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

WEI-YEH CHEN
Department of Information Management
Southern Taiwan University of Technology
Taiwan, R.O.C
cwycwy@mail.stut.edu.tw

JEAN-LIEN C. WU
Department of Electronic Engineering
National Taiwan University of Science and Technology
Taiwan, R.O.C
jcw@mail.ntust.edu.tw

HAW-YUN SHIN
Department of Information Engineering
Ming Chuan University
Taiwan, R.O.C
hyshin@mcu.edu.tw

Abstract

Future communication systems are expected to provide a broad range of multimedia services with guaranteed quality of service (QoS). Thus, effective management of radio resources is important to enhance the network performance. In this paper, we proposed an analytical model to 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. The strategy of buffer-only-for-preempted-RT-packets can achieve relatively low queueing delay of RT traffic. Moreover, 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.

Keywords: *cellular network; preemption; QoS; resource allocation*

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 radio resources, efficient management of radio resources becomes a key factor in enhancing the network performance [3].

Several resource allocation schemes [4-11] 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. To guarantee the QoS of voice traffic, voice traffic is given preemptive priority in their work. 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. Huang *et al.* [8] proposed to use silent periods of idle channels to provide high throughput and low average transmission delay for data traffic in GSM/GPRS networks. A study on the performance of a system equipped with finite buffers for integrated voice/data services was conducted by Huang *et al.* [9]. In their study, preemption was not considered to maintain the QoS of voice service.

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 [10]. In [11], the authors used the decomposition technique [12] to decompose a two-dimensional Markov chain into two one-dimensional Markov chains to evaluate the mean delay and 95% delay for the GPRS.

In this paper, we proposed an analytical model to analyze the performance of radio resource allocation for multimedia traffic in cellular networks. 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 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. Two buffers are provided respectively for RT traffic and NRT traffic to reduce their loss

probability. 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. In section 4, comparisons of numerical and simulation results are provided. Performance comparisons for different RT traffic ratio and number of reserved channels for RT traffic are also presented. The impact of maximum number of NRT users admitted to the system and different data packet lengths on system performance is also investigated. 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 [13]. The complete partitioning strategy divides the total cell capacity into two portions, one for voice traffic and one for data traffic. On the contrary, with complete sharing strategy the radio resources are completely shared by voice and data traffic. The partial sharing strategy divides the total cell capacity into three portions, one for voice traffic, one for data traffic and the rest are shared by voice and data traffic. The results in [13] 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 arriving NRT data user is dropped. To further improve the data packet loss probability, some channels are reserved specifically for RT data traffic. Fig. 1 depicts the proposed resource allocation model adopted in this work.

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. The idea of using buffer-only- for-preempted-RT-packets comes from [6].

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. Once being accepted, a voice call or a data packet will be allocated with one channel.

3 The Analytical Model

In this section, we will describe the analytical model based on a three-dimensional Markov chain. The analyses will be focused on a single cell in isolation and assume that the network is symmetric and the traffic is homogenous. 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. $p_{i,j,k}$ 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.

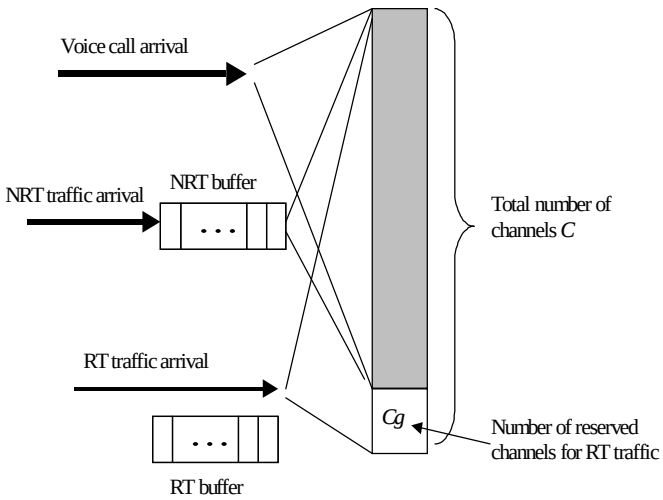


Fig. 1 The resource allocation model

To investigate the performance of radio resource allocation, several assumptions are made in the analytical model. The arrival of voice call requests forms Poisson process with a rate of l_v. The service time of voice calls is assumed to be exponentially distributed with a mean of $1/m_v$. The arrivals of RT and NRT data packets are assumed to be Poisson processes with a rate of l_2 and l_1, respectively, and $l_d = l_1 + l_2$ is the aggregate data arrival rate. The ratio $r = l_2 / l_d$ 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 $1/m_2$ and $1/m_1$, respectively. For simplicity, we let $m_1 = m_2 = m_d$ in the analysis. Fig. 2 shows the state transition diagram. Let *S* be the set of feasible states,

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

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

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

2) $M_v(i, j, k)$ is the transition rate from state (i, j, k) to $(i\text{-}1, j, k)$, *i.e.*, one voice user completes service while j NRT data users and k RT data users are in the system. Therefore, $M_v(i, j, k)$ can be written as

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

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

$$L_1(i, j, k) = \begin{cases} l_1, & \text{if } 0 \le j < N \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

4) $M_1(i, j, k)$ is the transition rate from state (i, j, k) to $(i, j\text{-}1, k)$, *i.e.*, one NRT data user completes service while i voice users and k RT data users are in the system. Note that if the network reserves *Cg* channels specifically for RT data traffic, the maximum transition rate from state (i, j, k) to $(i, j\text{-}1, k)$ is $(C–Cg)\mu_d$. Therefore, $M_1(i, j, k)$ can be written as

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

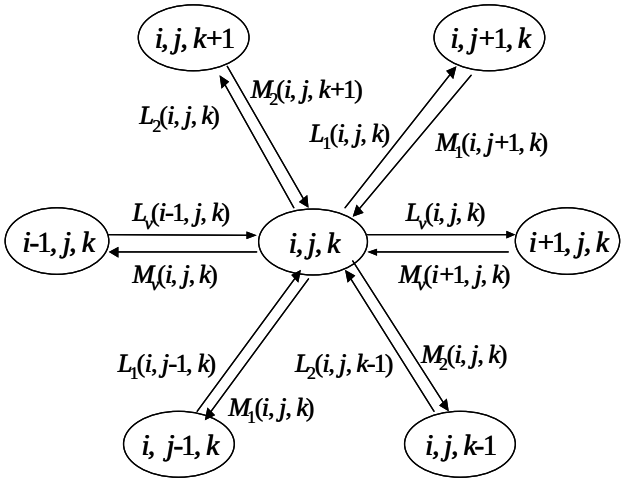


Fig. 2 The state transition diagram

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

$$L_2(i, j, k) = \begin{cases} l_2, & \text{if } i + k < C \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

6) $M_2(i, j, k)$ is the transition rate from state (i, j, k) to $(i, j, k\text{-}1)$, *i.e.*, one RT data user completes service while i voice users and j NRT data users are in the system. Therefore, $M_2(i, j, k)$ can be written as

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

The balance equation for the Markov process is expressed as

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

By applying the constraint $\sum_S p_{i,j,k} = 1$ to the set of balance equations, we can obtain the steady-state probability $p_{i,j,k}$ 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

$$
P_{vb} = \sum_{k=0}^{C} \sum_{j=0}^{N} p_{C-Cg,j,k}, \tag{9}
$$

Since voice traffic has preemptive priority over data traffic, the voice blocking probability will remain unchanged under certain voice traffic load and is not affected by the data traffic load. Therefore, we can use this property to verify the accuracy of the numerical results.

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

$$
P_{rt-loss} = \sum_{i+k \geq C} p_{i,j,k}, \tag{10}
$$

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

$$
P_{nrt-loss} = \sum_{k=0}^{C} \sum_{i=0}^{C-Cg} p_{i,N,k}, \tag{11}
$$

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

$$
W_{rt} = \frac{1}{l_2(1-P_{rt-loss})} \cdot \sum_S k \cdot p_{i,j,k} - \frac{1}{m_d}, \tag{12}
$$

$$
W_{nrt} = \frac{1}{l_1(1-P_{nrt-loss})} \cdot \sum_S j \cdot p_{i,j,k} - \frac{1}{m_d}. \tag{13}
$$

The mean queueing delay of RT data packets can be also calculated by

$$
W_{rt} = \frac{1}{l_2(1-P_{rt-loss})} \cdot \sum_{\substack{i+k>C \\ \& \, k \geq 1}} (i+k-C) \cdot p_{i,j,k} \tag{14}
$$

The condition $i+k > C$ in Eq. (14) implies that there are RT data packets queued in the buffer. In addition to simulation experiments, Eqs. (12) and (14) can be also used to validate the accuracy of numerical results.

4 Numerical and Simulation Results

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. Table 1 provides the parameters used for numerical calculations.

We will evaluate the performance under various RT traffic ratio and different number of reserved channels for RT data traffic. The impact of different number of maximum allowed NRT data users in the system and various mean packet length of data packets on the system performance are also investigated.

To validate the numerical results, an event-driven simulator is developed. Table 2 lists the comparisons of performance metrics for both analytical and simulation models. In the table, the errors between numerical and simulation results are less than 4%. The table indicates that the numerical results match closely with the simulation results. Similar results for different system parameters are obtained and are not presented.

Table 1. Parameters used for numerical calculations

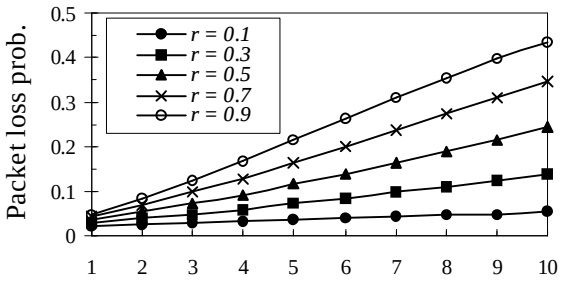
Total number of channel, *C*	16
Number of reserved channel, *Cg*	0, 1, 2
Mean arrival rate of voice calls, l_v	0.08183 calls/s
Mean service time of voice calls, $1/m_v$	120 s
Mean arrival rate of data packets, l_d	0.5 ~ 5 packets/s
Mean service time of data packets, $1/m_d$	2 s
RT traffic ratio, *r*	0.1 ~ 0.9
Max. allowed NRT users, *N*	20

Table 2. Comparison of the analytical and the simulation results. (l_v =0.08183 calls/s, $1/m_v$ =120 s, $1/m_d$ =2 s, *Cg*= 0, *r*= 0.5, *N* =20)

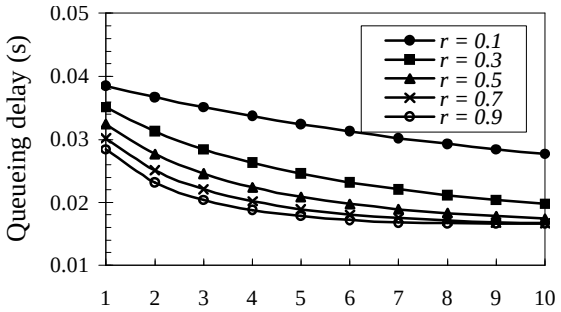
	l_d	Numerical	Simulation	Error
$P_{rt\text{-}loss}$	2 packets/sec	0.092	0.093	1.1%
$P_{nrt\text{-}loss}$		0.076	0.075	1.31%
W_{rt}		0.0224	0.0217	3.12%
W_{nrt}		3.9	3.88	0.51%
$P_{rt\text{-}loss}$	5 packets/sec	0.243	0.246	1.23%
$P_{nrt\text{-}loss}$		0.515	0.514	0.19%
W_{rt}		0.0174	0.0168	3.45%
W_{nrt}		12.21	12.16	0.41%

Fig. 3 shows the performance of RT data traffic for different RT traffic ratio when *Cg*=0 and *N*=20. As can be seen in Fig. 3(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. 3(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. This characteristic is similar to the result of Fig. 9(a) in [6]. The same reason can also be applied to explain the larger queueing delay for smaller RT traffic ratio.

Fig. 4 shows the performance of NRT data traffic for different RT traffic ratio when *Cg*=0 and *N*=20. It can be seen in Fig. 4(a) that the packet loss probability of NRT traffic first increases with increasing RT traffic ratio, and then decreases with increasing RT traffic ratio. The reason is that when the RT traffic ratio is small, *e.g.*, *r* < 0.5, the number

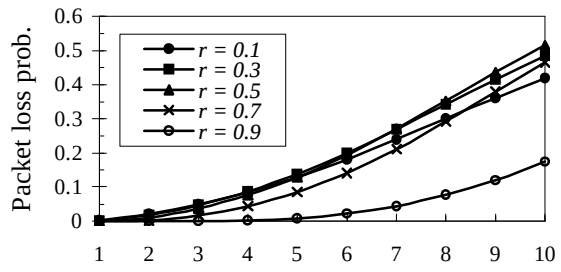


(a) Packet loss probability

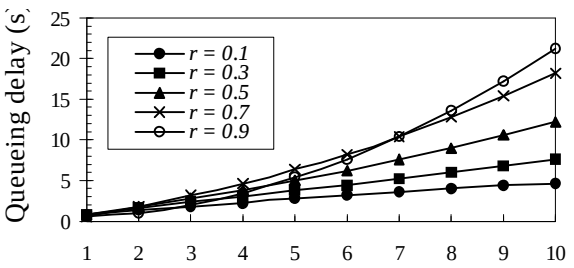


(b) Mean queueing delay

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



(a) Packet loss probability



(b) Mean queueing delay

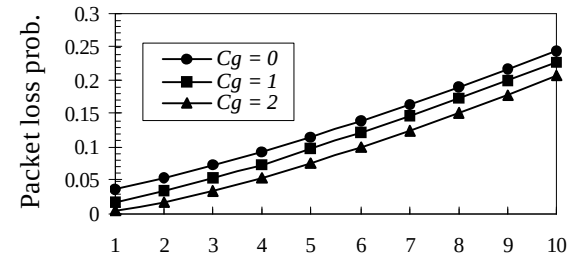
Fig. 4 Performance of NRT data traffic for different RT traffic ratio. (*Cg*=0, *N*=20)

of RT data users in the system is fewer than that of NRT data users. Therefore, as the RT traffic ratio increases, the RT data arrivals will have more chance to preempt the NRT data users when there are no channels available, causing the probability of buffer overflow to increase. Thus, the packet loss probability of NRT traffic increases with increasing RT traffic ratio. However, when the RT traffic ratio exceeds 0.5, the number of NRT data users in the system is less than that of RT data users. Thus, the NRT data arrival will be less likely to find its buffer full. Therefore, the packet loss probability is lower 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, as can be seen in Fig. 4(b).

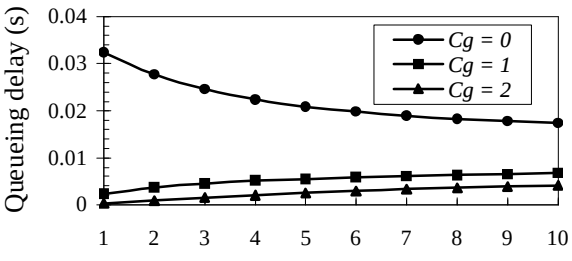
Fig. 5 shows the performance of RT data traffic for different number of reserved channels when r=0.5 and N=20. In the legend, Cg = 1 indicates that the network reserves one channel specifically for RT data users. It can be seen that both the packet loss probability and mean queueing delay decrease as the number of reserved channels increases. This is achieved at the expense of increasing the voice blocking probability from 2% for Cg = 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. 5(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. 6 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. 7 and 8 show the impact of different maximum allowed NRT data users on the system

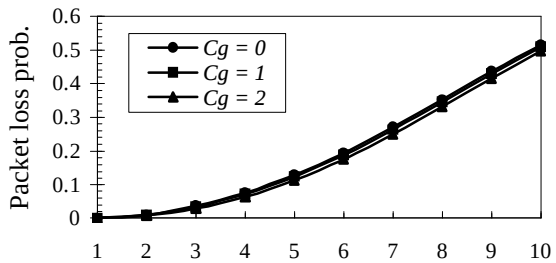


(a) Packet loss probability

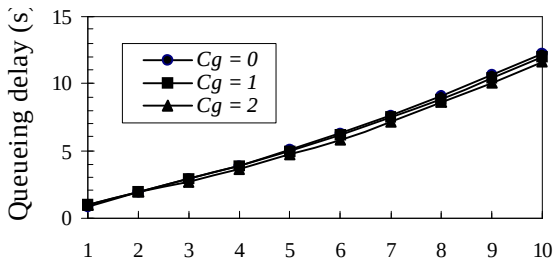


(b) Mean queueing delay

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



(a) Packet loss probability

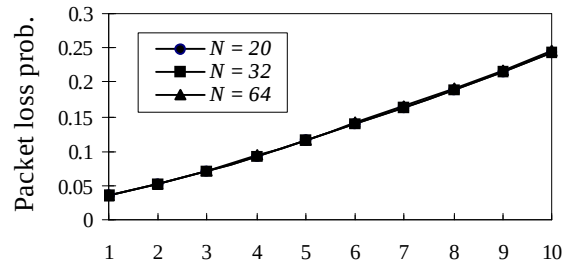


(b) Mean queueing delay

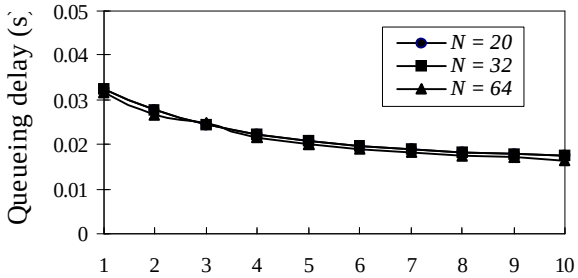
Fig. 6 Performance of NRT data traffic for different number of reserved channels. (r = 0.5, N=20)

performance when r=0.5 and Cg=0. Note that the results of N=64 are obtained by simulation. For large value of N, the number of states in the analytical model will be enormous, which makes the queueing analysis very difficult. As can be seen in Fig. 7, 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. 8. 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.

Figs. 9 and 10 show the impact of different mean packet length of RT traffic on the system performance when r=0.5, Cg=0 and N=20. len = 1 means the mean packet length is one unit long, correspond to the mean service time is 1 second in the analytical model. For comparative purpose, the traffic load of these two scenarios are kept the same, *i.e.*, the mean arrival rate of mean packet length being 1 is twice than that of mean packet length being 2. It can be seen that all performance metrics except the queueing delay of RT traffic remain unchanged irrespective to the mean packet length of RT traffic. Smaller queueing delay for shorter packet length is due to that shorter packet length will occupy the channels shorter than longer packet length, resulting in data packets waiting in queue being able to obtain services more quickly. Figs. 11 and 12 show similar characteristics for different mean packet length of NRT traffic.

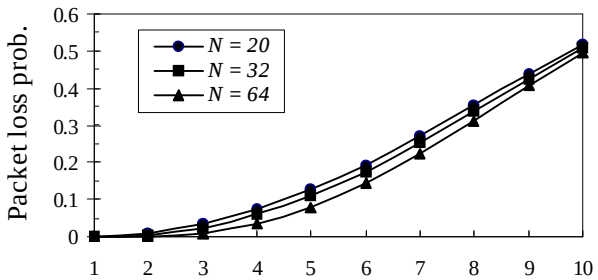


(a) Packet loss probability

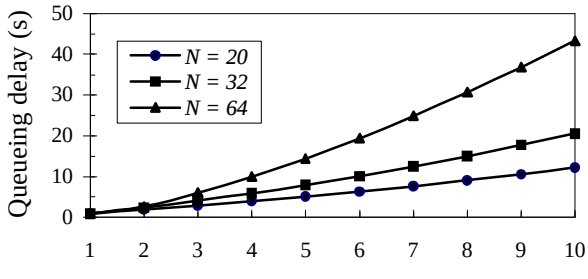


(b) Mean queueing delay

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



(a) Packet loss probability



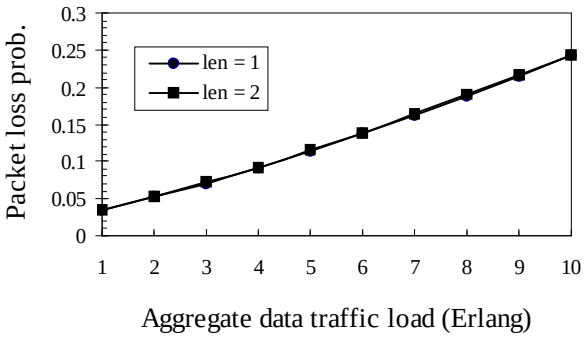
(b) Mean queueing delay

Fig. 8 Performance of NRT data traffic for different maximum allowed NRT data users in the system. (r=0.5, Cg=0)

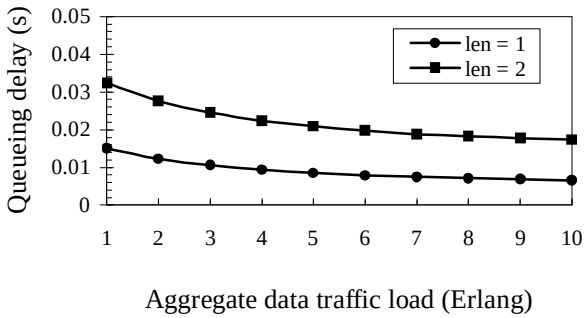
5 Conclusion

In this paper, we proposed an analytical model to analyze the performance of radio resource allocation for multimedia traffic in cellular networks. To achieve better channel utilization, the radio resources are completely shared by voice, RT and NRT data traffic. To guarantee the voice performance, voice calls can preempt data packets and RT data traffic can preempt NRT data traffic. The results are validated by simulation experiments.

The results show that with preemptive priority, the packet loss probability and mean queueing delay of RT data traffic remain constant irrespective to the maximum number of allowed NRT data users in the system. By reserving channels specifically for RT traffic, the packet loss probability and mean queueing delay can be reduced, meanwhile, the performance of

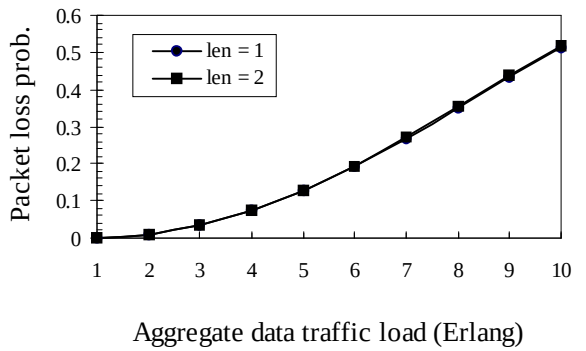


(a) Packet loss probability

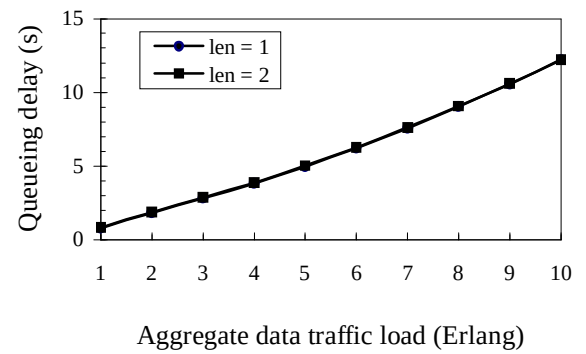


(b) Mean queueing delay

Fig. 9 Performance of RT data traffic for different mean packet length of RT traffic. (r=0.5, Cg=0 and N=20)

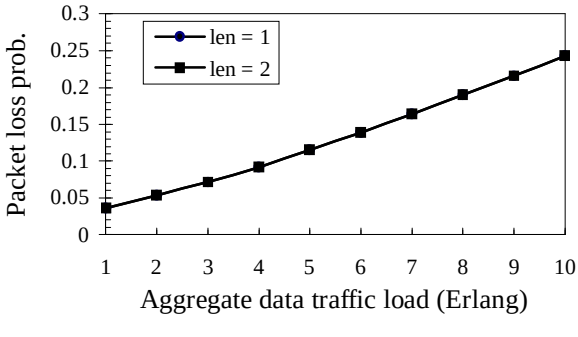


(a) Packet loss probability

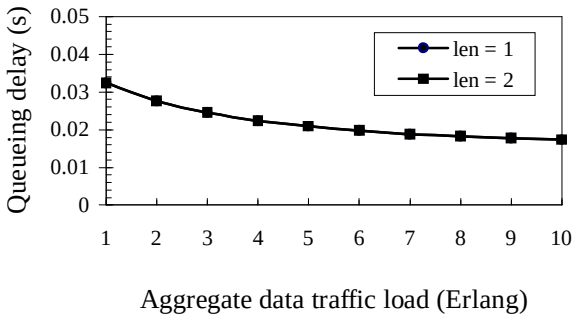


(b) Mean queueing delay

Fig. 10 Performance of NRT data traffic for different mean packet length of RT traffic. (r=0.5, Cg=0 and N=20)

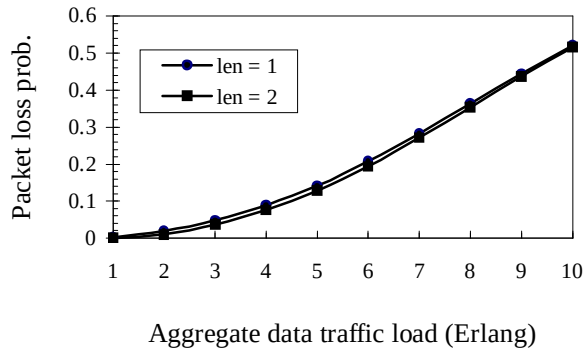


(a) Packet loss probability

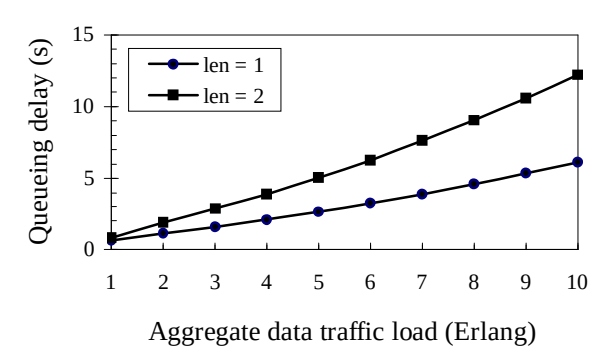


(b) Mean queueing delay

Fig. 11 Performance of RT data traffic for different mean packet length of NRT traffic. (r=0.5, Cg=0 and N=20)



(a) Packet loss probability



(b) Mean queueing delay

Fig. 12 Performance of NRT data traffic for different mean packet length of NRT traffic. (r=0.5, Cg=0 and N=20)

NRT traffic can also be improved. This is achieved at the expense of increasing voice blocking probability. 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. 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.

The results also show that 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. In addition, under the same data traffic load condition, shorter packet length gives shorter mean queueing delay while maintaining the packet loss probability of RT traffic and NRT traffic unchanged.

Acknowledgement

The authors thank the anonymous reviewers for their careful reviews and useful suggestions. This work was supported in part by the National Science of Council under Contract NSC 95-2221-E-218-014 and NSC 95-2221-E-130-004.

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, August 1998, pp. 858-873.

[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, March 2000, pp. 167-177.

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

[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, June 2000, pp. 1069-1083.

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

[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, Sept. 2002, pp.1461-1465.

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

[8] J. H. Huang, S. L. Su, and J. H. Chen, "Design and performance analysis for data transmission in GSM/GPRS system with voice activity detection," *IEEE Trans. on Vehicular Techno.*, vol. 51, no. 4, July 2002, pp. 648-656.

[9] Y. R. Huang, Y. B. Lin, and J. M. Ho, "Performance analysis for voice/data integration on a finite buffer mobile system," *IEEE Trans. on Vehicular Techno*., vol. 49, no. 2, March 2000, pp. 367-378.

[10] 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, April 2003, pp. 1224-1233.

[11] H. H. Liu, J. L. C. Wu and W. C. Hsieh, "Delay analysis of integrated voice and data service for GPRS," *IEEE Communications Letters*, vol. 6, no. 8, Aug. 2002, pp. 319-321.

[12] S. Ghani and M. Schwartz, "A decomposition approximation for the analysis of voice/data integration," *IEEE Trans. on Commun*., vol. 42, no. 7, July 1994, pp. 2441-2451.

[13] M. Ermel, K. Begain, T. Muller, J. Schuler, and M. Schweigel, "Performance of GSM networks with general packet radio services," *Performance Evaluation*, vol. 48, May 2002, pp. 285-310.