NON-LINEAR REGRESSION FOR BIVARIATE SELF-SIMILARITY IDENTIFICATION - APPLICATION TO ANOMALY DETECTION IN INTERNET TRAFFIC BASED ON A JOINT SCALING ANALYSIS OF PACKET AND BYTE COUNTS

Jordan Frecon[1], Romain Fontugne[2,3], Gustavo Didier[4], Nelly Pustelnik[1], Kensuke Fukuda[2], Patrice Abry[1,2,3]

[1] CNRS, Physics Department, ENS Lyon, France, firstname.lastname@ens-lyon.fr
[2] The National Institute of Informatics, Tokyo, Japan, firstname@nii.ac.jp
[3] Japanese French Laboratory for Informatics, Tokyo, Japan,
[4] Mathematics Department, Tulane University, New Orleans, USA, gdidier@tulane.edu

Supported by ANR BLANC 2011 AMATIS BS0101102,
PA and RF acknowledge specific supports from NII[1], ENS de Lyon[3] and JFLI[2]

ABSTRACT

Internet traffic monitoring is a crucial task for network security. Self-similarity, a key property for a relevant description of internet traffic statistics, has already been massively and successfully involved in anomaly detection. Self-similar analysis was however so far applied either to byte or Packet count time series independently, while both signals are jointly collected and technically deeply related. The present contribution elaborates on a recently proposed multivariate self-similar model, Operator fractional Brownian Motion (OfBm), to analyze jointly self-similarity in bytes and packets. A non-linear regression procedure, based on an original Branch & Bound resolution procedure, is devised for the full identification of bivariate OfBm. The estimation performance is assessed by means of Monte Carlo simulations. Further, an Internet traffic anomaly detection procedure is proposed, that makes use of the vector of Hurst exponents underlying the OfBm based Internet data modeling. Applied to a large set of high quality and modern Internet data from the MAWI repository, proof-of-concept results in anomaly detection are detailed and discussed.

***Index Terms*—** Internet traffic, anomaly detection, bivariate self-similarity, non linear regression, branch and bound.

1. INTRODUCTION

Internet traffic monitoring. Internet traffic monitoring and modeling constitute crucial tasks for network engineering and design, resource allocation, performance and service optimization, and for security assessment. Notably, anomalous traffic detection has received a considerable amount of attention and research efforts, as malicious traffic may have dramatic consequences both for users and operators. Anomaly detection in Internet traffic context is however highly challenging, because of the strongly heterogeneous natures of networks, applications, protocols and user behaviors. Normal or legit traffics may per se show a large diversity, whose actual definition hence remains a difficult issue. Anomalous behaviors may correspond to an even larger level of heterogeneity, from the known Distributed Denial of Service (DDoS) and port scanning to the potential occurrence of behaviors never encountered before, cf. e.g., [1, 2, 3] for reviews. In addition, the nature of the available data may depend on the network or on the operator, consisting either of systematic IP (Internet Protocol) packet timestamps, volume in bytes, and 5-tuples[1] or of higher granularity recordings (applications, protocol,...). The present work analyzes times series consisting of the counts of IP Pkt $Y_{Pkt,\Delta}(t)$ (or of bytes $Y_{Byt,\Delta}(t)$) in consecutive time bins, of size Δ, obtainable from packet header traces only.

Related works: anomaly detection. The literature dedicated to anomaly detection in Internet traffic is huge (cf. e.g., [1, 2, 3, 4, 5] for reviews) and an exhaustive description is out of the present scope. The focus here is on anomaly detection based on aggregated time series, thus respecting privacy, as opposed to techniques that rely on packet payload examination. Because of the ever changing nature of Internet traffic, extracting a reference for normal traffic against which anomalies could be compared is an almost impossible task. Instead, random projection tools, also referred to as *hashing* procedures or *sketching*, were used for the automated construction of self-referenced traffic [6] and [4, 7]. Often, anomaly detection relies on the statistical modeling of regular traffic. Amongst numerous attempts, self-similarity and fractional Brownian motion (fBm) [8] have been shown to provide relevant and robust models for Internet traffic statistics across a large variety of networks, traffics and from the early ages of Internet to the most recent data collections [9, 10, 11, 12, 13]. In essence, self-similarity implies that Internet traffic temporal dynamics are not driven by specific and characteristic time scales, but rather involve a large continuum of time scales, whose relation is quantified by the so-called Hurst parameter H, often empirically found in $H \in (0.8, 1)$. So far, however, in most anomaly detection procedures relying on self-similarity, analysis and modeling remained univariate: Byte and Pkt aggregated count time series were analyzed independently, cf. [9, 12, 4, 7, 14]. It is even often much debated and controversial to decide whether self-similarity analysis should be conducted on byte-counts or packet-counts time series. Often, parameters H measured on each type of data differ, thus calling into question the validity of the celebrated queuing mechanism that relates self-similarity to the heavy tailed nature of the distributions of the objects to be distributed on the Internet (cf. e.g., [15, 16, 17]) or raising questions such as: Do the mechanisms apply to packet, bytes, both? if so, why should H differ between packet and bytes? To address such issues, the joint availability of Pkt and Byte count time series has rarely been exploited in bivariate self-similarity analysis, see a contrario [18] and [7] for a

[1]The standard 5-tuple consists of five IP packet header fields: IP address and port number for source and destination, and IP protocol carried (TCP, UDP or ICMP).

preliminary comparisons of self-similarity exponents computed independently on $Y_{Pkt,\Delta}(t)$ and $Y_{Byt,\Delta}(t)$.

Goals, contributions and outline. The present contribution makes use of a recently proposed multivariate self-similar model, Operator Fractional Brownian motion (OfBm) [19, 20], to model joint self-similarity of bytes and packets as well as to construct an anomaly detection procedure that exploits OfBm parameters. The definition of OfBm is detailed in Section 2, and complemented with the study of its wavelet analysis, that incorporates an original tunable *fractional integration* parameter that permits coping with the specific nature of Internet data. Another specificity of the present work consists in formulating the estimation of the full set of OfBm parameters as a non linear regression (cf. Section 3). In addition, an original Branch & Bound procedure is devised to minimize the corresponding functional. The estimation performance is assessed by means of Monte Carlo simulations conducted on synthetic OfBm, that mimic Internet data properties. An anomaly detection procedure is then constructed on OfBm parameters estimated from bytes and packets time series, and applied to a large set of high quality and recent Internet data, from the MAWI repository [21], described in Section 4. Results are discussed in Section 5.

2. BIVARIATE OFBM AND WAVELET ANALYSIS

Definitions. The general definitions of OfBm can be found in [22, 20] as the only multivariate Gaussian self-similar process with stationary increments. The definitions here are restricted to the class of bivariate time-reversible OfBm $\{Y(t) = (Y_1(t), Y_2(t))\}_{t\in\mathbb{R}}$. Let $\{X(t) = (X_1(t), X_2(t))\}_{t\in\mathbb{R}}$ denote 2 fBms, with auto and cross-covariance functions written as: $\mathbb{E}X_p(t)X_{p'}(s) =$

$$\Sigma_X(p,p')/2(|t|^{H_p+H_{p'}} + |s|^{H_p+H_{p'}} - |t-s|^{H_p+H_{p'}}), \quad (1)$$

with $(p,p') \in \{1,2\}^2$, with $0 < H_1 \leq H_2 < 1$, and where $\Sigma_X \equiv \mathbb{E}X(1)X^*(1)$.

Process X is well defined if and only if [23, 19]:

$$g(H_1,H_2,\rho_{\mathrm{x}}) \equiv \Gamma(2H_1+1)\Gamma(2H_2+1)\sin(\pi H_1)\sin(\pi H_2) - \rho_{\mathrm{x}}^2\Gamma(H_1+H_2+1)^2\sin^2(\pi(H_1+H_2)/2) > 0. \quad (2)$$

Further, let W denote a 2×2 invertible matrix, then $\{Y(t) = (Y_1(t), Y_2(t))\}_{t\in\mathbb{R}}$ is defined as $\{Y(t)\}_{t\in\mathbb{R}} = \{WX(t)\}_{t\in\mathbb{R}}$. A parsimonious parametrization of $Y(t)$, in 7 parameters accounting for under-determinations, $\Theta = (H_1, H_2, \rho_{\mathrm{x}}, \sigma_{\mathrm{x}_1}, \sigma_{\mathrm{x}_2}, \beta, \gamma)$, has been proposed [24, 25, 26]:

$$W = \begin{pmatrix} \frac{1}{\sqrt{1+\gamma^2}} & \frac{\beta}{\sqrt{1+\beta^2}} \\ \frac{-\gamma}{\sqrt{1+\gamma^2}} & \frac{1}{\sqrt{1+\beta^2}} \end{pmatrix}, \Sigma_X = \begin{pmatrix} \sigma_{\mathrm{x}_1}^2 & \sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\rho_{\mathrm{x}} \\ \sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\rho_{\mathrm{x}} & \sigma_{\mathrm{x}_2}^2 \end{pmatrix}. \quad (3)$$

Wavelet analysis. Let Y^δ denote the increment process of Y:$Y_p^\delta(t) = Y_p(t+1) - Y_p(t)$, $p = 1, 2$. The multivariate discrete wavelet transform (DWT) of Y^δ, $(D_{\mathrm{y}_1}(j,k), D_{\mathrm{y}_2}(j,k))$, is defined as:

$$D_{\mathrm{y_p}}(j,k) = \int_{\mathbb{R}} \psi_{j,k}(t) Y_p^\delta(t)\mathrm{d}t, \quad (4)$$

$$\text{where } \{\psi_{j,k}(t) = 2^{-j(0.5-\mu)}\psi_0(2^{-j/2}t - k)\}_{(j,k)\in\mathbb{Z}^2} \quad (5)$$

denotes the collection of dilated and translated templates of ψ_0, an oscillating reference pattern with joint time and frequency localization. It is referred to as the mother wavelet and further characterized by its number of vanishing moments N_ψ, a positive integer, defined as $\forall n = 0, \ldots, N_\psi - 1$, $\int_{\mathbb{R}} t^k\psi_0(t)dt \equiv 0$ and $\int_{\mathbb{R}} t^{N_\psi}\psi_0(t)dt \neq 0$. For a detailed introduction to wavelet transforms, interested readers are referred to e.g., [27]. We have introduced an additional parameter μ (compared to classical definition) which acts as a fractional integration parameter [28] and whose practical crucial role is detailed in Section 4.

Combining Eq. (1) with $Y(t) = WX(t)$, and using $\eta_{j,h} = \frac{1}{2}\int_{\mathbb{R}^2}\left(|u+2^{-j}|^{2h} + |u-2^{-j}|^{2h} - 2|u|^{2h}\right)\psi_0(v)\psi_0(v-u)^*\mathrm{d}u\mathrm{d}v$, it can be shown that [25, 26]:

$$\mathbb{E}D_{\mathrm{y}}(j,k)D_{\mathrm{y}}(j,k)^* = \begin{pmatrix} (E_{11}(\Theta))_j & (E_{12}(\Theta))_j \\ (E_{12}(\Theta))_j & (E_{22}(\Theta))_j \end{pmatrix} \quad (6)$$

$$\text{with } (E_{11}(\Theta))_j = (1+\gamma^2)^{-1}\sigma_{\mathrm{x}_1}^2\eta_{j,H_1}2^{j(2H_1+1+2\mu)} + 2\beta(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2}\rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\eta_{j,\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1+2\mu)} + \beta^2(1+\beta^2)^{-1}\sigma_{\mathrm{x}_2}^2\eta_{j,H_2}2^{j(2H_2+1+2\mu)}, \quad (7)$$

$$(E_{12}(\Theta))_j = -\gamma(1+\gamma^2)^{-1}\sigma_{\mathrm{x}_1}^2\eta_{j,H_1}2^{j(2H_1+1+2\mu)} + (1-\beta\gamma)(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2}\rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\eta_{j,\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1+2\mu)} + \beta(1+\beta^2)^{-1}\sigma_{\mathrm{x}_2}^2\eta_{j,H_2}2^{j(2H_2+1+2\mu)}, \quad (8)$$

$$(E_{22}(\Theta))_j = \gamma^2(1+\gamma^2)^{-1}\sigma_{\mathrm{x}_1}^2\eta_{j,H_1}2^{j(2H_1+1+2\mu)} - 2\gamma(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2}\rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\eta_{j,\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1+2\mu)} + (1+\beta^2)^{-1}\sigma_{\mathrm{x}_2}^2\eta_{j,H_2}2^{j(2H_2+1+2\mu)}. \quad (9)$$

Estimation. In practice, the ensemble average $\mathbb{E}D_{\mathrm{y}}(j,k)D_{\mathrm{y}}(j,k)^*$ is replaced by the sample mean estimator (with N the sample size) $S(2^j) = \frac{2^j}{N}\sum_{k=1}^{N/2^j} D(2^j,k)D(2^j,k)^*$.

Univariate analysis. Univariate analysis of self-similarity consists in performing a linear regression of $\log_2 S_{p,p}(2^j)$ against $\log_2 2^j = j$ to yield univariate parameters H_p, $p = 1, 2$, [12, 4, 7]. This amounts to neglect the potential mixture of power-laws inherent to multivariate self-similarity and to assume a priori the absence of mixing, i.e., $\beta = \gamma = 0$. When mixing is present, this leads to a substantial bias in the estimation of the H_ps [25].

3. BIVARIATE-OFBM FULL IDENTIFICATION

Non linear regression. Elaborating on [25, 26], the originality of the present contribution is to formulate the full identification of Biv-OfBm (i.e., the estimation of Θ) as a non linear regression:

$$\hat{\Theta} = \arg\min_{\Theta\in\mathcal{Q}} \sum_{p,p'=1}^{2}\sum_{j=j_1}^{j_2}\left(\log_2|(S_{p,p'})_j| - \log_2|(E_{p,p'}(\Theta))_j|\right)^2, \quad (10)$$

where the use of the logarithm ensures that all scales $2^j, j \in \{j_1, \ldots, j_2\}$ contribute equally. Minimizing Eq. (10) is intricate because of the non convexities of both the implied functional (as a mixture of power laws) and the search space (due to Constraint 2):

$$\mathcal{Q} = \Big\{\Theta = (H_1, H_2, \rho_{\mathrm{x}}, \sigma_{\mathrm{x}_1}, \sigma_{\mathrm{x}_2}, \beta, \gamma) \in \mathbb{R}^7 \,|\, \Theta \in [0,1]^3 \times [0, \sigma_{\max}]^2 \times [-1,1]^2, g(H_1,H_2,\rho_{\mathrm{x}}) > 0, H_1 \leq H_2\Big\}. \quad (11)$$

Branch & Bound optimization. To find the global minimizer $\widehat{\Theta}$ in Eq. 10, the second originality of this work is to resort to a Branch & Bound (B&B) procedure [29], as devised and studied in [26]. The B&B procedure avoids greedy searches by smart enumerations that stem from the repetition of the following steps until a stopping criterion is reached: i) Partitioning: Choose any region $\mathcal{R}$ from the search

space and divide it into two smaller regions $\mathcal{R}_a$ and $\mathcal{R}_b$; ii) Bounding: Compute lower and upper bounds of the objective function in Eq. 10, on $\mathcal{R}_a$ and $\mathcal{R}_b$. Lower bounds are obtained by *interval arithmetic* (cf. [30]), a technique that combines elementary operations to produce rough lower bounds ; iii) Pruning: Discard regions that do not satisfy Eq. 2 and regions whose lower bound is larger than the smallest upper bound. The corresponding algorithm solving (10) is fully detailed in [26] and MATLAB routines will be made publicly available at the time of publication.

Estimation performance. The estimation performance of the proposed B&B procedure is assessed by Monte Carlo simulation. The procedure is applied to independent copies of synthetic OfBm, whose sample size ($N \simeq 3600$) and parameter settings ($1 \gtrsim \hat{H}_2 \geq \hat{H}_1 \gtrsim 0.8$, cf. Fig. 3 and $\hat{\rho} = 0.8$) match those observed or expected for the Internet traffic analyzed here $|H_2 - H_1| = 0$ or $|H_2 - H1| \simeq 0.2$, with or without mixing W. Synthesis and analysis procedures were devised by ourselves. The quality of the estimation performance is quantified in Fig. 1, which shows that whereas correlation and mixing parameters remain difficult to estimate for short time series, and settings that mimic Internet traffic, while H_1 and H_2 are always well estimated.

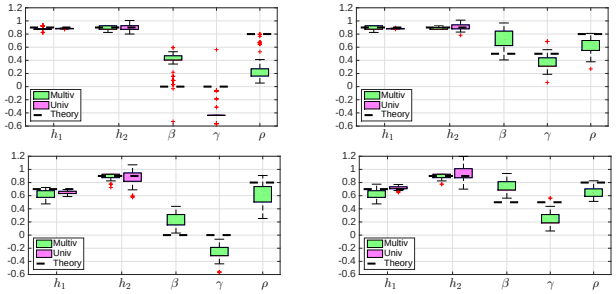

Fig. 1. **Estimation performance on synthetic OfBm,** for four configurations potentially matching Internet Traffic: $(H_1, H_2) = (0.9, 0.9)$ or $(0.7, 0.9)$, $(\beta, \gamma) = (0, 0)$ or $(0.5, 0.5)$.

4. MAWI DATABASE AND RANDOM PROJECTIONS

MAWI database. The MAWI repository [21, 31] is an on-going collection of Internet traffic traces, captured within the *WIDE* backbone network (AS2500) that connects Japanese universities and research institutes to the Internet. Packet 5-tuple and timestamps, collected daily from 14:00 to 14:15 (Japanese Standard Time), are anonymized and made publicly available. Each trace contains roughly 100 to 150 million IP packets.

Random projections. As discussed in Section 1, a major issue in anomaly detection consists in defining and computing a reference normal traffic for comparisons. Yet, the ever varying nature of Internet traffic precludes the use of traffic collected another day or from another network [4, 7]. Instead, the use of random projections (or sketches) [6] has been shown to be a relevant procedure to construct self-reference of normal traffic. In random projections, each IP packet is attributed to one of the M outputs of a hash table [32], acting on one selected element of the 5-tuple (here the IP Source address). Therefore, all packets of any given flow with the same IP Source address are allocated together to the same randomly chosen entry in the hashtable. Traffic is hence split into M sub-traffic. When traffic contains no anomaly, all sketches are expected to be statistically equivalent. When traffic contains some anomalies, associated to one same IP Source address, all corresponding packets are associated to the same sketch, while all other sketches are anomaly free. Comparisons across sketches can then be used to construct reference statistics for normal traffic and hence to detect anomalies.

Fractional integration. While in OfBm model, theoretically $0 < H_1 \leq H_2 < 1$, estimated H for normal traffic takes large values, close to 1, as well documented in [4, 7]. This practically raises severe issues with respect to the proposed B&B procedure described above, as some data may randomly and accidentally lead to H that exceeds the allowed range. This is illustrated in the reports of univariate based estimates of H (bottom row in Fig. 3), since univariate wavelet based estimation does not force $H \in (0, 1)$ (cf. e.g., [12, 4, 7]). To circumvent this issue, the present contribution introduces an extra fractional integration parameter μ in the definition of the wavelet coefficients, cf. Section 2. Instead of using a single and standard $\mu \equiv 1$ that matches synthetic OfBm with Internet data, wavelet coefficients are computed using $\mu_W = 1/2$ (in Eq. 5) while the B&B procedure is applied with $\mu_B = 1$ (in Eqs. 7 to 9): This amounts to fractionally integrate data to force H_1 and H_2 to live well in the middle of the $(0, 1)$ range. Estimation of H_1 and H_2 needs then only to be shifted a posteriori by $\mu_B - \mu_W$. Because, we will mostly use $H_2 - H_1$, this does not impact results and conclusions. We see the practical possibility of decoupling and tuning parameters $0 < \mu_W = 1/2 \leq 1$ and $\mu_B = 1$ as an original and practically efficient trick in the proposed procedure that permits to adjust to the specificities of real world data.

5. OFBM MODELING AND ANOMALY DETECTION

Setting. Results are reported here for four traces collected on four different days, in 2008, 2013, 2014 and 2015, as typical and very recent examples of Internet traffic for the case study intended here. Hash tables with $M = 16$ outputs are used. Sketches are aggregated at $\Delta_0 = 0.25$s, as it is now well documented that in Internet traffic, self-similarity develops across scales ranging from seconds to hours [12, 4, 7]. Examples of aggregated sketches are illustrated in Fig. 2 both for packets (first column) and bytes (last column). Wavelet analysis is conducted using least asymmetric orthogonal Daubechies wavelets with $N_\psi = 2$ vanishing moments [27]. The minimization of Eq. (10) is conducted using scales $j_1 = 3$ to $j_2 = 8$, corresponding respectively to time scales ranging from 2s to 1min (as available data are limited to 15min durations). Although the full identification of Biv-OfBm requires 7 parameters, we focus here on the estimation of the 5 most interesting parameters, namely $\Theta = (H_1, H_2, \beta, \gamma, \rho_\times)$ and set $\sigma_{\times_1} = \sigma_{\times_2} = 1$ by a priori data normalization. Estimation of Θ is done based on the proposed B&B procedure for each sketch of each dataset.

Metadata and ground truth. Metadata regarding anomalies in Internet MAWI traffic were provided to us by experts via the outputs of a computerized procedure, MawiLab [33], inspecting the content of Internet traffic in an automated and systematic manner: MawiLab relies on the combined use of several benchmark anomaly estimators. It is worth emphasizing that these metadata thus do not constitute the ground truth but only indications against which the outputs of the proposed detection procedure can be compared.

Internet traffic statistical modeling with Biv-OfBm. Fig. 2 illustrates a posteriori that the joint statistics of packet and byte counts, empirically estimated using $S_{p,p'}(j)$, match well Biv-OfBm model $E_{p,p'}(\hat{\Theta})$, with parameters $\hat{\Theta}$ estimated using the proposed B&B minimization procedure. It also shows that OfBm models equally relevantly sketches with and without anomalies, yet with obviously different estimated parameters $\hat{\Theta}$. Fig. 2 shows that the use of the chosen 5 parameter parametrization is satisfactory.

Anomaly detection. Parameters Θ estimated for each sketch and each trace were compared with the available MawiLab anomaly metadata. First inspections, not reported here for space reasons, show that the estimated correlation ρ and mixing β and γ parame-

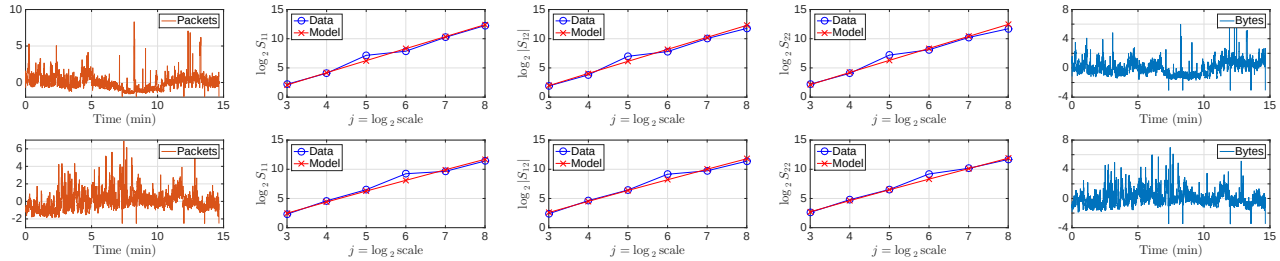

Fig. 2. Bivariate-OfBm model for packet and byte aggregated time series. First and last column: Normalized Pkt and Byte count time series ; 2nd to 4th column: comparisons of $\log_2 S_{p,p'}$ (blue) vs. $\log_2 E_{p,p'}(\hat{\Theta})$ (red) for $(p, p') = (1, 1)$, $(1, 2)$ and $(2, 2)$. Sketch without (top) and with (bottom) anomaly.

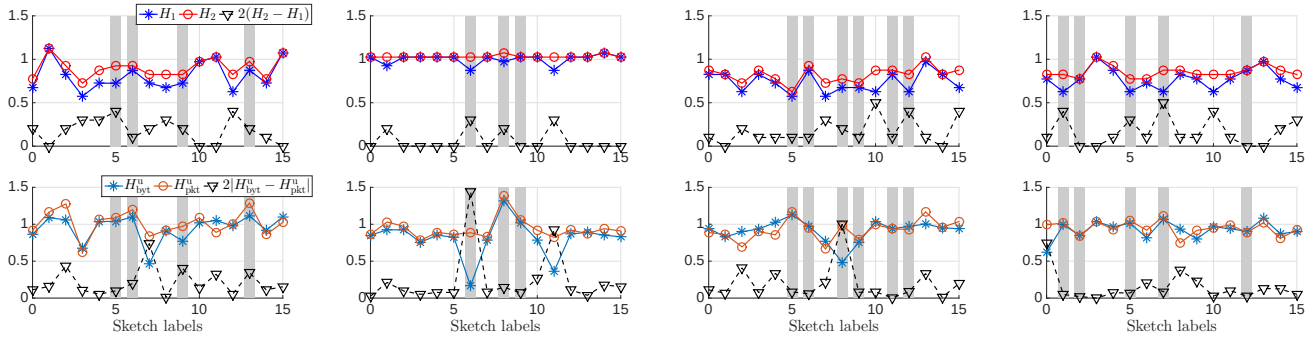

Fig. 3. $H_2 - H_1$ **as an anomaly detector.** Top row: bivariate analysis, H_2, H_1 and $|H_2 - H_1|$ (multiplied by 2 for readability) as functions of the sketch labels. Bottom row: univariate analysis, H^U_{Byt}, H^U_{Pkt} and $||H^U_{Byt} - H^U_{Pkt}||$ (multiplied by 2 for readability) as functions of the sketch labels. The vertical grey lines indicate sketches tagged by experts as containing the largest number of anomalous packets. From left to right: Data collected in 2008, 2013, 2014 and 2015.

ters do not seem to correlate well with the occurrence of anomalies. This may partly be explained by the poor estimation performance for these parameters, as observed in Monte Carlo simulations, in a comparable setting. This analysis however clearly indicates that mixing parameters β and γ significantly depart form 0, thus showing the need for bivariate estimation for a non-biased estimation of parameters H_1 and H_2. Comparison of top and bottom rows in Fig. 3 shows significant discrepancies between the univariate estimates $(H^U_{\mathrm{byt}}, H^U_{\mathrm{pkt}})$ and the bivariate ones (H_1, H_2), which clearly illustrate the limitations and biases of univariate analysis.

The inspection of H_1 and H_2 estimated from the bivariate B&B procedure tends to show that large deviations of $|H_2 - H_1|$ from 0, match a significant number of sketches marked by MawiLab as containing a large number of *anomalous packets*, cf. Fig. 3, top row. A close inspection of Fig. 3 shows that the agreement between departures from 0 and occurrences of anomaly is not perfect, which may have two origins: As mentioned above, metadata are not the ground truth, and some anomalies might have been missed by MawiLab procedure. Alternatively, some anomalies relevantly detected by MawiLab may not be signed by a departures of $|H_2 - H_1|$ from 0 which may in turn indicate specific subclasses of anomalies. These discrepancies, requiring deeper expert inspection, will be further investigated. However, expert inspections indicate that i) the *Trinocular* anomaly, specific to a computer network experiment being run on the MAWI network [34], is systematically detected ; ii) 96% of Deny-of-Service attacks were detected for 2013 ; and iii) 100% of Heavy Hitter anomalies were detected in 2014.

These case study results are altogether very encouraging and consistent with the queuing mechanism connecting self-similarity to heavy tail distribution of Internet objects proposed in [15, 16, 17] that leads to predict identical Hurst exponents for packets and bytes. They are also consistent with prior empirical results, relying on univariate analysis of bytes and packets that tend to comfort $H^{\mathrm{byt}} \simeq H^{\mathrm{pkt}}$ for normal traffic, while departures of H^{byt} from H^{pkt} may indicate anomalies [4, 7].

6. CONCLUSIONS AND PERSPECTIVES

The present contribution promotes the use of multivariate models for self-similarity, such as OfBm and proposes, to the best of our knowledge, the first procedure for the full identification of OfBm. It consists of a non-linear regression on log wavelet coefficients, solved by an original Branch and Bound procedure. Numerical simulations conducted on independent copies of synthetic OfBm show that the propose procedure achieves, in parameter settings that match those observed in Internet traffic, satisfactory performance for the estimation of the Hurst exponents H_1 and H_2, while the estimation of the correlation and mixing parameters turns more difficult in such settings. MATLAB procedures implementing analysis and synthesis will be made publicly available at the time of publication.

It is then shown that bivariate OfBm constitutes a relevant model to describe jointly the scale invariance properties observed in Internet traffic for both packet and byte time series. It also shows that Biv-OfBm is relevant both for regular or normal traffic and for traffic with anomalies. These per se original results permit to conduct an anomaly detection case study, which in turn provides significant evidence that certain types of anomalies are marked by a significant discrepancy between $H^{\mathrm{byt}} \simeq H^{\mathrm{pkt}}$. Results are satisfactory enough to call for a large scale systematic study, with enriched metadata, that may permit not only the detection of anomalies but also the classification of the types of anomalies that can be detected by a change in their scale invariance properties. This is under current investigation.

7. REFERENCES

[1] P. Barford, J. Kline, D. Plonka, and A. Ron, "A Signal Analysis of Network Traffic Anomalies," *ACM SIGCOMM IMW '02*, pp. 71–82, 2002.

[2] A. Lakhina, M. Crovella, and Ch. Diot, "Mining Anomalies Using Traffic Feature Distributions," *ACM SIGCOMM '05*, pp. 217–228, 2005.

[3] J. Mazel, R. Fontugne, and K. Fukuda, "Taxonomy of Anomalies in Backbone Network Traffic," in *Proceedings of the fifth International Workshop on TRaffic Analysis and Characterization (TRAC '14)*, 2014, pp. 30–36.

[4] G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, and K. Cho, "Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedure," in *Proc. ACM SIGCOMM Workshop on Large Scale Attack Defense*, 2007, pp. 145–152.

[5] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang, "An Empirical Evaluation of Entropy-based Traffic Anomaly Detection," *ACM IMC '08*, pp. 151–156, 2008.

[6] B. Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen, "Sketch-based change detection: Methods, evaluation, and applications," in *Proc. ACM SIGCOMM Conf. on Internet Measurement (IMC)*, 2003, pp. 234–247.

[7] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho, "Seven Years and One Day: Sketching the Evolution of Internet Traffic," *Proc. IEEE INFOCOM'09*, pp. 711–719, 2009.

[8] G. Samorodnitsky and M. Taqqu, *Stable non-Gaussian random processes*, Chapman and Hall, New York, 1994.

[9] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson, "On the self-similar nature of ethernet traffic," *IEEE Trans. on Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[10] Vern Paxson and Sally Floyd, "Wide area traffic: The failure of poisson modeling," *IEEE Trans. on Networking*, vol. 4, no. 3, pp. 209–223, 1995.

[11] W. Willinger, M. Taqqu, and A. Erramilli, *Stochastic networks: Theory and applications*, chapter A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for Modern High-Speed Networks, pp. 339–366, Clarendon Press (Oxford University Press), 1996.

[12] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch, "Multiscale nature of network traffic," *IEEE Signal Proc. Mag.*, vol. 19, no. 3, pp. 28–46, 2002.

[13] R. Fontugne, P. Abry, K. Fukuda, P. Borgnat, J. Mazel, H. Wendt, and D. Veitch, "Random projection and multiscale wavelet leader based anomaly detection and address identification in internet traffic," in *IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, Brisbane, Australia, April 2015.

[14] H. Gupta, V.J. Ribeiro, and A. Mahanti, "A longitudinal study of small-time scaling behavior of internet traffic," in *NETWORKING 2010*, Mark Crovella, LauraMarie Feeney, Dan Rubenstein, and S.V. Raghavan, Eds., Lecture Notes in Computer Science, pp. 83–95. Springer, 2010.

[15] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," *IEEE Trans. on Networking*, vol. 5, no. 6, pp. 835–846, 1997.

[16] M. S. Taqqu, W. Willinger, and R. Sherman, "Proof of a Fundamental Result in Self-Similar Traffic Modelling," *Comp. Commun. Rev.*, vol. 27, pp. 5–23, 1997.

[17] O.J. Boxma and J.W. Cohen, *Self-Similar Network Traffic and Performance Evaluation*, chapter The Single Server Queue: Heavy Tails and Heavy Traffic, Wiley-Interscience, 2000.

[18] F. Silveira, Ch. Diot, N. Taft, and R. Govindan, "ASTUTE: Detecting a Different Class of Traffic Anomalies," *ACM SIGCOMM '10*, pp. 267–278, 2010.

[19] G. Didier and V. Pipiras, "Integral representations and properties of operator fractional brownian motions," *Bernoulli*, vol. 17, no. 1, pp. 1–33, 2011.

[20] P-O. Amblard and J-F. Coeurjolly, "Identification of the Multivariate Fractional Brownian Motion," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5152–5168, Nov. 2011.

[21] "MAWI Traffic Archive," .

[22] G. Didier and V. Pipiras, "Exponents, symmetry groups and classification of operator fractional Brownian motions," *To appear in Journal of Theoretical Probability*, 2011.

[23] P-O. Amblard, J-F. Coeurjolly, F. Lavancier, and A. Philippe, "Basic properties of the multivariate fractional Brownian motion," *Bulletin de la Société Mathématique de France, Séminaires et Congrès*, vol. 28, pp. 65–87, 2012.

[24] G. Didier, H. Helgason, and P. Abry, "Demixing multivariate-operator self-similar processes," .

[25] P. Abry and G. Didier, "Wavelet estimation for operator fractional brownian motions," *Bernoulli*, pp. to appear, arXiv preprint arXiv:1501.06094, 2015.

[26] J. Frecon, N. Didier, G.and Pustelnik, and P. Abry, "Bivariate operational fractional brownian motion: A variational problem based full identification," p. preprint, 2015.

[27] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, Academic Press, 3rd edition, 2008.

[28] H. Wendt, P. Abry, and S. Jaffard, "Bootstrap for empirical multifractal analysis," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 38–48, 2007.

[29] E. Hansen, "Global optimization using interval analysis - the multi-dimensional case," *Numerische Mathematik*, vol. 34, no. 3, pp. 247–270, 1980.

[30] R. E. Moore, *Interval analysis*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1966.

[31] K. Cho, K. Mitsuya, and A. Kato, "Traffic data repository at the WIDE project," in *USENIX 2000 Annual Technical Conference: FREENIX Track*, June 2000, pp. 263–270.

[32] M. Thorup and Y. Zhang, "Tabulation based 4-universal hashing with applications to second moment estimation," in *ACM SIAM Symp. Discrete Algorithms (SODA)*, 2004, pp. 615–624.

[33] Romain Fontugne, Pierre Borgnat, Patrice Abry, and Kensuke Fukuda, "MAWILab: Combining Diverse Anomaly Detectors for Automated Anomaly Labeling and Performance Benchmarking," *Proc. ACM Co-NEXT*, 2010.

[34] Lin Quan, John Heidemann, and Yuri Pradkin, "Trinocular: Understanding Internet Reliability Through Adaptive Probing," *Proc. ACM SIGCOMM'13*, pp. 255–266, 2013.