A WAP TRAFFIC MODEL AND ITS APPLIANCE FOR THE PERFORMANCE ANALYSIS OF WAP OVER GPRS

Peter Stuckmann[1], Henrik Finck[1], Thomas Bahls[2]

[1] RWTH Aachen University of Technology, Germany
E-Mail: pst@comnets.rwth-aachen.de
WWW: http://www.comnets.rwth-aachen.de/˜pst
[2] Siemens AG, Berlin, Germany
E-Mail: thomas.bahls@icn.siemens.de

ABSTRACT

In this paper the Wireless Application Protocol (WAP) as one mobile data service application is examined with the aim to dimension next generation mobile radio networks. A new traffic model for WAP-based applications comprising the application characteristics and the user behaviour is introduced. Taking this model as the basis a WAP load generator and its integration into the General Packet Radio Service (GPRS) simulator GPRSim is presented. Simulation results of quality of service measures and GPRS system measures depict the performance of WAP over GPRS and the GPRS network capacity needed for WAP-based applications in comparison to mobile Internet applications based on HTTP over TCP.

I. INTRODUCTION

In the context of the exponential growth of the Internet market the growing demand of data services is also expected for mobile users. Besides conventional Internet applications like WWW and E-Mail, specific service platforms and applications are developed for future mobile data services like the *General Packet Radio Service* (GPRS) and the *Universal Mobile Telecommunication System* (UMTS). Reasons for this are different conditions namely limited bandwidth, higher delays and error rates on the radio interface and mobile terminals with limited user interfaces and lower processing power and memory capacity compared to PC platforms.

The *European Telecommunications Standardization Institute* (ETSI) as well as the *Wireless Application Protocol* (WAP) Forum address these specific requirements and characteristics by some of their work. While ETSI developed three so-called classmarks of the Mobile Station Application Execution Environment (MExE), the WAP Forum compiled a set of specifications targeting at mobile devices with limited capabilities. The current WAP specifications, including the June 2000 Conformance Release, also known as WAP 1.2.1, aim at optimizing operation in 2G networks. Therefore WAP 1.2.1 defines a distinct technology comprising protocols and content representation. Because of such an optimized design the generated traffic patterns cannot be mapped onto today's Internet models. Extra work is required to model and simulate WAP traffic. In the next release, some existing WAP protocols will be extended by new capabilities. One example is the ability to transfer large files over WAP. Future WAP Forum releases are expected to converge with widely used Internet protocols like Transmission Control Protocol (TCP) and Hypertext Transfer Protocol (HTTP). However, they still aim at wireless optimizations by profiling those protocols. In addition to protocol work, the WAP Forum will continue its work on service-enabling features for the mobile environment, like the push service or synchronization issues.

In this paper WAP traffic characteristics of typical applications using this platform are outlined and are mapped together with identified parameter sets on a complete traffic model taking WWW traffic models as a paradigm [6]. In the next step measurements of arrival times of requests within sessions and packet sizes of typical WAP applications are presented. With the results of these measurements the WAP traffic model is parameterized. The measured parameters are typical for WAP applications coming up in the short term. This traffic model is the basis for a load generator, which was developed within this framework, that is presented together with a full GPRS simulation environment GPRSim. Taking this simulation environment as the basis results are produced giving performance measures of several WAP-based application types compared with conventional Internet applications with larger content based on HTTP over TCP. The aim of this paper is not to compare the performance characteristics of the WAP protocol architecture with the conventional Internet protocols, but the performance analysis for network dimensioning purposes. The results show the multiplexing gain possible with GPRS compared to circuit switched GSM data services. They are directly usable for capacity planning and network dimensioning.

II. WIRELESS APPLICATION PROTOCOL (WAP)

WAP is a suite of specifications that defines an architecture framework containing optimized protocols (e.g., WDP, WTP, WSP), a compact XML-based content representation

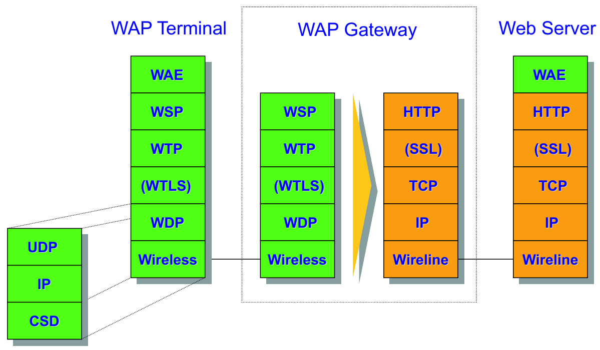


Figure 1: The WAP protocol architecture

(WML, WBXML) and other mobile-specific features like Wireless Telephony Applications (WTA) [1, 2]. In addition to the goal of optimized operation in 2G networks, WAP has been developed because today's graphics-enhanced web services cannot be brought to and displayed on thin clients, e.g., GSM mobile phones, and IP as network layer may not be applicable in some environments, e.g., WAP over Short Message Service (SMS) or Unstructured Supplementary Service Data (USSD). Because of the optimizations and different protocols it is not possible to run WAP end-to-end to a regular Internet site. Instead, a WAP Gateway must be used. The main services a WAP Gateway provides is protocol conversion between WAP stack and Internet stack. In addition to this standardized functionality, many gateway vendors provide a variety of value-added services that allow for personalization, for example.

The WAP protocol architecture (see Figure 1), which this paper focuses on, comprises the following components:

Wireless Application Environment (WAE) With WAE an environment for development and execution of services is provided. The core component of WAE is the micro browser that renders the respective WML content.

Session Layer (WSP) WSP contributes methods, which allow for communication between client and server. WSP provides mechanisms to retain a session context, if the call is dropped.

Transaction Layer (WTP) WTP is the transport layer of WAP. It allows for connectionless as well as connection-oriented communication. Today, WTP class 2 transactions are commonly used for most of the WAP traffic. WTP class 2 represents the acknowledged client-to-server communication.

Security Layer (WTLS) WTLS is an optional component, which works similar to SSL or TLS in the Internet. WTLS provides for authentication and privacy of an ongoing communication.

Transport Layer (WDP) WDP is an abstraction layer between WAP and the underlying bearer. In case of IP as the network layer, UDP is used as WDP layer.

Compared to the well-known Web protocols like TCP and HTTP, WAP protocols differ in the following aspects:

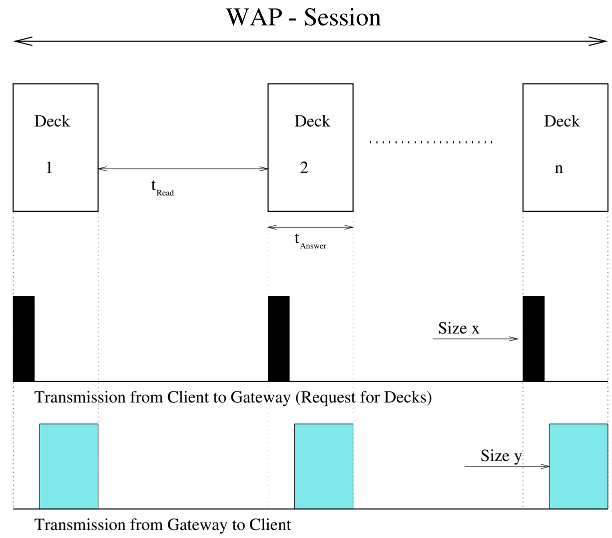


Figure 2: Sequence chart and important parameters for a WAP session

- WDP or UDP provide an unacknowledged datagram transport only. This approach avoids overhead and delay for services that operate on connectionless basis. One example is a WTP class 0 transaction.
- In-order delivery and retransmission are handled by a separate transport layer. It provides these functions only if they are requested by a higher-layer service (WTP class 1 or 2).
- WAP works transaction-oriented. The amount of data that can be transferred by one transaction defaults to 1400 bytes. It can be negotiated at WSP session setup time. If WTP segmentation and reassembly is used the amount of data per transaction is limited by the negotiated SDU size.
- WSP implements all features of HTTP 1.1 and enhances these by a binary encoding aiming at greater efficieny over the air.

III. WAP Traffic Model

A. Session Characteristics

A WAP session consists of several requests for a deck performed by the user. The sequence chart of a session is depicted in Figure 2.

WAP sessions are totally described by

- requests for a number of decks, *n*
- packet size in up- and downlink, *size x* and *size y*
- a reading time the user takes before requesting the next deck, t_{Read}
- a response time of the network, $t_{Response}$

The latter one is not determined by the user. This parameter is totally dependent of the underlying network. It

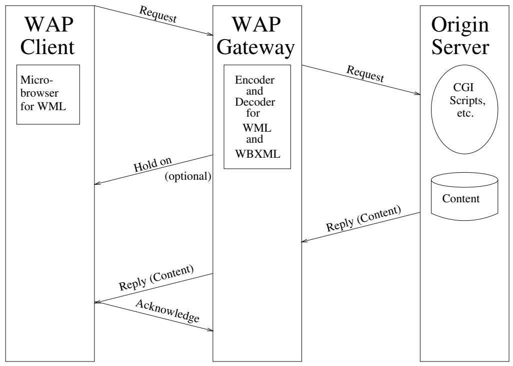


Figure 3: Sequence chart of a class 2 WTP transaction

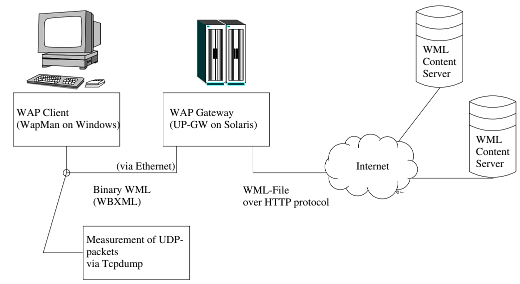
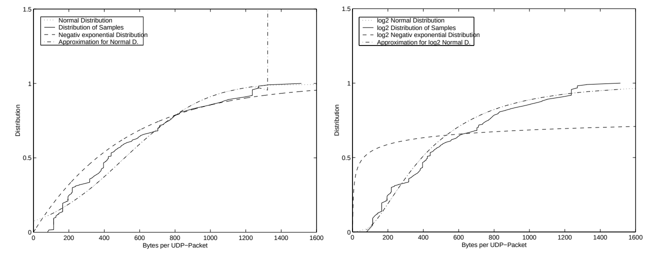


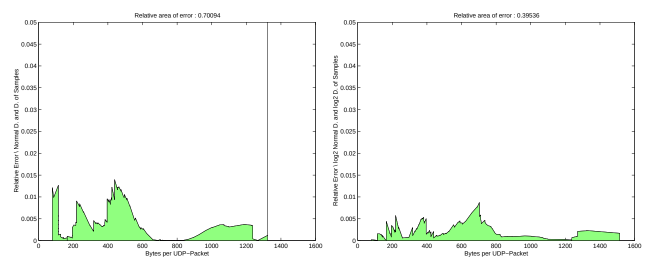
Figure 4: Setup for the traffic measurement



(a) Distributions of dowlink packet size (not transformed)

(b) Distributions of dowlink packet size (log2-transformed)

Figure 5: Downlink distributions



(a) Error function for normal distribution

(b) Error function for log2-normal distribution

Figure 6: Error functions for downlink distributions

characterizes the QoS seen by the user and is examined as an independent value in the evaluation of the simulations. The implemented method for retrieving a WML deck corrosponds to a WTP class 2 transaction (3).

Subsequent to [6], the number of decks is modelled by a *geometric distribution*, the reading time by a *negative exponential distribution* and the packet size by a *log2-normal distribution* (as later verified in Section III.C).

To cover the influence of different applications, different types of user profiles are introduced:

- E-Mail: different WAP services which give the possibility to read and write E-Mails
- News(paper): newspaper articles (often in shortened versions), news tickers, stock quotes, etc.
- M-Commerce: services where the user can, i.e., buy books, tickets for railway and flight, overall goods, reserve cinema or theatre tickets, etc.
- Common: mixed traffic traced from a WAP server in real operation

The parameters for the first 3 classes are derived by using just one application type during one measurement period. The amount of collected data comprises 127 samples for E-Mail, 266 for News and 100 for M-Commerce. The common traffic profile is introduced as the large database traced from a WAP Gateway in real service (5660 samples). These data traces could not be split into different application types. This in comparison to the other 3 services high amount of values of this case also represents a base for Section III.C. While the model of the session can be seen as general for WAP sessions the parameters are depending on the content. The parameters taken here are typical for WAP applications directly after introduction of WAP. Since the applications will change over the next years, the parameters of the model are subject to future updates.

B. Traffic Measurements

In Figure 4 the setup for the measurements is shown. The client application is running on a Windows-based PC connected over Ethernet to the WAP gateway. In order to trace the packets generated by the real application, the tool *Tcpdump* [3] was used, which is commonly available for the desired platform SOLARIS. TCPdump is able to trace packets on TCP or UDP level with their characteristics (size, content, time) on the communication path in TCP/IP-based networks. All parameters for the WAP traffic model that are depending on the underlying network can be gained with these traces.

C. Verification of the Packet Size Distribution

Since WAP is a new service platform, the distribution for the packet size may vary from the ones stated in [6].

In the following the measured distribution of the downlink packet size, which is the important parameter for performance analysis of asymmetric traffic, is compared with a normal, log-normal as well as an exponential distribu-

Table 1: Gained parameters for the WAP traffic model

Value	mean	stand. dev.
News		
Reading Time [s]	11.4	16.5
Size of ‘Get Request’- Packet [byte]	82.1	26.2
Size of ‘Content’ - Packet [byte]	638.4	389.9
E-Mail		
Reading Time [s]	10.0	22.2
Size of ‘Get Request’- Packet [byte]	112.0	57.7
Size of ‘Content’ - Packet [byte]	582.9	260.0
M-Commerce		
Reading Time [s]	10.7	13.1
Size of ‘Get Request’- Packet [byte]	84.3	16.5
Size of ‘Content’ - Packet [byte]	641.0	342.9
Common		
Reading Time [s]	13.8	116.8
Size of ‘Get Request’- Packet [byte]	108.2	84.7
Size of ‘Content’ - Packet [byte]	511.0	368.0

All packet sizes are defined by the amount of data the WTP layer generates

tion. Figure 5(a) shows the distribution of the measured samples in downlink for the packet size compared with a normal and an exponential distribution parameterized with the mean and standard deviation taken from the measurements. Figure 5(b) shows the respective log2-distributions compared with the measured samples. Figure 6 points out that in downlink direction the log2-normal distribution of the packet size always fits the distribution of samples best throughout the three distributions mentioned above. Since the packet size is limited in the WSP layer, e.g. to 1400 byte the distributions are not heavy-tailed contrary to models for the WWW [6]. The same was found for the uplink packet size.

Being aware of the presented error functions the log2-normal distribution is chosen as the distribution for the packet size of the WAP traffic model.

D. Parameters for WAP

Table 1 shows the parameters resulting from the measurements. Comparable are the first 3, because they correspond to one application type each. The parameters of the application type ‘Common’ correspond to the data gained from 2 weeks of varying usage of different users. Especially remarkable for the first 3 services are

- The ‘Get Request’ size of E-mail is comparatively large, since if the user sent a larger E-Mail this would be done in uplink direction
- The ‘Content’ size of News deviate most, because the user has reached the destination of content through several links to the desired information, the size of decks often reaches the maximum transferrable size in WAP
- The reading time of the user reaches the maximum at News, because the desired information that is displayed corresponds to the large deck sizes
- E-Mail has the largest standard deviation in reading time because the sending of an E-Mail includes to key it in on a device not quite suitable for text input

The parameter for the maximum deck size is chosen to 1397 byte, which is the limiting value of the first WAP phone on the market.

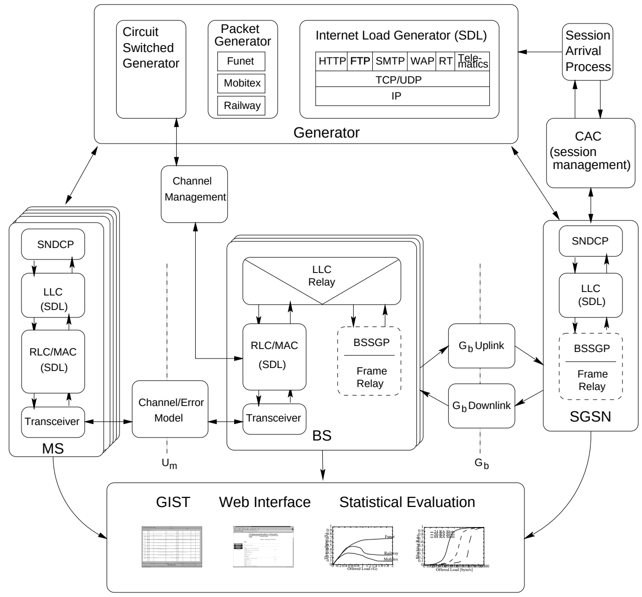


Figure 7: The GPRS Simulator GPRSim

IV. Simulation

For this study a simulation tool named GPRSim was used. This simulation environment allows to ascertain and to optimize properties of different protocols of the GPRS transmission plane. In addition, it gives the possibility of network capacity and quality of service planning by performance evaluation in certain simulation scenarios. The GPRS Simulator GPRSim is developed as a pure software solution in the programming language C++. Up to now in the GPRS architecture models of Mobile Station (MS), Base Station (BS), and Serving GPRS Support Node (SGSN) are implemented. The simulator offers interfaces to be upgraded by additional modules (see Figure 7). For implementation of the simulation model in C++ the Communication Networks Class Library (CNCL) [4], which was developed at the Chair of Communication Networks, is used. This allows an object oriented structure of programs and is especially applicable for event driven simulations. The complex protocols like LLC, RLC/MAC based on GPRS Release 99 and the Internet Load Generators including TCP/IP and UDP/IP are specified with the Specification and Description Language (SDL), are translated to C++ by the Code Generator SDL2CNCL [5] and are finally integrated into the simulator.

Different from usual approaches to building a simulator, where abstractions of functions and protocols are being implemented, the approach of the GPRSim is based on the detailed implementation of the standardized protocols. This enables to study the behaviour of GPRS in a realistic way. The WAP load generator developed for this study is also based on SDL and the tool SDL2CNCL. The WAP protocol architecture and a session model corresponding to the WAP traffic model introduced in Section III are implemented.

V. Simulation Results

A. Simulation Scenarios

The cell configuration is given by the number of Packet Data Channels (PDCHs) permanently available for GPRS. In this paper 1 and 4 fixed PDCHs have been regarded. For the simulations with Internet traffic, 3 and 4 fixed PDCHs are provided.

All simulations use a discrete random process with a block error probability (BLEP) that has previously been determined by measurements of the real channel in an urban environment. Thus, C/I values can be mapped on BLEP by means of reference functions. A C/I of 12 dB (13.5% BLEP) and an ideal channel, have been regarded.

LLC and RLC/MAC are operating in acknowledged mode. The multislot capability is 4 uplink and 4 downlink slots. The MAC protocol instances in the simulations are operating with 3 random access subchannels per 52-frame. All conventional MAC requests have the radio priority level 1 and are scheduled with a FIFO strategy. LLC has a window size of 16 frames. TCP/IP header compression in SNDCP is performed. The maximum IP datagram size is set to 256 byte for WAP. In the Internet stack for WWW and E-Mail TCP is operating with a maximum congestion window size of 8 Kbyte. The transmission delay in the core network and externel networks, i.e., the public Internet is ignored.

WAP sessions are parameterized with a traffic mix of 50% of the ‘Common’, 25% of the ‘E-Mail’, 15% of the ‘News(paper)’ and 10% of the ‘M-Commerce’ user profile. The conventional Internet traffic [6] regarded for comparison is composed of 70 % E-Mail sessions and 30 % WWW sessions (see Table 2). Here it becomes evident that different services with a different amount of data are compared and no protocol performance comparison is done.

The diagrams presented in the following section show inconsistencies for higher load situations owing to the lack of statistical reliability over limited simulated time. They are providing tendencies for performance estimation.

B. Performance Analysis

Starting point is the quality of service from the user’s point of view. For WAP applications this is the response time meaning the period when a request has been started until the information is displayed. For example, the average time for an acceptable deck response can be set to 4 seconds. Figure 8 shows the response time for WAP decks. The maximum number of concurrent users multiplexed onto the packet data channels are 20 for 1 PDCH (13.5% BLEP) and 25 for 4 PDCHs (13.5% BLEP) for this given QoS limit of 4 seconds average response time. Here it becomes clear that the performance gain achievable with the allocation of more than 1 PDCH for WAP over GPRS is low. The bad

Table 2: Model parameters of Internet applications (HTTP and SMTP)

HTTP Parameter	Distribution	Mean
Pages per session	geometric	5.0
Intervals between pages [s]	negative exponential	12.0
Objects per page	geometric	2.5
Object size [byte]	$\log_2$-Erlang-k	3700
Amount of SMTP data	**Distribution**	**Mean**
E-Mail Size [byte]	$\log_2$-normal	10000
base quota [byte]	constant	300

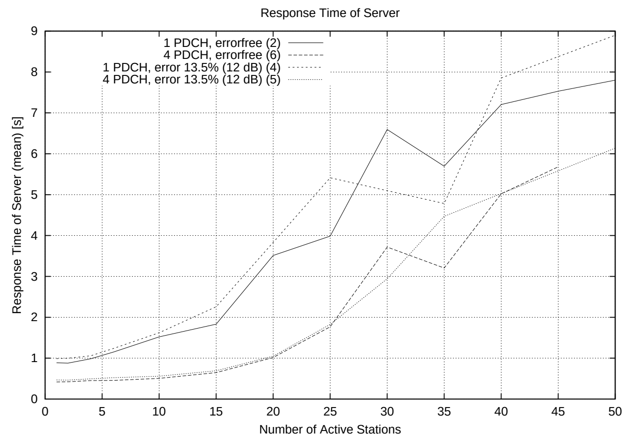


Figure 8: Response time for pure WAP traffic

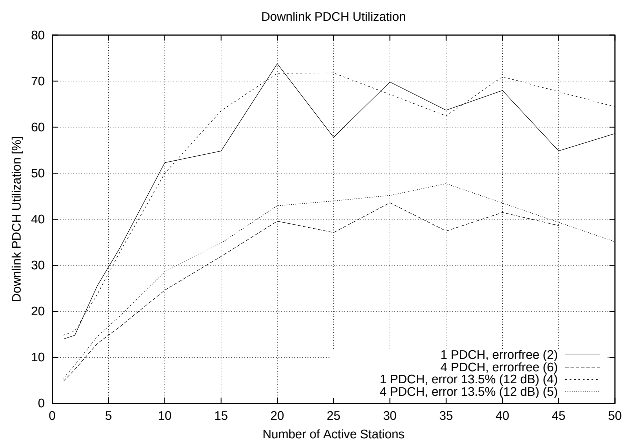


Figure 9: PDCH utilization for pure WAP traffic

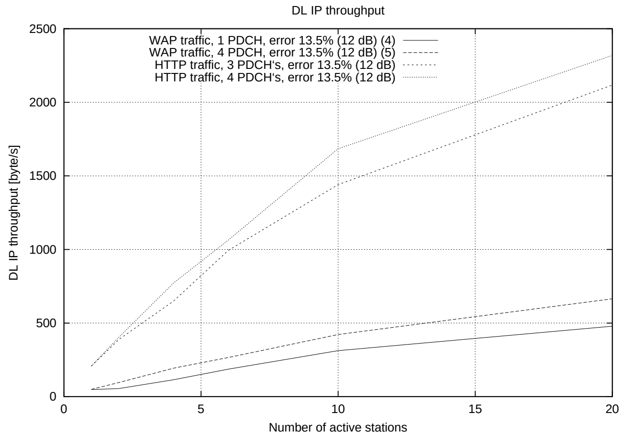


Figure 10: Downlink IP system throughput for pure WAP and pure WWW traffic

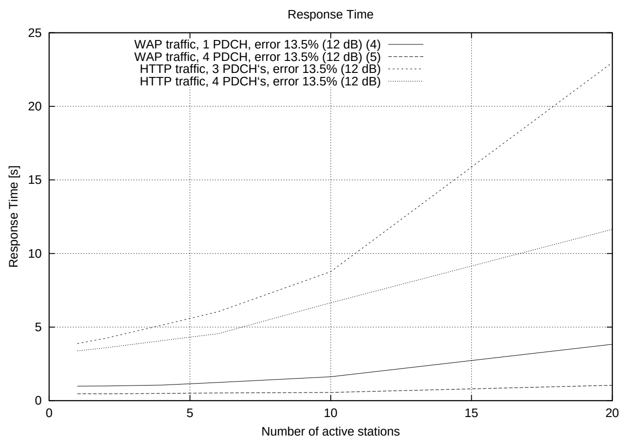


Figure 11: Response time for pure WAP and pure WWW traffic

performance gain of 4 PDCHs compared to 1 PDCH that is about factor 2 in response time can be explained by the constant signalling delay to set up a Temporary Block Flow (TBF) that decreases the performance especially for small offered packets typical for WAP. Therefore the PDCH utilization in the 4 PDCH case shown in Figure 9 is comparatively low. The PDCH utilization is the number of downlink radio blocks carrying data or control information divided by the total number of transmitted downlink radio blocks.

Figure 10 shows system throughput per radio cell for WAP users compared to conventional Internet users according to [6]. The comparison of the response time for WAP and for Internet (see Figure 11) shows that the number of users that can be served by a given number of PDCHs is much higher for WAP.

VI. Conclusions

In this paper a traffic model for WAP-based applications is introduced. The session characteristics and the process of parameterizing the model by traffic measurements are outlined. Applying this traffic model in the simulation environment GPRSim performance and capacity simulations are presented. They provide statements of the performance of WAP-based applications over GPRS and the multiplexing gain achieved by using GSM channels as Packet Data Channels for GPRS.

Here it becomes clear that WAP traffic sources are served in a very efficient way by GPRS since the multiplexing gain for a single PDCH is around 20 compared to circuit switched services (GSM 9.6 kbit/s), if the QoS limit of an average response time of 4 seconds is assumed. With more than one PDCH used for WAP over GPRS the multiplexing gain decreases to 6. This means that the allocation of more than one PDCH does not increase the performance significantly, since the average number of users per radio cell will not exceed 20 in the short term.

In addition the performance and system measures can be used for capacity planning and network dimensioning, since it becomes evident how many users can be served by a number of channels ensuring quality of service constraints.

Future examinations should concentrate on configuring the parameters of GPRS to optimize the behaviour for WAP-based applications. The traffic model and the load generator should be extended for future features of WAP, i.e., the push method or large data transfer. The parameters gained through measurements should be validated and updated when WAP is in use in a wider range and different services become available. In addition to examining the traffic seperately, a mixture of different traffic sources within one system should be regarded.

VII. References

[1] Wireless Application Protocol Forum. *Wireless Application Protocol - Wireless Application Environment Overview*. www.wapforum.org, 1999.

[2] Wireless Application Protocol Forum. *Wireless Application Protocol Architecture Specification*. www.wapforum.org, 1999.

[3] Van Jacobson, Steve McCanne, Craig Leres. *Tcpdump*. Network Research Group, Lawrence Berkeley Laboratory, 1996. ftp://ftp.ee.lbl.gov/tcpdump-*.tar.Z.

[4] M. Junius, S. Arefzadeh. *CNCL – ComNets Class Library*. Chair of Communication Networks, RWTH Aachen University of Technology, 1995.

[5] Martin Steppler, Wolfgang Olzem, Christoph Lampe. *SDL2CNCL 3.6: SDL-PR to C++ code generator using the ComNets Class Library*. Chair of Communication Networks, RWTH Aachen University of Technology, August 1997.

[6] P. Stuckmann, P. Seidenberg. *Quality of Service of Internet Applications over GPRS*. In *Proc. of the European Wireless'99, ISBN 3-8007-2490-1*, Munich, Germany, Oct. 1999.

[7] B. Walke. *Mobile Radio Networks - Networking and Protocols*. John Wiley & Sons, Chichester, 2nd edition, to be published in 2001.

[8] B. Walke, G. Brasche. *Concepts, services and protocols of the new GSM Phase 2+ General Packet Radio Service*. IEEE Communications Magazine, Vol. 35, 1997.