Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT®

Computer Communications xx (0000) 1–8

computer communications

www.elsevier.com/locate/comcom

Fluid-flow modelling of internet traffic in GSM/GPRS networks

Tim Irnich*, Peter Stuckmann

Chair of Communication Networks, Aachen University of Technology (RWTH), Kopernikusstr, 16, D-52074 Aachen, Germany

Received 5 September 2002; accepted 3 December 2002

Abstract

In order to find an analytical solution for performance evaluation of Internet access over the General Packet Radio Service (GPRS), the packet-switched extension of the GSM mobile radio network, we evaluate the well-known Fluid-flow modelling (FFM) approach. Mobile Internet users are modeled by an ON/OFF source with exponentially distributed sojourn times in the ON and OFF state.

We choose the IP datagram delay at the radio interface as quantity of interest. As Fluid-flow analysis only delivers the mean equilibrium buffer content, we present some additional methodology to calculate the sojourn time in the Fluid-flow model, taking specific characteristics of the GPRS into account. ON/OFF source parameters are determined representing the mean offered traffic in typical GPRS load scenarios by stochastic simulation, using the GPRS emulation system GPRSim. The GPRSim comprises load generators representing typical GPRS usage and a prototypical implementation of the GPRS protocols.

Comparison of simulation results and analysis is performed using the same traffic source model for both simulation and analysis, eliminating approximation of WWW traffic characteristics as a reason for deviations between simulation results and analysis. Additionally analytical results are compared with results of the GPRSim using a detailed WWW model for traffic generation. The results illustrate the influence of accurate source modelling and the limitations of the FFM to model the GPRS system with elastic traffic. It is shown that the FFM is not able to depict the elastic property of TCP-based Internet traffic, but is capable of modelling the multiplexing of inelastic traffic like generated by streaming or real-time applications over the GPRS radio link in sufficient detail for performance estimations.

Keywords: General packet radio service; Performance analysis; Fluid-flow model; ON/OFF source model; Radio network dimensioning

1. Introduction

In the context of the evolution towards 3rd Generation (3G) mobile radio networks, packet-switched data services like the General Packet Radio Service (GPRS) and the Enhanced GPRS (EGPRS) are presently introduced into GSM and IS-136 systems. For network operators, equipment vendors, and system integrators dimensioning procedures have to be developed to estimate the radio capacity that is needed to carry the predicted amount of additional user data [1,2]. For dimensioning of circuit-switched networks the Erlang theory has been successfully applied over decades, while for packet-switched cellular radio networks such an applicable traffic engineering model is still missing. Packet-based data transmission introduces additional complexity that has to be depicted in an accurate analytical approach. Where the Erlang theory focuses on user calls and determines blocking and loss probabilities, we now have to 'zoom into' the user call and regard the single packet. Thus, the Quality of Service (QoS) perceived by the user has to be defined on packet level rather than on connection level [3].

We investigate the applicability of Fluid-flow modelling (FFM) in this context. The FFM is based on a concept developed by Kosten [4], and was extended by Anick et al. [5]. For our analysis, we use the notation of Fiedler and Voos [6]. The main application of the FFM approach is the prediction of the equilibrium buffer size's *Cumulative Distribution Function (CDF)* under a given load scenario. Using Little's Law, we derive the mean waiting time of an IP datagram. We do not directly compare the FFM waiting time to the waiting time of data packets in our GPRS simulation system and real GPRS systems, but introduce the FFM sojourn Time as an analytical quantity which is equivalent to the IP datagram delay evaluated in our simulation system. It is derived by adding an offset for

* Corresponding author. Tel.: +49-241-8027928; fax: +49-241-8890385.
E-mail address: tim.irnich@comnets.rwth-aachen.de (T. Irnich).

0140-3664/03/$ - see front matter © 2003 Published by Elsevier Science B.V.
doi:10.1016/S0140-3664(03)00048-3

the mean transmission duration to the mean waiting time obtained by FFM analysis. This offset is based on evaluation of the mean IP datagram size, the mean number of Radio Link Control (RLC) blocks per IP datagram and the minimum time that is required for transmission of these RLC blocks.

The rest of this paper is organized as follows: In Section 2 we summarize the relevant characteristics of the GPRS, before we introduce the applied analytical methodology in Section 3. Section 3.1 introduces the analytical traffic source model and in Section 3.2 the key equations of the applied analysis are stated, additionally we present our methodology to derive a quantity that is comparable to the simulation's IP datagram delay from the results of the classic Fluid-flow analysis. In Section 4 our GPRS simulation environment GPRSim is described, which is used as a reference for comparison with our analytical results presented and discussed in Section 5. Section 7 concludes this paper.

2. General packet radio service

GPRS has been standardized by the ETSI as part of the GSM *Phase 2+* development to introduce a packet-switched extension to the GSM radio interface, which is essentially a circuit-switched technology. For a detailed description of GPRS please refer to Ref. [2]; we will limit our description to aspects that are particularly relevant in the context of this article.

Packet switching means that radio resources are used only when users are actually sending or receiving data. Rather than exclusively dedicating a radio channel to a mobile data user for a fixed—and, compared to the usual duration of a circuit-switched connection, short—period of time, the available radio resources can be concurrently shared between several users. Through multiplexing of several logical connections on one or more GSM physical channels, GPRS reaches a flexible use of channel capacity. The GPRS provides packet-switching logical channels (Packet Data Channel, PDCH). The basic transmission unit of a PDCH is a radio block that requires four time slots in four consecutive GSM Time Division Multiple Access (TDMA) frames. The length of a TDMA frame is 4.615 ms, and the length of a GPRS Multiframe is 18.46 ms. Every 13th burst is not used for transmission, see Fig. 1. Four different Coding Schemes (CS) are defined, providing data rates from 9.05 to 21.4 kbit/s per PDCH, see Table 1. Since in GPRS the access of all eight slots of a TDMA frame is foreseen, data rates up to 160 kbit/s can be achieved. For a single mobile station its Multislot Capability (MSC) defines how many slots within the TDMA frame may be used.

3. Fluid-flow model

In the scope of the FFM, arrival of data generated by a number of traffic sources at the point of interest is compared to water falling into a reservoir (the network element's buffer memory), which depletes at a constant rate C. The unit of data is assumed to be infinitely small. Traffic sources asynchronously alternate between an ON state and an OFF state. The sojourn times in the ON and OFF states are exponentially distributed. While in the ON state, a source transmits data at a constant rate h.

3.1. Source modelling

A single traffic source is modeled by a two state Markov Modulated Rate Process (MMRP), called Interrupted Rate Process (IRP). Multiple equal subscribers are depicted by superposition of multiple IRPs, called NIRP.

Each IRP is controlled by a two-state Markov chain (MC) with states Λ_0 and Λ_1. In state Λ_0 the source transmits packets at the rate $r_0 = 0$ and in state Λ_1 the source transmits packets at the rate $r_1 = h$. The transition rates between the ON state Λ_1 and the OFF state Λ_0 are λ and μ. The sojourn times in each state are negative exponentially distributed.

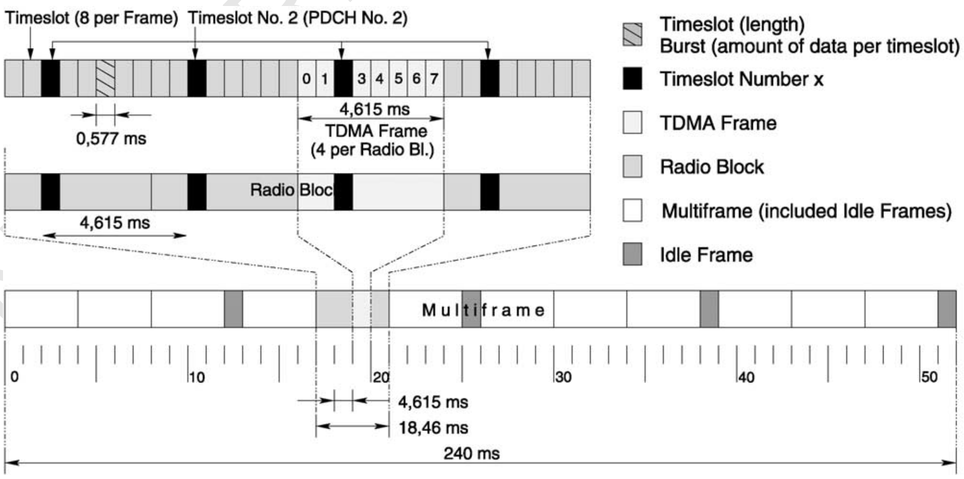

Fig. 1. GPRS multiframe structure.

Table 1
GPRS coding schemes (CS)

	CS 1	CS 2	CS 3	CS 4
PDCH data rate (kbit/s)	9.05	13.4	15.6	21.4
MAC block size (bit)	181	268	312	428
RLC block payload (bit (byte))	152 (19)	232 (29)	280 (35)	392 (49)

Each IRP needs the following parameters to be described completely:

- Activity factor α: fraction of time the source is active
- Mean burst length EN_B, $[EN_B] = byte$
- Bit rate during ON state h, $[h] = byte/s$

Thus, λ and μ can be derived to

$$\mu = \frac{h}{EN_B} \qquad \text{and } \lambda = \mu\frac{\alpha}{1-\alpha}. \tag{1}$$

The mean transmission rate of the source is $M = \alpha h$

Superimposing N equal IRPs, a so-called N Interrupted Rate Process (NIRP) can be defined [5,6], adding the number of superimposed IRPs N to the set of parameters described before. A NIRP can be described by a one-dimensional Markov chain (MC) (see Fig. 2). The state variable of this MC is the number of active IRPs, the total number of states is $N+1$.

For each state Λ_q of the NIRP's MC the corresponding transmission rate is $r_q = qh$ and the NIRP's mean transmission rate is $M = N\alpha h$.

Knowing the total rate r_q and the capacity C, the state space of a NIRP can be subdivided into:

- $\Lambda^u = \{\Lambda_q \in \Lambda \text{ with } r_q < C\}$: set of underload states
- $\Lambda^e = \{\Lambda_q \in \Lambda \text{ with } r_q = C\}$: set of uniform load states
- $\Lambda^o = \{\Lambda_q \in \Lambda \text{ with } r_q > C\}$: set of overload states

In an underload state, the buffer content depletes at the rate $C - r_q$, in an overload state the buffer content rises with rate $r_q - C$ and in a uniform load state the buffer content remains constant.

3.2. Fluid-flow analysis

In the following we shortly summarize the formulae that were used to obtain our results, for a complete derivation please refer to Ref. [6]. We regard the general case of a Fluid-flow multiplexer with buffer size K and N equal ON/OFF sources attached to it, which means that the complete arrival process is represented by a single NIRP.

Starting with the equilibrium probability of state Λ_q, and the buffer of maximum capacity K being filled with x bytes of data waiting for transfer

$$F_q(x,K) = Pr\{X \le x \text{ and } \Lambda_q\};\quad x \le K \tag{2}$$

a differential equation system can be set up, leading to an eigenvalue problem. Calculation of the equilibrium buffer size's CDF requires calculation of the eigenvalues z_q, the sum of each eigenvector's components and a set of coefficients to fit the solution to boundary conditions.

The eigenvalues z_q can be derived to:

$$\begin{aligned} z_q = {} & \frac{1}{2(C-qh)(C-(N-q)h)}(NC(\lambda+\mu) - N^2h\lambda \\ & + 2(N-q)qh(\lambda-\mu) + (2q-N)(C^2(\lambda+\mu)^2 \\ & + (Nh\lambda)^2 - 2NhC\lambda(\lambda+\mu) + 4(N-q)qh^2\lambda\mu)^{1/2}) \end{aligned} \tag{3}$$

Special cases that have to be treated separately are:

(1) $C = qh$:
The eigenvalue is undetermined.

(2) $C = (N-q)h$:
The eigenvalue is given by:

$$z_q = \frac{2(N-q)q(\lambda+\mu)^2}{N(2C-Nh)(\lambda+\mu) - (N-2q)^2(\lambda-\mu)} \tag{4}$$

The eigenvectors are calculated using the inverse eigenvalue problem (see Ref. [6]). The sum of the eigenvector components can be obtained by evaluating the generating function of the eigenvector components.

$$\Phi_q(1,z_q)=\sum_{i_1=0}^{\Gamma_1}\left(\binom{\Gamma_1}{z_q}(-res_{q,1}(z_q))^{\Gamma_1}\sum_{i_2=0}^{\Gamma_2}\binom{\Gamma_2}{i_2}(-res_{q,2}(z_q))^{\Gamma_2-i_2}\right) \tag{5}$$

with

$$\Gamma_1(z)=N_j-\Gamma_2(z)=\begin{cases} q & z>0 \\ (N-q) & z<0 \end{cases} \tag{6}$$

and

$$res_{q,1/2}(z)=\frac{1}{2\lambda}\left((\lambda-\mu-zh)\pm\sqrt{(\mu-\lambda+zh)^2+4\mu\lambda}\right). \tag{7}$$

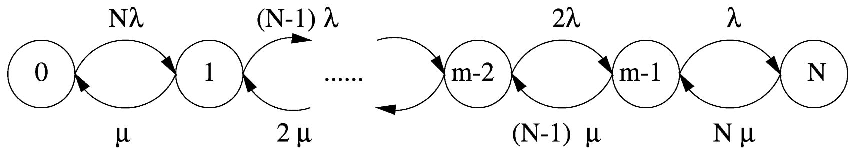

Fig. 2. NIRP state transition diagram.

In the scope of this article the available buffer memory is assumed to be unlimited, because there are no restrictions to the queue length in the simulation system as well. Accordingly, the coefficients that fit the solution to its boundary conditions are

$$a_q(\infty) = -\alpha^N \Phi_q(1,z_q) \prod_{\substack{s\in\Lambda^0 \\ s\neq q}} \frac{z_s}{z_s - z_q}. \tag{8}$$

The CDF $F(x)$ of the equilibrium buffer size is

$$F(x) = \sum_{\forall_q} a_q(\infty)\Phi_q(1,z_q)\mathrm{e}^{z_q x}. \tag{9}$$

Taking into account that the eigenvalue z_0 equals zero, the eigenvector $\vec{\varphi}_0$ equals the vector of the NIRP's steady state probabilities (see Ref. [5]), the sum of which is one and regarding the system's boundary conditions, which allow to derive $a_q(\infty)$ for some special cases (see Refs. [6,7] for a more detailed description), we finally receive for the buffer content's CDF and CCDF (denoted by $G(x)$):

$$F(x) = 1 + \sum_{q\in\Lambda^o} a_q(\infty)\mathrm{e}^{z_q x} \Rightarrow G(x) = -\sum_{q\in\Lambda^o} a_q(\infty)\mathrm{e}^{z_q x}. \tag{10}$$

Integration of $G(x)$ finally delivers the mean equilibrium buffer size:

$$E[x] = \int_0^\infty G(x)\mathrm{d}x = \int_0^\infty (1-F(x))\mathrm{d}x = \sum_{q\in\Lambda^o} \frac{a_q(\infty)}{z_q} \tag{11}$$

The mean waiting time can be obtained by Little's Law:

$$t_w = \frac{E[x]}{M} = \frac{1}{M}\sum_{q\in\Lambda^o} \frac{a_q(\infty)}{z_q} \tag{12}$$

3.3. FFM sojourn time

Up to this point we have been using the standard Fluid-flow analysis, which is well-established in the literature. Due to the different scheduling paradigms in the scope of FFM and GPRS, respectively, comparison of Fluid-flow analysis and our simulation system GPRSim cannot be based on backlogged traffic at the router or waiting time of IP datagrams. The GPRS MAC protocol does not apply strict FIFO queuing of the aggregated traffic, see Section 4.1 and [3,8]. Thus, we have to derive a quantity that can be compared with the IP datagram delay at the GPRS radio interface, based on the results obtained by Fluid-flow analysis. The basic idea is to add an offset to the mean waiting time that accounts for the minimum time an IP datagram transmission over the radio interface requires. We call the resulting quantity the FFM Sojourn Time.

The duration of a Radio Block Period (RBP) is $t_{\mathrm{RBP}} = 18.46$ ms and the duration of a TDMA frame is $t_{\mathrm{TDMA}} = 4.615$ ms. Evaluation of the IP datagram size in our simulation system leads to a mean IP datagram size $\overline{N_{\mathrm{IP}}}$ of 340 bytes. Division of $\overline{N_{\mathrm{IP}}}$ by $N_{\mathrm{RLC,CS}}$, the number of bytes contained in one RLC block (according to the applied coding scheme, see Table 1), leads to the mean number of RLC blocks needed for transmission of an IP datagram. The Multislot Capability (MSC) of a GPRS MS determines how many time slots the MS is allowed to use in a TDMA frame. Thus, the number of RLC blocks per IP datagram additionally has to be divided by the MSC, denoted by N_{msc}, to determine how many RBPs the transmission of an IP datagram takes. We denote the number of RBPs by R and write:

$$R = \left\lceil \left\lceil \frac{\overline{N_{\mathrm{IP}}}}{N_{\mathrm{RLC,CS}}} \right\rceil \frac{1}{N_{\mathrm{msc}}} \right\rceil, \tag{13}$$

where $\lceil\cdot\rceil$ denotes rounding to the next greater interger value. Multiplying with t_{RBP} then leads to the mean transmission delay of an IP datagram. Due to the fact that the scheduling at the MS is shifted by two TDMA frame durations, and after every three RBPs there is one idle frame (see Fig. 1), we have to add $(\lfloor R/3 \rfloor + 2)$ times the duration of one TDMA frame:

$$t_s = t_w + \left(Rt_{\mathrm{RBP}} + \left(\left\lfloor \frac{R}{3} \right\rfloor + 2\right)t_{\mathrm{TDMA}}\right) \tag{14}$$

For example for CS-2 we have 29 byte payload per RLC block (see Table 1)), this leads to a mean number of 12 RLC blocks per IP datagram. Furthermore assuming MSC 4 leads to 3 RBPs required for transmission ($R = 3$). Thus, the lower border for mean transmission duration is 69.225 ms.

4. Simulation environment

The (E)GPRS Simulator GPRSim [9] is a pure software solution written in C++. Models of Mobile Station (MS), Base Station (BS), Serving GPRS Support Node (SGSN), and Gateway GPRS Support Node (GGSN) have been implemented. The simulator offers interfaces to be upgraded by additional modules (see Fig. 3). Different from usual approaches to establish a simulator, where abstractions of functions and protocols are being implemented, the approach of the GPRSim is based on the detailed implementation of the GSM and (E)GPRS protocols. The TCP implementation includes slow start and congestion avoidance algorithms. This enables a realistic study of the behavior of EGPRS and GPRS. The protocol stacks of (E)GPRS are used during system simulation and are statistically analyzed under a well-defined and reproducible traffic load. Results gained with the GPRSim can be regarded as representative because the GPRSim has been validated by traffic performance measurements in operational GPRS networks [10]. A number of samples that is sufficient for performance evaluation have been used. Per simulation several thousands of web sessions have been

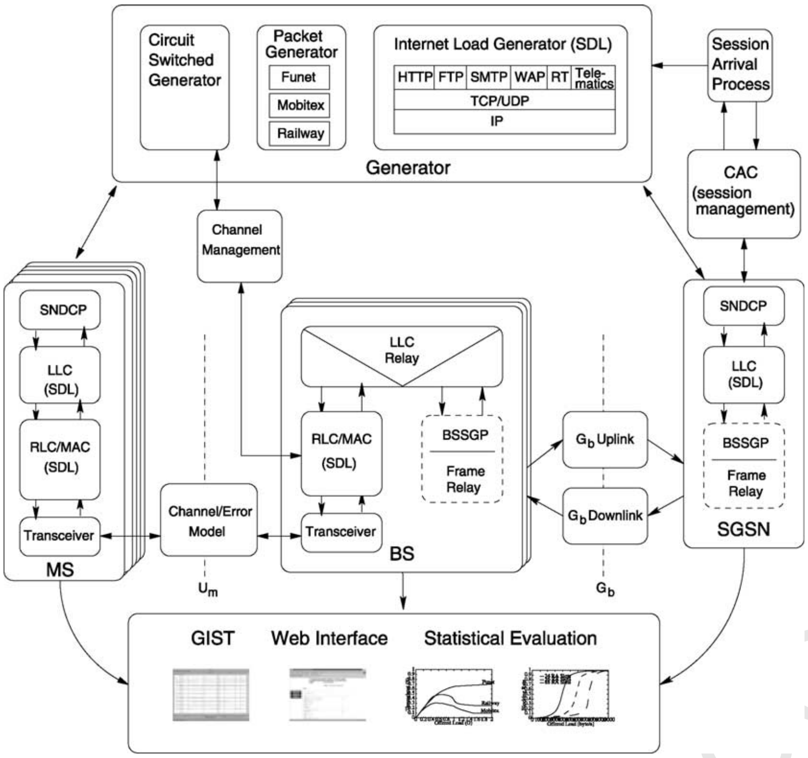

Fig. 3. Overview GPRSim structure.

simulated per mobile station in the scenario and the LRD algorithm [11] has been applied to ensure a relative error interval of less than 5% for the considered events with the lowest probability.

4.1. Scheduling of downlink traffic at the BSS

Scheduling of radio resources at the BSS can be subdivided into two steps: the selection of the next TBF and scheduling of the next RLC block of the selected TBF (see Fig. 4.1). Once a traffic class queue containing all TBF identifiers of this traffic class has been selected by the traffic class scheduler, the TBF scheduler selects one TBF of this TBF queue applying the TBF scheduling algorithm. This algorithm is also implementation-specific. As an example a round robin (RR) algorithm can be applied. The TBF scheduler only has the information that a TBF is established and has neither information on the amount of data to transmit nor if the TBF actually has data available. So the scheduler starts with the first TBF listed in the queue and checks if it has been allocated to there related PDCH. If not, the scheduler continues with the following TBF. In case the TBF is able to use the regarded PDCH the related RLC entity is polled for data until it reaches the predefined RR quantum or there are no more radio blocks to transmit. Then the following TBF of the same class queue is served if the same traffic class is still selected by the traffic class scheduler. Typically the RR quantum is in the order of 1–20 radio blocks.

In the next, step the RLC entity which has been polled for data by the MAC scheduler, checks if there are any data blocks available in the transmit buffer.

In case of RLC acknowledged mode the elements in V(B) indicate the acknowledgement status of related RLC data blocks. There are three possible states for each RLC data block:

- NACK indicates an RLC block which has not been transmitted yet, which has been negatively acknowledged or which has an expired timer
- PENDING_ACK indicates an RLC block which has been sent, but no acknowledgement has been received for this block yet
- ACK indicates data which has been sent and has already been acknowledged

The RLC block scheduling algorithm determines the order of transmission of the RLC blocks inside the RLC send buffer of a regarded TBF. The RLC data blocks in the RLC transmit window with the acknowledge state NACK, are forwarded to the MAC starting with the oldest one. If no NACK data block exists, the oldest RLC data block with the acknowledge state PENDING_ACK is retransmitted. The priority of NACK blocks to PENDING_ACK blocks inside one RLC entity is specified in ETSI's GPRS specifications.

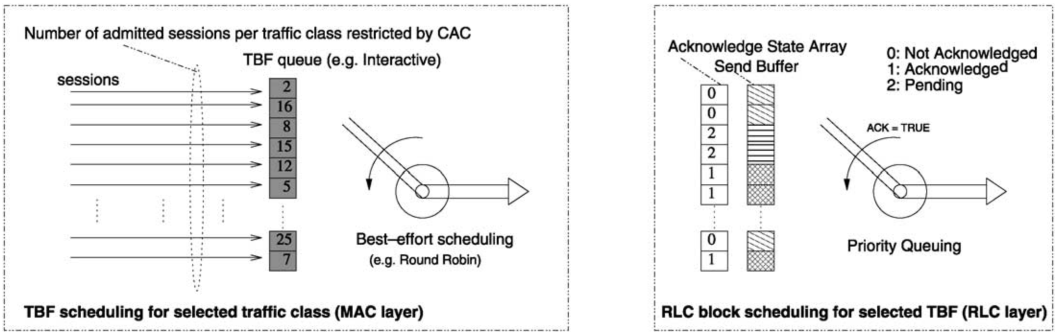

Fig. 4. Scheduling of TBFs and RLC blocks.

It is also specified that PENDING_ACK blocks should be transmitted if a RBP is scheduled for the regarded TBF and if no NACK block exists for this TBF.

4.2. WWW traffic generator

The WWW traffic generator is based on a empirically derived traffic model introduced in Ref. [12]. The parameters values of this models have been updated by values given by ETSI/3GPP propositions for the behavior of mobile Internet users [13]. For a detailed description of this Adapted Mosaic WWW Model see Ref. [2].

The Hypertext Transfer Protocol (HTTP) defines two ways setup and release of TCP connections can be controlled. There can be one TCP connection per requested Web object, or an existing connection can be 'kept alive' for the request of the next Web object belonging to the same Web request. This behavior is called keep-alive. Simulation can be performed with keep-alive alternatively enabled or disabled.

5. Evaluation of parameters for fluid-flow analysis

In order to determine FFM parameter values representing typical GPRS applications, the ON/OFF source parameters h, α and EN_B of the GPRSim WWW traffic generator module [9] have been evaluated, while the specific value for the FFM system capacity C was calculated from the data rate per PDCH and the number of PDCHs available in the regarded scenario.

The evaluation of the ON period is triggered by monitoring the TCP flags in the uplink direction. Once an uplink packet carrying the first segment of a TCP connection (SYN flag), the next downlink packet is counted as the first packet of an ON period. The last packet of this ON period is defined to be the last TCP segment of a TCP connection in the downlink direction (marked by the FIN flag). Thus, an ON period is equal to the lifetime of a TCP connection.

This evaluation procedure is affected by the HTTP keep-alive feature (see Section 4). If keep-alive is used, several Web objects belonging to the same Web page are transferred using the same TCP connection, which is equivalent to defining the download of a whole Web page as the ON phase of the WWW traffic model. Disabling keep-alive causes one TCP connection per Web object to be established. In this case an ON phase of the WWW model is defined as the download of a single Web object.

The IRP parameters have been evaluated using the following simulation scenario:

- 1 MS
- Traffic Mix 100% WWW
- 4 fixed PDCHs
- CS-2
- Error-free radio bearer
- Multislot capability 4

The resulting values are shown in Table 2. We see that in fact the HTTP keep-alive feature has a significant influence on the parameters that define an equivalent ON/OFF source. The enabled keep-alive feature results in a higher activity factor, caused by the longer duration of the TCP connections in this case. The mean transmission rate is higher, because the longer lifetime of a TCP connection leads to a higher maximum TCP send buffer size and thus reduces the impact of TCP flow control. Since the mean burst size is equal to the mean amount of data transferred in a TCP connection, a longer lifetime of a TCP connection and consecutive downloading of several Web objects leads to a higher mean burst size.

In case of an error-free radio channel the GPRS system capacity can be calculated to $C = N_{\mathrm{PDCH}} \cdot R_{\mathrm{PDCH}}$, where N_{PDCH} represents the number of PDCHs available to GPRS and R_{PDCH} the maximum IP throughput of a single PDCH depending on the applied Coding Scheme (CS).

Table 2
FFM parameters WWW over GPRS (1 MS)

Keep-alive enabled	Keep-alive disabled
$\alpha = 0.187346$ $h = 3272$ byte/s $EN_{\mathrm{B}} = 9149$ byte	$\alpha = 0.0859222$ $h = 2735$ byte/s $EN_B = 3677$ byte

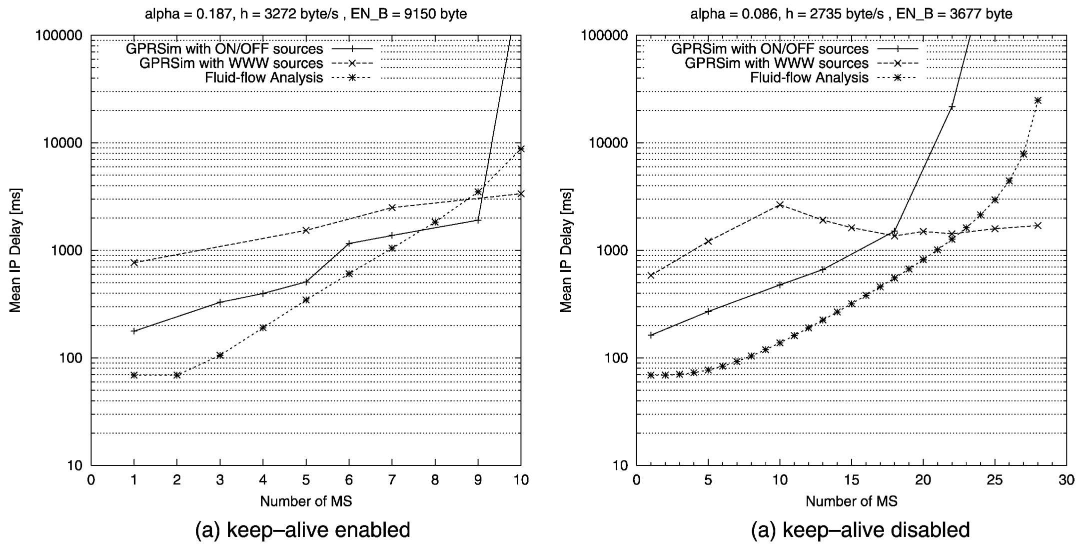

Fig. 5. Mean IP datagram delay, FFM parameters representing WWW over GPRS (1 MS).

6. Results and discussion

To eliminate inaccuracies evolving from deviations in source behavior between simulation and analysis, an implementation of the ON/OFF source model, parameterized to generate the same mean offered traffic as a single WWW traffic source was used on top of the GPRS protocol stack. Thereby an identical source behavior in both analysis and simulation was achieved, allowing an evaluation of the potential of the Fluid-flow approach to model the GPRS protocol stack and its performance. Each IRP instance was operating on its own LLC connection, which was kept established throughout the simulation. Data from all IRPs was served by a round robin algorithm; see Section 2 and [3], which is equivalent to the FFM's assumption that all data is served equally. In order to generate a mean arrival rate h during the ON state, we divide the given IP MTU by h to determine the required packet inter-arrival time. Once a source enters the ON state, a sample from the burst size random generator (negative exponential distribution with mean enb) is drawn and generation of IP datagrams with inter-arrival time as calculated before is started. Once the given burst size is reached, the source draws the length of the following OFF interval and switches to silence.

6.1. Performance evaluation

The parameter sets described in Section 5 have been used in the simulation with ON/OFF sources and the Fluid-flow analysis. The results of WWW with keep-alive are shown in Fig. 5(a) and for WWW without keep-alive in Fig. 5(b).

In both cases the Fluid-flow analysis and the simulation with ON/OFF sources show a similar tendency and the same order of magnitude. This shows that if the elastic property of WWW traffic is eliminated, the performance degradation of GPRS under an increasing traffic load can be depicted by FFM analysis. On the other hand it is visible in both cases that the FFM underestimates the IP datagram delay (more than 100%), which is caused by the conservative assumptions for deriving the transmission delay offset (see Section 3.3). The selected definition of the FFM sojourn time leads to a lower bound of the IP datagram delay and thus allows prediction of the theoretical maximum performance of GPRS under the given traffic load.

The third graph in both diagrams additionally represents a simulation series with the traditional WWW application model generating GPRS traffic. In the range of low system load the mean IP datagram delay is higher than predicted by FFM analysis. Advancing to higher system load a lower performance degradation than in both FFM analysis and simulation with inelastic traffic sources can be recognized. The TCP flow control procedures (i.e. TCP slow start and congestion avoidance, see Ref. [14]) applied by the WWW traffic source are responsible for this behavior, because they influence the burstiness of the IP arrival process. One single WWW load generator instance causes a higher load than the same instance would, when the same capacity is shared by several sources. To express this in terms of the Fluid-flow source parameters, the load of the GPRSim simulation scenario with 1 MS has a lower activity factor and a higher ON state transmission rate than the load of the simulation with 20 MSs. In the range of small load this causes higher delays than predicted by the FFM, because the source transmission rate in the active state is higher. Advancing into the range of high load, the delay does not rise as fast as the FFM predicts, because the traffic sources reduce the transmission rate during the active phases.

7. Conclusions

To eliminate inaccuracies evolving from deviations in source behavior between simulation and analysis, an implementation of the ON/OFF source model, parameterized to generate the same load as a single WWW traffic source, was used on top of the GPRS protocol stack. Thereby an identical source behavior in both analysis and simulation was achieved, allowing an evaluation of the Fluid-flow approach's capabilities to model the GPRS protocol stack and its various ways of influencing performance measures.

GPRS protocol behavior and performance degradation can be modeled using the FFM approach. Additionally the gap between the analytical curves and the GPRS simulation with identical source behavior is related to the overall loss of capacity that is inherent for a real GPRS environment. The performance degradation with increasing system load and the system's stability border are modeled quite well by the FFM.

Despite of that we have shown the limits of the IRP source model for elastic traffic types, the FFM approach seems to be an appropriate method of modelling statistical multiplexing of inelastic traffic like streaming applications (web radio, video streaming, etc.). Furthermore, the ability of the FFM approach to determine the sensitivity of the GPRS system to changes of system parameters like available bandwidth (i.e. the number of allocated PDCHs) or coding scheme used should be investigated. The influence of transmission errors can be taken into account by reducing the system capacity considering retransmissions.

Several extensions to the FFM are currently known, such as generally distributed state sojourn times, integration of batch arrivals or time-variant system capacity C, providing more accurate modelling of arrival and service process. Unfortunately the basic problem of modelling elastic traffic with an inelastic source model still remains unadressed by these enhancements.

References

[1] B. Walke, Mobile Radio Networks, Second ed., Wiley, New York, 2001.

[2] P. Stuckmann, The GSM Evolution—Mobile Packet Data Services, Wiley, New York, 2002.

[3] P. Stuckmann, Quality of service management in GPRS-based radio access networks, Telecommunication Systems 19:3, Kluwer Academic Publishers, 2002, pp. 531–546.

[4] L. Kosten, Stochasitc theory of a multi entry buffer, Delft Progress Report, vol. 1 of F, pp. 10–18, 1974.

[5] D. Anick, D. Mitra, M. Sondhi, Stochastic theory of a data-handling system with multiple sources, The Bell System Technical Journal 61 (1982) 1871–1894.

[6] M. Fiedler, H. Voss, Fluid-flow modelling of ATM-multiplexers (in German), Herbert Utz Verlag, Munich, Germany, 1997, ISBN 3-89675-251-0.

[7] T. Irnich, P. Stuckmann, Fluid-flow modelling of internet traffic in GSM-GPRS networks, International Symposium on Performance Evaluation of Computer and Telecommunication Systems, 2002, pp. 625–632.

[8] P. Stuckmann, Advanced scheduling and admission control techniques for cellular packet radio networks, Fifth European Personal Mobile Communication Conference (EPMCC) (2003) in press.

[9] P. Stuckmann, Simulation Environment GPRSim: Tool for Performance Analysis, Capacity Planning and QoS Enhancement in GPRS/EDGE Networks, Technical Report, http://www.comnets.rwth-aachen.de/~pst.

[10] P. Stuckmann, N. Ehlers, B. Wouters, GPRS traffic performance measurements, Proceedings of the IEEE Vehicular Technology Conference (VTC'02 fall), 2002.

[11] F. Schreiber, C. Görg, Stochastic simulation: a simplified LRE-algorithm for discrete random sequences, AEÜ, International Journal of Electronics and Communication 50 (4) (1996) 233–239.

[12] M. Arlitt, C. Williamson, A synthetic workload model for internet mosaic traffic, Proceedings of the 1995 Summer Computer Simulation Conference (Ottawa, Canada), 1995, pp. 24–26.

[13] ETSI 3GPP, Selection Procedures for the Choice of Radio Transmission Technologies of UMTS (UMTS 30.03, 3G TR 101 112), Technical Report, 1998.

[14] R. Stevens, TCP/IP Illustrated, vol. 1, Addison-Wesley, Reading, MA, 1996.