Proceedings of International Conference on Distributed
Systems, Software Engineering and Database Systems

Dynamic Resource Scheduling for GSM Data Services

Jeu-Yih Jeng*†, Chi-Wai Lin*, Yi-Bing Lin*

*Department of Computer Science and Information Engineering,
National Chiao Tung University, Hsinchu, Taiwan, R.O.C.,
Fax:+886-35-724176, Email:liny@csie.nctu.edu.tw.
†Telecommunication Laboratories, Chunghwa Telecom Co., Ltd.,
9 Lane 74, Hsin-Yi Road, Sec. 4, Taipei, Taiwan, R.O.C.,
Email:jyjeng@csie.nctu.edu.tw.

ABSTRACT

A new GSM data protocol called high speed circuit switched data (HSCSD) have been developed by European Telecommunications Standards Institute (ETSI) for high speed file transfer and mobile video applications. HSCSD increases data rate by using multiple TDMA time slots (up to 8) instead of one time slot in the current GSM implementation. The problem of multiple time slot assignment is that the blocking rate of the system will increase. This problem can be solved by flexible resource assignment where the service specifies the maximum and the minimum capacity. Based on the current available capacity of a base station, a user will be assigned any rate between the maximum and the minimum capacities. This article describes HSCSD protocol and presents four radio resource allocation strategies for HSCSD: always allocates maximum, always allocates minimum, allocates maximum unless available resources are not enough, and allocates resources according to the current blocking statistics of the base station. A simulation model is proposed to investigate the performance of these algorithms. The blocking probability, the call completion probability, and the quality of service are used to evaluate the effects of algorithms in different system behaviors.

KEYWORDS. GSM, HSCSD,Radio Resource Allocation, Blocking Probability, Quality of Service.

1 INTRODUCTION

GSM (Global System for Mobile Communications) [1] is a standard for the digital cellular telephone systems. Recently, data applications have been considered as important GSM services. Nokia foresees that half the information over GSM could be data by the beginning of the 21st century [2]. Current GSM standard supports two data service groups: GSM *Short Message Service* provides a connectionless transfer of messages with low-capacity and low-time performance. Each message can transmit up to 160 alphanumeric characters and operates like a paging service with the added capability that messages can pass in both directions and confirmation can be provided that the sent message has been received. The short messages are transported on the GSM signaling links. Thus, the messages can be received while the mobile users are in conversations. Two types of the short message services have been defined:

- *cell broadcast* service that periodically delivers short messages to all subscribers in a given area, and
- *Point-to-point* service that enables short messages to a specific user. This GSM feature can be considered as an enhanced two way paging service.

GSM *Bearer Services* are similar to the ISDN services (data circuit duplex, data packet duplex, and so on) except that the maximum data rates are limited to 9.6 kbps. These services can be employed on notebook PCs or PDA connected to mobile station via a PC card. To offer a bearer service, a circuit-switched connection is established in the GSM network to connect the mobile station and an interface of the Public Switched Data Network.

The current GSM data services do not support the fast access to sufficient radio resources on demand. Thus, GSM cannot effectively support Internet applications such as large volume FTP. To provide efficient data capabilities for GSM, a new GSM data protocol called *high speed circuit switched data* (HSCSD) have been developed by European Telecommunications Standards Institute (ETSI) as part of GSM phase 2+ standard. HSCSD is designed specifically for high speed file transfer and mobile video applications. The data rate has been increased by using multiple TDMA time slots instead of one time slot. It causes high blocking rate of the system. To reduce the blocking probability, some flexible assignment strategies are proposed.

2 HSCSD Protocol

High Speed Circuit Switched Data (HSCSD) is a circuit-switched protocol for large file transfer and multimedia applications. The physical layer of HSCSD is the same as that of the current GSM data services. The HSCSD architecture is illustrated in Figure 1. The data computing session is performed at a *terminal equipment* (TE; e.g., computer) connected to the mobile station using the *terminal adaption function* (TAF). TAF performs the adaption between TE and the generic radio transmission part of the mobile station. The *network interworking function* E and the generic radio transmission part of the mobile station. The *network interworking function* (IWF) supports adaption between GSM and the external networks. Thus, TAF and IWF are sensitive to the end-to-end services. On the other hand, the GSM entities between TAF and IWF are independent of the services. They only provide the bearer capabilities required to transport the corresponding data flow. The radio interface is the same as that of the current GSM system except that multiple, independent time slots can be utilized to provide high speed link as previously described. The *radio link protocol* (RLP) has been enhanced in HSCSD to support multi-link (time slot) operation. The protocol may or may not recover the frame errors between TAF and IWF. The problem of multiple time slot assignment is that the blocking rate of the system will increase (that is, less customers can share the GSM services if more radio resources are assigned to individual customers). This problem can be reduced by flexible resource assignment where the service (or the customers) specifies the maximum capacity and the minimum capacity. Based on the current available capacity of a base station, a customer will be assigned any rate between the maximum and the minimum capacities. This problem is similar to the *soft request* resource allocation [3] for a wireline network. A fundamental distinction between the two problems is that in HSCSD, before a request is completed, it may release resources in a base station BS_1, and request resources in another base station BS_2 because the user moves from BS_1 to BS_2 (this phenomenon is called *handover*). Let C be the total number of channels of the base station (BS). Assume that the minimum capacity required to make a HSCSD connection is C_{min} and the maximum capacity is C_{max}. This paper studies four resource allocation strategies for HSCSD. These strategies attempt to strike a balance between two common performance measures: the *blocking probabilities* for new/handover calls and the *quality of service* (QOS) for each request. The QOS is determined by several factors (and is formally defined in (1)) such as the average resources utilized when the call is connected and whether the call is completed or forced terminated. We consider four algorithms: H1 (always allocates maximum), H2 (always allocates minimum), S1 (allocates maximum unless available resources are not enough), and S2 (allocates resources according to the current blocking statistics of the base station). The details of the algorithms are described as follows.

Algorithm S1: Let $n < C$ be the number of free channels in the base station (BS) when a request arrives. One of the three cases are executed:

- **CASE I.** If $n \geq C_{max}$, then the BS allocates C_{max} channels for the request.
- **CASE II.** If $C_{min} \leq n < C_{max}$, the BS allocates n channels for the request.
- **CASE III.** If $n < C_{min}$, then the BS blocks the request.

Algorithm S2: The BS assigns C^* ($C_{min} \leq C^* \leq C_{max}$) channels to a request. The value C^* is adjusted dynamically based on the blocking statistics of the BS. Initially, C^* is set to C_{max}. When a call arrives, one of the two cases is executed:

- **CASE I.** If $n < C^*$, then the BS blocks the request. The BS updates the blocking statistics. If the blocking probability of the BS is larger than a threshold θ ($0 < \theta < 1$) and $C^* > C_{min}$, then C^* is decremented by 1.
- **CASE II.** If $n \geq C^*$, then the BS allocates C^* channels for the request. The BS updates the blocking statistics. If the blocking probability of the BS is smaller than the threshold θ and $C^* < C_{max}$, then C^* is incremented by 1.

Algorithm H1. H1 is similar to S1, except that it either allocates C_{max} to a request or rejects the request

Algorithm H2. H2 is the counterpart of H1. Instead of the maximum number, this algorithm only allocates the minimum number (C_{min}) for every request it accepts.

3 The Simulation Model

This section proposes a simulation model to investigate the performance of the four algorithms. We consider 64 BSs connected as an 8×8 wrapped mesh [4]. A mobile

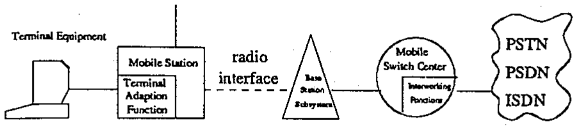


Figure 1 – The HSCSD Architecture

user resides in the coverage area of a BS for a period of time (in our study, the period is exponentially distributed with mean $1/\eta$). The user moves to the coverage area of one of the four neighboring BSs with the same routing probability (i.e., 0.25). In other words, the user movement follows a two-dimensional random walk model. The call arrivals to each BS form a Poisson process with arrival rate λ. The call holding times are exponentially distributed with mean $1/\mu$. The number of channels allocated to a call request depends on the resource allocation algorithm being exercised (in S2, the threshold is $\theta = 0.05$). In the experiments, Every BS has $C = 48$ channels. For every request, $C_{max} = 5$, and $C_{min} = 2, 3, 4$ are considered. If the user moves from a BS to another BS before the call is completed, then a handover is performed. The user first releases the resource of the old BS, and requests resources at the new BS. The new BS allocates the resources to the handover call the same as that to a new call. In each experiment, 100,000 incoming calls are simulated to ensure that the simulation results are stable. The following output measures are considered in our experiments.

- p_o : the probability that a request (either a new or a handover call) is rejected
- p_i : the probability that a call is connected but is eventually forced terminated
- p_c : the probability that a call is completed; i.e, $p_c + p_i + p_o = 1$.
- C_a : the expected number of channels allocated to a connected call
- QOS : the quality of service for a call, which is defined as:

$$QOS = (p_i t_i \alpha + p_c t_c)\left(\frac{C_a}{C_{\max}}\right) \quad (1)$$

where t_i is the expected holding time for an incomplete call, t_c is the expected holding time for a complete call, and α is the discount factor for an incomplete call. Note that $0 \leq \alpha \leq 1$. This discount factor is used to distinguish the QOS levels of the complete and the incomplete calls.

4 Results and Discussions

Based on the simulation experiments, we observe the following results.

Effects of the Resource Allocation Algorithms. Figure 2 shows the performance of the four algorithms.

In the experiments, $C_{\max} = 5$, $C_{\min} = 2$, $\eta = 0.5\mu$, $\alpha = 1$ (no discount for incomplete call), and the offered load (λ/μ) ranges from 1 to 31. H1 and H2 are two extremes in handling soft requests. Since H1 allocates maximum resources to every request, it has the worst p_o performance, as indicted in Figure 2 (a). On the other hand, H2 allocates minimum resources to every request, it has the best p_o performance. S1 behaves similarly to H1 except that it tries to partially allocates the remain resources to the "last" requests. Figures 2 (a) and (c) indicate that S1 has better p_o and QOS measures compared with H1. S2 dynamically allocates resources according to the current blocking statistics of the base station. This heuristic maintains p_o at an acceptable level like H2 (see Figure 2 (a)), but provides better resource utilization compared with H2 (see Figures 2 (b) and (c)). Thus, S2 also remedies the low QOS problem of H2 at light load. Figure 2 (b) shows that, to maintain low blocking probability, C_a of S2 drops sharply as the offered load increases. The QOS of S1 always outperforms S2 when the discount factor $\alpha = 1$ and when the initial call acceptance cost is not considered (to be elaborated).

Effects of Mobility. Figure 3 indicates that both p_o and p_c decrease as η increases (the offered load is 20). When η increases, the system experiences larger handoff arrivals and shorter channel occupancy times. One would expect that p_c increases as p_o decreases. Thus, our results are counter intuitive, which are due to the fact that a call will experience more handovers as η increases, and has more opportunities to be forced terminated. These phenomena are consistent with our previous studies for the voice cellular systems [5]. However the mobility η has the similar effects on the four algorithms.

Effects of Softness. As $C_{\min}$ increases (i.e., the "soft" requests become "harder"), the behavior of the four algorithms become similar. It is not surprising that when $C_{min} = C_{max}$, the four algorithms are the same.

Effects of the Discount Factor α. If a call is incomplete, it is reasonable to charge the call with a discount to increase the customer satisfaction. As the discount factor α decreases, S2 becomes a promising resource allocation strategy. Figure 4 (a) indicates that for $\alpha = 0.1$, S2 outperforms S1 when the offered load > 20.

Effects of the Initial Call Acceptance Cost β. In some telecommunication systems, an accepted request is charged a basic fee no matter how many channels it occu-

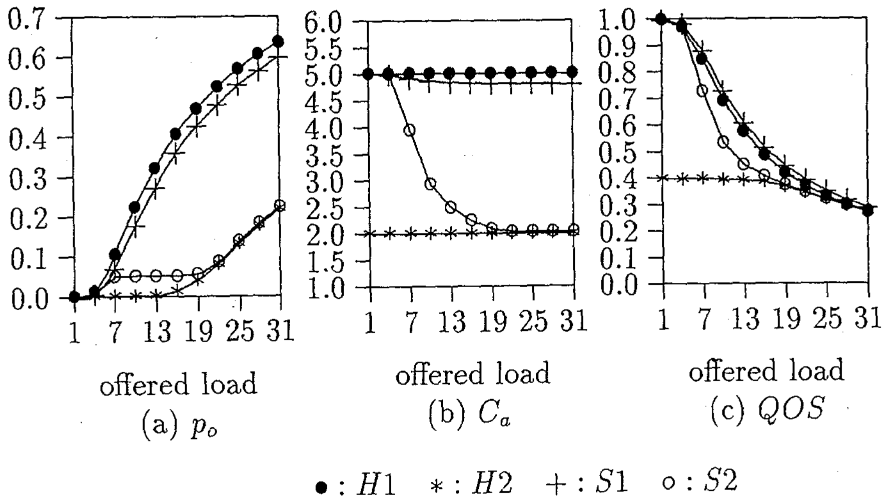


Figure 2 – **Effects of the Resources Allocation Algorithms**

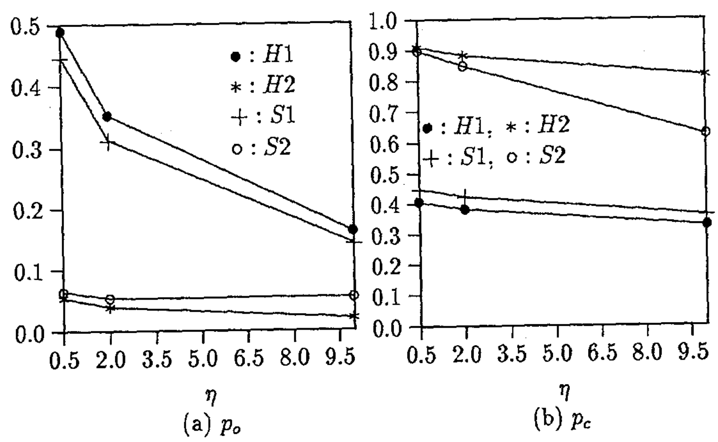
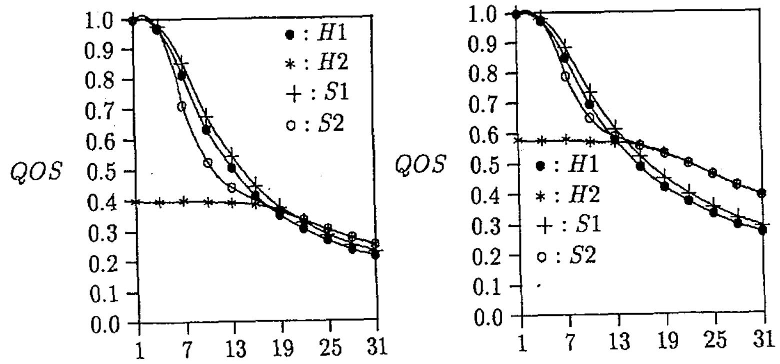


Figure 3 – **Effects of** η



(a) The effect of α

(b) The effect of β

Figure 4 – Effects of α and β

pies. In this case, it means that when a call is connected, a "basic" QOS has already offered to the user. Thus, the QOS measure may be modified as:

$$QOS = (p_i t_i \alpha + p_c t_c)\left[\beta + (1-\beta)\left(\frac{C_a}{C_{\max}}\right)\right]$$

where β ($0 \leq \beta \leq 1$) is the initial call acceptance cost (or the basic QOS) factor. Figure 4 (b) illustrates the QOS measures of the four algorithms when $\beta = 0.3$. Note that S2 outperforms the other three algorithms when the offered load > 15. Thus, as the β becomes a dominant factor, S2 will be a promising strategy for HSCSD resource allocation.

5 Conclusions

Flexible resource assignment is a good solution to reduce high blocking rate of HSCSD due to multiple time slot assignment. In this paper, we proposes four radio resources allocation strategies for HSCSD of GSM data services. Our study indicated that soft requests always outperform hard requests. S1 has better p_o and QOS measures compared with H1, and S2 remedies the low QOS problem of H2. When the discount factor or the initial call acceptance cost are not considered, S1 provides the best performance to HSCSD. If the discount factor is considered or the initial call acceptance cost becomes a dominant factor, then S2 is a good candidate for HSCSD resource allocation.

References

[1] Mouly, M. and Pautet, M.-B. "The GSM System for Mobile Communications," M.Mouly, 49 rue Louise Bruneau, Palaiseau, France, 1992.

[2] Debon, P. "The Future of GSM," Telecommunications, pp.54-58, March 1996.

[3] Lin, Y.-B., Lin, Y.-J., and Mak, V.W.K. "Allocating Resources for Soft Requests - A Performance Study," Information Sciences, 85(1):39-65, 1995.

[4] Lin, Y.-B.,Noerpel, A., and Harasty, D. "The Sub-rating Channel Assignment Strategy for PCS Handoffs," IEEE Trans. Veh. Techno., 45(1), February 1996.

[5] Lin, Y.-B., Mohan, S., and Noerpel, A. "Queueing Priority Channel Assignment Strategies for Handoff and Initial Access for a PCS Network," IEEE Trans. Veh. Techno., 43(3):704-712, 1994.