Internet Client Traffic Measurement and Characterisation Results

Joachim Charzinski
Siemens AG, Hofmannstr. 51, D-81359 Muenchen, Germany,
Tel. +49-89-722-46803, Fax +49-89-722-26877, e-mail: j.charzinski@ieee.org

Abstract

On the basis of recent long-term traffic traces, some traffic characteristics of Internet applications are presented and interpreted. The symmetry and asymmetry occurring in connections for Web access, e-mail and file transfer traffic is discussed and explained by TCP (the Internet transport layer protocol) characteristics. The paper concludes by introducing a "fun factor" as a measure for perceived quality of service and by discussing its quantification and application for network dimensioning purposes.
Index terms: Internet Traffic; Application Characteristics; Access Network; Asymmetry; Dimensioning; Bit Rate Distribution; Elastic Traffic; Fun Factor

1 Introduction

Capacity planning is an important part of network engineering. In particular, it is important to match the capacity of the trunk part of a network with the traffic offered on the access lines. Offering too much bandwidth on a common trunk is uneconomical whereas too little bandwidth compromises network performance as perceived by the users.

This paper concentrates on the traffic requirements of elastic traffic as caused by WWW access (HTTP), e-mail transfer (SMTP, POP) and file transfer (FTP). Combining results of different recent client side measurements allows distinguishing between user and application behaviour and requirements. Previously known measurements focus on backbone or server side traffic. In contrast to these, our measurement points were chosen to be close enough to the subscribers in order to capture the complete network traffic of each user.

As quality of service (QoS) for elastic traffic is mainly determined by the bit rate that can be sustained during a TCP connection, one focus of this paper is the distribution of mean TCP connection bit rates as observed under different circumstances. A second focus is on the symmetry of data volumes or bit rates within a connection as well as between different connections, highlighting cases of symmetry and asymmetry in access traffic streams.

The traces that are the basis for the numerical evaluations in this paper are described in sec. 2. Short and long-term bit rate distributions are presented in sec. 3 for different applications and the relation between the amount of traffic sent into the core network to the amount of traffic received from the core network is investigated in sec. 4 for HTTP, e-mail and FTP traffic. Finally, in sec. 5 the dimensioning of access multiplexers and concentrators for elastic traffic is discussed and an appropriate quality of service parameter is suggested.

2 Measurement

The results given in the following sections have been obtained from two different measurements, which will be referred to as "trace A" and "trace B". Both traces were collected at local Ethernet segments where all the traffic to and from individual users' computers could be observed.

Trace A was collected during seven months, when 100 students' PCs were connected to their university's backbone network in Münster, Germany, via ADSL lines. The lines were configured to 2.5Mbit/s downstream and 384kbit/s upstream bit rates. Usage of the Internet access service was free of charge and there was no dial-up procedure, i.e. computers could have an "always on" mode connection to the Internet. In total, during the six months of monitoring, 14 million IP packet headers belonging to HTTP were collected, covering around 480,000 HTTP/TCP connections.

Trace B was collected during five weeks when all traffic at a local Internet Service Provider (ISP) "Bürgernetz Fünfseenland" close to Munich, Germany, was monitored using the same `tcpdump` software [1] as with trace A. Around 300 mostly residential subscribers shared 30 dial-up lines reaching from low-speed modems to double ISDN lines at 128kbit/s plus compression. Apart from charges for the local telephone call needed to connect to the ISP, subscribers only paid a yearly flat rate. Here, for HTTP traffic alone 43 million IP packet headers were collected, covering around 1.6 million HTTP/TCP connections.

In the following, the terms "traffic volume" and "bit rate" are determined on the packet level including the overhead of Ethernet frames. Instantaneous bit rates are averaged over short time intervals whereas the traffic volume and mean bit rate of a "flow" are determined from the total packet traffic carried in the flow and its duration.

The number and characteristic parameters of HTTP flows detected in the two traces in different flow classes are summarized in Table 1. The concept of a *flow* is used in the Internet to denote a number of IP packets traversing the connectionless core network between the same source and destination. There are several ways of defining a flow [2, 3]. For a thorough discussion of flows in conjunction with Web traffic, see [4]. We use the notion of port to port (P2P), host to host (H2H) and total client traffic (CL) flows. A P2P flow comprises all traffic that is transmitted between the same two hosts using the same protocol and the same two ports on these hosts. The beginning and end of a P2P flow are denoted by the first SYN (TCP `synchronize`) and the last FIN (TCP `final`) packet. Defined in this way, a P2P flow is equivalent to a TCP connection. A host to host (H2H) flow comprises all traffic between the same two IP addresses, up to an inter-packet interval of 10 minutes, after which the flow is declared to have ended with the last packet seen. A H2H flow can consist of the packets of one or more TCP connections. A CL flow comprises all traffic for a selected application (WWW, FTP or e-mail) that a single client consumes or produces. As with the H2H flows, the end of a total client traffic flow is determined by a 10 minute timeout.

A comparison of the P2P, H2H and CL characteristics reveals that the mean bit rate of a flow could hardly be increased by clients having parallel TCP connections to the same server but that parallel connections to different servers resulted in an increase of mean bit rates by a factor of two, as can be seen by comparing the CL mean bit rates with H2H or P2P mean bit rates.

Besides the fairly similar values for mean bit rates, there is a big difference between the mean flow durations measured in both traces.

Table 1: HTTP flows in traces A and B: number of flows, mean and coefficient of variation of flow duration and mean downstream bit rate.

flow type	trace	A	B
P2P	#flows	480 794	1 576 151
	mean duration in s	57	19.5
	c_V of duration	15.7	4.2
	mean downstream bit rate in kbit/s	4.6	3.9
H2H	#flows	43 537	95 401
	mean duration in s	471	303
	c_V of duration	4.9	2.3
	mean downstream bit rate in kbit/s	5.0	4.1
CL	#flows	2 260	9 253
	mean duration in s	4 860	1 350
	c_V of duration	2.0	1.5
	mean downstream bit rate in kbit/s	10.5	8.8

This is partly due to the fact that trace A was recorded in a flat-rate environment without any usage charges whereas the users in trace B had to pay for the access via telephone lines, which reduced the total usage as well as the mean usage duration per session. In addition to the mean flow durations of Tab. 1, Fig. 1 gives the complementary distribution function (cdf) of flow durations[1], explaining another part of the difference between mean P2P flow durations of the two traces: A fit of a Pareto distribution to the P2P flow duration cdf $C_{P2P}(t)$ from trace A in Fig. 1 reveals that $C_{P2P}(t) \sim t^{-\alpha}$ for over three decades with $\alpha \approx 0.95$. At this value of α, the expectation and variance of the Pareto distribution, a so-called heavy-tailed distribution, are both infinite if it is continued until $t \to \infty$. Therefore, for any measurement of finite duration, the measured mean and variance grow with the measurement duration. It is this heavy-tailed distribution of burst durations that causes the long-range dependence and fractal patterns observed in Internet traffic [5].

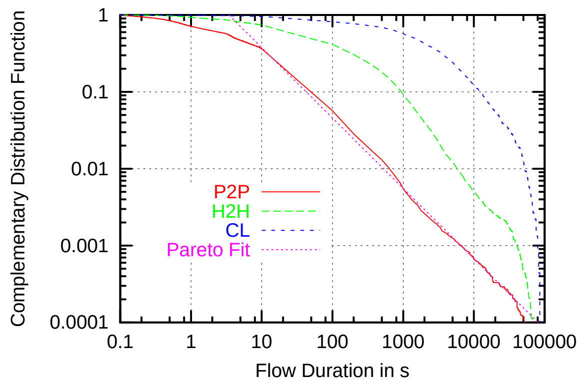

Figure 1: Complementary distribution function of flow durations as measured in trace A.

The distribution of H2H flow durations in Fig. 1 still exhibits a power tail between around 300 and 30 000s, with an exponent $\alpha \approx 1.13$, indicating a finite expectation but still an infinite variance if the distribution tail is continued until infinity. Finally, in the cdf of total client HTTP session durations, this power tail is not visible any more, but due to the small number of CL flows observed, the statistical significance in the corresponding cdf tail region is low, so that neither presence

[1] A cdf gives the probability $C_F(t) = \mathrm{P}\{T_F > t\}$ for the flow lifetime T_F to exceed a given duration t.

nor absence of a power tail in CL flows can be proved.

3 Bit Rate Distribution of Elastic Traffic Flows

In a packet based network, the instantaneous bit rate measured at one location can only take two discrete values: either it is equal to the line bit rate (ongoing packet transmission) or it is zero. Therefore, all meaningful bit rate figures must be averages over a certain time interval. A short averaging time interval leads to a good time resolution but also to a coarse granularity of bit rates. On the other hand, a long averaging interval produces fine granularity bit rates at the cost of a reduced time resolution.
The graph in Fig. 2 gives the complementary distribution function of 100ms average downstream bit rates measured per TCP connection for the application protocols HTTP, POP3 (e-mail reception) and FTP. The downstream access line rate of 2.5Mbit/s is reached in a small fraction of the time intervals whereas there is also a fairly high probability of the connection to be idle during a 100ms interval.

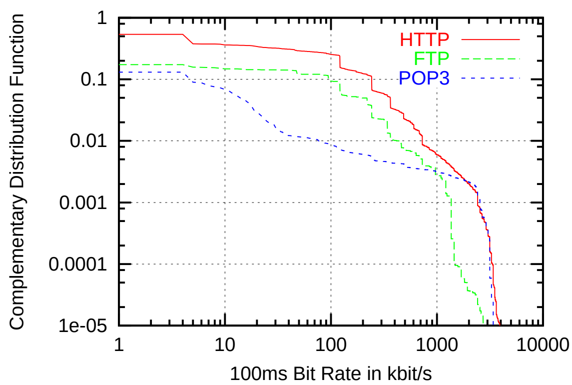

Figure 2: P2P flow (TCP connection) bit rate distribution of HTTP, FTP and E-Mail traffic in 100ms intervals from trace A.

This idle probability is increased depending on the observed service when the total client session traffic is considered in Fig. 3. Here also the effect of parallel connections is visible by a slight increase in the probabilities for high bit rates.

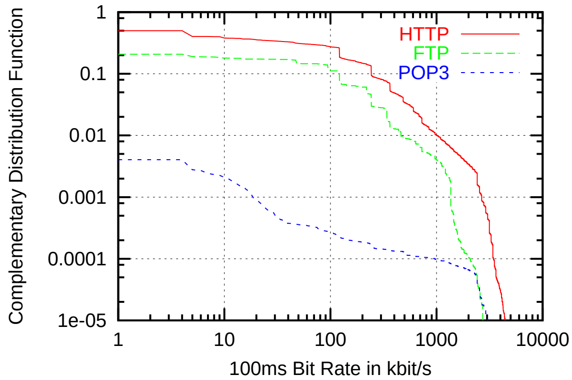

Figure 3: Total client session bit rate distribution for HTTP, FTP and E-Mail traffic in 100ms intervals from trace A.

In order to view the relation between upstream and downstream bit rates during an HTTP/TCP connection (P2P flow), the mean upstream bit rate has been plotted versus the mean downstream bit rate for each connection in Fig. 4. A lighter gray indicates the accumulation of more points on the same spot. Although the bit rates themselves are more variable, the ratio between upstream and downstream rates is limited to values between around 3:1 and 1:30. The same characteristics have been obtained with different applications and different access bit rates, e.g. from trace B or in the cases analyzed in [6], except that for lower access line rates, the achievable mean bit rates of a connection are lower, which limits the extension of the observed combinations at the top right corner in Fig. 4. This effect of limited asymmetry and its causes are investigated further in the following section.

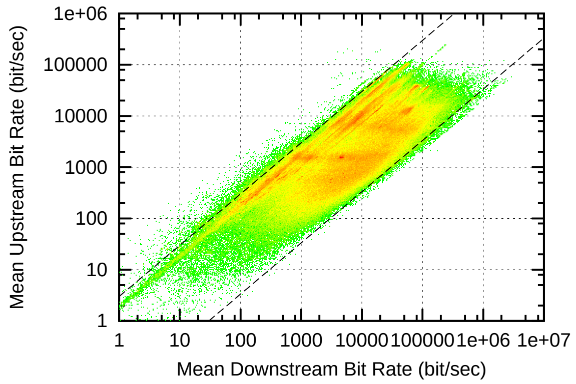

Figure 4: Bit rate symmetry of HTTP traffic

4 Traffic Asymmetry

4.1 HTTP Traffic

Fig. 5 gives the complementary bit rate asymmetry distribution for HTTP traffic on the different flow levels. This distribution is defined as the probability for the ratio of upstream to downstream mean bit rate of a flow to exceed a given value. Note that as the duration of a flow must obviously be the same for the upstream and downstream directions, the ratio of upstream to downstream mean bit rates is equivalent to the ratio of upstream to downstream data volume of a flow.

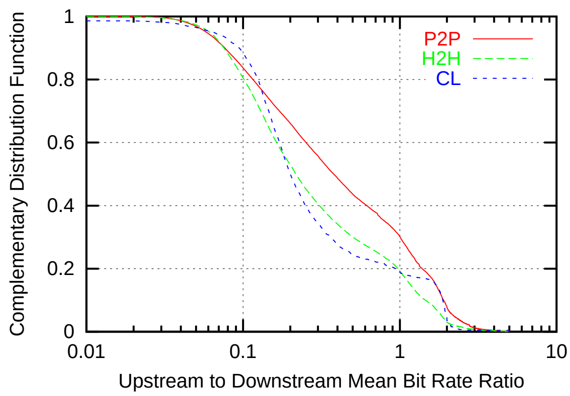

Figure 5: Distribution of upstream to downstream bit rate (or volume) ratios in HTTP/TCP flows.

Fig. 5 indicates that there is a significant fraction of HTTP flows on each flow level where the data volume transmitted upstream is larger than the data volume received from the network. If the total HTTP client traffic is considered, this fraction amounts to 30% of all Web sessions. In addition, the form of the curves indicates that the distributions can be described by lognormal distribution functions (note the log-lin plot).

In order to find out the reason behind this ratio distribution and the high percentage of symmetric or even upstream-oriented transmissions, the ratio of mean bit rates in each connection has been plotted against the downstream volume received from the network in the connection. The resulting graph in Fig. 6 reveals that connections with a small download volume are fairly symmetrical whereas the higher asymmetry ratios of 1:30–1:50 can only be reached in connections in which a large volume of data is retrieved from a server.

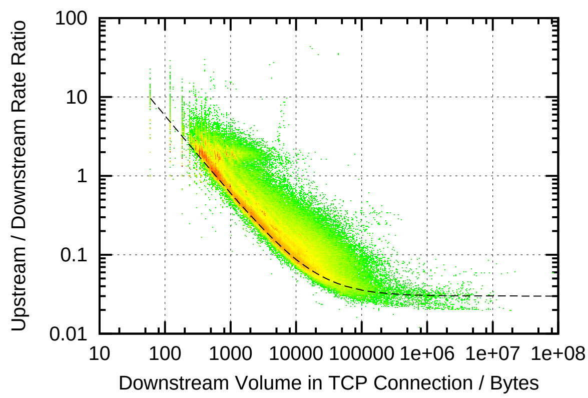

Figure 6: Correlation of upstream to downstream bit rate (or volume) ratios and downstream volume in HTTP/TCP connections.

Denoting the upstream volume in a connection by v_u, the downstream volume by v_d and assuming η acknowledgements per received downstream data packet with values of η between 0.5 and 1, the volume ratio can be derived from the basic HTTP/TCP message sequence depicted in Fig. 7 for the case of non-persistent connections in which only one GET request can be served per HTTP/TCP connection.

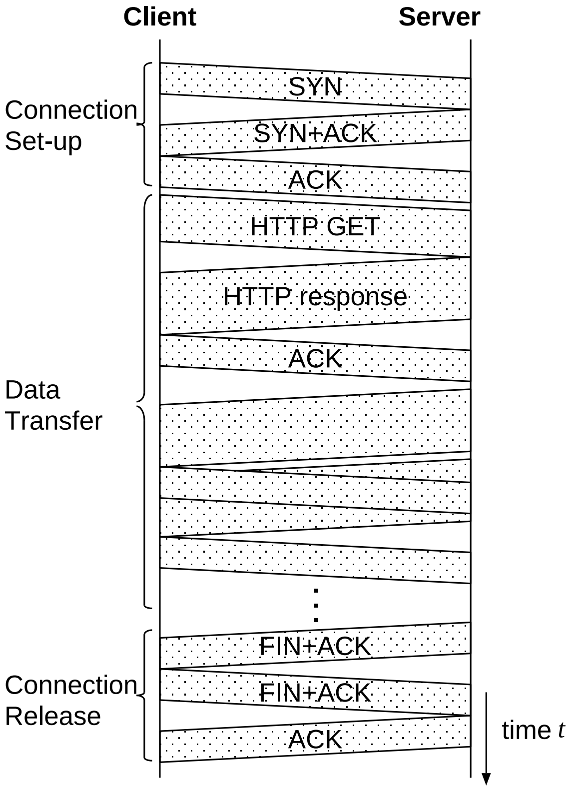

Figure 7: Basic HTTP/TCP message sequence.

As the packet sizes measured in the traces include the Ethernet headers, the sizes of the connection management and acknowledgement packets are 60 bytes each. A mean GET request packet has a size of around 400 bytes and the downstream data packets have a maximum size of 1500 bytes in larger downloads. Thus, the ratio of upstream to downstream volumes in a download is approximated by

$$\frac{v_u}{v_d} \approx \frac{580B}{v_d} + \eta \cdot \frac{60B}{1500B} \quad .$$

The corresponding values have been added as a dashed line in Fig. 6. Connections with a smaller maximum transmission unit (MTU) of 512 bytes e.g. can still be described by using an adapted effective η value. This general behaviour of any TCP based upload or download services has been confirmed with FTP and e-mail data (depending on the direction of transmission).

4.2 E-Mail and FTP Traffic

The complementary bit rate ratio distributions for P2P and H2H e-mail traffic flows in Fig. 8 have been split into two parts each: Residentials usually transmit outgoing e-mails via the simple mail transfer protocol (SMTP) over TCP whereas received messages are retrieved from the mail server via the post office protocol (POP3) over TCP. This fact is reflected in the observation of upstream to downstream ratios being nearly always greater than one for the SMTP traffic and less than one for the POP3 flows. Note that there is a concentration of probability mass at bit rate ratios between 0.64 and 0.74, which is due to the large fraction of POP3 connections that are only used to check if there are new mails available on the server and do not retrieve any data.

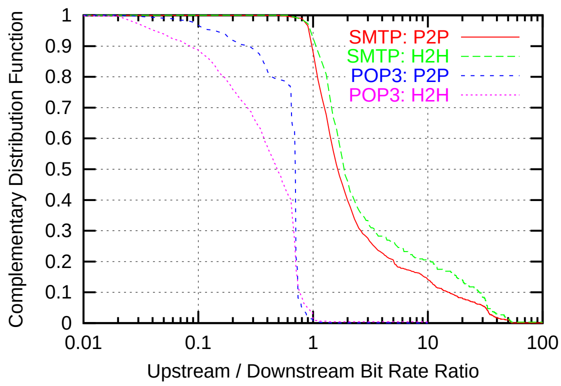

Figure 8: Distribution of upstream to downstream bit rate ratio for e-mail traffic.

In the FTP case depicted in Fig. 9, around 5% of all P2P and 14% of all H2H flows transmit more data into the network than they receive.
Note that the file transfer protocol (FTP) uses separate control and data connections. This leads to a split of the typical TCP

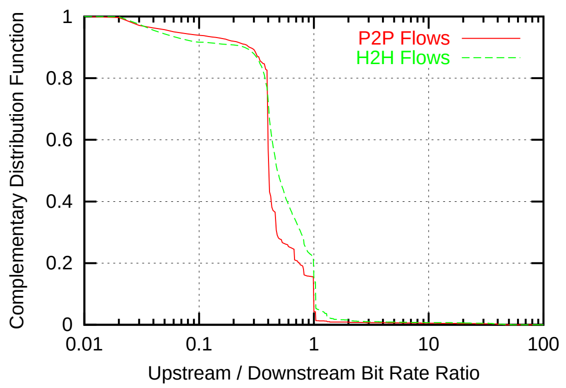

Figure 9: Distribution of upstream to downstream bit rate ratio for FTP traffic.

graph as observed in Fig. 6 into three distinct branches, as can be seen in Fig. 10. The lower branch is due to file downloads with the main part of the traffic in downstream direction. The upper branch is due to the (fewer) file uploads. It differs in form from the lower branch because it would have to be plotted versus the *upstream* instead of the downstream data volume in order to obtain the same form. More results not given here have shown that also the size distribution of uploaded items is highly similar to that of downloaded items – only the frequency of uploads and downloads differ. The third branch in the middle of Fig. 10 is due to the control connections which produce more or less symmetric traffic.

4.3 Access Session Traffic

Apart from the per connection symmetry observed above, the decision as to which access technology to offer to customers is also determined by the overall traffic symmetry that is expected. The asymmetry distribution of the total upstream and downstream traffic per access session as measured in trace B is depicted in Fig. 11. As most of the total access traffic is HTTP or e-mail traffic, we expect to see a mixture of the HTTP asymmetry distributions (Fig. 5) and both distri-

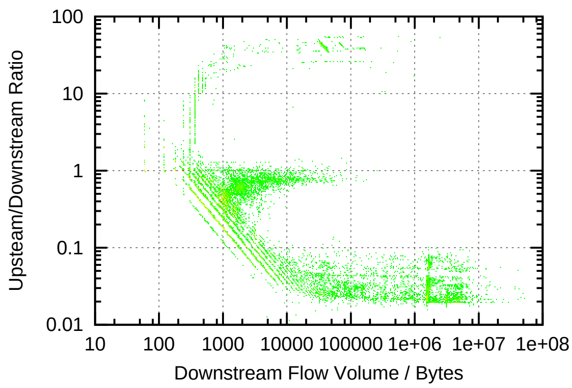

Figure 10: Correlation of upstream to downstream bit rate (or volume) ratios and downstream volume in FTP/TCP connections.

butions for e-mail traffic (Fig. 8). The plot in Fig. 11 contains separate evaluations for normal modem and (single or double B channel) ISDN access sessions, which do not show significant differences. Like in the HTTP and e-mail flow asymmetry distributions, there is a significant proportion of 15–20 % of access sessions in which there is more total upstream than downstream traffic.

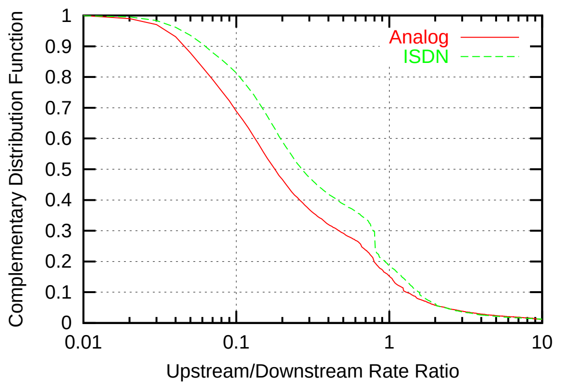

Figure 11: Upstream/downstream ratio distribution of total access traffic.

The decision as to which access technology to offer to customers will also be influenced by the overall traffic symmetry that is expected. In addition to the standard cases of teleworking business customers or server-at-home scenarios, also residential subscribers showing a high volume of upstream e-mail or FTP transmissions or a large fraction of short HTTP downloads will be in favour of a more symmetrical access system, whereas subscribers with an emphasis on large downloads or audio/video streaming applications will prefer to have an asymmetric access line which focuses on downstream line speed.

5 QoS Requirements of Elastic Traffic

Whereas most real-time services producing stream traffic need a certain guaranteed minimum bandwidth in order to be usable at all, applications producing elastic traffic (like WWW browsing, e-mail, file transfer or network news) can be used starting with a very low minimum bandwidth. However, as most of today's traffic in data networks is elastic traffic, Internet Service Providers (ISPs) can use good elastic traffic performance of their networks as a means of advertising. If a corresponding QoS measure, as defined in the following, is standardised and available as a reference or for comparing ISPs, it may be readily accepted by subscribers.

The Internet's transmission control protocol (TCP) provides the necessary flow control functions to optimally utilise the available bandwidth available in the network. Teletraffic models based on the Processor Sharing discipline, like the M/G/R-PS model, take this behaviour into account and allow to use a relative delay factor f_R to give the ratio of transmission time needed for a file to the ideal transmission time that would have occurred e.g. if the full link rate of the access line could have been utilized by the data transfer [7, 8]. This model is insensitive to the burst size distribution, which makes it even more appropriate also for the heavy tailed burst size distributions found in data traffic. In contrast, classical models for multiplexers that determine the delay and/or packet loss probability without regarding the effect of flow control will greatly over-estimate the amount of bandwidth required [9, 10].

In an attempt to define an understandable measure for user perceived quality of service, a "fun factor" φ is defined as the ratio of the transmission time for a given amount of data under ideal conditions to the transmission time actually needed:

$$\varphi = \frac{t_{ideal}}{t_{observed}}$$

This fun factor is for a single connection equal to $1/f_R$ but has the advantage of giving results positively correlated with a perceived quality: A fun factor of zero describes a completely unusable service whereas the upper bound $\varphi = 1$ denotes the best achievable service. Using this measure for QoS, realistic target values for φ can be estimated: In the past, residential subscribers have bought new Internet access equipment when they could increase the bandwidth by a factor of 2–3. Assuming that the old equipment was operated at the limits of its throughput capacity ($\varphi \approx 1$), the same bit rate would correspond to a fun factor of $\varphi =$0.3–0.5 with the new equipment. A noticeable improvement can therefore be felt if $\varphi >$0.7–0.8. For business customers, on the other hand, the target values should be much higher, e.g. 95 or even 99%, depending on the degree of service requested. For high-quality business access, one could also imagine determining the distribution of φ and demanding that the ratio of connections with $\varphi < 90\%$ be less than a certain percentage during busy hours.

For a conservative estimate of the required bandwidth of the common trunk line in an access network, an ON/OFF source model with the ON bit rate equal to the individual access line rate can be used as a reference for the ideal situation. Another approach to obtain more realistic estimates of the required trunk line capacity is to convolve the measured bit rate distributions as given in Fig. 3 for the number of simultaneously active subscribers and to determine the fun factor from the resulting distribution of the ideal aggregated bit rate and its reduction by limiting the sum of all bit rates to a given trunk line rate. However, this approach requires that up-to-date bit rate distribution measurements such as those displayed in Fig. 3 be available from a situation in which the Internet access was sufficiently overdimensioned.

6 Conclusions

Using recent long-term client side traffic traces from a high-speed (2.5Mbit/s ADSL) Internet access network and from a modem/ISDN access pool, several characteristics of Internet client traffic were analyzed. As expected, flows on different aggregation levels show significant idle phases and short-term mean bit rates can reach values around the access line rate. Investigations of the asymmetry in bit rates and data volumes of different flows revealed that although most single connections have a direction of preference, the overall traffic is a mix of upstream and downstream oriented data transfers. In addition, there are a lot of TCP connections used for HTTP transport where upstream and downstream data volumes are about equal, which is due to the symmetry of connection establishment and connection release procedures as well as the fairly big sizes of the packets containing the HTTP GET requests. On the other hand, connections used for transferring large volumes of data cannot exceed an asymmetry ratio of around 1:30–1:50 due to the size and frequency of the TCP acknowledgements transmitted in the reverse direction. The measurement results also revealed that the overall symmetry of the traffic is such that in 15–20% of all access sessions, e.g. modem or ISDN dialup sessions, more traffic is transmitted upstream than downstream. Taking the path from measurement results to dimensioning, the "fun factor" was introduced as an easy-to-quantify measure of perceived quality of service for elastic traffic and hints were given on how dimensioning can be achieved for elastic traffic assuming ideal conditions or even taking the bit rates available from the real Internet into account.

References

[1] S. McCanne, C. Leres and V. Jacobson, "tcpdump", LBNL Network Research Group, ftp://ftp.ee.lbl.gov/tcpdump.tar.Z

[2] P. Newman, W. L. Edwards, R. Hinden, E. Hoffman, F. Ching Liaw, T. Lyon, G. Minshall, "Ipsilon Flow Management Protocol Specification for IPv4," *IETF Request for Comments* RFC 1953, 1996, available at ftp://ftp.isi.edu/in-notes/rfc1953.txt

[3] J.A. Copeland, R. Abler and K.L. Bernhardt, "IP Flow Identification for IP Traffic Carried over Switched Networks," *Computer Networks and ISDN Systems*, Vol. 31, No. 5, 1999, pp. 493–504.

[4] A. Feldman, J. Rexford, R. Cáceres, "Efficient Policies for Carrying Web Traffic Over Flow-Switched Networks", *IEEE/ACM Trans. Networking*, Vol. 6, No. 6, Dec. 1998, pp. 673–685.

[5] M.E. Crovella, A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes", IEEE/ACM Trans. Networking, Vol. 5, No. 6, Dec. 1997, pp. 835–846.

[6] N. Vicari, S. Köhler, J. Charzinski, "The Dependence of Internet

User Traffic on Access Speed", Research Report No. 246, CS Dept., University of Würzburg, Germany, Jan. 2000. Available at http://www-info3.uni-wuerzburg.de/TR/tr246.pdf

[7] D.P. Heyman, T.V. Lakshman, A.L. Neidhardt, "A New Method for Analysing Feedback-Based Protocols with Applications to Engineering Web Traffic over the Internet" Proc. ACM SIGMETRICS'97, Seattle, WA, USA; ACM Perf. Eval. Review Vol. 25, No. 1, pp. 24–38.

[8] K. Lindberger, "Balancing Quality of Service, Pricing and Utilization in Multiservice Networks with Stream and Elastic Traffic" Proc. ITC 16, Edinburgh, UK, June 1999, pp. 1127–1136.

[9] Å. Arvidsson, P. Karlsson, "On Traffic Models for TCP/IP", Proc. ITC 16, Edinburgh, UK, June 1999, pp. 457–466.

[10] M. Grossglauser, J.-C. Bolot, "On the Relevance of Long-Range Dependence in Network Traffic", IEEE/ACM Trans. Networking, Vol. 7, No, 5, Oct. 1999, pp. 629–640.