Performance Evaluation of Data Subsystem in GSM/GPRS Using Complete Sharing of Bandwidth

M.Mahdavi, R.M.Edwards AMIEE, P.A.Ivey MIEEE, email: {m.mahdavi, r.edwards, p.ivey}@shef.ac.uk

The Centre for Mobile Communications Research (C4MCR), Department of Electronic and Electrical Engineering,

The University of Sheffield, Mappin Street, Sheffield, UK S1 3JD

Abstract: In this paper we evaluate the performance of data users in a GSM/GPRS system using queuing analysis and the QDB approach. GPRS has been introduced in GSM as a proposed data packet service over the air interface. Currently GPRS uses a fixed number of time slots, which are not shared with GSM voice users. Therefore the QoS of GSM users that are circuit switched is decreased. Here we propose a complete sharing policy that prevents decrease in QoS for GSM users. In this policy, the data message in the head of FIFO queue is permitted use of all available free channels that are not already in use by voice users. A novel analytical model yielding mean delay and number of data users in the queue is also presented. In our model we give right of pre-emption to the voice users to prevent any effect from data users on voice users.

1 Introduction.

Since Internet applications are becoming widely used the call for support of packet data over the air interface is now topical as a goal for mobile systems. GSM, although popular is based for voice on circuit switching and supports few data applications including SMS. An enhancement has therefore been introduced in the form of GPRS that is based on GSM but adds packet switching using a pool of GSM channels. Mixing of service is possible, is anticipated and has been researched in various aspects for a range of networks in [1- 4]. This area is highly significant since the radio resource is restricted and the demand for service provision is increasing. For voice users to keep their QoS one would gives right of priority and pre-emption over data services. Also to support more voice users the total bandwidth has to be available for such users. This still allows non real-time services to use any unused bandwidth. This scheduling method is referred to as a *complete sharing policy*. This paper presents a method for the analysis of such a mechanism along with results derived from that analysis. The organisation of this paper is as follow: section 2, gives a brief description of the model and both the voice and data subsystems. In section 3 we present an analytical approach based on the phase service time Markovian system (M/G/1-type model) and apply the QDB approach. The proposed model and analysis are applied to a TDMA system. However, the model is flexible and can also be used for a CDMA based system. In Section 4 we present examples and the numerical results from our analysis. Simulation results are also used for comparison and corroboration. Finally a conclusion is presented in section 5.

2 General Model description

We assume channel is slotted with fixed slot duration T_s, which is equal to a packet length. Also each F slot makes a frame. Some of slots are used for signalling propose and not taken into account. Now let the number of traffic slots be N. With regard to the systems in total we propose that NT_s<<1, and we model the total bandwidth with N parallel servers or channels. Continuous time analysis is used. In the uplink a pool of mobile data users are in contention to send their requests for the capture of a channel/s. Data is backlogged until the base station dedicates a channel/s. Typically many data users would transmit to one base station. In uplink each mobile has its own queue. However by convention we could assume a virtual queue and that all successful mobiles will be queued here. In the downlink, since the base station is aware of all mobile users, there is no contention mechanism and data is sent whenever a channel allocated for a transfer. In the downlink we consider only one queue. We further assume a FIFO system for all queues.

2.1 Voice subsystem

We have assumed that voice calls arrive as a Poisson process with arrival rate λ_v and their call holding time is a negative exponential with service rate μ_v. We further assume that network only dedicates one time slot per voice user and that if all channels are in use by other voice users the call will be blocked. In the case that there is no idle channel and some channels are already in use by data user, the network allows the new call to pre-empt one of the channels in use by a data user. The voice users subsystem can be modelled using *M/M/N/N: N* - server loss system.

2.2 Data subsystem

In this part of the system, a user initiates request to download a file. Such downloads can be assumed to experience three delays, namely *access delay*, *round trip delay* and *retrieve delay*. The round trip and retrieve delays are related to the wired part of the system. Since we concentrate on wireless only access delay is addressed here. Because an upper band analysis is considered there is no contention in the uplink. In the performance analysis, consequent requests are generated with negative exponential interarrival times with a rate of λ_d, which is the aggregate arrival rate of the requests from users. We have modelled file length modelled with a geometric distribution in terms of the number of packets of which each file consists, with parameter p and q. We assume that all packets have the same fixed length. So:

$$\text{Prob}\,(\textit{Burst Length}=n) = pq^{n-1},\ n \geq 1 \ \text{and} \quad \text{Mean burst length} = \frac{1}{p} \qquad (1)$$

Where $p = 1 - q$. We further assume that mean number of packets in a burst is much larger than unity. This assumption is reasonable for most of Internet applications an example of which is FTP.

3. Queuing Analysis

Figure 1 shows the model of our system. We use this figure to establish our analytical model. In the next analysis we do not consider signalling

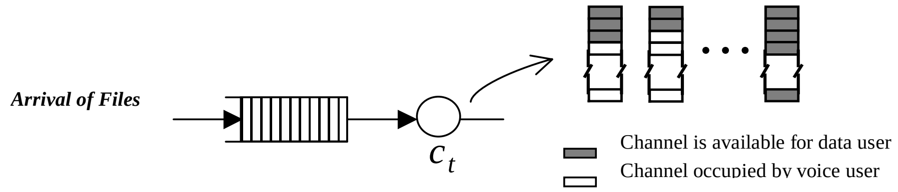


Figure 1:Model of the system

As the above figure shows, the available bandwidth or the service rate changes dynamically and depends on the number of voice users in the system. Using the same definition as in [5], we refer to the voice user subsystem as a "phase process" that operates independent of the data users behaviour or queue occupancy and has effect on data subsystem. The phase process is the birth-death process as we mentioned in section 2.1. With respect to the voice users' statistics the infinitesimal generator matrix Q, can be found using Equation (2). Since voice users have right of pre-emption, their QoS is considered to have a Blocking Probability that can be calculated using the Erlang-B formula.

$$Q = \begin{bmatrix} -\lambda_v & \lambda_v & 0 & \ldots & & & 0 \\ \mu_v & -(\lambda_v + \mu_v) & \lambda_v & 0 & \ldots & & 0 \\ 0 & 2\mu_v & -(\lambda_v + 2\mu_v) & \lambda_v & \ldots & & 0 \\ . & . & & \cdots & & & . \\ . & . & & \cdots & & & . \\ . & . & & \cdots & & & . \\ 0 & 0 & \ldots. & & (N-1)\mu_v & -(\lambda_v + (N-1)\mu_v) & \lambda_v \\ 0 & 0 & \ldots. & & 0 & N\mu_v & -N\mu_v \end{bmatrix} \qquad (2)$$

3.1 QDB approach

In this section we attempt to establish a Quasi-Birth-Death (QBD) analytical model in order to analyses our system. We define a negative exponential random variable b with parameter μ_d and $q = exp(-\mu_d)$. According to our assumption that mean burst length is higher than 1, so we could further approximate burst length distribution by continuous random variable b as follow:

$$\frac{1}{p} >> 1 \Rightarrow \mathrm{p} << 1 \ \Rightarrow \mathrm{q} = exp(-\mu_d) \approx 1 \qquad (3)$$

Consequently: $q \approx 1 - \mu_d$, considering the fact that (3) implies $\mu_d \approx 0$ and therefore $exp(-\mu_d) \approx 1 - \mu_d$. Thus:

$$\textit{Prob}\,(\textit{Burst Length}=n) = pq^{n-1} \approx \mu_d\, exp(-n\mu_d) \qquad (4)$$

Thus, (4) shows that burst length as a discrete random variable that can be approximated using a continuous random variable b. We use this approximation for our analysis. Let $i(t)$ be the number of retrieved files in the queue and $j(t)$ be the phase of the system at a time t. We claim that our process is a QBD process on the state space $(i(t), j(t))$. Later by comparing the results from our analysis with simulation, we justify our assumptions.

For QBD analysis we define: $P_{ij} = \lim_{t\to\infty} P(i(t),j(t))$. Figure 3 shows a general state diagram of the process. Referring to Figures 1 and 2 we can find that $\mu_j = (N\text{-}j)\mu_d$, for $j \le N$. Now let us assume that:

$$\pi_i = [P_{i0}\ P_{i1}\ \ldots\ P_{iN}] \quad \text{and} \quad S = diag(\mu_0, \mu_1, \mu_{2,\ldots,}\mu_N) \qquad (5)$$

Next using the matrix geometric method reported in [5].

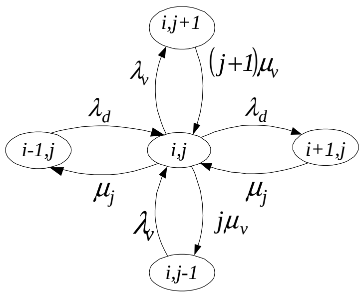


Figure 2: General State

And noting that in [5], there exist matrix R such that:

$$\pi_i = \pi_{i-1} R \quad \text{For } \forall i \ge 1 \qquad (6)$$

R can be found recursively using Equation (7) with initial condition R_0 as Equation (8). θ is some positive number. It has been shown that Equation (7) converges [5,6].

$$R_n = \theta R_{n-1}^2 S + R_{n-1}[I + \theta(Q - S - \lambda_d I)] + \theta\lambda_d I \qquad (7)$$

$$R_0 = diag\left(1/(M+1), 1/(M+1), \ldots, 1/(M+1)\right), \quad M > 0 \qquad (8)$$

The initial value of π_i which is π_0, can be obtained as:

$$\pi_0 = P_v(I - R) \qquad (9)$$

Where P_v is the solution of equation $P_v Q = 0$ as (10). Furthermore, the stability condition can be found as in Equation (11) where *e* is a row vector with all elements equal to one.

$$P_v = [P(j)],\ P(j) = \frac{\rho^j}{N!}\left(\sum_{k=0}^{N}\frac{\rho^k}{k!}\right)^{-1},\ \rho = \frac{\lambda_v}{\mu_v} \qquad (10)$$

$$\lambda_d < P_v.(e.S) \qquad (11)$$

Now having π_0, one can find π_i using (6)-(8). With π_i, the probability of having a certain number of bursts or corresponding data users in the queue obtained as:

$$P(i) = \pi_i.e^T \qquad (12)$$

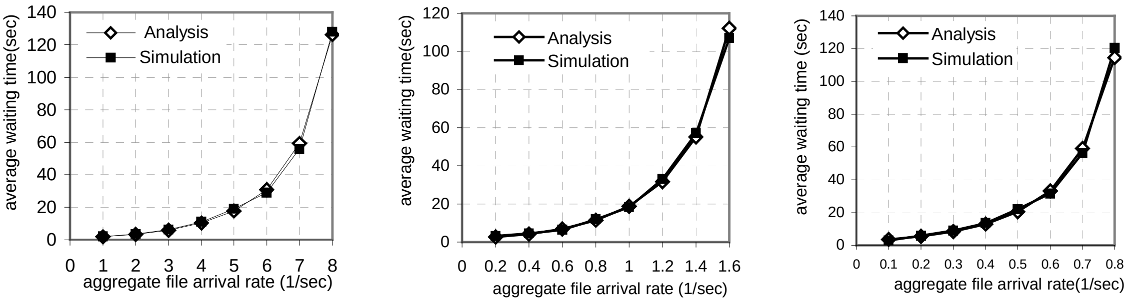
Also, the average queue length and waiting time in the system can be found [6], using (13) and Little's formula as in (14).

$$L_{ave} = \sum_{i=1}^{\infty} i.P(i) = P_v[I - R]^{-1} R.e^T \qquad (13)$$

$$W_{ave} = \frac{L_{ave}}{\lambda_d} \qquad (14)$$

4. Numerical Results for GSM/GPRS

For this paper to illustrate our analysis we have used a single cell of a GSM/GPRS mobile network. It is assumed that frequencies in neighbouring cells differ and therefore there is no interference effect. The GSM/GPRS standards indicate that both support eight time slots in a frame. For the case where both GSM and GPRS use the same pool of channels, we consider two time slots are used for signalling as in [4]. Consequently we have used $F = 8$ with two time slots be unavailable for traffic because of signaling. Therefore we have $N = 6$. We also consider CS-1 coding. Average holding time of voice is considered to be 120sec [7]. Initial blocking probability of 5% is considered. For data we have considered three different burst lengths: small file size of 1 Kbytes, medium file size of 5 Kbytes, and large file size of 10 Kbytes. Figure 3 shows the average waiting time in a queue as a function of the average arrival rate of data λ_d, with file size as parameter. It can be seem that the agreement between simulation and analysis is good. This lends weight to both the analysis and the simulation.



(a) (b) (c)

Figure 3: Average waiting time in the system for file sizes of: (a) 1 Kbytes, (b) 5 Kbytes, c) 10 Kbytes,

From these results and Equation (9) we can also note that there is a limitation on data traffic. Therefore knowing the file size, one can deduce the maximum arrival rate for which the system is remains stable. To highlight this, Table 1 presents approximate figures for the stability condition.

Table 1: Stability condition

File Size	Max λ_d
1kbyte	≈ 10 files/sec
5kbyte	≈ 1.92 files/sec
10kbyte	≈ 0.95 files/sec

5. Conclusion

In this paper we have presented an analysis method for mixed services of voice and a data in the GSM/GPRS system. To support more voice users, they can use all traffic channels and we give the right of priority and pre-emption to voice users. Therefore performance of voice subsystem can be found independent of data users. We also consider complete sharing of traffic channels. For data subsystem we assume that a data user in the head of buffer can use all unused channel by voice users. The results from our analytical approach and simulation show good agreement. The approach of ignoring the signalling channel/s and modelling TDMA channels with *N* parallel servers has been justified. The analytical model is principally for TDMA system and more specifically for the downlink, however it is quite wide and seems applicable for CDMA systems as well as for uplink. Use of the algorithm allows us to measure the upper band capacity of GSM/GPRS system with different file size in terms of delay in the queue.

References:

[1] I. Rubin, Z. Zhang," Message Delay for TDMA Schemes Using Contiguous-Slot Assignments ", IEEE Trans. On Communications, Vol. 40, No. 4, Apr 1992, pp. 730-737

[2] San-qi Li; Hong-Dah Sheng, "Discrete queuing analysis of multimedia traffic with diversity of correlation and burstiness properties" Communications, IEEE Transactions on, February-April 1994, pp. 1339 –1351

[3] C. Santivanez, I. Stavrakakis, " Study of various TDMA schemes for wireless networks in the presence of deadlines and overhead ", IEEE Journal on Selected Areas in Communications, Volume: 17 Issue: 7, July 1999, pp. 1284 –1304

[4] M.Mahdavi, R.M.Edwards, S.R.Cevetkovic, "Policy for Enhancement of Traffic in TDMA Hybrid Switched Integrated Voice/Data Cellular Mobile Communications Systems", IEEE Communications Letter, June 2001, pp.242-244

[5] M. F. Nuets," Matrix-Geometric Solutions in the Stochastic Models: An Algorithmic Approach ", Baltimore, MD: The Johns Hopkins University Press, 1981

[6] J. N. Daigle and D. M. Lucantoni, "Queuing systems having phase de-pendent arrival and service rates," in Numerical Solution of Markov Chains. New York: Marcel Dekkler, 1991, pp. 162–202

[7] ETSI TR 101 398 v3.0.1, "Universal Mobile Telecommunication system (UMTS); High level requirements relevant for definition of the UMTS Terrestrial Radio Access (UTRA) concept (UMTS 21.02 version 3.0.1)", October 1998