University of Würzburg
Institute of Computer Science
Research Report Series

# Measuring Internet User Traffic Behavior Dependent on Access Speed

**Norbert Vicari and Stefan Köhler**

Institute of Computer Science, University of Würzburg,
Am Hubland, D - 97074 Würzburg, Germany
Tel.: +49 931 888 6652, Fax.: +49 931 888 6632
e-mail: vicari|koehler@informatik.uni-wuerzburg.de

## Abstract

*A variety of Internet traffic characterization and source models already exist. Since, the origin of the analyzed data differs a comparison of the reported traffic characteristics is difficult. In this paper, we analyze a single IP-packet trace and relate the trace data to the user access speed. This approach leads to some interesting insights showing whether and how the access speed influences the Internet user behavior and thus the generated source traffic. The derived Internet traffic characterization could be seen as a first step to a generalized parametric Internet traffic model.*

# 1 Introduction

Today the Internet covers and integrates many different applications. In the next few years not only data traffic, but also time-critical applications demanding QoS will be transported. For dimensioning of networks, access routers and development of new QoS mechanisms more accurate and realistic source models of Internet user traffic are required. Many studies deal with backbone Internet traffic characterization, revealing interesting facets like Long-Range Dependence or multi-fractal behavior [1, 2, 3, 4, 5]. Since, these models are based on a traffic stream consisting on many superimposed traffic sources, scaling is a demanding task.

As the variety of applications for traffic models increases, i.e., different access speeds and new technologies, also the variety of models increases. Thus, generalized parametric source models of Internet traffic are an interesting approach to fulfill the dimensioning and planning needs.
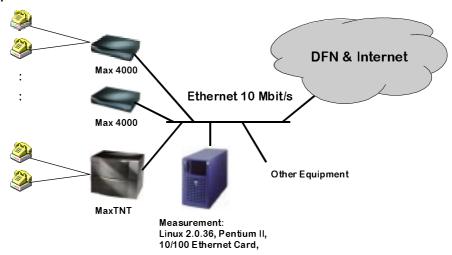
Generalized parametric source models could be derived from client-based Internet traffic traces. But it is not clear, how to compare different models, e.g. [6, 7], since the models are based on different traces. One step towards an generalized model would be to compare the statistics of users accessing the Internet with different modem speed by evaluating an single trace.

The key for obtaining statistics of the access speed as well as packet level characteristics of IP traffic is a parallel measurement of the session logfile of a dial-in router and the packet trace measured on the connection between the dial-in router and the Internet.

The paper is organized as follows: In the next Section the measurement configuration of the analyzed data is described. Section 3 gives the characteristic of the session behavior and Section 4 describes the characteristics of the TCP (HTTP) connections. Section 5 concludes the paper.

# 2 Measurement Setting and Trace Description

The measurement was done during 2 weeks in February and March 1999 in the computing center of the University of Würzburg. About 6000 students and staff members are subscribed to the dial-in Internet access of the university. The access is implemented by three Ascend dial-in routers, enabling simultaneously 240 access lines. All lines facilitate digital ISDN access at 64kbps, while 192 lines also facilitate analog modem access at speeds up to 56kbps. The dial-in routers are connected with a 10Mbps ethernet bus to the German Research Net (DFN) and the Internet.
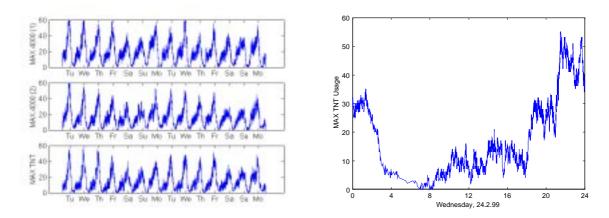
Figure 2: User activity during measurement period (left) and during one day (right).

Figure 1 depicts the measurement configuration. Basic access data were logged by the routers, mainly the session start time, session duration, negotiated up- and downlink speed, data volume and the dynamically assigned IP-address. In order to ensure privacy, all records allowing conclusions to single users were deleted from the logfile.

To obtain detailed information of the user behavior, a sniffer measurement with the tool TCP-Dump [9] running on a Linux PC was made. All IP-packet headers to and from the dial-in routers were logged. The analysis of this trace was performed with the help of several perl scripts and the tool TCPTrace [10]. The IP-header information contains among others the time stamp, source- and destination address including port number and the protocol used (TCP, UDP,...).

Aligning the router logfile and the TCPDump trace allows to securely discern sessions, even in the case of fast IP-address reuse. All traffic characteristics contained in a packet trace could be related to the dial-in speed of the modem used.

During the measurement period approximately 62000 sessions were logged, the total data volume sums up to 82 GB. Figure 2 shows on the left side the number of users during the measurement period. A periodic usage pattern in dependence of the daytime is observed. On the right side of Figure 2 the user activity for 24 hours is depicted. Obviously the users adopt to the phone tariff of the German Telekom AG by increasing activity at 6 p.m and 9 p.m, that is, the starting time of cheaper phone tariffs.
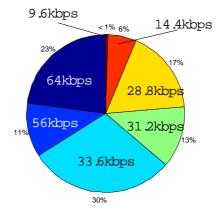


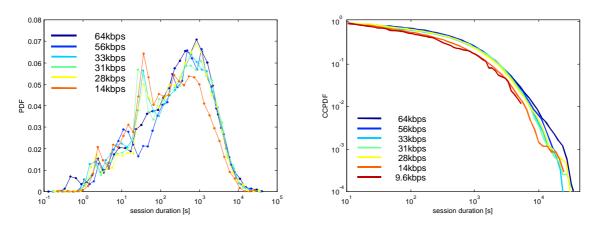Figure 3: Percentage of modem classes in the measurement.

Figure 4: Session duration distribution.

The pie chart in Figure 3 shows the percentage of modem classes in the trace. The main part of the users access the internet with a speed of 28kbps to 33.6kbps. One third of the sessions are using connections with up to date technology, that is ISDN (64kbps) or 56kbps modems. Note that the later class of modem connect asymmetrically, upstream with 33.6 kbps and downstream - depending on the quality of the line - between 40kbps and 56kbps. Only 7% of the sessions stem from slow 9.6kbps and 14.4kbps modems. The color code representation used in this pie chart will also be used for further figures.

# 3 Session Characteristics

The session characterization stands on top of all hierachical modeling approaches. In the following we will look at the distribution of the session duration and volume. Table 1 shows the average session volume and duration in dependence of the modem speed. While the average session duration increases only slightly with the modem speed the data volume increases proportionally to the modem speed. The coefficient of variation (CoV) is nearly independent of the modem speed.

| | | Duration | | Volume | |
|---|---|---|---|---|---|
| Speed | #Sessions | Mean [s] | CoV | Mean [kB] | CoV |
| 9.6 kbps | 273 | 536 | 1.93 | 241 | 2.68 |
| 14.4 kbps | 3624 | 644 | 2.20 | 540 | 2.85 |
| 28.8 kbps | 10602 | 885 | 1.97 | 1133 | 3.31 |
| 31.3 kbps | 7969 | 904 | 1.89 | 1135 | 2.80 |
| 33.6 kbps | 18538 | 896 | 1.91 | 1137 | 2.68 |
| 56 kbps | 6809 | 964 | 1.85 | 1578 | 2.44 |
| 64 kbps | 14268 | 979 | 2.14 | 1866 | 3.70 |

Table 1: Average and Variation of Session Volume and Duration.

The session duration distribution is depicted in Figure 4. On the right hand side of the figure the complementary cumulative probability distribution function (CCPDF) shows in double logarithmic plotting that the shape of the distribution is identical for all modem speeds. On the left hand side the probability density function (PDF) shows two peeks, which initiates to separately characterize short sessions. About 30% of the sessions do not last longer than 100s.

Besides the session duration the data volume transferred is used for the characterization of sessions. The up- and downstream volume distribution is shown in Figure 5. For all modem speeds the curves exhibit a similar shape, while the data amount increases with the modem speed. The tail of the distributions decays linearly in double logarithmic representation, which indicated the heavy tailed probability of these type of distributions. The ratio of down- and upstream session volume ranges from 4 to 7.
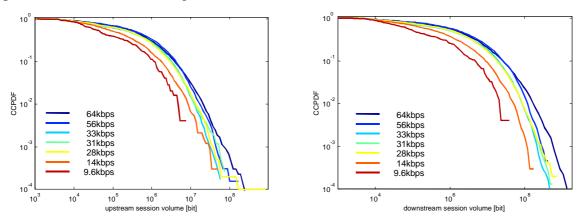


Figure 5: Session up- and downstream volume distribution.

It can be seen that the services (applications) used during a session influence strongly the behavior of the traffic-stream. E.g., in [8] two applications cause the complete correlation of an IP trace. Thus, we list the volume percentage of the services used in our measurement in Table 2. As expected, 80% of the traffic is caused by HTTP. Mail is the preferred application of users with slow 9.6kbps modems. The percentage of Mail reduces with increasing modem speed, while services like FTP or Games are utilized more often. About 10% of the TCP port numbers are used to rarely to show up in this statistic and/or could not be assigned to known services. Since the unidentified UDP port numbers take at most one percent of the traffic volume, it could be assumed that the application *Real Audio* is not attractive to modem users.

| Speed | Data Volume | HTTP | Mail | FTP | Telnet | News | Other TCP | DNS | Games | Other UDP |
|---|---|---|---|---|---|---|---|---|---|---|
| 9.6 kbps | 0.07 GB | 55.03 | 29.88 | 0.02 | 0.31 | 0.00 | 13.44 | 0.87 | 0.00 | 0.45 |
| 14.4 kbps | 1.9 GB | 80.41 | 9.27 | 1.13 | 0.47 | 1.32 | 6.29 | 0.69 | 0.00 | 0.42 |
| 28.8 kbps | 12.0 GB | 80.19 | 9.89 | 0.98 | 0.27 | 2.36 | 4.61 | 0.42 | 0.24 | 1.04 |
| 31.3 kbps | 9.9 GB | 79.18 | 5.89 | 1.96 | 0.08 | 0.83 | 10.35 | 0.41 | 0.43 | 0.86 |
| 33.6 kbps | 21.0 GB | 83.12 | 5.42 | 2.23 | 0.31 | 0.49 | 7.28 | 0.43 | 0.12 | 0.61 |
| 56 kbps | 10.7 GB | 79.97 | 4.58 | 2.85 | 0.15 | 0.55 | 10.49 | 0.39 | 0.00 | 1.03 |
| 64 kbps | 26.6 GB | 74.02 | 3.29 | 5.01 | 0.23 | 1.50 | 11.51 | 0.28 | 3.05 | 1.11 |

Table 2: Volume Percentage of Services.

The statistic in Table 2 could be somewhat misleading, since one might think that ISDN users read less Mail than users with slower modems. Figure 6 shows the average data volume per session for the services HTTP, Mail, FTP, and other. While the data volume of HTTP and FTP increase with modem speed, the volume of Mail is kept constant. The reason for this property

is, that with the exception of mailing lists, the usage of mail also implies to write and send mails. This reduces the volume of mail independently of modem speed.
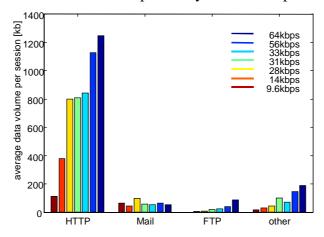


Figure 6: Average Data Volume per service (HTTP, Mail, FTP) and Session.

As indicated in Figure 4, around 30% of the sessions last less than 100s. Comparing the services used in these sessions with the average of all sessions the role of HTTP and Mail are found to be exchanged. Mail volume ranges from 80% to 50% in these short sessions. This results from the fact that users check mail or download mail and then read mail off-line. This could also imply the first peak in the session duration distribution.

## 4 TCP (HTTP) Connection Characteristics

Since 80% of the traced traffic volume are HTTP-connections we concentrate in the following on results of HTTP-connections.

The user activity alters in dependence of the daytime and the tariff structure of the German Telekom AG. Figure 7 depicts the TCP-connection data rate in upstream (left side) and downstream (right side) direction. The largest data rate is obtained between midnight and 8 a.m. During the business hours and the period of maximum user activity, the obtained data rate keeps constant, while the backbone load of the bearer DFN network [11] weakens. Nevertheless, the modem speed is also reflected in the obtained data rate. Further discrimination of the connection destinations in conjunction with a data rate analysis could give valuable hints for network bottleneck detection.
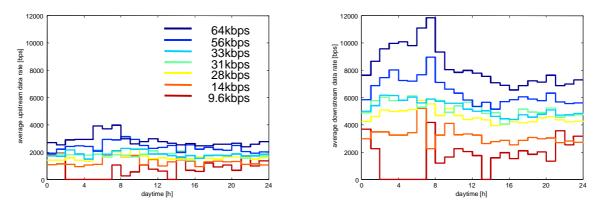


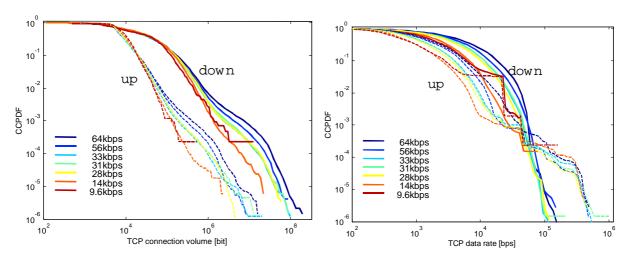Figure 7: TCP-connection data rate in dependence of daytime.

Figure 8: TCP-connection volume (left) and TCP data rate (right) distribution.

The TCP-connection volume distribution depicted in Figure 8 (left hand side) exhibits a similar shape for all access speeds. Also the shape of the volume in up- and downstream direction is similar, showing an heavy tailed decay over four orders of magnitude. The downstream volume is about one order of magnitude larger than the upstream volume. The volume also increases with the access speed. The shape of the data rate distributions, as plotted in Figure 8 (right hand side) is independent of the access speed but differs for up- and downstream direction. Some connections reach upstream data rates, that are 10 times higher than the modem speed. This could happen if a connection setup is requested with a SYN-packet and refused with a Reset-packet.

The throughput of TCP-connections depends on a closed-loop control cycle. Therefore, the up- and downstream data rate is correlated. Figure 9 shows the correlation of up- and downstream data rates for 14 kbps resp. 64 kbps access speed. The figures for the remaining modem speeds are omitted since no additional information is contained. The correlation looks similar for all modem speeds.
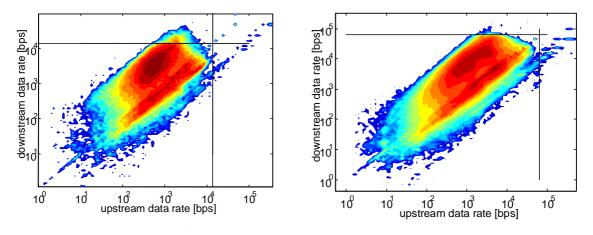


Figure 9: TCP data rate correlation for 14.4kbps (left) and 64kbps (right).

If up- and downstream data rate are highly correlated, the plot in Figure 9 would be a diagonal line. The fact that a significant part of the connections exhibits larger or smaller upstream rate than downstream rate is investigated in Figure 10, where the rate ratio is rendered over the downstream data volume. For small downstream data volume the upstream rate is higher than the downstream rate. For large connections a limit is reached at 1/50, since two data packets of
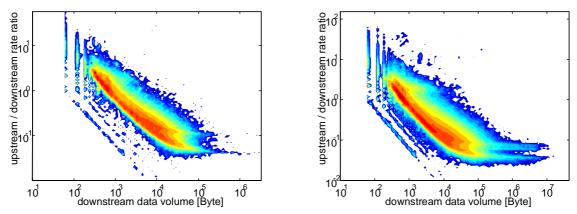
7

Figure 10: TCP data rate ratio in dependence of downstream data volume for
14.4kbps(left) and 64kbps (right).

1500 Bytes could be acknowledged with one packet of 60 Bytes. This also implies that for fast transmission of HTTP data, where about 50% of the downstream connection volumes are less than 3000 Bytes, a symmetric connection is required.

# 5 Conclusion and Outlook

In this paper we presented an analysis of an IP-packet trace measured at the dial-in access of the computing center of the University of Würzburg. By evaluating the packet trace in conjunction to the accounting data, we were able to relate the access speed and the packet data, in order to measure the user behavior in dependence of access speed. The intention of the approach was to identify general properties of internet traffic and dependencies of the subscriber behavior and the access speed. The results of this form the basic for developing a generalized Internet user model, becoming increasingly important in presence of the variety of future access technologies, which range from slow wireless IP access to high-speed HDSL connections.

We identified several general properties of Internet traffic, that are almost independent on the access speed. Among them are the session duration, the shape of the up- and downstream session volume, the shape of the TCP-connection volume and the TCP connection data rate. The correlation of up- and downstream data rate is also not influenced by the access speed.

The traffic volume and the obtained data rates were found to increase proportionally to the access speed. Also the services utilized change in dependence of the access speed. Nevertheless, the similar shape of the distributions should be helpful when developing a parametrized Internet traffic model.

Further investigation is required to find the limits of Internet usage: What is the influence of tariffs to the session duration? Will the traffic volume increase also proportional to access speeds in the range higher than 64kbit and is there a limit? If there is a limit, is it due to network congestion or to the ability of human perception?

# Acknowledgments

# References

1.  M. E. Crovella and A. Bestavros. *Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes.* Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, May 1996, Philadelphia, PA, USA.

2.  W. E. Leland, W. Willinger, M. S. Taqqu and D. V. Willson. *On the Self-Similar Nature of Ethernet Traffic.* Proceedings of the ACM SIGCOMM'93, September 1993, San Francisco, CA, USA.

3.  P. Karlsson and A. Arvidsson. *The Characteristics of WWW Traffic and the Relevance to ATM.* May 1997, Temporary Document COST 257(97)21, COST 257 Management Committee Meeting, Helsinki, Finland.

4.  S. Molnar and A. Vidacs. *On Modeling and Shaping Self-Similar ATM Traffic.* Proceedings of the 15th International Teletraffic Congress, June 1997, Washington, D.C., USA.

5.  B. A. Mah. *An Empirical Model of HTTP Network Traffic.* Proceedings of the IEEE INFOCOM '97, April 1997, Kobe, Japan.

6.  C. R. Cunha, A. Bestavros and M. E. Crovella. *Characteristics of WWW Client-based Traces.* June 1995, Technical Report TR-95-010, Boston University Computer Science Department.

7.  N. Vicari. *Models of WWW-Traffic: a Comparison of Pareto and Logarithmic Histogram Models.* April 1998, Proceedings of the 5th Polish Teletraffic Symposium / Project COPERNICUS 1463 ATMiN - Closing Seminar, Warsaw, Poland.

8.  V. Bolotin, J. Coombs-Reyes, D. Heyman, Y. Levi and D. Liu. *IP Traffic Characterization for Planning and Control.* Proceedings of the ITC 16, June 1999, Edinburgh, UK.

9.  TCPDUMP, available via http://ftp.ee.lbl.gov/

10. TCPTRACE, available via http://jarok.cs.ohiou.edu/software/tcptrace/tcptrace.html

11. DFN. *B-Win Statistics.* http://www-win.rrze.uni-erlangen.de/cgi-bin/auswahl-linkstat.pl.