Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT•

Computer Communications 26 (2003) 1931–1943

computer communications

www.elsevier.com/locate/comcom

Real-time packet scheduling in next generation radio access system

Chiang-Shiang Wan[a,c,*], Wei-Kuan Shih[b], Ruei-Chuan Chang[a]

[a]Department of Computer and Information Science, National Chiao Tung University, Hsinchu 300, Taiwan, ROC
[b]Department of Computer Science, National Tsing Hua University, Hsinchu 300, Taiwan, ROC
[c]Wireless Communication Technology Lab., Chunghwa Telecom Laboratories, Yang-Mei, Taoyuan 326, Taiwan, ROC

Received 21 January 2003; accepted 21 January 2003

Abstract

All IP-based radio access network enable packet-oriented connections to offer real-time applications. Radio access networks must provide efficient radio management for data session establishments because of scarce radio resources. In this paper, we present three real-time scheduling algorithms to support quality-of-service at IP-based radio access networks. The real-time generic scheduling (RTGS) algorithm applies the functionalities of the radio management framework to establish new data sessions for real-time service requests. The real-time bandwidth scheduling (RTBS) algorithm implements the early-deadline-first scheme to do the schedulability analysis and to schedule the data sessions to reduce power consumption. The RTBS algorithm can decrease the power consumption more than can RTGS. Based on the RTBS mechanism, we design the real-time code scheduling (RTCS) algorithm. In this algorithm, we apply the dynamic code assignment scheme to increase the probability of schedulable sessions and improve the radio resource utilization. Experimental results show that RTCS outperforms RTBS and, in turn, RTBS outperforms RTGS.

Keywords: Real-time scheduling; Radio access network; Quality of service; Channelization code; Early-deadline-first; Dynamic code assignment

1. Introduction

The mobile communication technologies advance rapidly. Although current mobile systems provide only speech and low-rate data services, the next generation mobile system will support high quality voice, high-rate data and real-time multimedia services. The Universal Mobile Telecommunications System (UMTS) proposed by the 3rd Generation Partnership Project (3GPP) is the most popular 3G mobile system, and complies with IMT-2000 standards. The Wideband Code Division Multiple Access (WCDMA) is an air interface for the UMTS, which adopts code division technology to increase spectral efficiency and enhance the system's capacity. Specifically, WCDMA can provide maximum 2 Mbps bearer when the subscriber is pedestrian at a speed of less 10 km/h [1,2]. Thus, using various terminals, for example, handhelds, PDAs and laptops, UMTS can provide high quality multimedia services for mobile users.

In the standard 3GPP Release 5, UMTS has emerged as IP multimedia architecture. The architecture enables session initial protocol-based call control of packet-oriented connections, to offer voice over IP (VoIP) and video telephony services. IP-based network entities integrated voice and data on unified IP backbone, which can increase the resource utilization over existing mobile networks. WCDMA radio access network must manipulate the delay-sensitive real-time packets to provide IP multimedia service. The resource management of the radio access network will become the most important issue for real-time packets quality-of-service (QoS) provisioning because of the scarcity of radio frequencies. Several researchers have recently proposed the concept of IP-based radio access networks [3,4] to utilize efficiently radio resources. In an IP-based radio access network, the real-time traffics are transmitted as packet-oriented connections over an air interface. Therefore, applying the radio access architecture will benefit radio resource utilization for multimedia services.

Several resource management polices have been proposed to improve radio resource utilization [5–8,14]. Jorguseski

* Corresponding author. Address: Wireless Communication Technology Lab., Chunghwa Telecom Laboratories, Yang-Mei, Taoyuan 326, Taiwan, ROC. Tel.: +886-3-4244746; fax: +886-3-4244920.
E-mail address: cswan@cht.com.tw (C.-S. Wan).

0140-3664/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0140-3664(03)00158-0

et al. [5] proposed a resource allocation algorithm, based on the CDMA capacity analytical models, which manipulates the system's states of power allocation. The allocation algorithm measures the radio link and calculates the received signal-to-interference ratio (SIR) for all existing services. The authors monitored the session bit rate to determine the transmission power that could fit the E_b/N_0 target ranges, by using the power control process iteratively. Thus, their algorithm reduces the number of blocked sessions and drops the probability for various traffic requirements.

Gürbüz and Owen [6] proposed a dynamic scheduling algorithm for radio resources, which supported the real-time traffic for QoS provisioning, with the help of power assignment and code hopping. A resource scheduler was implemented in the base station. It collects requests from all mobile users. The real-time traffic in a radio access network have priority serving mechanism and have a higher bit rate than can non-real-time traffic. Fitzek et al. [7] used multicode based link layer transmission strategies. Multiple channels were used to achieve a high spectral efficiency, only if the wireless link was less error-prone. However, using multiple channels leads to the degradation of the SIR and it results in raised power consumption.

Das et al. [8] presented a framework to support traffic services for various QoS requirements. The basic concept of this scheme was to treat real-time and non-real-time applications differentially on a wireless physical link layer. A bandwidth compaction method, that combined call admission control, bandwidth reservation and degradation, was employed to improve spectral utilization. However, the main drawback of bandwidth compaction is that it is too expensive and time consuming due to the large number of channel reassignments.

Most literature solves the congestion control of the switching element and allocated resources for the multiprocessor by using real-time packet scheduling algorithms [9–11]. The contribution of this paper is to apply real-time packet scheduling schemes in the IP-based radio access network for QoS provisioning. An early-deadline-first (EDF) scheduler in the access network of WCDMA is developed to manage channelization code allocation and thus to fulfill a data session's requirement. Notably, the data session will be assigned a new channelization code and the power resource is consumed only if the session is not a feasible schedule by the scheduler. In addition, applying the dynamic code assignment (DCA) scheme enables the scheduler to reduce the number of unfeasible sessions. Simulation experiments show a significant improvement in the radio resource utilization.

The rest of the paper is organized as follows. Section 2 defines a generic framework for manipulating real-time packets for QoS provisioning, within the IP-based radio access network. Based on the WCDMA capacity analysis model [12,13], we discuss issues of power saving while establishing and transmitting various bit rates in a session. Section 3 presents three real-time algorithms for packet scheduling. The operations of each algorithm, such as power allocation and real-time packet scheduling to guarantee QoS provisioning, are discussed in detail. Section 4 presents simulation results to demonstrate the performance of our packet scheduling algorithms. Concluding remarks are given in Section 5.

2. Real-time packet access

In a WCDMA system, the available channelization codes and base station power are the most important resources. The estimated system capacity will dominate both the cell interference and base station power amplifier since a WCDMA system is generally planned for use in dense urban and urban areas. In this section, a generic management framework is described within the IP-based radio access network to schedule real-time packet traffic. The power consumption and characteristics of each mobile user with varying bit rates within a base station is discussed based on a WCDMA capacity analysis model.

2.1. A generic framework for real-time packet QoS provisioning

We consider the implementation of management functions for real-time packet QoS provisioning. Fig. 1 shows the management framework implemented at IP radio access network, which contains three entities—Admission Control (AC), Load Control (LC) and Packet Scheduler (PS). The purpose of AC is to decide whether a new real-time request can be accepted into the access system. The AC uses

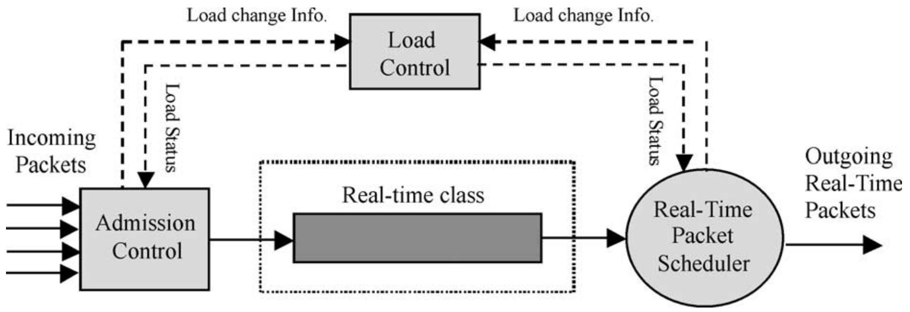

Fig. 1. A generic framework for real-time traffic in radio access network.

the downlink transmission power information that was analyzed by LC, to make this decision for assigned real-time bearers. The resource allocation is allowed if the estimated power consumption does not exceed a target threshold. The main tasks of LC are to measure the uplink interference and downlink transmission power periodically. In this way, LC can prevent overload of the radio system. If an overload situation occurs, the LC will execute the overload control that invokes either AC to drop a call or PS to decrease the transmitted bit rate for reducing the power consumption. Moreover, the mission of LC is to provide the current load status, and to interact with AC to make the decision concerning connection administration. The function of PS is to schedule the allocated radio frequency resource and to ensure availability of air interface capacity for all real-time packet traffic. The PS estimates the load change information and interacts with LC to notify of the current load status.

Let tx_{max} be the maximum transmission power of a base station; tx_{drop} is the threshold for blocked real-time traffic, and tx_{admit} is the threshold for admitting new real-time traffic. Notably, within a radio access network, the parameters are $tx_{\text{max}} > tx_{\text{drop}} > tx_{\text{admit}}$. Each entity of the management framework performs the following activities to satisfy a real-time packet requirement (Fig. 1).

1. In the beginning, AC decides whether the allocated power of a base station is larger than tx_{admit}. If the allocated power is smaller than the value of tx_{admit}, the AC allows establishment of the real-time data session and then invokes PS to allocate radio resources to the data session. LC will perform the estimation of load change. However, if the allocated power is larger than tx_{admit}, AC rejects the request.
2. Suppose that a base station consumes much power to serve real-time data sessions. The pre-allocated power should be increased to compensate the power requirement of mobile user increment and degeneration of circumstances. When the power consumption exceeds tx_{admit}, LC must ask PS to execute the real-time scheduling algorithm to reschedule the traffic.
3. If the interference is still increasing and it exceeds tx_{drop}, LC notifies PS to drop real-time traffic. This approach prevents the danger that the power consumption is over tx_{max} and avoids the crashing of the radio access network.

2.2. Simulation results of the power consumption

In a WCDMA system, many subscribers simultaneously use the same frequency. The WCDMA uses channelization codes to spread information for each mobile user based on the orthogonal variable spreading factor (OVSF) technique [15]. Different users employ different channelization codes for packet transmission. The channelization code is assigned according to the user-required bit rate and the preset service QoS. The simultaneous use of the same frequency by many users makes interference between the users likely. In the case of much interference, the base station must increase power of transmission to ensure that mobile users receive data correctly.

An air interface model is applied to analyze the interference of WCDMA and illustrate the power consumption. Based on the capacity analysis model [12], the base station output power can be written as

$$P_{\text{tot}} = \frac{[1 + nr(1-\beta)] + P_{\text{SCH}} + P_{\text{CCH}} + \sum_{k=1}^{n} r\frac{N}{G_k}}{1 - nr(\beta + \lambda)} \tag{1}$$

Assume that there are n active mobile users served by this base station. P_{SCH} is the transmitting power of the non-orthogonal synchronization channel; P_{CCH} is the transmitting power of the orthogonal common control channel, and N is a floor noise parameter. Suppose that the power transmitted of the dedicated channels and common channels of a base station has a β percent loss of their orthogonality, and λ is the rate of power reception from intra-cell to inter-cell. G_k denotes the path gain from the base station to mobile user k, and r denotes the SIR.

Numerical results concerning adding mobile users and supporting various bit rates, and evaluating the power consumption of the base station in the downlink are presented. Consider r is given from E_b/N_0

$$r = \frac{E_b}{N_0} - 10\log\left(\frac{W}{R_b}\right) \tag{2}$$

where E_b/N_0 denotes the ratio of bit energy to interference noise density; W/R_b denotes the spread gain that is spread factor; W is 3.84 Mcps chip rate, and R_b is bit rate. Table 1 shows E_b/N_0 with different bit rates for a mobile speed of 3 km/h and an FER target of 10% [2]. The values of r are calculated from Eq. (2).

Evaluating the power consumption of the base station, the values of the corresponding parameters are defined as [12]

n	1
P_{SCH}	0.2 W
P_{CCH}	3.8 W
N	− 99 dBm

Table 1
Values of E_b/N_0 and r with different bit rates

Bit rate (Kbps)	8	16	32	64	128	256	512
E_b/N_0 (dB)	4.2	3.2	2.8	2.4	2.3	2.2	2.1
r (SIR)	0.0103	0.0163	0.0301	0.0543	0.1061	0.2074	0.4055

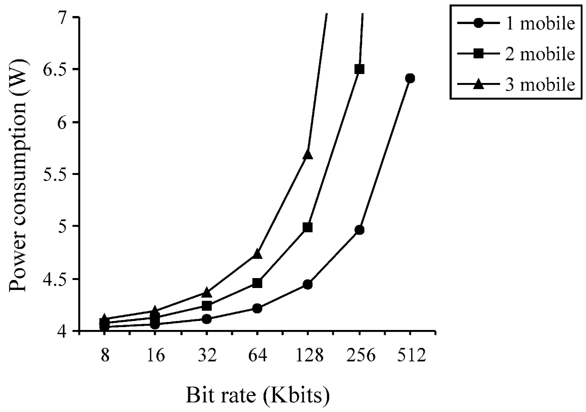

Fig. 2. Power consumption of mobile with different bit rate.

β	0.06
λ	0.84
G_k	0.2, where $k = 1, \ldots, n$

Fig. 2 illustrates the simulation results of the base station power consumption for various input parameters. In Fig. 2, the power consumption of the base station is increased with the allocated bit rate and the number of active mobiles. For example, the power consumption is considered when $n = 1$ and 2, and the allocated bit rate is $R = 16$ Kbps, and when the power consumption is $P_{tot} = 4.06$ and 4.13 W, respectively. The allocated bit rate is $R = 8$ and 512 K for one mobile user, and the required power is $P_{tot} = 4.04$ and 6.42 W, respectively. Hence, the base station power consumption can decrease by reducing the number of active mobiles and the bit rate allocation in a cell.

Fig. 3 depicts the variation of power consumption with number of active mobile users when the system allocated bandwidth is constant. For a total allocated bandwidth in a cell of 1 Mbps, more active mobile users consume more base station power. Fig. 3 relates to an eight-users mobile system, where all mobile users are assumed to use equal bandwidth (and each mobile user has a 128K bit rate). Thus, the base station transmitting power is 17.64 W. Similarly, if each cell includes only two mobile users and each user is allocated a 512K bit rate, the base station only requires 15.37 W. Moreover, via the previous analysis, preventing the simultaneous use of many active mobiles within a cell is advantageous. The WCDMA system is interference limited; that is, lower interference and lower power consumption allow higher capacity.

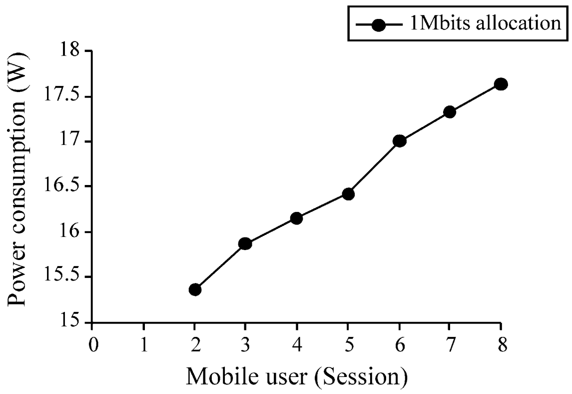

Fig. 3. Power consumption of constant bandwidth with various data sessions.

3. Real-time scheduling schemes

Three algorithms, which minimize the power consumption of the base station for real-time packet transmission, are proposed. The proposed algorithms are implemented on the management framework described in Section 2.1. For real-time packet transmission, these algorithms reduce power consumption using packet scheduling strategies based on the capacity analysis model. In addition, these algorithms can prevent the simultaneous transmission of large packets (high packet activity at the same time). Our algorithms are based on the tolerance of service delay and can decrease simultaneously active mobile users.

Recall that the real-time packet services are periodic multimedia traffic, such as VoIP, video conferencing and video streaming. In a mobile system, the following hypotheses are assumed to apply to the real-time packet services.

- For each service request, the mobile system creates a real-time data session and a QoS profile according to the service class.
- Each data session contains a periodic task with multiple instances. The instances occur regularly at a constant rate. An instance can be activated for transmission when it arrives. The period of an instance is the time interval between two active instances.
- All instances of a data session have the same processing time. The mobile system defines a deadline for each instance according to the QoS profile. The relative deadline is equal to the period of the instance.
- Each active instance must be assigned an appropriate channelization code that depends on the required bit rate and the maximum bit rate of the QoS profile.

3.1. Real-time generic scheduling

The real-time generic scheduling (RTGS) algorithm applies the functionalities AC, PS and LC for real-time packet QoS provisioning. Parameter tx_{admit} determines the power threshold for newly admitted real-time data sessions, based on the capacity analysis model of Eq. (1). If a data session request arrives, the management framework defines the instance's period, processing time, required bit rate and allowed maximum bit rate, according to the QoS profile of service. The RTGS algorithm works as follows.

```
Algorithm RTGS
  Initiate the setting of parameter tx_admit, generate FIFO prioritized queue;
  While (an event occurs)
     Case 1 : A session request arrivals
        AC makes power reservation;
        LC estimates BTS's power consumption P_tot;     /* By Eq. (1) */
        If (P_tot < P_admit) then
           Create new real-time data session;
           Assign channelization code and power resource;
           Update system load status;
        Else
           Reject real-time data session;
     Case 2 : An instance is ready
           Execute the instance;
     Case 3 : An instance is finished or its deadline is reached
        Release allocated channelization code and power resource;
        Update system load;
End Algorithm
```

Fig. 4. Pseudo code for the RTGS algorithm.

1. In the beginning, the PS maintains a prioritized queue in which instances of data sessions are ordered on a first-in-first-out basis.
2. When a data session request arrives, the AC reserves power according to the required bit rate. The LC estimates the power consumption using the capacity analysis model.
3. The AC decides whether the power, P_{tot}, exceeds the threshold, tx_{admit}, or not. If the power, P_{tot}, is smaller than the threshold, tx_{admit}, a new data session is admitted. Otherwise, the AC rejects the request.
4. If a new data session is established, the PS puts the session's instance into the queue, and schedules the instances of the queue for execution. The PS, based on the required bit rate, assigns the channelization code and required power for the instance. After the power allocation is finished, the PS asks the LC to change the system's load.
5. The PS terminates an instance when it is finished or when its deadline is reached. When an instance is terminated, the PS removes this instance from the queue and releases the occupied channelization code and the allocated power. Finally, the PS asks the LC to change the system's load.

Table 2
Selected input parameters

Session (S_i)	Arrival time (A_i)	Period (E_i)	Process time (P_i)	Required rate, Kbps (R_i)	Maximum rate, Kbps (R_{max})
S_1	1	6	4	8	16
S_2	2	8	4	16	32
S_3	4	8	4	8	16
S_4	5	6	4	16	32

The pseudo code of the RTGS algorithm is shown in Fig. 4. A scheduling decision is made when any of the following events occur—a session request arrives, an instance is ready, an instance finishes or an instance's deadline is reached.

Table 2 depicts a service profile to illustrate the RTGS algorithm. This table includes four data sessions $S_1 - S_4$. For each session S_i, A_i denotes the arrival time. T_{ij} denotes the jth instance. A_{ij} denotes the arrival time of instance T_{ij}. E_i denotes the arrival period. The instance's relative deadline D_{ij} is equal to E_i. P_i denotes the instance processing time. R_i and R_{max} denote the required bit rate and admitted maximum bit rate, respectively.

Fig. 5 illustrates a feasible schedule obtained from the RTGS algorithm. The figure specifies the parameters of the instances for each session, and the timing diagram shows the schedule of the sessions at time $t = 24$. Assume that the maximum power of the base station is $tx_{max} = 20$ W and the threshold for a newly created real-time data session is $tx_{admit} = 18$ W. In the beginning, instance T_{11} of S_1 arrives at time $t = 1$. The required bit rate R_1 of S_1 is 8K. The LC calculates the power consumption P_1 of the base station, based on the capacity analysis model of Eq. (1), to satisfy S_1 establishment. The PS assigns a channelization code of which bit rate is 8K for S_1. Since $P_1 = 4.04$ W which is less than 18 W. Additionally, the PS schedules instance T_{11} in the interval (1,5), instance T_{12} in the interval (7,11) and instance T_{13} in the interval (13,17). Notably, T_{11}, T_{12} and T_{13} are schedulable before their respective deadlines $D_{11} = 7$, $D_{12} = 13$ and $D_{13} = 19$.

Similarly, the data sessions S_2, S_3 and S_4 have arrival times of $(A_2, A_3, A_4) = (2, 4, 5)$ and required bit rates of $(R_2, R_3, R_4) = (16\text{K}, 8\text{K}, 16\text{K})$. The PS assigns three

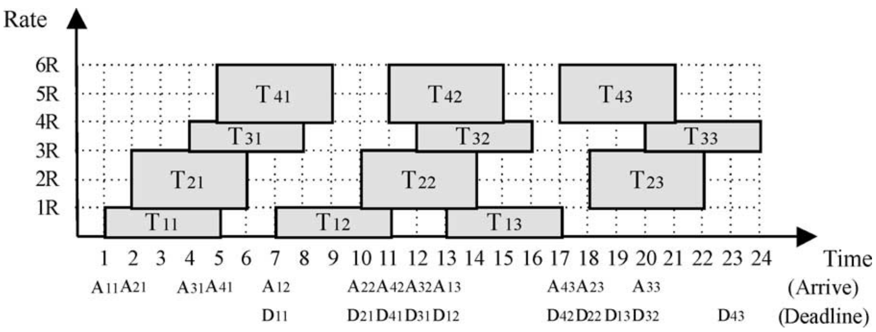

Fig. 5. An example illustrating the usage of RTGS algorithm.

channelization codes whose bit rates are 16K, 8K and 16K for S_2, S_3 and S_4, respectively, since the LC determines that the power consumption values $(T_2, T_3, T_4) = (4.10\ \text{W}, 4.14\ \text{W}, 4.21\ \text{W})$ are less than 18 W. For session S_2, PS schedules T_{21} in the interval (2,6), T_{22} in the interval (10,14) and T_{23} in the interval (18,22). For session S_3, PS schedules T_{31} in the interval (4,8), T_{32} in the interval (12,16) and T_{33} in the interval (20,24). Finally, instances T_{41}, T_{42} and T_{43} of session S_4 are scheduled in intervals (5,9), (11,15) and (17,21), respectively. Consequently, all data sessions are scheduled before their deadlines. The amount of the allocated bit rate for sessions $S_1 - S_4$ is 48K.

The advantage of the RTGS algorithm is that all instances can be finished prior to their deadlines. For a data session request, however, a new channelization code must be allocated to spread its instances. For example, RTGS requires four channelization codes to manipulate four real-time data sessions at time $t = 13$ (Fig. 5), for which the maximum power consumption is 4.21 W. This result implies that higher code consumption corresponds to higher power consumption.

3.2. Real-time bandwidth scheduling

Next, we present the real-time bandwidth scheduling (RTBS) algorithm, which overcomes the shortcoming and retains the advantage of the RTGS algorithm. Like RTGS, RTBS exploits the management framework and makes the decision based on the capacity analysis model for real-time session admissions. Furthermore, RTBS applies the schedulability analysis by using the EDF scheme [16] to schedule the real-time data sessions and save power. All instances of a data session are delivered in time for their deadlines due to the delay constraint. In our schedulability analysis, if an instance will miss its deadline because of limited power, it is discarded to prevent variation in the end-to-end delay. However, if the transmission of an instance can be finished before its deadline, RTBS can reschedule the instance's active time (but not miss the instance's deadline) if necessary. The base station will control channelization code allocations and prevent high simultaneous mobile user activity since instances can be rescheduled if necessary. Consequently, the base station can decrease the power consumption by using RTBS.

Now, consider that a data session has a feasible schedule, by using the EDF scheme. If a data session is a feasible schedule by EDF, all transmission of the session's instances can be finished before their deadlines. However, if a data session is not a feasible schedule, some of the session's instances will miss their deadlines anyway. RTBS performs the schedulability analysis based on the instance's period and processing time of a data session, and the allocated channel number, C, of a base station. Assume that P_i denotes the instance's processing time of session i by using one channelization code. C denotes the number of channelization codes that can be allocated by a base station. If a channelization code is used to spread an instance, the system can create C channelization codes for simultaneously spreading C instances with the available power resource; therefore, P_i/C denotes an instance's processing time of session i that uses C channelization codes to transmit the instance simultaneously. The fraction of radio resources, $(P_i/C)/E_i$, spent on transmitting an instance of session i since E_i denotes the period of the instance. Then, the resource utilization for n data sessions is given by

$$U = \sum_{i}^{n} \frac{P_i/C}{E_i}$$

Assume that each session contains one periodic task. Based on the schedulable analysis, a set of periodic tasks is schedulable with EDF in a base station if and only if the resource utilization U is less than 1. Thus, if $U \leq 1$, session i has a feasible schedule by the EDF scheme. We can schedule all sessions with the same channel number before the tasks miss their deadlines. However, if $U > 1$, session i cannot be scheduled by the EDF scheme. Then, a new channelization code must immediately be allocated for the session to manipulate the session's schedulability.

RTBS works as follows. For a data session request, the management framework defines a instance's period E, processing time P, required bit rate R_i, and allowed maximum bit rate $R_{\max}$, according to the QoS profile of service.

1. In the beginning, the PS maintains a prioritized queue in which instances of data sessions are ordered according to the EDF basis.
2. The PS makes schedulable analysis for the session when a data session request arrives. The utilization factor, U, is calculated using the instance's period E, processing time P and the presently assigned channel number C.
3. If $U \leq 1$, the data session is a feasible schedule by EDF. The AC allows for the session establishment. Go to Step 5.
4. If $U > 1$, the data session is not a feasible schedule by EDF. The PS increases the channel number, C. The LC makes power reservation to increase channel number. According to the required bit rate, the LC estimates power consumption P_{tot} of the base station. If the power, P_{tot}, is below the threshold tx_{admit}, then a new session is admitted. Otherwise, the AC rejects the request.
5. If a data session is admitted, the PS puts session's instance into the queue and schedules it. The scheduler chooses C instances with higher priority (early deadline instances) in the queue for execution.
6. The PS, based on the required bit rate, assigns the channelization codes and the required power for the instances. Following the power allocation, the PS asks the LC to change the system's load.
7. The PS terminates an instance when it is finished or its deadline is reached. When an instance is terminated,

the PS removes this instance from the queue and releases the occupied channelization code as well as the allocated power. Finally, the PS asks the LC to change the system's load.

The RTBS algorithm is presented by pseudo code as shown in Fig. 6.

Fig. 7 illustrates a feasible schedule constructed using the RTBS algorithm. Based on the service profiles (Table 2), the figure shows the schedule of four sessions at time $t = 24$. Assume that the maximum power of the base station is $tx_{\max} = 20$ W and the threshold for a newly admitted real-time session is $tx_{\text{admit}} = 18$ W. In Fig. 7(a), instance T_{11} of session S_1 arrives at time $t = 1$. The instance's required bit rate, R_1, is 8K. The PS doesn't make schedulability analysis for this new request because the allocated channel number $C = 0$. The channelization code of a 8K bit rate is assigned to S_1, and the channel number is $C = 1$. Thus, PS can only use one channelization code to schedule instances at any time. Finally, instances T_{11}, T_{12} and T_{13} are schedulable before their deadlines (Fig. 7(a)).

In Fig. 7(b), instance T_{21} of session S_2 arrives at time $t = 2$ and it requires a bit rate of $R_2 = 16$K. Consider the case that channel number $C = 1$, for which the utilization factor, U, is

$$U = \frac{P_1/C}{E_1} + \frac{P_2/C}{E_2} = \frac{4}{6} + \frac{4}{8} > 1$$

S_2 is not schedulable by the EDF scheme since $U > 1$. That is, the transmission of all instances of S_2 cannot be completed with channel number $C = 1$. In this case, PS assigns a new channelization code for S_2. Since LC estimates power, $P_2 = 4.14$ W, is less than 18 W, PS assigns a new channelization code with bit rate 16K for S_2, and sets the channel number $C = 2$. Finally, instances T_{21}, T_{22} and T_{23} can be scheduled before their deadlines (Fig. 7(b)).

In Fig. 7(c), instance T_{31} of session S_3 arrives at time $t = 4$. It requires a bit rate of $R_3 = 8$K. The utilization factor, U, is

$$U = \frac{P_1/C}{E_1} + \frac{P_2/C}{E_2} + \frac{P_3/C}{E_3} = \frac{2}{6} + \frac{2}{8} + \frac{2}{8} < 1$$

Since $U \leq 1$, S_3 is a feasible schedule by the EDF scheme. T_{31} can be scheduled at time $t = 5$. The PS schedules T_{31} in the intervals (5,9) before its deadline D_{31} at time $t = 12$. Similarly, instance T_{32} can be scheduled in the interval (12,13) and (14,17), and for instance T_{33} in (20,24) before their deadlines $D_{32} = 20$ and $D_{33} = 28$, respectively. Hereafter, RTBS can only apply two channelization codes to schedule sessions $S_1 - S_3$ at any instant.

In Fig. 7(d), instance T_{41} of session S_4 arrives at time $t = 5$ and requires a bit rate of $R_4 = 16$K. Thus

$$U = \frac{P_1/C}{E_1} + \frac{P_2/C}{E_2} + \frac{P_3/C}{E_3} + \frac{P_4/C}{E_4}$$

$$= \frac{4/2}{6} + \frac{4/2}{8} + \frac{4/2}{8} + \frac{4/2}{6} > 1$$

S_4 cannot be scheduled by the EDF scheme. PS allocates a new channelization code for S_4 and increases channel number, C, to 3. After the channel number is increased, all the instances of sessions $S_1 - S_4$ can be scheduled before their deadlines. In this case, RTBS only applies three channels to control four real-time data sessions, in which the amount of the bit rate allocation is 40K (at time $t = 5$), and the maximum power consumption is 4.17 W. Consequently, RTBS utilizes more bandwidth and consumes less power than RTGS (4.21 W).

```
Algorithm RTBS
  Initiate the setting of parameter tx_admit, generate EDF prioritized queue
  While (an event occurs)
      Case 1 : A session request arrivals
            Compute utilization factor U = Σ_i (P_i/C)/E_i ; /* Schedulable analysis for session*/
            If (U > 1) then
               AC makes power reservation;
               LC estimates BTS's power consumption P_tot;   /* By Eq. (1) */
               If (P_tot < P_admit) then
                   Create new real-time data session;     /* Increase active #MS */
                   Increase channel number C;
               Else
                   Reject real-time data session; Break;
            Else
               Put instance into queue;
               Schedule the instances of the queue by EDF basis;
      Case 2 : Get C high priority instances from queue for execution
            Assign channelization codes and power resources;
            Update system load status;
      Case 3 : An instance is finished or its deadline is reached
            Release channelization code and power resource;
            Update system load status;
End Algorithm
```

Fig. 6. Pseudo code for the RTBS algorithm.

3.3. Real-time code scheduling

Within the RTBS mechanism, the real-time code scheduling (RTCS) algorithm applies DCA scheme to improve radio resource utilization. In the WCDMA system, the DCA is used to assign a channelization code for task activation. The DCA scheme changes the bit rate during a data session that is based on the OVSF technique [17,18]. The DCA must assign channelization codes according to the adaptive spreading factor (SF) to satisfy variable-rate transmission. In a single-code transmission, however, the assigned basic rate code is limited to a multiple of 2^n. For example, if SF = 32 can provide a bit rate of up to 64 Kbps, then SF = 16 will provide a maximum bit rate of up to 128 Kbps such that a 3.84 Kbits task can be completely transmitted within 10 and 5 ms, using SF = 32 and 16, respectively.

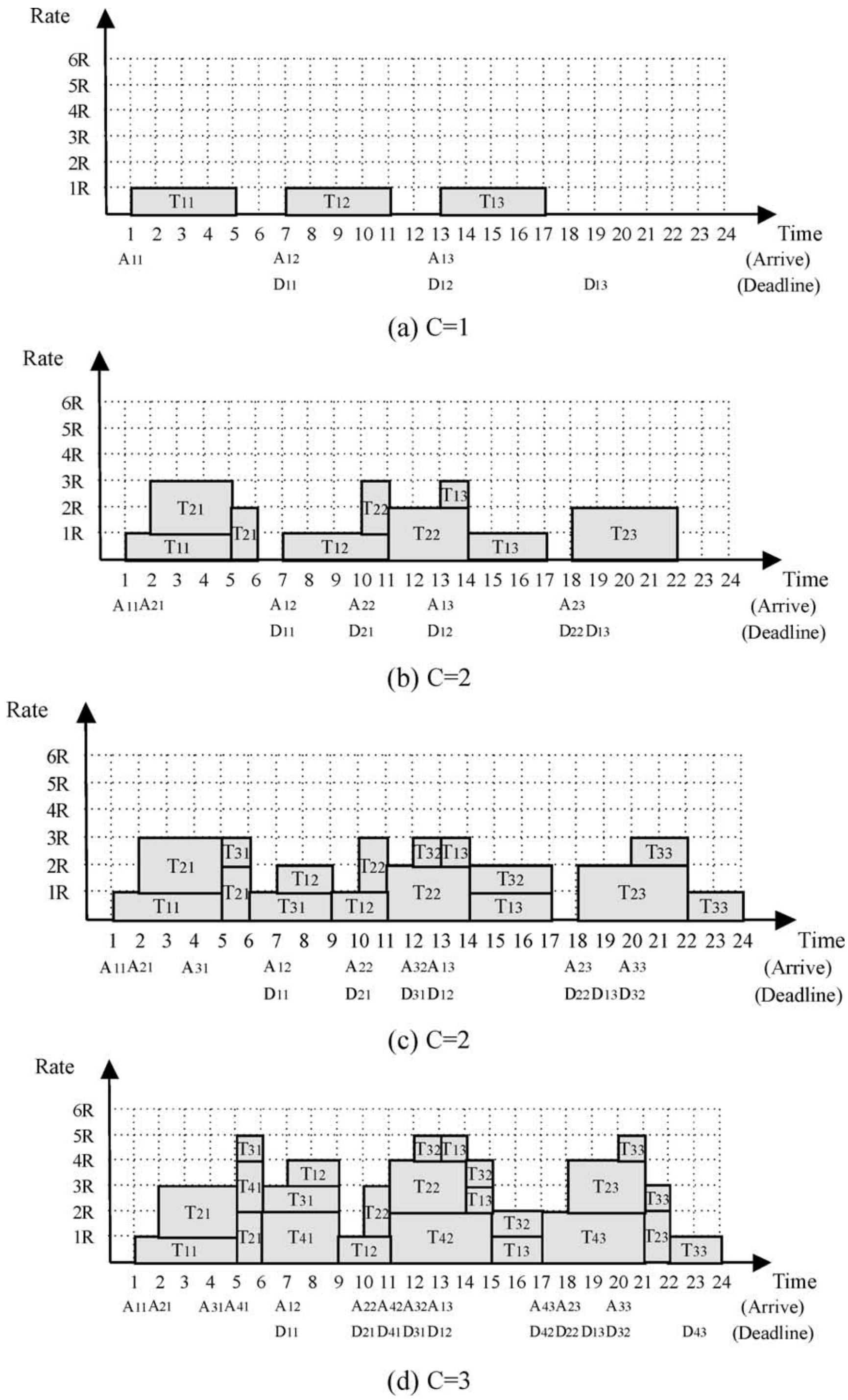

Fig. 7. An example illustrating the usage of RTBS algorithm.

RTCS, based on the DCA scheme, can assign a high bit rate channelization code to reduce an instance's processing time within a session, and prevent large simultaneous task transmissions within a cell. Thus, the RTCS can minimize the power consumption of a base station.

Suppose that a data session request arrives. The instance's period E, processing time P, required bit rate R_i and allowed maximum bit rate $R_{\max}$ are defined according to the QoS profile of service. Notably, the instance allocation bit rate R_c begins at $R_c = R_i$ and is limited by its maximum bit rate. Thus, the RTCS algorithm works as follows.

1. Initially, the PS maintains a prioritized queue in which instances of data sessions are ordered by EDF.
2. When a data session request arrives, the PS performs the schedulability analysis for the session. The utilization factor, U, is determined from the instance's

period E, processing time P and the presently assigned channel number, C.

3. If $U \le 1$, the data session can be feasibly scheduled by EDF. The AC allows the session to be established. Go to Step 6.
4. If $U > 1$, the data session cannot be feasibly scheduled by EDF. The PS makes the decision for the allocated bit rate, R_c. If $R_c < R_{max}$, then the PS changes the channelization code with high bit rate. The PS determines the utilization factor, U. Go to step 3. Otherwise, if $R_c = R_{max}$, then the PS has assigned the maximum bit rate for the session. Go to the next step.
5. The PS increases the channel number, C. The LC reserves power for channel number increasing. The LC estimates power consumption P_{tot} of the base station according to the required bit rate. If the power, P_{tot}, is below the threshold, tx_{admit}, a new session is admitted. Otherwise, the AC rejects the request.
6. If a session is admitted, the PS puts the session's instance into the queue and schedules it. The scheduler chooses C instances with higher priority in the queue for execution.
7. The PS assigns the channelization codes and the required power for the instances based on the required bit rate. After the power allocation is finished, the PS asks the LC to change the system's load.
8. The PS ends an instance when it is finished or its deadline is reached. The PS removes a terminated instance from the queue, and releases the occupied channelization code and the allocated power. Finally, the PS asks the LC to change the system's load.

Fig. 8 shows the pseudo code of the RTCS algorithm. Based on the same template as RTBC, the RTCS adds the DCA scheme to change the bit rate during a session and assign a channelization code with a higher bit rate to reduce the instance's processing time of a session.

The service profiles given in Table 2 are used to describe the RTCS algorithm. Fig. 9 shows a feasible schedule constructed using the RTCS algorithm to schedule the four sessions at time $t = 24$. Assume that the maximum power of the base station is $tx_{max} = 20$ W and the threshold for a newly admitted real-time session is $tx_{admit} = 18$ W. In Fig. 9(a), instance T_{11} of session S_1 arrives at time $t = 1$. The PS assigns a channelization code whose bit rate is 8K for S_1, and sets the channel number, $C = 1$. PS schedules T_{11} in the interval (1,5), T_{12} in the interval (7,11) and T_{13} in the interval (13,17).

In Fig. 9(b), instance T_{21} of session S_2 arrives at time $t = 2$ and requires a bit rate of $R_2 = 16$K. Consider the channel number $C = 1$, whose utilization factor, $U > 1$. S_2 is not a feasible schedule by the EDF scheme. RTCS applies the DCA scheme to change the channelization code of S_2 with a higher bit rate. The PS reassigns the channelization code with $R_c = 32$K to S_2 because the used bit rate R_c (16K)

```
Algorithm RTCS
  Initiate the setting of parameter tx_admit, generate EDF prioritized queue
  While (an event occurs)
      Case 1 : A session request arrivals
          Compute utilization factor U = Σ_i (P_i/C)/E_i ; /* Schedulable analysis for session */
          If (U > 1) then
              If (R_c < R_max) then                /* allocated bit rate < R_max */
                  Use DCA scheme to assign higher bit rate channelization code;
                  Decide new processing time according to channelization code;
              Else
                  AC makes power reservation;
                  LC estimates BTS's power consumption P_tot;   /* By Eq. (1) */
                  If (P_tot < P_admit) then
                      Create new real-time data session; /* Increase active #MS */
                      Increase channel number C;
                  Else
                      Reject real-time data session;
          Else
              Put instance into queue;
              Schedule the instances of the queue by EDF basis;
      Case 2 : Get C high priority instances from queue for execution
          Assign channelization codes and power resources;
          Update system load status;
      Case 3 : An instance is finished or its deadline is reached
          Release channelization code and power resource
          Update system load status;
End Algorithm
```

Fig. 8. Pseudo code for the RTCS algorithm.

is less than R_{max} (32K). Hence, the processing time of instance T_{21} can be reduced from four to two units. Now, the PS calculates utilization factor U

$$U = \frac{4}{6} + \frac{2}{8} < 1$$

After the code is reassigned, S_2 is a feasible schedule by the EDF scheme. The transmission of T_{21} starts at time $t = 5$. The PS schedules T_{21} in the interval (5,7) before its deadline $D_{21} = 10$. Similarly, instances T_{22} and T_{23} can be scheduled in the interval (11,13) and (18,20) before deadlines $D_{22} = 18$ and $D_{23} = 26$, respectively (Fig. 9(b)). Assume that S_3 arrives at time $t = 4$ (Fig. 9(c)). Then, the utilization factor, U, is

$$U = \frac{P_1/C}{E_1} + \frac{P_2/C}{E_2} + \frac{P_3/C}{E_3} = \frac{4}{6} + \frac{2}{8} + \frac{4}{8} > 1$$

S_3 cannot complete its transmission when the system uses only one channelization code. In this case, a new channelization code needs to be assigned. The PS assigns a new channelization code whose bit rate is 8K for S_3 since the LC estimates a power consumption of $P_3 = 4.08$ W that is less than 18 W. The PS sets the channel number at $C = 2$ and reschedules the sessions S_1–S_3 at time $t = 4$. The instances T_{21}, T_{22} and T_{23} are reassigned in the intervals (4,6), (10,12) and (18,20), respectively. Finally, instances T_{31}, T_{32} and T_{33} can be scheduled in the intervals (5,9), (12,16) and (20,24), respectively, before their deadlines (Fig. 9(c)).

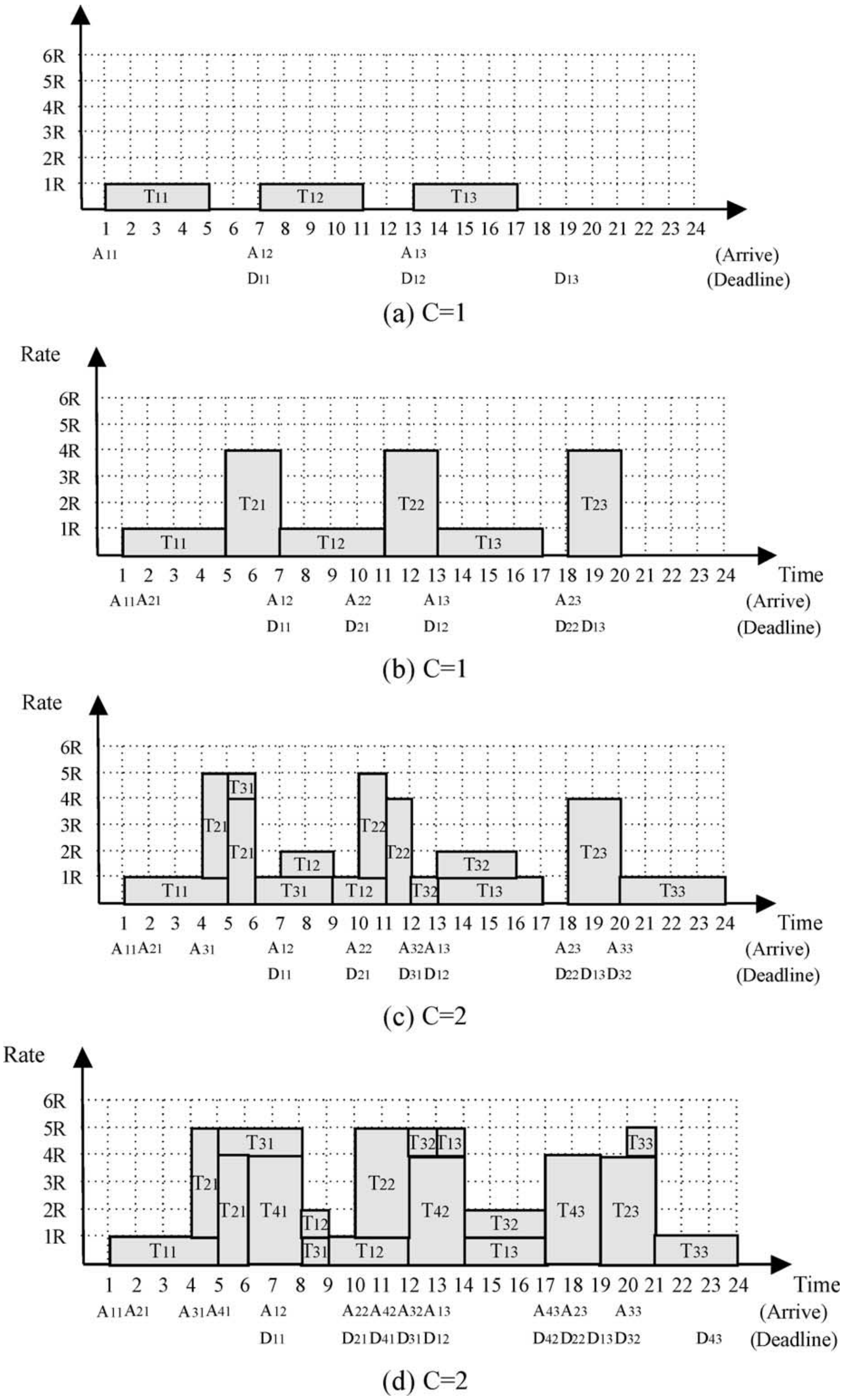

Fig. 9. An example illustrating the usage of RTCS algorithm.

In Fig. 9(d), instance T_{41} of session S_4 arrives at time $t = 5$, requiring a bit rate of $R_4 = 16\text{K}$. Thus

$$U = \frac{P_1/C}{E_1} + \frac{P_2/C}{E_2} + \frac{P_3/C}{E_3} + \frac{P_4/C}{E_4}$$

$$= \frac{4/2}{6} + \frac{2/2}{8} + \frac{4/2}{8} + \frac{4/2}{6} > 1$$

The channelization code of S_4 is reassigned as $R_c = 32\text{K}$. The processing time of instance S_4 can be reduced from four to two units. Then, the utilization factor, U, is checked:

$$U = \frac{4/2}{6} + \frac{2/2}{8} + \frac{4/2}{8} + \frac{2/2}{6} < 1$$

S_4 is schedulable after the code is reassigned. At time $t = 5$, the PS schedules the instances T_{41}, T_{42} and T_{43} of S_4 in the intervals (6,8), (12,14) and (17,19), respectively. Moreover, all instances of S_1-S_3 can be rescheduled before their deadlines (Fig. 9(d)). This example shows that RTCS only needs two channelization codes to schedule sessions S_1-S_4. Although the maximum bit rate allocation of RTCS is 40K (at time $t=5$), which is the same as in RTBS, the power consumption of RTCS is 4.15 W, which is less than the power consumptions in RTBS (4.17 W) and RTGS (4.21 W).

4. Simulation results

This section presents the simulations of the various real-time scheduling algorithms. Experiments are performed to examine power consumption, session drop rate and bandwidth utilization with traffic class to evaluate the performance of the proposed algorithms. The power consumption is measured when a base station allocates channelization codes for data session requests, based on the capacity analysis model of Eq. (1). The session drop rate is calculated from the probability of sessions rejected. The bandwidth utilization is defined as the allocated bit rate divided by the requested bit rate of all data sessions.

The experimental input parameters are as follows.

- The session requests are classified into five traffic classes. The occurrences of each class are uniformly distributed.
- Session arrivals are Poisson distributed with mean arrival rate of $A = 1-20$ sessions/s. Session duration time is exponentially distributed with mean value of $U = 100$ s. Thus, the traffic load L is $L = A \times U$.
- Each session involved contains several real-time instances whose arrivals are periodic. The deadline of each instance is equal to the period.

For each traffic class, Table 3 gives the input parameters of the instance's possible period, processing time, required bit rate and maximum bit rate. Figs. 10–12 display the experimental results concerning of power consumption, session drop rate and bandwidth utilization versus traffic load for three proposed algorithms. Each traffic load is simulated with 10,000 session requests.

RTGS consumes the most power. It will assign one channelization code for each session arrival. RTBS, applying the real-time EDF scheme to schedule sessions' instances, is more cost-effective comparing to RTGS. RTCS can change the bit rate allocation for a session based on DCA scheme to prevent large, simultaneous instance transmissions. Consequently, RTCS improves the power consumption over that of RTBS. Fig. 10 demonstrates the power consumption using RTGS, RTBS and RTCS. In this figure, the traffic load increases from 800 to 2600. The respective power consumptions of RTGS, RTBS and RTCS are 15.14, 13.49 and 12.87 W when traffic load is 800. The improvements in power consumption of RTGS, varies from approximately 11% compared to RTBS, to approximately 15% compared to RTCS. When the traffic load is 2600, the power consumption is improved from approximately 7.28% compared to RTBS, to approximately 7.85% compared to RTCS.

Fig. 11 plots session drop rate against traffic load for RTGS, RTBS and RTCS. The RTCS, consuming less power, also has a lower session drop rate. As shown in Fig. 11, the session drop rate of RTGS is 0.05, and that of both RTBS and RTCS are 0 for a traffic load of 600. In the case of heavy load (traffic load = 2200), the session drop rates are 0.71 for RTGS, 0.63 for RTBS and 0.60 for RTCS.

Table 3
Selected input parameters for simulations

Session	Period (ms)	Processing (ms)	Required rate (Kbps)	Maximum rate (Kbps)
S_1	60	40	8	16
S_2	80	40	16	32
S_3	80	40	8	16
S_4	60	40	8	16
S_5	100	60	16	32

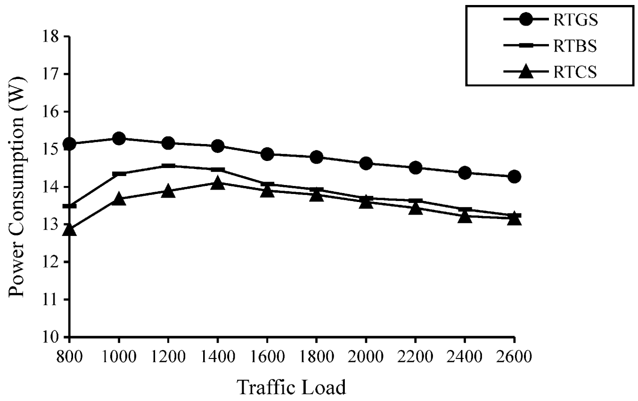

Fig. 10. Power consumption of algorithms versus traffic load.

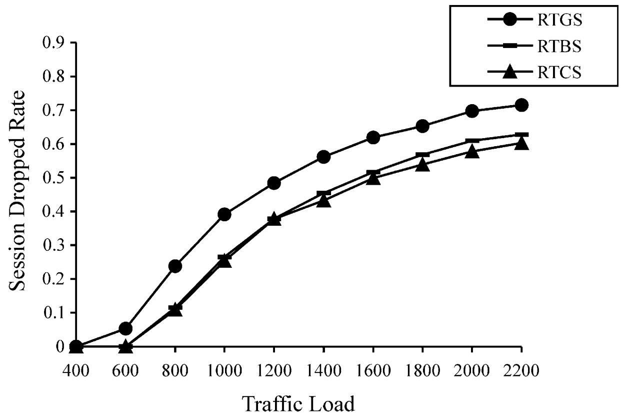

Fig. 11. Session drop rate of algorithms versus traffic load.

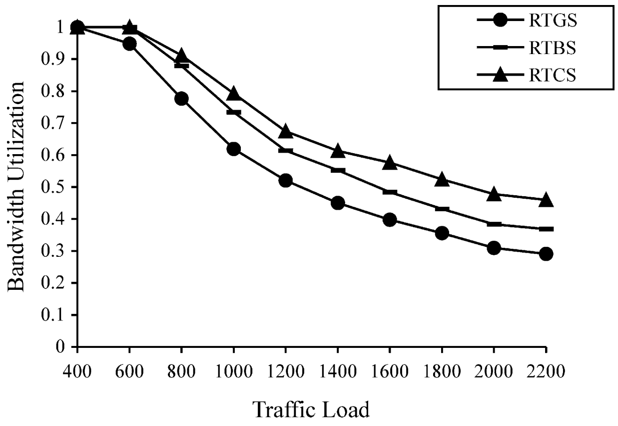

Fig. 12. Bandwidth utilization of algorithms versus traffic load.

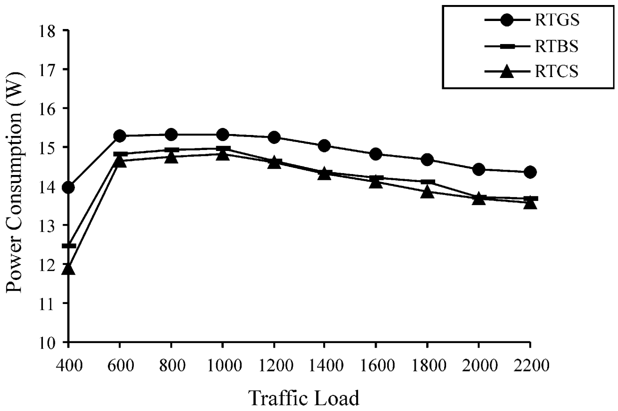

Fig. 13. Power consumption of algorithms versus traffic load with higher data rate.

The bandwidth utilization shows that the mobile system prefers to drop data sessions with higher bit rates because the high bit rate sessions consume more power. Fig. 12 indicates that the respective bandwidth utilizations of RTGS, RTBS and RTCS are 0.95, 1 and 1 for a traffic load of 600. Even at heavy traffic load of 2200, RTCS still performs best with respect to bandwidth utilization.

Next, the performance of the three real-time algorithms is examined with higher data rate requirements (Table 4). As listed in Table 3, the input parameters of our simulation are the same as before, except that the bit rate for session requirements is doubled. Figs. 13–15 present the experimental results of power consumption, session drop rate and bandwidth utilization versus traffic load for the three proposed algorithms under new bit rate. Each simulation again involves 10,000 data sessions.

Fig. 13 displays the power consumption of RTGS, RTBS and RTCS as the load increases from 400 to 2200. The power consumption is 13.96 W for RTGS, 12.47 W for RTBS and 11.91 W for RTCS at a traffic load of 400. Compared to RTGS, the improvement in power consumption varies from 8% for RTBS to 15% for RTCS. When the traffic load is 2200, the power consumption is still improved by 4.67% for RTBS and 5.37% for RTCS.

Fig. 14 depicts that the session drop rates are 0.13 for RTGS, 0.06 for RTBS, and 0.05 for RTCS at a traffic load of 400. In the simulation with heavy load (traffic load = 2200), the session drop rates of RTGS, RTBS and RTCS are 0.83, 0.78 and 0.76, respectively. Fig. 15 shows that the respective bandwidth utilizations of RTGS, RTBS and RTCS are 0.87, 0.93 and 0.95 for a traffic load of 400. As the traffic load increases to 2200,

Table 4
Selected input parameters with higher data rate

Session	Period (ms)	Process (ms)	Initiated (Kbps)	Maximum (Kbps)
S_1	60	40	16	32
S_2	80	40	32	64
S_3	80	40	16	32
S_4	60	40	16	32
S_5	100	60	32	64

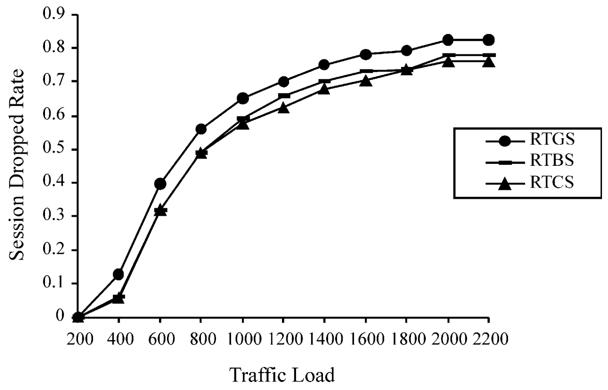

Fig. 14. Session drop rate of algorithms versus traffic load with higher data rate.

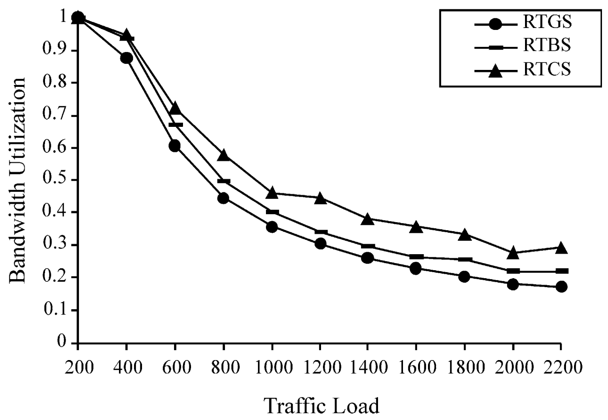

Fig. 15. Bandwidth utilization of algorithms versus traffic load with higher data rate.

the observed improvement of bandwidth utilization is 0.29 for RTCS, which is greater than 0.17 for RTGS and 0.22 for RTBS.

5. Conclusions

The development of an all IP-based radio access network is the future for the next generation mobile system. The network enables packet-oriented connections to offer real-time applications. The radio access system must provide efficient management for QoS provisioning, due to the scarcity of radio resources.

This work proposed real-time algorithms for packet scheduling. RTGS algorithm applies the functionalities of the management framework. It uses the capacity analysis model to decide the threshold power for newly admitted real-time sessions. RTBS algorithm retains the advantages of RTGS. For the power resource consumption, RTBS implemented the EDF scheme to perform the schedulability analysis and schedule the real-time sessions. The algorithm consumes less power than that of RTGS. RTCS algorithm applied the DCA scheme, which increases the schedulable probability of the sessions to improve the utilization of radio resources.

Experimental results show that, under various traffic loads, RTCS performs best in terms of power consumption, session drop rate and bandwidth utilization. It also shows that RTBS outperforms RTGS. In the case of higher bit rate requirements, RTCS still outperforms RTBS, and RTBS outperforms RTGS.

To consider a WCDMA system, the macro diversity occurs when the mobile user may use cells belonging to different base stations. The macro diversity starts soft handover, which consumes more radio resource because the mobile user is occupying more than one radio link. In this circumstance, our algorithms need more work to consider the macro diversity and handover control.

References

[1] S. Dixit, Y. Guo, Z. Antoniou, Resource management and quality of service in third-generation wireless networks, IEEE Commun. Mag. 40 (2001) 125–133.
[2] H. Holma, A. Toskala, WCDMA for UMTS, Wiley, Nokia, Finland, 2000.
[3] N. Musikka, L. Rinnbäck, Ericsson's IP-based BSS and radio network server, Ericsson Rev. 4 (2000) 224–233.
[4] G.A. Eriksson, B. Olin, K. Svanbro, D. Turian, The challenges of voice-over-IP-over-wireless, Ericsson Rev. 1 (2000) 20–31.
[5] L. Jorguseski, E. Fledderus, J. Farserotu, R. Prasad, Radio resource allocation in third-generation mobile communication systems, IEEE Commun. Mag. 40 (2001) 117–123.
[6] Ö. Gürbüz, H. Owen, Dynamic resource scheduling strategies for QoS in W-CDMA, Proc. IEEE GLOBECOM '99 1a (1999) 183–187.
[7] F.H.P. Fitzek, R. Morish, A. Wolisz, Comparison of multi-code link-layer transmission strategies in 3G wireless CDMA, IEEE Commun. Mag. (2000) 58–64.
[8] S.K. Das, R. Jayaram, N.K. Kakani, S.K. Sen, A call admission and control scheme for quality-of-service (QoS) provisioning in next generation wireless networks, Wireless Networks (2000) 17–30.
[9] W.K. Shih, J.W.-S. Liu, Algorithms for scheduling imprecise computations with timing constraints to minimize maximum error, IEEE Trans. Comput. 44 (3) (1995) 466–471.
[10] V. Millan-Lopez, W. Feng, J.W.-S. Liu, Using the imprecise-computation technique for congestion control on a real-time traffic switching element, International Conference on Parallel and Distributed Systems, 1994, pp. 202–208.
[11] A. Mittal, G. Manimaran, Integrated dynamic scheduling of hard and QoS degradable real-time tasks in multiprocessors systems, Fifth International Conference on Real-Time Computing Systems and Applications, Proceedings, 1998, pp. 127–136.
[12] K. Hiltunen, R.D. Bernardi, WCDMA downlink capacity estimation, IEEE VTEC2000 (2000) 992–996.
[13] K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver Jr, C.E. Wheatley III, On the capacity of a cellular CDMA system, IEEE Trans. Veh. Technol. 40 (1991) 303–312.
[14] N. Dimitriou, R. Tafazolli, G. Sfikas, Quality of service for multimedia CDMA, IEEE Commun. Mag. (2000) 88–94.
[15] R. Kohno, R. Meidan, B. Milstain, Spread spectrum access method for wireless communications, IEEE Commun. Mag. (1995) 58–67.
[16] J.W.-S. Liu, Real-time Systems, Prentice-Hall, Englewood Cliffs, NJ, 2000.
[17] R.G. Cheng, P. Lin, OVSF code channel assignment for IMT-2000, IEEE VTEC2000 (2000) 2188–2192.
[18] F. Adachi, M. Sawahashi, K. Okawa, Tree-structured generation of orthogonal spreading codes with different lengths for forward link of DS-CDMA mobile, Electron. Lett. 33 (1) (1997) 27–28.