Aggregated Impulses: Towards Explanatory Models for Self-Similar Alpha Stable Network Traffic

Jorge Gonzalez
Dept. of Mathematical Sciences
Florida Atlantic University
Boca Raton, FL, USA
jorgegonzale2013@fau.edu

Joshua Clymer
SEAP Intern
Naval Postgraduate School
Monterey, CA, USA

Chad A. Bollmann
Dept. of Electrical & Computer Engineering
Naval Postgraduate School
Monterey, CA, USA
cabollma@nps.edu

***Abstract*—Heavy-tailed models of computer network traffic have been shown to more accurately reflect actual traffic distributions of many traffic features than methods based on exponential distributions. The power-law tail inherent to the alpha-stable distribution better accommodates network traffic properties such as impulsiveness, self-similarity, and long-range dependence, enabling more precise models and more accurate network anomaly detection. Beginning from individual traffic processes, this work presents two explanatory mathematical methods for network aggregation which lead to either Gaussian or alpha-stable traffic distributions. The first method, based on the generalized central limit theorem, shows how self-similarity originates from an impulsive-noise-based representation of individual processes. A second method based on renewal theory supports the predictions of the first method while also permitting estimation of rates of convergence. We develop working models of these methods to empirically validate our aggregation approach and provide an explanation for the heavy tails and varieties of scaling observed in network traffic.**

***Index Terms*—alpha-stable, computer network traffic model, heavy-tail, long-range dependent, self-similar**

I. Introduction

While network traffic is known to be bursty, self similar (SS), and long-range dependent (LRD), it remains an open problem to develop simple models that both explain and reproduce these features [1]. Another open question involves the distribution of aggregated network traffic: Heavy tails of many features (e.g., packet rate, flow size, inter-arrival times) have been well documented, but disagreement exists regarding an overall best-suited distribution to characterize features for either modeling or anomaly detection [2]–[4].

The explanatory models described by Willinger et al. require key milestones of discovery, construction, and validation [1]. The discovery examined in this work is the tendency, in larger networks, of certain network traffic features including packet rate to trend towards alpha-stable distributions [3], [4]. As power-law distributions in network traffic are well-known, the more interesting part of this discovery is that the authors have frequently found non-parametric or Gaussian distributions of the same features in smaller networks. The focus of our efforts thus becomes identifying mechanisms that can deliver Gaussian as well as alpha-stable distributions and that also reflect characteristics of actual network traffic. Our results can also provide justification for the coexistence of fine and coarse scaling recently observed in a longitudinal study of network traffic [5].

The milestone of construction begins with the observation that, for a given device, traffic processes occur at a few typical rate levels and can be characterized as impulses. This is illustrated in Figure 1, a traffic rate plot of traffic to a single, centrally-managed device on a medium-sized campus network.

Over the nearly 15 minutes of minimal user activity, most traffic events are automated, periodic backup and sync processes, mostly at low rates. When we add a human doing typical actions such as streaming videos or music, the number and magnitude of these impulses change while their overall impulsive nature is preserved, as shown in Figure 2.

The aggregated traffic of these impulses in a network has long been modeled using Lévy processes $\{X_t : t \geq 0\}$ such as Poisson and fractional Brownian motion [6]. Leveraging previous work [3], [4], we propose the alpha-stable distribution as an alternative Lévy-based model for this aggregated traffic due to its strong theoretical connections to SS and LRD. As previously discussed, many features of network traffic are known to be heavy-tailed [7]; random variables (RVs) with heavy-tailed distributions belong to the domain of attraction

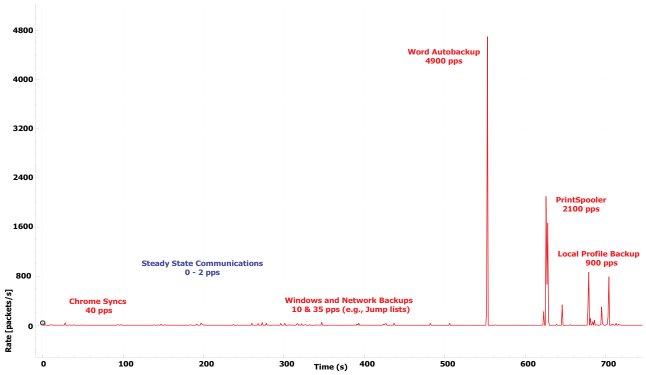

Figure 1. Rate plot of 12 minutes of traffic received at a single host with minimal automated and no intentional user activity.

This work was funded by the NPS Naval Research Program under proposal NRP-18-039A. Joshua Clymer is with the ASEE Science and Engineering Apprenticeship Program.

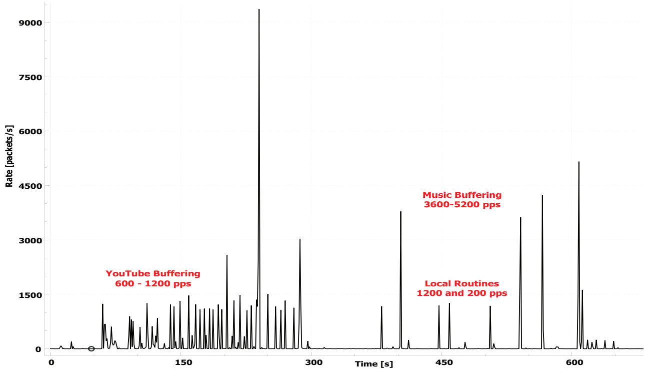

Figure 2. Rate plot of nearly 12 minutes of traffic received at a single local host with some user interaction and typical streaming activity (e.g., YouTube and Amazon music). Note that even with user interaction and more activity, traffic remains impulsive and loosely consistent in magnitudes.

of stable processes per the Generalized Central Limit Theorem (GCLT) of Gnedenko and Kolmogorov and they are the *only* distributions with a non-empty domain of attraction. Self-similarity is similarly well documented [7], [8]; Lévy processes are SS if and only if alpha-stable [9].

While the macro properties of network traffic and empirical observation support an alpha-stable approach to modeling aggregated traffic, to our knowledge the *mechanism* of aggregation that can result in alpha-stable and Gaussian traffic distributions has not been defined in the literature. The primary contributions of this work are to explain how individual device traffic aggregates to larger flows that can have exponential- *or* heavy-tailed characteristics, and how the origin of SS impacts the final aggregated traffic distributions. This theory requires no assumptions regarding inter-arrival ties or packet rates and is thus more general than other heavy-tailed modeling approaches.

To approach these problems, we propose two general, end-to-end, and explanatory models for network traffic based on individual sources. By *end-to-end*, we mean that these models specify theoretical limiting distributions as Gaussian or alpha-stable for aggregated traffic based on the characteristics of the inputs.

We begin by exploring two complementary theories that show how heavy-tailed inputs can aggregate to either Gaussian or alpha-stable outputs. The first theory, based on impulse aggregation under the GCLT, predicts aggregation to alpha-stable processes when the inputs are heavy-tailed, IID, and have power-indices (i.e., tail decay values) less than 2. The second theory, grounded in renewal process theory of Taqqu and Levy, predicts alpha-stable aggregation based on the population ratios of heavy-tailed IID inputs.

We then develop two models, impulse and renewal, based on these theories and provide preliminary evaluations of their accuracy in replicating 4 unique datasets from networks of varying size and device populations. Finally, we perform preliminary evaluations of our proposed models to validate this theoretical approach and evaluate their ability to reproduce LRD.

The remainder of this work is organized as follows: Section 2 describes datasets, background theory, and prior work. Section 3 validates foundational IID and ergodicity assumptions, then describes the theory of the impulse model. Section 4 contains the renewal model, Section 5 describes our simulation process and assessment results, and conclusions and future work items are contained in Section 6.

II. Background and Prior Work

A. Datasets used in this work

To provide a rigorous comparison using real-world network traffic, this work used network traffic data (i.e., traces) from three different sources; these sources roughly fall into a three different categories of typical traffic. The traces and some specific attributes are summarized in Table I, where the average traffic rate of each trace is given in gigabits per second.

Table I
Network Traffic Sources

Name	Type	Rate [Gbps]	ID	Source
WAND	Residential DSL	0.065	20090106-04	[10]
NPS	Academic Campus	1.1	2019Jul01	-
MAWI Nov	Backbone	0.43	2017Nov11	[11]
MAWI Apr	Backbone	0.46	2016Apr28	[11]

The WAND data is a capture of residential traffic from a New Zealand internet service provider [10]; this dataset was selected for its low rate and anticipated homogeneity of traffic. The Naval Postgraduate School (NPS) trace is a capture of inbound traffic to a relatively small campus network; on a typical day the number of active devices is in the low thousands, but device and process diversity is expected to be greater than that of the WAND trace due to the mix of student and professional services. The most diverse traces are expected to be the MAWI traces, as these are captures of bi-directional Internet traffic between Japan and the United States [11].

The WAND and MAWI traces are publicly-available; as of the time of publication, we are working to make the NPS traces used in this work available as well via the authors' NPS website. Due to space constraints and further descriptions of the WAND and MAWI datasets available in existing literature, we will now provide background on alpha-stable (i.e., stable, Lévy stable, or Pareto-stable) processes and their relationship to the properties of SS and LRD.

B. Alpha-Stable Processes

By considering the definitions of alpha-stable processes and their sources, we can develop intuition regarding the types of inputs that would aggregate to an alpha-stable result.

Definition II.1. A random variable X is said to have a stable distribution if, for any positive number A and B, there is a positive number C and a real number D such that

$$AX_1 + BX_2 \overset{\mathrm{d}}{=} DX + D \tag{1}$$

where X_1 and X_2 are independent copies of X and $\overset{\text{d}}{=}$ indicates equality in the distribution sense.

See [12] for equivalent definitions. A fundamental result is that A, B, C satisfy $A^\alpha + B^\alpha = C^\alpha$ for some $\alpha \in (0, 2]$. For a proof see [13]. This alpha is of singular importance in the theory of stable processes and very relevant to our application as we will slowly uncover in the progression of the paper.

A definition in terms of the domain of attraction of a stable process is possible thanks to the GCLT [14]. This definition is the first indication that alpha stable processes provide a suitable framework to model network traffic where many signals aggregate, and it thus becomes fundamental to our approach to anomaly detection.

Definition II.2. For a a random variable X. We define the *domain of attraction* of X, denoted by $\mathscr{D}(X)$ to be the set of random variables Y such that there exists $d_n > 0$, $a_n \in \mathbb{R}$ and

$$\frac{Y_1 + Y_2 + ... + Y_n}{d_n} + a_n \overset{\text{d}}{\to} X \tag{2}$$

for $Y, Y_1, ..., Y_n$ IID random variables. The symbol $\overset{\text{d}}{\to}$ expresses convergence in distribution.

Definition II.3. A random variable X is a stable if $\mathscr{D}(X) \neq \varnothing$

The set $\mathscr{D}(X)$ is characterized in [14]. Explicit expressions for the distribution of stable processes are unknown except for the following three classical examples: the Gaussian distribution where $\alpha = 2$, the Lévy distribution for $\alpha = 1/2$, and the Cauchy distribution where $\alpha = 1$. However, it is possible to define stable processes in terms of characteristic functions using the four parameters of α, β, σ, and μ.

$$E(e^{i\theta X}) = \exp\{-\sigma^\alpha |\theta|^\alpha (1 - i\beta \text{sign}(\theta)\omega(\theta, \alpha)) + i\mu\theta\} \tag{3}$$

where

$$\omega(\theta, \alpha) = \begin{cases} \tan(\frac{\pi\alpha}{2}) & \alpha \neq 1 \\ \frac{2}{\pi}\ln(|\theta|) & \alpha = 1 \end{cases}$$

and

$$\text{sign}(\theta) = \begin{cases} 1 & \theta > 0 \\ 0 & \theta = 0 \\ -1 & \theta < 0 \end{cases}$$

These formulas were classically obtained through the study of infinitely divisible processes and their *Lévy-Khintchine representation* [9], but other approaches have been discovered [15].

A quick exploration of this expression reveals the effects of these parameters [12]: $\alpha \in (0, 2]$ characterizes tail size; $\sigma \in [0, \infty)$ determines spread; $\beta \in [-1, 1]$ describes skewness; and $\mu \in \mathbb{R}$ gives location. We write $X \sim S_\alpha(\sigma, \beta, \mu)$ if X has characteristic function given by (3).

Having defined alpha-stable processes, we now link this distribution to the closely-related property of SS. The manifested burstiness in network traffic at different scales can be described with the introduction of the concept of SS. This is the second indication that stable processes offer the correct framework to modeling network traffic.

Definition II.4. A process $X = \{X(t) : t \in \mathbb{R}\}$ is SS if for any $a > 0$, there is $b > 0$ such that the finite-dimensional distributions of X are the same as $\{bX(at) : t \in \mathbb{R}\}$

Surprisingly, there is $H > 0$ such that for each a, $b = a^{-H}$. H is called the SS index or Hurst exponent [9]. There are many methods to approximate H [16]. The original rescaled range (R/S) method discovered by Hurst in the context of hydrology sparked the introduction and study of SS by Mandelbrot et al. Another noticeable property of traces is long-range-dependence.

Definition II.5. A second order stationary time series $X = \{X_n : n \in \mathbb{Z}\}$ is long-range dependent (LRD) or is said to have long memory if the auto-covariance function γ of X is not absolutely summable, i.e

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| = \infty \tag{4}$$

Equivalently, if $\gamma(k) = L(k)k^{2d-1}$ where $d \in (0, 1/2)$ and L is a slow varying function at infinity, that is, L is positive on $[c, \infty)$ for some $c \geq 0$ and for any $a > 0$

$$\lim_{x \to \infty} \frac{L(ax)}{L(x)} = 1$$

See [17] for other equivalent definitions.

For a second order SS process $\{Y_n : n \in \mathbb{Z}\}$ with stationary increments and index $1/2 < H < 1$ the process $\{X_n = Y_n - Y_{n-1} : n \in \mathbb{Z}\}$ is LRD with $d = H - 1/2$, see [17] for details. This fact also reinforces our ongoing support for stable models.

The permissible (α, H) region for non-degenerate α- stable and SS processes of index H with stationary increments is described in [12] pg 317.

With the alpha-stable distribution and its relationships to SS and LRD described, we now examine how the alpha-stable distribution has been applied in prior work and applies to our datasets.

C. Prior modeling work

Only a brief overview of key network traffic models is warranted, as exhaustive discussion is readily available in the literature, including [6], [18]. Poisson-based models for aspects of aggregated traffic were shown to be inaccurate in the mid-1990s [8]. To reflect the observed heavy tails, SS, and LRD in the network core, numerous models were subsequently developed using hybrid approaches such fractional ARIMA, fractional Gaussian, fractional Brownian, and Pareto burst processes, among others [6], [19]. Work in the related area of anomaly detection improved detection accuracy using Gamma and then alpha-stable distributions as traffic models [2], [3]. Our previous work confirmed the alpha-stable detection results in [3], finding that the heavy tail and four parameters of alpha-stable distributions most accurately described a variety of simulated and real datasets, even in the presence of severe noise in the form of cyber attacks [4].

For this work we decided to evaluate two new datasets as well as two new MAWI traces. As we began evaluating the data, we identified that their characteristics mirror the disagreements in the literature regarding the "best" models: The packet rates for 3 traces are described by non-Gaussian, alpha-stable (i.e., heavy-tailed) distributions, while the WAND trace is nearly Gaussian. This can be seen in Figure 3, which compares the Gaussian and alpha-stable maximum-likelihood (ML) fits of per-subwindow packet count for a randomly-selected 5 s window of each of our four datasets. The stable fit parameter α and normalized negative log-likelihood value of the ML fit are given in the figure.

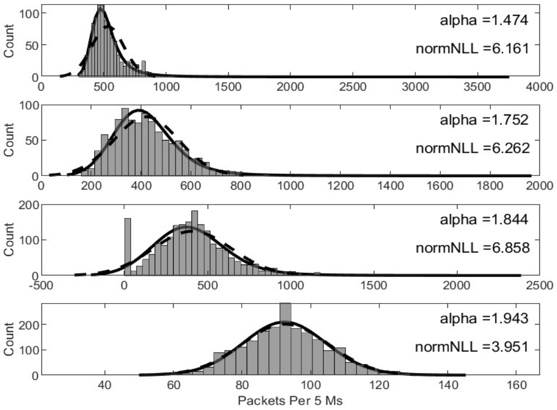

Figure 3. Packet count per subwindow histograms for a randomly-chosen 5 s window of the MAWI Apr, MAWI Nov, NPS, and WAND data sets (listed from top to bottom). The ML Gaussian and Stable fits are shown by dashed and solid lines, respectively.

The near-Gaussian fit of the WAND trace is likely due to its overall low volume (7 impulses per typical subwindow). At such a low aggregation level, the fat tail is less likely to significantly affect the distribution in the sense that one is less likely to sample outliers from the heavy tail (a window is about 1000 samples). This sampling effect may also affect our modeling and simulation results in subsequent sections. Note that, as shown in Figure 6, the tail exponent of the WAND trace is estimated to be much larger than the other datasets, and larger tail exponent values predict that the aggregated distribution will converge to Gaussian.

D. *Heavy tails in examined datasets*

To apply our impulse and renewal models by approaching these datasets as aggregations of heavy-tailed impulses, we must first confirm the heavy-tailed nature of our inputs and verify this characteristic is persistent. To accomplish this, we measured the slope of the tail impulse volumes for sub-window sizes between 1 and 10 ms. We found the slopes to be fairly invariant with respect to the length of the sub-window, as shown by the upper plots in Figure 4.

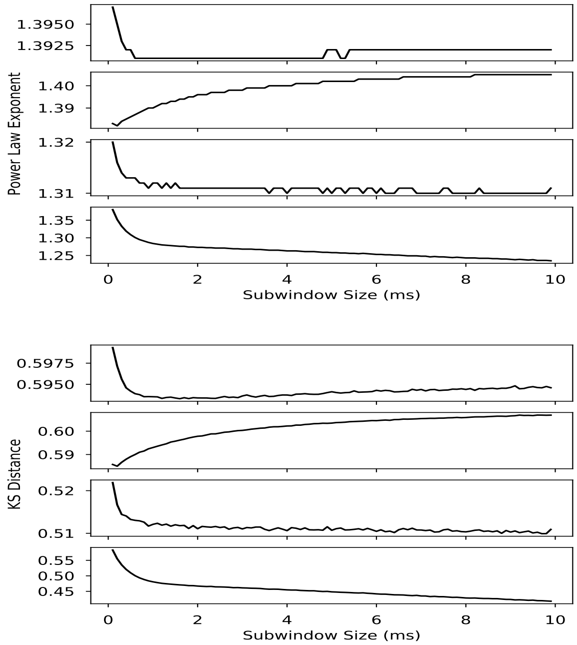

Figure 4. Plots of tail slope values (top) and fit sensitivity (bottom) to sub-window size for MAWI Apr, MAWI Nov, NPS, and WAND data sets (top to bottom).

These plots asses the consistency of the tail-size slope and quality of Pareto fit (as given by Kolmogorov-Smirnov (K-S) distance) for a randomly-chosen 1 s window from each trace. Note that, the WAND network is more sensitive to sub-window size due to the network's small size and limited variety of inputs, as previously discussed.

As part of this analysis, we observed that trace fits tend towards Gaussian (i.e., $\alpha \to 2$ at large sub-windows (e.g., 30 ms or more). In contrast, changing window size had little effect on fits, thus increasing the size of the window should only improve the fit of our model due to the increased number of aggregation samples.

For a fixed window, gradually increasing the sub-window size effectively increases the magnitude of impulses belonging to packet flows that were previously segregated into a different sub-window, introducing new impulses to the aggregation. The increase of sub-window size could also be understood as a re-scaling and studied using a self-similarity mindset. In practice, larger sub-windows can lead to decreased variation in the aggregations and smaller populations of samples. We note that this sub-window size dependency, if consistent across other datasets, may lead to (possibly) inappropriate conclusions of Gaussianity and the application of statistical measures such as mean that are not appropriate for heavy-tailed distributions.

Having characterized our dataset inputs, we will now propose an impulse aggregation model for heavy-tailed data.

III. The impulse model

The goal of this simple model is to explain the self-similarity tendency of network traffic by describing the aggregation of packets inside the sub-windows. We will define an *impulse* as a group of time stamps (packets) within a sub-window that are related by a unique source IP and destination IP pair $(\text{IP}_0, \text{IP}_1)$.

For a given window size (e.g., 5 s) we can characterize a trace based on the impulses in each sub-window (typically on the order of 5 ms). Impulses are ordered in such a way that $\mathbb{P}(Y_i = a)$ does not depend on i, where Y_i is the volume of the ith impulse. The total volume of traffic in a sub-window is given by the sum of the impulses within it, the number of which is described by a distribution E.

We expect the center of this distribution to shift in the positive direction as the size and complexity of the network increases, while the variance should stay bounded. In other words, the distribution of the volumes of all impulses Y_i and should have comparable tail decay for similar networks. We can get a sense of the complexity of our network datasets by counting the instances of each process in each subwindow, defined as impulses with a unique $\{\text{IP}_{\text{source}}, \text{IP}_{\text{destination}}\}$ pair. These results are shown in Figure 5. The results in Figure 6, a plot of the packet counts per sub-window for each of our four datasets on a log-log plot, confirm our expectations; even as network size increases between the NPS and MAWI traces, the variance (e.g., tail slopes) remains equivalent.

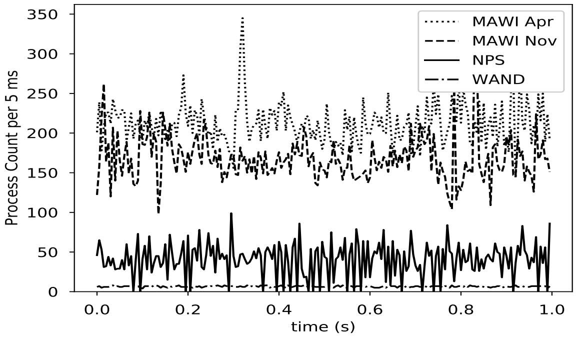

Figure 5. Process count per 5 ms sub-window plotted over a randomly-selected one second interval for four different networks.

The aggregation of traffic for a generic sub-window is then expressed by

$$S = Y_1 + Y_2 + ... + Y_e \tag{5}$$

where e is sampled from E.

A. Verifying assumptions: Independence and identical distribution

The IID assumption is critical to aggregation using the GCLT and renewal theory; we will first evaluate this assumption with respect to our data.

The independence assumption of processes and their impulses Y_i within a sub-window can be assumed to be independent. This is a reasonable assumption based on the diversity of communicating devices and processes in larger networks, particularly when further randomized by user action and network effects.

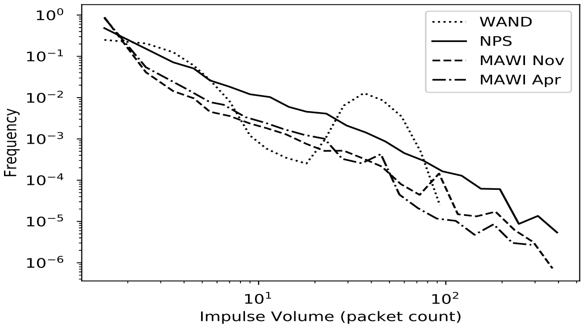

Figure 6. PDF of packet count per 5 ms sub-window on a log-log scale. The distributions are constructed from one second windows from each capture. The linearity of the plots suggests a power law distribution with estimated PDF tail slopes of 1.31, 1.40 and 1.40 for the NPS, MAWI Nov, and MAWI Apr data sets respectively. Note that the WAND slope is not estimated due to its variability.

The common distribution of impulses Y_i is given by V; these impulses must be ID. While a given activity may not necessarily ID across devices over a long period of time, this assumption is more plausible if we restrict ourselves to relatively short windows (e.g., over a window with length shorter than a typical video). For instance, the YouTube communication in Figure 2 looks fairly ID across several 5 s windows.

We can also quantitatively estimate the strength of our ID assumption. To evaluate impulse distribution for each of our four traces, we randomly selected two different indices i in 1 ms sub-windows. For each sub-window over 800 s of data, we then counted the impulses associated with these indices and compared their histograms using a Kolmogorov-Smirnov (K-S) test. Based on the observed small K-S distances between randomly-selected indices observed in Figure 7, the ID assumption is justified.

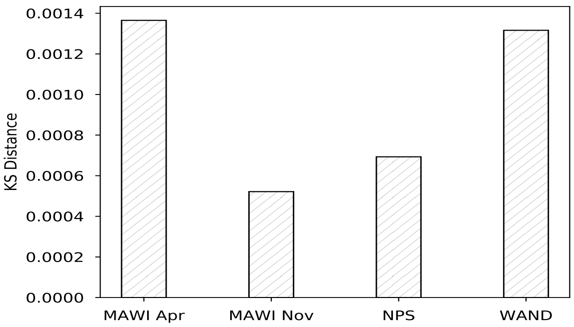

Figure 7. Plot of K-S distance between actual and randomized impulse volume distributions for each of our datasets.

Now that we have established the IID and heavy-tailed nature of our input processes, we can examine how they aggregate to alpha-stable (or Gaussian) network traffic.

B. Impulse aggregation

It is well documented in the literature that V can frequently be accurately approximated by heavy-tailed functions belonging to the domain of attraction of alpha stable distributions [7]. Heavy-tailed processes are also evident in three of our four traces, as shown in Figure 6. Heavy-tailed inputs are known to aggregate to stable distributions in accordance with the following theorem.

Theorem III.1. *Let $Y_1, ..., Y_n$ be IID with cumulative distribution F. Then $Y_1 \in \mathscr{D}(X)$ with $X \sim S_\alpha(1, \beta, 0)$ if and only if*

$$i)\ x^\alpha[1 - F(x) + F(-x)] = L(x) \text{ is slowly varying at infinity}$$

$$ii)\ \frac{F(-x)}{1 + F(-x) + F(-x)} \to \frac{1-\beta}{2} \text{ as } x \to \infty \tag{6}$$

In terms of Definition II.2, $d_n = n^{1/\alpha} L_0(n)$, where $L_0(n)$ is a slow varying function at infinity. (See [12] for explicit conditions on d_n and b_n.) In this application, where $F(-x) = 0$ for $x > 0$, condition $i)$ reduces to what we will refer to as *fat right tail* or simply *heavy tail.*

Given that we have established IID and heavy-tailed inputs, the asymptotic behavior of the aggregation in (5) can now be studied using the GCLT. Specifically, we can estimate the convergence rate under the stronger assumption that Y_1 lies in the strong domain of attraction of a stable distribution [20].

Convergence rate estimates in terms of the Mallows distance are given in [21].

The convergence rate permits evaluating tolerable errors in the stable fits used in modeling and anomaly detection. Note that classical analysis using higher-order moments is unavailable for non-Gaussian alpha-stable distributions [22]; alternatively, convergence can be studied via truncated moments [23], log-statistics [24], or fractional moments [25]. Further investigation of convergence rate and sensitivity to input populations are items for continuing work.

C. From sub-windows to aggregated traffic at the window level: Ergodicity

The sub-window aggregation model can provide information about the distribution of a generic window under the assumption of ergodicity of the trace. Consistent with many models in the literature, our proposed models rely on assumptions regarding ergodicity of the sub-window aggregations (or impulse superposition) as a discrete stochastic process; in this section we evaluate that assumption empirically.

Intuitively, we think of a window as a set of consecutive samples of sub-windows (typically between 600 and 1000). The distribution of the aggregation of the random variables defined above determines the outcome of randomly selecting sub-windows within a stationary trace. We can assume stationarity based on data windows and trace lengths in this work being shorter than empirical thresholds in the literature [3], [8], but a condition stronger than stationarity is required to determine the distribution of the aggregated result.

When interpreted as a Bernoulli scheme (permissible by the sub-window based modeling and the discrete volume variables) the model inherits the ergodic property which implies that the distribution of a large enough window approximates the *sample* distribution of the aggregation (5), the appropriateness of this assumption can be evaluated qualitatively using Figure 8.

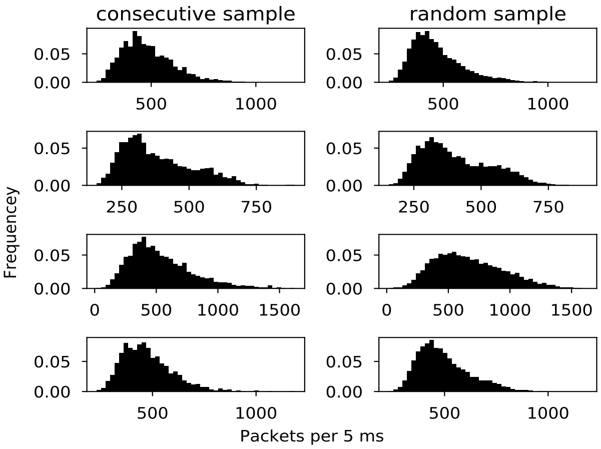

Figure 8. Qualitative evaluation of ergodicity through comparison of the distributions of measured and generated packet rates during a 30 second period of traffic for the Mawi Apr, Mawi Nov, NPS, and WAND traces (top to bottom).

Figure 8 compares the distribution of sampled to generated traffic sub-windows for 30 second portions of our four traces. The left-side histograms represent the packet rate density for 2,000 consecutive 5 ms sub-windows, while the right-side histograms were obtained by sampling 10,000 sub-windows at random (with overlap permitted).

The relative equivalence between the left and right figures demonstrates that the Bernoulli basis for ergodicity can be supported, and justifies application of the impulsive model theory to our datasets. The celebrated Ergodic Theory has many applications in the fields of Dynamical Systems, Stochatic Processes, Number Theory and many others. We refer the interested reader to [26] for a formal introduction to the subject.

IV. The Renewal Process Model

In this section, we interpret traffic as renewal processes whose aggregation is studied by Taqqu and Lévy in [27]. Specifically, they look at processes of the form

$$X^*(T, M) = \sum_{t=1}^{T} \sum_{m=1}^{M} X_{m,t} \tag{7}$$

where for each t, the random variables $X_{m,t} : 1 \leq m \leq M$ are IID copies of a renewal process. Two such processes are considered and the asymptotic behavior of $X^*(T, M)$ is explored. They discovered that for one of the considered processes, the accumulation $X^*(T, M)$ approaches a Gaussian fractional Brownian motion when $T << M$, and a stable process when $T >> M$. See [27] for a quick note on how these two SS processes differ.

A. Adaptation to the network traffic case

This venue of modeling SS was first proposed by Mandelbrot. In this paper, we simply offer an implementation of the processes in [27] with possible physical interpretations of their T, M parameters. The predictions made in that paper are empirically supported.

For a given $1 \leq t \leq T$, we think of the $\{X_{m,t} : 1 \leq m \leq M\}$ as a set of similar processes (say YouTube activity on a network due to many different users), whereas we interpret the index t as ranging across different processes or network activities in the distribution sense.

B. The impact of ON and OFF times

For each $k \geq 0$ we think of W_k as an independent copy of the random variable of number of packets over time rates with common distribution R, which we now assume to be truncated (W_k are assumed to posses finite second moments). U_k represents an independent copy of the packet flow duration with distribution U (the ON durations) and similarly F_k denotes the OFF period duration with distribution F. The variables F_k are absent in Taqqu's and Levy's considerations but it will promptly be clear that their results are still applicable. U_k will be assumed to satisfy the same conditions as in [27], namely they are IID and have finite variance or belong to the domain of attraction of a stable distribution with $1 \leq \alpha \leq 2$. These conditions are also extended to F_k. In addition, W_k is independent of U_k and F_k. Figure 9 shows how activity and inactivity periods are shadowed by power decaying distributions for a considerably long period of time; nevertheless, a sharp deviation from this trend is clearly expected at some point. This truncation imposed by physical constraints appears to fall under the term *soft truncation* introduced in [23].

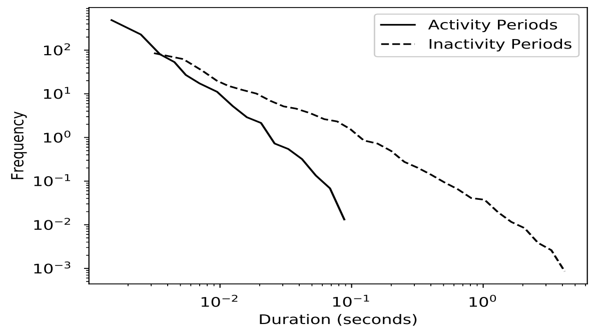

Figure 9. Plotted above are the probability distributions of packet flow duration and the interruption periods between packet flows related to the same process.

In order to compute the ON durations, we first define a packet flow as a string of packets related by $(\mathrm{IP}_0, \mathrm{IP}_1)$ possibly extending over several sub-windows (i.e a consecutive group of impulses). Two packets belong to the same flow if they are less than one sub-window apart. We also define the random variables S_k and E_k given by

$$\begin{aligned} S_k &= S_0 + \sum_{j=0}^{k-1} U_j + \sum_{j=0}^{k-1} F_j \qquad k \geq 1 \\ E_k &= S_k + U_k \qquad k \geq 0 \end{aligned} \tag{8}$$

representing the start-time and end-time of a packet flow respectively, analogously to [27]. $I_k = (S_k, E_k]$ denotes the kth ON interval.

Finally, we define the random variables

$$\delta_k = \begin{cases} 1 & w_k \cap (\bigcup I_j) \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

where w_k refers to the kth sub-window in the trace.

A signal is now expressed as $X_t = \sum_{k=0}^{\infty} W_k \delta_k$ and we interpret the expression $\sum_{t=1}^{T} X_t$ as the superposition of the volume of several impulses at a sub-window under the assumption of stationarity.

The sum of m copies of $X(t)$ in a given sub-window suggests the traffic of m "similar" processes. We expect that the relation $T >> M$ is satisfied in large networks and in traces captured at busy nodes due to the increased effect of perturbations and noise.

In this case, the finite distribution of

$$\frac{X^*([Ty], M)}{(MT)^{-1/\alpha} L(T)}$$

converges to an α-stable process when $T \to \infty$ first and then $M \to \infty$. L again denotes a slow varying function at infinity, $\alpha \in (1, 2)$ and $y \in [0, 1]$. See [27] for the complete theorem including the reverse order of the limits and the convergence to fractional Brownian motion.

V. Model Validation

In this section we show how these simple models capture some of the main properties of real network traffic. One of our goals was to describe traffic in the following way:

$$\begin{aligned} \text{Real Traffic}(x_1, ...x_L) &\overset{\mathrm{d}}{=} \text{Toy Model}_1(V, E) + \text{error}_1 \\ &\overset{\mathrm{d}}{=} \text{Toy Model}_2(T, M, U, F) + \text{error}_2 \end{aligned}$$

where both error_1 and error_1 go to zero asymptotically, while the models are as simple as possible but can still capture the main features of network traffic. Notice that the above models are very much related; and in fact, we think of them in terms of the relation

$$\text{Toy Model}_2(T, M, U, F) \overset{\mathrm{d}}{=} \text{Toy Model}_1(V, E) + \text{error}(U, F)$$

Their asymptotic convergence is demonstrated in Figure 13 and will be discussed in more detail later in this section.

Model 2 is strongly related to the $M/G/\infty$ construction due of Cox in the sense that similar conditions are assumed for the ON/OFF durations; however, our method does not assume that the volumes are heavy-tailed *and* or that the packets arrival rates is constant. See [28].

A. Model description

The predicting power of both models lie on their foundations on asymptotic results. By understanding the limiting behavior of the aggregation and the convergence rate in a given metric, it is possible to compute an upper bound on the observed error at a fixed aggregation with serves as a tolerance level in the anomaly detection phase.

These traffic descriptions can then be used to model traffic by following the process described in Algorithm 1. Note that this algorithm describes the renewal process-based model and is straightforward to modify for the impulse-based model.

```
Result: Catalog signal impulses from dataset
for each sub-window $\in$ window $\in$ dataset do
    get timestamp and packet count of each
    {Source IP, Destination IP} pair;
end
// Determines average number of unique
   signals over all windows (N), a
   placeholder for M and T in the
   current simplified simulation.
Result: get ON and OFF durations from a signal
for each window $\in$ dataset do
    record length of consecutive packets (packet flow)
    and length of consecutive null packet count
end
Result: Generate a sample signal
while position is not the last subwindow do
    sample ON durations;
    sample OFF durations;
    sample a corresponding number of volumes and
      assign to subwindows
    position = position + ON + OFF
end
Result: Generate renewal process-based traffic model
for $i = 1 : N$ do
    generate a signal
    aggregate signal to trace
end
```

Algorithm 1: Simplified renewal processes-based traffic model algorithm.

We note that the major intended contributions of this work are complete at this point: We have established a causal theoretical connection between heavy-tailed process inputs and SS and LRD that can result in Gaussian *or* alpha-stable aggregated network traffic, depending on specific conditions of the inputs. Grounded in this theory, we have also outlined two complementary models that utilize the observations of heavy-tailed and explain the alpha-stable marginal distributions of certain features of aggregated, large-network traffic.

Refinement of the model implementations and their outputs remains a work in progress, but we can present initial results that support our overall methodology.

B. Model assessment

As a preliminary check, we can assess the quality of the simulated traces generated by our two models both in terms of visual similarity to the parent trace and in terms of their ability to manifest LRD. The parent trace for this validation, shown in Figure 10, is a plot of packet count per sub-window for a randomly-chosen 5 s window of the MAWI 2017Nov11 trace.

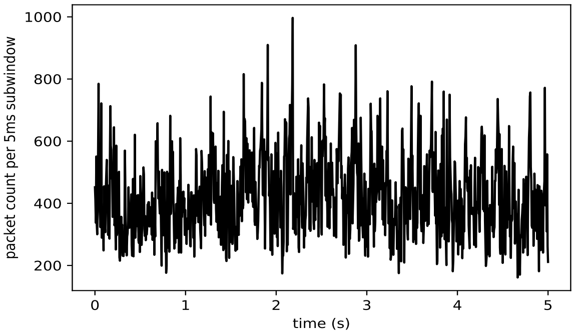

Figure 10. 5s of the MAWI Nov 11 trace

Due to space constraints, we only present the results for the renewal model. We note that this model presents slightly better *visual* results, but the autocorrelation and power spectral density analyses results are essentially identical.

The renewal model's reconstruction of the parent trace is shown in Figure 11. The renewal model provides slightly more fidelity, both in terms of variation and aperiodicity, than the impulse model and thus gives a better appearance of self-similarity.

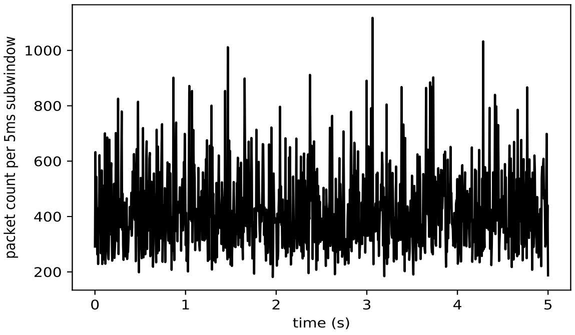

Figure 11. Generated packet rate over time of the simplified renewal processes model using only a historical distribution of ON durations for 5 s of the MAWI 11 Nov trace.

Both models do not possess the same magnitudes of aperiodic short-term volume displacement evident in parent trace. This is potentially the result of only using 5 s of parent trace data. These results are still promising, particularly given that this plot was generated using the simplified renewal model, still in development, that does not incorporate OFF periodicity.

Also, incorporating more than 5 s of process history into the model library may increase accuracy. These are both items of future work.

The autocorrelation results of the impulsive model are encouraging (but not shown); there is slow variation at increasing lags consistent with existing literature evaluations of heavy-tailed traces exhibiting LRD [29].

We can assess the ability of the renewal model to reproduce LRD by examining the model's power spectral density; this is shown in Figure 12.

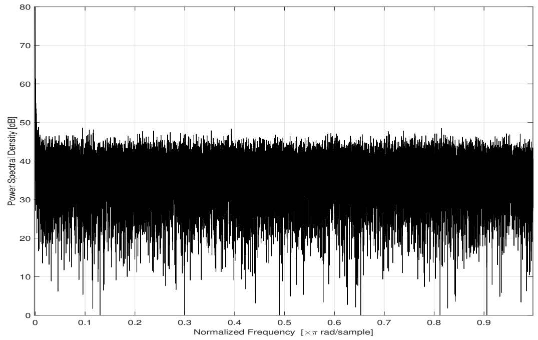

Figure 12. Power spectral density plot of the renewal processes model for five minutes of data.

LRD, consistent across both models, is implied by the non-zero asymptotic result at high lag. Figure 12 was produced from a generated 5 minutes of trace. We note that the initial rate of spectral decay is higher than expected and that the tail decay appears to be smaller than expected; these results are attributed to the in-progress model and will be re-examined as part of future work.

Finally, it is important to assess the flexibility of our modeling method by comparing generated and parent traces across all datasets. These results are summarized in Figure 13, which shows the K-S distance between the CDFs of the trace impulses of the parent traces (given by solid lines) and the model-generated traces (given by dashed lines) for the impulse model only.

To create Figure 13, we first selected a random 5 s window from the parent trace and applied the *Catalog impulses* step of Algorithm 1 to determine the distribution of impulse volumes. For each respective capture, these volumes were randomly divided into sub-windows with 212, 167, 40, and 7 impulses in each sub-window; impulse counts were determined by the mean number of impulses per sub-window in a one-second period for each trace. Finally the volumes of these impulses were summed for each sub-window to provide a distribution of packet count per sub-window for the overall window. The impulse distribution for the parent and generated traces were then compared to determine the K-S distances shown in the figure.

To explain this observation in terms of the renewal process description, we first notice that WAND was obtained from a residential network. For such a trace, the sample distribution of impulse volumes is sparse while the number of similar processes appearing in the network is comparably high (estimations of T and M for several networks will be included in future work), under these assumptions we anticipate the relation $T << M$ which [27] assures converges to Gaussian fractional Brownian motion.

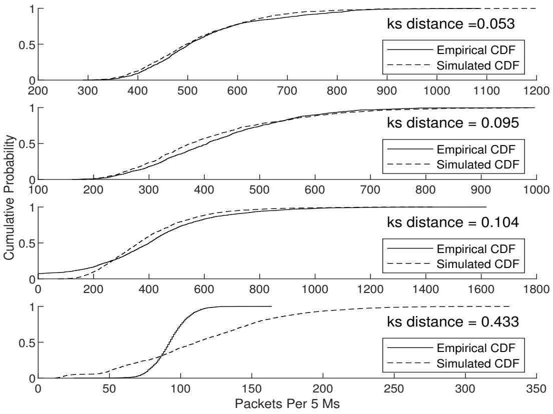

Figure 13. Solid lines indicate the cdfs of packet count per 5 ms across randomly chosen 5s windows for the MAWI APR, MAWI Nov, NPS, and WAND data sets (listed from top to bottom). Dotted lines indicate the simulated distributions, which were generated in the following manner.

We note that the comparison for the renewal model is not shown (again due to space); the results are similar and K-S distances are 0.054, 0.078, 0.17, and 0.522, respectively. Model performance improves in terms of K-S distance as the observed process count increases; it is intuitive that a better sample population improves the accuracy of the model. The tail accuracy of the model is significant, however, as many existing traffic models suffer in this region. Sensitivity of modeling accuracy to input population and the nature of trace activity (e.g., heavy hitters, mice, etc.) is an item of future work.

VI. Conclusions

This work establishes conditions for the alpha-stable aggregation of network traffic from individual device processes in larger networks. The alpha-stable distribution is closely tied to SS and LRD.

At many scales, process traffic can be characterized as impulses defined by large variations in amplitude with small on- and large off-periods. A model consisting of a small subset of these impulsive processes observed on a typical, centrally-managed campus network was created; the individual processes were found to rapidly aggregate, creating alpha-stable distributed network traffic.

The results of this model empirically validate the proposed two complementary, theoretical aggregation mechanisms resulting in alpha-stable distributed traffic: Renewal processes and impulsive processes leading to self-similarity. These two models show how features of network traffic can tend to exhibit Gaussian characteristics in small networks while growing heavy-tailed in larger networks (i.e., campus-sized or greater). This result also provides an alternative explanation for the SS and LRD traffic characteristics observed in large-volume traces.

We note that should alpha-stable distributions of network traffic become more widely observed and accepted, a re-examination of traffic measurement conventions may be warranted. Alpha-stable distributions lack higher-order moments, implying that methods using measures such as standard deviation, variance, power (and in some cases, mean) should be exchanged for those reflecting the nature of the traffic. The gain in performance from using appropriate measures in the presence of alpha-stable distributions is well documented [24], [25].

Items for future work include both extending the breadth of granularity of our aggregation models, more rigorously assessing the assumptions inherent to applying these models, and ultimately applying these findings to improve existing alpha-stable based network anomaly detectors. Processes from personal devices such as laptops and mobile phones can be characterized and added to our models, which would permit extending these results to wireless networks. These process aggregation models can be further enhanced by incorporating processes of typical network attacks (e.g., denial-of-service); this should provide a forecasting mechanism to differentiate normal from anomalous conditions. Ultimately, by identifying *when* network traffic features should be alpha-stable distributed and the effects of a variety of network attacks, distribution-appropriate anomaly detection algorithms can be deployed that more accurately reflect the actual, observed traffic characteristics with reduced false positive rates.

References

[1] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, "Scaling phenomena in the internet: Critically examining criticality," *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 1, pp. 2573–2580, 2002.

[2] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry, "Non-gaussian and long memory statistical characterizations for internet traffic with anomalies," *IEEE Transactions on Dependable and Secure Computing*, vol. 4, no. 1, pp. 56–70, 2007.

[3] F. Simmross-Wattenberg, J. I. Asensio-Perez, P. Casaseca-de-la Higuera, M. Martin-Fernandez, I. A. Dimitriadis, and C. Alberola-Lopez, "Anomaly detection in network traffic based on statistical inference and alpha-stable modeling," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 4, pp. 494–509, 2011.

[4] C. Bollmann, M. Tummala, J. McEachen, J. Scrofani, and M. Kragh, "Techniques to improve stable distribution modeling of network traffic," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[5] R. Fontugne, P. Abry, K. Fukuda, D. Veitch, K. Cho, P. Borgnat, and H. Wendt, "Scaling in internet traffic: a 14 year and 3 day longitudinal study, with multiscale analyses and random projections," *IEEE/ACM Transactions on Networking (TON)*, vol. 25, no. 4, pp. 2152–2165, 2017.

[8] V. Paxson and S. Floyd, "Wide area traffic: the failure of poisson modeling," *IEEE/ACM Transactions on networking*, 1995.

[6] 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.

[7] W. Willinger, V. Paxson, and M. S. Taqqu, "Self-similarity and heavy tails: Structural modeling of network traffic," *A practical guide to heavy tails: statistical techniques and applications*, vol. 23, pp. 27–53, 1998.

[9] K. Sato, *Levy Processes and Infinitely Divisible Distributions*. Cambridge Stud. Adv. Math. 68, 1999.

[10] J. Micheel, I. Graham, and N. Brownlee, "The auckland data set: an access link observed," in *Proceedings of the 14th ITC specialists seminar on access networks and systems*, 2001, pp. 19–30.

[11] R. Fontugne, P. Borgnat, P. Abry, and K. Fukuda, "MAWILab: Combining Diverse Anomaly Detectors for Automated Anomaly Labeling and Performance Benchmarking," in *ACM CoNEXT '10*, Philadelphia, PA, December 2010.

[12] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. CRC Press, 1994.

[13] W. Feller, *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, Inc., 1971, vol. 2.

[14] B. Gnedenko and A. Kolmogorov, "Limit distributions for sums of independent random variables," *Addison-Wesley Publishing Company, Inc*, 1968.

[15] E. J. G. Pitman and J. Pitman, "A direct approach to the stable distributions," *Advances in Applied Probability*, vol. 48, 2016.

[16] M. S. Taqqu and V. Teverovsky, "On estimating the intensity of long-range dependence in finite and infinite variance time series," *A practical guide to heavy tails: statistical techniques and applications*, vol. 177, p. 218, 1998.

[17] V. Pipiras and M. Taqqu, *Long-Range Dependence and Self-Similarity*. Cambridge University Press, 2017.

[18] W. Willinger, V. Paxson, R. H. Riedi, and M. S. Taqqu, "Long-range dependence and data network traffic," *Theory and applications of long-range dependence*, pp. 373–407, 2003.

[19] D. Ammar, T. Begin, and I. Guerin-Lassous, "A new tool for generating realistic internet traffic in ns-3," in *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques*. ICST (Institute for Computer Sciences, Social-Informatics and . . . , 2011, pp. 81–83.

[20] S. Manou-Abi, "Rate of convergence to alpha stable law using zolotarev distance: technical report," *arXiv preprint*, 2017.

[21] O. Johnson, R. Samworth *et al.*, "Central limit theorem and convergence to stable laws in mallows distance," *Bernoulli*, vol. 11, no. 5, pp. 829–845, 2005.

[22] V. M. Zolotarev, *One-dimensional Stable Distributions*. American Mathematical Soc., 1986.

[23] A. Chakrabarty and G. Samorodnitsky, "Understanding heavy tails in a bounded world or, is a truncated heavy tail heavy or not?" *Stochastic Models*, vol. 28, no. 1, pp. 109–143, 2012.

[24] J. G. Gonzalez, D. W. Griffith, and G. R. Arce, "Zero-order statistics: a signal processing framework for very impulsive processes," in *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*. IEEE, 1997, pp. 254–258.

[25] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, 1993.

[26] K. Petersen, *Ergodic Theory*. Cambridge University Press, 1983.

[27] M. Taqqu and J. Levy, "Using renewal processes to generate long-range dependence and high variability," *Dependence in Probability and Statistics. Progress in Probability and Statistics*, vol. 11, 1986.

[28] B. Cox, J. G. Laufer, S. R. Arridge, P. C. Beard, B. Cox, A. J. G. Laufer, A. S. R. Arridge, and P. C. B. A, "Long range dependence: A review," 1984.

[29] M. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar vbr traffic," in *SIGCOMM Symposium on Communications Architectures and Protocols*, pp. 269–280.