Towards An Explanatory Model for 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*—This work presents two explanatory mathematical models explaining how network traffic features that display Gaussian tendencies in single devices and small networks aggregate to alpha-stable processes in larger networks. The first model shows how self-similarity originates from an impulsive-noise-based representation of individual processes. A second model uses renewal processes to justify impulsive process aggregation to alpha-stable or Gaussian end states and permits estimating network traffic alpha-stable rates of convergence. We develop a model based on this first method to empirically validate this aggregation approach.**

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

I. Introduction

We propose explanatory, complementary models for aggregated network traffic, starting from the device level. Our purpose is to understand how network traffic aggregates from individual sources and investigate the tendency of certain network traffic features including packet rate to trend towards alpha-stable [1], [2] in larger networks, while the same features exhibit non-parametric or Gaussian distributions in smaller networks.

The central theme of our approach is an impulse-based model of traffic from individual processes on a single device. We observe that these traffic processes frequently are relatively periodic and consistent in terms of volume, as illustrated in Figure 1, a rate plot of traffic to a single, centrally-managed device. The impulse characterization of traffic is justified by the fact that even long transmissions (e.g., a video on YouTube) are sent as short bursts of traffic rather than constant-rate, lengthier transmissions.

Intuitively, the aggregation and perturbation of these traffic impulses from hundreds of devices leads to the frequently-observed self-similarity in aggregated network traces [3], [4]. These models have historically relied on Lévy processes $\{X_t : t \geq 0\}$, including Poisson and fractional Brownian motion [5]; we propose an alternative Lévy process, the alpha-stable, due to its deep theoretical connections with well-documented characteristics of traffic. More specifically, Lévy processes are self-similar (SS) *if and only if* they are alpha-stable [6]. Moreover, the domain of attraction of stable variables consist of random variables with heavy-tailed distributions per the Generalized Central Limit Theorem (GCLT).

The primary contribution of this work is the development of device traffic aggregation methods that can lead to alpha-stable distributed traffic under certain input conditions, and Gaussian-distributed traffic under other conditions.

II. The impulse model

The goal of this model is to explain the aggregation of traffic by looking inside the sub-windows where aggregations occur. We define an *impulse* as a group of time stamps (packets) within a sub-window that are related by a unique source and destination IP pair $(\mathrm{ip}_0, \mathrm{ip}_1)$. Different impulses are assumed to be independent.

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. V denotes the distribution of the volumes of all impulses. Traffic in a generic sub-window is given by $S = Y_1 + Y_2 + ... + Y_e$, where e is sampled from E. The Y_i's are assumed to be independent since they represent the volume of different processes within a sub-window and are also identically distributed by construction. Their common distribution is approximated by V. We provide

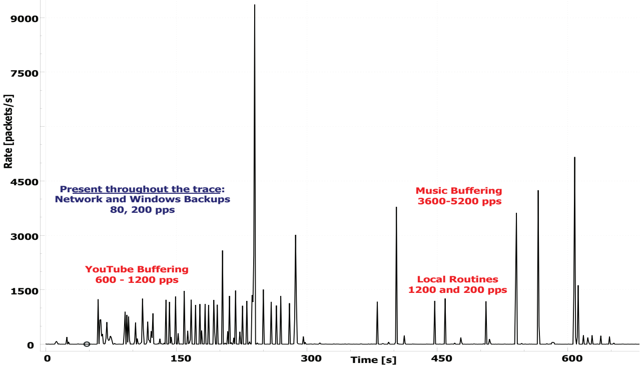

Figure 1. Rate plot of several minutes of traffic received at local host with periods of music and video streaming.

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

empirical evidence that V is well approximated by heavy-tailed functions belonging to the domain of attraction of alpha-stable distributions.

The asymptotic behavior of the aggregation S 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 [7].

The convergence rate serves to estimate the expected error in the stable fits.

A. Aggregation from sub-windows to window

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 at randomly selected sub-windows within a stationary trace.

When interpreted as a Bernoulli scheme the model inherits the stronger condition of ergodicity, which implies that the distribution of a large enough window approximates the *sample* distribution of the aggregation S.

Although the classical argument using moments is unavailable, we can justify this convergence using several other methods. The property is also demonstrated empirically in a longer work (in progress).

B. Aggregation using Renewal processes

We can also model traffic as the aggregation of renewal processes per Taqqu and Lévy in [8]. Specifically, they look at processes of the form

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

where for each t, the random variables $X_{m,t} : 1 \leq m \leq M$ are i.i.d. copies of a renewal process, and are interpreted as a set of similar processes (say streaming from many different users), whereas we interpret the index t as ranging across different processes or network activities in the distribution sense.

The accumulation $X^*(T, M)$ approaches Gaussian fractional Brownian motion (GfBm) when $T << M$, and a stable process when $T >> M$. See [8] for a quick note on how these two SS processes differ. We can assume that a certain process is identically distributed across devices if we restrict ourselves to relatively short windows.

For a given $k \geq 0$ we think of W_k as an independent copy of the random variable of number of packets over time 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; similarly, F_k denotes the OFF period duration with distribution F. The variables F_k are absent in Taqqu and Levy's considerations but it is clear that their results are still applicable. U_k will be assumed to satisfy the same conditions as in [8]; namely, they are i.i.d. 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 2 shows how activity and inactivity periods are shadowed by power-law distributions for a considerable period of time; nevertheless, a sharp deviation from this trend is clearly expected at some point.

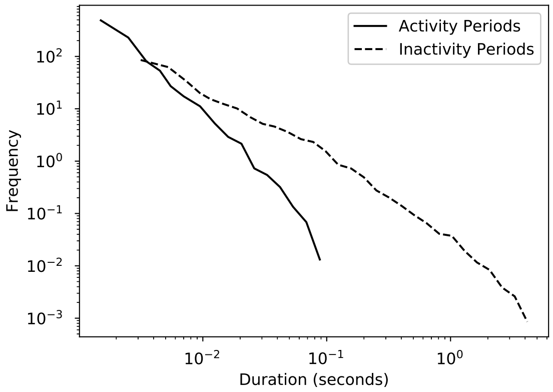

Figure 2. The probability distributions of packet flow duration and the interruption periods between packet flows related to the same process.

A *packet flow* is a string of packets related by $(\text{ip}_0, \text{ip}_1)$ possibly extending several sub-windows (i.e a consecutive group of impulses). Two packets belong to the same flow if they are less than two sub-windows apart. We also define the random variables S_k and E_k given by

$$\begin{aligned} S_k &= S_0 + \sum_{j=0}^{k-1} U_j + \sum_{j=0}^{k-1} F_j \qquad k \geq 1 \\ E_k &= S_k + U_k \qquad k \geq 0 \end{aligned} \tag{2}$$

representing the start-time and end-time of a packet flow respectively. $I_k = (S_k, E_k]$ 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 processes at a sub-window. The sum of M i.i.d. copies of $X(t)$ in a given sub-window suggests the traffic of M "similar" processes. We expect the relation $T >> M$ to be satisfied in large networks and in traces captured at busy nodes due to the increased effect of noise associated with the multi-tasking nature of devices using multiple network sockets.

C. Simulations

These simple models capture some of the main properties of real network traffic which depends on a great number of hard to quantify factors by describing it in the following way:

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

where both error_1 and error_2 go to zero asymptotically. Notice that these models are very much related. We think of them in terms of the relation:

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

Model 2 is analogous to the $M/G/\infty$ construction of Cox [9], in the sense that similar conditions are assumed for the ON/OFF durations. However, Model 2 does not require heavy-tailed volumes or constant packet arrival rates.

At low aggregation levels the models are expected to be inaccurate but the errors can be bounded by the use of convergence rate estimates.

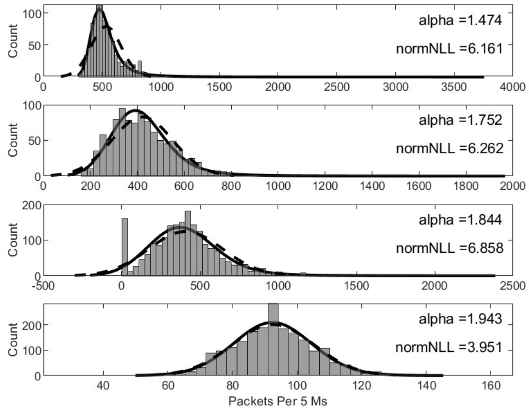

Figure 3. Gaussian (dotted line) and Stable (solid line) fits are shown above for the distributions of packet count per 5 ms across randomly selected 5 second windows. The MAWI Apr, MAWI Nov, NPS, and WAND data sets are shown in order of decreasing network size (listed from top to bottom).

In Figure 3, we observe that the small network WAND trace appears Gaussian. This is intuitively satisfying, as the tail decay parameter is estimated to be 2.40, and thus predicted to converge to a Normal distribution by the GCLT. For small residential networks, the variety of impulse volumes T is sparse while the number of similar processes M is comparably high; under this assumption $T << M$ which implies convergence to GfBm [8].

Errors can be large at low aggregation levels because the fat tail is less likely to significantly affect the distribution, (a window contains 1000 samples). As shown in Figure 4, the error in terms of the Kolmogorov–Smirnov test improves by almost an order of magnitude when the aggregation level increases from 7 processes in the WAND trace to 212 in the MAWI April trace. Moreover, Model 1 does not take into account ON/OFF periods, which leads to greater errors for small networks.

III. Conclusions

This work establishes conditions for the aggregation of network traffic from individual device processes where features of network traffic can tend to exhibit Gaussian characteristics in small networks and alpha-stable characteristics in larger networks. At many scales, process traffic can be characterized as impulses defined by large variations in amplitude with small ON- and large OFF-periods.

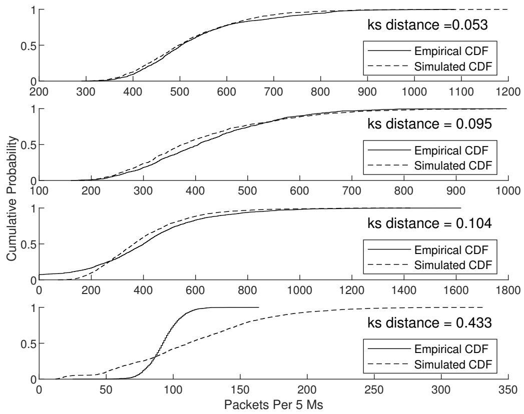

Figure 4. Solid lines indicate the CDFs of packet count per 5 ms across randomly-chosen 5s windows for the MAWI APR, MAWI Nov, NPS, and WAND data sets (listed from top to bottom). Dotted lines indicate the simulated distributions, generated using observed volumes and 212, 167, 40, and 7 impulses representing the observed mean number of processes.

Simulations of the considered traces validate the two proposed theoretical aggregation mechanisms. Alpha-stable distributed network traffic emerges at relatively low aggregation levels.

Items for future work include both extending the breadth of granularity of our aggregation models through considering additional sources of traffic as well as developing alpha-stable based methods of traffic measurement and anomaly detection.

References

[1] 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.

[2] 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.

[3] 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.

[4] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, “On the self-similar nature of ethernet traffic,” *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[5] 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.

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

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

[8] 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.

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