University of Würzburg
Institute of Computer Science
Research Report Series

QoS of Internet Access with GPRS

Dirk Staehle[1], Kenji Leibnitz[1], and Konstantin Tsipotis[2]

Report No. 283 January 2002

[1] University of Würzburg,
Dept. of Computer Science,
Am Hubland, 97074 Würzburg, Germany
[staehle, leibnitz]@informatik.uni-wuerzburg.de

[2] Lucent Technologies,
Windmill Hill Business Park,
Swindon, Wiltshire, SN5 6PB, UK
tsipotis@lucent.com,

Work was done while author
was with Libertel-Vodafone, Netherlands

QoS of Internet Access with GPRS

Dirk Staehle, Kenji Leibnitz
University of Würzburg,
Dept. of Computer Science,
Am Hubland, 97074 Würzburg,
Germany,
[staehle, leibnitz]
@informatik.uni-wuerzburg.de

Konstantin Tsipotis
Lucent Technologies,
Windmill Hill Business Park,
Swindon, Wiltshire, SN5 6PB, UK
tsipotis@lucent.com,

Work was done while author
was with Libertel-Vodafone, Netherlands

Abstract

In this paper we examine the effects of Internet traffic on the quality of service (QoS) in GPRS wireless networks. With a stochastic source traffic model describing the user behavior, we will derive subjective and objective quality of service measures in terms of WWW downloading time and the transmission bandwidths on TCP and TBF level. Comparing the obtained values with wireline network modems and ISDN yields a slowdown factor which indicates the subjective degradation that the wireless user experiences.

1 Introduction

The recent development in telecommunication networks has revealed two important tendencies. The first one is the huge drive in Internet traffic. Many applications like browsing the World Wide Web (WWW) or Electronic Mail (Email) have become a part of everyday life. Secondly, the demand for wireless services has caused wireless networks to extend at an enormous rate. With the introduction of *General Packet Radio Service* (GPRS) for wireless networks operating with the *Global System for Mobile Communication* (GSM) network providers aim at closing the gap between these two trends.

GPRS aims at accommodating packet data connections with variable bit rates (up to several ten kbps) operating simultaneously with GSM circuit switched connections and permits a broad range of existing and new applications [1, 2]. Packets originating from a GPRS mobile station can be directly transmitted to IP networks, since the GPRS support nodes (GSN) will use IP as the backbone protocol for transfer and routing of the protocol data units. This feature makes it very appealing for TCP/IP based applications.

Additionally, the GSM architecture only needs to be enhanced by introducing these additional support nodes, most of the existing architecture is left untouched. In general, GPRS aims at coexisting with GSM by utilizing the capacities not used by GSM and may assign currently available channels to packet data users. The higher bandwidth efficiency is achieved by a statistical multiplexing of the data connections on the packet data channels. The result is a better utilization of the radio frequencies. The dynamic behavior of GPRS with its *capacity on demand* concept requires the GSM network operators to modify their current network design approaches. While the planning process of a GSM network is dominated by radio propagation and frequency planning aspects, optimal GPRS design involves several issues. At first, the parameters of the radio interface, e.g. degree of multiplexing, ratio of fixed/on-demand channels, must be chosen to suit the traffic demand [3]. Secondly, the wired network infrastructure should be properly dimensioned such that there are no capacity bottlenecks on the transport links [4]. And finally, the service disciplines play an important role [5]. In [6] the service quality for Internet applications was investigated for the allocation of both on-demand and fixed packet data channels.

Considering these challenges, the network provider faces the problem that the existing GSM network should not be influenced by the additional service. Therefore, it is expected that in the first phases of the introduction of GPRS most providers will not assign any fixed capacity, but operate purely with on-demand channels. At a later stage, additional capacity in terms of further carriers can be assigned to cells with capacity shortage.

Another problem that arises is that the most important input parameter for planning, i.e. the actual traffic behavior of a GPRS user, is yet unknown. From the current situation it can be expected that most of the traffic will be generated from the applications WAP (Wireless Application Protocol), WWW, and Email. However, it is not clear how the behavior of a typical GPRS user will differ from that of a wireline access user in the circumstances of an erroneous radio link with lower access bandwidth. Certainly, this depends on the quality of service of his Internet access, which is influenced by the number of other GPRS users in the cell he shares his bandwidth with and other external factors, e.g. billing. The aim of an optimal network planning is, therefore, to identify the point where the QoS of a user is regarded as unsatisfactory and to add appropriate carriers in this case.

A growing number of papers on GPRS design and dimensioning can be found in the literature. In [7] an analytical Markov chain model is presented to compute the basic performance measures in the cell. However, the main problem is that a lot of approximations must be made in order to get an analytically tractable model. For example, the model in [7] considers only exponentially distributed file sizes and no feedback behavior from

TCP. For such reasons, performance evaluation of GPRS is often done on the basis of simulations, e.g. [3] and [6]. The advantage of this approach is that a detailed mapping of the protocol can be done which is capable of illustrating the influences of the parameters on the system performance. A further point of interest is the interaction of TCP and the RLC protocol of GPRS. In [8] the performance of TCP over GPRS was investigated and it was shown that GPRS and TCP cooperate well. This was also stated in [9] where the focus was on the quality of WWW access over EDGE (Enhanced Data Rates for GSM Evolution). In [10] the performance of streaming services over EDGE was investigated. The work in [11] studies the influence of dedicated data channels on the throughput of web traffic. The design of the RLC and MAC layer of EGPRS (Enhanced GPRS) were the subject of [12].

In this paper, we will therefore examine the effects of Internet traffic on the GPRS quality of service (QoS) by means of simulation. With a detailed stochastic source traffic model describing the user behavior, we will derive subjective and objective quality of service measures in terms of WWW downloading time and the transmission bandwidths at TCP and radio link layer. Comparing the obtained values with wireline network modems and ISDN yields a slowdown factor which indicates the subjective degradation that the wireless user sees.

This paper is organized as follows. In Section 2 we describe the stochastic source traffic model and give details on the simulation of the GPRS protocol architecture. This is followed in Section 3 by the evaluation of the quality of service of Internet traffic. We also compare the results with those from wired Internet access and obtain a measure to evaluate GPRS bandwidth and downloading time.

2 Modeling User-Network Interaction

The simulation aims at determining the quality of the wireless access to the Internet using GPRS under certain load conditions in a cell. The traffic load consists of voice traffic and data traffic. The amount of voice traffic offered in a cell is defined by an Erlang value.

The situation concerning data traffic is more complicated, since its amount and its characteristics are influenced by the traffic load itself. The problem when simulating a wireless Internet user is that currently this kind of access does not exist, and therefore no data resulting from measurements is available. The solution is to take results from wired networks and to use them in the wireless environment, see e.g. [3, 13].

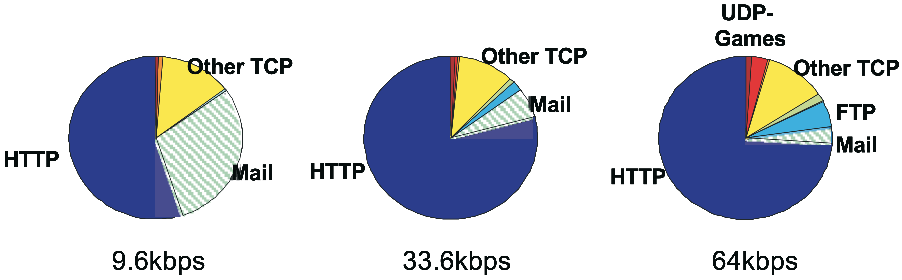

Figure 1: Composition of Internet traffic dependent on access speed

First the applications under consideration have to be specified. Vicari [14] performed measurements of the Internet modem access offered to students by the University of Würzburg. Fig. 1 shows the measured traffic composed of different applications. Obviously, the importance of an application depends on the available access bandwidth. For GPRS, the bandwidth will be between 9.6kbps and 33.6kbps. The traffic mixture for those modem speeds is given in the left and the middle pie chart. Due to these charts HTTP and Email are considered to be the most relevant applications for GPRS. With rising bandwidth FTP is gaining importance, as well; however, in a wireless environment it should be negligible. Additional to the common wired network applications the *Wireless Application Protocol* (WAP) will be of some importance to GPRS in the future. At the moment, WAP exists for GSM only, and additionally with a rather bad performance. For these reasons no measurements appropriate to derive the WAP user behavior exist and henceforth this new service is not included in the simulation.

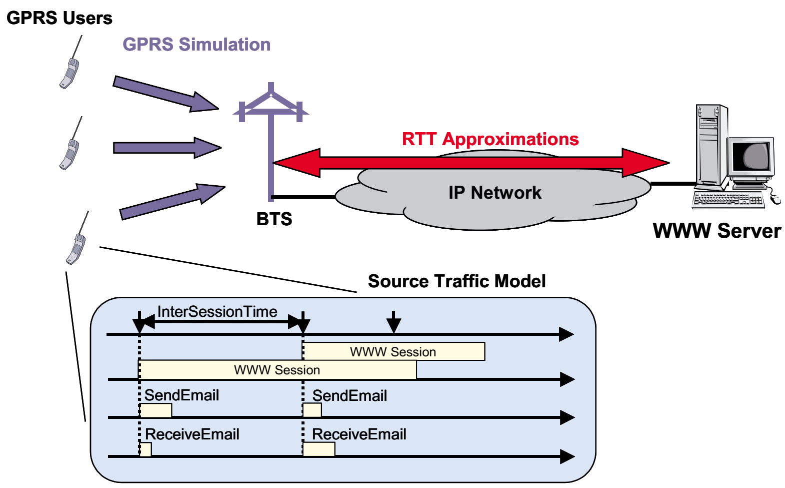

Figure 2: Basic simulation concept

All in all, the simulation is performed for two applications, WWW and Email. The basic simulation concept is illustrated in Fig. 2. Both considered applications are built on the *Transmission Control Protocol* (TCP). Besides offering a reliable connection to upper layers flow control is performed by TCP. Therefore, the characteristics of the IP packet stream arriving at the wireless link, strongly depend on TCP, such that a detailed implementation of the TCP stack is required. Furthermore, the bandwidth with which TCP offers packets to the wireless link is influenced by the latency in the wired part of the network simulated by RTT approximations. In the following sections the source traffic model for WWW and Email traffic, the simulation of TCP and the wired part of the network, and the GPRS simulation are described.

2.1 Stochastic Source Traffic Model

Currently, no measurements of wireless data users exist. Therefore, the source traffic model used as the basis for the data traffic generator implemented in the simulation relies on measurements in wired environments. In particular for HTTP traffic a lot of publications are available which characterize the behavior of a WWW user. In [15] the web traffic generator *SURGE* is described which relies on the traffic measurements of [16]. Other frequently referenced publications are [17] and [18]. Standardization organizations, like the ETSI [19] or the ITU [20], recommend traffic models for performance evaluation of wireless Internet access, as well, although without any validation from measurements.

The used source traffic model relies on two further publications, [21] and [13]. The model of [21] was chosen since it gives a detailed description of the structure and the size of a web page. The model of [13] contains a description of the session level, i.e. the user is not only observed while browsing the web but also while being passive. The resulting stochastic user traffic model combined from these publications is depicted in Fig. 3.

The source traffic model consists of two layers. The session layer contains the arrival process for web sessions and the number of pages in a web session. According to [13] the arrival process is Poisson and the number of pages belonging to one session follows a Lognormal distribution.

In the page layer, the structure of a web page is defined following [21]. A web page consists of the HTML code, referred to as *main object*, and optionally of images or other files referenced in the HTML code called *inline objects*. The size of *main and inline objects* as well as the number of inline objects per page are random variables. Furthermore, [21] gives the distribution of the time between two consecutive web pages, typically called *reading time* or *viewing time*. The used distributions and parameters are summarized in Table 1.

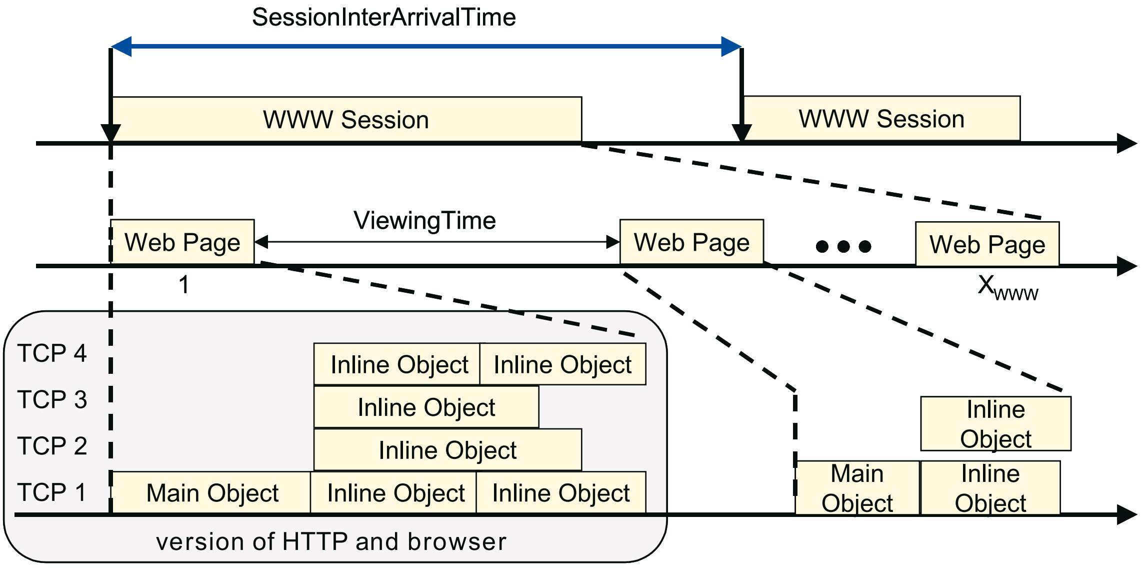

Figure 3: Source model for WWW traffic

random variable	distribution	mean	std. dev.
number of pages per web session	Lognormal	25	100
page viewing time	Weibull	39.5s	92.6s
number of inline objects per web page	Gamma	5.55	11.4
size of main object	Lognormal	10kBytes	25kBytes
size of inline object	Lognormal	7.7kBytes	126kBytes

Table 1: Parameters of the WWW traffic model

Such a detailed stochastic description of a web page is necessary due to the properties of TCP and the different HTTP versions. In particular, the *slow start* mechanism of TCP has a considerable influence on the performance of HTTP, see [22], and therefore on the downloading time for WWW pages which is one of the key quality of service parameters for the Internet access. HTTP 1.0 [23] caused a loss of performance when interacting with TCP, since every object was downloaded in an individual TCP connection. Small object sizes led to a small TCP bandwidth and long downloading times. In HTTP 1.1 [24] several changes were made to improve the performance. A comparison of HTTP versions 1.0 and 1.1 can be found in [25]. Additionally, the web browser influences the way a page is loaded, since the number of TCP connections established for a web page varies. The used HTTP implementation is demonstrated in Fig. 3. First, one TCP connection is opened to download the main object. To transmit the inline objects in parallel up to three further connections are established. Furthermore, *persistent* HTTP is implemented, which means that not only one but multiple objects are transmitted in a single TCP connection.

The second application considered in the simulation is Email. Publications describing the characteristics of Email traffic focus on the distribution of Email sizes, see e.g. [26, 1]. In Fig. 4 these Email sizes are depicted as cumulative distribution functions (CDF). Additionally, curves resulting from measurements at the University of Würzburg with about 35000 Emails are shown. The black curve which is used to generate Email sizes in the simulation represents all Emails while the grey curves correspond to the mails of single persons. The various protocols like SMTP, POP, IMAP, etc. used to send or read mail, are based on TCP. The simulation does not distinguish between the different protocols but simply transfers Emails either from the mobile to the server or vice versa. The sizes of sent and received Emails are assumed to follow the same distribution and for both the arrival process is Poisson, as it is for web sessions.

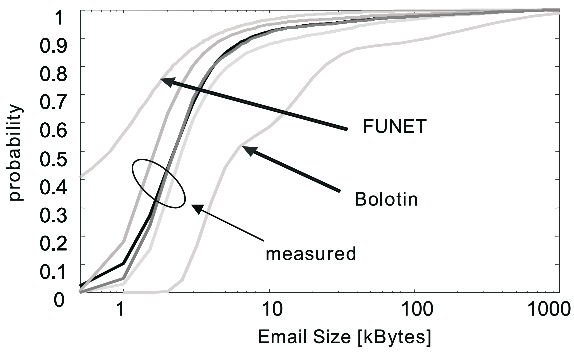

Figure 4: Distribution of Email sizes

Therefore, it is possible to combine the three arrival processes, i.e. web traffic, sending Email, and receiving Email, to a single Poisson process as illustrated in Fig. 5. For every arrival with probability p_{WWW} a web session is initiated, with probability $p_{Email}/2$ an Email is sent, and with $p_{Email}/2$ an Email is received. According to the measurements of Vicari [14] the data volume produced by WWW depends on the access speed and exceeds the Email volume by a factor of 2 to 16. For GPRS a factor of 12.5 is assumed. The amount of data in a WWW session is about 50 times larger than the size of an Email. As a consequence, Email sessions take place 4 times more frequently than WWW sessions, i.e. $p_{WWW} = 0.2$ and $p_{Email} = 1 - p_{WWW} = 0.8$. A more detailed description of the applied source traffic model can be found in [27].

2.2 TCP and HTTP Models

The applications which are of most importance to a user accessing the Internet via GPRS, HTTP and Email, are both based on TCP. Therefore, a detailed simulation of TCP is

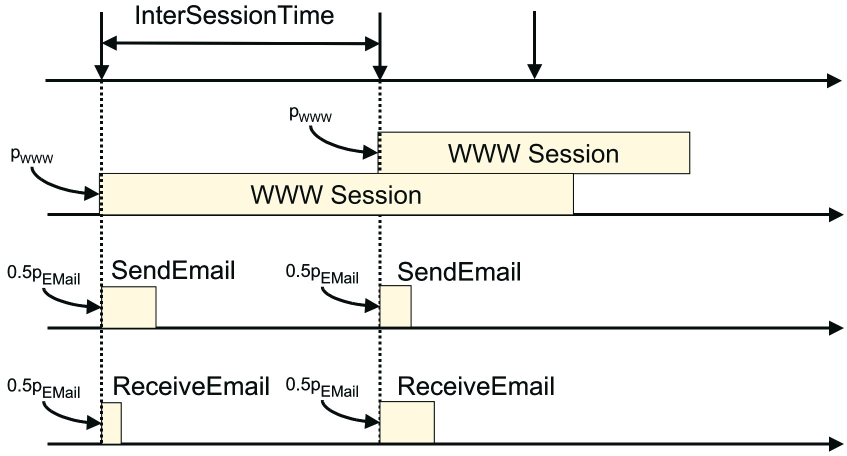

Figure 5: Illustration of the Poisson arrival process

required. Several versions of TCP may be running in the Internet depending on the configuration of the two communicating systems. In the simulation TCP *New Reno* [28] was implemented.

TCP provides a reliable connection between two systems using a *sliding window* type error recovery mechanism. By varying the size of the sender window, flow control is performed, as well. In [29] a description of the TCP mechanism is given, for more details [30] describes the code of a TCP implementation.

The flow control of TCP is reactive, i.e. the sender transmits new data only when acknowledgements arrive. The arrival rate of IP packets to the wireless link is therefore determined by the bandwidth available on this link and additionally by the transmission time of an IP packet from the base station to a server and back. From now on, this time is referred to as *wired network transmission time*.

In the simulation this time is assumed to be constant while the mobile is communicating to the same server, i.e. either sending or receiving an Email from a mail server or downloading one or more WWW pages from the same web server. In the first case, the time is rather short and assumed to be equal to 20ms for all Email servers. The wired network transmission times for web servers differ considerably depending on their location. From the view of users in Europe, servers located in the US correspond to a long delay while the access to servers in Europe works faster. Someone in the US browsing the web will experience similar transmission times, however, in the other direction.

The wired network transmission times used in the simulation result from quite simple *ping* measurements made at the University of Würzburg. A selection of web servers located in

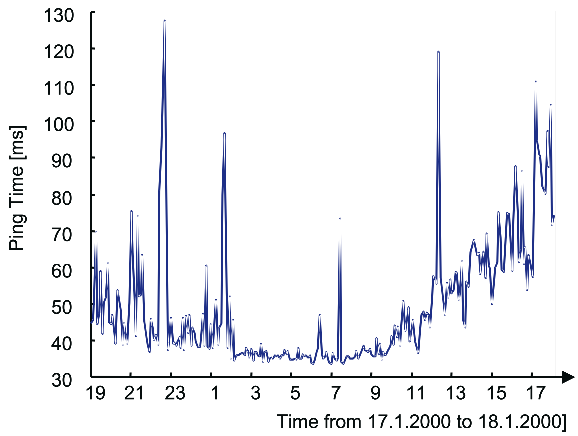

Figure 6: Ping times to www.libertel.nl

Germany, Europe, the US, and Japan were pinged for one week. Figure 6 shows the results of the measurements for the web server of Libertel in the Netherlands (www.libertel.nl) for one day. The ping times roughly reside in an interval between 35ms and 130ms. Similar times for other servers in Europe and Germany could be observed while the time to servers in the US proved to be mostly between 100ms and 350ms.

server location category	probability	time interval
near	0.3	10ms - 50ms
middle	0.4	50ms - 150ms
far	0.3	150ms - 350ms

Table 2: Wired network transmission times

The results are used to determine the wired network transmission times by two random variables. The first one describes the server location with an associated time interval and the second one the exact time within this interval. The server locations are divided into three categories near, middle, and far and intervals of 10ms-50ms, 50ms-150ms, and 150ms-350ms are associated. 30% of the web servers are near, 40% are in the category middle, and the remaining 30% are far. The times within each of these intervals follow a uniform distribution. Table 2 summarizes these values.

2.3 GPRS Network Model

The previous sections contained the communication between the mobile station and an Email or WWW server. Furthermore, the wired part of the network, i.e. from the BTS to the selected server, is simulated by a simple transfer delay. The remaining part is the transmission of IP packets over the wireless link. Actually, GPRS does not only define the transmission over the radio interface, but includes the architecture and protocol stack for the wired network, as well. Figures 7 and 8 show the GPRS logical network architecture and the GPRS protocol stack as defined by the ETSI [31] with minor changes. The components simulated in detail are marked by the fields in dark grey.

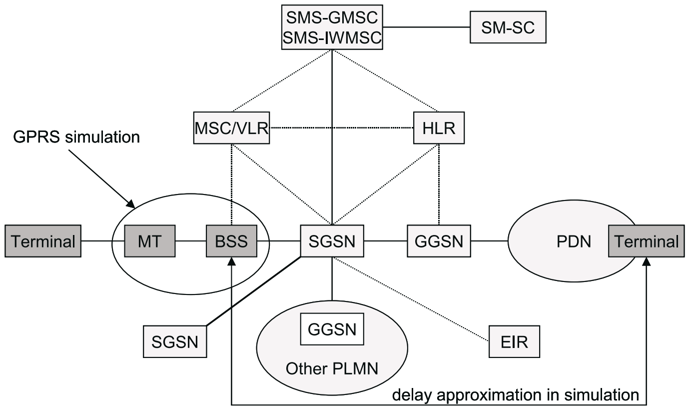

Figure 7: Overview of the GPRS logical architecture

Fig. 7 shows that the applications, as well as the TCP/IP stack, are located in the two terminals. These communicate by exchanging IP packets over the radio interface, see [32], and the wired part of the network, consisting of the Serving GPRS Support Node (SGSN), Gateway GPRS Support Node (GGSN), and Packet Data Network (PDN). The assumption of a constant wired network delay includes the transmission times at the SGSN, GGSN, and PDN. Therefore, only the GPRS protocols highlighted in Fig. 8, i.e. the *Radio Link Control* (RLC) protocol and the *Medium Access Control* (MAC) protocol both defined in [33], are simulated.

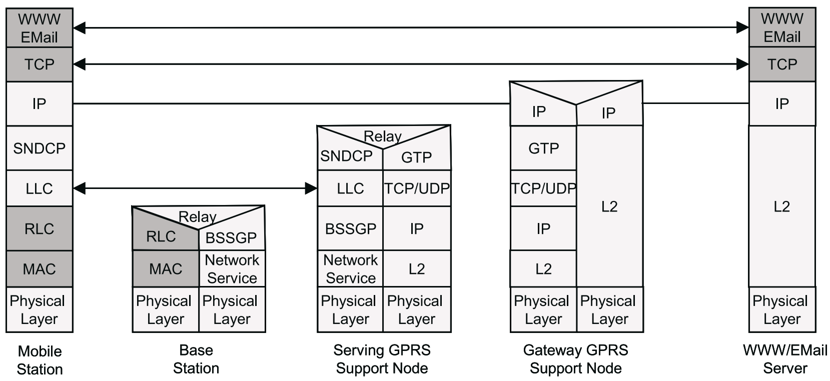

Figure 8: Protocol Stack

The transmission of an IP packet over the air interface is illustrated in Fig. 9. The RLC layer receives an LLC frame which comprises an IP data packet of 576 Bytes or an ACK of 40 Bytes and additionally the SNDCP header, LLC header, and the LLC Frame Check Sequence, altogether 15 Bytes, are added. This LLC frame is padded such that its size becomes a multiple of the RLC data block size which depends on the Coding Scheme, i.e. the used rate for the convolutional code. The Coding Scheme also defines the BLER (Block Error Rate) which is assumed to be constant in the simulation.

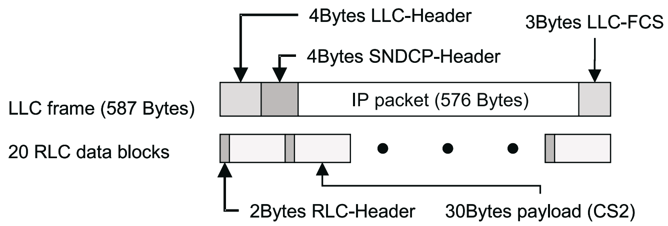

Figure 9: Segmentation of an IP packet into RLC data blocks

The RLC data blocks are transmitted in a *Temporary Block Flow* (TBF), which exists only if a data transfer is taking place. Each MS can have one TBF on the uplink and one on the downlink. The MAC protocol controls the assignment of resources, i.e. one or more *packet data channels* (PDCHs), to the TBFs. If one PDCH is assigned to several TBFs they share it equally. The resources of a TBF define the time required to transmit an RLC data block. The example in Fig. 10 shows a possible slot allocation of a radio bearer. Slots 0-2 are assigned to voice users and slots 3-7 are PDCHs belonging to different TBFs. The time to transmit an IP packet over the wireless link is explained for the example of TBF1. The allocations of TBF1 consist of the complete slot 7, slots 5 and 6 are each shared with another TBF. Altogether, TBF1 has allocations corresponding to 2 full PDCHs.

An RLC data block is transmitted in 4 bursts, i.e. with one PDCH the data block is transmitted in 4 frames. Therefore, with a frame length of $8 \cdot 577\mu s = 4.616ms$ we need $18.464ms$ for the data block. In the case of TBF1 with 2 PDCHs only half of the time is required. Considering Fig. 9, the time to transmit an IP packet can be derived. With a size of 576Bytes 20 RLC data blocks have to be transmitted yielding a total time of $184.64ms$. In this example block errors and resulting retransmissions are not taken into account. However, the simulation includes the required retransmissions by increasing the number of RLC data blocks for the IP packet. Whenever the resources of a TBF change the remaining transmission time has to be recalculated. This may occur if new TBFs are established, existing TBFs end, or in the case of preemption, i.e. one of the PDCHs is needed for a voice call.

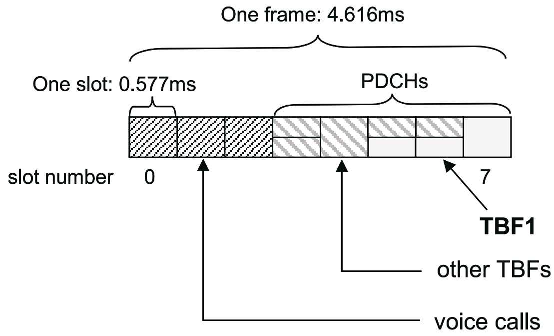

Figure 10: Example for slot allocations on a radio bearer

3 GPRS Network Performance

The intention behind the simulation study is to investigate the quality of the Internet access with GPRS in its first phase. At the beginning GPRS will have to cope with *on-demand* PDCHs only, i.e. the channels which are not used by voice users. Currently, radio cell dimensioning aims at keeping voice blocking below an upper bound of e.g. 1%. Consequently, a difference between the available and the mean used voice channels exists which corresponds to the PDCHs available for GPRS.

The following sections present simulation results for a radio cell with three frequencies and an offered voice traffic of 12.9 Erlangs, i.e. according to a blocking probability of 1%. Hence, an average of 8.23 PDCHs are available for GPRS.

Further simulation parameters are the Coding Scheme (CS) and the *multislot* class of the mobiles. With the chosen CS2, an RLC data block contains a payload of 30 octets which equals a data rate per PDCH of about 13 kbps and a block error rate of 7.5% is assumed. The multislot class defines how many neighbored PDCHs a mobile is able to handle in one direction. The simulated 4-slot mobiles allow a maximum transmission rate over the radio interface of 52 kbps. The simulation results are presented by the means of selected statistical values together with the 95%-confidence intervals.

The traffic load is scaled by the SessionArrivalRate, which is the inverse of the SessionInterArrivalTime. The average data volume transmitted over the wireless link according to the SessionArrivalRate is illustrated in Fig. 11 for both the up- and the downlink. The linear shape of the curves proves that the SessionArrivalRate is proportional to the average data volume except for very high loads which exceed the maximum possible bandwidth of 107kbps corresponding to the 8.23 PDCHs and CS2. In the linear part the transported traffic corresponds to an average of 313kBytes per session on the downlink and an average

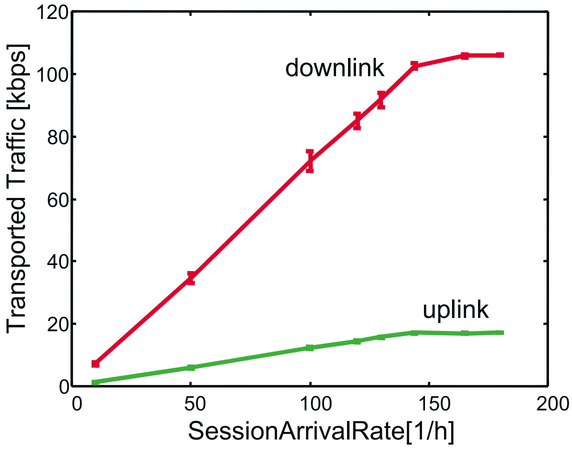

Figure 11: Transported traffic

of 54kBytes per session on the uplink. These values coincide with the parameters of the source traffic model.

Fig. 12 displays the mean number of users which are simultaneously active in the radio cell. The term “active users” has two different meanings here; users with an active web or Email session and users actually transmitting data over the wireless link. Contrary to the data rate these curves have no linear shape. The users simultaneously active in one cell share the available bandwidth and so, the data transmission takes the longer the more users are active which explains the nonlinear increase.

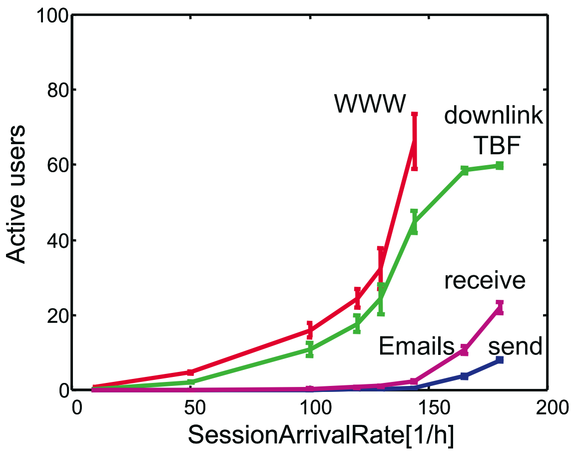

Figure 12: Number of active users

An upper bound for the number of downlink TBFs is 64 if each of the eight PDCHs carries eight TBFs and each TBF has resources on one PDCH only. The average number of sent and received Emails is clearly below the other two curves due to their smaller data volume. However, the shape of the curves is similar to the web session curve.

The ratio of the number of web and Email (receiving) sessions to the number of downlink TBFs is interesting to observe. It expresses which percentage of a session a user is actually transmitting data over the downlink. For only ten sessions per hour the activity on the downlink is less than 35 percent of a session. With increasing load the activity rate reaches over 60 percent for 130 sessions per hour. With a low load, transmissions over the wireless links are faster and hence the time on the wireless link decreases relative to the wired network delay which is independent of the load in the simulation.

The data rate over the wireless link is the value which will be measured in the future GPRS network and due to the linear shape of the curve it is possible to map it to a corresponding SessionArrivalRate. The QoS values corresponding to this parameter are defined and illustrated in the next section.

3.1 QoS Values for Internet Access

While for voice traffic the call blocking probability defines the QoS, for the Internet access it is more complicated and depends on the application. Concerning WWW, the user estimates its QoS by the time required to download a web page. However, this value is rather subjective since it depends on the size of the page. A more objective QoS value is provided by the TCP bandwidth, which can be thought of as the speed for downloading web pages or transmitting Emails and is less dependent on the page volume. It is defined as the ratio of the data volume transmitted in the TCP connection to the duration of the connection from its setup to the reception of the last data packet.

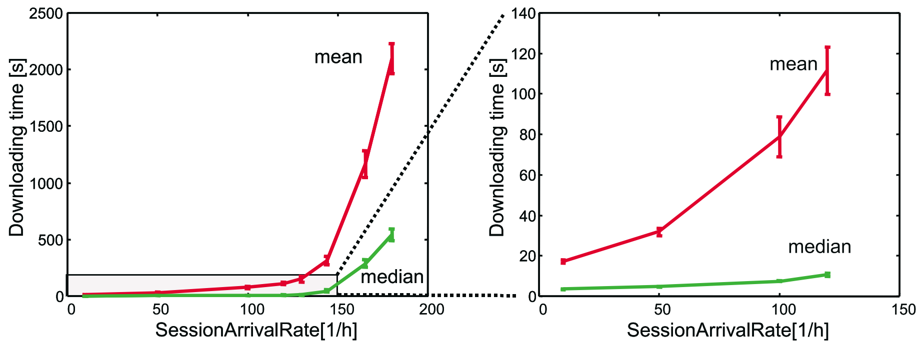

Figure 13: Web page downloading time

Fig. 13 shows the mean and the median of the time required to download a web page; the right graphic is a detailed view of the left graphic. Considering the scope of all SessionArrivalRates demonstrates the nonlinear behavior of the mean and the median downloading times. According to the web page model the mean page size with 50kBytes is much larger than the median page size with 9kBytes. Consequently, the mean downloading time is about five times longer than the median for low arrival rates. For higher loads the ratio of mean to median loading times increases since large web pages are affected stronger by the heavy traffic than smaller pages. For an arrival rate between 100 and 130 sessions per hour the mean outnumbers the median by ten. For even higher loads, small and large pages experience the same load again and as a consequence the ratio of the mean and median downloading times becomes about five.

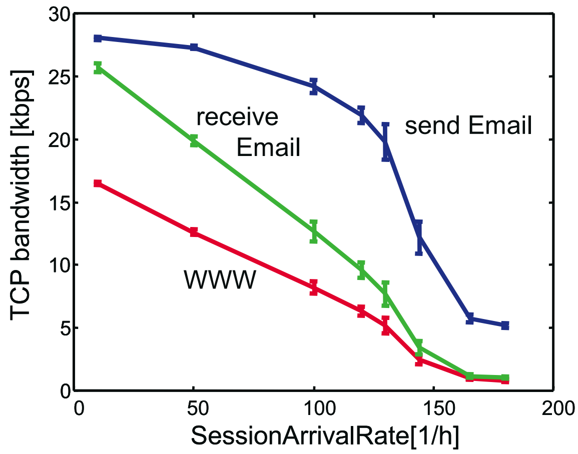

Figure 14: Average bandwidth of TCP connections

The web page downloading times depend on the underlying source traffic model with a strong variation in the sizes of web pages. A more objective impression of the performance of web browsing with GPRS is provided by the TCP bandwidth as depicted in Fig. 14 for HTTP traffic, Emails to the mobile, and Emails to the server. We can see that sent Emails sustain the greatest bandwidth and web traffic the least bandwidth. Emails from the server to the mobiles are in between, for low loads they are near the sent Emails and for high loads they are near the HTTP traffic. The behavior of TCP explains this feature. For a low load and short transmission times on the wireless link, the wired network delay dominates the TCP bandwidth in particular for connections with small data volume. Such connections emerge more frequently for web traffic with very small objects than for Emails. Furthermore, the bandwidths on uplink and downlink are almost equal such that Emails in both directions experience a similar TCP bandwidth. For higher loads, the transmission times on the wireless link dominate the TCP bandwidth. Therefore, the TCP connections with more data on the downlink, i.e. HTTP traffic or received Emails, experience a lower and similar bandwidth.

Both the page downloading time and the TCP bandwidth depend not only on the quality of the wireless access but also on the wired network transmission time. The performance of the radio part of the GPRS network is seen better when observing the bandwidth of a TBF as depicted in Fig. 15. It shows the TBF bandwidth for up- and downlink TBFs according to two different ways of calculating the mean bandwidth referred to as “while” and “per” TBF. For the latter one the bandwidth of each TBF is determined and from them the mean is calculated. In the case of “while” TBF, the mean TBF bandwidth for a simulation state is computed and then the mean bandwidth is found by weighting this TBF bandwidth with the probability of the corresponding simulation state. The difference is that in the case of “per”, all TBFs have the same weight whereas for “while”, longer TBFs have a greater influence than shorter ones. Since longer TBFs experience a lower bandwidth the curves marked as “while” are below the ones marked as “per”. The bandwidth “per” uplink TBF is almost constant and slightly below the maximum value of 52kbps. The bandwidth on the downlink, however, decreases strongly to below 2kbps for very high loads.

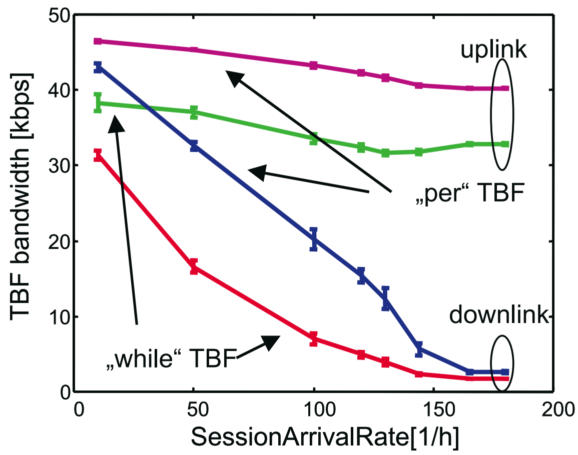

Figure 15: Average bandwidth of up- and downlink TBFs

The observed bandwidth “while” downlink TBF gives an impression of the quality of the GPRS access compared to a modem with a fixed access speed. In the next section the quality of GPRS in a cell with a certain load is compared to the modem access.

3.2 Comparison to Wireline Internet Access

The QoS values defined and plotted up to now describe the performance of the Internet access a GPRS user can expect in the first phase. However, the results strongly depend on the stochastic source traffic model and the approximation of the wired parts of the network. Furthermore, it is difficult to classify the quality of a resulting mean or median downloading time.

In order to evaluate the QoS and to give an impression what to expect from GPRS a comparison of the GPRS Internet access with the Internet access using a modem with fixed access speed is provided. The expectation from GPRS tells us that comparisons to modem speeds between 10kbps and 64kbps, corresponding to ISDN, are of interest.

In order to obtain QoS values of the wireline access, one user is simulated using the same source traffic model and the same wired network transmission times as for GPRS. Actually, the only change to the GPRS simulation is that the wireline access with a constant rate for one user replaces the wireless GPRS access with a varying and shared bandwidth. So the emphasis lies on the actual quality provided by GPRS.

At the first look the wireline access speed can be compared with the bandwidth "while" TBF on the downlink in GPRS. Considering Fig. 15 one can expect the performance of GPRS to be similar to a wireline access with speeds between 10kbps and 30kbps, dependent on the SessionArrivalRate in the cell.

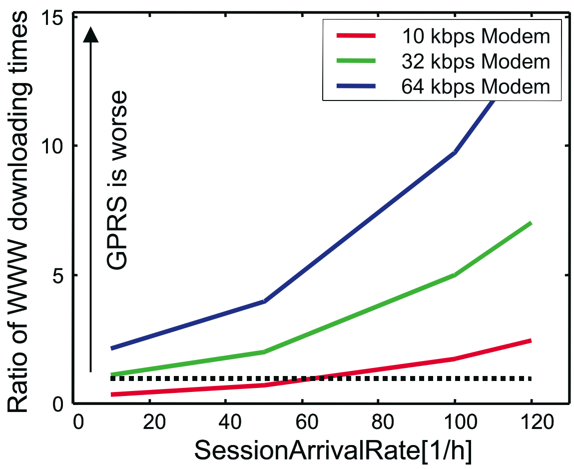

Figure 16: Comparison of the average web page downloading times for wireline and GPRS Internet access

For a more sophisticated comparison the downloading time for a web page and the TCP bandwidth are considered. Fig. 17 and Fig.16 illustrate the results for wireline access speeds of 10kbps, 32kbps, and 64kbps. The plotted values are the ratio of GPRS simulation results dependent on the SessionArrivalRate to the corresponding outcome of the wireline simulation.

Fig.16 shows the ratio of the mean downloading times for web pages. The dotted line marks where the performance of GPRS equals the performance of the wireline access. Above the line, the wireline access is better and below the line GPRS is better. The mean downloading time with GPRS will be clearly shorter than that with a 10kbps modem and approach the time of a 32kbps modem as long as the load in the cell is not too high.

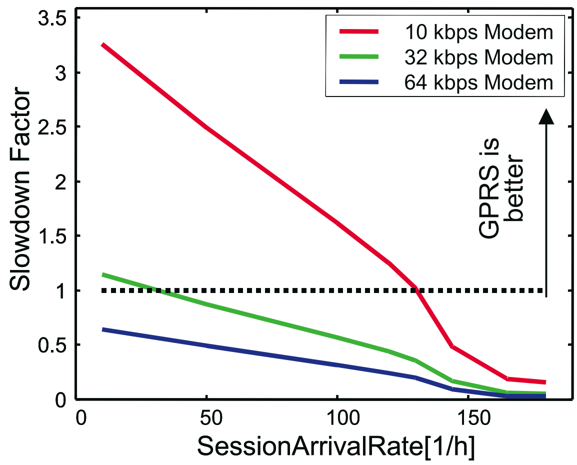

Figure 17: Comparison of the average TCP bandwidth for wireline and GPRS Internet access

Fig.17 shows the ratio of the TCP bandwidths. Again the dotted line divides the figure in two halves, but now above the line GPRS is better and below, the wireline access yields a higher TCP bandwidth. The comparison of the average TCP bandwidth shows a somewhat different result than the comparison of the mean downloading times. Concerning the TCP bandwidth GPRS performs in the low load case better than a 32 kbps modem. The reason for this is that, in the case of downloading times, very long times for large web pages have a greater influence than TCP connections with a low bandwidth. Since the TCP bandwidth is less dependent on the data volume than the downloading times are, it is better suited for the evaluation of the performance of GPRS. Therefore, the *Slowdown Factor* of the GPRS Internet access to a wireline Internet access with fixed bandwidth is defined as the *Average GPRS TCP bandwidth* divided by the *Average Modem TCP bandwidth*.

A Slowdown Factor less than one means that GPRS behaves worse than the considered modem and above one indicates that it performs better. The main result of our simulation study is, therefore, Fig. 17 which illustrates that the wireless Internet access using GPRS corresponds to the access with a modem of about 32kbps, as long as the SessionArrival-Rate stays below 30 sessions per hour.

4 Conclusion

This paper presented a simulative study of the performance of the wireless Internet access in a GPRS network. Based on a substantial number of publications and additionally on measurements made at the University of Würzburg a sophisticated traffic model describing the stochastic behavior of a GPRS user was derived. Using this model as traffic generator a GPRS cell with voice and Internet traffic was simulated. The transmission of the Internet traffic was considered very detailed including the simulation of TCP and HTTP as well as the implementation of the radio interface of GPRS. Certain QoS parameters for the Internet access were presented dependent on the traffic load in a GPRS cell. These included e.g. the downloading time of a web page or the achieved bandwidth of a TCP connection.

Furthermore, the performance of GPRS was compared to the performance experienced by an Internet user dialing in with a modem of fixed speed. The quality of GPRS related to modem quality was demonstrated by the ratio of web page downloading times and of TCP bandwidths. Due to the high variation in web page volumes the TCP bandwidth ratio is better suited for the comparison. Therefore, the indicator of how good GPRS performs under certain load conditions was defined as the Slowdown Factor which is the ratio of the GPRS TCP bandwidth to the wireline TCP bandwidth. From the simulation emerged that for medium load conditions GPRS in its first phase will perform similar to a modem with a speed of about 32kbps. This does not only hold for the considered case of a cell with three frequencies, but also for cells with different numbers of frequencies.

Acknowledgements

The authors would like to thank the Strategic Radio Network Development group of Libertel-Vodafone, Netherlands, for initiating and funding of this research project.

References

[1] G. Brasche and B. Walke, “Concepts, services, and protocols of the new GSM phase 2+ general packet radio service,” *IEEE Communications Magazine*, vol. 35, pp. 94–104, August 1997.

[2] J. Cai and D. Goodman, “General packet radio service in GSM,” *IEEE Communications Magazine*, vol. 35, pp. 122–131, October 1997.

[3] R. Kalden, I. Meirick, and M. Meyer, “Wireless Internet Access Based on GPRS,” *IEEE Presonal Communications*, vol. 7, pp. 8–18, April 2000.

[4] S. Malomsoky, S. Nádas, G. Tóth, and P. Zarándy, "Simulation based GPRS Network Dimensioning," in *Proc. of ITC Specialist Seminar on Mobile systems and Mobility*, (Lillehammer, Norway), pp. 237–248, March 2000.

[5] W. Ajib and P. Godlewski, "Service disciplines performance for best-effort policies in packet-switching wireless cellular networks," in *Proc. of IEEE VTC Spring 2000*, (Tokyo, Japan), May 2000.

[6] P. Stuckmann and F. Müller, "GPRS radio network capacity and quality of service using fixed and on-demand channel allocation techniques," in *Proc. of the Vehicular Technology Conference (VTC spring 2000)*, (Tokyo, Japan), May 2000.

[7] R. Litjens and R. Boucherie, "Radio Resource Sharing in a GSM/GPRS Network," in *Proc. of ITC Specialist Seminar on Mobile systems and Mobility*, (Lillehammer, Norway), pp. 261–274, March 2000.

[8] M. Meyer, "TCP performance over GPRS," *IEEE Wireless Communications and Networks Conferance 1999 (WCNC'99)*, vol. 3, pp. 1248–1252, 1999.

[9] Z. Kostic, "The performance of the www traffic due to interactions between tcp and rlp protocols in a cellular system," in *Proc. of IEEE VTC Fall 2001*, (Atlantic City, NJ), October 2001.

[10] M. Zhang and A.MacDonald, "Performance simulation for shared voice/data edge channel," in *Proc. of IEEE VTC Fall 2001*, (Atlantic City, NJ), October 2001.

[11] C. Johansson, H. Nyberg, and P. de Bruin, "Streaming services in gsm/edge - radio resource management concepts and system performance," in *Proc. of IEEE VTC Fall 2001*, (Atlantic City, NJ), October 2001.

[12] X. Qiu, K. Chawla, L. Chang, J. Chuang, N. Sollenberger, and J. Whitehead, "Rlc/mac design alternatives for supporting integrated services over egprs," *IEEE Personal Communications*, April 2000.

[13] A. Reyes-Lecuona, E. González-Parada, E. Casilari, and A. Díaz-Estrella, "A page-oriented WWW traffic model for wireless system simulations," in *Proc. of ITC'16*, (Edinburgh, UK), pp. 1271–1280, June 1999.

[14] N. Vicari and S. Köhler, "Measuring Internet User Traffic Behavior Dependent on Access Speed," in *Proc. of the 13th ITC Specialist Seminar on Internet Traffic Measurement and Modelling*, (Monterey, CA), September 2000.

[15] P. Barford and M. Crovella, “Generating representative workloads for network and server performance evaluation,” in *Proc. of ACM SIGMETRICS 1998*, (Madison, WI), pp. 151–160, June 1998.

[16] C. Cunha, A. Bestavros, and M. Crovella, “Characteristics of WWW Client-based trace,” Technical Report TR-95-010, Boston University, Dept. of Computer Science, April 1995.

[17] B. Mah, “An Empirical Model of HTTP Network Traffic,” in *Proc. of the IEEE INFOCOM 97*, (Kobe, Japan), pp. 592–600, April 1997.

[18] N. Vicari, “Measurement and Modeling of WWW-Sessions,” Technical Report 184, Institute of Computer Science, University of Würzburg, 1997.

[19] ETSI, “Universal Mobile Telecommunication System (UMTS); selection procedures for the choice of radio transmission technologies of the UMTS.” Technical Report TR 101 112 v3.2.0, 1998.

[20] R. C. S. G. ITU: US TG 8/1, “The radio cdma2000 RTT candidate submission.” TR45-5, 1998.

[21] H. Choi and J. Limb, “A behavioral model of web traffic,” in *Proc. of International Conference of Networking Protocol 99’ (ICNP 99’)*, (Toronto), September 1999.

[22] J. Heidemann, K. Obraczka, and J. Touch, “Modeling the Performance of HTTP Over Several Transport Protocols,” *ACM/IEEE Transactions on Networking*, vol. 5, pp. 616–630, October 1997.

[23] T. Berners-Lee, R. Fielding, and H. Frystyk, “Hypertext transfer protocol - HTTP/1.0.” IETF RFC 1945, 1996.

[24] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, and T. Berners-Lee, “Hypertext transfer protocol - HTTP/1.1.” IETF RFC 2068, 1997.

[25] H. Nielsen, J. Gettys, A. Baird-Smith, E. Prud’hommeaux, H. Lie, and C. Lilley, “Network Performance Effects of HTTP/1.1 , CSS1, and PNG,” in *Proc. of the ACM SIGCOMM ’97*, (Cannes, France), pp. 155–166, September 1997.

[26] V. Bolotin, Y. Levy, and D. Liu, “Characterizing data connection and messages by mixtures of distributions on logarithmic scale,” in *Proc. of the ITC’16*, (Edinburgh, UK), pp. 887–894, June 1999.

[27] P. Tran-Gia, D. Staehle, and K. Leibnitz, “Source Traffic Modeling of Wireless Applications,” *AEÜ, Int. J. Electron. and Commun.*, vol. 55, pp. 27–36, January 2001.

[28] S. Floyd and T. Henderson, “The NewReno Modification to TCP’s Fast Recovery Algorithm.” IETF RFC 2582, 1999.

[29] W. Stevens, *TCP/IP Illustrated*, vol. 1. Addison-Wesley, 1994.

[30] G. Wright and W. Stevens, *TCP/IP Illustrated*, vol. 2. Addison-Wesley, 1995.

[31] ETSI, “GSM 3.60, GPRS service description, Stage 2.”

[32] ETSI, “GSM 3.64, GPRS, Overall Description of the radio interface.”

[33] ETSI, “GSM 4.60, GPRS, MS-BSS interface, Radio Link Control/Medium Access Control (RLC/MAC) Protocol.”