The Effect of Statistical Multiplexing on the Long Range Dependence of Internet Packet Traffic (Extended Abstract)

Jin Cao, William S. Cleveland, Dong Lin, Don X. Sun

I. Introduction

A. Packet Traffic and Statistical Multiplexing

The packets transmitted over an Internet link result from simultaneous active connections, each sending its packets as part of a communication among two or more hosts. The packets of the different connections are intermingled on the link — for example, a packet is transmitted for one connection, then a packet for a second, another for the first, then two packets in a row for a third, and so forth. This intermingling is referred to as "statistical multiplexing" in the packet network literature and as "superposition" in the mathematical literature of point processes. We use the terms interchangeably here because we rely on both literatures.

This study treats the statistical properties of packet traffic and how the properties change with the magnitude of the statistical multiplexing. We investigate the properties theoretically, invoking the mathematics of marked point processes; we investigate the properties empirically, analyzing measurements of packet variables from live traffic; and we investigate the properties by network simulation, analyzing the packet traffic variables for synthetic HTTP traffic generated by NS plus statistical models that describe the application request process. Those interested only in the conclusions and a discussion of them can read this section and then proceed to Section II.

We study four packet traffic variables: packet sizes, packet inter-arrival times, byte counts in 100 ms intervals, and packet counts in 100 ms intervals. But we do not present the analyses of the byte counts in the interest of space because the conclusions are very nearly the same as those for the packet counts. We study the long-range dependence of the inter-arrivals, sizes, and packet counts, and we study the amount of statistical variability relative to the mean for the packet counts. The dependence and amount of variability is related to the number of active connections sending packets over the link (NAC), which measures the amount of statistical multiplexing. As the NAC increases from its smallest values, major changes occur. The change is complicated by an increase in the percent of back-to-back packets on the link due to increased upstream queueing.

The authors are in the Computing and Mathematical Sciences Research Division, Bell Labs, Murray Hill, NJ. {cao, wsc, dong, dxsun}@bell-labs.com

B. Previous Results

Packet counts and byte counts have been widely studied in the literature; the important discovery that these variables are long-range dependent dates back to two papers in the early 1990s ([1], [2]), the first using data on an Ethernet LAN, and the second corroborating the result for two links connecting LANs to the rest of the Internet. Since then, other writings (e.g., [3], [4], [5]) have corroborated and investigated the long-range dependence with a variety of other Internet measurements.

Models of source traffic have been put forward to explain the long-range dependence in the packet and byte counts [3], [6], [7], [8]. The mathematics can get intricate but the intuition for these on-off aggregation models is straightforward. The sizes of transferred files utilizing a link vary immensely; to a good approximation, the upper tail of the file size distribution is Pareto with a shape parameter that is often between 1 and 2. This means there is a nonnegligible probability of the transfer of a file that is considerably larger than the sizes typically transferred. If this happens with a high enough frequency for files large enough to have a substantial impact amidst the other aggregate link traffic, then the result is bursty activity. This burstiness increases the lengths of packet queues compared with Poisson arrivals and independent sizes, although there has been some debate about the magnitude and detailed causes of the effect ([9], [10], [11], [12]). However, the heavy tailed filesize distribution is not a necessary condition for long-range dependence. TCP by itself can create positive autocorrelations that taper off slowly [13], [14].

The statistical properties of the two traffic variables, inter-arrival and size, have been less studied, but nevertheless, long-range dependence has been convincingly demonstrated. The inter-arrivals in one study have a marginal distribution whose upper tail is longer than the exponential [2]. In three others, the long-range dependence of the inter-arrivals is demonstrated [15], [16], [17]. In a fourth, the inter-arrivals are long-range dependent, and wavelet modeling is used to investigate the properties of the dependence [5]. Thus, the arrival process is not Poisson for two reasons: inter-arrivals are not independent and their marginal distribution is not exponential. In [15], [16], the packet sizes are shown to be long-range dependent.

The potential importance of statistical multiplexing was understood at the very onset of studies of long-range dependence; an early discussion concludes with the statement [9]: "This is clearly an issue of practical importance, and there is considerable scope for further work." However, until recently, there has been little study of the effect of multiplexing; most of the work has been theoretical studies of queueing behavior [9], [18], [19], [20], [21], [22]. Also, through the experiences of network operators, it has been appreciated that the statistical variability of packet counts and byte counts, relative to the mean, decreases with the NAC; this is sometimes referred to as increased *smoothness* of the traffic.

Recently, the statistical properties of a number of traffic variables, measured on an Internet link between a large local network and the rest of the Internet, have been related to the NAC, establishing a pervasive nonstationarity in the variables in which fundamental statistical properties, such as shapes of marginal distributions and autocorrelations, change with the NAC [15], [23]. Included in the work is a study of packet arrivals and sizes. In this article we also focus on these traffic variables, plus packet and byte counts, and extend the work on nonstationarity, adding a theoretical framework and studying data with much higher NACs. Other past, recent, or ongoing results, somewhat different in approach, but relevant to issues studied here, also demonstrate that important changes in the statistics of packet traffic occur with increases in statistical multiplexing [24], [25], [26], [17].

C. Overview of Complete Article

The on-off aggregation model described above provides insights into the statistical behavior of the packet and byte count variables, in particular the long-range dependence. We put forward an alternative mathematical framework: the theory of the superposition of marked point processes. The packet process, arrival times and sizes, is a *marked point process*; the size of a packet is a mark at the arrival time of the packet. The theory leads to a collection of hypotheses about the changes in the statistical properties of the traffic variables as the NAC changes.

To test the hypotheses, we develop methods of analysis of empirical measurements of sizes, inter-arrivals, packet counts, and byte counts, and apply them to measurements from 2526 packet traces, each either 90 sec or 5 min in duration, from 6 Internet monitors: a 100 mbps Ethernet link, a 100 mbps Ethernet switch, two 156 mbps ATM links, and two 622 mbps packet-over-sonet links.

We also test the hypotheses by generating synthetic HTTP packet traffic. Statistical models that describe the HTTP request process generate values of application variables such as the request arrival process. The values of the request variables are handed off to NS where synthetic packet traffic is generated by simulating multiplexed TCP connections. Then the packet traffic variables are computed from the synthetic traffic and analyzed using the same methods used to study the live traffic. In the article, we describe the results briefly, in the interest of space; a fuller account is given in [24].

II. Discussion

A. Conclusions

The conclusions of this study involve the statistical properties of the packet traffic transmitted on an Internet link. Results are based on the following: (1) the mathematical theory of marked point processes; (2) empirical study of 2526 packet traces, 5 min or 90 sec in duration, from 6 Internet monitors measuring 15 interfaces ranging from 100 mbps to 622 mbps; (3) simple statistical models for the traffic variables; and (4) network simulation with NS.

Four packet traffic variables are studied as time series — sizes, inter-arrivals, packet counts in 100 ms intervals, and byte counts in 100 ms intervals. The last three series are transformed to make the resulting series as close to Gaussian as possible. The inter-arrivals are transformed by sixth roots, and the two count variables are transformed by logarithms. The conclusions describe the change in the statistical characteristics of the four variables as the magnitude of statistical multiplexing of packets from different active connections increases. Magnitude is measured by the number of active connections (NAC).

The sizes and inter-arrivals have components of long-range dependence at all levels of the NAC. This long-range dependence decreases as the NAC increases. The reduction is not reversed when the NAC becomes large enough to cause a significant percent of back-to-back packets from upstream queueing. Simple fractional sum-difference (FSD) statistical models describe the behavior; an FSD model consists of a white noise component plus a long-range dependent fractional ARIMA component. For the sizes, the percent of variance due to the fractional ARIMA component tends to zero; that is, the sizes tend to independent. For the inter-arrivals, either the percent of variance due to the fractional ARIMA tends to zero, or the fractional ARIMA tends to short-range dependent; that is, the inter-arrivals tend either to independent or short-range dependent.

The standard deviation relative to the mean for the packet counts and for the byte counts decreases toward zero like the inverse square root of the NAC; this means the statistical variability of the count series becomes small. The long-range dependence of the log counts, also fitted by an FSD model, does not change with the NAC. But this long-range dependence becomes less and less salient as the NAC increases because the standard deviation decreases relative to the mean; this means the sizes of the bursts due to long-range dependence become small relative to the overall traffic level. In theory, these results are altered when the NAC is large enough to cause a large percent of back-to-back packets; byte counts tend toward to a constant, so the standard deviation tends to zero, and the standard deviation of the packet counts tends to a positive constant. But the empirical study and the simulation study, which used 100 ms intervals, shows no evidence of the counts approaching this limiting behavior.

The results for the inter-arrivals and sizes might appear to be in conflict with those for the packet counts and byte counts. As the NAC increases, the long-range dependence of the inter-arrivals and sizes is reduced, but the counts retain the same long-range dependence. In fact, the FSD models for the sizes and inter-arrivals show the results are consistent. A small component of long-range dependence remains in the inter-arrivals and sizes, but there is less of the component for a higher NAC than a lower one, and more of the noise component. We go from arrivals and sizes to counts by summation of arrivals and sizes in fixed intervals of time such as 100 ms. There is more summation for a higher NAC. The summation operator is a low pass filter that tends to preserve the long-range dependent component and reduce the noise, so more summation results in more long-range dependence in the counts. The greater summation for the higher NAC increases the long-range dependence by more than that for the lower NAC, bringing the counts into alignment in terms of long-range dependence for the two NACs.

B. *Caveats*

These results do not account for sudden surges in demand due to events that create a surge in user demand for transfers due to some event such as an TV advertisement with a Web link or a major news event.

These results address what typically happens and not what can happen. A host with a path to an Internet link that is everywhere at or near the link speed could alter drastically the statistical properties of the packet traffic on the link. This was rare in our packet traces. For example, many users can can access the BELL 100 mbps Ethernet link at 100 mbps and so could employ an application that could dramatically alter the traffic. However, there are very few occurrences of this in the BELL data.

At very low values of the NAC, TCP dynamics can have a dramatic affect on the correlation properties of traffic variables. But these effects disappear quickly as the NAC increases from very low values. The TCP dynamics appear as strong peaks in the spectrum of the inter-arrivals. The FSD models do not fit these effects fully, but do provide a reasonable summary for studying long-range dependence.

C. *Implications for Internet Engineering*

The results presented here have important implications for Internet engineering. Long-range dependence of sizes and inter-arrivals creates buffer queueing distributions whose heights are much greater and much more variable than when the long-range dependence is removed [9]. For the most part, hardware design and network operations have been carried out under the presumption of a ubiquitous long-range dependence of sizes and inter-arrivals that causes large and highly variable queue heights at all levels of the NAC. For example, long-range dependence has led to setting maximum 15–60 min average utilizations of Internet links to be substantially less than they would be for independent or short range dependent sizes and inter-arrivals — often in the 50% to 60% range — to achieve high quality of service. The results presented here show that design and operations should be re-assessed for links with large peak NACs.

For example, the results make it likely that utilizations can increase even higher than the above range on these links. The reason is that other studies, described next, have demonstrated what one might expect of queueing distributions, given the results here — queue-height distributions relative to the traffic load decrease with the NAC.

One quite interesting theoretical queueing study compares three pairs of input traffic streams from multiplicative multifractal models with different long-range dependence characteristics: (1) both with the same degree of long-range dependence; (2) one with greater long-range dependence than the other; and (3) one with long-range dependence and the other with Poisson arrivals [26]. At 90% utilization, the queueing from the three multiplexed output streams are the following: (1) less queueing than the two inputs; (2) more queueing than the less dependent input and less queueing than the more dependent input; (3) less queueing than the dependent input, but more than the Poisson. In another theoretical study, it is shown that the multiplexing of more and more long-range-dependent homogeneous marked point processes results in queueing characteristics like that of a marked point process with Poisson arrivals and independent service times [25]. Recent empirical results show that open-loop queueing, using measured packet traces from the BELL database as input, also tends to that of Poisson arrivals with independent service times as the NAC increases [15]. In another open-loop empirical study, but not involving a changing NAC, link input arrivals are long-range dependent and packet sizes are independent and exponential and are independent of the inter-arrivals [17]. As the input link speed divided by the output link speed increases beyond 1, the long-range dependence of the queue output inter-arrivals goes to independence. This is consistent with our hypothesis that the inter-arrivals tend to the packet sizes divided by the link speed once the upstream queueing is sufficiently large.

References

[1] Will Leland, Murad Taqqu, Walter Willinger, and Daniel Wilson, "On the Self-Similar Nature of Ethernet Traffic," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1–15, 1994.

[2] Vern Paxson and Sally Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226–244, 1995.

[3] M. E. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," in *Proc. ACM SIGMETRICS*, 1996, pp. 160–169.

[4] A. Feldman, A. A. Gilbert, and W. Willinger, "Data Networks as Cascades: Explaining the Mulifractal Nature of Internet WAN Traffic," in *Proc. ACM SIGCOMM*, 1998, pp. 42–55.

[5] Rudolf H. Riedi, Matthew S. Crouse, Vinay J. Ribeiro, and Richard G. Baraniuk, "A Multifractal Wavelet Model with Application to Network Traffic," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 992–1019, 1999.

[6] Sally Floyd and Vern Paxson, "Why We Don't Know How to Simulate the Internet," Tech. Rep., LBL Network Research Group, 1999.

[7] K. Park, G. Kim, and M. Crovella, "On the Relationship Between File Sizes, Transport Protocols, and Self-Similar Network Traffic," in *Proceedings of the IEEE International Conference on Network Protocols*, 1996.

[8] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 71–86, 1997.

[9] Ashok Erramilli, Onuttom Narayan, and Walter Willinger, "Experimental Queueing Analysis with Long-Range Dependent Packet Traffic," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 209–223, 1996.

[10] D. P. Heyman and T. V. Lakshman, "What Are the Implications of Long-Range Dependence for VBR-Video Traffic Engineering?," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 301–317, 1996.

[11] K. Park, G. Kim, and M. Crovella, "On the Effect of Traffic Self-similarity on Network Performance," in *Proc. SPIE Intl. Conf. Perf. and Control of Network Systems*, 1997.

[12] B. K. Ryu and A. Elwalid, "The Importance of Long-Range Dependence of VBR Traffic in ATM Traffic Engineering: Myths and Realities," in *Proc. ACM SIGCOMM*, 1996, pp. 3–14.

[13] D. R. Figueiredo, B. Liu, V. Misra, and D. Towsley, "On the Autocorrelation Structure of TCP Traffic," Tech. Rep., Computer Science Technical Report 00-55, University of Massachusetts, 2000.

[14] A. Veres and M. Boda, "The Chaotic Nature of TCP Congestion Control," in *Proc. INFOCOMM*, 2000.

[15] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "On the Nonstationarity of Internet Traffic," in *Proc. ACM SIGMETRICS 2001*, 2001, pp. 102–112.

[16] J.B. Gao and I. Rubin, "Multiplicative Multifractal Modeling of Long-Range-Dependent Network Traffic.," *Int. J. Comm. Systems*, vol. 14, pp. 783–201, 2001.

[17] Ilze Ziedins, "On the output process from a finite buffer with long range dependent input," Tech. Rep., Waikato University, 2000, http://wand.cs.waikato.ac.nz/wand/publications/.

[18] D. D. Botvich and N. G. Duffield, "Large Deviations, the Shape of the Loss Curve, and Economies of Scale in Larger Multiplexers," *Queueing Systems*, vol. 20, pp. 293–320, 1995.

[19] G. L. Choudury, D.M. Lucantoni, and W. Whitt, "Squeezing the Most Out of ATM," *IEEE Transactions on Communications*, vol. 44, no. 2, pp. 203–217, 1996.

[20] N. G. Duffield, "Economies of Scale in Queues with Sources Having Power-Law Large Deviation Scaling," *Queueing Systems*, vol. 33, pp. 840–857, 1996.

[21] K.R. Krishnan, "A New Class of Performance Results for a Fractional Brownian Traffic Model," *Queueing Systems*, vol. 22, pp. 277–285, 1996.

[22] I. Saniee, A. Neidhardt, O. Narayan, and A. Erramilli, "Performance Impacts of Multi-Scaling in Wide Area TCP/IP Traffic," *Journal of Communication Networks*, to appear.

[23] William S. Cleveland, Dong Lin, and Don X. Sun, "IP Packet Generation: Statistical Models for TCP Start Times Based on Connection-Rate Superposition," in *Proc. ACM SIGMETRICS 2000*, 2000, pp. 166–177.

[24] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "FSD and PackMime: Open-Loop and Closed-Loop Modeling of Internet Packet Traffic," in *Nonlinear Estimation and Classification*, C. Holmes, D. Denison, M. Hansen, B. Yu, and B. Mallick, Eds. Springer, New York, 2002, to appear.

[25] Jin Cao and Kavita Ramanan, "A Poisson Limit for the Unfinished Work of Superposed Point Processes," in *Proceedings INFOCOMM*, 2002, to appear.

[26] J.B. Gao and I. Rubin, " Superposition of Multiplicative Multifractal Traffic Streams," in *Proceedings ICC2000*, 2001.