Connection-Level Modeling of Network Traffic

Xin Wang, Shriram Sarvotham, Rudolf Riedi, Richard Baraniuk

Abstract— **Aggregate network traffic exhibits strong burstiness and non-Gaussian marginals, which popular models like fractional Gaussian noise (fGn) fail to capture. To better understand the cause of traffic burstiness, we look into connection-level information of traffic trace. A careful study reveals that traffic burstiness is directly related to the heterogeneity in connection bandwidths (and round-trip times), and that a small number of high-bandwidth connections are solely responsible for burstiness. This leads to a new model for network traffic: the mixture of fGn and stable Lévy noise, which captures both long-range dependence and burstiness. This decomposition has far-reaching implications on network control.**

I. Introduction

Network traffic analysis and modeling play a major role in characterizing network performance. Models that accurately capture the salient characteristics of traffic are useful for analysis and simulation, and they further our understanding of network dynamics and so aid design and control.

Numerous studies have found that aggregate traffic exhibits *long-range dependence* (LRD) [1], and that traffic can be extremely bursty, resulting in a non-Gaussian marginal distribution [2]. These findings are in sharp contrast to classical Markovian type traffic models and their predictions [1], [3], [?].

Fractional Gaussian noise (fGn) is a stationary Gaussian process with LRD and is a popular model for network traffic. It owes its credibility to the fact that it can be obtained as the limit of the superposition of a large number of independent individual ON/OFF sources that transmit at the same rate but with heavy-tailed ON durations [1]. Here, it should be mentioned what is often overlooked, namely that the superposition of ON/OFF sources may also tend to i.i.d. stable Lévy noise. Roughly speaking, the Gaussian limit is obtained for very large numbers of sources (more precisely, when the number of sources is kept larger than a certain power of the time scale when both going to infinity), and the Lévy limit for very large time scales (see [4] for an overview).

In this paper we argue through a careful analysis of several traces of real network traffic that it is indeed a mixture of both limits that matches reality best. We explain this phenomenon through the heterogeneity of connection bandwidths.

II. Connection-Level Traffic Analysis

When modeling aggregate traffic as the superposition of ON/OFF sources, traffic bursts arise by definition only from a "constructive interference", i.e., large number of connections transmitting data simultaneously. However, Figures 1(a) and (b) demonstrate that this is not the case. In fact in most cases only *one* connection[1] dominates (see Figure 1(c)). This surprising finding has far-reaching implications for traffic analysis and modeling.

Thus motivated, we call any connection which sends more than a threshold of bytes during any time interval of a given size T an *alpha connection*. The (large) threshold is chosen based on the mean of the aggregate traffic at time-scale T plus a few standard deviations. We call all bytes sent by alpha connections the *alpha* traffic component. The residual traffic is called the *beta* component.[2] Our procedure thus decomposes aggregate traffic into

$$\text{total traffic} = \text{alpha traffic} + \text{beta traffic}. \quad (1)$$

We have applied the alpha/beta traffic decomposition to many real-world traffic traces at time scales T varying from 50 to 500 ms and found tremendous consistency in our results [5]. The statistical properties of the components can be summarized as follows.

Beta traffic: At time-scales coarser than the round-trip time, the beta component is very nearly Gaussian and strongly LRD (i.e., approximately fGn), provided a sufficiently large number of connections are present. Moreover, the beta component carries the same fractal scaling (LRD) exponent as the aggregate traffic.

Alpha traffic: The alpha component constitutes a small fraction of the total workload but is entirely responsible for the bursty behavior. Alpha traffic is highly non-Gaussian.

III. Origins of Alpha and Beta Traffic

While peculiarities in TCP's slow start phase, or random effects such as route changes could potentially cause dominant alpha connections, we identified in [5] the sole systematic cause of alpha connections to be the heterogeneity in connection bandwidths.

Department of Electrical Engineering, Rice University, Houston, TX 77005. Email: {xinwang, shri, riedi, richb}@rice.edu. This work was supported by NSF, DARPA/AFOSR, DOE, and ONR.

[1]In this analysis, a connection is identified through source and destination IP addresses and port numbers.

[2]By analogy to the dominating *alpha males* and submissive *beta males* observed in the animal kingdom.

Number of bytes per time   Number of connections per time   Bytes per connection during a burst

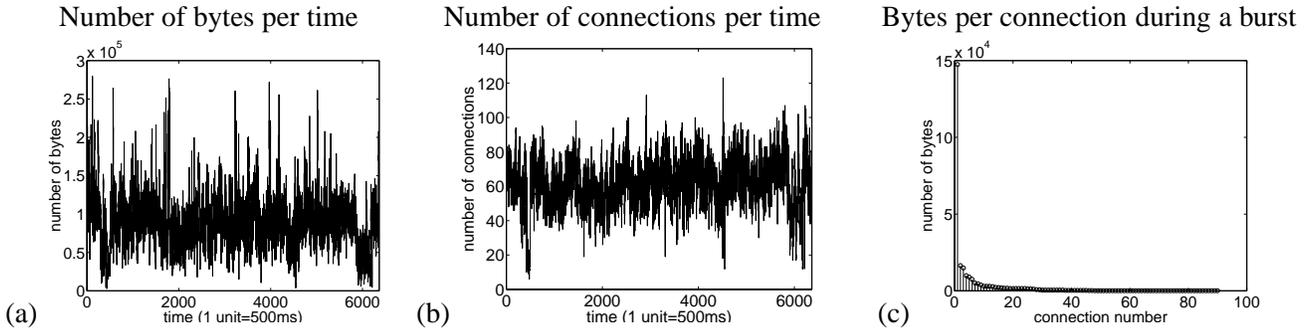(a)   (b)   (c)

Fig. 1. (a) Bytes-per-time and (b) number of connections-per-time of an aggregate traffic trace collected in Auckland. (c) Number of bytes per connection (sorted in decreasing order) during a typical burst. Clearly one connection dominates all others.

To support this claim we grouped all connections according to their source-destination IP addresses into end-to-end groups. We observed the following: If such a group contains any alpha connection then all other connections in the group are either alpha as well or transmit virtually all data in one single time bin of T ms. This is clear indication that this end-to-end path enjoys an extremely high bandwidth limit. On the other hand, if all connections in a group are beta, then all these connections experienced a clear common limit to the sending rate. With such a consistent picture, we can exclude re-routing and TCP peculiarities as a systematic cause for bursts, since it would be highly unlikely that all connections in a group would be systematically affected by it.

In summary, we conclude that the majority of burst-causing connections are due to *large file transmissions over high bandwidth end-to-end paths*.

IV. ALPHA AND BETA TRAFFIC MODELING

Since connection bandwidth is inversely proportional to the round-trip time (RTT) while RTT is intimately related to the network topology, the heterogeneity in connection bandwidth has its root in the topology of wide area network. We propose here a simple model to demonstrate the influence of RTT distributions on network traffic model.

Consider a scenario where a relatively small number of hosts locate close to the measured link while the majority of hosts are far away. As a consequence, the distribution of RTT is bimodal: a small number of connections with small RTTs and the remaining with large RTTs. Let us further assume that the TCP congestion windows of all connections are equal. The latter is certainly a rough approximation but draws some validity from our observations.

Applying the superposition on-off model or the infinite source Poisson model [4] to the scenario above, we obtain the following results. The alpha traffic is composed of the few connections with small RTTs and, therefore, high sending rates. The TCP clock ticks faster for these sources, therefore we approximate this component as the

on-off limit at infinite time, i.e., by i.i.d. Lévy stable noise. So the alpha traffic brings in burstiness but not LRD.

The beta traffic is made up from the bulk of remaining connections with large RTTs and low rates and is well approximated by the on-off limit at infinite source number, i.e., by fGn. It inflicts LRD on the overall traffic but is not as bursty as the alpha traffic.

While empirical evidence for consistency of measured network traffic with stable noise or fGn is not convincing at present, this paper claims that a closer match is obtained by an appropriate mixture of the two processes. The approach sketched in this paper explains the relevance of the two components in terms of an LRD background traffic produced by the "big crowd" and a bursty but light weight component due to a few "alpha connections". This has implications not only for modeling, but also for simulation, synthesis, estimation, prediction, performance evaluation and understanding of traffic dynamics.

This approach also opens a clear alley towards studying the influence of RTT distributions on network traffic dynamics which is currently under way.

REFERENCES

[1] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Trans. Networking (Extended Version)*, vol. 5, no. 1, pp. 71–86, Feb. 1997.

[2] R. H. Riedi, M. S. Crouse, V. Ribiero, and R. G. Baraniuk, "A multifractal wavelet model with application to TCP network traffic," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 992–1018, April 1999.

[3] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent traffic," *IEEE/ACM Trans. Networking*, pp. 209–223, April 1996.

[4] W. Willinger, V. Paxson, R. Riedi, and M. Taqqu, *Long range dependence : theory and applications*, chapter Long range dependence and Data Network Traffic, Wiley, 2002, Doukhan, Oppenheim and Taqqu eds.

[5] S. Sarvotham, R. Riedi, and R. Baraniuk, "Connection-level analysis and modeling of network traffic," *Proc. IEEE/ACM Network Measurement Workshop*, 2001.