Performance Analysis of Radio Resource Allocation for Multimedia Traffic in Cellular Networks

Wei-Yeh Chen
Department of Electronic Engineering
Northern Taiwan Institute of Science and Technology
wychen@ntist.edu.tw

Jean-Lien C. Wu
Department of Electronic Engineering
National Taiwan University of Science and Technology
jcw@et.ntust.edu.tw

Abstract

In this paper, we study the performance of radio resource allocation for cellular networks where real-time (RT) and non-real-time (NRT) data traffic are considered in addition to voice traffic. To guarantee the performance of voice traffic, preemptive priority is used so that voice calls can preempt data packets and RT data traffic preempts NRT data traffic. By reserving channels specifically for RT data, the packet loss probability and mean queueing delay can be reduced, meanwhile, the performance of NRT data can also be improved. This is achieved at the expense of increasing voice blocking probability. Thus, the number of reserved channels can be adjusted to control the voice blocking probability while improving the performance of data traffic. Besides, the packet loss probability of RT traffic is smaller for smaller RT traffic ratio. On the contrary, the smaller the RT traffic ratio, the larger the queueing delay.

1. Introduction

Next generation cellular networks are expected to support multimedia services, such as voice, video, and data [1,2]. Due to the rapid growth in mobile users and the limited scarce radio resources, efficient management of radio resources becomes a key factor in enhancing the network performance [3].

Several resource allocation schemes [4-9] have been proposed to improve system performance in integrated voice/data wireless networks. Lee et al. [4] proposed a scheme which combined the queueing strategy and priority control to improve the performance of multiclass calls in multiservice personal communications services. Lin *et al*. proposed several resource allocation algorithms

to investigate the impact of GPRS on the GSM network [5]. In their study, buffers are used to queue the delay-sensitive traffic only, and preemption is not considered. In [6], three resource allocation strategies, depending on whether to apply buffer for data packets, are studied in GSM/GPRS networks. Huang *et al*. [7] exploited call admission control and resource reservation schemes to adaptively allocate resources to meet the different service demands in wireless multimedia networks. Kim used a two-dimensional Markov chain model on voice and stream data services and the residual capacity concept on packet data service to investigate the performance of an integrated voice/stream-data/packet-data CDMA mobile system [8].

In addition to traditional voice traffic, real-time (RT) and non-real-time (NRT) data traffic are considered in this paper. The performance analysis of radio resource allocation for multimedia traffic in cellular networks is based on a three-dimensional Markov chain. To guarantee the voice performance not being affected by the introduction of data traffic, preemptive priority is applied for voice calls. Besides, RT data traffic is given preemptive priority over NRT data traffic. Our focus is on the performance of RT traffic and NRT traffic, in terms of packet loss probability and mean queueing delay.

The remainder of the paper is organized as follows. Section 2 describes the resource allocation scheme adopted in this study. The performance analysis is given in section 3. Section 4 provides the performance evaluation in numerical results. Section 5 concludes this work.

2. Radio Resource Allocation

There are three different resource allocation strategies for integrated voice/data services, i.e., complete partitioning, complete sharing, and partial sharing [9]. The results in [9] shows that the complete sharing strategy

gives better system utilization than the other two strategies. Therefore, the complete sharing strategy is adopted in this work.

To guarantee the voice performance, we assume that voice traffic has preemptive priority over data traffic. When there are no channels available upon a voice arrival, one of the NRT data packets in service is preempted. If there are no NRT data packets in service, then one of the RT data packets in service is preempted. If there are no data packets in service, the voice arrival will be blocked. Furthermore, we assume that the preempted data packets are buffered in queue.

Similarly, RT data traffic is assumed to have preemptive priority over NRT data traffic to maintain the QoS of RT traffic. When there are no channels available upon an RT data packet arrival, one of the NRT data packet in service is preempted. If there are no NRT data packets in service, the RT data arrival is dropped. When there are no channels available and the number of NRT data users in the system, in service and in queue, is below a maximum allowed number upon a NRT arrival, it is queued in the buffer. When the number of NRT data users in the system exceeds the maximum allowed number, the arrival is dropped. To further improve the data packet loss probability, some channels are reserved specifically for RT data traffic.

Two buffers are used to accommodate RT and NRT data packets, respectively. One is for the preempted RT data packets caused by voice preemption. The other is used to accommodate NRT data packets, new and preempted.

Being buffered in queue, the preempted RT data packets have priority over NRT data packets to obtain services. Besides, the preempted packets have priority over new data packets to obtain services, and are served in the first-come-first-served (FCFS) manner.

3. The Analytical Model

In this section, we will describe the analytical model based on a three-dimensional Markov chain. Let the state (i, j, k) denote that there are i voice calls, j NRT data packets, and k RT data packets in the system. $\pi_{i,j,k}$ denotes the state probability of the system in state (i, j, k). The total number of channels in the system is C and the number of reserved channels for RT traffic is Cg. The maximum number of NRT data users accepted into the system is N. The buffer size for RT traffic is set to be the total number of channels which is large enough to assure no preempted RT data packets being dropped due to buffer overflow.

The arrival of voice call requests forms Poisson process with a rate of λ_v. The service time of voice calls is assumed to be exponentially distributed with a mean of $1/\mu_v$. The arrivals of RT and NRT data packets are assumed to be Poisson processes with a rate of λ_2 and λ_1, respectively, and $\lambda_d = \lambda_1 + \lambda_2$ is the aggregate data arrival rate. The ratio $r = \lambda_2 / \lambda_d$ indicates the portion of RT data traffic among aggregate data traffic and is defined as RT traffic ratio. The service time of RT and NRT data packets is exponentially distributed with a mean of $1/\mu_2$ and $1/\mu_1$, respectively. For simplicity, we let $\mu_1 = \mu_2 = \mu_d$ in the analysis. Fig. 1 shows the state transition diagram. Let S be the set of feasible states,

$$S = \{(i, j, k) \mid 0 \le i \le C - Cg,\ 0 \le j \le N, \\ \text{and } 0 \le k \le C\} \tag{1}$$

For all $(i, j, k) \in S$, the transition rates of the Markov process are explained in the following:

1) $L_v(i, j, k)$ is the transition rate from state (i, j, k) to $(i+1, j, k)$. A voice user is admitted into the system as long as the number of voice users in the system is less than the number of available channels for voice traffic, i.e., C-Cg. Therefore, $L_v(i, j, k)$ can be written as

$$L_v(i, j, k) = \begin{cases} \lambda_v, & \text{if } 0 \le i < C - Cg \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

2) $M_v(i, j, k)$ is the transition rate from state (i, j, k) to $(i-1, j, k)$, which can be written as

$$M_v(i, j, k) = \begin{cases} i\mu_v, & \text{if } 1 \le i \le C\text{-}Cg \\ 0, & \text{otherwise} \end{cases} \tag{3}$$
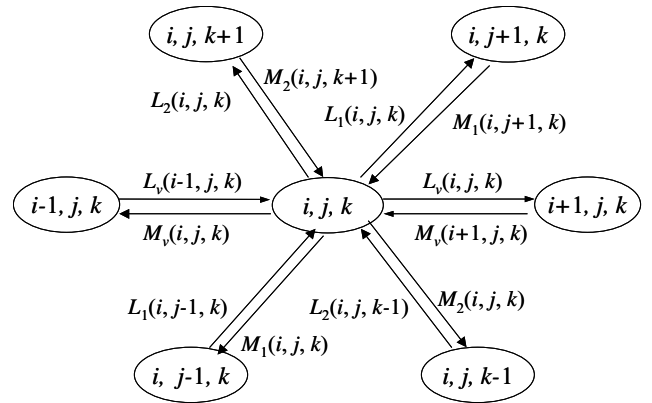

Fig. 1 The state transition diagram

3) $L_1(i, j, k)$ is the transition rate from state (i, j, k) to $(i, j+1, k)$. A NRT data user is admitted into the system as long as there is buffer space upon arrival. Therefore, $L_1(i, j, k)$ can be written as

$$L_1(i,j,k) = \begin{cases} \lambda_1, & \text{if } 0 \le j < N \\ 0, & \text{otherwise} \end{cases} \qquad (4)$$

4) $M_1(i, j, k)$ is the transition rate from state (i, j, k) to $(i, j-1, k)$. Note that if the network reserves Cg channels specifically for RT data traffic, the maximum transition rate from state (i, j, k) to $(i, j-1, k)$ is $(C-Cg)\mu_d$. Therefore, $M_1(i, j, k)$ can be written as

$$M_1(i,j,k) = \begin{cases} j\mu_d, & \text{if } i+j+k \le C \text{ and} \\ & \quad j \le C - Cg \\ (C-Cg)\mu_d, & \text{if } i+j+k \le C \text{ and} \\ & \quad j > C - Cg \\ (C-i-k)\mu_d, & \text{if } i+j+k > C \text{ and} \\ & \quad i+k < C \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

5) $L_2(i, j, k)$ is the transition rate from state (i, j, k) to $(i, j, k+1)$. An RT data user is admitted into the system as long as either there are free channels upon arrival or there are at least one NRT data users in service. In other words, an RT data user is admitted into the system when the total number of voice users and RT data users in the system is less than the total number of channels upon arrival. Therefore, $L_2(i, j, k)$ can be written as

$$L_2(i,j,k) = \begin{cases} \lambda_2, & \text{if } i+k < C \\ 0, & \text{otherwise} \end{cases} \qquad (6)$$

6) $M_2(i, j, k)$ is the transition rate from state (i, j, k) to $(i, j, k-1)$, which can be written as

$$M_2(i,j,k) = \begin{cases} k\mu_d, & \text{if } i+k \le C \\ (C-i)\mu_d, & \text{if } i+k > C \\ 0, & \text{otherwise} \end{cases} \qquad (7)$$

The balance equation for the Markov process is expressed as

$$\begin{aligned} \pi_{i,j,k} \cdot & [L_v(i,j,k) + M_v(i,j,k) + L_1(i,j,k) \\ & + M_1(i,j,k) + L_2(i,j,k) + M_2(i,j,k)] \\ = & \pi_{i-1,j,k} \cdot L_v(i-1,j,k) + \pi_{i+1,j,k} \cdot M_v(i+1,j,k) \\ & + \pi_{i,j-1,k} \cdot L_1(i,j-1,k) + \pi_{i,j+1,k} \cdot M_1(i,j+1,k) \\ & + \pi_{i,j,k-1} \cdot L_2(i,j,k-1) + \pi_{i,j,k+1} \cdot M_2(i,j,k+1) \end{aligned} \qquad (8)$$

By applying the constraint $\sum_S \pi_{i,j,k} = 1$ to the set of balance equations, we can obtain the steady-state probability $\pi_{i,j,k}$ to evaluate the performance metrics of the system.

A voice call will be blocked when the number of voice calls in the system equals to C-Cg upon arrival. Thus, the blocking probability of voice calls, P_{vb}, can be expressed as

$$P_{vb} = \sum_{k=0}^{C} \sum_{j=0}^{N} \pi_{C-Cg,j,k}, \qquad (9)$$

The packet loss probability of RT data traffic, $P_{rt\text{-}loss}$, is the probability that a data packet arrival finds there are no free channels and no NRT data packets in service, and can be obtained as

$$P_{rt-loss} = \sum_{i+k \ge C} \pi_{i,j,k}, \qquad (10)$$

Packet loss of NRT data traffic occurs when the number of NRT data users in the system equals to N upon arrivals. Therefore, the packet loss probability of NRT traffic, $P_{nrt\text{-}loss}$, can be obtained as

$$P_{nrt-loss} = \sum_{k=0}^{C} \sum_{i=0}^{C-Cg} \pi_{i,N,k}, \qquad (11)$$

The mean queueing delay of RT data packets, W_{rt}, and NRT data packets, W_{nrt}, can be obtained respectively by Little's formula,

$$W_{rt} = \frac{1}{\lambda_2(1-P_{rt-loss})} \cdot \sum_S k \cdot \pi_{i,j,k} - \frac{1}{\mu_d}, \qquad (12)$$

$$W_{nrt} = \frac{1}{\lambda_1(1-P_{nrt-loss})} \cdot \sum_S j \cdot \pi_{i,j,k} - \frac{1}{\mu_d}. \qquad (13)$$

4. Performance evaluation

This section investigates the performance of the proposed resource allocation scheme for multimedia traffic. The total number of channels in a cell is set to be 16. The mean voice arrival rate is taken to be 0.08183 calls/sec, and the mean service time of voice calls is 120 seconds. The voice traffic load is chosen to be 9.82 Erlang corresponding to a 2% blocking probability for 16 channels. The arrival rate of data packets is a system parameter and is chosen to set the aggregate data traffic load in the range of 1 to 10 Erlang. The mean service time of data packets is 2 seconds. The maximum number of NRT data users allowed in the system is 20.

Fig. 2 shows the performance of RT data traffic for different RT traffic ratio when Cg=0 and N=20. As can be

seen in Fig. 2(a), the packet loss probability is smaller for smaller RT traffic ratio. This is because small RT traffic ratio implies that a large portion of data traffic in the system is NRT traffic. Therefore, RT data packet arrivals will have more chance to preempt the NRT data packets when there are no channels available, causing the packet loss to decrease. On the contrary, the smaller the RT traffic ratio, the larger the queueing delay of RT data users, which can be seen in Fig. 2(b). In addition, the queueing delay also decreases with increasing data traffic load. The reason is that when the aggregate data traffic load is very low, once an RT data packet is preempted, it must wait for a voice call completion before resuming its service. As the data load increases, there will be some data packets in service, a preempted packet may wait for either a voice call or a data packet completion before resuming its service. Since the mean service time of data packets is much smaller than that of voice calls, the mean queueing delay in the latter case will be smaller than that in the former one. With further increased traffic load, the queue begins to build up and the queueing delay increases with increasing traffic load.

Fig. 3 shows the performance of NRT data traffic for different RT traffic ratio when Cg=0 and N=20. It can be seen in Fig. 3(a) that when the data traffic load is very low, *e.g*., aggregate data traffic load is below 2 Erlang, the packet loss probability for high RT traffic ratio is smaller than that for low RT traffic ratio. The reason is that when the data traffic load is very low, most data packets will be admitted into the system upon arrivals, and preemptions rarely occur. Therefore, the packet loss of NRT data traffic is mainly blocked by its own traffic. Since the number of NRT data packets for high RT traffic ratio is fewer than that for low RT traffic ratio, the probability of buffer overflow will be lower for the former case. However, as the data load increases, there are more RT data users in the system. Preemptions occur more frequently for larger RT traffic ratio, thus NRT data arrivals will be more likely to find its buffer full. Therefore, the packet loss probability is higher for larger RT traffic ratio. On the other hand, the queueing delay of NRT data users increases with both the data traffic load and the RT traffic ratio.

Fig. 4 shows the performance of RT data traffic for different number of reserved channels when r=0.5 and N=20. In the legend, $Cg = 1$ indicates that the network reserves one channel specifically for RT data users. It can be seen that both the packet loss probability and mean queueing delay decrease as the number of reserved channels increases. This is achieved at the expense of increasing the voice blocking probability from 2% for Cg
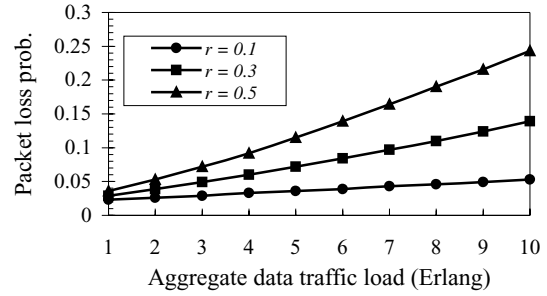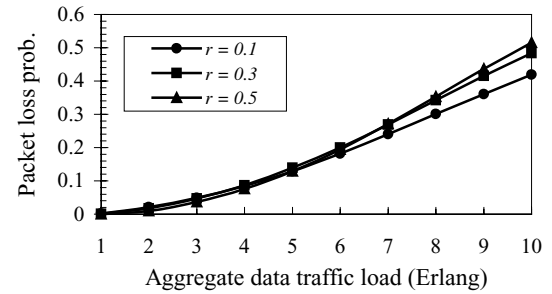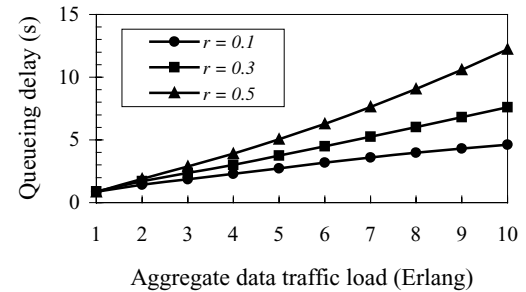

(a) Packet loss probability
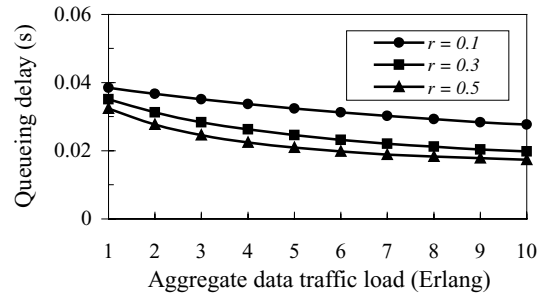
(b) Mean queueing delay

Fig. 2 Performance of RT data traffic for different RT traffic ratio. (Cg=0, N=20)

(a) Packet loss probability

(b) Mean queueing delay

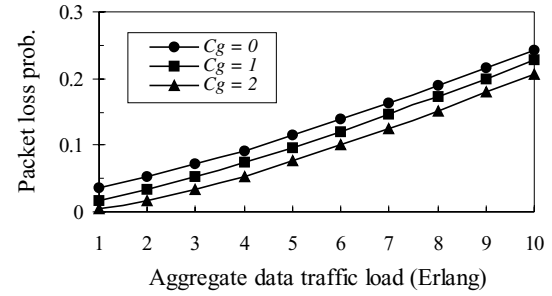Fig. 3 Performance of NRT data traffic for different RT traffic ratio. (Cg=0, N=20)

4

= 0 to 3.3% and 5.2% for Cg = 1 and Cg = 2, respectively. It is interesting to note that in contrast to the case of Cg = 0, the queueing delay increases with increasing data traffic load for Cg > 0 as shown in Fig. 4(b). The reason is that for Cg > 0, if there are preempted RT data packets, there are always RT data users in service. Since the mean service time of data packets is much smaller than that of voice calls, the preempted data packets can quickly obtain services. Therefore, the queueing delay is small at light load. As the traffic load increases, more packets will be preempted by voice calls which find no channels available upon arrivals, causing the queueing delay to increase. No matter what the value of Cg is, the strategy that buffer-only-for-preempted-packets can achieve relatively low queueing delay.

Fig. 5 shows the performance of NRT data traffic for different number of reserved channels when r=0.5 and N=20. Although the difference is negligible, the packet loss probability and mean queueing delay of NRT data users decrease with increasing number of reserved channels. This is because reserving channels for RT data traffic will cause voice calls less likely to obtain services. Since less voice calls are admitted into the network, the queued packets will have more chance to obtain services. Therefore, both the packet loss probability and queueing delay of NRT data traffic decrease.

Figs. 6 and 7 show the impact of different maximum allowed NRT data users on the system performance when r=0.5 and Cg=0. As can be seen in Fig. 6, the performance of RT data users is not affected by the maximum allowed NRT data users in the system. This is because RT data traffic has preemptive priority over NRT data traffic. However, increasing the maximum allowed NRT data users can reduce the packet loss probability of NRT traffic at the expense of increasing the queueing delay as can be seen in Fig. 7. Since large delay is tolerable for NRT traffic, the maximum allowed NRT data users can be set to a high value to reduce the packet loss probability of NRT traffic while keeping the performance of voice traffic and RT data traffic unchanged.

5. Conclusions

In this paper, we analyzed the performance of radio resource allocation for multimedia traffic in cellular networks. The results show that the packet loss probability of RT traffic is smaller for smaller RT traffic ratio. On the contrary, the smaller the RT traffic ratio, the larger the queueing delay. However, at high data traffic load, both the packet loss probability and mean queueing delay of NRT traffic increase with increasing RT traffic ratio. By
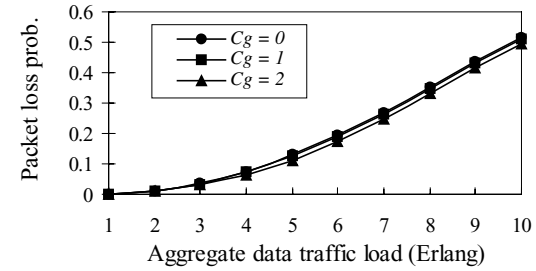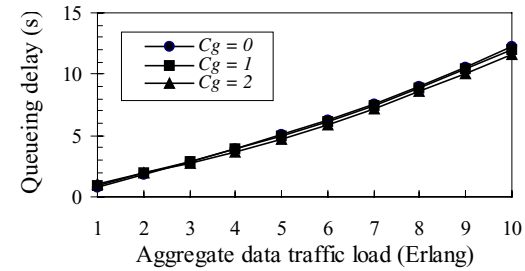

(a) Packet loss probability
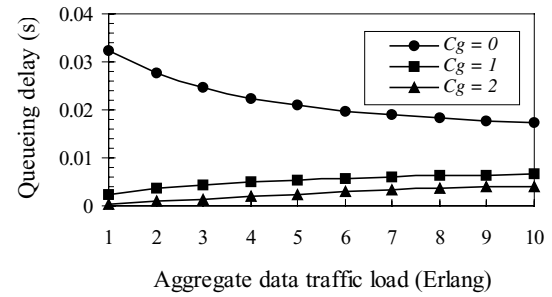
(b) Mean queueing delay

Fig. 4 Performance of RT data traffic for different number of reserved channels. (r = 0.5, N=20)

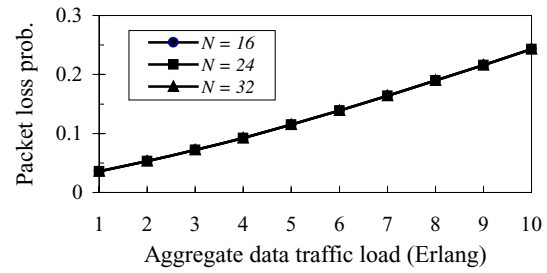(a) Packet loss probability

(b) Mean queueing delay

Fig. 5 Performance of NRT data traffic for different number of reserved channels. (r = 0.5, N=20)

5

reserving channels specifically for RT traffic, the packet loss probability and mean queueing delay can be reduced, meanwhile, the performance of NRT traffic can also be improved. This is achieved at the expense of increasing voice blocking probability.

With preemptive priority, the packet loss probability and mean queueing delay of RT data traffic remain constant irrespective of the maximum number of allowed NRT data users in the system. Moreover, the strategy of buffer-only-for-preempted-packets can achieve relatively low queueing delay and is useful for RT traffic to meet its strict delay requirement.

6. References

[1] C. Oliveria, J. B. Kim, and T. Suda, "An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks," *IEEE JSAC*, vol. 16, no. 6, pp. 858-873, August 1998.

[2] M. Marsan, S. Marano, and C. Mastroianni, and M. Meo, "Performance analysis of cellular mobile communication networks supporting multimedia services," *Mobile Network and Applications*, vol. 5, no. 3, pp. 167-177, March 2000.

[3] S. Kim and P. K. Varshney, "An adaptive bandwidth reservation algorithm for QoS sensitive multimedia cellular network," *IEEE VTC 2002-Fall*, Vancouver, Canada, pp. 1475-1479 , Sept. 2002.

[4] D. C. Lee, S. J. Park, and J. S. Song, "Performance analysis of queueing strategies for multiple priority calls in multiservice personal communications services," *Computer Communications*, vol. 23, pp. 1069-1083, June 2000.

[5] P. Lin and Y. B. Lin, "Channel allocation for GPRS," *IEEE Trans. on Vehicular Techno*., vol. 50, no. 2, pp. 375-387, March 2001.

[6] Wei-Yeh Chen, Jean-Lien C. Wu, and Hung-Huan Liu, "Performance Analysis of Radio Resource Allocation in GSM/GPRS Networks," *IEEE VTC 2002-Fall*, Vancouver, Canada, pp.1461-1465, Sept. 2002.

[7] L. Huang, S. Kumar, and C. C. J. Kuo, "Adaptive resource allocaiton for multimedia QoS management in wireless networks," *IEEE Trans. on Vehicular Techno*., vol. 53, no. 2, pp. 547-558, March 2004.

[8] B. H. Kim, S. H. Shin, and K. S. Kwak, "Performance analysis of multimedia CDMA systems supporting voice, stream data and packet data services," *IEICE Trans. Commun.* vol. E86-B, no. 4, pp. 1224-1233, April 2003.

[9] M. Ermel, K. Begain, T. Muller, J. Schuler, and M. Schweigel, "Performance of GSM networks with general packet radio services," *Performance Evaluation*, vol. 48, pp. 285-310, May 2002.
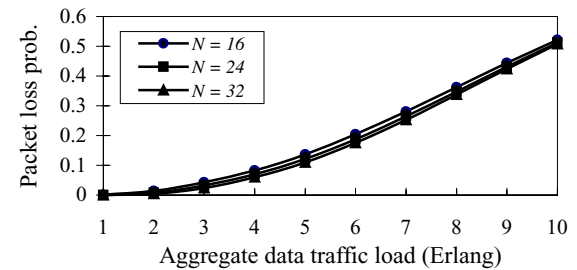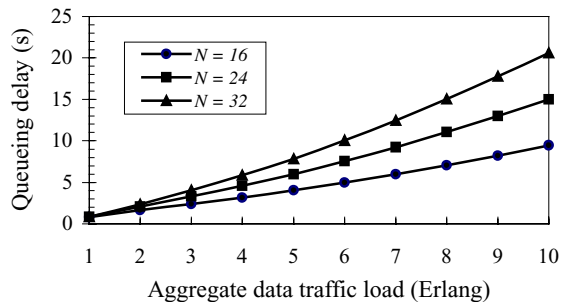
(a) Packet loss probability
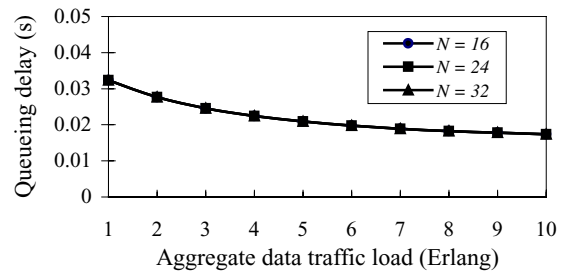
(b) Mean queueing delay

Fig. 6 Performance of RT data traffic for different maximum allowed NRT data users in the system. (r=0.5, Cg=0)

(a) Packet loss probability

(b) Mean queueing delay

Fig. 7 Performance of NRT data traffic for different maximum allowed NRT data users in the system. (r=0.5, Cg=0)