The WWW-Service in the AMUSE Field Trials: Usage Evaluation and Traffic Modelling

Johannes Färber [1], Matthias Frank [2], Joachim Charzinski [1]

[1] *University of Stuttgart, Institute of Communication Networks and Computer Engineering Pfaffenwaldring 47, D-70569 Stuttgart, email: {faerber|charzinski}@ind.uni-stuttgart.de*

[2] *University of Bonn, Institute of Computer Science IV Römerstraße 164, D-53117 Bonn, email: matthew@cs.uni-bonn.de*

Abstract

In this paper results of the WWW service evaluation of the first field trials of the AMUSE project are presented. The log data of the WWW service usage allows to characterize the user activity concerning online time and the amount of requested data. Size and interarrival time of the requested documents are used to describe the properties of the monitored traffic of each user. A simple model is presented that generates WWW user traffic of similar characteristics for the purpose of simulative system evaluation.

1 Introduction

Within the European ACTS project AMUSE (Advanced Multimedia Services to Residential Users), multimedia field trials have been carried out in six European cities. Their goal was to offer entertainment and information services over ATM-based broadband access networks to residential users. Via their Set Top Boxes (STB) the users had access to Video on Demand (VoD) and News on Demand (NoD) services as well as the World Wide Web (WWW). At this moment most trials have finished their first phase and first results are available.

In this paper we present the results on user behaviour of the most popular service, World Wide Web browsing. After a short description of the AMUSE trials and their configuration we present the observed usage data and traffic characteristics. These characteristics are used to build a simple traffic model for the purpose of the simulation of a larger user population.

2 The AMUSE project and trials

The AMUSE consortium comprises several companies, telecom operators, research centers and universities from all over Europe and started in 1995 to specify common guidelines and build up the necessary infrastructure for the planned trials. The project aims to develop, implement and offer advanced multimedia services to residential users. All six field trials of phase 1 in Milan, Aveiro, Munich, Basle, Reykjavík and Mons where these new services are tested, are based on an ATM infrastructure. The trials differ in terms of the services offered (VoD, NoD, WWW, SVB - Switched Video Broadcast, Teleshopping, Games, and others), the access network technology (Hybrid Fibre Coax HFC, Asymmetric Digital Subscriber Line ADSL, Fibre To The Curb FTTC, Hybrid Fibre Radio HFR) and the customer premises equipment (STB connected to the TV, a PC or a Network Computer NC). Figure 2-1 shows the reference configuration. More information on the network architecture may be found in [1].

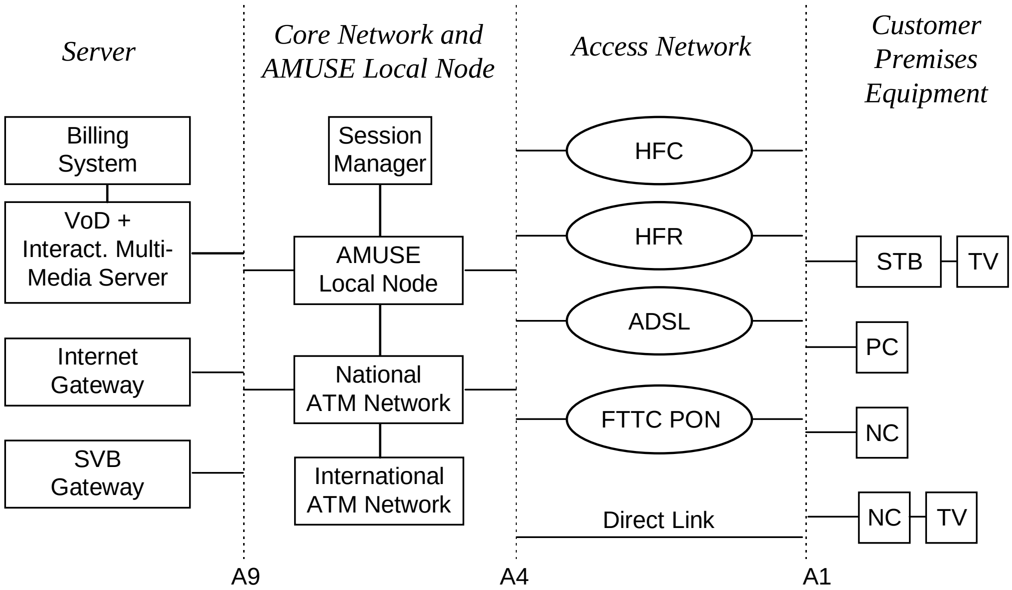

Figure 2-1: AMUSE Reference Configuration

The trials are carried out in two phases. Phase one is used to build up infrastructure and contents. Focus lies on the feasibility and the deployment of available technology. The first phase 1 trial started in May 1996 and the last will be finished in September 1997. A detailed description of the trial configurations, the services offered and experiments performed in phase 1 may be found in [5].

In phase two the trial islands will deploy more sophisticated services and technology like advanced Set Top Boxes capable of MPEG2 video or a HFR access network (Hybrid Fiber Radio) to link to the homes. All phase 2 trials will have their operation phase in 1998, a detailed description of these trials may be found in [6].

The first phase of the AMUSE trials has almost finished with overall success, the operation phase of the Belgium trial in Mons will be finished soon. A report on the major results of the service evaluation of the first AMUSE trial in Munich, Germany, has been published in [2]. A comprehensive summary of all phase 1 trial results will be presented after the final trial evaluations in [8].

3 User behaviour

The trials have been an excellent opportunity to gather data on real residential usage of multimedia services over broadband access networks. Apart from a user survey based on questionnaires and interviews to learn about attractiveness and acceptance of the offered services and interfaces, all requests issued by the users have been automatically monitored and logged. The following evaluation focuses on the analysis of the logging data for the WWW service and is intended to describe the actual usage of the system as well as the usage behaviour of a user. The analysis of the log data is divided into two parts: The first part basically deals with the usage of the service in terms of number of http requests, amount of data transferred and time spent on consuming the service. In the second part, stochastical properties are derived from the log data.

To learn about the overall acceptance and the user behaviour of the trials, different profiles of usage (like profile for the 7 days of a week or hourly intervals of a day) are built. They allow to determine service rush-hours for network and resource planning purposes and to derive parameters for the traffic model presented in the following section.

The logging functionality of the WWW proxy server records each http request that is received from the residential users or from external sources. Each logfile entry contains a time stamp indicating the time when the http daemon has served this request (also indicating the source address, data amount, URL requested and others). However, due to the centralized logging within the server, the logfile contains no information about the ending time of the WWW session (either by selecting a different service or switching off the set top box). Thus, from the raw log data produced by the proxy server it was not possible to derive any information about session length of the users on the WWW service.

This drawback has been overcome by implementing a program parsing the WWW logfiles and extracting the log data for individual residential users. The end of a session is recognized when the interarrival time between two consecutive requests coming from the same user is larger than a pre-defined idle interval. Of course, the selection of the idle interval as an input parameter for the parse program has a strong impact on the reliability on the output results of the program: The number of total sessions analysed may vary between a very high value (i.e. each request equals a separate session) and low values (i.e. the whole trial is assumed to be one session). Thus, the most appropriate selection of the idle interval would be a small value, where there is no major change of results when varying the interval in small steps.

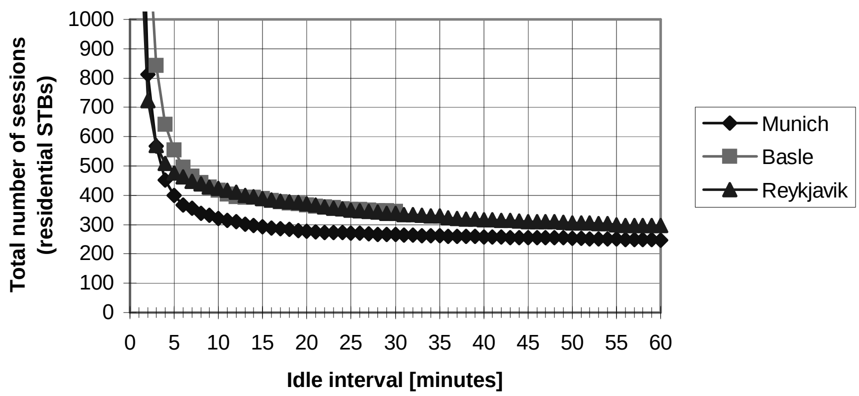

Figure 3-1: Selection of idle interval value, Munich trial evaluation

Figure 3-1 depicts this aspect for the analysis of the WWW session length for all trials where a WWW session analysis has been performed so far, namely Munich, Basle and Reykjavík. The plots show the total number of sessions (analysed for all residential STBs) produced by the parse program for a varying idle interval from 1 to 60 minutes (where the values larger than 1.000 sessions are cut off for reasons of readability of the figure). Although the total number of sessions is different between the trials and even the length of the WWW logging period is different, a value of about 16 minutes for the idle interval seems to be an appropriate value for the session analysis in all three trials.

The following WWW session averages for different issues have been evaluated from all phase 1 trials where a logging of WWW usage has been available (except for the Mons trial, where a complete analysis of WWW logging was not performed yet):

Table 3-1 indicates that the average values vary between the trials caused by the limited number of users and obvious differences in the user behaviour. Thus, the selection of values for traffic modelling based on these results has to be done with considerable care.

	Munich	Basle	Reykjavík
average session length [h:m:s]	0:28:37	00:44:27	00:13:44
average # requests per session	236	483	87
average amount of data per session [Mbyte]	1,276	3,585	0,603
total # sessions	288	382	385
total usage duration [h:m:s]	137:23:00	283:04:50	88:12:03
length of evaluation period [days]	98	105	46
number of residential users	11	10	10

Table 3-1: WWW session averages for phase 1 trials

The following three figures show usage profiles in terms of requested WWW data volume for each day of the week and for each hour of the day, separately for the three trials. The numbers are averages of the data volume accumulated for residential STBs over all weeks for each day of the week (Sa - Fr) of the corresponding trial (for the week profile) and all days of the trial for each interval of 1 hour (for the day profile). All figures also show the 90% confidence levels of the average.

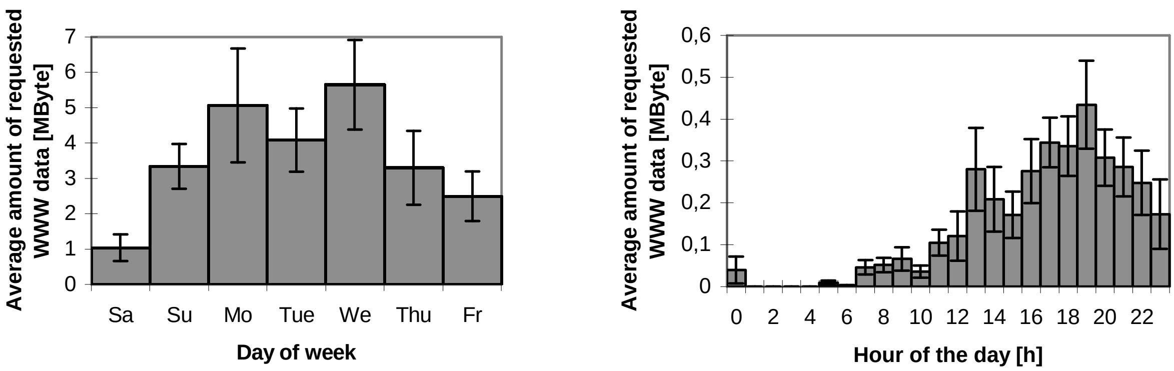

Figure 3-2: Profile of WWW usage, (a) per week, (b) per day (from Munich analysis)

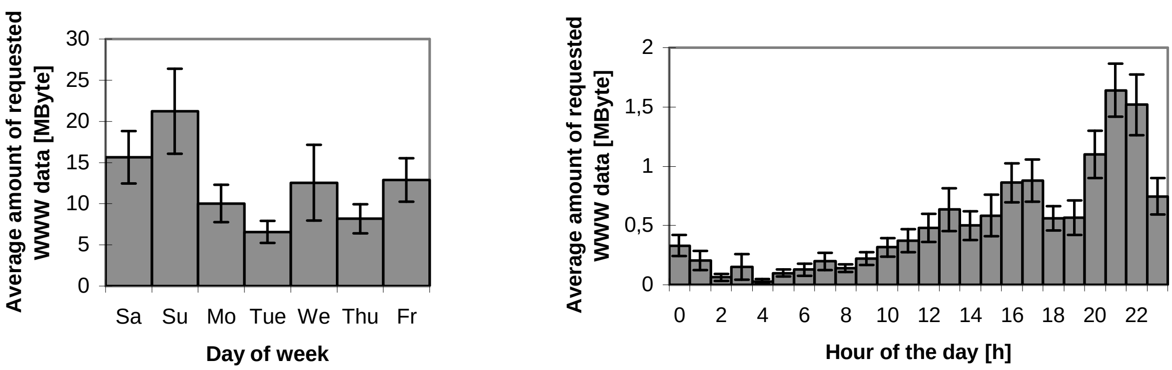

Figure 3-3: Profile of WWW usage, (a) per week, (b) per day (from Basle analysis)

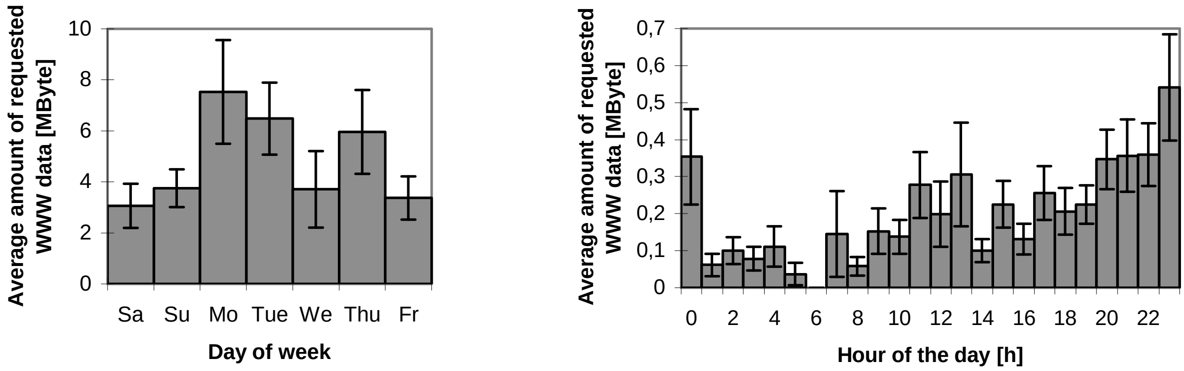

Figure 3-4: Profile of WWW usage, (a) per week, (b) per day (from Reykjavík analysis)

Again due to the limited number of residential users in each trial, the week profiles differ significantly between the trials and hardly any general conclusion can be drawn or model parameters be extracted from these profiles. Although the day profiles also vary on the average numbers for the intervals of 1 hour, a general tendency of higher usage during the late afternoon and the night hours is visible. This property and possibly the averages observed of requested data amount per hour interval may be used for a long term traffic model or for a traffic characterisation of peak hours of service usage.

In order to describe the characteristical behaviour of WWW traffic generated by a user, the most efficient approach is to just look at the interarrival time and the size of the requested documents. These two measures are sufficient to describe the traffic load of a system. Both values follow a certain probability distribution and can be represented by the corresponding distribution functions. This approach considers only interarrival time but no correlation between consecutive requests or different users. Any information on the relation of traffic behaviour and hour of the day or day of the week is lost. However, the resulting description is useful for basic performance evaluations.

The interarrival time of requested packets is an indicator for the frequency of requests. Figure 3-5 shows the probability of an interarrival time being greater than the value on the x-axis (complementary distribution function). The curves can be divided into two regions. First the curves drop quickly for small interarrival times, but then almost stagnate at a certain probability level. This behaviour results from the fact that the logging data contains the interarrival times within a WWW session as well as the time spans between the sessions. The heavy tails of these curves impose a problem for the modelling of such traffic since a mathematical function has to be found that fits the given form.

However, the heavy tail is based on few very long pauses between sessions. Since the trials were active for about three months each, it is only natural that long pauses in the logging data are rare. Thus the longterm data are not representative and the significance of the differences in the stagnation level of the heavy tail of each user is questionable. In the following section this longterm behaviour will not be considered for the design of the model.

Besides the interarrival time, the request sizes have to be considered to estimate the generated traffic load. The complementary distribution functions for the request sizes of all STBs of each trial are depicted in Figure 3-6. The curves show the probability of a request being of greater size than the value on the x-axis.

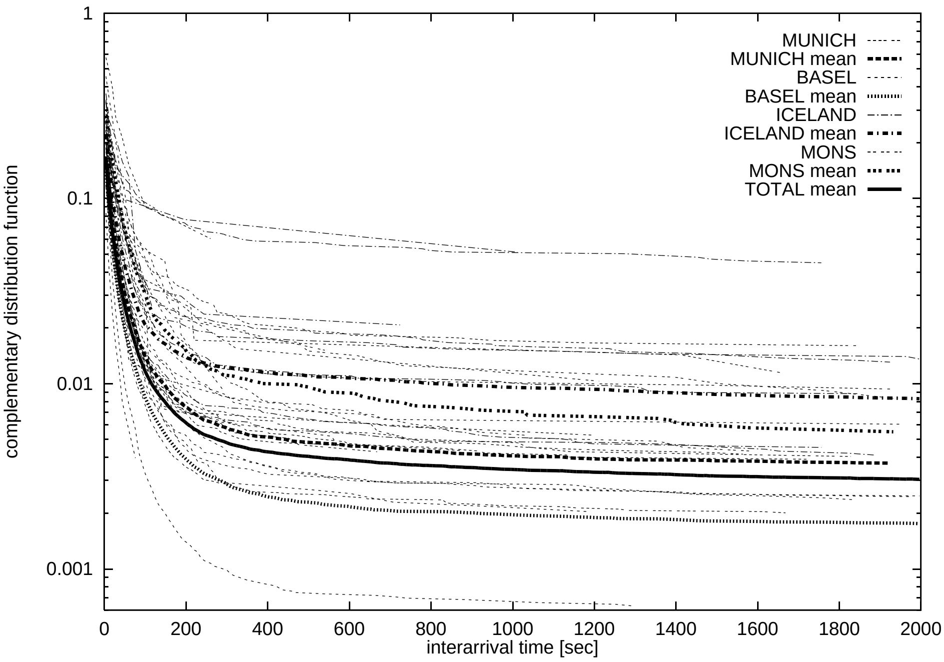

Figure 3-5: Complementary Distribution Functions for http request interarrival times

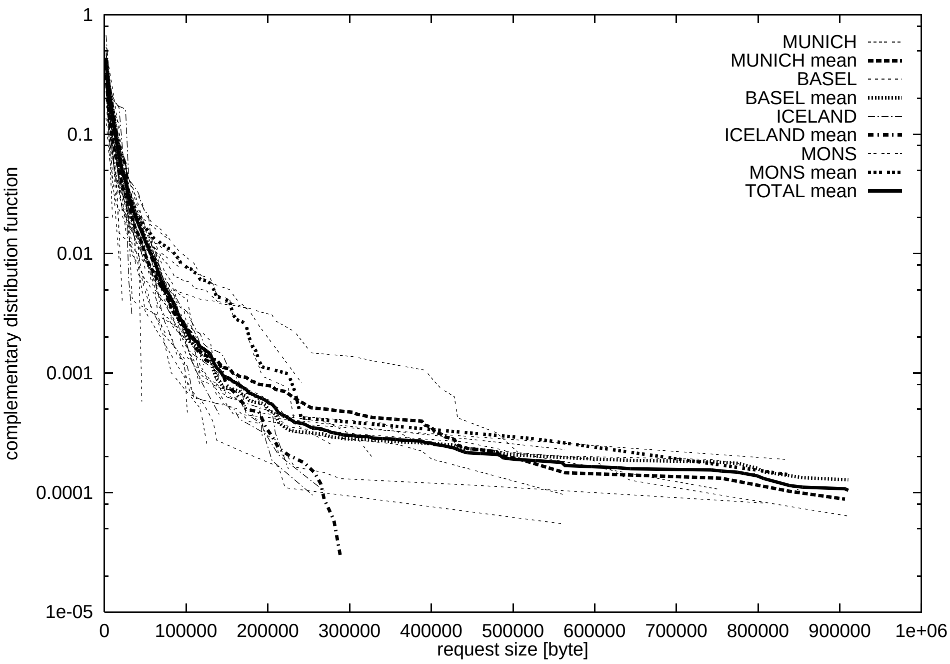

Figure 3-6: Complementary Distribution Functions for http request sizes

While sizes of most requested packets range between 0 and 200 kbytes, there are few packets of real large sizes up to 8 Mbytes (large text files, program files and images). Again, the shapes of the probability curves have an unfortunate characteristic, since they don't indicate any upper bound for request sizes. It is difficult to define a mathematical function of a similar shape for the modelling of the WWW-traffic.

4 Traffic model

Due to the small number of participating users in the AMUSE trials, it is difficult to make empirical statements on system performance of the selected configurations. Since the focus lies on reliability and optimized services for the users, a measurement under heavy load will not be done within the future AMUSE trials either. With a model that produces characteristic traffic of a user, it is possible to perform simulations of large user populations as described in [7].

Such models have already been suggested e.g. in [3] and [4]. They focus on a detailed description of burstiness within a single session. However, here the correlation between consecutive requests of the logged data was found to be very small. Thus a model neglecting these correlations seems to be sufficient, even if it will not provide on/off characteristic like the ones mentioned above do.

The model that was designed based on the analysis of the AMUSE log data generates similar requests to those generated by a real user concerning size and interarrival time. It consists of two generators for request sizes and for interarrival times. Since both measures where found to be uncorrelated, the model for their arrival processes can be based on iid random processes i.e. their output does not depend on the history of previous output. This property allows to follow the selected approach.

The complementary distribution functions of request size and interarrival time presented in the previous section can be used for such a model. For this, mathematical functions are fitted to the complementary empirical distribution funtions. They are used to calculate the interarrival time and request sizes of each newly generated request.

First, a suitable probability distribution function describing the arrival process of the interarrival time has to be found. If the fit is performed to the complete set of log data, a good choice is a hyperexponential function of the order four. To keep the characteristic shape of the complementary distribution function, one of the four exponential function must be fitted to the heavy tail. It is excluded from the following fitting operation since the heavy tail would be neglected otherwise because of small probabilities and the limited fitting range. The four exponential functions lead to four mean interarrival times that can be interpreted as pauses while the user is idle, active, browsing through pages or initiating a burst (e.g. if inline images of a document are loaded simultaneously).

However, the selection of the parameters determining the long term behaviour is arbitrary. Since the time scales differ enormously the mean interarrival time of the idle state does not affect the other parameters. For example a fit to a function fixed to 3 days as mean long term interarrival time would lead to almost identical results than a value of 10 days. Since the logged longterm data are not representative as discussed above, it is not possible to estimate a suitable mean value with a tolerable confidence level.

To avoid this arbitrariness, focus is put on the session period only. By omitting the idle state, the model is reduced to a session model that produces traffic like an active user. A hyperexponential probability density function (pdf) of the order three is now sufficient to describe the characteristics of the interarrival time:

$$pdf\,(t) = p_1\lambda_a e^{-\lambda_a t} + p_2\lambda_w e^{-\lambda_w t} + (1 - p_1 - p_2)\lambda_b e^{-\lambda_b t}$$

$$CDF(t) = \int_t^{\infty} pdf\,(t)dt = p_1 e^{-\lambda_a t} + p_2 e^{-\lambda_w t} + (1 - p_1 - p_2)e^{-\lambda_b t}$$

In the previous section a value of 16 min was suggested as a reasonable boundary between two sessions. Thus the complementary distribution function (CDF) of the selected function is fitted to the CDF of the filtered log data containing only interarrival times smaller than 16 minutes (Figure 4-7). Of course the exact cut at 16 minutes causes an error which is indicated by the sudden decrease of the functions tail at the right end. By performing a fitting operation without considering this part of the function, this error is reduced and can be neglected.

The fitting was realized with the Marquardt Levenberg algorithm with fixed parameters for the tail. The deviation of the fitted function at 100 seconds is due to the fact that a better fit of smaller values (1-50 sec) with higher probabilities is much more important.

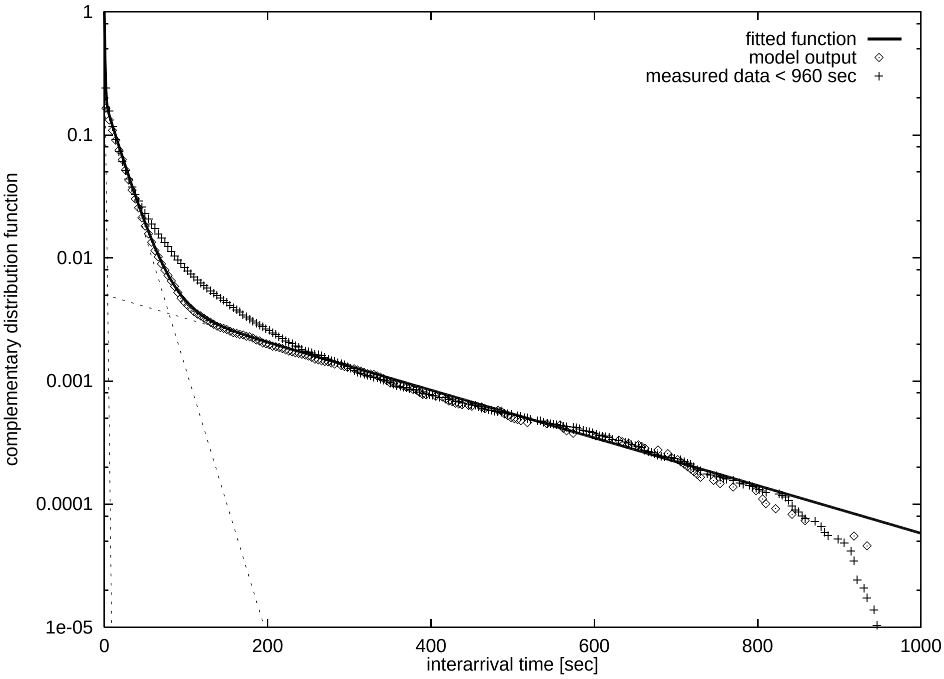

Figure 4-7: A fit of a hyperexponential function of the order 3 to filtered log data leads to a complementary distribution function of the interarrival time of an active session. A model based on these results produces traffic with a very similar characteristic.

Table 4-1 depicts the parameters for the hyperexponential function resulting from the fitting operation. Since the chosen function consists of the sum of exponential functions, it is easy to derive a simple corresponding state model which is shown in Figure 4-8.

State	Probability	mean interarrival time
Active	p_1 = 5.02646575e-03	$1/\lambda_a$ = 2.240881444e+02 sec
Browse	p_2 = 1.88319316e-01	$1/\lambda_w$ = 1.991215082e+01 sec
Burst	1- p_1 - p_2	$1/\lambda_b$ = 0.793234805e+00 sec

Table 4-1: Parameters for the generation of the interarrival time within a session

If a new interarrival time has to be determined, the model selects with the given probability state A (active), state W (browse) or state B (burst) and returns a value according to the states mean interarrival time. After the request was issued (state R) a new cycle will begin until the model is stopped.

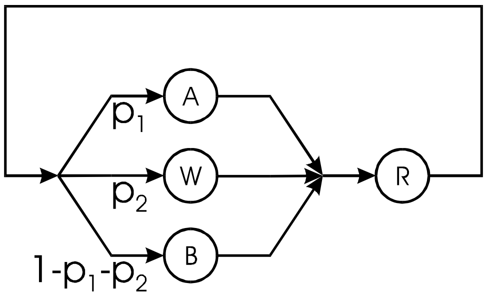

Figure 4-8: Generator model for interarrival times during an active session

The output of the model was evaluated in the same way as the measured data. The CDF resulting from this output is shown as well in Figure 4-7. Note that the probability for large interarrival times within a session is quiet small. To receive the depicted CDF, the model had to simulate a 7 day session to get enough large interarrival times for an evaluation. This is the reason why the CDF of the model output differs from the curve of the fitted function at the right side of the figure.

The second part of the traffic model is the generation of request sizes for each issued request. We assume that request sizes and interarrival times are not correlated and that we can use an iid random source for request size generation as well.

Since the complementary distribution function of the request size has a similar shape as the one of interarrival time, a similar function can be used for the modelling. As above, a hyperexponential function of the order 3 is sufficient to produce a very similar output. In Figure 4-9 the fitted function as well as the output of the model based on this function are depicted.

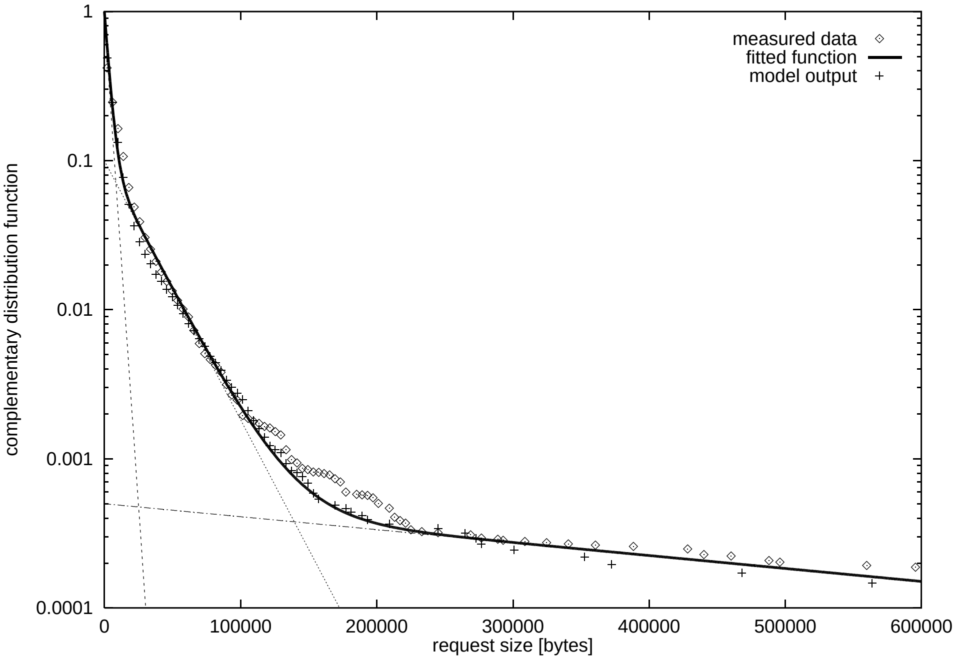

Figure 4-9: Fitting of a hyperexponential function of the order 4 to the complementary distribution function of the request size with Marquardt-Levenberg method

state	probability	mean request size
small	0.937	5.1 kbyte
medium	0.062	30.3 kbyte
large	0.001	673.2 kbyte

Table 4-2: Parameters for the generation of the request size

The fitting operation to the measured data leads to the parameters which are listed in Table 4-2 and are used for the corresponding state model (Figure 4-10). If a request is issued and the corresponding size has to be determined, the model selects with the given probability state S (small), state M (medium) or state L (large) to determine the request size according to the selected states mean request size.

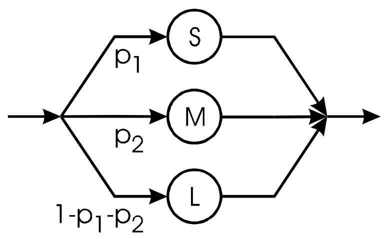

Figure 4-10: Model for request sizes

The presented model generates values for interarrival time and request size according to the fitted probability distribution functions. These values should serve as input of an event driven simulation of a system under evaluation. Due to the simplicity of the model, it is suited for the simulation of many simultaneously active users. Note that the model is based on log data on the proxy server and therefore simulates downstream data and not the request behaviour of the user.

The model simulates traffic of an average user and is based on a single mean profile. Further studies are needed to verify whether more user types lead to significantly other results when simulating multiple users. The model can be easily parameterized to support any other user type if the corresponding characteristics can be extracted.

5 Conclusion

The paper presented the evaluation of the WWW traffic monitored during the first field trials of the ACTS project AMUSE. Although the user base was rather small, clear tendencies in user behaviour and session characteristics have been found. When focusing on interarrival times in order to extract common characteristics, long term interarrival times should be ignored. However, by cutting off valid data, a small statistical error occurs. Log data on the session duration i.e. starting and ending time would be preferable to avoid this error.

Based on the session characteristics a simple model was designed and described. The model is suited to simulate traffic of multiple simultaneous active users on a system under evaluation and can be easily modified to support more types of user behaviour. All parameters can easily be interpreted as reasonable values like e.g. mean packet size and probability for certain data types or mean interarrival time during a burst period. When using the model it has to be considered that the output needs to be processed (e.g. segmentation into ATM cells or IP packets) in order to provide useful input for the system under evaluation. The suggested usage of the model for simulative performance evaluation is explained in [7].

Further studies on the accuracy of the model have to be done. The traffic characteristics are to be compared with data from following trials within the project as well as with results of other trials if available. This will also allow for a comparison between AMUSE users (residential users) and a more technical oriented user group (university staff) or maybe user groups of other field trials. The model based on the AMUSE data is not limited to the use within the AMUSE project. Its potential for improvement has been pointed out and more detailed traffic output as well as different user types will be supported.

6 References

[1] J. Charzinski, J. Färber, J. Enssle, M. Frank. The AMUSE Residential Multimedia Trials: Network Architecture and First Results. NOC´97, Antwerp, Belgium, 17-19 June 1997

[2] H. Elbert, J. Enssle, M. Frank, H. Hessenmüller, E.T. Hvannberg, T. Theimer, Multimedia Service Provisioning in the Munich Trial of the AMUSE Project. ECMAST´97, Milano, Italy, 21-23 May 1997

[3] S. Deng, Empirical Model of WWW Document Arrivals at Access Link, Proceedings of ICC'96, September 1996, pp. 1797-1802.

[4] E.A. Miltiades, J.-A. Sanchez-P, I.S.Venieris, A Multiservice User Descriptive Traffic Source Model, IEEE-Transaction on Communications Vol.44 No.10, Oktober 1996, pp.1243-1246.

[5] AMUSE Project Deliverable D3. Objectives, guidelines, platform definition and detailed planning for demonstration Phase 1. AC011/ITL/WP21/DS/P/003/b3, February 1996

[6] AMUSE Project Deliverable D11. Objectives, guidelines and final platform definition for demonstration Phase 2. AC011/ITL/WP21/DS/P/011/b3, July, 1997

[7] AMUSE Project Deliverable D12. Performance Evaluation of the Delivery System. AC011/UBN/WP21/DR/P/012/c1, August 1997

[8] AMUSE Project Deliverable D14. Final Report on the service related and technical/functional outcomes of the Phase1 demonstrations. AC011/SGCS/WP24/DR/P/014/??, to be released in October 1997