Measuring Probing Response Curves over the RON Testbed

Xiliang Liu[1], Kaliappa Ravindran[1] and Dmitri Loguinov[2]

[1] City University of New York, New York, NY 10016
xliu@gc.cuny.edu, ravi@cs.ccny.cuny.edu
[2] Texas A&M University, College Station, TX 77843
dmitri@cs.tamu.edu

Abstract. The probing response curves of an Internet path depict the first-order statistics of packet-train output signals (dispersion or rate) as a mathematical function of the input signal. Recent theoretical analysis showed that several useful path characteristics (e.g., raw capacity, cross-traffic rate, and residual bandwidth of the tight link) can be inferred from probing response curves [1]. To gain insight into the response curve properties of real Internet paths, we present a measurement study over the RON testbed covering 272 different paths among 17 RON hosts. We show that manually assisted estimation of the tight link capacity is possible using relatively short packet-trains and that for 98% of the paths measured, the available bandwidth is limited by an edge link with low utilization.

1 Introduction

In today's Internet, packet-train probing has become one of the major mechanisms for measuring useful path characteristics such as residual bandwidth, cross-traffic, and location of the available-bandwidth bottleneck link (also called the *tight* link of the path). Existing techniques often use first-order statistics (i.e., the mean) of packet-train output signals (dispersion or rate) and their functional relationship to the input signal to infer the desired information. We refer to this relationship as the *probing response curve* of the path. Recent analysis [1] shows that the response curve of a multi-hop path measured using sufficiently long packet-trains contains more information about the available-bandwidth bottleneck link than existing techniques have attempted to obtain. However, before we propose an additional measurement tool to infer these path characteristics, some experience and insights of how the response curves of real Internet paths behave are needed. Such an understanding can potentially lead to more effective measurement heuristics and to a better choice of the various parameters in the estimation algorithm.

Let us first define the concept of a probing response curve and several related path characteristics more formally. Consider an N-hop network path $\mathcal{P}$, where the capacity of link i is denoted by C_i and the long-term average of the cross-traffic arrival rate at that link is given by λ_i, the latter of which is assumed to be less than C_i. The hop available bandwidth of link i is $A_i = C_i - \lambda_i$. The tight link, with index b, is the link that carries the minimum available bandwidth:

$$b = \arg \min_{1 \leq i \leq N} (C_i - \lambda_i). \tag{1}$$

Suppose that we inject a probing train of n packets with equal inter-packet dispersion G_0 and packet size s into path $\mathcal{P}$. The output dispersion G_N is defined as the average inter-packet spacing within the train when it exits the path from link N. The input and output rates of the packet-train are given by $R_0 = s/G_0$ and $R_N = s/G_N$. Note that even though the input signal G_0 is deterministic, the output signal G_N is random due to cross-traffic burstiness. Assuming that random variable G_N has a time-invariant distribution for a given value of input dispersion G_0, the *gap-response curve* of path $\mathcal{P}$ is a functional relationship between $E[G_N]$ and G_0. Similarly, the *rate-response curve* is a relationship between $s/E[G_N]$ and R_0, where $s/E[G_N]$ is the average output rate at the receiver (which is the *harmonic mean* of output rate R_N) and R_0 is the input rate at the source. In this paper, we adopt a transformed version of the rate-response curve $\tilde{\mathcal{Z}}(R_0)$ that depicts the relationship between $E[R_0/R_N]$ and R_0. The advantages of this choice are that $\tilde{\mathcal{Z}}(R_0)$ has a direct association with the bandwidth characteristics of interests and that $\tilde{\mathcal{Z}}(R_0)$ exhibits piece-wise linearity when train length n is large, which facilitates information extraction from the curve.

It is shown in [1] that for any given probing packet size s, as packet train length n increases, the response curve $\tilde{\mathcal{Z}}(R_0)$ approaches a lower bound $\tilde{\mathcal{F}}(R_0)$. This lower bound is a piece-wise linear function of R_0 with increasing slopes. The first two linear segments in $\tilde{\mathcal{F}}(R_0)$ in the low input rate region can be expressed as

$$\tilde{\mathcal{F}}(R_0) = \max\left(1, \frac{\lambda_b + R_0}{C_b}\right) = \begin{cases} 1 & R_0 \leq A_b \\ \dfrac{\lambda_b + R_0}{C_b} & A_b \leq R_0 \leq \Lambda \end{cases}, \tag{2}$$

where Λ is the input rate at the end point of the second linear segment in $\tilde{\mathcal{F}}(R_0)$.

This result established the possibility in theory to obtain both C_b and λ_b of any path using long packet-trains. However, the practical feasibility of such measurements depends on the answers to the following questions. First, how long should the packet-trains be so that curve $\tilde{\mathcal{Z}}(R_0)$ is sufficiently close to $\tilde{\mathcal{F}}(R_0)$? Obviously, a requirement of using prohibitively long trains would render such measurements impractical. Second, how many linear segments usually appear in the curve $\tilde{\mathcal{F}}(R_0)$ and how prominent is the one that carries useful information? Understanding this problem helps us develop effective heuristics to locate the proper portion of the curve. Furthermore, if the second segment turns out in practice to be too short to be detected correctly, one should be advised against applying this theory to real measurements.

In this paper, we address the above questions using an extensive measurement study over the RON testbed. Our study covers 272 paths among 17 RON nodes. For each path, we measure the response curves using various packet-train

Table 1. RON nodes and their geographic locations

Node	Location	Node	Location	Node	Location
ana1-gblx	CA	ccicom	UT	chi1-gblx	IL
cmu	PA	coloco	MD	cornell	NY
digitalwest	CA	intel	CA	lon1-gblx	unknown
lulea	Sweden	mit-main	MA	nyu	NY
roncluster5	MA	speakeasy	MA	ucsd	CA
utah	UT	vineyard	MA		

length. We then eyeball the measured curves that have sufficiently approached their piece-wise linear lower bounds and manually identify the second segments to extract C_b and λ_b of the path. Our results show that for 98% of the paths measured, the response curve lower bound $\tilde{\mathcal{F}}(R_0)$ has no more than two observable linear segments, which enables easy visual identification of the second linear segment. This is an indication of the fact that the tight link is located at one edge of the path and is also the link with the minimum capacity. This allows us to compare our capacity measurement results with those of pathrate [2], a well-known public tool that measures the minimum capacity (i.e., bottleneck bandwidth) of end-to-end paths. We find that there are 39 paths whose true bottleneck capacity can be accurately measured by our method and not by pathrate. Our results also reveal that the packet-train length needed to assure measurement accuracy is only moderate, usually on the order of several dozen packets.

Note that the number of paths we examined is small and we do not attempt to generalize our observations to the whole Internet. Nevertheless, these results make a first step towards a new promising avenue of Internet bandwidth measurement.

2 Measurement Methodology

Our measurements of rate-response curves were conducted in the RON testbed [3] during a two-month period from April to June of 2005. We selected 17 RON nodes, mainly based on their availability to us at that time. Table 1 shows the selected RON nodes and their geographic locations. Among the 17 hosts, 16 are located in the United States and one is from Sweden. For each of the 272 possible source-destination pairs, we conducted two experiments on the corresponding path at approximately the same time. We first ran pathrate to estimate the minimum capacity of the path and then ran a packet-probing experiment to measure the rate-response curves of the path. For the measured response curves of each path, we visually identified the proper linear segment and applied linear regression to compute C_b and λ_b based on (2). We next describe the details about our probing experiments and our manually assisted bandwidth estimation from the measured response curves.

Our packet probing experiments operate as follows. We use packet-trains with packet size $s = 1500$ bytes and train length $n = 129$ to probe the path at 29 input rates, from 10 mb/s to 150 mb/s with a 5 mb/s increasing step. For each input rate, we use 200 trains to measure the average output dispersion $E[G_N]$. To avoid continuously congesting the path, the probing source idles for a random amount of time after sending a packet-train through the UDP channel. It then sends a control message through an additional TCP connection to inform the receiver of the upcoming transmission of the next train. Upon receiving the control message, the receiver saves into a trace file the arrival time of each packet received in the previous train.[3] The receiver then sends back an acknowledgement to the sender to initiate the transmission of the next train. The whole durations of the experiments range from 20 minutes to two hours, depending on the RTT and the available bandwidth along the path. To ensure that packet-trains of different rates sample the path within the same time interval, we interleave the packet-trains of different input rates so that the 200 packet-trains for each input rate are approximately equally separated during the whole measurement period.

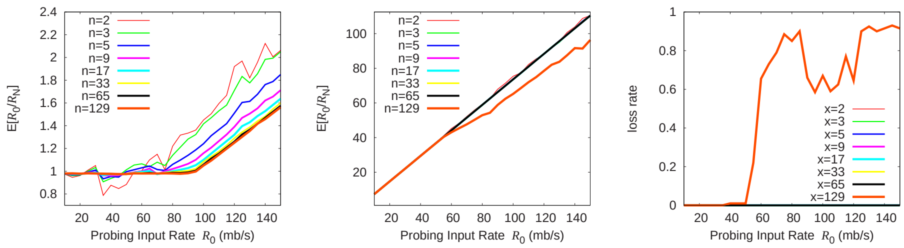
We make several comments on our experiments. First, the probing input range and the number of samples for each input rate are chosen based on our previous experience [1]. It is admittedly not an optimal choice for all 272 paths that exhibit diverse bandwidth characteristics. However, this decision greatly facilitates batch processing of all experiments. Second, note that our experiments not only allow us to obtain the response curve for $n = 129$, but also for any packet-train length k such that $2 \le k \le 129$, by simply taking the dispersion of the first k packets in each train. Finally, our experimental design requires sending a total of $748,200$ packets for each path. Even though this is a large amount of traffic, the purpose of this paper is not to advocate our method as another technique to be used directly in real measurements. Instead, we hope that the insights gained in this study will lead to the development of more efficient and accurate measurement tools.

After each probing experiment is finished, we compute from the trace file eight rate-response curves for train lengths $2, 3, 5, 9, 17, 33, 65$ and 129. To compute the rate response $\tilde{\mathcal{Z}}(R_0)$ with packet-train length k , we use all packet trains whose input rate was R_0 and whose first and k^{th} packets arrived at the receiving host successfully.

3 Manually Assisted Bandwidth Estimation

To estimate the tight link bandwidth characteristics, we eyeball the eight plotted response curves and confirm their convergence to a piece-wise linear curve. A simple and effective heuristics is that when packet-train length is sufficiently large, further increase in n only produces overlapping curves. We use an example to illustrate this manual identification process. Fig. 1(a) plots the rate-response

[3] For each packet, the trace file records the identification number and the input rate of the packet-train to which they belong, the sequence number of the packet within the train, and the arrival time of the packet.



(a) (b) (c)

Fig. 1. (a) Rate-response curves of the path from `cornell` to `cmu`; (b) Rate-response curves of the path from `roncluster5` to `speakeasy`; (c) Packet loss rate of the path from `roncluster5` to `speakeasy`, for a given input rate and packet sequence number x.

curves of the path from the Cornell host to the CMU host, measured on May 25, 2005. It is clear from the figure that when $n \geq 33$, the response curves overlap and remain piece-wise linear. There are only two linear segments observable in the plot, where the second linear segment falls into the input rate range between 96 mb/s to 150 mb/s. By conducting linear regression on this segment, we get $C_b = 96.32$ mb/s and $\lambda_b = 0.59$ mb/s. This result is partially confirmed by pathrate, which measures the minimum capacity (bottleneck bandwidth) of the path to be 96-99 mb/s. These results show that the available bandwidth of this particular path is limited by an empty Fast-Ethernet link, which is also the link with the minimum capacity (namely, the narrow link).

Even though the measured response curves for quite a large number of paths are similar to those shown in Fig. 1(a) in terms of smoothness and clear conformance to the theoretical prediction in [1], there are also paths whose measured response curves behave differently due to various measurement noises. We next briefly discuss one of the typical response curve anomalies caused by packet loss and the corresponding counter measures we take in our manual bandwidth estimation process.

Packet-loss is not considered in the theory [1], which assumes an infinite amount of buffer space at each hop. Among the 272 paths we measured, 168 paths (61%) experienced negligible packet loss ($\leq 1\%$), 56 paths (20%) suffered from moderate packet loss ($1\% \sim 10\%$), and 46 paths (17%) exhibited significant packet loss ($50\% \sim 86\%$) in the packet trains.

One example is the path from `roncluster5` to `speakeasy` measured on June 5, 2005. Fig. 1(b) shows the response curves of the path. This is a path with moderate overall probing packet loss (7%) where almost all lost packets belong to some packet-trains with an input rate above 50 mb/s and the lost packets within each train have sequence numbers greater than 65. Fig. 1(c) plots the conditional packet loss rate with respect to the input probing rates R_0 for several different packet sequence numbers. Notice that when the input probing rate is higher than 60 mb/s, the packets with sequence number $x = 129$ start to experience high loss rate (60% $\sim$ 90%). Recall that in our methodology, lost packets are simply ignored. This causes the measured response curves with $n = 129$ to fall below where it should be. Intuitively, by ignoring lost packets, a number of samples with *large* output dispersions are not taken into account in computing $E[G_N]$. This leads to an underestimation of $E[G_N]$ and also of the rate response $\tilde{\mathcal{Z}}(R_0)$, which is equal to $E[G_N]/G_0$.

For the path in Figure 1(b), there is only one linear segment in the whole probing range for all train lengths from 2 to 65. Using the curve with $n = 65$, we get $C_b = 1.36$ mb/s and $\lambda_b = 0$ mb/s. This clearly indicates that a link with capacity 1.36 mb/s and limited buffer space cannot accommodate long trains at high input rates. This leads to the perceived measurement packet loss that has a strong correlation with either the input rate R_0 or the packet sequence number in the probing train and causes the measured curve to fall below its theoretically predicted position. Our counter measure to this problem is to identify such curves and avoid using them in bandwidth estimation. This can usually be done by a visual examination.

We also examined the 46 paths with overall packet loss more than 50% and found that they all resemble the case we just showed. Specifically, there were 32 paths with the RON node `vineyard` as either the sending host or the receiving host. `Vineyard` has an incoming capacity of 1.53 mb/s [4] and an outgoing link capacity slightly smaller than that. The buffer space in both the incoming queue and the outgoing queue is quite small, which causes the probing packets with sequence numbers above 33 to experience frequent loss. The other 14 paths involve `lon1-gblx` as the receiving host, which has Ethernet access bandwidth with limited buffering in the incoming queue.

4 Major Findings

We now summarize several relevant findings regarding the properties of response curves and tight link bandwidth characteristics of the paths we measured.

4.1 Response Curve Properties

The first question we answer is *how large should packet train lengths be in order to converge the measured curves to their piece-wise linear lower bounds*? Our visual examination shows that the curves measured using train length $n \geq 33$ usually overlap with each other unless packet-loss-related anomalies occur. To provide a more accurate assessment, we pick the 168 paths where packet loss was negligible. For each path, we compute the tight link capacity using $n = 33$ and $n = 129$ and get two estimations C_1 and C_2. Fig. 2(a) plots the CDF of the metric $2|C_1 - C_2|/(C_1 + C_2)$, where we see that for over 80% of the 168 paths, the relative difference between C_1 and C_2 is less than 5%. This shows that a train of several dozen packets (with packet size $s = 1500$ bytes) is usually enough to guarantee response curve convergence and measurement accuracy.

[4] We explain later how to get this estimate and how to confirm its correctness.

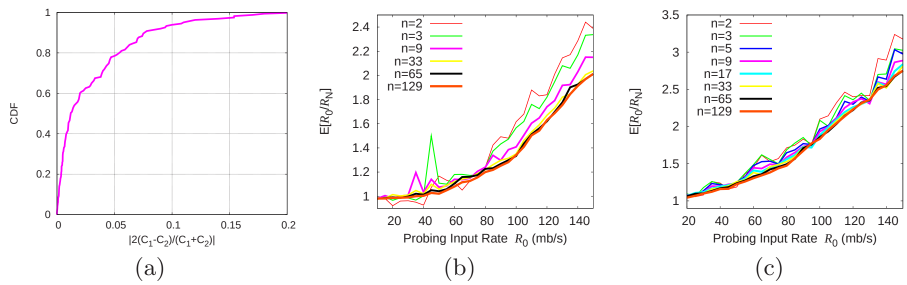


Fig. 2. (a) Relative difference between the measured capacities using $n = 33$ and $n = 129$; (b) Rate-response curves of the path from `ana1-gblx` to `cornell` on April 29, 2005; (c) Rate-response curves of the path from `ccicom` to `cornell` on April 18, 2005.

The second question of interest is *how many linear segments should one expect from the response curves measured using sufficiently long trains and how easily detectable are the second linear segments?* Among the 272 paths measured, we only have four paths where we can observe more than two linear segments in their response-curve lower bounds $\tilde{\mathcal{F}}(R_0)$. Fig. 2(b) and Fig. 2(c) show two such examples. The other two are `chi1-gblx`→`ana1-gblx` measured on April 13, 2005 and `chi1-gblx`→`ucsd` measured on June 3, 2005. The remaining 268 paths all had no more than two visible linear segments. It is also noteworthy that for a total number of 76 paths (mostly those involving `vineyard` and `lon1-gblx` as the sending or receiving host, or `speakeasy` as the receiving host), we can only see one segment (the second one) because the tight link capacity is smaller than the minimum input probing rate used in our measurement. Fig. 1(b) shows one such example.

If we assume that every link along a given path has different available bandwidth, then having only two linear segments in the response curve suggests that only the tight link can be congested by the probing traffic. However, it is very likely that a packet train with input rate $R_0 = 120$ mb/s can often congest at least the access links (which usually have capacity less than 100 mb/s) of both sending and receiving hosts. Hence, observing only two linear segments can have the following two possible explanations: 1) the second linear segment is too short to be detected and the one observed is a segment that corresponds to more than one link getting congested by the probing traffic; and 2) the third linear segment has a slope very close to the second one, which makes it difficult to differentiate between the two.

We indeed observed these two cases in our measurements. Fig. 3(a) shows the response curves of the path from `digitalwest` (with a T3 access capacity) to `cornell` (with a Fast-Ethernet access capacity). Although we can only observe two segments with a turning point at the input rate 32 mb/s, there are indeed three. The second turning point falls somewhere in the range $80 - 100$ mb/s. To confirm this, we apply linear regression on the input rate ranges $[40, 80]$ and $[100, 150]$, respectively, and we get two slightly different capacities 44.61 mb/s

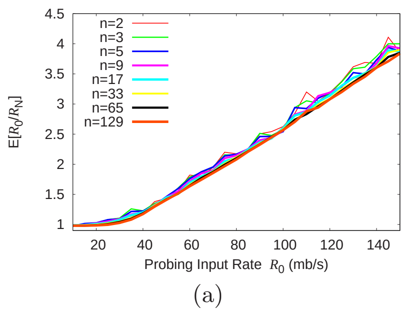


Fig. 3. (a) Rate-response curves of the path from `digitalwest` to `cornell` on June 6, 2005; (b) Rate-response curves of the path from `cornell` to `nyu` on May 27, 2005.

and 39.74 mb/s. The former is more accurate while the latter is slightly smaller than the true T3 link speed (taking into account the IP and UDP overhead). This is because the segment in the input rate range $[100, 150]$ corresponds to two congested links. It is easy to show that the slope difference of the two segments positively correlates to the link utilization of the second congested link. In this case, it is the access link of the `cornell` host. The low utilization of this link leads to a very small slope difference between the two linear segments. Such scenarios frequently happen when the access capacities of the two end-hosts are substantially different.

The second example in Fig. 3(b) shows the response curves of the path from `cornell` to `nyu`. Both hosts have a Fast-Ethernet access capacity with low but non-negligible utilization. Hence, the second segment becomes short and undetectable. The segment in the input rate range $[100, 150]$ corresponds to at least two congested links. Applying linear regression on this segment, we get $C = 89.74$ mb/s, which is somewhat smaller than the true Fast-Ethernet capacity. The two examples in Fig. 3 turn out to be the typical cases where identification of the proper segment becomes difficult and tight link capacities are often underestimated.

4.2 Tight Link Bandwidth Measurement

We have mentioned the access bandwidth of a number of nodes in previous sections. How did we get these results and why are we confident about their correctness? Note that we did *not* have a chance to confirm the access bandwidth of each node with site owners. Our confidence of these results comes from an examination of certain patterns from the *whole* set of measurement data. For example, we know that `speakeasy` has an incoming access capacity of 1.36 mb/s because all 16 paths with `speakeasy` as the receiving host have tight link capacity of 1.36 mb/s, as measured from their response curves. Similarly, `lon1-gblx` has $9.6 - 9.8$ mb/s bidirectional access speed, because all paths involving it have a tight link capacity of $9.6 - 9.8$ mb/s except when the other host is either `speakeasy` or

Table 2. Incoming and outgoing access bandwidth of RON nodes (mb/s)

Host	In	Out	Host	In	Out	Host	In	Out
ana1-gblx	100	100	ccicom	40-45	40-45	chi1-gblx	–	–
cmu	100	100	coloco	40-45	40-45	cornell	100	100
digitalwest	40-45	40-45	intel	12-13	12-13	lon1-gblx	10	10
lulea	100	100	mit-main	100	100	nyu	100	100
roncluster5	–	–	speakeasy	1.36	100	ucsd	100	100
utah	100	100	vineyard	1.53	1-1.5			

`vineyard`. Similar rules work for 262 paths with about ten exceptions. Four of them are the paths with a bottleneck at a non-access link. The rest are possibly because of access link changing at the end-hosts during our measurement. Table 2 provides the incoming and outgoing access bandwidths for all hosts except `roncluster5` and `chi1-gblx`, whose access capacities usually fall in the range from 80 to 100 mb/s and do not appear to be stable.

We provided the nominal values in Table 2 when we were sure of the fact that the access link type was Fast Ethernet or regular 10-mb/s Ethernet. The actual measured values for these two types of tight links were 96-100 and 9.6-9.8 mb/s, respectively. Also note that the access link was usually *not* the very first or last link along each path, but was frequently within 3 hops from the end host.

The major impediment of accurate bandwidth measurement of the tight link appears to be the case when both endpoints have similar access capacities and non-negligible utilization as exemplified by the path in Fig. 3(b). In such a case, the second linear segment is not detectable, which forces one to use the third segment in bandwidth estimation. This leads to an underestimation of both capacity C_b and cross-traffic intensity λ_b. However, this case did not happen frequently in our experiments meaning that usually at least one of the access links had utilization close to zero. Overall, we observed less than 10 paths whose Fast-Ethernet tight link capacities were underestimated to a value between $85-90$ mb/s. On the other hand, there were a total of 60 paths whose Fast-Ethernet tight-link capacities were accurately measured to fall between $95-100$ mb/s.

Since 98% (268/272) of the paths had a bottleneck at one of their access links, their tight links coincide with their narrow links. We hence can compare our measurements with those of pathrate. During this process, we found 39 paths with tight-link capacity $40-44$ mb/s (most likely a T3 link) as measured from the response curves. However, pathrate for some unknown reasons produces an estimate of $95-100$ mb/s along these particular paths. This happened for all paths from a host with Fast-Ethernet access link to a host with T3 access link and some paths from a host with a T3 access link to a host with a Fast-Ethernet access link. Apart from these paths and several others, our measurement results usually agree with pathrate very well. Fig. 4(a) shows the CDF of the relative difference between the two measurements for a total of 215 paths excluding the 39 paths we just mentioned and a handful of others where the tight link

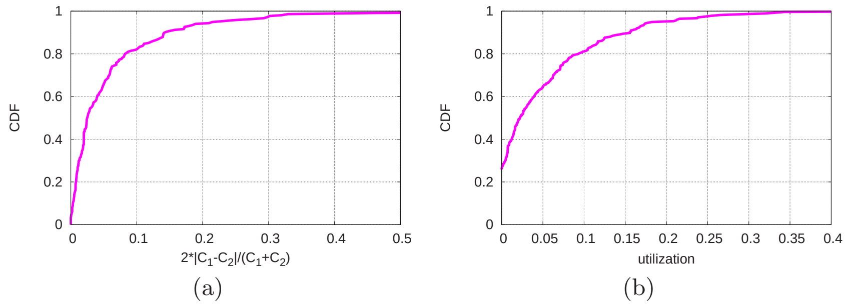


Fig. 4. (a) Relative difference between pathrate and our manual estimation for 215 paths; (b) Tight-link utilization for 256 paths.

apparently differs from the narrow link. We see that for more than 80% of the paths, the difference between the two techniques is within 10%.

Finally, Fig. 4(b) plots the CDF of tight-link utilization for 256 paths, which excludes a number of paths whose tight-link capacity and cross-traffic rates were obviously underestimated by our method (e.g., negative values of cross-traffic rate). We can see that close to 30% of the 256 paths had an empty tight link and produced curves similar to the one in Fig. 1(a). More than 60% of the paths were very lightly utilized ($\leq 5\%$) and more than 80% of the paths had utilization less than 10%.

5 Summary

In this paper, we presented a measurement study of probing-response curves for 272 paths over the RON testbed. Our results demonstrated the feasibility of measuring tight-link bandwidth characteristics (including capacity and utilization) from the corresponding response curves using moderate packet-train lengths (several dozen). This experience is necessary in the future development of an automated measurement tool. Our study also showed that tight links in the RON testbed mostly coincided with access links that experienced very little cross-traffic. Consequently, the path available bandwidth was frequently limited by the capacity of the access link.

References

1. Liu, X., Ravindran, K., Loguinov, D.: Multi-hop probing asymptotics in available bandwidth estimation: Stochastic analysis. In: USENIX/ACM IMC. (2005)
2. Dovrolis, C., Ramanathan, P., Moore, D.: Packet dispersion techniques and a capacity estimation methodology. IEEE/ACM Transactions on Networking **12** (2004) 963–977
3. RON tesbed. `http://nms.lcs.mit.edu/ron/`.