Taqqu, V. Teverosky, and W. Willinger, "Is network traffic self-similar or multifractal ?," *Fractals*, vol. 5, no. 1, pp. 63–73, 1997.

[36] Z. Zhang, V. Ribeiro, S. Moon, and C. Diot, "Small time scaling behavior of internet backbone traffic: an emprirical study," *INFOCOM*, Mar. 2003.

[37] K. Maulik and S. Resnick, "The self-similar and multifractral nature of a network traffic model," *Stochastic models*, vol. 19, no. 4, pp. 549–577, 2003.

[38] Benjamin Melamed, "An overview of TES processes and modeling methodology," in *Performance/SIGMETRICS Tutorials*, 1993, pp. 359–393.

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

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

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

[42] B. Tsybakov and N.D. Georganas, "Self similar processes in communications networks," *IEEE Trans. on Info. Theory*, vol. 44, no. 5, pp. 1713–1725, 1998.

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

[44] I. Norros, "On the use of fractional Brownian motion in the theory of connectionless networks," *IEEE journal on Selected Areas in Communications*, vol. 13, no. 6, 1995.

[45] 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*, vol. 8, no. 51, pp. 2229–2244, Oct. 2003.

[46] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. Kaye, "Modeling and simulation of self-similar Variable Bit Rate compressed video: a unified approach," in *ACM SIGCOMM*, Cambridge, UK, Aug. 1995.

[47] S.B. Lowen, S.S. Cash, M. Poo, and M.C. Teich, "Quantal neurotransmitter secretion rate exhibits fractal behavior," *The journal of Neuroscience*, vol. 17, no. 15, pp. 5666–5677, Aug. 1997.

[48] S.B. Lowen and M.C. Teich, *Fractal-Based Point Processes*, Wiley, Hoboken,NJ, 2005.

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

[50] A. Scherrer and P. Abry, "Marginales non gaussiennes et longue mémoire : analyse et synthèse de trafic Internet," in *Colloque GRETSI-2005*, Louvain-la-Neuve, Belgique, Sept. 2005.

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

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

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

[54] P. Abry and D. Veitch, "Wavelet analysis of long-range dependent traffic," *IEEE Trans. on Info. Theory*, vol. 44, no. 1, pp. 2–15, Jan. 1998.

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

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

[57] D. Veitch and P. Abry, "A statistical test for the time constancy of scaling exponents," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2325–2334, Oct. 2001.

[58] D. Veitch, P. Abry, and M. S. Taqqu, "On the automatic selection of the onset of scaling," *Fractals*, vol. 11, no. 4, pp. 377–390, 2003.

[59] S. Kandula, D. Katabi, M. Jacob, and A. Berger, "Botz-4-sale: survicing organized DDoS attacks that mimic Flash Crowds," in *USENIX' NSDI'05*, A. Vahdat and D. Wetherall, Eds., Boston, MA, May 2005.

[60] S. Farraposo, K. Boudaoud, L. Gallon, and P. Owezarski, "Some issues raised by DoS attacks and the TCP/IP suite," in *SAR' 2005*, Batz-sur-mer, France, June 2005.

[61] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Processing*, vol. 18, pp. 349–369, 1989.

Antoine Scherrer is a PhD student in computer sciences at the Ecole Normale Supérieure of Lyon. He obtained a Master degree in Computer sciences at the University of Lyon in 2003. His works mainly concern analysis and synthesis of Internet traffic as well as on-chip traffic. He will defend his PhD in december 2006.

Nicolas Larrieu is an Assisant Professor at ENAC (the French Civil Aviation University) in the network department. He completed his Ph.D. in July 2005 in Computer Science at LAAS-CNRS (Laboratory for Analysis and Architecture of Systems), in Toulouse, France. He received his M.S. also in Computer Science at INSA (Institut National des Sciences Appliques) of Toulouse in 2002. His main research interests include monitoring based characterization and modeling of Internet traffic to improve networking and QoS.

Philippe Owezarksi Philippe Owezarski is a full time researcher of CNRS (the French center for scientific research), working at LAAS (Laboratory for Analysis and Architecture of Systems), in Toulouse, France. He got a PhD in computer science in 1996 from Paul Sabatier University, Toulouse III. His main interests deal with high speed and multimedia networking and more specifically on IP networks monitoring, and Quality of Service and security enforcement based on measurements.

During year 2000, he spent 9 months working for Sprint ATL in Burlingame, California. There he has been working on the Sprint monitoring IPMON project, and focused mainly on actual TCP flows analysis. Back to LAAS, Philippe Owezarski has been one of the main contributors of a monitoring project in France – METROPOLIS – and has been leading a French steering group on IP networks monitoring. Now, he is contributing to the European EuQoS and E-NEXT projects related to QoS enforcement in next generation networks, and leading the French MetroSec project aiming at increasing the robustness of the Internet against DoS and DDoS attacks. He was also TPC co-chair of the first edition of CoNEXT, a conference which aims at becoming one of the largest events in networking in the world.

Pierre Borgnat was born in Poissy, France, in 1974. He made his studies at the cole Normale Supérieure de Lyon, France, receiving the Prof.-Agrg de Sciences Physiques degree in 97, a Ms. Sc. in physics in 99 and defended a Ph.D. degree in physics and signal processing in 2002. In 2003/2004, he spent one year in the Signal and Image Processing group of the IRS, IST (Lisbon, Portugal). Since October 2004, he has been a full-time CNRS researcher with the Laboratoire de Physique, NS Lyon. His research interests are in statistical signal processing of non-stationary processes (time-frequency, time deformations) and scaling phenomena (time-scale, wavelets) for complex systems (turbulence, networks, biological systems). He is currently involved in projects on Internet traffic measurements and modeling, especially for security enforcement based on measurements.

Patrice Abry was born in Bourg-en-Bresse, France in 1966.

He received the degree of Professeur-Agrégé de Sciences Physiques, in 1989 at Ecole Normale Supérieure de Cachan and completed a PhD in Physics and Signal Processing, at Ecole Normale Supérieure de Lyon and Université Claude-Bernard Lyon I, in 1994.

Since october 95, he is a permanent CNRS researcher, at the laboratoire de Physique of Ecole Normale Supérieure de Lyon.

Patrice Abry received the AFCET-MESR-CNRS prize for best PhD in Signal Processing for the years 93-94 and is the author of a book "Ondelettes et Turbulences - Multirésolution, algorithmes de décompositions, invariance d'échelle et signaux de pression", published in october 97, by Diderot, éditeur des Sciences et des Arts, Paris, France. He also is the coeditor of a book in French entitled "Lois d'échelle, Fractales et Ondelettes", Hermès, Paris, France, 2002.

His current research interests include wavelet-based analysis and modelling of scaling phenomena and related topics (self-similarity, stable processes, multi-fractal, 1/f processes, long-range dependence, local regularity of processes, inifinitely divisible cascades, departures from exact scale invariance…). Hydrodynamic turbulence and the analysis and modelling of computer network teletraffic are the main applications under current investigation. He is also involved in the study of baroreflex sensitivity with a French medical group at University Claude Bernard Lyon I. He recently started a wavelet based detection/analysis of Acoustic Gravity Waves in Ionosphere.

TABLE II: **Traffic with anomalies.** Description of the studied traces containing anomalies. Runs R stands for the reference attack, and FC-1 and FC-2 for the flash crowd used to validate our model on traffic with anomaly. Upper part, FC-1 and FC-2 are flash crowd anomalies (FC-1 is plotted in 2 and used to validate our model on traffic with anomaly). Middle part, DDoS attacks performed with IPerf in 2004 and 2005; run R stands for the reference attack (see Fig. 1). Lower part, attacks performed with Trinoo in 2006. For each trace and attack, t_i, t_a, T, T_A stand respectively for the starting times (in local time) and durations (in second) of the whole trace and of the anomalies. D, V and I refer respectively to the controlled throughput of each attacking source (in Mbps), the length of each attack packet (in bytes), and the attack relative intensity (i.e., the ratio between the sum of all attack flows and the average throughput on the LAAS link during attack).

table

Id	t_i	T(s)	t_a	T_A(s)	D	V	I (%)
FC performed by human							
FC-1	13:45	7200	14:30	1800	-	-	31.27
FC-2	15:00	7200	15:45	1800	-	-	18.35
DDoS performed with Iperf							
R	17:30	60000	20:00	20000	0.5	60	33.82
I	09:54	5400	10:22	1800	0.25	1500	17.06
II	14:00	5400	14:29	1800	0.5	1500	14.83
III	16:00	5400	16:29	1800	0.75	1500	21.51
IV	10:09	5400	10:16	2500	1.0	1500	33.29
V	10:00	5400	10:28	1800	1.25	1500	39.26
A	14:00	5400	14:28	1800	1	1000	34.94
B	16:00	5400	16:28	1800	1	500	40.39
C	10:03	5400	10:28	1800	1	250	36.93
X	14:00	5400	14:28	1800	5	1500	58.02
DDoS performed with Trinoo							
tM	18:21	5400	18:58	601	0.1	300	4.64
tN	18:22	3600	18:51	601	0.1	300	15.18
tT	18:22	3600	18:51	601	8	300	82.85

TABLE III: **Detection rates.** For for each attack, Detection probabilities obtained for a fixed False Alarm probability set at 10% (left) and 20% (right). We recall here the type and name of each run, and the intensity of the anomaly (computed as in Table II).

table

Type of Anomaly	performed with	Id	Intens. (%)	P_D 10%	P_D 20%
FC	Humans	1	31.27	04	14
FC	Humans	2	18.35	19	25
DDoS	Iperf	R	33.82	91	93
DDoS	Iperf	I	17.06	51	64
DDoS	Iperf	II	14.83	48	54
DDoS	Iperf	III	21.51	48	58
DDoS	Iperf	IV	33.29	33	50
DDoS	Iperf	V	39.26	18	40
DDoS	Iperf	A	34.94	21	50
DDoS	Iperf	B	40.39	81	87
DDoS	Iperf	C	36.93	52	58
DDoS	Iperf	X	58.02	93	96
DDoS	Trinoo	tM	4.64	27	50
DDoS	Trinoo	tN	15.18	54	54
DDoS	Trinoo	tT	82.85	82	82