

A Combined Methodology for Measurement of Available Bandwidth and Link Capacity in Wired Packet Networks

Khondaker M. Salehin and Roberto Rojas-Cessa

Networking Research Laboratory

Department of Electrical and Computer Engineering

New Jersey Institute of Technology

Newark, New Jersey 07102

Email: {kms29, rojas}@njit.edu

Abstract

Accurate measurement of network parameters such as available bandwidth, link capacity, delay, packet loss and jitter are used to support and monitor several network functions, e.g., traffic engineering, Quality-of-Service (QoS) routing, end-to-end transport performance optimization, and link capacity planning. However, proactive network measurement schemes can impact both the data traffic and the measurement process itself, affecting the accuracy of the estimation if a significant amount of probe traffic is injected into the network. In this work, we propose two measurement schemes, one for measuring available bandwidth and the other for measuring link capacity, both of them use a combination of data probe packets and Internet Control Messaging Protocol (ICMP) packets. Our schemes perform available-bandwidth and link-capacity measurements in a short time and with a small amount of probe traffic. We show a performance study of our measurement schemes and compare their accuracy to those of other existing measurement schemes. We show that the proposed schemes achieve shorter convergence time than other existing schemes and high accuracy.

Keywords: Network measurement, available bandwidth, link capacity, joint queuing region.

1 INTRODUCTION

Traffic engineering statistically ensures efficient use of existing network infrastructure based on information of network state through accurate measurement. Unused link-capacity, also called available bandwidth (ABW) is one of the parameters used to determine the condition of the network for service providers to efficiently manage resources and services. ABW is the difference between the link capacity and the traffic load on that link at a given time. In this paper, we focus on available-bandwidth and link-capacity measurements. Various network measurement schemes have been recently proposed to measure individual network parameters, such as link capacity, available bandwidth, and delay, separately [1]. A few other schemes are aimed to measure multiple network parameters (e.g., link capacity, link latency and queuing delay), such as Pathchar [2, 3]. Estimation techniques can be categorized based on their different properties such as measurement process, probe pattern, and probing technique [1, 4, 5]. Furthermore, measurement schemes can be coarsely divided into passive and active methods. Passive measurement methods use ongoing data traffic flows through a measuring node to estimate the network characteristics. Measurement is only possible with the administrative control and the existence of network traffic on the link under measurement. The Multi Router Traffic Grapher (MRTG) [6] is an example of a passive

measurement tool. Active measurement methods proactively shoot probe packets from a source node towards a destination node to estimate various network parameters. In these methods, different probe packet characteristics, such as probe size, number of probes, and inter-probe gaps determine the features of the measurement. These probes add extra traffic into the network and may affect data traffic when the ABW is small. In addition to this, extra probe load can also affect the measurement process itself [7, 8]. To reduce the measurement time and probe load, this paper proposes two schemes: one scheme measures the ABW of a source-destination path, and the other measures per-hop link capacities and the narrow-link capacity (i.e., the smallest link capacity of an end-to-end path [9]). The paper presents a performance study in terms of accuracy and the amount of probe load of both schemes and shows the measurement limit of the link-capacity measurement.

The remainder of this paper is organized as follows. Section 2 discusses related work on the existing network measurement schemes. Section 3 introduces the proposed algorithms to measure ABW and link capacity. Section 4 presents the simulation results of the proposed schemes. Section 5 presents our conclusions.

2 RELATED WORK

Controllability of the active measurement (e.g., rate and size of probe packets) makes it an attractive alternative. Several proactive measurement schemes have been proposed [10]-[19]. Most of them are targeted to the measurement of ABW and link capacity. Proactive measurement schemes inject probe packets either in probe pairs or probe trains on a source-destination path and check the modification that the probes undergo to collect link-state information. ABW and link-capacity measurement schemes can be further categorized into either probe-gap or probe-rate models [1]. The probe-gap model is generally used to measure tight-link (i.e., the smallest available bandwidth of an end-to-end path [9]) capacity between end-to-end nodes through measurement of the dispersion of the initial gap of a probe pair at the destination. A probe train, consisting of multiple probe pairs, is injected into the network to get an estimate over a large period of time. IGI [7] and Spruce [14] are two examples of probe-gap model schemes that measure ABW in similar manner. IGI uses the probe-gap model with two conditions: that both the tight-link and the narrow-link of a path lie on the same hop (or single hop assumption) and that queues operate in a so-called joint queuing region (JQR). The queue operation region concept is elaborated after the following terms are introduced.

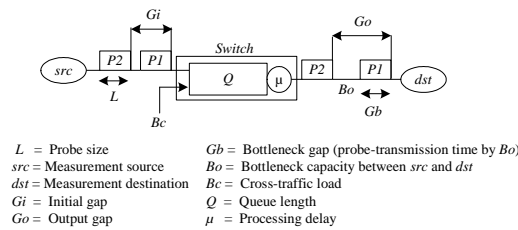


Figure 1: Single-hop model for probe-gap mechanism.

Figure 1 shows a single-hop model with parameters involved in the measurement process at an intermediate node (*Switch*) where the link in between *Switch* and the destination (*dst*) is considered as both the tight-link and narrow-link (i.e., bottleneck). Here, the source (*src*) shoots two fixed-size probe packets of L bytes each, $P1$ and $P2$, in pair with an initial

gap (G_i) towards the destination dst . The probe pair experiences a delay caused by the processing time, the transmission time, and the propagation time, on their way toward dst , which determines the output gap (G_o) at the destination. However, because of the variable nature of the processing time, including the burstiness and the computation capacity of network nodes, this value is not constant. The bottleneck capacity (B_o) and the cross-traffic load (B_c) shape up the final gap of the pair at the destination. As B_c is an unknown parameter, the change in the initial gap as the difference between G_i and G_o , is proportional to the amount of link capacity used by the cross traffic. Therefore, to notice an increase of G_o , the probe pair is required to operate under JQR, which is the presence of $P1$ in *Switch* (i.e., queue) during the arrival of $P2$. This is defined as the JQR operation. This condition ensures that G_i is only affected by B_c , giving place to the the initial gap at the destination (G_o). If $P2$ does not arrive in *Switch* before the departure of $P1$, the queuing operation is considered to be in a Disjoint Queuing Region (DQR). At DQR operation, G_i is not affected as there is no traffic in the network and therefore, the nodes solely process the probes.

In the gap model, Equations (1) and (2) are used to measure output gap and available bandwidth, respectively, as long as probe packets operate in JQR. Moreover, the gap model requires the knowledge of bottleneck capacity (B_o) in advance. G_o is calculated by measuring the effect the ratio of used bandwidth by the cross traffic B_c out of the bottle link capacity B_o has on the minimum probe gap, which is equal to the probe length G_b . This method requires knowledge of B_o and to ensure operation in JQR to perform the measurement of ABW. Therefore, ABW is calculated in function of the effect the ratio of the gap increase has on the bottleneck capacity B_o . The difference between the B_o and the effect determines ABW.

$$G_o = G_b + \frac{B_c * G_i}{B_o} \quad (1)$$

$$ABW = |B_o - B_o * \frac{G_o}{G_i}| \quad (2)$$

The probe-rate model is another model to measure ABW of an end-to-end path. In this model, probe packets are sent toward a destination node at a controlled transmission rate. Generally, the source node sends out a long train of probe packets with an initial rate and adjusts the rate during each iteration until the ending rate equals the receiving rate at the destination node. The determination of the probe rate in the rate model is not explicitly dependent on a particular queuing region operation as the gap model and the probe rates are adjusted by two simple parameters: probe size and inter-probe gap length (i.e., the transmission rate is determined as $R=L/T$ bps, where T is the time period). The same process iterates and adjusts T , and sends probe packets at different rates. However, this scheme must conform with the packet specification size of the lower layer (e.g., Maximum Transmission Unit, MTU). Such a restriction defines the lowest possible values of available bandwidth that can be estimated with this method. This model does not require prior knowledge of the bottleneck capacity to measure ABW.

PTR [7], Pathload [15], and TOPP [13] are examples of schemes that measure available bandwidth based on the rate model. PTR and TOPP use a generic probe-rate model, and Pathload and Self-Loading Periodic Streams (SLoPS) [18] use an additional mechanism with the rate model to compare the transmitting and receiving probe rates by computing One Way Delay (OWD) at the destination. Pathload uses a binary search algorithm [20] to adjust the probe rate at each iteration and determines a range of the ABW values as a final outcome unlike all other ABW measurement schemes that

provide a single ABW value. Bprobe [10], and Pathrate [11], are proactive schemes that measure link capacity. In Bprobe, the sender sends out a sequence of Internet Control Messaging Protocol (ICMP) ECHO packets to a destination node and waits for the replies to measure the inter-arrival time of the consecutive probe packets. In this scheme, it is considered that the dispersion between a probe pair is inversely proportional to the narrow-link capacity where probe pair dispersion with JQR operation, no packet loss, no intervening cross traffic, and no congestion can represent the actual link capacity. Bprobe applies either an intersection-set or a union-set operation to filter out the probe gap errors (i.e., error caused by DQR operation, cross-traffic, and congestion) to determine the minimum output gap for estimating the link capacity. IGI uses Bprobe for bottleneck capacity estimation to measure ABW [7].

3 PROPOSED SCHEMES FOR THE MEASUREMENT OF AVAILABLE BANDWIDTH AND LINK CAPACITY

In this section, we introduce two new measurement schemes, one is for estimating ABW and the other is for estimating link capacity along a source-destination path.

3.1 Available-bandwidth Measurement Scheme

The proposed ABW measurement scheme iteratively shoots two trains of a fixed number of probes from a source node to a destination node at a specific transmission rate and checks the change of probe train size at the destination to estimate packet transmission capacity. The change in probe train size is the cumulative gap difference of the probe pairs at the source and destination ends. The change in the cumulative gap at the destination (i.e., the output gap) is affected by the tight-link capacity during the estimation period. If the probe transmission rate is lower than or equal to the ABW, no change in the cumulative gap occurs [18]. To determine the probe rate at each iteration, a ternary search algorithm [21] is adopted. This search algorithm sends two probe trains with two different transmission rates, during each iteration. These two probing rate values are one third and two thirds of the difference between the minimum and maximum probe rate limits of the algorithm. Figure 2 shows the flowchart for the proposed ABW measurement scheme.

Figure 2 shows a flow chart that describes the ABW measurement scheme. In this scheme, the source node shoots two probe trains towards the destination at two different transmission rates, i.e., g -rate and h -rate. The source node computes the cumulative gaps of the trains affected at the destination, i.e., g -outgap and h -outgap, upon receiving the probes' replies, and compares them with the cumulative gap at the source, i.e., g -ingap and h -ingap, respectively. The source determines the next probing rates in the following iteration based on the change of g -increment (i.e., g -increment = g -outgap - g -ingap) and h -increment (i.e., h -increment = h -outgap - h -ingap). The measurement process terminates by providing an ABW value when either the g -increment or h -increment rates have 1%-1.5% of change or when the g -rate and h -rate difference is smaller than or equal to 1.5 Mbps. Here, the selected minimum measurable ABW is 0.5Mbps. This minimum is selected according to the complexity to comply with the termination condition of the search scheme and in function of the expected link capacity. Here we chose to adopt 2.5% of the link capacity (e.g. 20 Mbps) for the considered simulations examples. The maximum value is defined by the tight-link capacity of the path.

The proposed scheme is an improvement over both the IGI and Pathload methodologies concerning the probe rate adjustment technique (i.e., the methodology to determine probe transmission rate during each iteration) and the way the

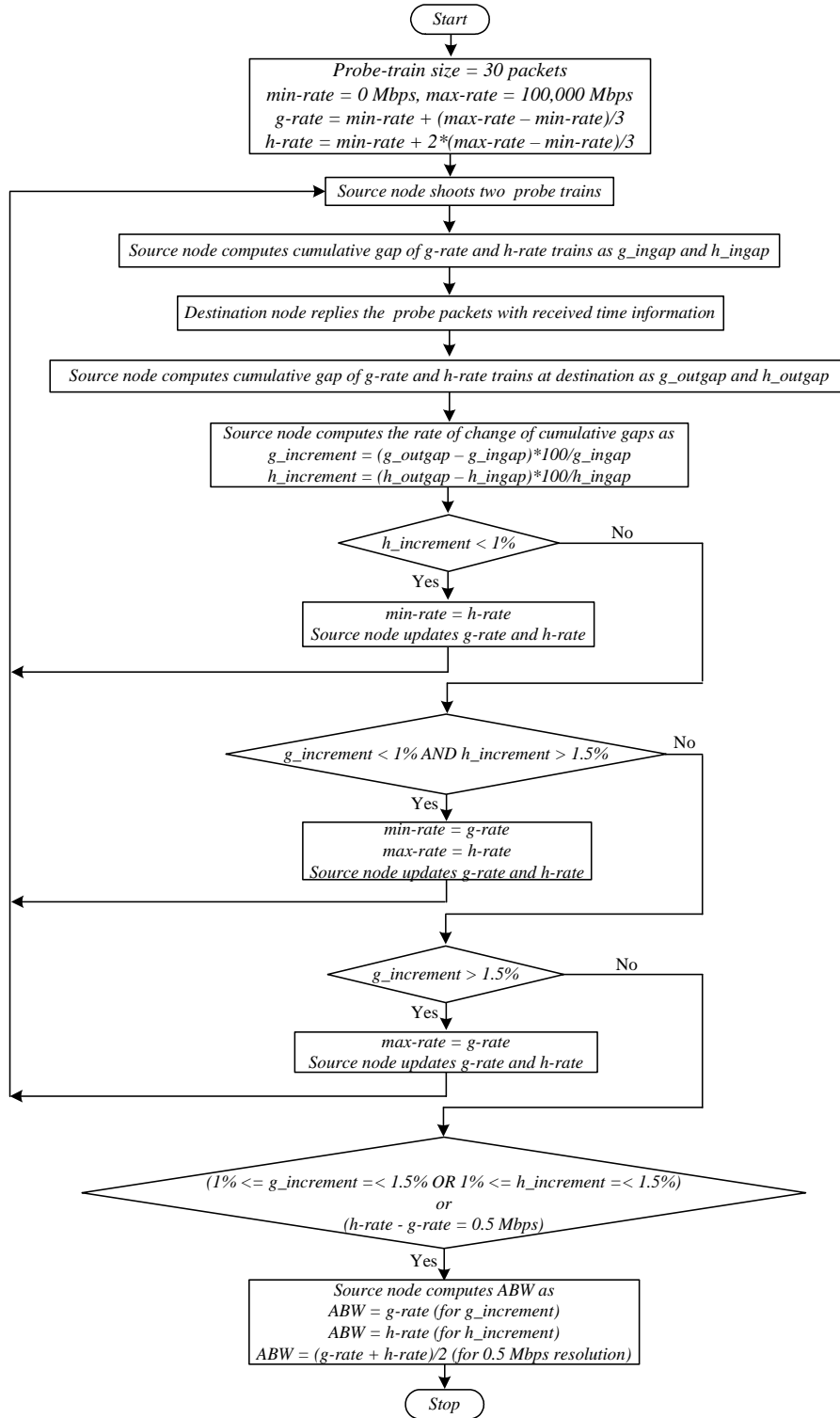


Figure 2: Proposed ABW measurement algorithm.

probes' delay (i.e., cumulative output gap or probes' OWD measurement) is estimated. The proposed scheme compares the probe train's gap values at the source and destination for adjusting the next probing rates. Therefore, this gap model scheme resembles a probe-rate model and achieves high accuracy. Figure 3 illustrates how the hardware limitation affects the accuracy of probe-gap model schemes.

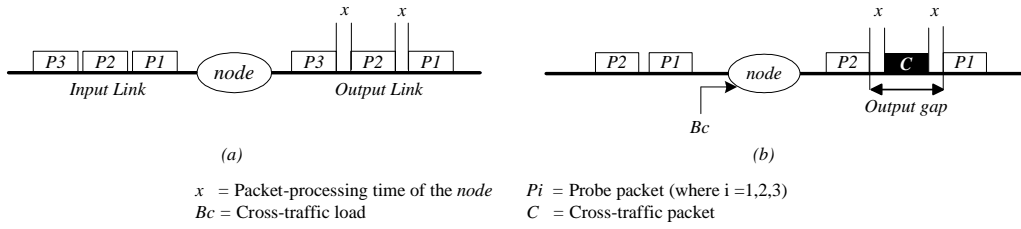


Figure 3: (a) Packet dispersion at the output link and (b) output gap in JQR operation created by a cross-traffic packet.

Figure 3.(a) shows an example of three consecutive probe packets that experience dispersion at the output link in the absence of cross traffic. The gaps in between the probe packets (i.e., P_1 , P_2 , and P_3) on the output link are directly dependent on the switching speed of that node, which can be in the range of 10 to 40 microseconds, depending on the node operation level (i.e., kernel level or user-mode level [7]). The switching speed of a node (e.g., Internet router or packet switch) is limited by hardware and it is assumed to be constant [22]. A minimum processing time of 40 microseconds is then assumed for the remainder of this paper.

Figure 3.(b) illustrates the additional packet processing time, x , in between P_1 and P_2 , required by the node at the output link and induced by the cross-traffic packet, C . The probe-gap model overestimates the cross-traffic load calculation due to the extra gap induced by the cross traffic. Moreover, the output gap increases as the cross-traffic load increases. As shown in Figure 3.(b), the gap between probes (P_1 and P_2) proportionately increases as $(n+1) * \text{packet-processing time}$ for each cross-traffic packet found within a probe pair, where n refers to the number of cross-traffic packets between P_1 and P_2 .

3.2 Link-capacity Measurement Scheme

The proposed scheme to measure link capacity uses a probe train of novel compound probes separated by an inter-probe gap, which is determined by the ABW of the source-destination path. The proposed scheme measures link capacity in a hop-by-hop manner and estimates the narrow-link capacity of the path from the link capacities measured from source to destination. Figure 4 shows a compound probe structure.

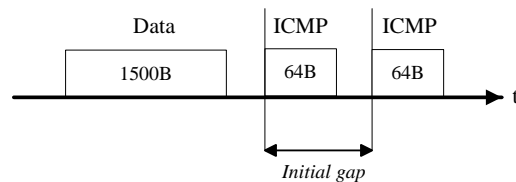


Figure 4: Structure of a compound probe.

As shown in Figure 4, there are two ICMP timestamp request packets following a 1500B dummy data-packet. The selection of the 1500B length follows the maximum transmission unit that a data-link layer protocol (e.g., Ethernet). The set of these three packets is named a compound probe. The initial gap (i.e., the gap between the first bits of any of the two

probes in the probe structure) is 40 microseconds. At each destination, the 1500B data-packets are discarded by using an appropriate Time-to-Live (TTL) value. The 1500B data-packet in the compound probe ensures that the two ICMP packets operate in the JQR. Figure 5 shows the flowchart for the proposed link-capacity measurement scheme.

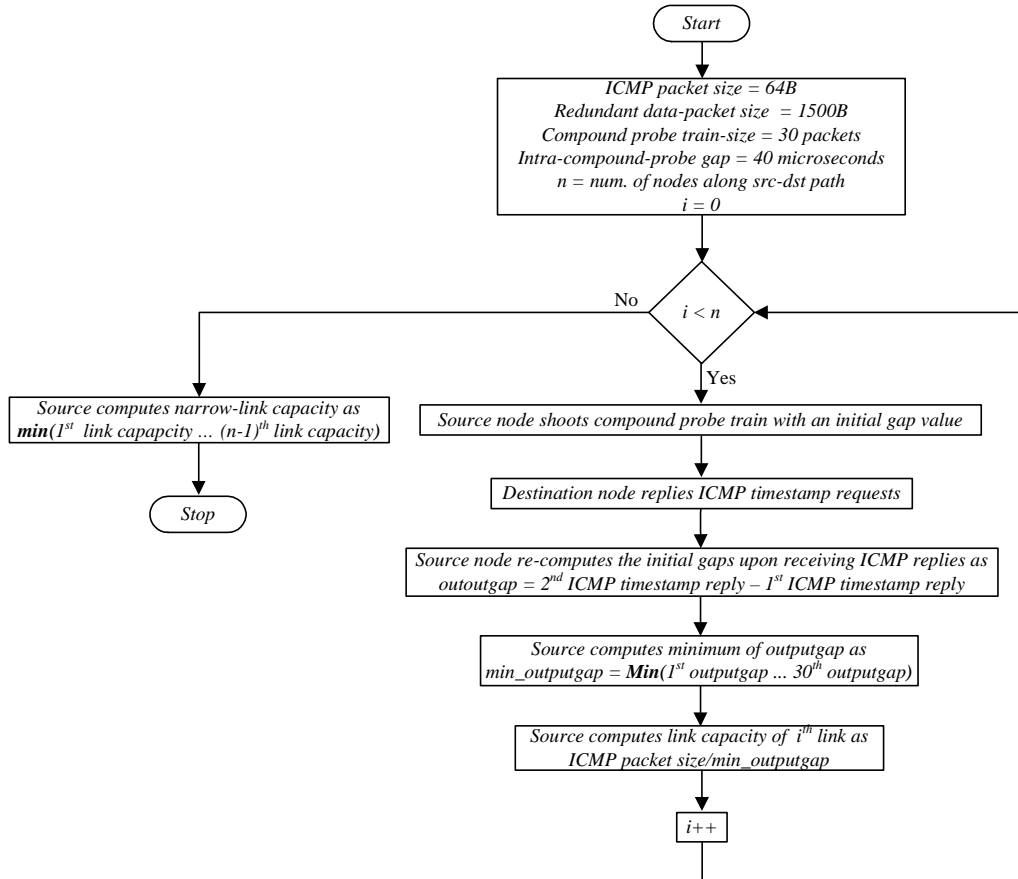


Figure 5: Proposed link-capacity measurement algorithm.

Figure 5 shows that the source node shoots a 30-packet train of compound probes with 40 microseconds of initial gaps toward each intermediate node until it reaches the destination node of the path. For link-capacity measurement, the source node waits for the ICMP timestamp replies from the destination nodes and re-computes the initial gap at the destination (i.e., *outputgap*) upon receiving the timestamp values. The minimum *outputgap* of the 30 compound probe packets is used to measure link-capacity for each hop using the packet size of an ICMP probe.

4 SIMULATION RESULTS

4.1 ABW Results

The proposed ABW scheme and two other comparable schemes, IGI and Pathload, were simulated in the ns2 network simulator [23] using cross traffic, which was modeled as constant bit rate (CBR) traffic. Previous works [24]-[26] proved that the stationary trend of traffic (i.e., constant rate of traffic load) over Internet paths for a long interval makes network measurement possible over long periods of time. Moreover, IGI and other schemes [27] also adopt the CBR traffic model. We follow this practice.

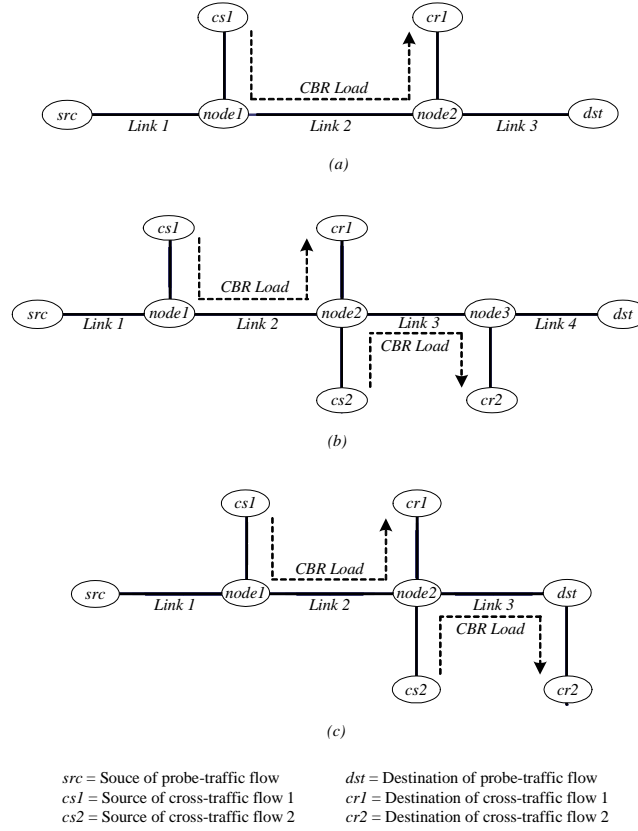


Figure 6: (a) Single-hop topology, (b) and (c) multiple-hop topologies.

Figure 6.(a) shows the single-hop scenario used in the simulation. In this figure, the *src* node shoots probe packets to the *dst* node via *node1* and *node2*. The traffic between *cs1* and *cr1* is the cross-traffic load. We used different CBR load values on *Link2* and keep the other two links with no load. Here, *Link2* represents both the narrow-link and the tight-link of this topology. We also tested the proposed scheme in two multiple-hop scenarios as shown in Figure 6.(b). These topologies have multiple cross-traffic sources on different nodes. In this case, the cross-traffic generated in different links provides possibilities for tight-link to appear in any hop of the network rather than on a single link only.

4.1.1 Probe-train Size

We investigated how the accuracy is affected by the probe-train size for ABW measurement using the single-hop scenario in Figure 6.(a). Table 1 shows the measurement results for probe-train sizes of 30, 60, and 100 packets. The results show that probe-train sizes of 30 and 60 packets produce similar accuracy, while the 100-probe packet train overestimates the measurement for small ABW. The reason is that the higher rate of the probes negatively affects the cross traffic along the source-destination path. Therefore, a probe-train size of 30 packets is adopted in the proposed scheme to minimize probe load and intrusion rate. The probe-train size for IGI and Pathload are 60 and 100 packets, respectively. For comparison purposes, we adopted the packet size that IGI uses, or 800B. Furthermore, packet sizes of 1000B and 1500B were tested and the results indicate that 800B are sufficient to provide a JQR, without inducing a high intrusion rate [7], [28].

<i>Ternary search, Topology: 6.(a), Probe packet size: 800B, Train size: 30, Iterations: 100</i>				
(Mbps)		Measured ABW(Mbps)		
Link Capacity (1,2,3)	Load (Link2)	Train Size 30	Train Size 60	Train Size 100
20	19	1.28	1.31	1.31
20	18	2.16	2.23	2.23
20	17	3.13	3.19	3.23
20	16	4.05	4.08	4.16
20	15	5.05	5.11	5.14
20	14	6.08	6.21	6.18
20	13	6.92	7.11	7.12
20	12	7.94	8.12	8.17
20	11	8.87	9.07	9.14
20	10	9.84	9.98	10.12
20	9	10.76	11.07	11.20
20	8	11.70	12.06	12.10
20	6	13.41	13.92	14.12
20	5	14.67	15.03	15.14
20	4	15.47	16.03	16.15

Table 1: ABW measurement using single-hop topology with different probe-train sizes.

4.1.2 ABW Measurement Accuracy

Table 2 shows the ABW measurement results of the proposed scheme for the multiple-hop scenario in Figure 6.(a). We tested this scenario with different combinations of cross-traffic load values. In this paper, the error is the percentage error in reference from the actual value, which is calculated as the ratio of the difference between the actual value and the measured value and the actual value. For the proposed scheme, the error is below 10%, even under high load conditions. As an example, for an ABW of 16 Mbps, the measured output is 15 Mbps. This is an error of 6.25%.

<i>Ternary search, Topology: 6.(b), Probe packet size: 800B, Train size: 30, Iterations: 25</i>				
Link (1,2,3,4)	Load (Link2)	Load (Link3)	Actual ABW	Measured ABW
(Mbps)	(Mbps)	(Mbps)	(Mbps)	(Mbps)
20	1	1	19	17.0
20	1	2	18	16.5
20	1	3	17	16.0
20	2	3	17	16.0
20	4	2	16	15.0
20	4	3	16	15.0
20	5	3	15	14.0
20	5	4	15	14.0
20	6	4	14	13.0
20	5	6	14	13.0
20	5	7	13	13.0
20	10	3	10	10.0
20	4	10	10	10.0
20	11	5	9	9.0
20	13	12	7	7.0
20	14	10	6	6.0
20	15	4	5	5.0
20	16	10	4	4.0

Table 2: ABW measurement using a multiple-hop topology.

Figure 7 shows an accuracy comparison graph of IGI, Pathload, and the proposed scheme on the single-hop scenario with

200 Mbps of link capacity for each link. Here, the x-axis denotes the cross-traffic load during the measurement process and the y-axis denotes the measured ABW. The values shown in this graph were obtained from 25 measurement iterations per ABW value for each of the three schemes.

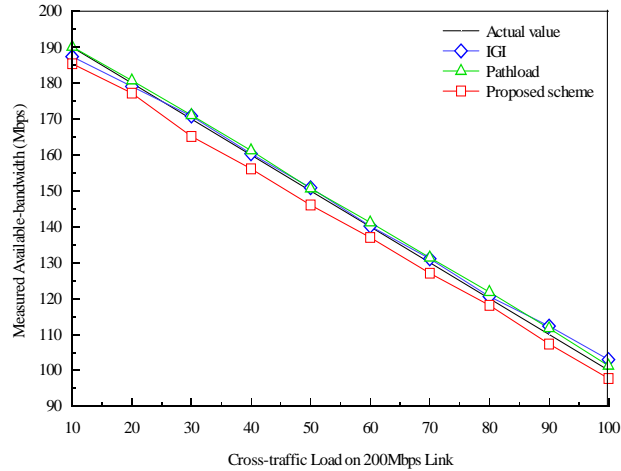


Figure 7: Accuracy comparison of different ABW measurement schemes.

The simulation results for the proposed scheme show an error smaller than or equal to 3% in every case. Even though IGI and Pathload deliver higher accuracy (i.e., smaller error) than the proposed scheme, the output of the proposed scheme is comparable to the other two schemes. Moreover, the proposed scheme has the shortest estimation time and does not require prior knowledge of the bottleneck capacity.

4.1.3 Measurement Time and Probe Load

We investigated the measurement time and overhead of the proposed scheme and compared them to those of IGI, which is considered the fastest existing ABW measurement scheme [7]. Although the proposed scheme requires 30-packet probe trains for optimal performance, train sizes of 60 packets each were used in this experiment for a fair comparison with IGI. Both schemes shoot 800B probe packets and terminate the simulation after 1 measurement iteration with similar termination conditions (i.e., IGI and the proposed scheme terminate with 1% and 1%-1.5% resolution, respectively). Figure 8.(a) shows that the proposed scheme has a smaller measurement time in 9 out of 15 instances where the difference of time is significant in most cases. This result validates the faster measurement capability of the proposed scheme when compared to IGI. In this simulation, the proposed scheme has a mean measurement time and standard deviation of 1.57 sec and 0.43 sec, respectively.

Figure 8.(b) shows the comparison of the number of probes of IGI and that of the proposed scheme in the same experiment. This graph shows that the proposed scheme generates a smaller number of probe packets than IGI to achieve comparable measurement results. In general, the proposed scheme requires at least 984 fewer probe packets than IGI for every case. In addition, IGI requires the value of the link capacity to measure ABW. Moreover, IGI shoots probe packets more frequently than the proposed scheme as its initial gap increases with small increments, $Gb/8$ [7], and this makes IGI perform a large number of iterations in a measurement process. On the other hand, the probe rate adjustment step value for the proposed

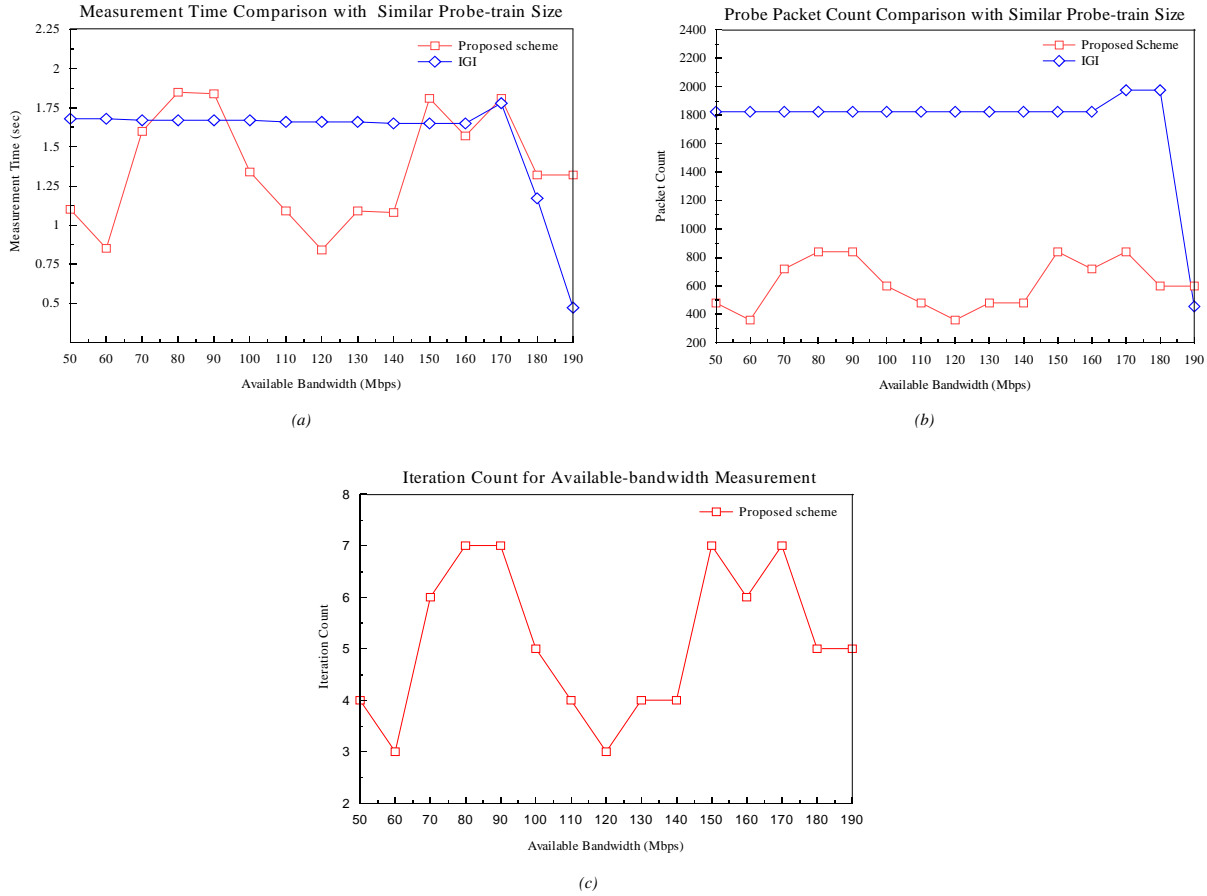


Figure 8: (a) Measurement time of the proposed scheme and IGI, (b) probe load of the proposed scheme and IGI, and (c) iteration counts of the proposed scheme.

scheme is larger than that of IGI as this is determined by the minimum and maximum probing rates at each iteration.

Figure 8.(c) shows that the number of iterations required for measurement with the proposed scheme is 6. In ternary search, there is a measurement window defined by the *min-rate* and *max-rate* values. The location of the window and the variation of the number of iterations for each ABW value contribute to the fluctuation in the proposed measurement scheme. More precisely, the inter-probe gaps in the two probe trains depend on the current *g-rate* and *h-rate* values, where the cumulative times for sending the probe packets of the two probe trains vary during each iteration. As the location of the window is different for every iteration, based on the capacity and load conditions, and an additional iteration requires sending two more probe trains with different rates, these fluctuations are, therefore, expected.

4.1.4 Average Measurement Accuracy

Figure 9 shows the accuracy for different number of iterations of the proposed scheme. This graph shows that the proposed scheme has similar accuracy for both small and large number of measurement iterations. It also shows that this scheme can measure ABW with high accuracy in short time and a small number of iterations.

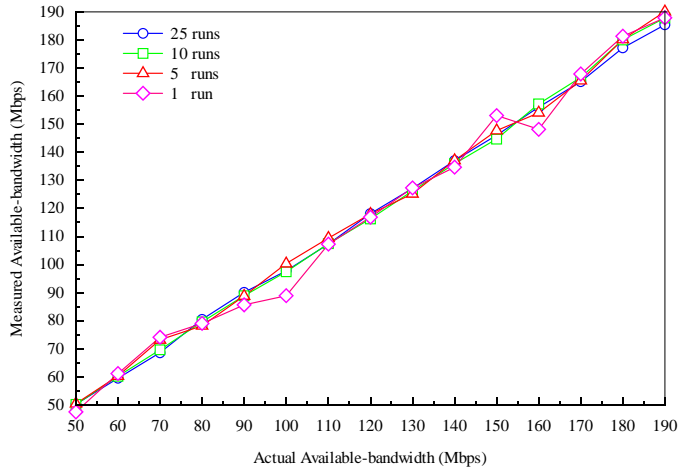


Figure 9: Accuracy comparison of the proposed scheme with different number of iterations.

4.2 Link-capacity Results

We tested the proposed link-capacity measurement scheme in a multiple-hop scenario, as shown in Figure 8, to validate accuracy in different network conditions.

4.2.1 Link-capacity Measurement

Table 3 shows the results for the multiple-hop scenario in Figure 6.c. The results show that the proposed scheme for measuring link capacity has high accuracy for links with low and high capacities (e.g., 15 Mbps and 150 Mbps, respectively), and with high cross-traffic loads (e.g., 12 Mbps over a 20 Mbps link). In every case, the measurement error is below 1%. The results also show that the amount of cross-traffic does not affect the link capacity and narrow-link estimation processes. The resolution for the link-capacity measurement scheme is finer than that of the ABW measurement scheme as these schemes use different probing techniques.

4.2.2 Impact of Redundant Data-packet

To show the effect of the redundant data-packet, we performed an experiment with 2-packet (without the 1500B packet) and 3-packet (with the 1500B packet) compound probes. According to the results shown in the left columns of Table 4, the measurement outcomes have low accuracy for small link capacities (e.g., 12 Mbps) with 2-probe compound probes without cross-traffic load. However, with 3-packet compound probes, the accuracy approaches 100%, as shown on the right columns of Table 4.

Topology: 6.(c), Compound probe size: 64+64+1500 Bytes, Iterations: 10									
Actual Capacity (Mbps)			Load (Mbps)	Measured Link Capacity (Mbps)			Error (%)		
Link1	Link2	Link3	Link2+Link3	Link1	Link2	Link3	Link1	Link2	Link3
20	15	20	10+10	19.99	14.99	20.00	0.050	0.066	0
20	15	20	12+12	19.99	14.99	20.00	0.050	0.066	0
20	15	20	14+14	20.00	15.00	20.00	0	0	0
20	20	20	10+10	19.99	19.99	20.00	0.050	0.050	0
20	20	20	12+12	19.99	20.00	20.00	0.050	0	0
20	20	20	14+14	19.99	20.00	20.00	0.050	0	0
20	20	20	16+16	20.00	20.00	20.00	0	0	0
20	20	20	18+18	20.00	20.00	20.00	0	0	0
20	150	20	6+6	19.99	149.99	20.00	0.050	0.006	0
20	150	20	8+8	19.99	149.99	20.00	0.050	0.006	0
20	150	20	10+10	19.99	149.99	20.00	0.050	0.006	0
20	150	20	12+12	19.99	149.99	20.00	0.050	0.006	0

Table 3: Hop-by-hop link-capacity measurement using a multiple-hop topology.

No cross-traffic, Topology: 6.(c), Iterations: 10														
Actual Link Capacity (Mbps)			Compound probe structure: 64B+64B						Compound probe structure: 1500B+64B+64B					
			Measurement (Mbps)			Error (%)			Measurement (Mbps)			Error (%)		
Link1	Link2	Link3	Link1	Link2	Link3	Link1	Link2	Link3	Link1	Link2	Link3	Link1	Link2	Link3
20	20	20	12.79	13.51	13.52	36.05	32.45	32.45	19.99	19.99	20.00	0.05	0.05	0
20	18	20	12.79	13.31	13.32	36.05	26.05	33.40	19.99	17.99	20.00	0.05	0.05	0
20	16	20	12.79	13.11	13.11	36.05	18.06	34.45	19.99	15.99	19.99	0.05	0.06	0.05
20	14	20	12.79	12.91	12.91	36.05	7.78	35.45	19.99	13.99	19.99	0.05	0.01	0.05
20	12	20	12.79	11.99	11.99	36.05	0.08	40.05	19.99	12.00	19.99	0.05	0	0.05
20	10	20	12.79	10.00	10.00	36.05	0	50	19.99	10.00	19.99	0.05	0	0.05
20	8	20	12.79	7.99	7.99	36.05	0.12	60.05	19.99	7.99	19.99	0.05	0.12	0.05
20	6	20	12.79	6.00	6.00	36.05	0	70.00	19.99	5.99	19.99	0.05	0.16	0.05
20	4	20	12.79	3.99	3.99	36.05	0.25	80.05	19.99	3.99	19.99	0.05	0.25	0.05
20	2	20	12.79	1.99	1.99	36.05	0.50	90.05	19.99	2.00	20.00	0.05	0	0

Table 4: Hop-by-hop link-capacity measurement using 2-packet and 3-packet compound probes in a single-hop topology.

4.2.3 Redundant Data-packet Dependency in Link-capacity Measurement and Scalability

Due to heterogeneities of links in the Internet, the 1500B data-packet might not be sufficient for all link capacities. Therefore, the proposed scheme has an upper limit for link-capacity measurement. This limitation is also dependent on the nodes' packet processing time. In the proposed scheme, a maximum processing time of 40 microseconds (in user-level operation mode) have been considered for intra-compound-probe gap in link-capacity measurement. Therefore, a link capacity of an output link in a node that transmits a 1500B data-packet earlier than the intra-compound-probe gap would be infeasible due to DQR operation for the proposed scheme to measure accurately. As such, determining the upper measurement limit for a 1500B packet is required in this scheme.

A single-link topology was used to determine dependency on the redundant data-packet. Table 5 shows the measurement results obtained by using compound probes with the 1500B redundant packet with and without cross-traffic loads. In both load conditions, the measurement process shows low accuracy for link capacities beyond 300 Mbps. This limitation value follows the theoretical upper measurement limit determined by the use of a 1500B redundant packet and an intra-probe gaps of 40 microseconds. Therefore, the upper limit of the link-capacity measurement, called the Maximum Measurable

Link-capacity measurement limitation test with 1500B redundant data packet					
No-load Scenario			Load Scenario		
Actual value (Mbps)	Measured value (Mbps)	Error (%)	Actual value Mbps	Measured value Mbps	Error (%)
40	40	0.000	40	40	0
80	79.99	0.012	80	79.99	0.012
100	99.99	0.010	100	99.99	0.010
150	150	0	150	150	0
200	200	0	200	200	0
250	249.55	0.180	250	249.99	0.003
295	294.99	0.003	295	294.99	0.003
296	295.99	0.003	296	295.99	0.003
297	297	0	297	297	0
298	297.99	0.003	298	297.99	0.003
299	299	0	299	299	0
299.5	299.49	0.003	299.5	299.49	0.003
300	299.99	0.003	300	299.99	0.003
300.5	289.2	3.760	300.5	293.72	2.250
301	279.18	7.249	301	294.45	2.176
302	261.18	13.516	302	285.67	5.407
303	245.46	18.990	303	279.98	7.597
304	231.61	23.812	304	267.8	11.907
305	219.32	28.809	305	262.16	14.045
310	174.03	43.861	310	269.21	13.158

Table 5: Maximum measurable link capacity under different load conditions.

Capacity (MMC), for the proposed scheme is:

$$MMC = \frac{\text{Packet length (bits)}}{\text{Initial gap (sec)}} = \frac{1500B * 8}{0.00004 \text{ sec}} = 300Mbps \quad (3)$$

In case the link capacities are larger than the MMC , the source node probes the measured link with a redundant data-packet size associated with the first measurement outcome. If the second outcome equals the first measurement value, the measured link capacity is accepted and is within the range marked by the MMC . Otherwise, the link capacity is declared over the range indicated by the MMC . For example, consider that the actual link capacity is 310 Mbps. In Table 5, the measurement value for the 310 Mbps with load condition is 269.21 Mbps. Hence, the source node would probe the same link with a 1350B (using $L=R*T$, where $R = 269.21$ Mbps and $T = 0.00004$ sec) redundant data-packet in the compound probes. Any measurement value with the 1350B data-packet except 269.21 Mbps would prove that the MMC range has been exceeded. Thus, the link-capacity measurement would be discarded from consideration.

5 CONCLUSIONS

Network measurement is a challenging task considering the current growth of networks and applications, and the high data rates on broadband links. This paper proposed a scheme for the measurement of ABW and another for the measurement of link capacity. These schemes can be bundled up as a tool kit. The ABW measurement scheme adopts a ternary search algorithm for measurement of available bandwidth. This scheme was simulated in ns2 in both single and multiple-hop scenarios to compare its accuracy against IGI and Pathload schemes. The new link-capacity measurement scheme uses a new probe-packet structure, which we have named compound probe. This probe structure has the potential to measure both per-hop link capacities and the narrow-link capacity accurately, regardless of the cross-traffic condition

on the end-to-end path. A relationship between the redundant data-packet length of the compound probe structure and highest possible link capacity measurement is also shown. The proposed schemes show high accuracy for measuring link capacity and an acceptable accuracy for estimating ABW with a short convergence time and low probe load. The two schemes can be used to evaluate link states in single- and multiple-hop paths for wirelined networks.

Acknowledgement

The authors are thankful to the anonymous reviewers for their insightful comments that helped to improve this paper.

References

- [1] Prasad, R., Murray, M., Dovrolis, C., and Claffy, K.: 'Bandwidth estimation: Metrics, measurement techniques, and tools', *IEEE Network*, 2003, 17, (6), pp. 27–35
- [2] Jacobson, V.: 'Pathchar - A tool to infer characteristics of Internet paths'. Available from <ftp://ftp.ee.ebl.gov/pathchar>, April 1997
- [3] Downey, A.: 'Using Pathchar to estimate Internet link characteristics', *Proc. ACM SIGCOMM*, Cambridge, USA, August 1999, pp. 241–250
- [4] Anagnostakis, K.G., and Greenwald, M.B.: 'Direct measurement vs indirect inference for determining network internal delays', *Performance Evaluation*, 2002, 49, (1-4), pp. 165–177
- [5] Johnsson, A.: 'On the comparison of packet pair and packet train measurement', *Proc. Swedish National Computer Networking Workshop*, Stockholm, Sweden, September 2003, pp. 241–250
- [6] <http://people.ee.ethz.ch/oetiker/webtools/mrtg>, accessed 2007
- [7] Hu, N., and Steenkiste, P.: 'Evaluation and characterization of available bandwidth probing techniques', *IEEE JSAC Special Issue in Internet and WWW Measurement, Mapping, and Modeling*, 2003, 21, (6), pp. 879–864
- [8] Timotijevic, T., Leung, C., and Schormans, J.: 'Accuracy of measurement techniques supporting QoS in packet based Intranet and Extranet VPNs', *IEE Proc. Communications Special Edition on VPNs and Broadband Technology*, 151, February 2004, pp. 89–94
- [9] Defining network capacity (RFC). Available: <http://www.ietf.org/rfc/rfc5136.txt>
- [10] Carter, R. and Crovella, M.: 'Measuring bottleneck link speed in packet-switched networks', *Performance Evaluation*, 1996, 27-28, pp. 297–318
- [11] Dovrolis, C., Ramanathan, R., and Moore, D.: 'What do packet dispersion techniques measure?', *Proc. IEEE INFOCOM*, Anchorage, USA, 2001, pp. 905–914
- [12] Pasztor, A., and Veitch, D.: 'Active probing using packet quartets', *Proc. IEEE INFOCOM*, Marseille, France, November 2002, pp. 293–305

- [13] Melander, B., Bjorkmann, M., and Gunningberg, P.: 'A new end-to-end probing and analysis method for estimating bandwidth bottlenecks', Proc. IEEE GLOBECOMM - Global Internet Symp., San Francisco, USA, November 2002, pp. 415–420
- [14] Strauss, J., Katabi, D., and Kaashoek, J.: 'A measurement study of available bandwidth estimation tool', Proc. ACM SIGCOMM Internet Measurement Conference, Miami, USA, October 2003, pp. 39–44
- [15] Jain, M., and Dovrolis, C.: 'Pathload: A measurement tool for end-to-end available bandwidth', Proc. Passive and Active Measurement, Colorado, USA, March 2002, pp. 14–25
- [16] A one-way active measurement protocol (RFC). Available: <http://www.ietf.org/rfc/rfc4656.txt>
- [17] Anagnostakis, K.G., Greenwald, M.B., and Ryger, R.S.: 'cing: Measuring network-internal delays using only existing infrastructure', Proc. IEEE INFOCOM, San Francisco, USA, March-April 2003, pp. 2112–2121
- [18] Jain, M., and Dovrolis, C.: 'End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput', Comp. Comm. Rev., 2002, 32, (4), pp. 295–308
- [19] Hu, N., Li, L., Mao, Z., Steekiste, P., and Wang, J.: 'Locating Internet bottlenecks: Algorithms, measurements, and implications', Comp. Comm. Rev., 2004, 34, (4), pp. 41–54
- [20] <http://www.nist.gov/dads/HTML/binarySearch.html>, accessed May 2007
- [21] Hesselink, W.H.: 'Ternary Search'. Available from <http://www.cs.rug.nl/wim/pub/whh303.pdf>
- [22] De Schepper, B., Steyart, B., Wittevrongel, S., Moeneclaeys, M., Bruneel, H.: 'Constant hardware delay in integrated switching elements with multiserver output queues', IEE Proc. Communications, 153, October 2006, pp. 664–670
- [23] <http://www.isi.edu/nsnam/ns>, accessed July 2007
- [24] Balakrishnan, H., Seshan, S., Stemm, M., and Katz, R.H.: 'Analyzing stability in wire-area network performance', Proc. ACM SIGMETRICS, June 1997, pp. 2–12
- [25] Paxson, V.: 'End-to-end Internet packet dynamics', IEEE/ACM Trans. Networking, June 1999, pp. 277–292
- [26] Zhang, Y., Duffield, N., Paxson, V., and S. Shenker. Y.: 'On the consistency of Internet path properties', Proc. ACM SIGCOMM Internet Measurement Workshop, November 2001, pp. 197–211
- [27] <http://www.cs.unc.edu/jasleen/research/ab-estimation>, accessed September 2007
- [28] Pasztor, A., Veitch, D.: 'The packet size dependence of packet pair like methods', Proc. IEEE/IFIP Int. Workshop Quality of Service, Miami, USA, May 2002, 204–213