University of Würzburg
Institute of Computer Science
Research Report Series

Measurement and Modeling of WWW-Sessions

Norbert Vicari

Report No.: 184 September 1997

Institute of Computer Science, University of Würzburg,
Am Hubland, D - 97074 Würzburg, Germany
Tel.: +49 931 888 5505, Fax.: +49 931 888 4601
e-mail: vicari@informatik.uni-wuerzburg.de

Abstract

In this paper we present measurements of WWW-traffic, the analysis of the measured data, and derive a simple abstract model, which could be used to describe WWW-traffic for analysis and simulation purposes. The data was measured in the Ethernet segment of the department of computer science at the University of Würzburg. We analyze client WWW-sessions which are characterized by the size of the response and inter-response intervals. The samples of both categories are found to be approximately Pareto-distributed and exhibit small dependencies. Thus we model WWW-traffic by two independent Pareto distributions. The model is evaluated by simulated transmission of the modeled traffic over an ATM link using the VBR service category. With respect to the simple modeling approach we obtain a good fit between the measured and modeled data set.

1 Introduction

In the last five years an exponential growth of the Internet was observed in different Internet surveys, c.f. [9]. Most of the traffic volume is originated by data transfers in the WWW (World-Wide Web). Growing bandwidth demand for WWW-applications is expected due to high resolution graphics workstations, multimedia applications and network computers. Therefore WWW-traffic is considered to be an important traffic source for future ATM based B-ISDN networks.

The characterization and modeling of WWW-traffic gained a lot of attention in the last years. Numerous studies deal with inquiries of accumulated traffic streams. Either the data-rates of Ethernet-traffic [5] or WWW-traffic [1][2][6][7][8] are considered as traffic sources. The main result of this investigations is the evidence of self-similarity of this type of traffic. Other publications deal with the modeling of WWW-request traffic [3] and the locality of WWW-references [4], which is an important measure for the performance of proxy-servers.

In this study we concentrate on traffic characteristics of single client WWW-sessions. We derive these characteristics from measurements of WWW-traffic in a local ethernet segment at the Department of Computer Science at the University of Würzburg. Currently, all WWW-traffic is influenced by the TCP/IP protocol stack and slow ethernet links, but this influence is expected to be of less importance in future networks. An abstract and simple model of single WWW-sessions is derived from the measured data. The model can be applied for the evaluation of connection technologies which cover the last mile to the user, e.g. HFC (hybrid fiber coax) systems or ADSL (asymmetric digital subscriber line) modems [10]. An other interesting application of the model is the evaluation of the applicability of different ATM service categories for the transmission of WWW-traffic.

The paper is organized as follows: In Section 2 we describe the measurement of WWW-traffic and the environment of the measurement. The third section deals with the analysis of the measured data. The characteristics of WWW-sessions and WWW-pages are derived. Section 4 describes how the measured WWW-traffic is modeled. The model is evaluated in comparison to measured data obtained by simulating the transmission of both traffic types over an ATM link with the VBR service category. The paper is concludes with a summary and a description of aspects on future work.

2 Measurement Environment and Data Set

The investigations presented are based on a measurement of WWW-traffic in an ethernet segment of the Department of Computer Science at the University of Würzburg. Connected to this segment are about 20 workstations including 1 file-server and 2 WWW-servers. The equipment is used by 9 employees and several students. To ensure that a reasonable number of users utilize the equipment and produce sufficient traffic the data was captured for two weeks before end of term.

Technically, the measurement was carried out with the TCPDUMP software [12] on a SUN Sparc 4 workstation. This tool logs the headers of IP-packets. The logged information includes the source address and port, destination address and port, the time and the size of the packet. Further flags indicating the initialization and termination of TCP-connections and the TCP-window sizes are recorded. Options of TCPDUMP allow to filter traffic with respect to the ports used. Since the well-known port number of WWW-servers is 80, packets carrying WWW-traffic can be logged.

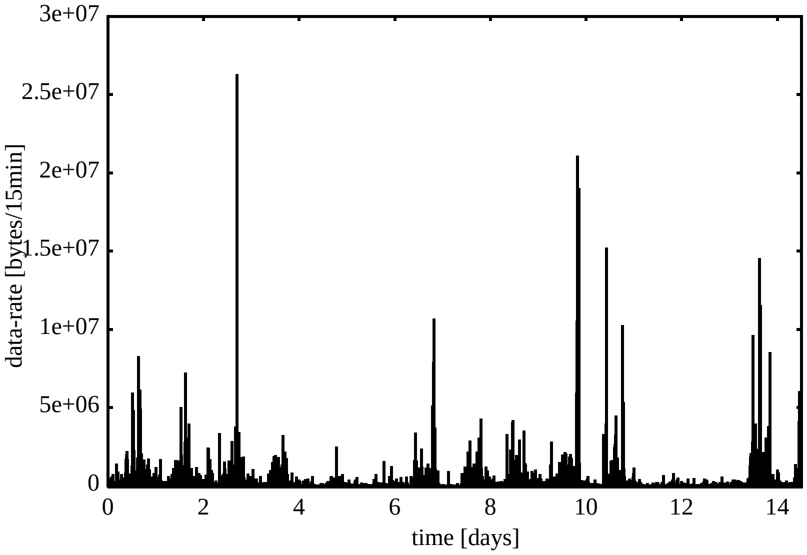

Figure 1: Trace of WWW-traffic in the Ethernet segment of the department.

In Figure 1 the trace of the WWW-traffic recorded is depicted. The total traffic of 944 MBytes is gathered in 15 minute bins. Obviously the main fraction of traffic is measured during busy hours, while during the nights and weekends less traffic was produced. In measurement intervals where nobody was in the office, traffic is caused be external downloads from the WWW-servers in the department.

3 Results and Data Analysis

To obtain information on the user behavior in single WWW-sessions the trace is analyzed. In Figure 2 the hierarchical components of a WWW-session are presented. A WWW-session is the period starting at the time a user launches his WWW-browser and ending when the user quits the WWW-browser. Therefor, the traffic sources of a single WWW-session include only one client and many WWW-servers. Since these events are not logged in the measured trace we introduce the concept of sub-sessions. A sub-session is defined to be the interval in which a users creates WWW-traffic without being silent for more then a interval named time-out. In most cases it could be assumed that session and sub-session are identical.

A sub-session consists of pages, which are the traffic a user originates with one mouse click. One page might induce several TCP-connection to one server and from the server. We define the response-size to be the sum of the size of all IP-packets sent from the server to the client to display a single WWW-page.

We divide the trace in single sub-sessions and extract the start time and size of the pages in the order requested by the user. We concentrate on client sub-sessions, only taking into account data requested by local clients. Data requested from outside represents no complete WWW-session and thus is ignored.

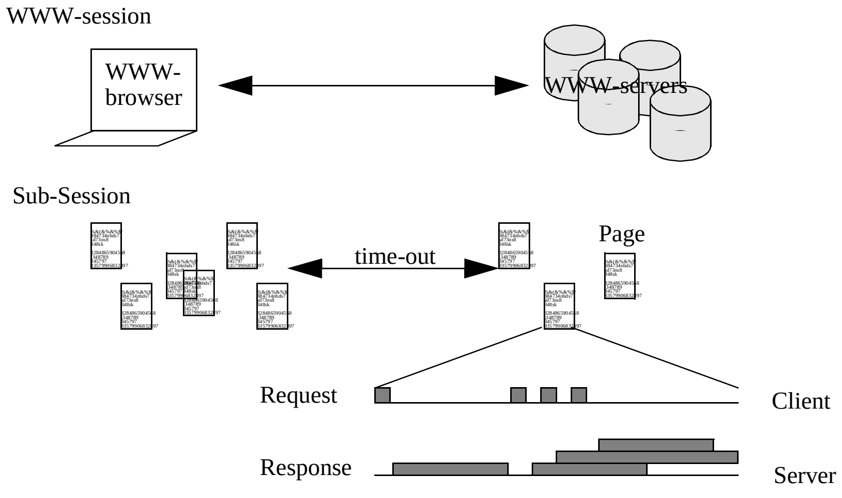

Figure 2: Hierarchical components of WWW-sessions

3.1 Sub-Session Detection

To detect sub-sessions, we assume that at most one WWW-browser is launched on one workstation. Since all workstations are equipped to allow effective working it would be no advantage to use browsers remotely. We can not exclude that some users might open several WWW-sessions at a time. The data belonging to these WWW-sessions are assumed to belong to one sub-session.

The following algorithm is used for detecting sub-sessions: The start of a sub-session is given by the transmission of the first IP-packet from a workstation, called the client, to a WWW-server. All subsequent packets sent from the client to the server and packets sent from the server to the client are assumed to belong to the sub-session. The sub-session is assumed to end if no packets are sent for a certain time. This timeout is chosen to cover the time a user might spend reading a document without requesting a new document, but should be short enough to detect the start of new sub-sessions.

The sub-session detection algorithm considers only client sessions, since we are primarily interested in the user behavior. The volume of data transferred is approximately half amount of the trace. The remaining data volume is caused by requests to the local WWW-servers from clients outside the department. The sub-session detection algorithm shows high stability with regard to the timeout. For timeouts ranging from 15 to 45 minutes, the same 373 sub-sessions have been detected from the trace.

In Figure 3 the number of parallel sections detected in the 14-day trace are depicted. The pattern is strongly related to the busy hours, since a sub-session normally requires human interaction. One sub-session from the 8th to the 9th day of the trace lasts over night, which was caused by the HTML push/pull mechanism.

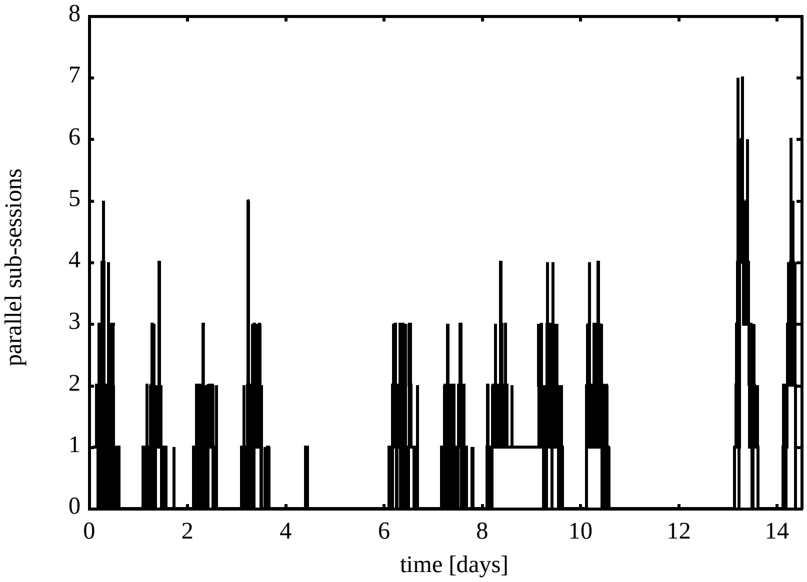

Figure 3: Parallel WWW-sessions detected during measured time interval.

3.2 Characteristics of Sub-Sessions

Figure 4 shows the histograms of properties of the detected sub-sessions. On the left-hand side, the number of sub-sessions are depicted over the sub-session-sizes, which are gathered in 10kB bins. The average sub-session has a size of 1.28MB and the coefficient of variation is 3.2. On the right-hand side, the histogram of the sub-session durations gathered in 60s bins is illustrated. The mean sub-session duration is 29 minutes with a coefficient of variation of 3.0.

All measured sessions caused the transmission of 480MB of data. About 10% of the traffic (the requests) was directed from the clients to WWW-servers while the main part of the traffic was caused by responses on requests. Therefore, we concentrate our further investigation on the characteristics of response traffic, which originates the major part of the traffic.

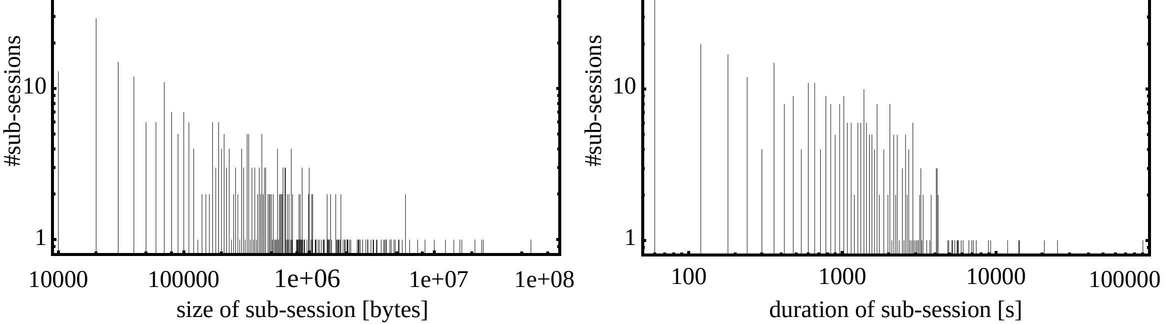

Figure 4: Histogram of data volume (in 10k bins) of sub-sessions (left) and duration (in 60s bins) of sub-sessions (right).

3.3 Characteristics of WWW-page Sizes

According to the current HTTP/1.0 standard [11] WWW-pages are downloaded in several TCP/IP connections. For each inline graphic or other object a separate connection is opened. A similar algorithm to detect WWW-sessions can be used to extract the download-times and sizes of WWW-pages. The start-time of a WWW-page is set to the first IP-packet of a new connection. All subsequent packets of connections between the identical host and client are assumed to belong to the WWW-page if the time between the connections is less than a time-out of 3 seconds. This selection of the timeout showed the best performance. During the 14 day trace a total of 7480 WWW-pages have been downloaded. We define the size of a response onto a WWW-request as the sum of the sizes of all packets which are down-loaded from a WWW-server to the client upon a request. In the average one response contains four separate files - the actual WWW-page and inline objects. On average 19.6 WWW-pages are loaded in one sub-session.

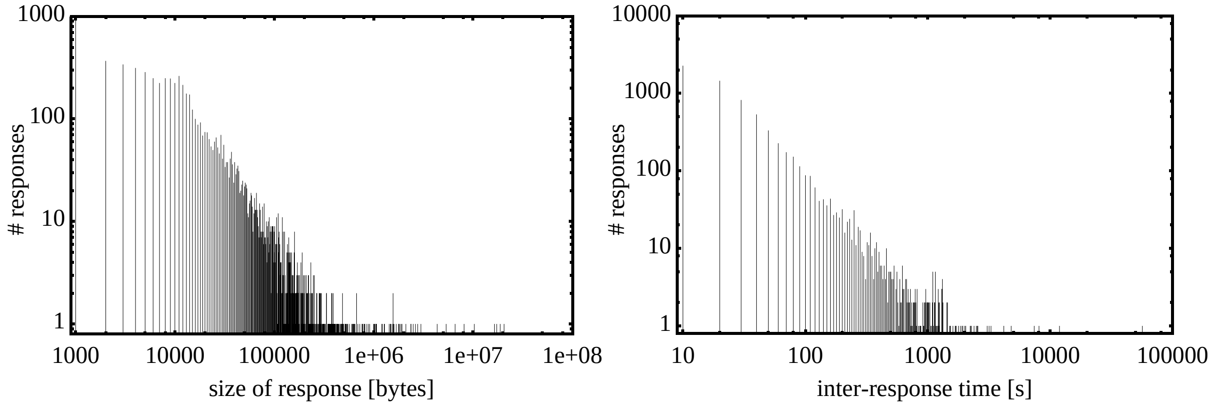

Figure 5: Histogram of response size of WWW-requests in 1k bins (left) and histogram of inter-response times in 10s bins (right).

On the left-hand side of Figure 5 the histogram showing the response sizes gathered in 1kB bins is illustrated. The average response size is 54kB with a coefficient of variation of 9.1. The graph on the right-hand side of Figure 5 shows the histogram of the inter-response times, which are gathered in 10s bins. For the computation, times between subsequent sub-sessions are not taken into account. The mean inter-response time is 81s and the coefficient of variation of the inter-response times is 9.0. Both the incidence of response sizes and inter-request times plotted on double logarithmic axes exhibit a linear decay. This property leeds to the assumption that the distributions of the inter-response time and response size could be modeled with Pareto-distributions.

The scatter plot in Figure 6 shows the dependency between the time to the next response and the size of the current page. Again only intervals within sub-sessions have been considered. The area covered by the pairs of inter-response time and current response size is quite large. Consequently the axes are scaled logarithmically. As shown in the figure no particular relation between large response sizes and large inter-response times or small response size and small inter-response time can be found. The coefficient of covariance of the samples is 0.04. These properties indicate the independence of the response-size and inter-response time. An explanation for these properties can be given by user behavior and WWW characteristics. Often users utilize large WWW-pages as starting point without really reading this pages, which explains the missing relation of large responses and inter-response times. On the other hand the combination of large inter-response times and small WWW-pages might be caused by congested WWW-servers and Internet links.

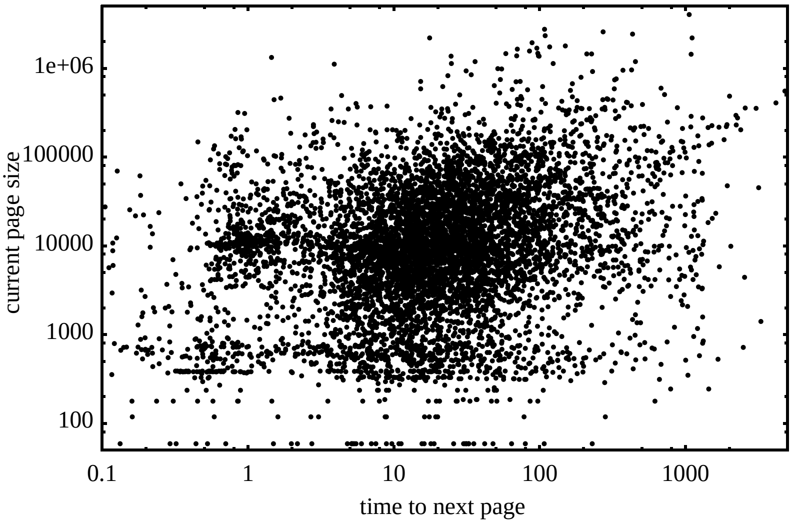

Figure 6: Dependence of the time to the next response and the current response size.

4 Modeling

As indicated before, we model both the distribution of the response size and the distribution of the inter response time with a normalized Pareto-distribution. To obtain a sample of the artificially modeled WWW-traffic, samples of both distributions are independently combined.

4.1 Model Description

Since the samples in the measured data have finite minimum and maximum values and exhibit a high but finite variance, we introduce a Pareto-distribution with similar properties. The well-known Pareto-distribution is normalized to cover values from a minimum k to a maximum m. The gradient of the distribution is given by a parameter α.

We obtain the following equation for the probability density function of the modified Pareto-distribution:

$$(x) = \frac{1}{1-\left(\frac{k}{m}\right)^{\alpha}} \alpha k^{\alpha} x^{-\alpha - x}, \qquad \alpha, k > 0, k \le x \le m \tag{1}$$

and the corresponding probability distribution function:

$$F(x) = \frac{1}{1-\left(\frac{k}{m}\right)^{\alpha}} \left(1-\left(\frac{k}{x}\right)^{\alpha}\right), \qquad \alpha, k > 0, k \le x \le m. \tag{2}$$

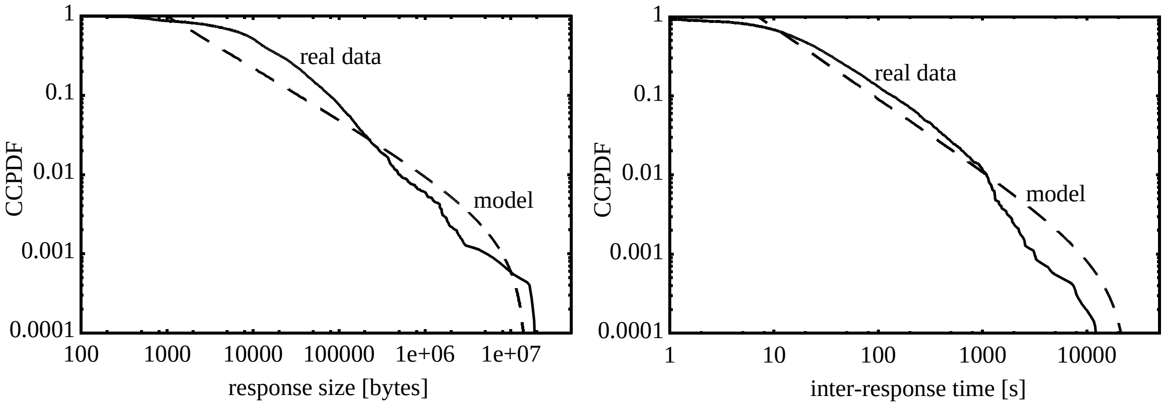

Figure 7: Comparison of measured and modeled distribution functions of the response size and inter-response time.

Figure 7 shows the complementary distribution function of measured and modeled inter-response time (right) and response size (left). The solid lines indicate the empirical distribution functions while the dashed lines depict the fitted modified Pareto-distribution functions. The gradient parameter of the distributions was determined by a least square optimization and the minimum and maximum were chosen to fit the mean and variance of the empirical distributions. The selection of the parameters allows a high degree of freedom. The estimation of the gradient depends strongly on the choice of the minimum of the distribution. Further work is required to automate the parameter estimation in order to provide a more exact modeling of the distributions. For the first modeling approach, which depends on the assumption of independence of the inter-response time and response size, the parameters used in the graph are sufficiently exact.

We used for the distribution of the response size the parameters $\alpha = 0.65$, $k = 1000$ and $m = 1.5 \cdot 10^7$ to obtain a mean of 52000 and a CoV of 8.7. For the distribution of the inter-

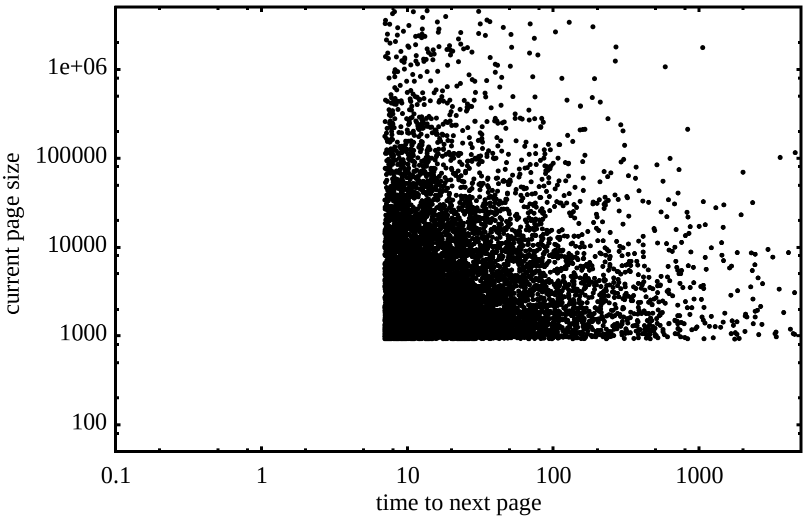

Figure 8: Scatter plot for the model data.

response time we used the parameter set $\alpha = 0.9$, $k = 7$ and $m = 2.5 \cdot 10^4$. The mean of the modeled distribution is 79.8 and the CoV is 7.6.

In Figure 8 the scatter plot of 10000 pairs of inter-response time and response size of the model are depicted. In comparison to the scatter plot of the measured values in Figure 6 the area of small values is not covered, which is caused by the selection of the minimum values of the distributions.

4.2 Model Validation

To check the accurateness of the model, we simulate a transmission of the modeled and measured WWW-pages over an ATM-link using the VBR service category. The number of cells required for each page is determined and submitted to the link of speed PCR. Cells which do not conform to the connection traffic descriptors SCR and BT are assumed to be lost.

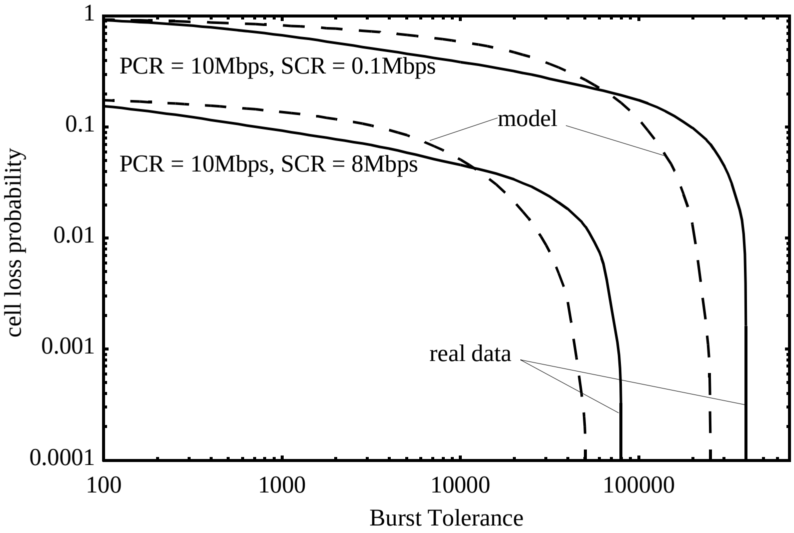

Figure 9: Dimensioning of connection traffic descriptors.

In Figure 9 the cell loss probabilities for a PCR of 10Mbps and a SCR of 8Mbps (0.1Mbps) are depicted. The average data rate of both samples is less than 5.4kbps. The dashed lines show the blocking probabilities of the modeled samples while the bold lines indicate the blocking probabilities of the measured response sizes and inter-response times. For small values of the Burst Tolerance, the modeled WWW-traffic exposes higher blocking probability than the measured traffic. This is an indication for overrepresenting short-term dependencies in the modeled WWW-traffic. On the other side, with large values of the Burst Tolerance the blocking probability of the modeled traffic decays faster than the blocking probabilities of the original traffic. The reason for this behavior is an insufficient modeling of long-range dependencies, which have found to be characteristic for Internet-traffic.

5 Outlook

In the future it is expected that WWW-communications will be an important traffic source to be carried on emerging broadband networks. Thus, modeling this kind of traffic is required to evaluate the applicability of different ATM service categories for the transmission of WWW-traffic.

In the investigation presented in this paper we have measured WWW-traffic in the local Ethernet segment of the department of computer science of the University of Würzburg. The measured data was analyzed and found to be in good accordance with other measurements published in other papers. The inter-response time and the response size proof to be the most important characteristics of WWW-traffic. The samples of both values are approximately Pareto-distributed and independent. Thus we model the inter-response time and response size as independent and normalized Pareto-distributions. The model is validated by the simulated transmission of data over an ATM-link utilizing the VBR service category. The model exhibits stronger short-range dependencies and lacks the long range dependency of the measured data set. Further work has to be carried out to reflect the dependencies of the measured data more correctly.

Acknowledgment

The author thanks Georgios Aivalis for providing the measured WWW-data set and his work on the analysis of the data set. Further, the financial support of the Deutsche Telekom AG (Technologiezentrum Darmstadt) is appreciated.

References

1. 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.
2. 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.
3. J. Judge, H. W. P. Beadle and J. Chiocharo. *Modeling World-Wide Web Request Traffic.* Proceedings of the Multimedia Computing and Networking Conference 1997, February 1997, San-Jose, CA, USA.
4. V. Almeida, A. Bestavros, M. Crovella and A. de Oliveira. *Characterizing Reference Locality in the WWW.* Proceedings of the IEEE Conference on Parallel and Distributed Information Systems, December 1996, Miami Beach, FL, USA.
5. 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.
6. B. A. Mah. *An Empirical Model of HTTP Network Traffic.* Proceedings of the IEEE INFOCOM '97, April 1997, Kobe, Japan.
7. 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.
8. P. Karlsson and A. Arvidsson. *The Characteristics of WWW Traffic and the Relevance to ATM.* May 1997, Technical Document COST 257(97)21, COST 257 Management Committee Meeting, Helsinki, Finland.
9. Network Wizards Inc. *Internet Domain Survey.* available via: http://www.nw.com/zone/WWW/top.html
10. P. J. Kyees, R. C. McConnell and K. Sistanizadeh. ADSL: *A New Twisted-Pair Access to the Information Highway.* April 1995, IEEE Communications Magazine.
11. RFC 1945. *Hypertext Transfer Protocol -- HTTP/1.0.*
12. TCPDUMP, available via http://ftp.ee.lbl.gov/

Preprint-Reihe
Institut für Informatik
Universität Würzburg

Verantwortlich: Die Vorstände des Institutes für Informatik.

[128] N. Gerlich. *A Toolkit of Octave Functions for Discrete-Time Analysis of Queuing Systems.* Dezember 1995.

[129] M. Ritter. *Network Buffer Requirements of the Rate-Based Control Mechanism for ABR Services.* Dezember 1995.

[130] M. Wolfrath. *Results on Fat Objects with a Low Intersection Proportion.* Dezember 1995.

[131] S. O. Krumke and J. Valenta. *Finding Tree–2–Spanners.* Dezember 1995.

[132] U. Hafner. *Asymmetric Coding in (m)-WFA Image Compression.* Dezember 1995.

[133] M. Ritter. *Analysis of a Rate-Based Control Policy with Delayed Feedback and Variable Bandwidth Availability.* Januar 1996.

[134] K. Tutschku and K. Leibnitz. *Fast Ray-Tracing for Field Strength Prediction in Cellular Mobile Network Planning.* Januar 1996.

[135] K. Verbarg and A. Hensel. *Hierarchical Motion Planning Using a Spatial Index.* Januar 1996.

[136] Y. Luo. *Distributed Implementation of PROLOG on Workstation Clusters.* Februar 1996.

[137] O. Rose. *Estimation of the Hurst Parameter of Long-Range Dependent Time Series.* Februar 1996.

[138] J. Albert, F. Räther, K. Patzner, J. Schoof, J. Zimmer. *Concepts For Optimizing Sinter Processes Using Evolutionary Algorithms.* Februar 1996.

[139] O. Karch. *A Sharper Complexity Bound for the Robot Localization Problem.* Juni 1996.

[140] H. Vollmer. *A Note on the Power of Quasipolynomial Size Circuits.* Juni 1996.

[141] M. Mittler. *Two-Moment Analysis of Alternative Tool Models with Random Breakdowns.* Juli 1996.

[142] P. Tran-Gia, M. Mandjes. *Modeling of customer retrial phenomenon in cellular mobile networks.* Juli 1996.

[143] P. Tran-Gia, N. Gerlich. *Impact of Customer Clustering on Mobile Network Performance.* Juli 1996.

[144] M. Mandjes, K. Tutschku. *Efficient call handling procedures in cellular mobile networks.* Juli 1996.

[145] N. Gerlich, P. Tran-Gia, K. Elsayed. *Performance Analysis of Link Carrying Capacity in CDMA Systems.* Juli 1996.

[146] K. Leibnitz, K. Tutschku, U. Rothaug. *Künstliche Neuronale Netze für die Wegoptimierung in ATG Leiterplattentestern.* Juli 1996.

[147] M. Ritter. *Congestion Detection Methods and their Impact on the Performance of the ABR Flow Control Mechanism.* August 1996.

[148] H. Baier, K.W. Wagner. *The Analytic Polynomial Time Hierarchy.* September 1996.

[149] H. Vollmer, K.W. Wagner. *Measure One Results in Computational Complexity Theory.* September 1996.

[150] O. Rose. *Discrete-time Analysis of a Finite Buffer with VBR MPEG Video Traffic Input.* September 1996.

[151] N. Vicari, P. Tran-Gia. *A Numerical Analysis of the* $Geo/D/N$ *Queueing System.* September 1996.

[152] H. Noltemeier, S.O. Krumke. *30. Workshop Komplexitätstheorie, Datenstrukturen und effiziente Algorithmen.* Oktober 1996.

[153] R. Wastl. *A Unified Semantical Framework for Deductive Databases.* Oktober 1996.

[154] R. Wastl. *A Vectorial Well-Founded Semantics for Disjunctive, Deductive Databases.* Oktober 1996.

[155] G. Niemann. *On Weakly Growing Grammars.* Oktober 1996.

[156] W. Nöth, U. Hinsberger, R. Kolla. *TROY — A Tree Oriented Approach to Logic Synthesis and Technology Mapping.* November 1996.

[157] R. Wastl. *Lifting the Well-Founded Semantics to Disjunctive, Normal Databases.* November 1996

[158] H. Vollmer. *Succinct Inputs, Lindström Quantifiers, and a General Complexity Theoretic Operator Concept.* November 1996

[159] H. Baier. *On the Approximability of the Selection Problem.* Dezember 1996

[160] U. Hafner, S.W.M. Frank, M. Unger, J. Albert. *Hybrid Weighted Finite Automata for image and video compression.* Januar 1997

[161] N. Gerlich. *On the Spatial Multiplexing Gain of SDMA for Wireless Local Loop Access.* Januar 1997

[162] M. Dümmler, A. Schömig. *Discrete–time Analysis of Batch Servers with Bounded Idle Time and Two Job Classes.* Januar 1997

[163] U. Hinsberger, R. Kolla, M. Wild. *A parallel hybrid approach to hard optimization problems.* Januar 1997

[164] M. Ritter. *Analysis of a Queueing Model with Delayed Feedback and its Application to the ABR Flow Control.* Januar 1997

[165] R. Wastl. *Unfolding in Disjunctive Deductive Databases with respect to 3-Valued Stable Models.* Januar 1997

[166] W. Nöth, R. Kolla. *Node Normalization and Decomposition in Low Power Technology Mapping.* Februar 1997

[167] R. Wastl. *Tableau Methods for Computing Stable Models and Query Answering in Disjunctive Deductive Databases.* März 1997

[168] S. Bartelsen, M. Mittler. *A Bernoulli Feedback Queue with Batch Service.* März 1997

[169] M. Ritter. *A Decomposition Approach for User-Network Interface Modeling in ATM Networks.* April 1997

[170] N. Vicari. *Resource-Based Charging of ATM Connections.* April 1997

[171] K. Tutschku, T. Leskien, P. Tran-Gia. *Traffic estimation and characterization for the design of mobile communication networks.* April 1997

[172] S. Kosub. *On cluster machines and function classes.* Mai 1997

[173] K. W. Wagner. *A Note on Parallel Queries and the Difference Hierarchy.* Juni 1997

[174] S. Bartelsen, M. Mittler, O. Rose. *Approximate Flow Time Distribution of a Queue with Batch Service.* Juni 1997

[175] F. Duckstein, R. Kolla. *Gültigkeitsmetriken für animierte gerenderte Szenen in der Echtzeitcomputergraphik.* Juni 1997

[176] O. Rose. *A Memory Markov Chain Model For VBR Traffic With Strong Positive Correlations.* Juni 1997

[177] K. Tutschku. *Demand-based Radio Network Planning of Cellular Mobile Communication Systems.* Juli 1997

[178] H. Baier, K. W. Wagner. *Bounding Queries in the Analytic Polynomial-Time Hierarchy.* August 1997

[179] H. Vollmer. *Relating Polynomial Time to Constant Depth.* August 1997

[180] S. Wahler, A. Schoemig, O. Rose. *Implementierung und Test neuartiger Zufallszahlengeneratoren.* August 1997

[181] J. Wolff von Gudenberg. *Objektorientierte Programmierung im wissenschaftlichen Rechnen.* September 1997

[182] ???. *???.* September 1997

[183] S. Kosub, H. Schmitz, H. Vollmer. *Uniformly Defining Complexity Classes of Funcions.* September 1997

[184] N. Vicari. *Measurement and Modeling of WWW-Sessions.* September 1997