

What Level of Estimating Accuracy Does TCP Need and Can TCP Achieve

Lin Cong¹, Guohan Lu², Yang Chen³, Beixing Deng¹, Xing Li¹

¹ Department of Electronic Engineering, Tsinghua University, Beijing, P. R. China

² Microsoft Research Asia, Beijing, P. R. China

³ Institute of Computer Science, University of Göttingen, Germany

ABSTRACT

Accurate and effective estimation of TCP operating point impacts directly on TCP performance, especially in modern high-speed networks. In this paper, we focus on two crucial questions: 1) In order to guarantee the performance and stability of TCP, how accurately does the TCP operating point need to be estimated? 2) Ahead of the occurrence of network congestion, how accurately can TCP estimate its operating points at the end host? We declare that the estimator which is based on TCP packet series can satisfy the accuracy TCP required.

Categories and Subject Descriptors

C.2.2 [Network Protocols]: Protocol verification – TCP/IP.

General Terms

Measurement, Performance

Keywords

TCP operating point; available bandwidth estimation.

1. INTRODUCTION

Since TCP accounts for the vast majority of Internet traffic, TCP congestion control has been a major objective of networking research for a long term, especially for the modern high-speed network. According to literature, a congestion control method should achieve high utilization, small queueing delay, freedom from oscillations, and fairness in bandwidth allocation [1]. To satisfy these requirements, an accurate and effective estimator of TCP operating point is the key prerequisite for all current TCP variances, such as CUBIC [2], Compound TCP [3] and etc.

Traditionally, under the end-to-end architecture, TCP has only two signals to estimate its operating point without modifying switchers: packet loss and delay. Obviously, these estimators can work only after the occurrence of packet delay or loss, when network congestion has happened already. On the other hand, according to information carried in the TCP packet series, end hosts can estimate the operating point ahead of the packet delay and loss. We study this consequential question: is this kind of estimator accurate enough to meet the requirements of TCP? As far as we know, it's pioneered in this literature.

2. ACCURACY LIMITATION

TCP operating point, termed as bandwidth-delay production (*BDP*) is the product of round-trip time (*rtt*) and available bandwidth (*av-bw*), while *av-bw* is the product of bottleneck link capacity (*lc*) and link utilization ratio (ρ). *rtt* and *lc* are considered

to be stable in the time scale of TCP transmission. Hence how to estimate ρ is the critical problem for TCP operating point estimation. Suppose that TCP sender transmits $N+1$ packets to network. We denote data set Δt_M as the N time-gaps between every two consecutive packets, while $\Delta t'_M$ as the arrival time-gap series at receiver. Fundamentally, without the occurrence of packet loss or delay, no matter what special estimation methods we use, TCP estimates its operating point only based on the differences of the distribution of $\Delta t'_M$.

Now we investigate the accuracy limitation by using this kind of differences to estimate link utilization ratio ρ . Focus on single-hop scenario first. If ρ is determined, the probability distribution of $\Delta t'_M$ p is also determined. For example, if the service time of router is constant and the background traffic is Poisson flow, we can calculate this distribution p under different ρ using *M/D/1* queueing model [4]. Effort [5] has been devoted to describe the relationship between the difference of distributions and the probability of estimating error. Let p_1/p_2 be two distributions of $\Delta t'_M$ under different ratios ρ_1/ρ_2 . The Directed Divergence I between p_1

and p_2 is defined as,
$$I(p_1; p_2) = \sum_{t \in \Delta t'_M} p_1(t) \log \frac{p_1(t)}{p_2(t)}.$$

(2.1)

We consider a simple hypothesis testing problem on the ρ parameters of this probability distribution, where $H_1: \rho_1$ and $H_2: \rho_2$. Using Chernoff information [5], which is believed to be the best asymptotic achievable exponent for hypothesis testing, we can obtain the maximum probability of error when we determine whether H_1 or H_2 is accepted as,
$$P_e = 2^{-NI(p_2; p_1)} = 2^{-NI(p_1; p_2)}.$$

(2.2)

p_i is the distribution of $\Delta t'_M$ which guarantees that $I(p_i; p_1) = I(p_i; p_2)$ while its existence can be proved. This formula indicates: if the directed divergence between two distributions is positive, there must exist an N that is larger enough to guarantee the error small enough. We extend this result under two hypotheses testing to multiple hypotheses testing. Notice that our discussion is based on the initial behavior of TCP packet series, which is independent of any details on congestion avoidance level, so the analysis of accuracy fits for all of TCP variances.

Theorem 1: Suppose a finite discrete set of ρ , $[\rho_1, \rho_2, \dots, \rho_m]$, with each element belonging to $[0, 1]$. If N , the length of TCP packet series is large enough, the accuracy of estimation between these different ρ can be guaranteed.

As for multi-hop scenario, the utilization ratio of each link forms vector $\mathbf{\rho}$, which also determines the distribution of $\Delta t'_M$ completely. Thus we have the following inference.

Corollary: If the length of TCP series N is large enough, the estimating accuracy of vector $\mathbf{\rho}$ can be guaranteed.

Copyright is held by the author/owner(s).

CoNEXT Student Workshop'09, December 1, 2009, Rome, Italy.

ACM 978-1-60558-751-6/09/12.

3. TCP REQUIREMENT

We obtain the up and low bounds of the TCP operating points (*BDP*) estimation. Our market-based approach is studied in [6][7][8]: several links indexed by l , shared by a set of source-destination pairs, indexed by i . Each source i has an associated transmission rate $x_i(t)$. The set of transmission rates determines the aggregate flow $y_l(t)$ at each link l , by the equation:

$$y_l(t) = \sum_i x_i(t - \tau_{li}^f). \quad (3.1)$$

In this equation, the forward delays between sources and links are termed as τ_{li}^f . Each link has capacity c_l . The feedback mechanism is modeled as follows. Each link has an associated congestion price $p_l(t)$, which is proportional to the congestion state in each link. Sources can access to $p_l(t)$ of all links to estimate the congestion state $q_{ii}(t)$ in network: $q_i(t) = \sum_l p_l(t - \tau_{li}^b)$. (3.2)

We set the backward delays as τ_{li}^b in the feedback path from destinations to sources. Note vectors (x, y, p, q) collect the above quantities across destinations and links. The total *rtt* for the source thus satisfies:

$$\tau_i = \tau_{li}^f + \tau_{li}^b. \quad (3.3)$$

We assume sources have a demand curve that specifies their rate as a decreasing function of price:

$$x_i(t) = f_i(q_i) \quad (3.4)$$

And we use this classical algorithm to set prices at each link,

$$p_l = \gamma_l \int y_l(t) dt. \quad (3.5)$$

Here the γ is a positive constant to be chosen for each link, and could be estimated by every source.

We recognize that equilibrium considerations (x_0, y_0, p_0 and q_0) of system are meaningful that system is operating at or near operating point. We pursue local stability results, based on small perturbations $x=x_0+\Delta x, y=y_0+\Delta y, p=p_0+\Delta p, q=q_0+\Delta q$ around this point and studied via linearization. Then the linearization of (3.4), (3.5) around operating point in the Laplace domain are:

$$dx_i = -\kappa_i dq_i, dp_l = \frac{\gamma_l}{s} dy_l.$$

The measurement units for each parameter (x, y, q, p, κ and γ) are (bps, bps, 1, 1, bps and bps⁻¹). Our question is how to choose the gains γ and κ so that the feedback loop remains stable while keeping high performance.

Considering the time-shifting characteristic of Laplace, we can obtain the Laplace transform of (3.1) and (3.2) in the following forms:

$$R_f(s) = e^{-\tau_f s}, R_b(s) = e^{-\tau_b s}.$$

Considering that sources can estimate τ , we set $\kappa = \alpha/\tau$, where α is the actual *BDP*. The Closed-Loop Transfer Function of this system is denoted as,

$$L(s) = R_f(s) R_b(s) \kappa \frac{\gamma}{s} = \kappa \gamma \frac{e^{-\tau s}}{s} = \alpha \gamma \frac{e^{-\tau s}}{\tau s}. \quad (3.6)$$

We study the stability of this closed loop with Nyquist criterion method. This loop would be unstable for large τ , unless the gain $\alpha\gamma$ compensates for it as,

$$\alpha\gamma \leq \pi/2 \quad (3.7)$$

On the other hand, considering that τ_{ii}^f is approximate to $\tau/2$, the Open-Loop Transfer Function is,

$$T(s) = \frac{\bar{R}_f(s)}{1 + L(s)} \approx e^{-\frac{\tau s}{2}} \left(1 + \alpha \gamma \frac{e^{-\tau s}}{\tau s} \right)^{-1} \quad (3.8)$$

We calculate Range-Frequency Property of (3.8) with different $\alpha\gamma$. The max value of each curve indicates the maximum gain for this open loop system, which can guarantee the performance of transmission. We can see these curves first touch a rational negative gain with parameters 0.8, which implies that, $\alpha\gamma \geq 0.8$ (3.9)

As mentioned before, the physical meaning of α is *BDP* and $1/\gamma$ is the operating point estimated by sources, so combined (3.7) and (3.9) we obtain:

Theorem 2: The low and up bounds for the estimation of operating points which can guarantee both the performance and stability of TCP transmission are:

$$\frac{2}{\pi} BDP \leq \text{OperatingPointEstimation} \leq \frac{1}{0.8} BDP. \quad (3.10)$$

Numerical simulation validate that: we can use the difference of Δt_M distributions to estimate TCP operating point while satisfying the accuracy demand illustrated by (3.10) with the estimating error less the 5% and the series length N less than 2⁸.

4. CONCLUSION ANF FUTURE WORK

Devoting to *Information Theory*, we discuss the limitation of accuracy of operating points we can estimate. On the other hand, devoting to *System Theory*, we determine the requirement of accuracy for TCP operating points. Our work shows that the estimator which applies the differences of distribution of packet time series can accommodate the requirements of TCP. Fundamentally speaking, our work is also one step to answer how far we can go in achieving better performance of data transmission within the existing end-to-end philosophy of Internet. But our work is not merely an incremental improvement but rather an insight into TCP congestion estimation. Our future work will focus on how to design an estimator to approach the limitation of estimating accuracy and improve TCP performance.

5. ACKNOWLEDGMENTS

This work is supported by National Basic Research Program of China (973 Project, No.2007CB310806) and National Science Foundation of China (No.60850003).

6. REFERENCES

- [1] F. Paganini, Z. Wang, J. Doyle and S. H. Low, "Congestion control for high performance, stability and fairness in general networks," IEEE/ACM Transactions on Networking, vol. 13, iss.1, pp. 43 - 56, February 2005.
- [2] S. Ha, I. Rhee and L. Xu, "CUBIC: A New TCP Friendly High-Speed TCP Variant," In Proc. of ACM SIGOPS, 2008.
- [3] K. Tan, J. Song, Q. Zhang and .etc, "A Compound TCP Approach for High-speed and Long Distance Networks," In Proceedings of IEEE INFOCOM, 2006.
- [4] D. Gross and G. M. Harris, "Fundamentals of Queueing Theory", Wiley Press, Fourth Edition, 2008.
- [5] Thomas M. Cover, Joy A. Thomas. "Elements of Information Theory", Wiley Press, Second Edition, 2006.
- [6] J. Wang, D. X. Wei, and S. H. Low. "Modeling and Stability of FAST TCP," In Proc. of IEEE INFOCOM, March 2005.
- [7] Lee J.W., and .etc, "Price-based distributed algorithms for rate-reliability tradeoff in network utility maximization," IEEE JSAC, vol. 24, iss. 5, pp. 962-976, May 2006.
- [8] A. Tang, J. Wang, S. H. Low and M. Chiang, "Equilibrium of Heterogeneous Congestion Control: Existence and Uniqueness," IEEE/ACM Transactions on Networking, 2007.