Resource Allocation in Multimedia CDMA Communication Systems*

Vincent Huang† and Weihua Zhuang‡
Centre for Wireless Communications (CWC)
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Abstract

The future wireless communication systems are expected to support a broad range of multimedia services to mobile users with guaranteed quality-of-service (QoS). With the increasing demands for wireless services, the available frequency bandwidth of the systems is very limited and should be used efficiently. In this paper, we consider a wireless code-division multiple access (CDMA) system and develop an algorithm to allocate system resources to multimedia users for QoS provisioning and for high resource utilization. We propose a medium access control (MAC) protocol which exploits both time-division and code-division multiplexing. The MAC protocol uses fair packet loss sharing (FPLS) scheduling to guarantee the QoS requirements. The FPLS scheduler uses the information of traffic rate distribution and QoS requirements of the users to assign priorities to the users and determines an efficient accommodation of the packets in the time slots of each frame, so that the number of the served users is maximized under the QoS constraints. The resource utilization problem is formulated as an optimization problem. For practical solution, we propose the sub-optimal bin-packing approach for the single-cell system. Numerical results are presented to demonstrate the performance of the single-cell approach.

Index Terms — code-division multiple access (CDMA), medium access control (MAC), multimedia traffic, packet scheduling, quality-of-service (QoS) provisioning.

---

* This research was supported by research grants from the Communications Information Technology Ontario (CITO) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.
† Email: huang@bbcr.uwaterloo.ca
‡ Contact author. Email: wzhuang@bbcr.uwaterloo.ca

I. Introduction

The freedom of movement has motivated the increasing use of mobile phones and portable computers. Wireless communication services have been extending from pure voice services to multimedia services. A broad range of multimedia services which include voice, data, and video will be supported by the third generation (and beyond) wireless communication systems. Unlike wireline communication systems, wireless systems have very scarce bandwidth of available frequency spectrum. The limited network resources have to be used efficiently to provide satisfactory services to the users.

Wideband code-division multiple access (CDMA) has been selected as the standard multiple access technique for the third generation wireless multimedia networks. Unlike time-division multiple access (TDMA) and frequency-division multiple access (FDMA) where the communication channels for different users are separated in time and frequency respectively, in a CDMA system, by assigning each user a specific set of codes, all the users share the same frequency spectrum and more than one user can transmit at the same time. Therefore, in a CDMA system, power control has to be used to combat the near-far problem. In supporting voice only services, the received signal power from each of the mobile users in the cell is maintained at a constant level in the uplink at the base station [1]. However, in a multimedia CDMA system, different traffic classes require different transmission accuracies specified by bit error rate (BER). If the received power is kept at the same level for all traffic classes, the capacity is limit by the most stringent BER requirement and cannot be used efficiently. Recently, several approaches have been proposed for optimal power control for multimedia traffic to maximize the capacity or to minimize the total transmit power [2][3][4].

Optimal resource management for wireless multimedia CDMA systems has been an active research topic in recent years. For a single-cell system, minimizing the total transmitted power and maximizing the total transmission rate in the cell are treated as two separate optimization problems for the reverse link transmission [5][6][7][8]. In [2], resource management (transmission power and rate allocation) is combined with base station assignment for the reverse link transmission in a multi-cell system. The previous work on resource management assumes a continuous-time transmission with the processing gain depending on the allocated time-varying transmission rate. The system model cannot efficiently accommodate bursty data traffic with short bursts. Furthermore, it requires high implementation complexity for the variable processing gain. As data services become more and more important, it is necessary for a wireless system to efficiently support data traffic for applications such as web browsing, email, file transfer, fax, and remote log-in.

Packetized transmission over wireless links makes it possible to achieve a high statistical multiplexing gain. However, the packet transmissions have to be arranged by medium access control (MAC) protocols. Packet flows generated by mobile users can be classified to several traffic classes. Each of these classes has its unique quality-of-service (QoS) requirements and traffic characteristics. Due to heterogeneous nature of multimedia traffic flows, the traditional voice-based MAC protocols do not perform well in a multimedia environment. A flexible MAC protocol, which can efficiently accommodate multimedia traffic is required. One important MAC issue is the packet scheduling. The order of packet transmissions for multimedia traffic has a great impact on the efficiency and performance for wireless networks. However, the design of a packet scheduler involves balancing a number of conflicting requirements. Common criteria for the packet scheduler design include maximization of throughput, QoS provisioning, scheduling according to a pre-defined priority structure, and low implementation complexity as packet scheduling is implemented in real time. Most packet scheduling strategies, such as first-in-first-out (FIFO), round robin, and generalized processor sharing (GPS) [9][10] have been originally proposed for wireline networks. Random access protocols have been widely used in the past for wireless networks. Packet reservation multiple access (PRMA) [11] is a well-known TDMA based protocol proposed for voice and data traffic. For CDMA based systems, some methods have been proposed without considering TDMA component [12]. Hybrid time-division/code-division multiple-access schemes have been proposed in [13][14]. In [13], only traffic rate and delay constraints are considered in resource allocations, packets with different BER requirements are not differentiated. In [14], a wireless multimedia access control protocol with BER scheduling called WISPER is proposed, where packets with the same or similar BER requirements are transmitted in the same time slot with the same received power level for all the packets.

In this paper, a MAC protocol is proposed for CDMA multimedia communications. The MAC protocol exploits both time-division and code-division multiplexing for efficient resource utilization. The resource allocation is formulated as an optimization problem. The objective of the optimal scheduler is to schedule the transmission of the packets so that the system throughput is maximized and the QoS requirements of each user are guaranteed. Since the optimization problem can not be solved in real time, a realistic packet scheduler called fair packet loss sharing (FPLS) is proposed. FPLS is a QoS requirement based packet scheduling algorithm. The objectives of the scheduling are to provide QoS guarantees in terms of delay and BER and to maximize the system resource utilization. In a wireless environment, a packet is expected to be delivered to the destination within required time frame and with certain accuracy. Any violation of these two requirements will cause the packet to be useless and will be discarded. As QoS satisfaction and high resource utilization are in general conflicting goals, high utilization of the limited wireless bandwidth often means that the system resources cannot accommodate the traffic load from time to time, and some packets have to be dropped occasionally. To support as many satisfied users as possible, fair sharing of the dropped packets among all users is essential. With the FPLS, the bandwidth is shared among all users in such a way so that, when the QoS requirements are guaranteed for one user, they will be guaranteed for all other users at the same time. No user will be allocated more bandwidth than needed if the bandwidth is not enough for other users.

The remainder of this paper is organized as follows. In Section II, we describe the system model, including the proposed MAC protocol and the QoS provisioning. The optimization problem formulation is described in Section III. The proposed FPLS scheduler is presented in Section IV. In Section V, the performance of the FPLS scheduler is studied by computer simulation and is compared with the GPS and WISPER scheduling schemes. Finally, Section VI summarizes this research.

II. System Model

We consider a hybrid time-division/code-division multiple access (TD/CDMA) cellular system with packetized transmission [15]. The system operates in the time division duplex (TDD) mode, as

TDD can best accommodate asymmetric traffic [16]. In a multimedia system, the traffic loads for uplink and downlink are highly asymmetric. The boundary between uplink and downlink is controlled by software. Time is partitioned into frames of a constant duration. Each frame is divided into time slots. Multiple access within each time slot is accomplished by assigning unique pseudo-random noise (PN) code sequence(s) to each user. The source information from and to mobile users is segmented into packets of equal length. The packets are transmitted at a constant bit rate, each packet requiring a time slot for transmission. Packet transmission from and to mobile users is synchronized in time. The decision on packet transmission in each time slot for both uplink and downlink is made at the base station and is broadcast to the mobile users as described in the following.

MAC Protocol

Figure 1 shows the TDD multi-code TD/CDMA for multi-rate packet transmission. In the uplink transmission, there are several request access mini-slots of a constant duration in the beginning of the frame. All packet transmission requests are sent in these slots. When there are active video traffic sources, some of the request access slots are reserved as request update slots. The request update slots are used exclusively by video traffic. Since the video traffic has a variable packet generation rate, it has to constantly inform the base station the number of packets arrived at the user terminal during the previous time frame. The request slots are followed by a number of packet transmission slots of a constant duration. The downlink transmission in each frame starts with a control slot. The control slot is a broadcast time slot that consists of a request acknowledgment (ACK) slot to acknowledge that a request from the terminal has been received and a transmission permission (TP) slot to broadcast the packet scheduling for the uplink transmission in the next frame. For the request access slots, a multi-code direct sequence (DS)-CDMA with slotted ALOHA random access protocol is used. Dedicated codes are used for the requests. These codes belong to a common code pool and are not associated with any terminal. The base station broadcasts these codes to the users. When a user is ready to send a request, it chooses randomly a code from the code pool and a request access slot for transmission. More than one user can send their requests in the same request access slot if different codes are used. The user ID will be included in the request. If the request has been received successfully, the base station will broadcast the user ID in the ACK slot in the current frame. If the terminal does not receive its ID in the ACK slot, it will retransmit the request in the next frame. When the terminal has received its ID in the ACK slot, it will listen to the TP slot for transmission permission. The base station uses the TP slot to inform each user in which slot(s) and how many packets are going to be transmitted in each allocated slot. Requests for transmission will be sent at the beginning of each frame for packets arrived at the terminal buffer in the previous frame. The transmission scheduling for next frame is sent in the control slot of the current frame. During the call admission control phase, the base station obtains all the related information of each mobile user, including the traffic class, the initial number of packets at the terminal, the terminal buffer size, the transmission delay requirement, and the BER requirement.

QoS Provisioning

The QoS parameters considered here are transmission delay and BER requirements. In multimedia network, a user requires its packets to be delivered to the destination within a certain time period; otherwise, the packets will be worthless. Here we consider the delay over the wireless link only. The delay requirement can be represented by the *life span* of each packet, which is a set of frames from the moment that the packet is generated to the moment that the delay bound is reached. The *residual delay bound* is the difference between the delay bound and the total accumulated queuing delay and is also called the *time-out value* of the packet. In a wireless environment, the transmission error is caused not only by packet loss due to scheduling and buffer overflow, but also by transmission through the fading dispersive medium. The required BER can be decomposed into two parts. The BER due to wireless transmission will be referred to as transmission BER (TBER) and the BER due to buffer overflow and exceeding delay bound will be referred to as packet loss probability (PLP). Since in the system model all packets have the same length, the rate of bit loss due to scheduling is the same as the rate of packet loss. Consider a single-cell system where there is no inter-cell interference. Given the traffic load in the single cell and the background noise, the maximum number of packets to be transmitted in a time-slot from sources of the same type traffic can be determined in order to achieve the required TBER value. The maximum number decreases with a more stringent TBER requirement. Let $N_{\max}$ denotes the maximum number of packets requiring the least stringent TBER, and $P_{\min}$ denotes the corresponding required received signal power level for each packet. All other required power levels from a more stringent BER can be represented in terms of the minimum power level. For example, the required received power level for the *i*th packet can be represented as $p_i P_{\min}$, where $p_i (\geq 1)$ is a constant. With a known propagation path gain between the mobile and the base station, the received signal power level can be translated into the transmitted power level at the mobile. Mobile users of different service classes have different TBER requirements, which can be guaranteed by controlling the number of simultaneously transmitted packets and the power level of each packet. To transmit packets with different TBER requirements in the same time slot, the received power level for each packet and the number of packets should be determined properly so that the TBER requirements of all packets are met. For example, if a packet can tolerate $(N_{\max} - 1)$ other simultaneously transmitted packets with power $P_{\min}$, it can tolerate $(N_{\max} - 1)/p_i$ other simultaneously transmitted packets with power $p_i P_{\min}$. Transmission of a packet with $P_{\min}$ is referred to as one code slot, and transmission of a packet with $p_i P_{\min}$ requires p_i code slots. The summation of the code slots from all the packets transmitted in each time slot cannot be larger than $N_{\max}$ for satisfactory transmission accuracy.

In summary, the TBER requirements are to be guaranteed by properly arranging simultaneous packet transmission and controlling their received power levels, and the delay and PLP requirements are to be guaranteed by proper packet scheduling. In the following, we focus on the uplink transmission. The downlink transmission is controlled by the base station and can be done in a similar way.

III. Optimization Problem Formulation

For the uplink transmission of the single-cell system, given the total allocated frequency bandwidth and the time duration for the uplink transmission in each frame, the total system resources are fixed. The objective of the optimal scheduler is to maximize the system throughput while guaranteeing QoS requirements. For a single-service system, the throughput can be defined as the average number of packets transmitted in each frame. The definition cannot be directly applied to the multi-service system under consideration, where packets of different services require different transmission BER, resulting in different requirements on the system resources. As the system resources used to transmit each packet is proportional to the received signal power level, one way to define the throughput is the average total transmitted packets weighted by the associated received power levels in each frame, where the minimum received power level for each packet is used to satisfy the required BER. From the viewpoint of a service provider, if the revenue generated by transmission of one packet is proportional to the minimum resources required subject to QoS satisfaction, then maximizing the throughput corresponds to the maximum profit. Given the constant total available resources, the objective is to achieve the maximum throughput. This means maximizing the total code slots of all transmitted packets over all the frames under consideration. Let K denote the total number of time frames to be optimized, M_K denote the total number of packets to be transmitted during the K frames, and L_u denote the total number of time slots in each frame for the uplink. The objective function is to

$$\text{maximize} \sum_{k=1}^{K} \sum_{i=1}^{M_K} \sum_{l=1}^{L_u} p_i S_{i,l}^k$$

where $S_{i,l}^k$ is equal to 1 if packet i assigned to time slot l in frame k and is equal to 0 otherwise. The constraints in the scheduling are: a) once a packet is scheduled into a time slot, it cannot be assigned to any other time slots; b) for the BER requirements, the sum of all occupied code slots in a time slot cannot exceed $N_{\max}$; and c) for the delay requirements, packet i should be scheduled to transmission within its life span, denoted by E_i, which is a subset of $\{1,\ldots,K\}$ frames. E_i is determined based on the transmission delay requirement and the terminal buffer size, so that the delay requirement can be guaranteed and packet loss due to buffer overflow can be avoided. A dynamic programming problem can be formulated for this problem. As shown in Figure 2, we want to place M_K packets into KL_u slots. Each packet has its own life span, with its own required p_i value represented by the size of the packet. Each time slot has a fixed available capacity, i.e. $N_{\max}$. The objective in placing the packets is to maximize the total resource utilization in all the time slots subject to the QoS constraints.

One possible solution to the dynamic programming problem is to search through all possible combinations of packet assignments. This solution has two drawbacks: a) the total number of all possible combinations is large and the number increases exponentially with the total number of the time slots; and b) it is based on the information of all the packets, which corresponds to a scheduling delay up to K frames. As a result, this is an unrealistic formulation since the scheduler has to work in real time. To overcome the drawbacks, probabilistic dynamic programming approach can be used, which schedules packets to transmit in the next time frame based on the information currently available and treat the future packet arrivals as a random process. The objective is to maximize the expected resource utilization over a given number of time frames. However, this approach requires the statistical information of the packet arrivals for each traffic type, which can be difficult to obtain as it depends on characteristics of each service class and user mobility pattern in a practical system. The difficulties in the optimal resource allocation result from the fact that it is necessary to consider the resource allocation over a large number of the frames together, because of the various delay requirements of the packets and the dynamics in new packet arrivals. For a practical solution, we propose the following FPLS packet scheduling.

IV. FPLS Scheduler

The packet scheduling depends on factors such as available resources, number of users, traffic characteristics, and QoS requirements. All these factors have to be weighted and balanced to achieve the fair sharing of the available resources by all the users. As high QoS requirements will result in low resource utilization with bursty traffic, when the system resources are just enough to accommodate the QoS requirements of all the admitted users, over allocating bandwidth to one user will cause a failure in satisfying QoS requirements of all other users. Knowing the rate characteristics, the proposed scheduler allocates the minimum amount of resources to satisfy the QoS requirements. It first decides the priorities for users to transmit their packets, and then determines in which time slots the packets will be transmitted so that the total number of scheduled packets is maximized.

Packet Loss Calculation

Here we discuss how to calculate the number of the packets to be dropped for each user in the current frame for a given fixed total capacity. To focus on the PLP requirement, we first assume that all the packets have the same TBER requirements and therefore represent the system capacity (total resources), denoted by C, as the maximum number of packets that can be transmitted in a frame. The effect of different TBER requirements (represented in terms of the different received power levels) will then be considered in the next subsection. As the scheduler schedules the packet transmission for the next frame, the packets with the time-out value equal to one are referred to as the *most urgent packets* (MUPs). The MUPs must be scheduled for transmission in the next frame; otherwise, they will be dropped. Under the assumption that the terminal buffer size for each user is large enough, the packet loss happens only during scheduling when the number of MUPs exceeds the system capacity. To guarantee the PLP requirements, we need to control the dropped MUPs for each and every user. Even though the overload of MUPs is caused by some bursty traffic sources during their bursty periods, it is fair to share the packet loss according to the PLP requirements of all the users who can tolerate some degree of packet loss. The packets from a user with more strict delay requirement become MUPs sooner and also should be scheduled sooner. Only if all the time slots in the frame are not fully utilized after all MUPs are scheduled, will the scheduler consider non-MUPs in the order of sequentially increased time-out values, starting with the packets having a time-out value equal to two. Consider the radio cell with I users in service. In order to determine the number of the MUPs to be dropped for each user based on the PLP requirements of all the users, first we need to establish a relation between the overall PLP requirements (with respect to all the packets including both MUPs and non-MUPs) and the packet loss probabilities with respect to only the MUPs. Let the integer random variable R denote the rate of MUPs (in packets/frame) from all the users. Let $P_L^{(i)}(>0)$ denote the PLP upper bound required by user i, $1 \leq i \leq I$. Given the MUP traffic load in a frame, R, the conditional MUP packet loss probability for user i is denoted by $\hat{P}_M^i(R)$. Thus, the actual PLP for user i, $\hat{P}_L^i$, is the average number of the lost MUPs divided by the average number of the generated packets in each frame. When the user number I is maximized, the packet loss of all the users will be at their limits, i.e., $\hat{P}_L^i = P_L^i$. If one more user is admitted to the system, the QoS requirements of all the users cannot be satisfied. Since all lost packets are MUPs, we can choose the value of $\hat{P}_M^i(R)$ in such a way that the PLPs of all the users will reach their limits at the same time. The choice of $\hat{P}_M^i(R)$ is described in [17]. This procedure is to achieve fair packet loss sharing. It is *fair* in the sense that the packet losses are arranged according to the PLP requirements of all the users.

Time-Slot Assignment

To consider both TBER and PLP requirements, we propose a bin-packing scheduling algorithm. The original bin-packing problem is a well-known combinatorial problem, which deals with the way of packing a set of indivisible blocks into the minimum number of bins. In the packet scheduling for each time frame, we consider the time slots as bins and the packets as blocks. The size of each bin is $N_{\max}$ code slots, and the size of each block is the number of code slots required for the packet. The number of bins is fixed. We want to pack as many blocks as possible in the bins without splitting and without exceeding the size of each bin. Figure 3 illustrates a heuristic algorithm developed for this problem, where i is the user index, $1 \le i \le I$; l is the time slot index in each frame; p_j is the packet size of user j; p^l is the total size of the scheduled packets in time slot l, and $\mathbf{p} = (p^1, p^2, ..., p^{L_u})$ with an initial value (0,0,…,0). In the algorithm, we assign the size of each packet according to the required received power level. The packets are scheduled according to their urgency. For packets with the same time-out value, we will schedule them according to the number of lost packets for each user calculated using the FPLS method. Since the packet sizes are different, the total system capacity C is unknown before the packet scheduling. To overcome the uncertainty, we start with the assumption of C=0. Then we increase the capacity by one packet gradually until we cannot schedule any more packets. For each capacity increase, all the users have a share (in packet) in using the resource. Let $\Delta\kappa_i$ denote the share of user i. The value of $\Delta\kappa_i$ is calculated based on the FPLS principle. Since a packet has to be transmitted in whole and a fraction of a packet cannot be transmitted, we will allocate the resource to the user with the largest share for transmission of an MUP. The shares are accumulated for all the users respectively. Let κ_i denote the accumulated share of user i. It indicates the difference between the number of packets that should be scheduled according to FPLS and the actual number of scheduled packets. User i has been over-scheduled if $\kappa_i < 0$ and under-scheduled if $\kappa_i > 0$. As a result, κ_i is used as a priority index to determine the order of transmission for user i among all the users. If user i has the largest priority index value, then the capacity increase of one packet is used to schedule an MUP from user i. After that, κ_i is decreased by one. Note that the priority index κ_i is not used to determine the priority for each packet, but rather the priority for each user to transmit their MUPs. When the user with the highest κ value does not have any more MUP to transmit, packets from the user with the next highest κ value will be scheduled. However, this should be recorded and carried over to the scheduling in the next frame.

V. Performance Evaluation

Comparison with WISPER and GPS

The performance of the proposed packet scheduler is demonstrated by comparison with other packet scheduling schemes. Unfortunately, there are very few previously proposed MAC protocols operate in the hybrid TD/CDMA scenario. Also, most packet schedulers are analyzed for only certain traffic types, such as the on-off voice traffic model. Here we consider the comparison with (a) the WISPER [14] which is a MAC protocol with packet scheduling for multimedia traffic in the system model similar to the one considered here and (b) the discretized GPS which is a well known work conserving protocol for wire-line bandwidth allocation. The three packet scheduling schemes are compared via computer simulation using the same system model and guaranteeing the same QoS requirements for the same traffic flows. The original GPS assumes that the server can serve multiple sessions simultaneously and that the traffic is infinitely divisible. However, in the hybrid TD/CDMA system, the slots are defined both in time and code domains. Hence, discretized GPS is considered for the large common pool of available resources. Although many methods have been reported recently for choosing the GPS bandwidth allocation weight factors, a fair and efficient algorithm has yet to be developed. Here, the GPS weights are chosen to be proportional to the effective bandwidths of the traffic flows [18]. The effective bandwidth is the bandwidth required for a given performance objective. For real-time traffic without buffering, the effective bandwidth for user i is given in [19]. However, for non real-time traffic with delay and PLP requirements, the calculation of effective bandwidth is very complex. In [18], the effective bandwidth for lossless multiplexing is given. In the following comparison for non real-time traffic, the bandwidth required for homogeneous traffic to guarantee the given QoS requirements is obtained by computer simulation, and then this bandwidth is used as the weighting factor in the simulation with heterogeneous traffic. The three schemes to be compared (i.e., the FPLS, WISPER, and GPS) schedule packets based on different principles. The WISPER does not consider the PLP requirement and, therefore, the scheduling remains the same for different PLP requirements. Transmission order is determined according to the packets' time-out values and the number of packets ready for transmission at each mobile terminal. Packets with equal or similar BER requirements are transmitted in the same slots. As to the GPS, using the effective bandwidth as the weighting factor takes into account both delay and PLP requirements. However, GPS does not take the current traffic load into consideration. The packet loss happens mainly during bursty periods. The main feature of the FPLS scheduler is to even out the packet loss over a large time period for each user and to let all the users share the packet loss depending on their PLP requirements. When a user is in a bursty period and the system capacity is not sufficient to accommodate the total traffic requirement, all other users will have a share of packet loss with this user; on the other hand, this user will share the loss when any other users are in their bursty periods. Thus, instead of a large number of lost packets during a bursty period, the packet loss for each user is smoothed over a longer time period. Based on the real-time information of the transmission rate requirement, the FPLS scheduler allocates the resources to the users according to their actual needs and QoS requirements. The difference between the GPS and FPLS is illustrated in Figure 4 for two users as an example, where packet loss happens whenever the allocated rate is below the input traffic rate. Using the GPS, there will be little (or no) packet loss for a non-bursty traffic flow even though the user may tolerate packet loss to some degree. In the FPLS, a non-bursty traffic flow will experience packet loss in order to give more resources to other traffic flows at their bursty periods, as long as the user's PLP requirement can be guaranteed. In GPS, when the bandwidth assigned to a user cannot be used, it will be shared among all other users, and the user will not be compensated. In FPLS, when a user has dropped too many packets, it will be compensated in the future by keeping a low packet dropping. This gives the fair share of system resource among all users. In this way, FPLS scheduling is expected to achieve high resource utilization than the GPS. In fact, the FPLS can be considered as an improved GPS scheme. It dynamically allocates the bandwidth frame by frame. The bandwidth is allocated optimally among all users to satisfy their PLP requirements. For a long message period, it should be possible to find a particular weighting factor assignment (related to the PLP) for GPS to achieve the same performance as FPLS, and the FPLS is one way to find the weighting factors in the resource allocation.

The Simulation Environment

Consider a single-cell system with two traffic types — voice and video. Each time frame is 10 ms in length. The on-off model is used to simulate voice traffic. During the on-state, one packet is generated in each frame which is equivalent to a rate of 100 packets per second. The video traffic has a variable rate, which varies among four rates (0,4,8,12), in packets per frame, with a probability of $(\frac{1}{6},\frac{1}{3},\frac{1}{3},\frac{1}{6})$, respectively. Both real-time and non real-time traffic types are considered. For the case of real-time traffic, the time-out value is chosen to be 1 frame for both voice and video traffics. All the packets are MUPs and must be transmitted immediately or will be dropped. For non real-time case, the time-out value is 2 frames for voice traffic and 10 frames for video traffic. The simulation parameters are given in Table 1. Using the FPLS scheduling requires the rate distribution $\bar{r}_i(n)$ given R=n, which can be very complex to calculate. For the real-time traffic, Figure 5 shows the rate distributions of each voice user and video user, respectively, with a total of 200 voice users and 15 video users, based on computer simulation. It is observed that the distribution for user i can be approximated by $\bar{r}_i(n) \approx n\frac{r_{i\max}}{R_{\max}}$ to reduce the implementation complexity of the scheduling algorithm, where $r_{i\max} = 1$ and 12 packets/frame for voice and video users, respectively, and $R_{\max} = 380$ packets/frame is the maximum number of MUPs generated in a frame. In the following simulations, the approximation is used, and the packets are generated for each individual (voice or video) user in each frame. In packet scheduling, all the voice packets are scheduled together; so are all the video packets. In other words, packets from different homogeneous traffic sources are not differentiated in the scheduling to reduce computer simulation time. In the simulations, using each scheduling method, the maximum numbers of users that can be supported by the system under the QoS constraints are found by searching through all possible combinations of the user numbers with different traffic types. For simplicity, the resource overhead necessary for signaling and control in MAC is not considered.

Real-Time Traffic

Figure 6 shows the maximum numbers of the voice and video users that can be supported by the system with QoS satisfaction. FPLS outperforms GPS and WISPER in most situations. When there is only one traffic type (voice or video) in the system, the total number of lost packets is the same for all three algorithms and, therefore, there is no performance difference among the algorithms. However, if there exists a mixture of the two traffic types in the system with different QoS requirements, FPLS performs better because of the more effective statistical multiplexing. Since all traffic is real-time, all arrived packets are MUPs, which means all packets are either transmitted or dropped. The average rate of MUPs is the same as the average rate of the arrived packets. The system capacity depends mainly on the proper choice of packet dropping. FPLS balances the dropped packets between the two traffic classes constantly according to their PLP requirements to achieve the best performance. GPS uses PLP to obtain the effective bandwidth as the weights, the effective bandwidth itself is an approximation and it will not be perfect for all combination of traffics. WISPER does not take PLP into consideration in the scheduling. Therefore, we have FPLS outperforms both GPS and WISPER and GPS gives better result than WISPER. Figure 7 shows the corresponding resource utilization efficiency with guaranteed QoS requirements for all the users. The FPLS provides a higher resource utilization to accommodate more users than the GPS and WISPER. It is expected that, with an increase of the system capacity (e.g. L_u), the performance improvement of the FPLS over the GPS and WISPER will increase because a higher statistical multiplexing gain can be achieved with more voice and video users in service.

Non Real-Time Traffic

The less stringent delay requirements of the traffic are expected to increase the resource utilization efficiency. Figure 8 shows the maximum numbers of the voice and video users can be supported by the system using FPLS, GPS, and WISPER, respectively, for the same QoS requirements. FPLS can provide service for more users. When the maximum number of fourteen video users is reached, FPLS can still provide service for about forty voice users while GPS and WISPER can barely support any voice users. Figure 9 shows the corresponding resource utilization efficiency. The FPLS scheduling clearly outperforms both GPS and WISPER. In particular, when the difference between traffic loads is large, WISPER gives a higher priority to the traffic with high load, thus causes high packet loss for users with the lower traffic load. In the WISPER, as the order of packet transmission remains the same for different packet loss requirements, the system capacity is limited by the most stringent packet loss requirement. This problem is taken care in the FPLS where all QoS requirements are used in determining the order for transmission; thus the resource utilization is greatly improved. As the number of the traffic types and/or the difference in the PLP requirements increase, the further performance improvement achieved by the FPLS scheduling over that of the WISPER is expected. On the other hand, in the GPS scheduling, the allocated resource to each traffic type is proportional to its effective bandwidth and depends on the total traffic load in the frame. Since video users can tolerate a large delay, effective statistical multiplexing among video users can be achieved in both time and code domains; however, the small number of video users in service limits the multiplexing gain. Voice users have the most stringent delay requirement. The statistical multiplexing is achieved mainly in the code domain. However, when there are a much larger number of voice users in service (due to the relatively low rate from each voice user) than that of video users, a larger multiplexing gain can be achieved. As a result, in Figure 9 the resource utilization efficiency decreases when the number of video users increases (corresponding to a significant decrease in the number of voice users, as shown in Figure 8). The resource utilization efficiency fluctuates when the number of video users is close to the maximum value. This is because the large number of video users dominates the resource usage and the voice traffic may not get its fair share in the resources from frame to frame due to the work-conserving discipline and discrete nature of the GPS protocol used in the simulation. As a result, the QoS for the video users can be higher than required, which translates to a reduced resource utilization efficiency.

VI. Conclusion

In this paper, we investigate resource allocation at the link layer for wireless multimedia CDMA communications. The proposed MAC protocol achieves statistical multiplexing in both time domain and code domain. Transmission accuracy requirement over the wireless link is guaranteed by proper transmit power allocation, while packet loss probability and delay requirements are guaranteed by proper packet scheduling. The resource allocation problem is formulated as an optimization problem and a sub-optimal real-time FPLS scheduler for the problem is developed. Using the bin-packing principle, the FPLS scheduler allocates a minimum amount of resources to each user for QoS provisioning by letting each user have a fair share in packet loss and by assigning a minimum required transmit power level and having a maximum multiplexing in the code domain for each time slot, based on the transmission rate statistics and real-time traffic load information. Simulation results demonstrate that the FPLS scheduler outperforms both WISPER and discrete GPS scheduling schemes in terms of resource utilization. The FPLS scheduler requires the rate statistics to calculate the packet loss rate over a long time period. For traffic flows with a relatively short transmission period and/or unknown traffic rate distribution, the FPLS scheduler should be used in combination with other scheduling methods. The extension of this research to multiple-cell systems is currently under investigation.

References

[1] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver Jr., and C. E. Wheatley III, "On the Capacity of a Cellular CDMA System", *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, May 1991, pp. 303-312.

[2] M. Soleimanipour, W. Zhuang, and G.H. Freeman, "Optimal Resource Management in Multimedia WCDMA Systems", *Proceedings of IEEE Global Communications Conference (Globecom'00)*, Nov.-Dec. 2000, pp. 1544-1547.

[3] L. C. Yun and D. G. Messerschmitt, "Variable Quality of Service in CDMA Systems by Statistical Power Control", *Proceeding 1995 IEEE International Conference on Communications. Gateway to Globalization*, vol. 2, pp. 713-719.

[4] S. Yao and E. Geraniotis, "Optimal Power Control Law for Multi-media Multi-rate CDMA Systems", *Proceeding IEEE 46th Vehicular Technology Conference (VTC'96)*, vol. 1, 1996, pp. 392-396.

[5] R. D. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems", *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, Sep. 1995, pp. 1341-1347.

[6] Jens Zander, "Performance of Optimum Transmitter Power Control in Cellular Radio Systems", *IEEE Transactions on Vehicular Technology*, vol. 41, no. 1, Feb. 1992, pp. 57-62.

[7] A. Sampath, P. S. Kumar and J. M. Holtzman, "Power Control and Resource Management for a Multimedia CDMA", *Proceedings IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PRMIC'95)*, vol. 1, 1995, pp. 21-25.

[8] R. D. Yates and Ching-Yao Huang, "Integrated Power Control and Base Station Assignment", *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, Aug. 1995, pp. 638-644.

[9] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case", *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, Jun.1993, pp. 344-357.

[10] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple Node Case", *IEEE/ACM Transactions on Networking*, vol. 2, no. 2, Apr. 1994, pp. 137-150.

[11] D. J. Goodman R. A. Valenzuela K. T. Gayliard and B. Ramamurthi, "Packet Reservation Multiple Access for Local Wireless Communications", *IEEE Transactions on Communications*, vol. 37, no. 8, Aug. 1989, pp. 885-890.

[12] M. J. Karol, Z. Liu and K. Y. Eng, "Distributed-Queueing Request Update Multiple Access (DQRUMA) for Wireless Packet (ATM) Networks", *Proceedings of IEEE International Conference on Communications (ICC'95)*, vol. 1, Jun. 1995, pp. 1224-1231.

[13] A. E. Brand and A. H. Aghvami, "Multidimensional PRMA with Prioritized Bayesian Broadcast -- A MAC Strategy for Multiservice Traffic over UMTS", *IEEE Transactions on Vehicular Technology*, vol. 47, no. 4, Nov. 1998, pp. 1148-1161.

[14] I. F. Akyildiz D. A. Levine and I. Joe, "A Slotted CDMA Protocol with BER Scheduling for Wireless Multimedia Networks", *IEEE/ACM Transactions on Networking*, vol. 7, no. 2, Apr. 1999, pp. 146-158.

[15] V. Huang and W. Zhuang, "Optimal resource management in packet-switching TDD CDMA systems", *IEEE Personal Communications*, Dec. 2000, pp. 26-31.

[16] G. J R Povey and M. Nakagawa, "A Review of Time Division Duplex - CDMA Techniques", *1998 IEEE 5th International Symposium on Spread Spectrum Techniques and Application - Proceedings*, vol. 2, 1998, pp. 630-633.

[17] V. Huang and W. Zhuang, "Fair Packet Loss Sharing (FPLS) Bandwidth Allocation in Wireless Multimedia CDMA Communications", *Proc. 2001 International Conference on Third Generation Wireless and Beyond (3Gwireless'01)*, pp. 198-203.

[18] A. Elwalid and D. Mitra, "Design of Generalized Processor Sharing Schedulers Which Statistically Multiplex Heterogeneous QoS Classes", *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3, 1999, pp. 1220-1230.

[19] M. Schwartz, *Broadband Integrated Networks*, Prentice-Hall International, Inc, 1996.

Authors' Biographies

Vincent Huang received the M.Sc. degree in electrical engineering in 1994 from the Chalmers University of Technology, Gothenburg, Sweden. He is currently working toward the Ph.D. degree at the University of Waterloo, Canada. During 1994-1999, he was a research assistant in the Department of Signals and Systems at the Chalmers University of Technology. His current research interests are MAC protocols and resource management for multimedia wireless communications.

Weihua Zhuang received the Ph.D. degree in electrical engineering in 1993 from the University of New Brunswick (Canada). Since 1993 she has been a faculty member at the University of Waterloo, Canada, where she is a Professor in the Department of Electrical and Computer Engineering. Her current research interests include resource management and wireless networking for multimedia personal communications. She received the Premier's Research Excellence Award (PREA) in 2001 from the Ontario Government.

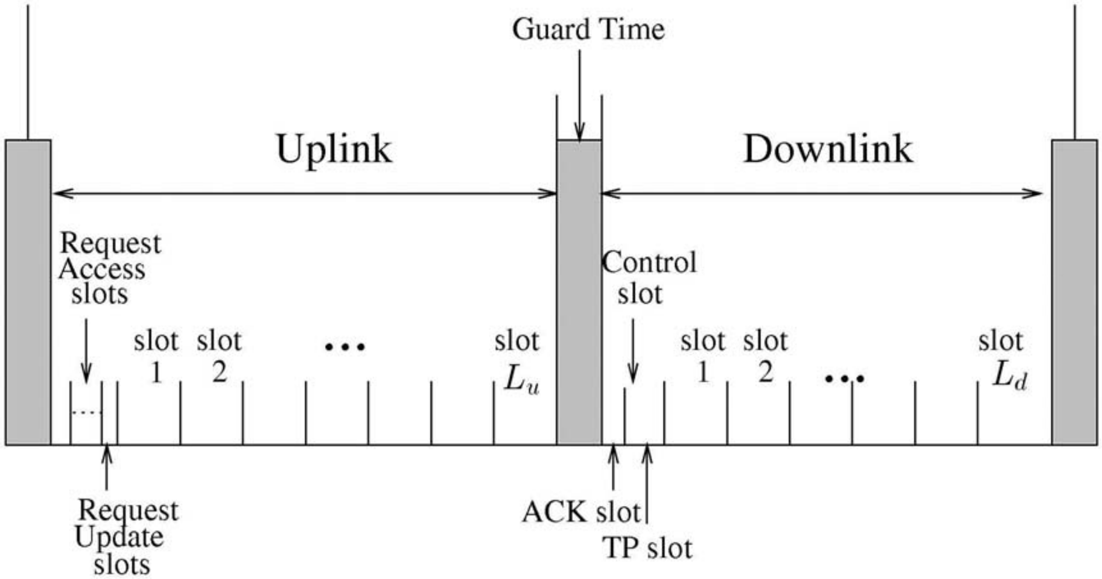

Figure 1 Time slots of uplink and downlink

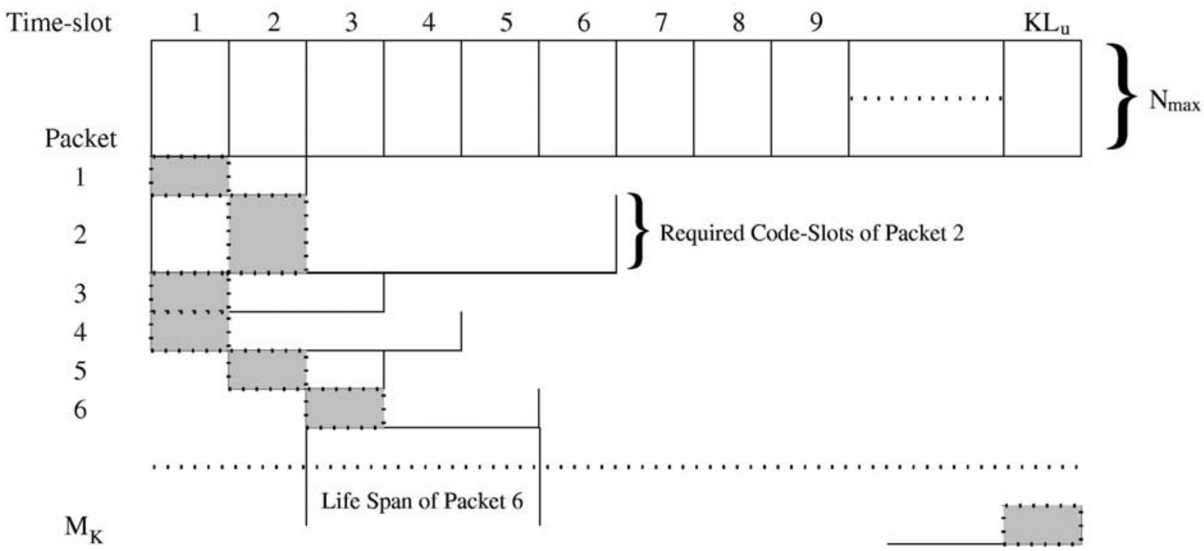

Figure 2 Dynamic programming problem formulation

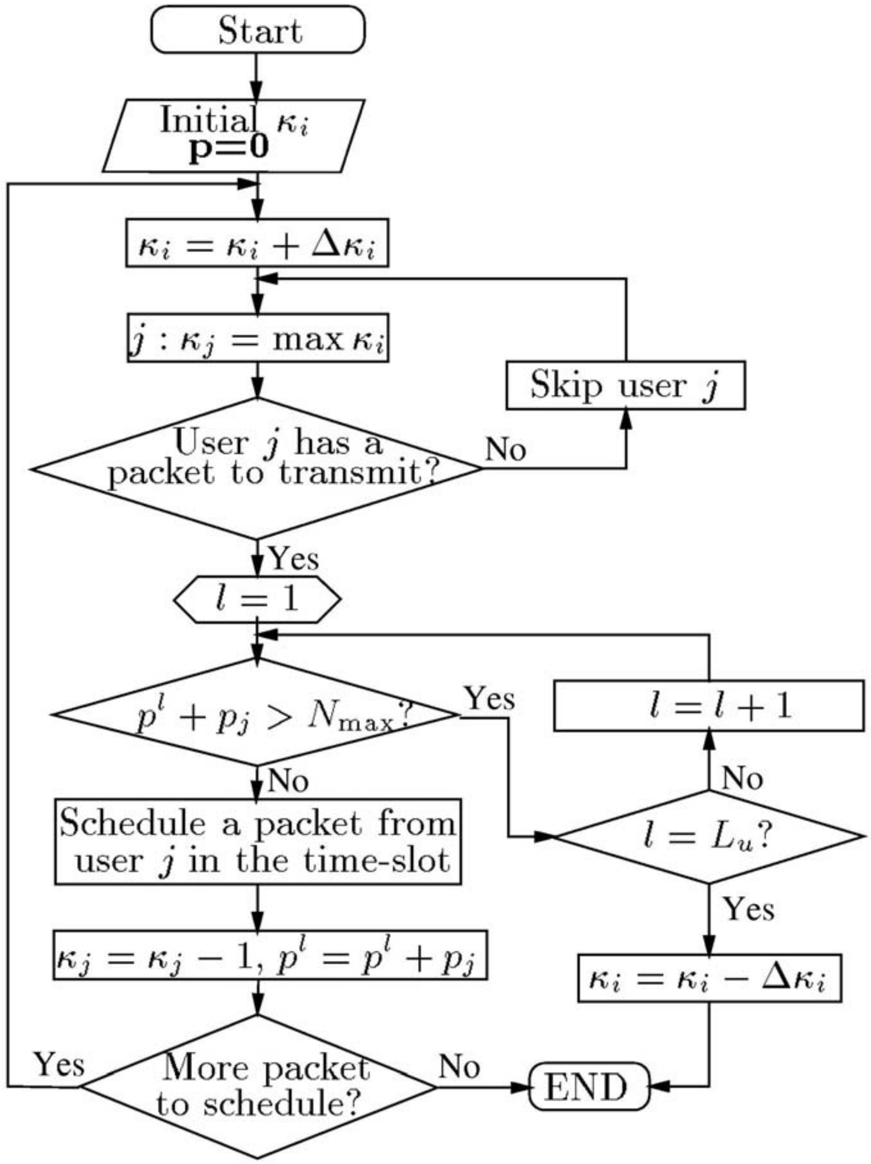

Figure 3 The proposed FPLS bin-packing packet scheduling algorithm for each frame.

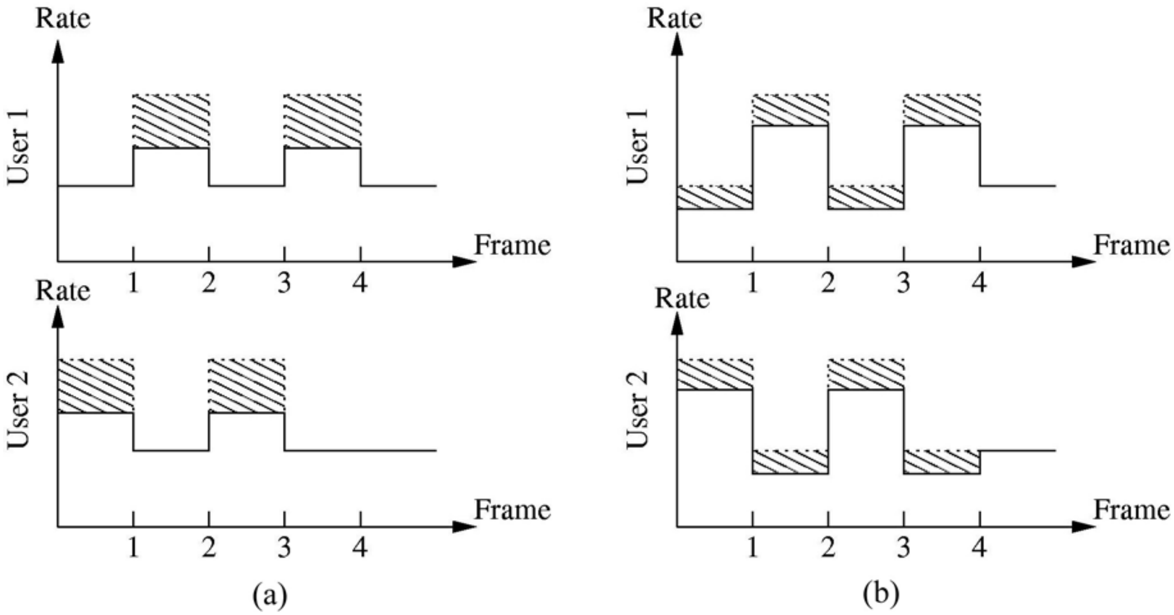

Figure 4 Comparison of resource allocation using (a) GPS and (b) FPLS. Solid lines represent the allocated rates and dotted lines represent the input traffic rates.

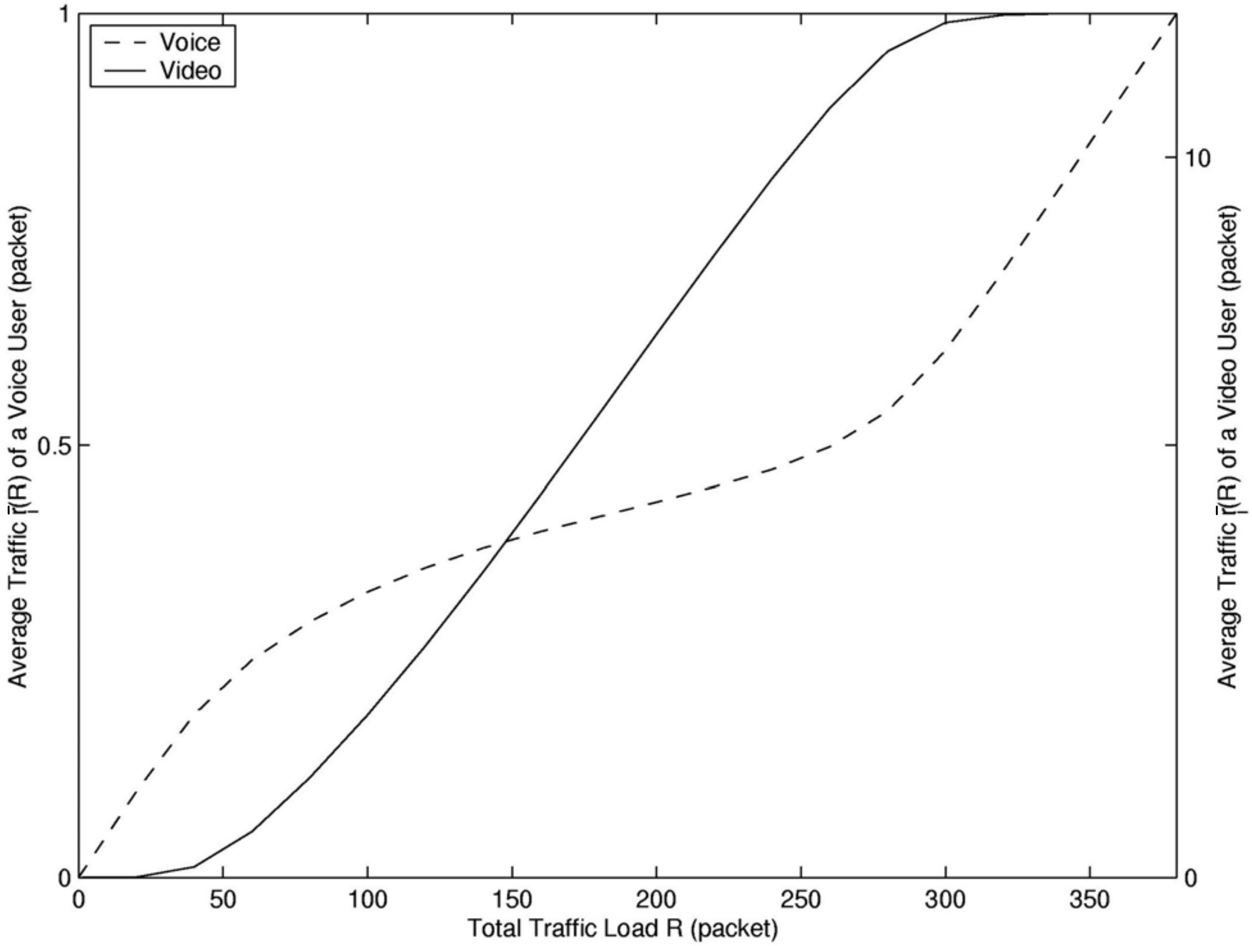

Figure 5 The conditional MUP rate distribution for each voice and video user given the total MUP traffic load.

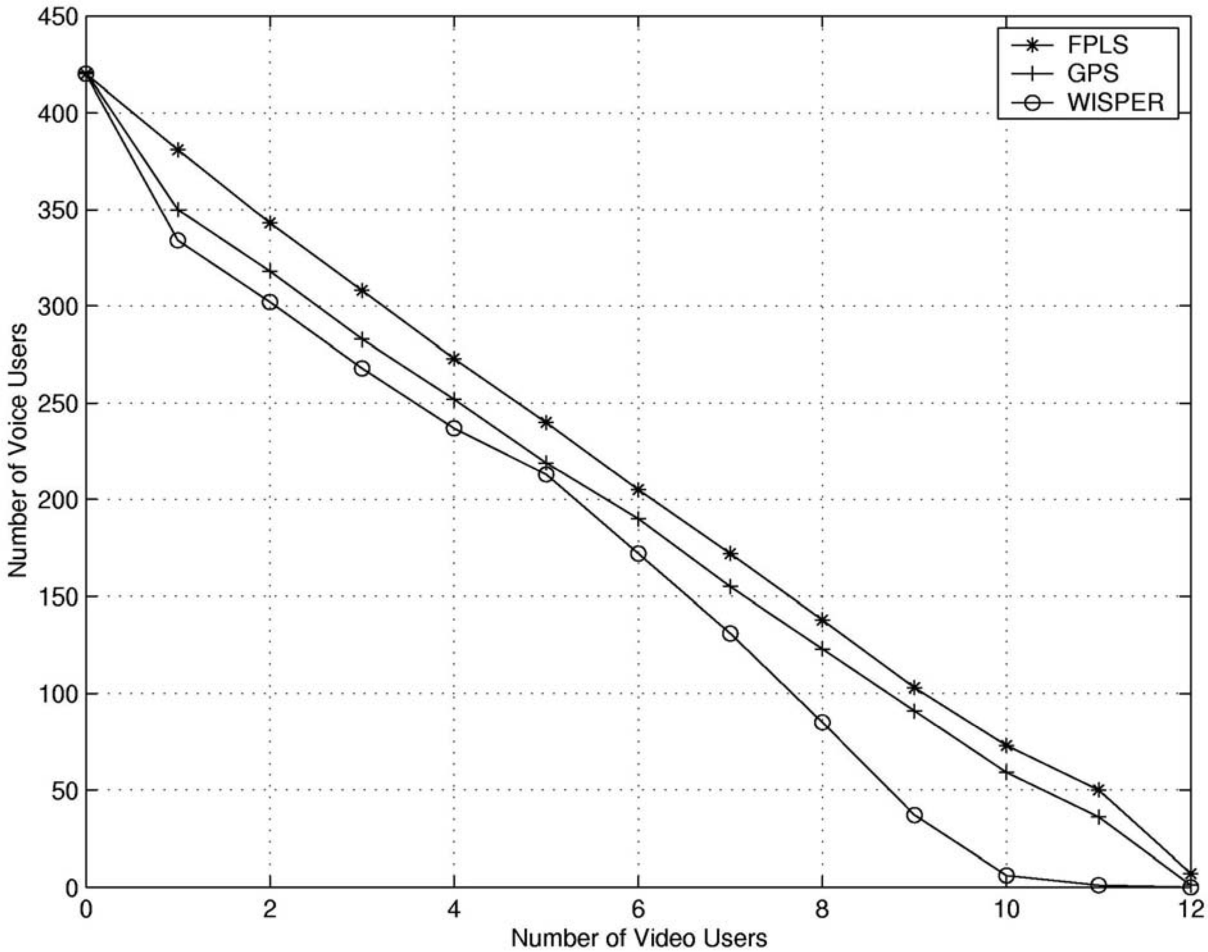

Figure 6 The number of real-time traffic users supported using FPLS vs WISPER and GPS.

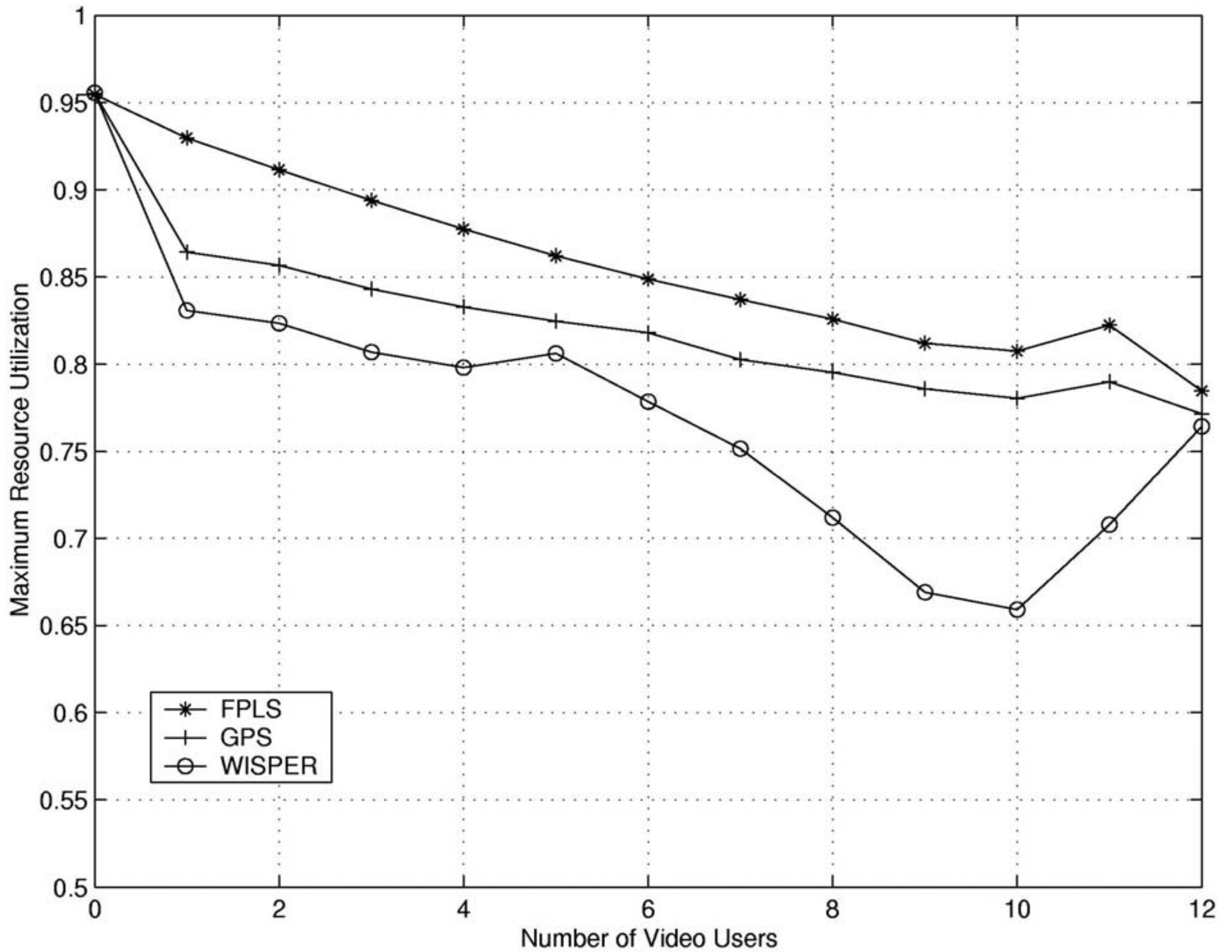

Figure 7 The percentage of system resources utilized for real-time traffic using FPLS vs WISPER and GPS.

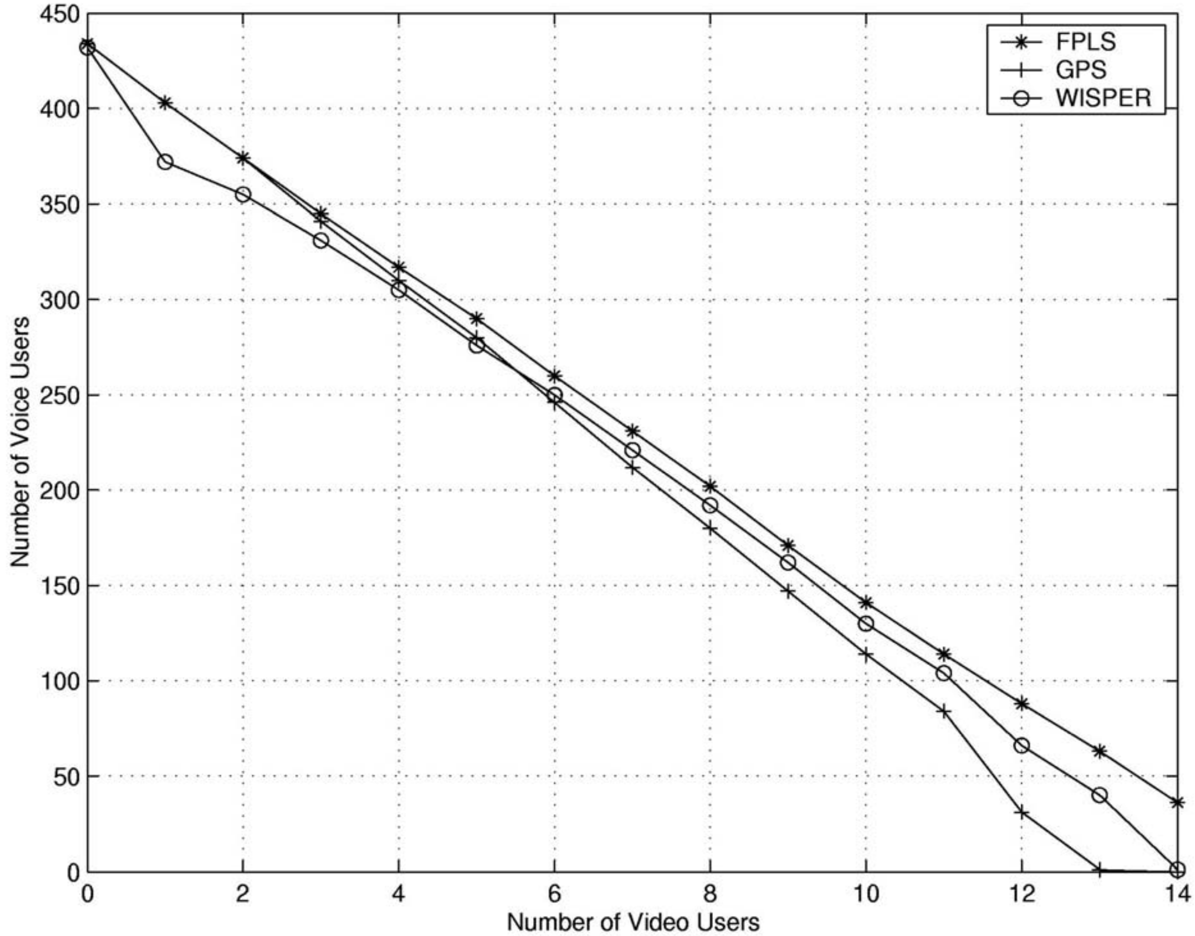

Figure 8 The number of non real-time traffic users supported using FPLS vs WISPER and GPS.

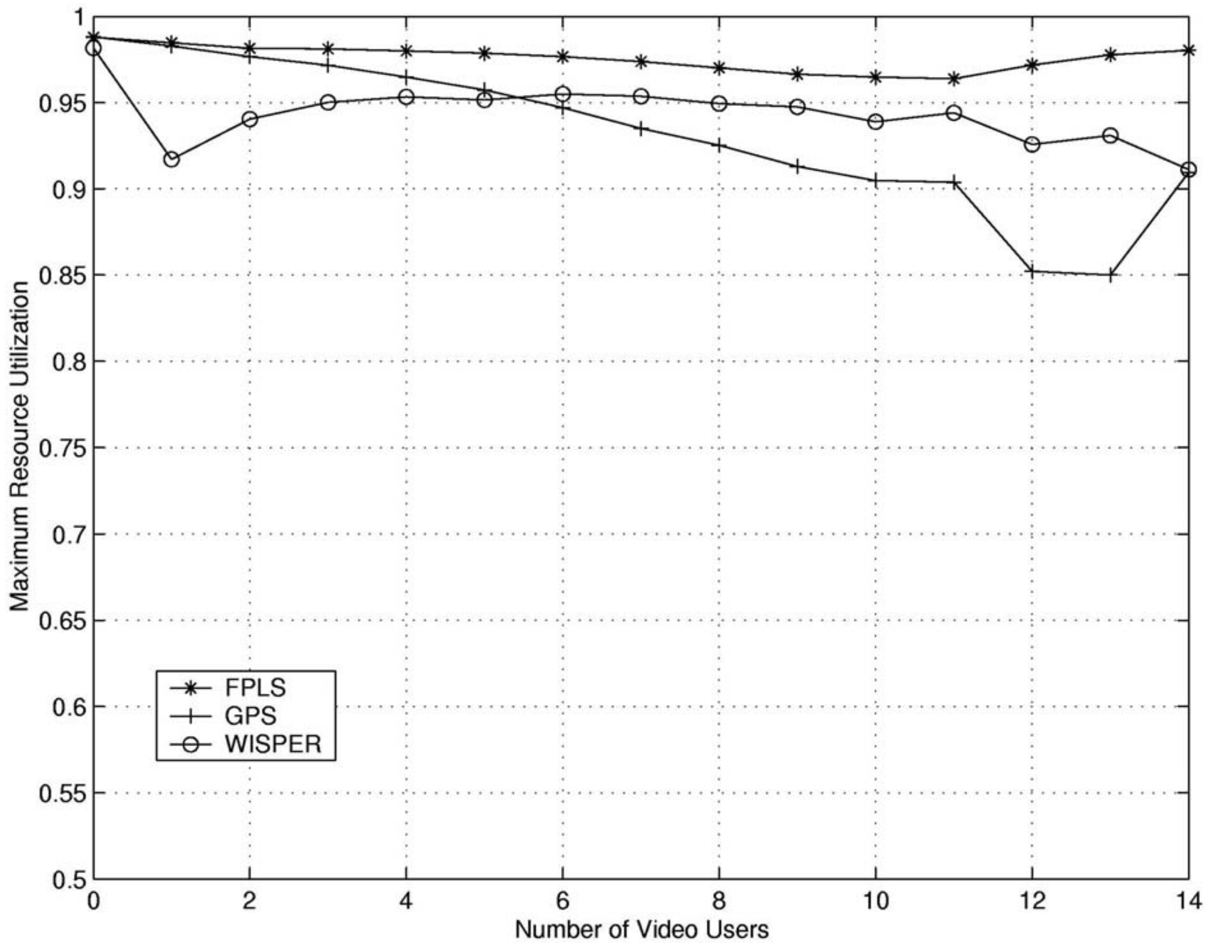

Figure 9 The percentage of system resources utilized for non real-time traffic using FPLS vs WISPER and GPS.

Table 1 System parameters used in the FPLS vs GPS and WISPER simulations

Parameter		Value
Number of time-slots per frame (L_u)		8
Maximum number of code slots per time-slots ($N_{\max}$)		22
Simulation time (frames)		10000
Voice traffic:	Time-out value (D_i)	1 or 2
	Required received power (p_i)	1
	Required packet loss probability upper bound ($P_L^{(i)}$)	10^{-2}
	Average talk spurt length (frames)	10
	Average silent period (frames)	15
Video traffic:	Time-out value (D_i)	1 or 10
	Required received power (p_i)	1.9
	Required packet loss probability upper bound ($P_L^{(i)}$)	10^{-2}
	Average rate (packets/frame)	6
	Peak rate (packets/frame)	12