UNICAST INFERENCE OF NETWORK LINK DELAY DISTRIBUTIONS FROM EDGE MEASUREMENTS

Meng-Fu Shih *Alfred Hero*

Department of EECS
University of Michigan
Ann Arbor, MI 48109-2222, U.S.A.
mfshih@umich.edu, hero@eecs.umich.edu

ABSTRACT

Inference of network internal link characteristics has become an increasingly important issue for operating and evaluating large telecommunication networks. Since it is usually impractical to directly monitor each link along a specific path, end-to-end probes are sometimes used to collect link characteristic information at edge nodes of the network. This paper deals with unicast probing methods for estimation of link delay characteristics. Unicast traffic is easy to generate and is supported by almost every network currently in operation. Under the assumptions that link delays are spatially and temporally independent, we propose a bias corrected estimator for the internal link delay cumulant generating function (CGF) based on unicast probe end-to-end delay measurements. Through simulation we show that the proposed estimator attains a level of mean squared error comparable to link delay CGF estimates obtained from directly measured link delay statistics. We can use these CGF estimates to estimate delay mean, variance and level exceedance probabilities for each link.

1. INTRODUCTION

Network monitoring, prediction and diagnosis are very important issues for network operators and designers. However, these are challenging problems due to several factors: (1) direct measurement of packet transport statistics are usually impossible - internal nodes may not support such diagnostics or these diagnostics may be disabled to minimize overhead; (2) the internal parameters of ISP controlled links are usually inaccessible to outsiders. Network monitoring algorithms can be categorized into two different groups: passive and active methods. Active methods send special packets, called probes, over known routes through the network to measure end-to-end delay, packet error rates, and packet loss probabilities. Vardi [1] proposed a network tomography method to estimate source-destination traffic intensities by monitoring the link count data. This problem was further investigated by Cao, et al [2] for the case of time varying network transport characteristics. The methods of [1] and [2] are passive as no probes are used. Active probing methods are an alternative way to collect internal link statistics. Despite the fact that active probes may perturb the network traffic, these methods can give more reliable information about link behavior than passive methods. There are two kinds of active probing schemes: unicast and multicast probing. Several papers on link bandwidth measurement propose unicast probes (e.g., Jacobson [3], Downey [4], Lai & Baker [5]). Multicast probes are also widely used to perform inference of link delay distributions, internal loss characteristics, and network connectivity. (e.g., see [6] [7], [8], [9]).

In this paper, we focus on estimating the internal link delay cumulant generating function (CGF). Packet delays are due to three factors: (1) queueing delays, which depend on service times and buffer occupancy; (2) transmission delays, which depend on packet size and link data rate; (3) propagation delays, which depend on the transmission medium. The sum of these delays over a route can be measured by end-to-end delays of unicast probes sent over the route. After collecting a sufficient number of these probes, an overdetermined system of equations is constructed for the delay CGFs. Based on a least-square approximation, we propose a bias corrected estimator for each internal link delay CGF. We evaluate performance of the algorithm using the `ns` network simulation program [10]. Several measures of performance are investigated, including overall mean-square goodness of fit of the estimated CGF to the empirical CGF and correct bottleneck detection probability and bottleneck localization.

The paper is organized as follows. In section 2, we describe the network delay model. Section 3 presents the bias corrected internal link delay CGF estimator. Section 4 presents the result of computer simulations comparing the bias corrected to the sample mean link delay CGF estimates. In section 5, we present several applications and extensions of our technique.

2. NETWORK DELAY MODEL

Let a communication network consist of m internal links. Identical probe packets are sent through n paths across the network. Suppose we know the routing of each of the probes which specifies the nxm probe routing matrix A. A has elements a_{ij} equal to 1 when probe path i intersects link j, and equal to 0 otherwise. Let M_i denote the set of link indices which compose the ith probe path, $i = 1, \ldots, n$. Then $Y_i = \sum_{j \in M_i} X_{ij}$ is the measured end-to-end delay of a probe transmitted along the ith path where X_{ij} is the delay encountered by probe i across link j and $i = 1, \ldots, n$. Define the end-to-end probe delay CGF $K_{Y_i}(t) = \log E[e^{tY_i}]$ and the link delay CGF of the jth link $K_{X_{ij}}(t) = \log E[e^{tX_{ij}}]$, $j \in M_i$, with CGF parameter t, $t \in (-\infty, \infty)$. We make the following spatial independence and stationary assumptions, respectively:

A1) The link delays X_{ij} are mutually independent, $j \in M_i, i = 1, \ldots, n$.

A2) If paths of probe i and probe k both contain a common link j, then X_{ij} and X_{kj} have identical CGF denoted K_{X_j}.

The CGF of Y can therefore be expressed as

$$\begin{aligned} K_{Y_i}(t) &= \log E\left[e^{tY_i}\right] \\ &= \log E\left[e^{t\left(\sum_{j \in M_i} X_{ij}\right)}\right] \\ &= \log\left\{\prod_{j \in M_i} E\left[e^{tX_{ij}}\right]\right\} \\ &= \sum_{j \in M_i} \log E\left[e^{tX_{ij}}\right] \\ &= \sum_{j=1}^{m} a_{ij} \cdot K_{X_j}(t) \\ &= A_{(i)} \cdot K_X(t) \qquad (1) \end{aligned}$$

where $A_{(i)}$ denotes the ith row of the routing matrix A and $K_X(t) = [K_{X_1}(t), \ \ldots, K_{X_m}(t)]^T$ (^{T} denotes transpose). Thus we can express the vector of end-to-end CGF's $K_Y(t) = [K_{Y_1}(t), \ \ldots, K_{Y_n}(t)]^T$ by the linear relation

$$K_Y(t) = A \cdot K_X(t). \qquad (2)$$

When $n \geq m$ and A is full rank, the relation (2) is invertible and thus $K_X(t)$ can be determined from $K_Y(t)$ by the formula $K_X(t) = (A^TA)^{-1}A^TK_Y(t)$. Let $B = (A^TA)^{-1}A^T$, then we have

$$K_{X_j}(t) = \sum_{i=1}^{n} b_{ji} \cdot K_{Y_i}(t). \qquad (3)$$

A full rank matrix A can be ensured by making $n \geq m$, and selecting distinct probe paths which cover the network, i.e., every link is contained in some path. When A is not full rank, only linear combinations of link CGF's lying outside of the null space of A can be determined from (2).

3. ESTIMATION OF CGF

Let N_i be the number of probes collected for a given path i, $i = 1, \ldots, n$. Define

$$\hat{M}_{Y_i}(t) = \frac{1}{N_i} \sum_{k=1}^{N_i} e^{tY_{ik}}, \qquad (4)$$

where Y_{ik} is the measured end-to-end delay of the kth received probe along path i. We obtain estimates of the vector $K_X(t)$ from $\hat{M}_Y(t) = [\hat{M}_{Y_1}(t) \ldots \hat{M}_{Y_n}(t)]^T$ by the method of least-squares (LS). Note that as $\hat{M}_{Y_i}(t)$ is an unbiased estimate of the moment generating function $M_{Y_i}(t) = e^{K_{Y_i}(t)}$, a plausible estimator for $K_{X_j}(t)$ in (3) would be the method-of-moments estimate (MOM):

$$\hat{K}'_{X_j} = \sum_{i=1}^{n} b_{ji} \cdot \log\left(\hat{M}_{Y_i}(t)\right) \qquad (5)$$

Unfortunately, this estimator is biased due to non-linearity of the log. In order to obtain a bias corrected estimator for $K_{X_j}(t)$, we apply a technique similar to that of Gibbens [11]. In [11] linearization was used to derive bias corrected estimators for *effective bandwidth*, which is of similar mathematical form as the cumulant generating function. Observe that as $\log(1+u) = u - \frac{u^2}{2} + H.O.T.$

$$\begin{aligned} &\hat{K}'_{X_j}(t) = \log\left\{\prod_{i=1}^n \left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}\right\} = \\ &\log\left\{\prod_{i=1}^n E^{b_{ji}}\left[\hat{M}_{Y_i}(t)\right] - \left(\prod_{i=1}^n E^{b_{ji}}\left[\hat{M}_{Y_i}(t)\right] - \right.\right. \\ &\left.\left.\prod_{i=1}^n \left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}\right)\right\} = \\ &\log\left\{\prod_{i=1}^n E^{b_{ji}}\left[\hat{M}_{Y_i}(t)\right]\left(1 - \left(1 - \frac{\prod_{i=1}^n \left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}}{\prod_{i=1}^n E^{b_{ji}}\left[\hat{M}_{Y_i}(t)\right]}\right)\right)\right\} \\ &\approx K_{X_j}(t) - w_j - \tfrac{1}{2}w_j^2, \end{aligned}$$

where $w_j = 1 - \frac{\prod_{i=1}^n \left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}}{\prod_{i=1}^n \left(E\left[\hat{M}_{Y_i}(t)\right]\right)^{b_{ji}}}$. This suggests that a reasonable way to correct the bias is to use (3) with an estimate of w_j:

$$\hat{K}_{X_j}(t) = \sum_{i=1}^{n} b_{ji} \log\left(\hat{M}_{Y_i}(t)\right) + \hat{E}[w_j] + \frac{1}{2}\hat{E}[w_j^2], \quad (6)$$

where $\hat{E}[\cdot]$ denotes empirical average for which we use MOM estimates

$$\hat{E}[w_j] = 1 - \frac{\prod_{i=1}^{n} \hat{E}\left[\left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}\right]}{\hat{M}_{X_j}(t)} \tag{7}$$

$$\hat{E}[w_j^2] = 1 - \frac{2\prod_{i=1}^{n} \hat{E}\left[\left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}\right]}{\hat{M}_{X_j}(t)} + \frac{\prod_{i=1}^{n} \hat{E}\left[\left(\hat{M}_{Y_i}(t)\right)^{2b_{ji}}\right]}{\hat{M}_{X_j}^2(t)}. \tag{9}$$

$\hat{M}_{X_j}(t)$ is an estimate of the moment generating function of link delay at link j, which can be obtained from

$$\hat{M}_{X_j}(t) = \prod_{i=1}^{n} \left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}. \tag{10}$$

We obtain the empirical average $\hat{E}\left[\left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}\right]$ by implementing a sliding window method with window size W and step size S. Define the number $N_w = \lfloor \frac{N_i - W}{S} \rfloor$ of window increments

$$\hat{E}\left[\left(\hat{M}_{Y_i}(t)\right)^{b_{ji}}\right] = \sum_{l=1}^{N_w} \frac{1}{N_w} \left(\frac{1}{W} \sum_{k=(l-1)S+1}^{(l-1)S+W} e^{t y_{ik}}\right)^{b_{ji}}. \tag{11}$$

We obtain the empirical average $\hat{E}\left[\left(\hat{M}_{Y_i}(t)\right)^{2b_{ji}}\right]$ in a similar manner.

4. EXPERIMENTAL RESULTS

We used the ns network simulator program to perform a *TCP/UDP* simulation of the network in Fig. 1. Probes were sent through 5 different paths in order to estimate delay CGF for 4 links. The topology is shown in Fig. 1, and the corresponding routing matrix A is

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{12}$$

We set up a similar test environment to that reported in Presti and Duffield [9]. All the links to be estimated had bandwidth 4Mb/sec with latency 50ms. Each link was modeled as a Drop-Tail queue (FIFO queue with finite buffer). The queue buffer sizes were 50 packets. We generated probes

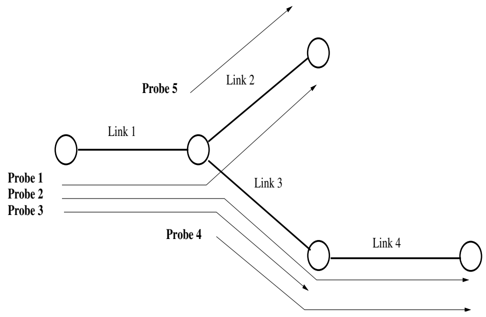

Fig. 1. Probe routing paths for the experiment described in Section 4

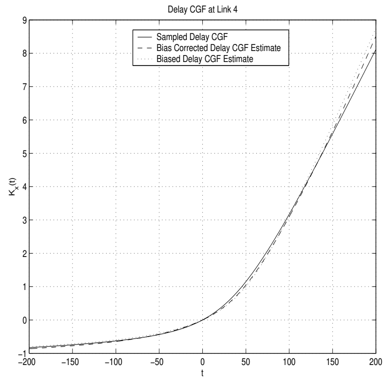

Fig. 2. Link delay CGF at link 4 in the ns simulation described in Section 4

as 40 byte UDP packets. The probe transmissions were generated independently at each source node according to a Poisson process with mean interarrival time being 16ms and rate 20Kb/sec. The background traffic consisted of both Exponential on-off UDP traffic and FTP traffic.

N probes were collected for each path for a total of 5xN probes. We estimated each probe queueing delay by substracting the minimum probe delay over the N trials. This provides a biased estimate of queueing delay across the probe path since the minimum probe delay is a biased estimator of transmission delay plus latency. However, the bias decreases as $1/N$. In order to estimate the expected values in (8) and (9), we set the window size W to be 2/3 of N, and the window shift step size S to be 10 probe delay samples.

We compared the proposed bias corrected estimator to the biased estimator (5) for $K_X(t)$. We evaluated the CGFs over the range $t = -200$ to $t = 200$. Comparing the estimates of CGF of sampled link delays with and without bias correction in Table 1, we can see that the proposed estimator achieves lower MSE. The corresponding estimated CGF for Link 4 is shown in Fig. 2. These results were obtained by using $N = 1500$ probes per route.

Table 1. MSE of $\hat{K}_{X_j}$ (bias corrected) and $\hat{K}'_{X_j}$ (no bias correction)

Link	1	2	3	4
MSE of $\hat{K}_{X_j}$	0.0086	0.0247	0.0483	0.0096
MSE of $\hat{K}'_{X_j}$	0.0060	0.0326	0.0644	0.0325

5. APPLICATIONS AND EXTENSIONS

Each link delay CGF preserves all the statistical information of the delay since it is the log of the Fourier transform of the link delay probability density function. We can accurately estimate many features of the delay distribution from the delay CGF. Here we give results for *Bottleneck link detection* We define a bottleneck as the event that the probability of a link delay exceeding some delay threshold δ exceeds a prespecified threshold P. By the Chernoff bound,

$$P(X_j \geq \delta) \leq e^{-t\delta} E\left[e^{tX_j}\right] = P_j. \tag{13}$$

By appropriately selecting the threshold δ and a threshold P close to 1, we can detect a bottleneck link by testing whether $\max_{j=1,\dots,m} P_j > P$ In Table 2, we show the Chernoff bounds for $P(X_j \geq \delta = 0.02s)$ which were estimated from the computer simulation in Section 4. By setting threshold P to be 0.95, we can identify link 3 as the bottleneck link.

Table 2. Chernoff bound and empirical estimate of $P(X_j \geq 0.02)$ for each link delay in Section 4

Link	1	2	3	4
P_j	0.7517	0.4030	0.9620	0.9012
$P(X_j \geq \delta)$	0.2504	0.1921	0.3447	0.2790

6. CONCLUSION AND FUTURE WORK

In this paper we proposed a unicast method to perform inference on internal link delay characteristics. We derived a bias corrected estimator for internal link delay cumulant generating functions based on LS approximation. The proposed estimator was evaluated by `ns` simulations with TCP/UDP background traffic and FIFO finite buffer link queues. The MSE of the proposed estimator is lower than that of the direct biased sample mean estimator.

In the future, we will look into the following issues. First, our proposed estimator assumes stationarity of the network over the probing period (Assumption A2), which may be violated in real applications. Some adaptive estimation must be done in order to track the true link delay distributions. Besides, if the internal link delays are spatially dependent, a more sophisticated model must be used.

7. REFERENCES

[1] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *J.Amer.Statist.Assoc.*, vol.91, pp.365-377,1996.

[2] J. Cao, D. Davis, S.V. Wiel, B. Yu, "Time-varing network tomography: router link data," *lucent Bell Lab Technical Memo*, Feb. 2000. `http://plan9.bell-labs.com/cm/ms/departments /sia/cao/htmls/pub.html`

[3] Jacobson. "`pathchar`-a tool to infer characteristics of Internet paths," Presented at the *Mathematical Sciences Research Institute*(MSRI), Apr. 1997.

[4] A. Downey, "Using `pathchar` to estimate Internet link characteristics," Proceedings of *ACM SIGCOMM '99*.

[5] K. Lai, M. Baker, "Measuring link bandwidths using a deterministic model of packet delay" Proceedings of *ACM SIGCOMM 2000*.

[6] A. Ziotopoulos, A. Hero, W. Wasserman, "Estimation of Network Link Loss Rates Via Chaining In Multicast Trees," submitted to *ICASSP* 2001.

[7] R. Caceres, N.G. Duffield, J. Horowitz, D.F. Towsley, "Multicast-based inference of network-internal loss characteristics," *IEEE Trans. on Information Theory*, vol.45, no.7, Nov. 1999.

[8] N.G. Duffield, F.F. Presti, "Multicast inference of packet delay variance at interior network links," Preprint from MINC, `http://www-net.cs.umass.edu/minc/`.

[9] F.L. Presti, N.G. Duffield, J. Horowitz, D. Towsley, "Multicast-based inference of network-internal delay distributions," Preprint from MINC, `http://www-net.cs.umass.edu /minc/`.

[10] *ns* - Network Simulator. See `http://www-mash .cs.berkeley.edu/ns/ns.html`.

[11] R.J. Gibbens "Traffic characterisation and effective bandwidths for broadband network traces," *Stochastic Networks, Theory and Applications*, pp. 169-179, Oxford Science Pub., 1996.