Towards Modeling Effect of Packet-level TCP Dynamics on Internet Operations and Management

V. Marbukh
National Institute of Standards and Technology
100 Bureau Drive, Stop 8920
Gaithersburg, MD 20899-8920
E-mail: marbukh@nist.gov

***Abstract*—This paper reports on work in progress on modeling of the effect of packet-level TCP congestion avoidance dynamics on flow-level Internet performance and management including buffer sizing under arriving/departing flows. The proposed model accounts for the packet-level TCP burstiness due to congestion avoidance dynamics by incorporating the entire effective bandwidth of TCP-controlled traffic as opposed to the conventional model which accounts for the average TCP rate only. In particular, we demonstrate that conventional Markov model of Internet performance is applicable only in a case of sufficiently heavy average load corresponding to long flow-level queues and file transfer times. In a case of lighter average load the conventional Markov performance model significantly underestimates flow-level queues and file transfer times, and surprisingly a simpler mean-field performance model may be more accurate. While increasing buffer sizes causes proportional increase in the worst-case packet-level delays, the positive effect on the average capacity utilization and flow-level performance may be muted due to increase in the TCP burstiness. These conclusions may explain and quantify observed self-similarity of TCP traffic, and have important implications for the Internet design, operations, and management.**

Keywords-packet-level TCP dynamics, flow-level Internet performance, buffer sizing.

I. INTRODUCTION AND MOTIVATION

Flow-level models of TCP performance trade accuracy of packet level models for tractability. While packet-level models are inherently discrete since they are based on modeling TCP dynamics at the packet level, flow-level models are inherently continuous since they are formulated in terms of source rates or throughputs obtained by averaging over a large number of packets. Given a fixed set of carried flows, conventional TCP flow models assume that as time progresses flow rates reach equilibrium with constant flow rates, which depend on the TCP version [1]. Markov model of Internet performance [2] is based on assumption of time scale separation: TCP reaches equilibrium bandwidth allocation for a given set of carried flows much faster than set of carried flows changes due to flow arrivals/departures. In addition to this assumption, conventional Markov model [2] assumes constant instantaneous TCP flow rates by disregarding the inherent packet-level TCP burstiness resulted from congestion avoidance dynamics.

Figure 1 from [3], which plots the average number of file transfer TCP flows in progress on a single link vs. average link utilization under assumption that files of exponential size arrive according to a Poisson process, clearly demonstrates that this conventional Markov model [2] (red curve) significantly underestimates the average number of flows in progress obtained from ns2 simulations (blue curve).

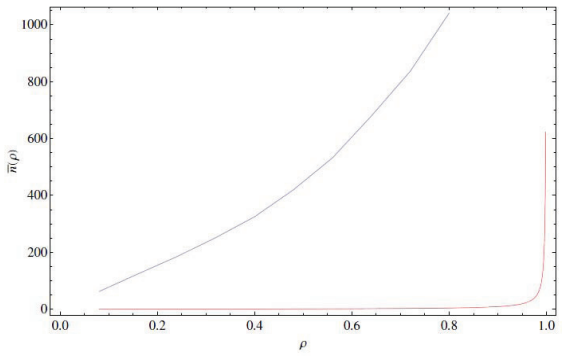

Figure 1. Average numbers of flows in progress

Paper [3] suggested that this discrepancy results from disregarding TCP burstiness by the conventional Markov model.

Recent paper [4] used a combination of simulations and approximations to demonstrate that packet-level TCP burstiness significantly impacts how router buffer sizing affects the capacity utilization and flow-level performance in the Internet under arriving/departing flows. In particular, the squared coefficient of variance, characterizing the relative burstiness of the aggregate TCP traffic traversing a link through a buffer of size B, is an increasing and concave function of B, and the buffer overflow probability is a power function of B: $p \sim B^{-\gamma}$ for some $\gamma > 0$. These traits indicate TCP traffic self-similarity and long-range dependencies since for "regular" traffic buffer overflow probability decays exponentially in buffer size [5]-[7].

Nature of TCP-controlled traffic has important implications for the Internet design, operations, and management, affecting TCP stability, trade-off between packet and flow-level

978-1-4673-0269-2/12/$31.00 ©2012 IEEE

performances, etc. [7]. The challenge is developing a performance model which combines consistency with empirical observations [3], [6], applicability to various TCP versions, and tractability. This paper suggests that this can be achieved by incorporating the entire effective bandwidth [8] of TCP traffic into performance model. Effective bandwidth accounts not only for the average rate, as conventional models, but also for the packet-level burstiness due to congestion avoidance TCP dynamics [8]. The tractability is achieved by using asymptotic for TCP traffic in a practically important regime of small buffer overflow probabilities $p \to 0$ [9].

The paper is organized as follows. Section II proposes a framework for incorporating packet-level TCP congestion avoidance dynamics into performance model of a TCP network. Section III discusses this framework, and on an example of TCP network comprised of a single buffered link demonstrates that this framework results are consistent with simulation and empirical results reported earlier. Finally, Section IV concludes and identifies directions of future research.

II. From Packet to Flow-Level TCP Performance

Following [9] consider TCP protocols with window sizes evolving according to the following homogeneous in time Markov chain $\{W_n, n = 0,1,..\}$:

$$W_{n+1} = \begin{cases} W_n + \lceil \kappa(p) W_n^k \rceil, & \text{with probability } e^{-pW_n^h} \\ \lceil \chi^{\upsilon_n} W_n \rceil, & \text{with probability } 1 - e^{-pW_n^h} \end{cases} \quad (1)$$

where parameters $k \in [0,1], h \geq 0, \chi \in (0,1)$, and i.i.d. sequence of the numbers of losses in n-th clump $\{\upsilon_n\}$ is independent of W_0. Case $k = 0$ corresponds to the Additive Increase Multiplicative Decrease (AIMD) TCP, and $k = 1$ corresponds to scalable TCP. TCP flow rate at moment t is $\xi_t = W_{\rfloor t/T \lfloor}/T$, where T is the round trip time, $\rfloor x \lfloor = \max\{n : n \leq x, n = -1,0,1,..\}$, and $n = \rfloor t/T \lfloor$. Define effective bandwidth of a TCP flow as follows [8]:

$$eff(s,t) = \frac{1}{st} \log E[e^{s\xi_{[0,t]}}] \quad (2)$$

where the amount of work generated by a TCP flow in the interval $[0,t]$ is $\xi_{[0,t]} = \sum W_n$, and $n = 0,..,\rfloor t/T \lfloor$.

Scaling [9] for Markov chain (1) can be reformulated in terms of the effective bandwidth (2) as follows:

$$eff(s,t|p) \sim A(p)\varepsilon\left[\frac{A(p)}{M(p)}s, M(p)\frac{t}{T}\right], \quad p \to 0 \quad (3)$$

where the average flow rate scales as $A(p) = [\kappa(p)/p]^{\frac{1}{1+h-k}}$, time scales as $M(p) = [\kappa(p)]^{\frac{h}{1+h-k}} p^{\frac{1-k}{1+h-k}}$, and the effective bandwidth of the corresponding limit process is $\varepsilon(s,t) = (st)^{-1} \log E[e^{s\chi_{[0,t]}}]$. Due to limited space we make a number of assumptions which can be relaxed. We assume that $\kappa(p) = \kappa$ is independent on p, and use the following approximation for the effective bandwidth (2):

$$eff(s,t|p) \approx \left(1 + \frac{s\omega^{2-k}}{2} x^{2-k}\right) x \quad (4)$$

which can be obtained from (3) assuming $p, s \to 0$, where the average flow rate $x \sim A(p)$ and ω is some positive parameter independent on s, t and p.

Consider a TCP network carrying a fixed number of flows N_r on feasible routes $r \in R$. Effective bandwidth of a TCP flow carried on a route $r \in R$ is $eff(s,t|p_r)$, where the corresponding end-to-end packet loss rate is p_r. Under assumption of small and independent packet loss rates on different links l: $p_l << 1$, one can approximate the effective bandwidth of the aggregate traffic carried on a link l as follows:

$$Eff_l(s,t) = \sum_{r: l \in r \subset R} N_r eff(s,t|p_r) \quad (5)$$

where

$$p_r \approx \sum_{l \in r} p_l \quad (6)$$

Consider a case when each link l of capacity C_l serves aggregate traffic through a completely shared buffer of size B_l according to the Fist-In-First-Out (FIFO) scheduling discipline. It is known [8] that aggregate effective bandwidth (5) can be used to approximate link l buffer overflow probability p_l. With such an approximation one can express the link overflow probabilities p_l as functions of the vector of the average end-to-end flow rates x_r and the numbers of carried flows N_r on feasible routes $r \in R$:

$$p_l = p_l(N_r, x_r, r : l \in r \subset R) \quad (7)$$

In a practically important case of small packet losses $p_l << 1$ it follows from scaling (3) that

$$x_r = [\kappa(p_r)/p_r]^{\frac{1}{1+h-k}} \chi_r \quad (8)$$

where χ_r does not depend on p_r. Expressions (6)-(8) form a closed system of fixed-point equations for vector of link l buffer overflow probabilities (p_l), given a vector of numbers of carried flows (N_r). After solving this system, whose dimension is equal to the number of links in the network, end-to-end flow rates x_r are explicitly determined by (8).

Here, following [4], we only consider approximation for buffer overflow probability, which uses two first terms in expansion for the aggregate effective bandwidth (5):

$$Eff_l(s,t) \approx \left(1+\frac{s\sigma_l^2}{2}\right)X_l \qquad (9)$$

where average rate of the aggregate traffic carried on link l is $X_l = \sum_{r:l\in r} N_r x_r$ and parameter $\sigma_l^2 = X_l^{-2}\sum_{r:l\in r} N_r \omega_r x_r^{3-k}$ characterizes relative burstiness of the aggregate traffic carried on link l. Approximation [4] for link l buffer overflow probabilities p_l is

$$p_l \approx \frac{\theta_l e^{\theta_l B_l}}{e^{\theta_l B_l}-1}\left(\frac{1+\sigma_l^2/\rho_l}{2}\right) \qquad (10)$$

where $\theta_l = 2(\rho_l - 1)/(\rho_l\sigma_l^2+1)$ and the average link l utilization is $\rho_l = X_l/C_l$. Assuming $2(1+\sigma_l^2/\rho_l) << B_l << \theta_l^{-1}$ we obtain from (10)

$$p_l \approx \frac{1+\sigma_l^2/\rho_l}{2B_l} \qquad (11)$$

Following [2] consider files transfer flows arriving on feasible routes r according to Poisson processes of rates λ_r. The size of a file arriving on a route r is distributed exponentially with average b_r. All arrival processes and file sizes are jointly statistically independent. Assuming that flow control operates on much faster time-scale than the process of flow arrivals/departures, vector of the numbers of flows in progress on all feasible routes $r\in R$, $n(t)=(n_r(t))$ is a Markov process. Probabilities $P(t,n) = \Pr ob\{n(t)=n\}$ are determined by the corresponding Kolmogorov equations. According to the Little theorem [10], the average file transfer time on a route r is $\tau_r = \lambda_r^{-1}\overline{n}_r$, where $\overline{n}_r$ is the steady-state average number of flows carried on route r.

III. Discussion

It is natural to expect that the proposed framework, which takes into account the packet-level burstiness of TCP-controlled traffic, results in worse packet and flow-level performances as compared to the conventional framework, which does not take into account the packet-level burstiness. Worse packet-level performance manifests itself in higher buffer overflow rates and lower link utilization. Worse flow-level performance manifests itself in more carried flows and longer file transfer times.

Expression (9) indicates that as the number of flows in progress increases, the burstiness of the aggregate TCP traffic decreases to zero, causing the corresponding effective bandwidth to decrease and approach the average aggregate rate. Thus the difference between the Markov model of TCP network under arriving/departing flows, which takes into account the packet-level TCP busrstiness, and the corresponding conventional Markov model, which does not take this bursiness into account, diminishes as the numbers of flows in progress increases. Since ergodicity of the Markov model of a TCP network is determined by the network behavior for large numbers of carried flows, one may expect that both Markov models have the same ergodicity conditions: the average utilization of each network link should be less than unity [11]:

$$\rho_l \overset{def}{=} \frac{1}{C_l}\sum_{r:j\in r\subset R}\lambda_r b_r < 1 \qquad (12)$$

In the rest of this Section we illustrate and quantify some of these statements for a case of TCP network comprised of a single link of capacity C serving TCP traffic through a buffer of size B, assuming applicability of approximation (11).

We consider a natural asymptotic regime of small packet loss rate: $p\to 0$ and large number of carried flows $N\to\infty$. It follows from (9) that in this regime $\sigma^2 \sim Nx^{3-k}/X^2 = x^{2-k}/X$ and $X\approx C$. Combining these relations with scaling $x\sim(1/p)^{\frac{1}{1+h-k}}$ we obtain

$$\sigma^2 \approx (\omega/C)p^{-\frac{2-k}{1+h-k}} \qquad (13)$$

Substituting (13) into (11) we obtain the following fixed-point equation for p:

$$p \approx \frac{1}{2B}\left(1+\frac{\omega}{\rho C}\left(\frac{1}{p}\right)^{\frac{2-k}{1+h-k}}\right) \qquad (14)$$

Under our assumptions equation (14) yields:

$$p \approx \frac{1}{2B}\left(1+\frac{\omega}{\rho C}(2B)^{\frac{2-k}{1+h-k}}\right) \qquad (15)$$

Approximation (15) describes hyperbolical decay of buffer overflow probability as buffer size increases with range of interest. Since this phenomenon is indicative of traffic self-similarity and long-range dependencies [5]-[6], the proposed performance model seems to be in agreement with empirical data. In a particular case of AIMD TCP, when $h=1, k=0$, we obtain from (15), (13): $p\approx 1/(2B)+\omega/(\rho C)$ and $\sigma^2 \approx 2\rho B/(2B+\rho\omega^{-1}C)$. These results are at least qualitatively consistent with empirical observation [4]. It can be shown that more accurate calculations lead to better agreement of the theoretical and empirical results.

Figure 2 sketches the link aggregate load $X=Nx$ and goodput G vs. the number of carried flows N. As N increases, the relative burstiness of the aggregate instantaneous load decreases causing increase in X and G. Combining these qualitative considerations with observation that the aggregate link goodput cannot exceed the link capacity: $G\le C$, we propose the following approximation

for the aggregate goodput $G \approx \widetilde{G} = \min\{Nx, C\}$, or equivalently, $\widetilde{G} = Nx$ for $N \le N^*$ and $\widetilde{G} = C$ for $N > N^*$, where N^* is the solution to equation $Nx = C$. It can be shown that in a case of a single link fixed-point model (6)-(8) yields $N^* \approx mCT$, where m is a TCP version specific coefficient and T is the round-trip time.

Figure 2. Aggregate throughput and goodput

In a case of a single link the proposed Markov model results in the following steady-state distribution for the number of flows in progress: $P(N) = Z^{-1}(\lambda b)^N \Big/ \prod_{n=1}^{N} G(n)$, where Z is the normalization constant. Figure 3 sketches the average number of flows in progress $N_{ave} = \sum_N NP(N)$, the average number of flows in progress calculated according to the conventional Markov model $\overline{N} = (1-\rho)^{-1}$ where link utilization $\rho = \lambda b / C < 1$, and solution to the mean-field equation $\widetilde{G}(\widetilde{N}) = \lambda b$ which states that the link goodput is sufficient for sustaining the average load.

Figure 3. Average number of flows in progress.

Analysis shows that $\widetilde{N} > \overline{N}$ for $\rho < \rho^*$, and $\widetilde{N} < \overline{N}$ for $\rho > \rho^*$, where $\rho^* \approx 1 - (mCT)^{-1}$, and T is round-trip time. Thus, the flow-level queuing performance is dominated by the effect of packet-level congestion avoidance TCP dynamics for sufficiently light load, and is dominated by the effect of flow-level arrivals/departures for sufficiently heavy load. This result is qualitatively consistent with Figure 1, since conventional flow-level queuing performance Markov model [2] accounts for random flow arrivals/departures but disregards packet-level TCP randomness. Threshold value $\rho^* \approx 1 - (mCT)^{-1}$ indicates that the applicability region of the conventional Markov model shrinks with increase in the delay-bandwidth product.

IV. Conclusion and Future Research

This paper has proposed a framework for combing packet and flow-level burstiness into Internet performance modeling. The proposed framework has advantage of generality as compared to performance modeling of specific TCP versions [12]-[13]. We demonstrated that the proposed framework is tractable and consistent with known empirical results. This combination of traits makes the proposed framework suitable for comparative performance analysis of the existing and emerging TCP versions as well as Internet design, operations, and management. Future research should evaluate accuracy and reveal potential of the proposed framework through a combination of simulations and analytics.

References

[1] R.J. Gibbens, S.K Sargood, C. Van Eijl, F.P. Kelly, H. Azmoodeh, R.F. Macfadyen, and N.W. Macfadyen, "Fixed-point model for end-to-end performance analysis of IP networks," 13th ITC Specialist Seminar: IP Traffic Modeling, Measurement and Management, 2000.

[2] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, and J. Roberts, "Statistical bandwidth sharing: a study of congestion at flow level," SIGCOMM 2001.

[3] D. Genin and V. Marbukh, "Bursty fluid approximation of TCP for modeling Internet congestion at the flow level," 47th Annual Allerton Conference on Communication, Control, and Computing, 2009.

[4] A. Lakshmikantha, C. Beck and R. Srikant. Impact of File Arrivals and Departures on Buffer Sizing in Core Routers, IEEE/ACM Transactions on Networking, April 2011.

[5] B. Tsybakov and N. D. Georganas, "On self-similar traffic in ATM queues definitions, overflow probability bound, and cell delay distribution", *IEEE/ACM Trans. on Networking*, 5(3): 397-409, 1997.

[6] M. Crovella and A Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *IEEE/ACM Trans. on Networking*, 1997.

[7] D. Wischik and N. McKeown, "Part I: Buffer sizes for core routers," *Comput. Commun. Rev.*, vol. 35, no. 3, pp. 75–78, Jul. 2005.

[8] F.P. Kelly, "Notes on effective bandwidth," In Stochastic Networks: Theory and Applications (Editors F.P. Kelly, S. Zachary and I.B. Ziedins), Royal Statistical Society Lecture Notes Series, 4. Oxford University Press, 141-168, 1996.

[9] K Maulik and B. Zwart, "An extension of the square root law of TCP," Annals of Operations Research, 170, No. 1, 217-232, 2009.

[10] Bertsekas, D. P. and Gallager, R. G., Data Networks, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1987; Edition 2, 1991.

[11] G. de Veciana, T.J. Lee, and T. Konstantopoulos, "Stability and performance analysis of networks supporting elastic services," *IEEE/ACM Trans. on Networking*, No. 9, pp. 2-14, 2001.

[12] C. Graham, P. Robert, and M. Verloop, "Stability properties of networks with interacting TCP flows," arXiv:0906.2615v1.

[13] E. Altman, K. Avrachenkov, C. Barakat, A.A. Kherani, and B.J. Prabhu, "Analysis of MIMD congestion control algorithm for high speed networks," The international Journal of Computer and Telecommunications Networking, Vol. 48, Issue 6, 2005.