GPRS performance estimation in GSM circuit switched services and GPRS shared resource systems*

Shaoji Ni and Sven-Gustav Häggman

Helsinki University of Technology, Institute of Radio Communications, Communications Laboratory,
P.O. Box 2300, Otakaari 8, 02015 FIN-HUT, Finland, E-mail: ni@tiltu.hut.fi.

***Abstract*—GPRS is designed for transmitting packet data and supposed to take its radio resource from the pool of channels unused by GSM voice services. The number of channels allocated to GPRS is a random variable depending on the voice traffic. In this paper, an approximation method is used for evaluating the GPRS performance of single-slot service in the variable radio resource. The method could be used for evaluating the GPRS performance when the average service time of circuit switched services is much longer than that of GPRS. The multi-slot services cause higher blocking probability and longer delay to the network than the single-slot service. However, those effects can be reduced by implementing a GPRS resource allocation scheme with flexible multi-slot services.**

I. INTRODUCTION

The current method of data transmission in the pan-European Global System for Mobile Communications (GSM) and the American Advanced Mobile Phone Standard (AMPS) cellular networks is circuit switching. This technique reserves the traffic channel for the entire communication time, and wastes the radio resource when data traffic occurs in bursts with long silent intervals. In the development of GSM phase 2+, the European Telecommunications Standard Institute (ETSI) has specified a general packet radio service (GPRS) over the GSM to increase the utilization efficiency of the radio resource. The physical channels unused by circuit switched services are allocated dynamically to the GPRS according to the actual needs for packet transfers.

Earlier studies of GPRS performance found in the literature [1-2] focus on the protocol behavior with a fixed number of channels used for data transmission. However, the number of channels available to GPRS is a random variable depending on the voice traffic and the voice channels' occupancy, thus, the service statistics is a movable boundary Markov process [3-4]. The analysis of GPRS performance is a complicated problem especially as multiple classes of quality of service and multiple classes of users are supported in GPRS. In this paper, the GPRS performance, e.g., the average queueing time and blocking probability, in the variable resource is evaluated by an approximation method and simulations. The paper is organized as following: in section II, the principles of the radio resource allocation for GPRS calls are described; in section III, an approximation method for evaluating the performance of single-slot service is introduced and the performance of GPRS is discussed by the approximation method and simulations; the multislot services are discussed in detail; the conclusions are given in section IV.

II. PRINCIPLES OF THE RESOURCE ALLOCATION FOR GPRS

GPRS is designed to support from intermittent and burst data transfers to occasional transmission of large volume of data. The GPRS and GSM circuit switched services share the same radio resource. Whenever a channel is not used by circuit switched services, it may be utilized by GPRS. The allocation of physical channels for GPRS can be based on the needs for actual packed transfers which is referred to as “capacity on demand” principle. When the packet data channels (PDCHs) shared by all GPRS users are congested due to the GPRS traffic load and more resource available in the cell, the network can allocate more physical channels as PDCHs.

The GPRS does not require permanently allocated PDCHs. The operator can decide to dedicate permanently or temporarily some physical channels for GPRS traffic. However, the high speed circuit switched data (HSCSD) service supports as well multiple slot services and has higher priority to access the physical channels. As the introduction of HSCSD service into the GSM system, it might be difficult to guarantee the quality of service of GPRS if no channel is dedicated to GPRS.

The number of allocated PDCHs in a cell can be increased or decreased according to demand. In order to implement this principle, a load supervision function, which monitors the load of the PDCHs and the number of allocated PDCHs in a cell can be increased or decreased, must be used in the system. Upon resource demand for circuit switched services, some PDCHs must be released as soon as soon as possible. The release can have two alternatives:

1) *Immediate Release*: the GPRS user is forced to stop its transmission until resource is available for GPRS again and the channel released by GPRS is allocated to circuit switched services.
2) *Delayed Release*: the GPRS user can continue its transmission up to some frames or until the ending of packet transmission, before the channel is allocated to circuit switched services.

In the paper, the immediate release is assumed in order to investigate the relation between the average interruption time and interrupting probability and the GPRS traffic. If the average interruption time and interrupting probability are low,

* The project is financially supported by Nokia Research Center, Sonera, Technology Development Center of Finland and Helsinki Telephone Company.

the “Delayed Release” protocol can be considered to simplify the system.

III. GPRS PERFORMANCE EVALUATION

A. System model

For a system with m physics channels, m_v channels are shared by voice and GPRS services and m_d channels are dedicated to GPRS (Fig. 1). In the pool of m_v channels, when channels are not used by voice services, those channels are used for GPRS transmission. The voice services own preemptive priority over GPRS, i.e., whenever channels used by the GPRS service are needed by voice services, the GPRS transmission in those channels is stopped until some channels are available for GPRS. The users interrupted service have higher priority for resource allocation than those in queue.

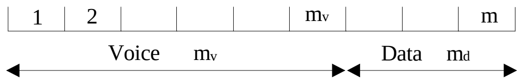

Fig. 1. Illustration of radio resource structure allocated to GSM voice services and GPRS. The number of m_v channels are shared with voice and GPRS services and m_d channels are dedicated to GPRS.

Assume that voice users arrival is a Poisson process with a rate of λ_v and the call service time is exponentially distributed with a mean of $1/\mu_v$. All GPRS users share the physical channels unused by the voice services. The arrival of GPRS data users are assumed to be a Poisson process with rate λ_d and the service time is exponentially distributed with a mean of $1/\mu_d$. The maximum number of data users accepted into the system (in service and queue) is N. GPRS calls are served according to the first in first out (FIFO) principle. The arriving GPRS user is allowed to transmit data if a sufficient number of free channels is available; otherwise it is queued or blocked.

B. An Approximation Method for Performance Evaluation

The voice services are independent of GPRS. Because GPRS is mainly designed to transmit intermittent and burst data, the service time of GPRS is rather smaller than that of voice services. As an approximation, the decomposition technique [4] can be used to analyze the GPRS performance. The essential of this technique is to use the voice services probability distribution to describe the interaction of voice services to GPRS. Thus, the GPRS performance in the dynamically variable resource is obtained by combining this distribution with the performance in a fixed resource.

For the voice services, the probability of n users in service is

$$r_n = r_0 \left(\frac{\lambda_v}{\mu_v}\right)^n \frac{1}{n!}, \quad n = 1, 2, ..., m_v \tag{1}$$

where $r_0 = \left[\sum_{n=0}^{m_v} \left(\frac{\lambda_v}{\mu_v}\right)^n \frac{1}{n!} \right]^{-1}$

The channels unused by the voice services may be used for the data services. The probability of x channels available for the data services is equal to that of m_v-x channels used by voice services and is obtained as by (1):

$$g(x) = r_0 \left(\frac{\lambda_v}{\mu_v}\right)^{m_v - x} \frac{1}{(m_v - x)!}, \quad x = 1, 2, ..., m_v \tag{2}$$

For the transmission of single slot GPRS in a fixed number of C channels, the average queueing time can be obtained from the $M/M/C/N$ queueing system, where N is the maximum number of users in the system (in service and in queue). The steady-state probability p_n is:

$$p_n = \begin{cases} p_0 \dfrac{\rho^n}{n!}, & n < C \\ p_0 \dfrac{\rho^n}{C!C^{n-C}}, & C \le n \le N \end{cases} \tag{3}$$

where n is the number of users in the system, $\rho = \lambda_d/\mu_d$ and

$$p_0 = \left[1 + \sum_{n=1}^{C-1} \frac{\rho^n}{n!} + \sum_{n=C}^{N} \frac{\rho^n}{C!} \frac{1}{C^{n-C}} \right]^{-1}$$

$$= \left[\sum_{n=0}^{C-1} \frac{\rho^n}{n!} + \frac{\rho^C \left(1 - \left(\frac{\rho}{C}\right)^{N-C+1}\right)}{C!(1-\rho/C)} \right]^{-1}$$

A new arrival is accepted into the system only if the number of GPRS users in the system is below the maximum accepted number N. Otherwise, the new arrival is blocked. The blocking probability is

$$P_N(C) = p_0 \frac{\rho^N}{C!C^{N-C}} \tag{4}$$

The average number of user in the system is obtained as

$$W(C) = \sum_{n=1}^{N} n p_n = p_0 \left(\sum_{n=1}^{C} \frac{\rho^n}{(n-1)!} + \frac{C^C}{C!} \sum_{n=C+1}^{N} \frac{n\rho^n}{C^n} \right)$$

$$= p_0 \left\{ \sum_{n=1}^{C} \frac{\rho^n}{(n-1)!} + \frac{C^C}{C!} \cdot \frac{(C+1)\left(\frac{\rho}{C}\right)^{C+1} - C\left(\frac{\rho}{C}\right)^{C+2} - (N+1)\left(\frac{\rho}{C}\right)^{N+1} + N\left(\frac{\rho}{C}\right)^{N+2}}{\left(1 - \frac{\rho}{C}\right)^2} \right\} \tag{5}$$

Combining (2) with (4) and (5), the average blocking probability, throughput and average queueing time of single slot GPRS in a dynamically varied resource are obtained as following expressions respectively:

$$\overline{P} = \sum_{x=0}^{m_v} g(x) P_N(x + m_d) \tag{6}$$

$$\overline{TH} = \lambda_d (1 - \overline{P}) \tag{7}$$

$$\overline{T} = \frac{1}{\lambda_d (1 - \overline{P})} \sum_{x=0}^{m_v} g(x) W(m_d + x) - \frac{1}{\mu_d} \tag{8}$$

For multiple class (slot) services, it is very difficult to analyze the state distribution probability of queueing system even in a fix number of transmission channels because of the large size of the state space. Here, we use simulation to evaluate the performance.

C. Numerical and Simulation Results

For single slot service, the results from the approximation and simulation are presented, but for multiple slot service, only results from simulation are given.

In the numerical calculations and simulations, 4 carriers, i.e., 4×8=32 channels in a cell are assumed, from which 1 channel is reserved for GPRS data and 31 channels are shared by circuit switched services and GPRS. In the simulations, when a new circuit switched call arrives, if no free channel is available and the number of circuit calls in service is below 31, one of GPRS calls is stopped its transmission in order to allocate one channel to the new circuit call. When resources are available, the interrupted GPRS calls have higher priority to be allocated resource than the queueing calls. The average interruption time and probability of interruption are simulated. The average service time of circuit switched services is exponentially distributed with a mean of 180 s. This traffic load of circuit services is 22.83 Erlang corresponding to 2% blocking probability for 31 channels. The maximum number of GPRS users allowed into network is 40.

C1 coding scheme is mainly used for signaling traffic and C2, C3 and C4 coding schemes are used for user data traffic. The C2 scheme corresponding with a transmission rate of 13.4 kb/s is assumed to be used. The GPRS message size is exponentially distributed with means of 2×13.4 kb, 5×13.4 kb, 10×13.4 kb respectively, corresponding to the mean service time (1/μ) of 2s, 5s and 10s with single slot transmission. The approximation results of single slot service are calculated from (6) and (8).

Fig. 2, 3 and 4 show the mean queueing time and blocking probability of single-slot service distributed to the traffic load for the average service time (1/μ) of 2s, 5s and 10s respectively. In this paper the average interruption time is included into the mean queueing time. Comparing the approximation results with simulation results, we find the approximation method could be used for evaluating the GPRS performance when the average service time of circuit switched services is much longer than that of GPRS, e.g., μ_d/μ_v>100. As the average service time of GPRS increases, the error becomes larger. The reason of error might be that the approximation method over-estimates the blocking probability (Fig. 4). The simulations show that the interruption probability of GPRS calls depends on the average message size (service time) more strongly than on the traffic load.

In the simulations of multislot services, the traffics in Table 1 are used. Two schemes are used for resource allocation to GPRS calls.

- *scheme*-1: On the arrival of a multislot (two or three slots) call from the queue, if the available channels are not enough for the call's requirement, it still stays in queue until its resource requirement is fulfilled.
- *scheme*-2: On the arrival of a multislot (two or three slots) call from the queue, if the available channels are not enough for the call's requirement, the available channels are allocated to the call.

TABLE1: GPRS traffics

Traffic-1	All arrival traffics are required single-slot service.
Traffic-2	In the arrival traffics, 70%, 20%, 10% of them are required single-slot, two-slots and three-slots services respectively, i.e., $\lambda=\lambda_1/0.7=\lambda_2/0.2=\lambda_3/0.1$.
Traffic-3	In the arrival traffics, 50%, 30%, 20% of them are required single-slot, two-slots and three-slots services respectively, i.e., $\lambda=\lambda_1/0.5=\lambda_2/0.3=\lambda_3/0.2$.

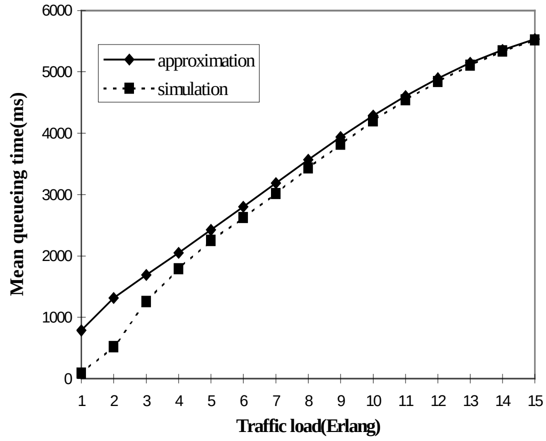

Fig2. The mean queueing time of single-slot service for the average service time (1/μ) of 2 s.

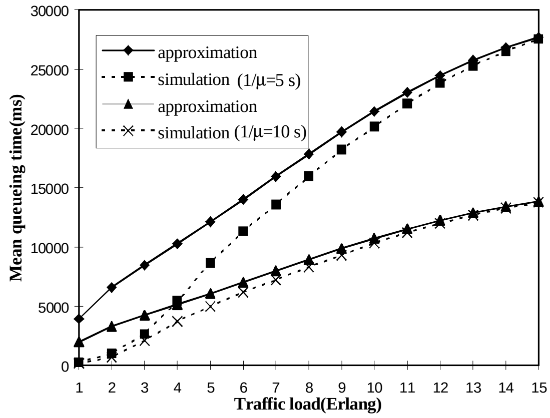

Fig3. The mean queueing time of single-slot service for the average service time (1/μ) of 5 s and 10 s respectively.

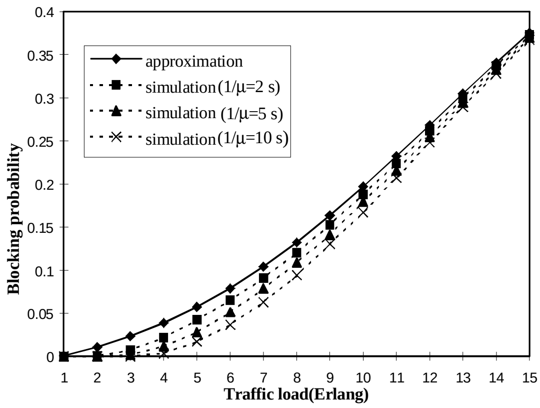

Fig. 4 The blocking probability of single-slot service for the average service time (1/μ) of 2 s, 5 s and 10 s respectively.

Fig. 5 and 6 show the mean queueing time of the three types of services with the resource allocation *scheme*-1 distributed with the call arrival rate for the average message size of 2×13.4 kb and 5×13.4 kb respectively. The solid lines are used for the traffic-3 and the dashed lines for traffic-2. In the same traffic type, the mean queueing time of the three types of services does not have much big difference, but it is much higher than that of single-slot service.

Fig. 7-10 show the mean queueing time (average for all served calls) and blocking probability with resource allocation *scheme*-1 and *scheme*-2 distributed with the call arrival rate for the average message size of 2×13.4 kb and 5×13.4 kb respectively. The performance of scheme-2 is better than that of scheme-1 and is similar to that of the single-slot-service system.

To investigate the multislot service, a parameter of the multislot served rate is defined as:

$$\text{served rate} = \frac{\text{number of calls served for three slots (or two slots)}}{\text{number of calls required for three slots (or two slots)}}$$

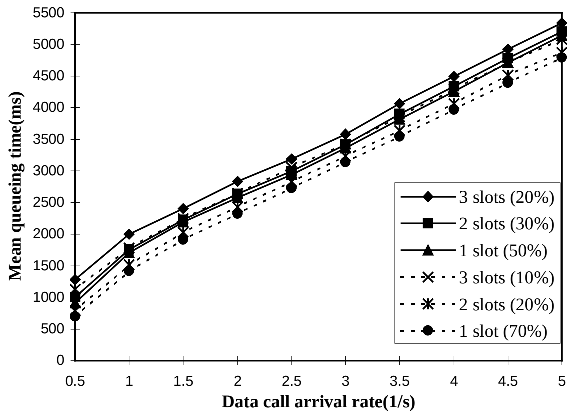

Fig. 5 The mean queueing time of the three types of services with the resource allocation *scheme*-1 for the average message size of 2×13.4 kb. The solid lines are used for the traffic-3 and the dashed lines for traffic-2.

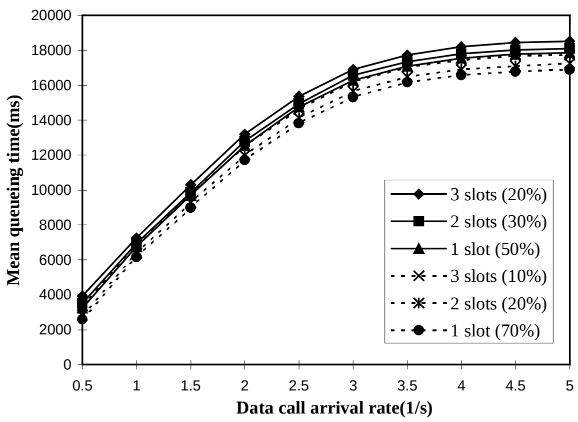

Fig. 6 The mean queueing time of the three types of services with the resource allocation *scheme*-1 for the average message size of 5×13.4 kb. The solid lines are used for the traffic-3 and the dashed lines for traffic-2.

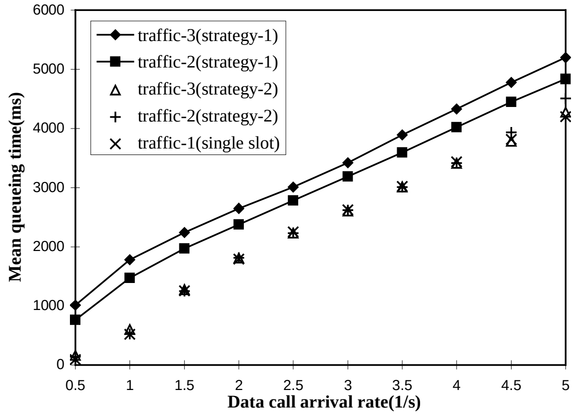

Fig. 7 The mean queueing time (average for all served calls) with resource allocation *scheme*-1 and *scheme*-2 for the average message size of 2×13.4 kb.

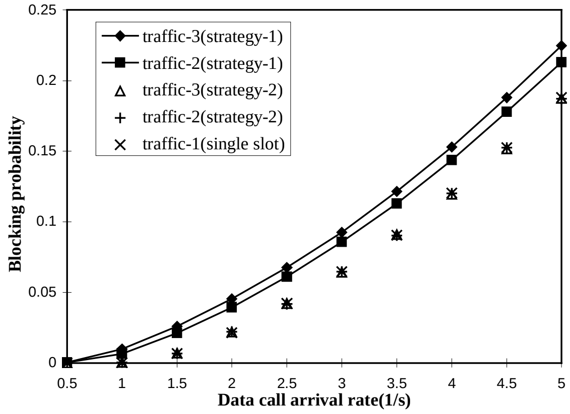

Fig. 8 The average blocking probability with resource allocation *scheme*-1 and *scheme*-2 for the average message size of 2×13.4 kb.

Fig. 11 shows the served rate of multislot services with resource allocation *scheme*-2 distributed with call arrival rate for the average message size of 2×13.4 kb and 5×13.4 kb respectively. The solid lines are used for an average message size of 2×13.4 kb and dashed lines for an average message size of 5×13.4 kb. For the traffic with the same average message size, the served rate for the same type of multislot service is almost same even though those traffics have different percentage of multislot service traffics. The served rate decreases as the increment of call arrivals and the average traffic size. Therefore, the resource allocation *scheme*-2 not only can reduce the blocking probability and delay caused by multislot services, but also can adapt to the GPRS traffic to provide the optimal performance to the network

The simulations of the multislot services also show that the interruption probability of GPRS calls depends on the average message size more strongly than on the traffic load.

Those simulations show that the higher blocking probability and longer delay caused by multislot services could be reduced by using a flexible multislot service scheme. The protocol of a flexible multislot service scheme is:

on the arrival of a multislot call,

- if the available resource is enough to provide its required service, the call is allowed to transmit its required rate;
- if the available resource is not enough to provide its required service, the network negotiates with the call to reduce its transmitted rate to the value which the network can provide;
- if the call agrees to transmit with the rate which the network can provide, the network further inquires the call if it wants to restore its required transmission rate when the network can provide;
- if the call does not agree to reduce its transmission rate, put it into the queue until the network can provide its required transmission rate.

IV. Conclusions

In this paper, an approximation method is used for evaluating the GPRS performance of single-slot service in the variable radio resource. By the comparison of numerical results and simulated results, it shows that the method could be used for evaluating GPRS performance when the average service time of circuit switched services is much longer than that of GPRS, e.g., $\mu_d/\mu_v>100$. The simulations show that the interruption probability of GPRS calls due to releasing its channel to the demand of circuit switched services depends on the average message size more strongly than on the traffic load. The multi-slot services cause higher blocking probability and longer delay to the network than the single-slot service. However, according to the simulation those effects can be reduced by implementing a GPRS resource allocation scheme with flexible multi-slot services. In this scheme, when the available network resource cannot provide a call with its required transmission rate, the network negotiates with the user and agrees on a transmission rate which the network can provide.

References

[1] G. Bianchi, A.Capone, L. Fratta, L. Musumeci, "Packet Data Service over GSM Networks with Dynamic Stealing of Voice Channels," in proceedings of IEEE Globecom 95, 13-17 Nov. 1995, Singapore, pp.1152-1156.

[2] D. Turina, P. Beming, E. Schster, A. Andersson, "A Proposal for Multi-slot MAC Layer Operation for Packet Data Channel in GSM," in proceedings of ICUPC'96. Sep 29-Oct 2 1996, Cambridge, USA, pp. 572-576.

[3] J.E. Wieselthier and A. Ephremides, "Fixed- and Movable Boundary Channel-Access Schemes for Integrated Voice/Data Network," *IEEE Trans.on Communi*. Vol. 43, No. 1, January 1995, pp. 64-74.

[4] S. Ghani and m. Schwartz, "A Decomposition Approximation for the Analysis of Voice/Data Integration," *IEEE Trans.on Communi*. Vol. 42, No. 7, July 1994, pp. 2441-2451.

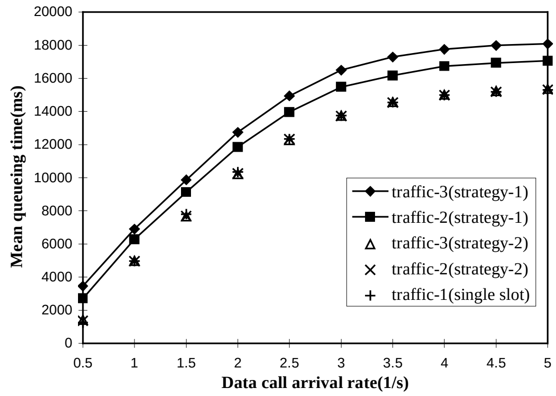

Fig. 9 The mean queueing time (average for all served calls) with resource allocation *scheme*-1 and *scheme*-2 for the average message size of 5×13.4 kb.

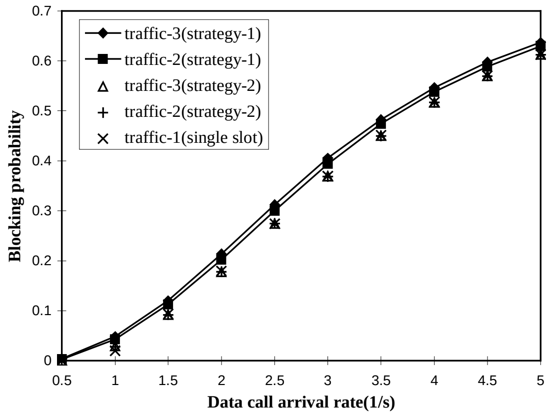

Fig. 10 The average blocking probability with resource allocation *scheme*-1 and *scheme*-2 for the average message size of 5×13.4 kb.

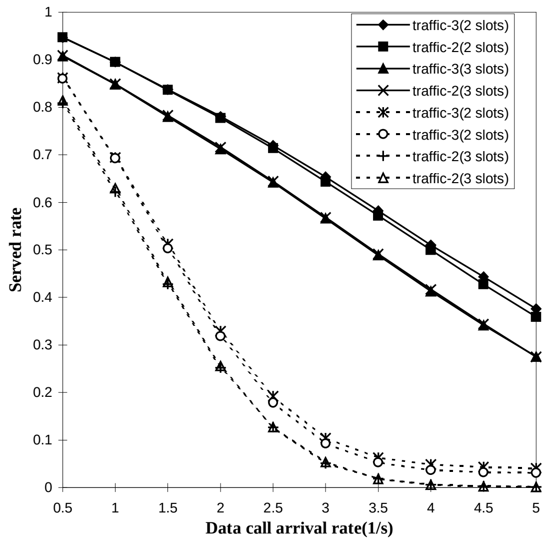

Fig. 11 The served rate of muitlslots services with resource allocation *scheme*-2 for the average message size of 2×13.4 kb and 5×13.4 kb respectively. The solid lines are used for the average message size of 2×13.4 kb and dashed lines for the average message size of 5×13.4 kb.