Crossover Scaling Effects in Aggregated TCP Traffic with Congestion Losses*

Michael Liljenstam
Institute for Security Technology Studies
Dartmouth College
45 Lyme Rd., Suite 200
Hanover, NH 03755
mili@ists.dartmouth.edu

Andy T. Ogielski
Renesys Corporation
Hanover, NH 03755
ato@renesys.com

ABSTRACT

We critically examine the claims that TCP congestion control contributes to the observed self-similar traffic rate correlations. A simulation model is designed to analyze aggregated traffic of many TCP file transfers, with network topologies large enough so that each transfer has independent packet losses due to competition with other TCP traffic. To separate the effects of session-level variability from network-level variability we examine traffic consisting of small fixed-size files, and of heavy-tailed distribution of file sizes, with small variance of inter-session periods.

We find that, with increasing packet loss rate, traffic rate scaling crosses over from the regime dominated by file size distribution to another scaling regime that is independent of file sizes. That loss-dominated scaling stretches over the timescales from RTT to the longest consecutive TCP timeouts (hundreds of seconds), and is not asymptotic. Analysis at the flow level exposes the mechanism of the crossover, from scaling dominated by variability of the flow ON-periods to that dominated by variability of the OFF-periods.

However, it is unlikely that TCP timeouts contribute much to observed Internet traffic correlation structure, as they would matter only if widespread congestion losses exceeding 10% dominated the typical behavior of the Internet.

Keywords

Self-similarity, TCP

1. INTRODUCTION

Explanation of ubiquitous observations of packet traffic self-similarity on a single link [14, 20] in terms of high variability of data transfer sizes at the session layer has been well supported by network measurements (LAN traffic [30], WAN traffic [20], and WWW traffic [5]).

However, recently there have been claims that congestion losses in the TCP sessions constituting the largest fraction of the Internet traffic could be responsible for self-similarity [26], or for scaling effects that can be confused with self-similarity [9, 12] on the time scales reported in measurements.

Veres and Boda [26] claimed on the basis of simulations of a few TCP connections competing in a single link that heavy losses induce chaotic behavior and self-similarity in individual connections, and that this contributes to self-similar traffic scaling at a more fundamental level than session-layer data size distributions. Guo et al. [12] and Figueiredo et al. [9] followed up on this observation by analyzing an approximate mathematical model of an infinitely long TCP connection with an externally imposed Bernoulli packet loss process. These authors confirmed the calculations with simulations of a single connection using a detailed TCP Tahoe model. They correctly recognized that loss-induced correlations cannot be asymptotically self-similar because TCP timeouts are bounded. Nevertheless, Figueiredo et al. also claimed that the time range of what they called "pseudo-self-similar" scaling (cf. [15]) [1] can be comparable to the range of scaling observed in traffic measurements, and thus TCP loss effects can be a plausible explanation for observed self-similarity.

These are interesting claims, challenging an established explanation, and they deserve evaluation under more realistic conditions: It has to be noted that they are based on studies of a *single* connection subjected to a rather artificial loss process. First, in a network consisting only of a single link [26], the behaviors of TCP connections are very highly correlated with one another, and exhibit peculiar periodic phase effects [11, 13]. Second, the assumption [12, 9] of constant probability, independent packet loss is lacking an experimental justification (see [19]). In real networks the aggregate traffic on any link mixes packets from connections that may have independently suffered from losses somewhere else along their paths, queueing losses are bursty, and most connections are rather short.

In this paper we present the analysis of a simulation experiment that has been carefully designed to compare the evolution of TCP traffic rate scaling behavior with increas-

*This work has been partially supported by the Defense Advanced Research Projects Agency (DARPA), under grant N66001-00-8065 from the U.S. Department of Defense. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Defense.

[1] "Pseudo-self-similarity" is not well defined: a Hurst parameter estimate obtained over a limited segment of a variance-time or wavelet plot can be inconsistent with the stationarity condition [29]

ing congestion losses, at the two extreme session-layer traffic scenarios: one with small, constant-size file transfers, another with heavy-tailed, Pareto file size distributions. We designed a network model suited to study aggregate TCP traffic, where the following requirements are satisfied: First, each TCP connection in the aggregate is subject to losses independently of others, at some other network location, and the losses are caused by a different set of competing TCP connections rather than artificially imposed from the outside. Second, packet loss rate can be controlled by changing the number of competing TCP connections. Third, the effects of TCP timeouts on the scaling behavior can be validated in a control experiment by varying the maximum timeout duration in the TCP implementation.

We find that, with increasing packet loss rate, traffic rate scaling crosses over from the regime dominated by file size distribution to a scaling regime that is dominated by TCP timeouts, and is independent of file size distribution. That loss-dominated scaling stretches over the timescales from the round-trip time (RTT) to the longest sequence of 12 consecutive TCP timeouts before the connection is dropped. In the heavy loss regime the range of loss-dominated scaling is quite large, reaching times on the order of hundreds of seconds. We expose the mechanism of the crossover behavior by traffic analysis at the flow level, showing that the scaling crossover is due to the change of the dominant contribution to traffic variability, from variability of the flow ON-periods to variability of the inter-flow OFF-periods.

The quantitative analysis shows that it is unlikely that TCP timeouts in the Internet can contribute much to observed self-similar Internet traffic correlations: that would require widespread, long-lasting congestion losses exceeding 10%. This is not the usual operating regime of existing networks.

The paper is organized as follows: In the following section we describe the experiment design, that allows to compare non-variable to highly variable traffic processes under varying loss conditions. This is followed by the presentation of the traffic rate scaling under such conditions. The changing mechanism of scaling is explained in the sections on the analysis of IP flow statistics.

We have made use of some fundamental definitions and theorems from the following sources: *self-similarity* [14], *heavy-tailed distributions* [16], and an ON/OFF source superposition model for *Fractional Brownian Motion* [30]. We also use the following statistical tools: scaling analysis using *Variance-Time (VT) plots* [14], *Multi-Resolution Analysis (MRA) based on wavelets* [1], and the *Rescaled adjusted range (R/S) statistic* [2]. Finally, we detect heavy-tailed distributions using Log-Log plots of the *Complementary Cumulative Distribution Function (CCDF)* [5]. Details of the experimental design can be found in the Appendix.

2. EXPERIMENT DESIGN

Network topology: In order to study aggregate TCP traffic with uncorrelated losses we must design network models that are more complex than one-link "dumbbell" topologies commonly used in studies of TCP dynamics. A parsimonious design satifying the requirements of our experiments is shown in Figure 1.

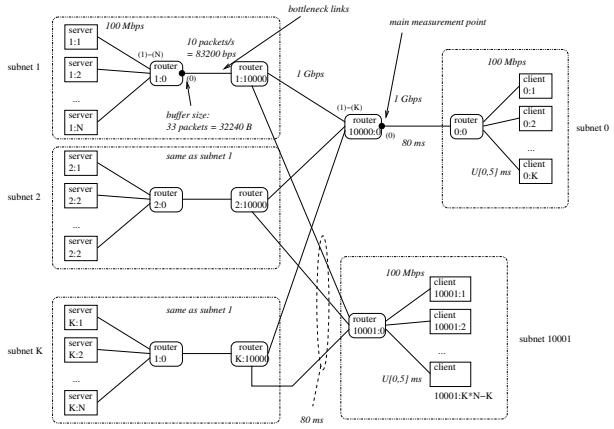

Figure 1: A family of network topologies used in simulation experiments described in this paper. K server networks, each with N server hosts, are shown on the left. The bottleneck links are between the two routers in the server networks. Two client networks are shown on the right. The client-server TCP connections are established so that in each server network exactly one server has one connection with one client in the upper right, while other servers' connections are with clients in the lower right. In this design, the TCP connections through the router labeled 10000:0 have independent losses.

The two building blocks are a server network and a client network. Each client host requests a file from one server host. We multiplex independent TCP connections on a link between the routers 10000:0 and 0:0 as follows[2]: For each of the K server networks, a single connection is established between one server and one host in client network 0 (upper right in the diagram), while the remaining $N-1$ servers have connections with clients in the network 10001 (lower right in the diagram). Thus, K connections always follow the path through the router 10000; and if the congestion losses are on the bottleneck links between two routers in each server network, these K connections will have independent losses.

The network is dimensioned so that for $N = 1$ a single TCP connection can flow without losses through a bottleneck link. The congestion loss rate can be systematically increased by increasing N: as more connections are added, there will be competition between them in the bottleneck link. We use the TCP Tahoe version because it was used in the prior work [26, 9, 12] that is addressed here. Details on the values of network parameters and the TCP parameters are listed in the Appendix A.

We measure packet statistics on the network interface (NIC) number 0 on router 1:0 to monitor the packet loss probability on the bottleneck links. The losses in the other server networks are similar due to symmetry. We measure the ag-

[2]We use the SSFNet's "Network, Host, Interface" topological naming convention for network elements. Thus a two-level hierarchy of networks, where network N_1 contains a subnetwork N_2 is written as $N_1 : N_2$; a host H in network N_2 is referred to as $N_1 : N_2 : H$; and a network interface I on host H is specified as $N_1 : N_2 : H(I)$.

gregate traffic at the NIC 0 on router 10000:0. At each measurement NIC we log packets arriving at the output queue *before* packet losses are counted.

The network models are constructed and simulated using the SSFNet simulator release 1.3 (source code, validation tests[3], and manuals are available at `http://www.ssfnet.org`).

Traffic models: Three traffic cases are studied. In each case every client repeatedly requests file transfers with a random, exponentially distributed wait time between the completion of one transfer and the start of the next one. Thus, the client-server traffic process is an alternating renewal ON/OFF-process [4] with exponentially distributed OFF-periods. The ON-period distributions are as follows:

repeated fixed size file transfers File sizes are fixed to 12000 bytes (12 packets).

heavy-tailed file sizes, modest tail weight File sizes are drawn from a Pareto distribution (same mean as for fixed size), and the shape parameter $\alpha = 1.8$ (close to 2 so not given much weight).

heavy-tailed file sizes, high tail weight File sizes are Pareto distributed with the same mean, and the shape parameter $\alpha = 1.2$.

With such an experiment design we can compare the contributions to aggregate traffic rate correlations due to network layer variability (congestion) and to session layer variability (file transfer renewal processes).

We note that in previous analytic studies [9, 12] of the loss-induced scaling behavior of a *single* TCP connection it was assumed that packet losses are independent, i.e., packet losses follow a Bernoulli process. However, we found that when packet losses are due to competition among multiple TCP connections in a congested queue, then for each connection the consecutive losses are definitely *not* independent. We performed the χ^2 tests for independence of two consecutive packet drop decisions (for a single connection) in the simulation experiments, and we found that the independence hypothesis is false at the 1% significance level, basically meaning 99% certainty in the conclusion.

Statistical significance: Mathematically speaking, self-similarity of a superposition of K ON/OFF alternating renewal processes is an asymptotic phenomenon, reached as first K and then the time scale approach infinity, in this order [23, 27], when either ON, or OFF, or both periods have a heavy-tailed distribution.

In practical experiments, this limit is gradually approached, with both K and the measurement time affecting the variability of the superposition and the range of time scales manifesting an approximate power law behavior. Since the corrections to scaling for finite K and finite times are not known in analytic form, in this work we have been gradually increasing K and the simulated time until sufficiently robust statistics were obtained. See Section 6 for more details.

To obtain sufficiently robust traffic variability in finite simulations with Pareto-distributed ON periods, we consider the interplay of two factors: the mean of the exponential OFF period distribution, and the number K of multiplexed TCP connections. It is a simple observation that if the mean is larger, one needs a larger K to achieve sufficiently high traffic rate variablity in a finite time interval.

In turn, degree of traffic variability influences the duration of simulated time and the number of independent experiments required for good statistics: statistics involving long-range dependent processes and heavy-tailed distributions converge very slowly with time and the number of samples [6, 24]. Therefore, both K and the length of simulated time must be large enough in order to achieve sufficiently small sampling errors in the wavelet, VT and R/S plots and to expose the scaling properties over several decades of time. Note that the required simulated time length depends on the RTT and bandwidth in the network model, because the number of sampled ON-periods increases with the number of TCP connections.

The remaining two concerns are: *i)* the initial start-up transient, and *ii)* that traffic load does not decrease in time due to dropped connections. We verified that the initial traffic transient is so short that it has no effect on the results. The simulated client implements error recovery: After connection drop due to 12 consecutive retransmission attempts or a timed-out connection attempt, the client requests a new file transfer after an exponentially-distributed inter-transfer wait time. Therefore, even with connection failures the traffic load is constant throughout the experiment.

In the data presented in this paper the values of K, exponential mean, and the simulated time range were chosen so that the crossover of scaling becomes clearly visible, and its mechanism can be unambiguously identified. For $K = 10$, and the OFF-period mean of 1.1 seconds the experiments running for 100,000 seconds (over 27 hours) of *simulated time* produce acceptable statistics for the time scales spanning three to four decades in time, from 1 to 10^3 or 10^4 seconds. The adequacy of our choice of the K parameter and other issues related to parameter sensitivity will be discussed in more detail in section 6.

3. SCALING ANALYSIS

We study the scaling behaviour of the packet rate process for each traffic scenario and loss rate, with the smallest packet count bin size of 100 ms. Figure 2 summarizes the results in terms of the wavelet energy plots and variance-time plots. All wavelet plots use the same Haar wavelet (D1) base to facilitate comparability of qualitative aspects. Other bases were tried but led to only small differences in estimates for reasonable fits. The R/S plots were also generated, but are not shown.

The Hurst parameter estimates $\hat{H}$ obtained using the three distinct methods have been summarized in Table 2. The table also shows the theoretical value of the Hurst parameter for a corresponding Fractional Brownian Motion (FBM)

[3]In particular, the SSFNet TCP implementation validation includes the ns-2 test suite [10], and testing with a version of the *tcpanaly* tool [18].

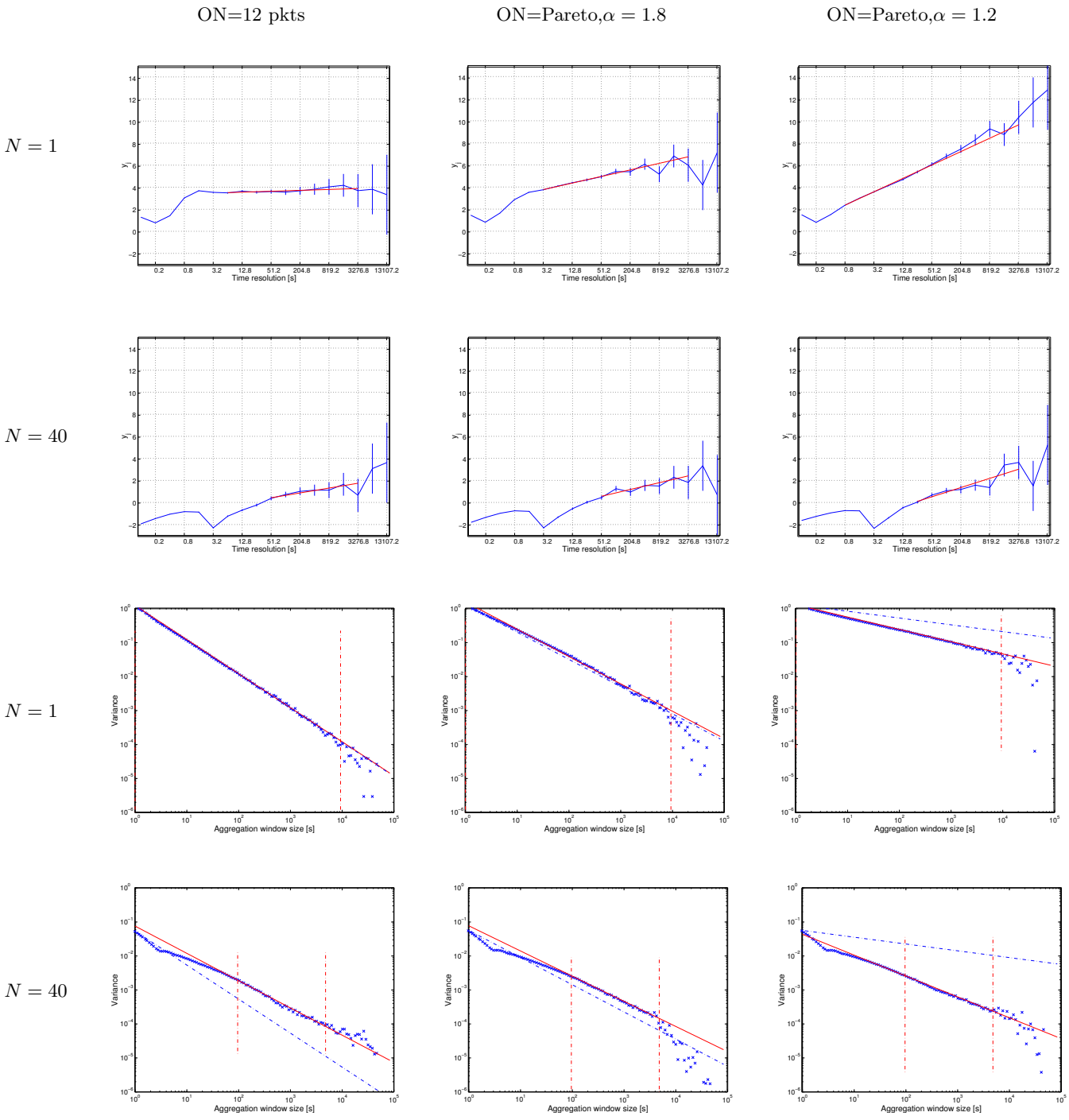

Figure 2: **Wavelet and variance-time (VT) scaling plots of the aggregate traffic rate ($K = 10$). Columns, from left to right: small fixed sized transfers (12 packets long), Pareto distributed file sizes with $\alpha = 1.8$, and Pareto with $\alpha = 1.2$. Rows labeled $N = 1$ correspond to aggregate traffic of K TCP connections without any packet losses, while rows labeled $N = 40$ correspond to aggregate traffic where each connection is suffering from packet losses exceeding 10% (see Table 1). Initial bin width is 100 ms. The slopes of the dashed lines in the VT plots follow the power law for the corresponding idealized FBM processes.**

Scenario		*Router 1:0(0) statistics*			*Router 10000:0(0)*		*Global statistics*		
File size	Server Hosts (N)	Packets	Dropped Packets	P[drop] estimate	Packets	Netflows	TCP connections attempted	failed TCP connection attempts	dropped TCP connections
fixed	1	523183	0	0	5234381	140081	348962	0	0
size	40	1477604	219944	0.149	313250	180040	808866	0	1
Pareto	1	539782	0	0	5405933	139885	349367	0	0
$\alpha = 1.8$	40	1527169	224393	0.147	324659	184304	814319	0	1
Pareto	1	525673	0	0	5247843	147961	369202	0	0
$\alpha = 1.2$	40	1562748	207911	0.133	328751	188043	937462	0	1

Table 1: Packet loss, netflow, and connection statistics for each traffic model.

process, where H is related to the shape of the heavy-tailed ON distribution by $H = (3 - \alpha)/2$.

For lossless traffic cases ($N = 1$) the wavelet plots in Figure 2 behave as expected: a flat plot for fixed-size file transfers, and increasing slopes for Pareto transfers with increasingly heavy tails. The dip at a fine scale (0.2 s for $N = 1$) is consistent with the presence of a periodic traffic component about the RTT value [8]. Note that the dip shifts to the right as traffic load increases ($N = 40$) as expected for an increase in RTT due to queuing delay: In this model, when packets encounter a full queue most of the time the RTT is about 3.3 seconds.

The Hurst parameter estimates for lossless traffic are close to theoretical values for the corresponding FBM processes. However, in the case of Pareto file size distribution with $\alpha = 1.2$ the estimate $\hat{H} \approx 0.8$ is smaller than the expected theoretical value $H = 0.9$. This negative bias can be attributed to undersampling of the heavy tail, even with simulation time of 10^5 seconds, as long range correlations are produced by rare instances of extremely large file transfers [24]. Nevertheless, this bias does not affect the analysis of the effects of TCP timeouts on the scaling properties.

Having established the baseline behavior of the aggregate traffic of lossless TCP connections, we proceed to analyze the effects of heavy packet losses. Recall that increasing packet losses are generated by increasing the number N of competing connections in the server network. Having verified the evolution of new scaling behavior with gradually increasing N, in this paper we only present the results for $N = 40$ which corresponds to packet loss rates of 13–14% (Table 1).

The most striking feature of the wavelet plots in Fig. 2 for an aggregate traffic of TCP connections with packet losses is that they become qualitatively similar for *both* fixed-size and Pareto distributions. Clearly, some loss-related mechanism affects the short–to–intermediate range scaling behavior, regardless of the file size distribution. Moreover, the same mechanism depresses long-range correlations for the heavy-tailed (especially $\alpha = 1.2$) Pareto distribution, as if the long file transfers became significantly rarer than in the lossless case. We also experimented with lower loads resulting in lower loss rates (results not shown here) but found that the scaling features found for fixed-sized transfers only become pronounced for high loss rates, typically above 10%.

Scenario		*Estimates*			*Th.*
File size	N	$\hat{H}$ (wavelet)	$\hat{H}$ (VT)	$\hat{H}$ (R/S)	H
fixed	1	0.52 [0.50,0.54]	0.50	0.55	0.5
size	40	0.61 [0.56,0.66]	0.61	0.60	
Pareto	1	0.65 [0.64,0.66]	0.60	0.62	0.6
$\alpha = 1.8$	40	0.65 [0.60,0.70]	0.63	0.64	
Pareto	1	0.80 [0.80,0.81]	0.82	0.78	0.9
$\alpha = 1.2$	40	0.71 [0.67,0.74]	0.70	0.85	

Table 2: Hurst parameter (H) estimates for the three distinct traffic processes, and theoretical values for the corresponding idealized Fractional Brownian Motion processes.

In the following sections we explain the structural mechanism of this phenomenon in terms of flow fragmentation by TCP timeouts.

4. TCP TIMEOUTS AND SCALING CROSSOVER

In this section we demonstrate that the range of TCP-induced scaling in the heavy packet loss regime is determined by the longest "gaps" in the TCP transmission, that are due to the accumulation of TCP timeouts during consecutive packet losses.

In order to determine how the maximum timeout duration relates to the range of loss-induced scaling we experiment with modifications of the timeout computation in the TCP implementation. Recall that in BSD-style TCP implementations the timeout value is calculated as $RTO \cdot 2^k$, where $0 \leq k \leq 6$, and then is restricted to be between 1 and 64 seconds. Therefore, before a connection is dropped due to 12 consecutive packet losses, the longest transmission gap is at least 383 seconds. Note that with Karn's algorithm, two retransmissions are considered consecutive if there was no RTT calculation between them. In other words, there are two consecutive retransmissions if a segment was retransmitted, and either the same segment is retransmitted again, or the next segment is retransmitted. Thus a case of, say, 11 consecutive retransmissions may occur in multiple ways, e.g., when the same segment is retransmitted 11 times, or when 11 consecutive segments are sent and each of them is retransmitted once, and so on.

It is desirable to reduce the statistical errors in the scal-

	$\hat{H}$		
	orig t.o. factor=64	max t.o. factor=8	max t.o. factor=2
Wavelet plot	0.61 [0.56,0.66]	0.70 [0.68,0.72]	0.53 [0.48,0.58]
VT plot	0.61	0.54	0.51
R/S plot	0.60	0.67	0.62

Table 3: Hurst parameter estimates $\hat{H}$ for TCP with reduced timeouts, $N = 40$.

ing plots at the scale corresponding to the estimated location of the crossover from loss-induced to asymptotic scaling. Therefore, we designed simulation experiments where the maximum value of the timeout factor was reduced from $2^6 = 64$ to 8 and 2, respectively. Figure 3 shows the scaling plots obtained with these modifications, and the unmodified scaling plot for comparison, for the repeated fixed-size file transfer traffic scenario. Table 3 gives the Hurst parameter estimates.

For a maximum timeout factor of 8 the VT plot shows a broad "knee" at the scale of about 100 seconds, and for a maximum timeout factor of 2, the location of the "knee" is at the scale of about 30 seconds. The "knee" in a VT plot demonstrates the crossover from short-time to asymptotic scaling regime. The corresponding feature in the wavelet plots is the onset of the asymptotic zero-slope regime. Thus, data indicates that the range of TCP timeouts is indeed responsible for the observed range of loss-induced scaling. For packet loss rate over 10% the crossover region is approximately at the time scale corresponding to the maximum duration of accumulated consecutive timeouts, which for standard TCP implementations and typical RTT values would be about 400 seconds.

5. FLOW STATISTICS AND MECHANISM OF CROSSOVER

A network *flow* on a link is defined as a sequence of IP packets traveling from a given source to a given destination, where arrivals of successive packets are separated by the time interval smaller than a predefined *flow threshold* Δt_f. In the model discussed here the source and destination are uniquely identified by the IP addresses of the corresponding hosts. Note that a flow is intrinsically defined relative to the time scale Δt_f, and the characterization of flow-level behavior needs to be reasonably robust with respect to the choice of Δt_f.

In this section we discuss the distributions of flow durations (ON intervals) and of inter-flow gaps (OFF intervals) for the three traffic scenarios, both without and with packet losses. As in previous sections, we consider side by side the cases of aggregate traffic without packet losses ($N = 1$), and with heavy packet losses ($N = 40$). Figure 4 shows the log-log plots of the complementary cumulative distribution functions (CCDF) for the ON-intervals.

Without packet losses, and at longer time scales, we see that for each traffic scenario the distribution of the ON intervals corresponds to the distribution of file sizes. There is a sharp cutoff on flow lengths in the case of fixed-size file transfers, and the heavy power-law tail for the case of Pareto-distributed file sizes, with the same value of the shape parameter α.[4]

With heavy packet losses ($N = 40$), however, the flow length distributions are significantly different, with much shorter flows dominating even for Pareto-distributed file sizes. In presence of packet losses and TCP timeouts the long file transfers are broken into many shorter flows, whose distribution depends on the value of flow threshold Δt_f.

The corresponding log-log plots of the complementary cumulative distribution functions (CCDF) for the OFF-intervals are shown in Figure 5. With no packet losses, and Δt_f greater than RTT, the OFF-interval distribution is generated exclusively by the exponential inter-flow distribution (mean 1.1 seconds). With heavy packet losses, however, at the longer time scales the inter-flow OFF-intervals are generated by TCP timeouts. Notice the presence of a substantial tail of the distribution reaching times in excess of 300 seconds, but given that there is an upper bound on the duration of the longest cumulative TCP timeout this tail does not stretch much further. Moreover, this tail is not well approximated by a (truncated) power law.

Asymptotic robustness of the ON- and OFF-interval distributions with respect to the choice of the flow threshold parameter Δt_f is a subtle matter. An elegant analysis of the Internet data presented in [30] illustrated the robustness of the measured tail behavior of ON- and OFF-interval distributions for high variability traffic by insensitivity to the choice of the flow threshold Δt_f as long as it is large enough. Note, however, that such analysis requires that at large time scales Δt_f is still smaller than the tail values of the inter-flow OFF-intervals, for otherwise for any finite duration dataset one would always end up with a single long flow when Δt_f is large enough.

Therefore, it is easy to explain why the flow size distributions under heavy loss condition shown in Figure 4 do not manifest any Pareto tail behavior when the file size transfers are Pareto distributed. Mathematically speaking, they should remain heavy-tailed since it's an asymptotic property, while the retransmission timeouts are bounded. Thus, in theory we should increase Δt_f above the range of the largest TCP timeouts, (approximately 400 seconds) to observe the robust behavior of the tail distribution of the ON-intervals.

Alas, what theory recommends is impossible in practice if the OFF-interval distribution is not heavy-tailed, and its mean is smaller than 400 s. In this paper the tradeoff for generating sufficiently high variability of aggregate traffic is that the mean OFF-interval must be rather small (1.1 s, but even if it were an order of magnitude larger the argument below still applies). Thus, if the inter-session OFF-intervals are random but not heavy-tailed, e.g., exponen-

[4]The maximum TCP window size in the model network is smaller than the end-to-end bandwidth-delay product. But since the flow measurement is carried out after the bottleneck link and *before* the large bandwidth link, after the slow start the packets will be arriving in a back-to-back fashion.

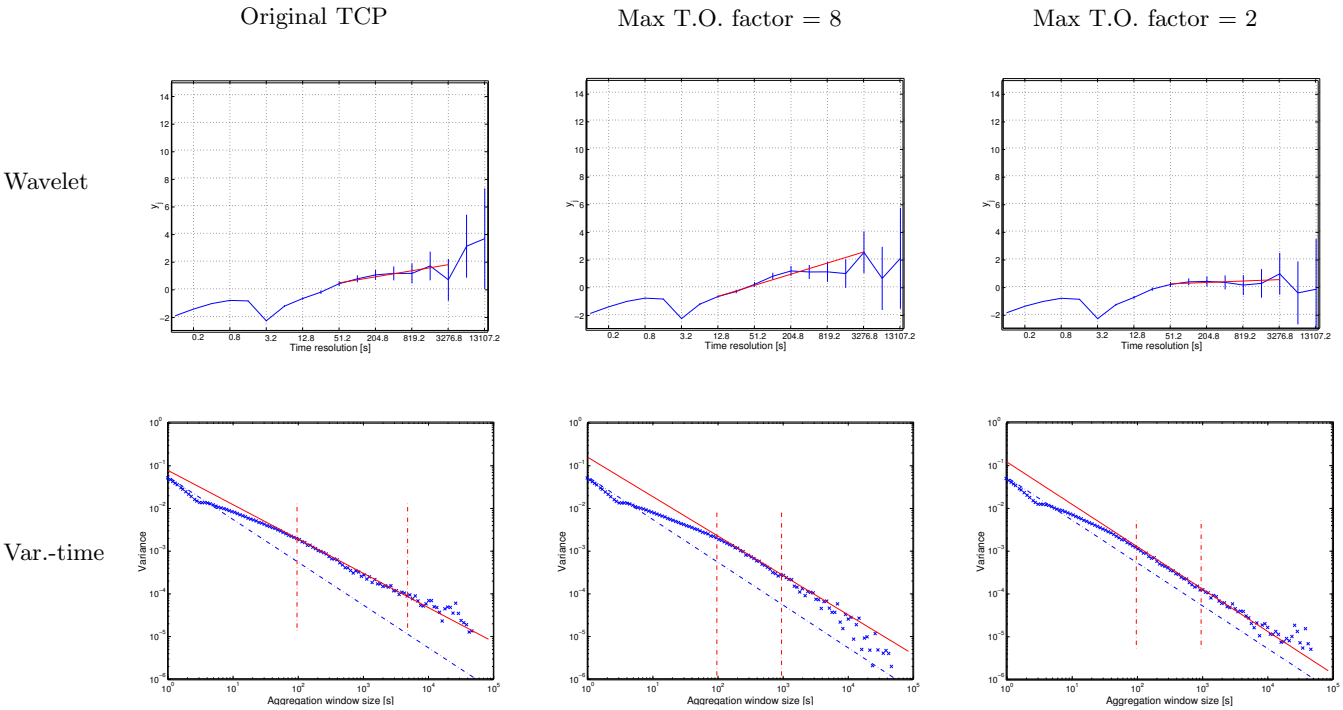

Figure 3: Scaling analysis with reduced maximum TCP timeouts, left to right: maximum timeout factor 64, 8, and 2; for TCP traffic of fixed-size (12 packets) data transfers, at heavy loss conditions ($N = 40$). Dashed lines in VT plots correspond to the slope with $H = 0.5$, i.e. short-range dependence. Note the reduction of the time range of loss-induced scaling with the reduction of the maximum timeout factor.

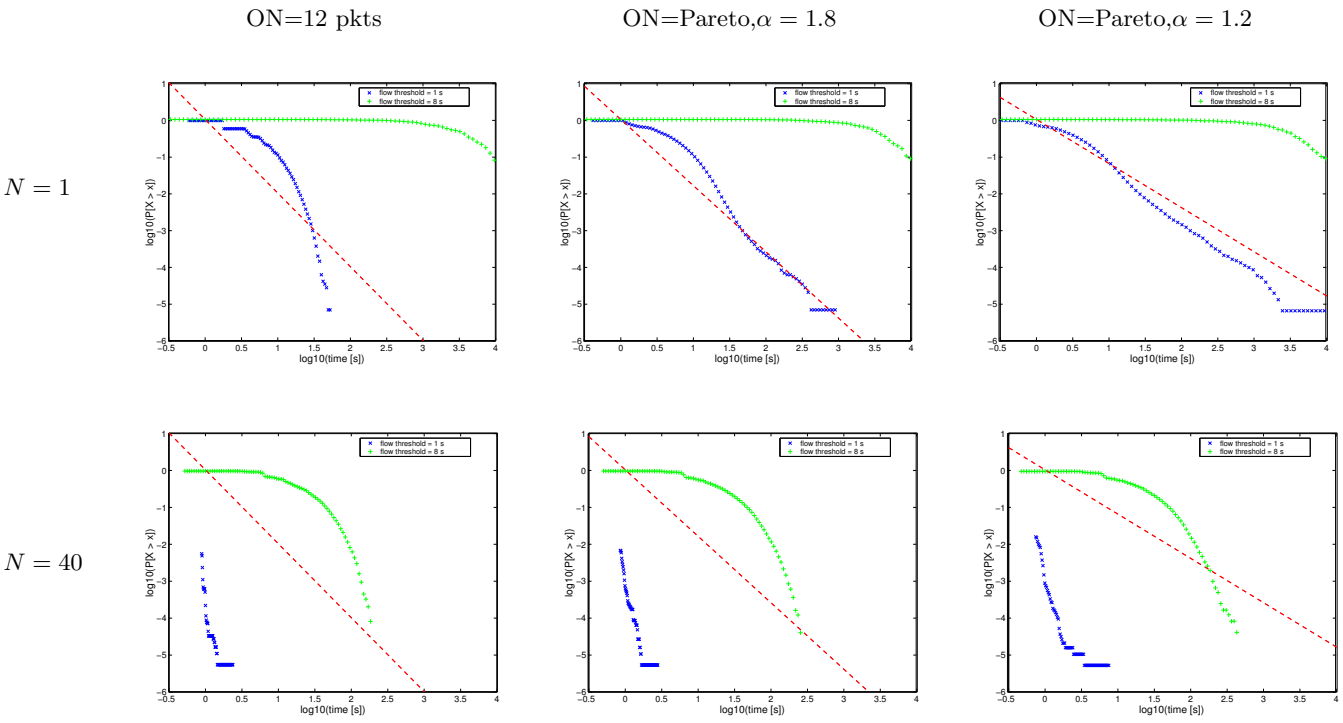

Figure 4: Complementary cumulative distribution function (CCDF) of flow ON-intervals for the three traffic scenarios ($K = 10$) and two different flow thresholds. Crosses show distribution for $\Delta t_f = 1$ s, and plus signs for $\Delta t_f = 8$ s. Columns from left to right: small fixed sized transfers (12 packets long), Pareto distributed file sizes with $\alpha = 1.8$, and Pareto with $\alpha = 1.2$. Top row: lossless traffic ($N = 1$), bottom: heavy packet losses ($N = 40$).

$N = 1$

$N = 40$

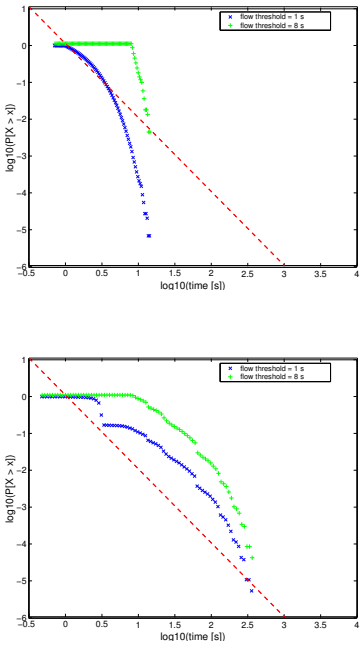

Figure 5: Complementary cumulative distribution function (CCDF) of flow OFF-intervals for the traffic scenario of small fixed-size transfers. The other two traffic scenarios generate identical plots, as the CCDFs are determined only by the packet losses and not by file sizes. Top: lossless traffic, bottom: heavy traffic losses.

tially distributed with mean $1/\lambda = 1.1$ s, then the probability of encountering an OFF-interval larger than 400 s is $P[X > 400] = e^{-400\lambda} \approx 1.2 \cdot 10^{-158}$. Even though theoretically it is true that the flows are heavy-tailed, for all practical time-scales we will never experimentally observe any tail.

In summary, in cases when inter-session OFF intervals have small mean and little variability, and TCP session sizes have large variability, the high packet losses will result in strongly reduced variability of the flow ON-intervals for the practical range of Δt_f. But this effect is counter-acted by the traffic correlations induced by a broad—but not heavy tailed—TCP timeout distribution over practically observable time scales. Now it is easy to see the reason why experiments appear to show self-similarity of TCP traffic under heavy loss conditions.

Since it is not feasible to beat the statistics by executing infinitely long simulations, we will expose this effect in a control experiment where the variability of TCP timeouts has been artificially eliminated.

5.1 Removal of TCP Timeout Variability

In order to separate the two effects of TCP timeouts: *i)* reduced variability of flow sizes (ON-intervals), and *ii)* increased variability of OFF-interals, we again experiment with a modification to the TCP protocol. In this experiment we use the Pareto-distributed file sizes ($\alpha = 1.2$), and reduce the variability in the OFF-intervals by *i)* setting all TCP timeouts to 4 seconds, and *ii)* setting the inter-session

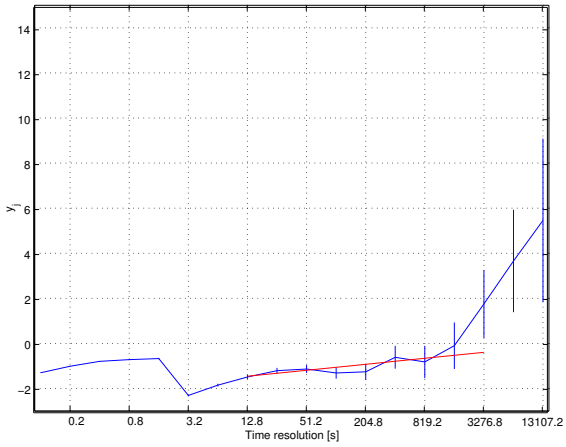

Figure 6: Wavelet plot for the Pareto file size distribution traffic scenario ($\alpha = 1.2$) with heavy packet losses, but with a modified TCP with the constant timeout length fixed at 4 seconds. The onset of asymptotic scaling is at the scale of longest consecutive timeouts; compare with Figure 2.

OFF-intervals constant, to 1.1 seconds. The resulting OFF-interval distribution for high packet losses ($N = 40$) shows that the maximum observed OFF-interval length has been reduced to less than 40 seconds. The reason that we observe OFF-intervals longer than 4 seconds is due to consecutive losses and timeouts.

The effect this modification has on the scaling is most noticeable in the wavelet scaling plot, shown in figure 6. High losses lead to a complete destruction of Pareto behavior for short to medium time scales, and the asymptotic Pareto scaling sets in only at time-scales above approximately 1000 seconds.

In a previous work [16] it was asserted that the TCP transport plays an important role in preserving self-similarity induced by session-level variability under packet loss conditions. Using simulations, these authors compared reliable and unreliable transport (TCP vs. UDP) and found that packet losses can erode self-similarity in UDP traffic, but concluded that reliable TCP transport preserves the long range dependence. However, we have shown here that TCP's effect on scaling is more subtle and caution is required in the analysis of the resulting scaling: loss-induced scaling can stretch to time scales on the order of 10^2 seconds, and even if scaling plots look deceptively like a power law in this range, the exponent is, in general, different from the asymptotic behavior. The reliable TCP transport does not simply preserve self-similarity under heavy packet losses, but leads to two *separate* scaling regimes with the crossover at the timescale of the longest timeouts.

6. STATISTICAL ERRORS AND PARAMETER SENSITIVITY

The network model parameters used in the experiments were chosen so that the simulations could unambiguously reveal the scaling phenomena, while remaining computationally practical for very long time scales that we studied. However, the experiments constitute only a few points in a vast parameter space, and we need to address the robustness of our conclusions with respect to changes in the network parameters.

The first set of questions concerns the convergence of statistics in the ON/OFF traffic superposition with heavy-tailed distributions: *i*) is $K = 10$, i.e. a superposition of ten traffic ON/OFF processes sufficient to approximate a self-similar aggregate process, and *ii*) how significant is the undersampling bias in the finite samples drawn from the heavy-tailed file size distributions [6, 24].

Convergence of the superposition model: We examined the effect of finite K on the Hurst parameter estimates by directly simulating the superposition of K ON/OFF alternating renewal processes (FBM model), with K increasing from 10 to $5 \cdot 10^3$, for the simulated time of 10^6 seconds.[5] ON/OFF-period distributions were chosen as in the network simulations: OFF-periods exponential with mean 1.1 s, and ON-periods Pareto with mean 1.8 s and $\alpha = 1.2$. We have observed that in each of these experiments the VT plot shows virtually no deviation from a pure power law for aggregation windows ranging from 1 to $5 \cdot 10^3$ seconds. However, the exponent estimate is a *random variable*. Figure 7 shows the derived Hurst parameter estimates together with the error bars.

As $K \to \infty$, the limit theorem predicts that the process converges to Fractional Gaussian Noise (FGN) with $H = 0.9$. Both VT and wavelet estimates show good agreement about $H = 0.85$ which is a little lower than the theoretical value. In the case of the VT estimator this could be explained by the negative estimator bias [22], for FGN with $H = 0.9$ of approximately -0.05. We attribute the scatter of the Hurst parameter estimates to undersampling: even for $K = 10,000$ and 10^6 seconds the extremely long ON-periods that produce long-range correlations are still not adequately sampled.

Given the stability of wavelet-based statistics, we conclude that $K = 10$ is sufficient to produce a *qualitatively* correct approximation to the self-similar aggregate traffic in this case. Our estimates from the FBM model should also be compared with the estimates obtained from Figure 2 (Table 2) for the lossless case with $\alpha = 1.2$. By changing the wavelet base (not shown) we get $H = 0.85$, i.e. a good match with the FBM model.

Timescales repesented in the data: We again consider simulated time-series of length $T = 10^6$ s, with K ON/OFF-processes. As in the previous paragraph let the mean ON-period $t_{ON} = 1.8$ s, the mean OFF-period $t_{OFF} = 1.1$ s, and let $t_{ON/OFF} = t_{ON} + t_{OFF} = 2.9$ s. The expected number of transfers for one client/server pair (C/S pair) would be $n = E[\#\text{transfers}] = \frac{T}{t_{ON/OFF}} \approx 3.4 \cdot 10^4$. With K simultaneous C/S pairs and no contention we get $n \cdot K$ transfers, which is close to numbers in Table 1, so n is a fairly good estimate of the number of Pareto samples drawn for each process in the lossless case. As stated in [6], the expected largest sample observed $E[Y]$ from n Pareto samples with mean $E[X]$, is $E[Y] \approx E[X]n^{1/\alpha}$, giving us here $E[Y] \approx 7.2 \cdot 10^3$ s. [24] also recommend examining the sampled data to keep track of the largest samples collected. We can do this by studying the rightmost top-row graph of figure 4. This indicates reasonable samples up to approximately $10^{3.5} \approx 3.2 \cdot 10^3$ with some samples up to 10^4. Thus, starting from 1 s timescale, we could expect to have useful data somewhere between three and four decades, which agrees with the VT plots in figure 2.

[5] An initial warmup period of 10^6 seconds was also simulated and discarded. This was important for stationarity in the case of $K = 10^4$, and thus for VT estimates, but insignificant for other cases.

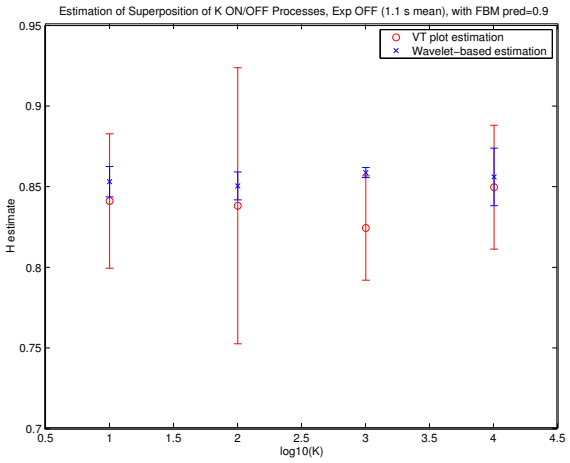

Figure 7: Estimates of the Hurst parameter H obtained in simulations of FBM by superposition of K alternating ON-OFF renewal processes, with Pareto-distributed ON periods and exponentially-distributed OFF periods. Circles: average H using VT plot estimation, Δt_fcrosses: average H using wavelet estimation method. Each data point represents the average of H estimates obtained from 5 independent time series of length 10^6 s with 90% confidence intervals. The VT plot used aggregation window sizes from 1 to $5 \cdot 10^3$ s.

For high loss, the number of transfers per C/S pair is drastically reduced, to about $2.8 \cdot 10^3$ (from table 1), leading to a largest expected sample of only $8.9 \cdot 10^2$ s. (Here we cannot rely on figure 4.) However, the VT plots show reasonable statistics to somewhat larger scales. This could be simply and artifact of the VT plot, but a possible explanation is that there are now many competing C/S pairs (total 390) that also draw Pareto samples, and could induce more long-lasting correlations through the competition in the bottleneck link (by suppressing other flows for long times).

Parameter sensitivity: Whereas the previous papers have focused on the loss probability as the sole factor determining the traffic rate correlation structure imposed by TCP, our experiments indicate that the situation is more complex. We examined the sensitivity to the following model parameters: *i)* increasing the bottleneck buffer size to 100

and 200 packets, and *ii)* increasing the inter-transfer OFF-period mean by a factor of 10 and 100. In both cases the load factor N was increased to maintain a comparable loss probability, and in both cases a break-down of asymptotic scaling became more evident. The latter is attributed to increased undersampling of the process due to the sparsening of the traffic, as each of these changes increases the mean of the flow OFF-periods.

By increasing these parameters, the distance between the smallest and the largest OFF-intervals (induced by time-out) tends to decrease, which reduces the range of the loss-induced scaling. In the case of the buffer size there is a connection between the buffer size and the inter-transfer times and TCP timeouts. For larger buffers and constant mean inter-transfer time, the buffer will not have time to empty between transfers or short time-outs. Thus, the flows become "compressed" and the OFF-periods shrink below the flow separation threshold.

7. CONCLUSIONS

In this report we have used a simulation model to illustrate and compare traffic rate correlations induced by the TCP protocol with those induced by high-variability at the session level.

We found that under heavy packet loss conditions the TCP timeouts induce strong traffic rate correlations stretching far into medium range time scales. Over their range, these correlations decay approximately like a power law with an exponent that is only weakly dependent on the session-level file size distribution. Under certain traffic scenarios there is a crossover in scaling from the time range dominated by TCP timeouts to the asymptotic time range dominated by session size distribution. We studied packet flow distributions for a structural explanation of our findings and observed *i)* the OFF-period distribution develop a noticeable tail as losses induce longer and longer timeouts, and *ii)* we were unable to detect heavy-tailed ON-periods as losses increased. By experimenting with a modification to the TCP mechanism where the timeouts are limited or fixed, we *i)* experimentally link the flow OFF-period distribution to the TCP induced scaling (in the absense of a mathematical link) and *ii)* couple the broken ON-periods to an "erosion" of self-similarity over medium-range time-scales.

We conclude that TCP in combination with extreme losses does not induce self-similarity in the traffic. Instead, heavy losses may lead to a form of "erosion" of self-similarity up to medium-range time-scales. This "erosion" is masked by medium-range scaling effects due to TCP timeouts.

Therefore, it is highly improbable that TCP contributes much to scaling measured in the Internet traffic, as *i)* the effect is sensitive to parameters such as source-level OFF-period distribution and buffer sizes, and *ii)* it would happen only if the normal day-to-day operating condition of the Internet manifested steady, widespread and extreme congestion losses exceeding 10% over many hours.

Acknowledgments

The authors thank Walter Willinger, AT&T Labs, for a number of illuminating discussions on the finer points of scaling analysis.

8. REFERENCES

[1] P. Abry and D. Veitch, "Wavelet Analysis of Long-Range-Dependent Traffic", *IEEE Transactions on Information Theory*, vol. 44, No. 1, Jan. 1998.

[2] J. Beran, "Statistics for Long-Memory Processes", Chapman & Hall, New York, NY, 1994

[3] L. Brakmo and L. Peterson, "Experiences with Network Simulation", *in Proceedings of the ACM SIGMETRICS conference on Measurement & modeling of computer systems*, pp. 80-90, Philadelphia, PA, May 1996.

[4] D. R. Cox, *Renewal Theory*, Chapman and Hall, New York, 1982.

[5] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes", *IEEE/ACM Transactions on Networking*, vol. 6, pp. 835-846, Dec. 1997.

[6] M. Crovella and L. Lipsky, "Long-Lasting Transient Conditions in Simulations with Heavy-Tailed Workloads", *in Proceedings of the 1997 Winter Simulation Conference*, pp. 1005–1022, Atlanta, GA, Dec. 1997.

[7] A. Erramilli, Onuttom Narayan, and Walter Willinger, "Experimental Queueing Analysis with Long-Range Dependent Packet Traffic", *IEEE/ACM Transactions on Networking*, 4(2):209-223, April 1996.

[8] A. Feldmann, A. Gilbert, P. Huang, and W. Willinger, "Dynamics of IP traffic: A study of the role of variability and the impact of control", *in Proceedings of the ACM/SIGCOMM'99*, Cambridge, MA, Aug. 1999.

[9] D. Figueiredo, B. Liu, V. Misra, and D. Towsley, "On the Autocorrelation Structure of TCP Traffic", *Tech. Report 00-55*, Dept of Computer Science, University of Massachusetts, Amherst, MA, Nov. 2000.

[10] S. Floyd, "Simulator tests", *http://www-nrg.ee.lbl.gov/nrg-papers.html*, July 1995.

[11] S. Floyd and V. Jacobson, "On Traffic Phase Effects in Packet-Switched Gateways" *Computer Communication Review (ACM)*, Vol. 21, No. 2, April 1991.

[12] L. Guo, M. Crovella, and I. Matta, "How does TCP generate Pseudo-self-similarity?", in *Proceedings of the Ninth International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunications Systems*, pp. 215-223, Cincinnati, OH, Aug. 2001

[13] Y. Joo, V. Ribeiro, A. Feldmann, A. Gilbert, and W. Willinger, "TCP/IP Traffic Dynamics and Network Performance: A Lesson in Workload Modeling, Flow Control, and Trace-Driven Simulations", *SIGCOMM Computer Communication Review (ACM)*, Vol. 31, No. 2, April 2001.

[14] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic" (Extended Version), *IEEE/ACM Transactions on Networking*, Vol. 2, No. 1, Feb. 1994.

[15] S. Manthorpe, I. Norros, and J. Y. L. Boudec, "The Second-Order Characteristics of TCP", *in Proc. of Performance'96*, Presented in the Self-Similarity hot-topic session, Lausanne, Oct. 1996.

[16] K. Park, G. Kim, and M. Crovella, "On the Relationship Between File Sizes, Transport Protocols, and Self-similar Network Traffic", *in Proc. IEEE International Conference on Network Protocols*, pp. 171–180, 1996.

[17] K. Park and W. Willinger, "Self-Similar Network Traffic: An Overview", in *Self-Similar Network Traffic and Performance Evaluation*, Edited by K. Park and W. Willinger, Wiley – Interscience, New York, NY, pp. 1-38, 2000

[18] V. Paxson, "Automated Packet Trace Analysis of TCP Implementations", *in Proceedings SIGCOMM Symposium (ACM)*, pp. 167-179, Cannes, France, Sept. 1997.

[19] V. Paxson, "End-to-End Internet Packet Dynamics", *IEEE/ACM Transactions on Networking*, pp. 277-292, Vol. 7, No. 3, June 1999.

[20] V. Paxson and S. Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, Vol. 3, pp. 226-244, June 1995.

[21] V. Paxson and S. Floyd, "Why we don't know how to simulate the Internet", *in Proceedings of the 1997 Winter Simulation Conference*, Atlanta, GA, Dec. 1997.

[22] M. Taqqu and V. Teverovsky, "On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series", *in A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Adler, Feldman, Taqqu, Editors, Birkhauser, Boston, 1998.

[23] M. S. Taqqu, W. Willinger and R. Sherman, "Proof of a Fundamental Result in Self-Similar Traffic Modeling", *Computer Communication Review (ACM)*, vol. 25, pp. 5–23, 1997.

[24] M. Roughan, J. Yates and D. Veitch, "The mystery of the Missing Scales: Pitfalls in the Use of Fractal Renewal Processes to Simulate LRD Processes", *ASA-IMA Conference on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics, Washington, DC, June 1999.*.

[25] D. Veitch, `http://www.emulab.ee.mu.oz.au/~ darryl/secondorder_code.html`, June 2001.

[26] A. Veres and M. Boda, "The Chaotic Nature of TCP Congestion Control", *in Proc. IEEE INFOCOM'2000*, (Tel Aviv, Israel), Apr. 2000.

[27] W. Willinger, V. Paxson and M. S. Taqqu, "Self-Similarity and Heavy Tails: Structural Modeling of Network Traffic" *in A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Adler, Feldman, Taqqu, Editors, Birkhauser, Boston, 1998.

[28] W. Willinger, DARPA Network Modeling and Simulation Principal Investigators Meeting, La Jolla, CA, March 2001.

[29] W. Willinger, personal communication, June 2001.

[30] 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*, 5:71-86, 1997.

APPENDIX

A. EXPERIMENT PARAMETERS

A.1 Network Dimensioning

The TCP loss-induced traffic correlations will only show up in high packet loss regimes. It is well known that it is difficult to provoke TCP into high loss rate regimes since it will back off when congestion is detected. In order to be able to place a high load on the bottleneck links (meaning many TCP connections), simulate multiple bottlenecks, and still keep the simulations resource demands feasible, we set the bottleneck bandwidth quite small, 10 packets per second. Other link capacities were set high so that they would not interfere with the traffic (100 Mbps for individual connections or 1 Gbps for aggregate loads). A link delay is set to 80 ms for the measurement link and for all deflection links going to alternate clients. All other links have zero delay. The bandwidth-delay product is large: 10 million Bytes, or about 10,000 packets. The RTT is about 270 ms. The bottleneck link buffer size was set to 33 packets, which is one larger than the TCP window size, to ensure that a single connection will not experience any packet losses. Thus, losses are only induced due to competition among TCP connections. Bottleneck buffers use a droptail overflow policy. Other buffers are infinite. All server-subnetworks are equivalent. The links connecting the client hosts to the router in each of the client-subnetworks are given a random delay, uniform distribution on the interval $[0, 5]$ ms, at network initialization time.

A.2 TCP Parameters

Table 4 lists the TCP parameters used in the experiments. The model of the TCP protocol used is the Tahoe version (no fast recovery), with delayed ack option switched off for compatibility with prior studies (see Section 1).

Table 4: TCP parameters

Parameter	Value
Max Segment Size *MSS* (bytes)	1000
Receive Window Size (MSS)	32
Send Window Size (MSS)	32
Send Buffer Size (MSS)	128
Max Number of Retransmissions	12
Slow timer granularity (seconds)	0.5
Fast timer granularity (seconds)	0.2
Max Segment Lifetime (seconds)	60.0
Max connection idle time (seconds)	600.0
Delayed ACKs	no
TCP version	Tahoe

Table 5: File transfer traffic parameters

Parameter	Value
start time (seconds)	10.0
start window (seconds)	1.0
request size (bytes)	4

A.3 Traffic Startup, Routing, and Probes

The simulated model employs a static version of the OSPF routing protocol to establish routes. Enough time is left between the simulation start and the first TCP file transfers to allow OSPF to converge. In each client host the first transfer request time is randomized (uniform distribution) to occur within a delayed start window (see Table 5) in order to avoid artifacts due to synchronized connection initiations.