QoS-Oriented Access Control for 4G Mobile Multimedia CDMA Communications

Vincent Huang and Weihua Zhuang, University of Waterloo, Canada

ABSTRACT

In this article an efficient medium access control protocol with fair packet loss sharing packet scheduling is proposed for wireless code-division multiple access communications. The proposed MAC protocol exploits both time-division and code-division statistical multiplexing. The FPLS scheduler uses the information of traffic rate distribution and quality of service requirements to assign priorities to the users and determines an efficient combination of the packets for transmission in the time slots of each frame, so the number of the served users is maximized under the QoS constraints. Simulation results demonstrate the effectiveness of the FPLS scheduler, in comparison with other previously proposed scheduling algorithms.

INTRODUCTION

The demand on information exchange has pushed the development of wireless communication systems in an unprecedented pace. The transition of the second generation (2G) to the third generation (3G) has begun. System designers are already thinking about fourth-generation (4G) technology. Although there are no solid specifications for 4G systems yet, it is clear that 4G will support higher data rates than 3G and will be able to more efficiently integrate different modes of wireless communications. The 3G wireless systems will provide high data rate up to 2 Mb/s and support a broad range of multimedia services including voice, data, and video to mobile users. In 4G systems, data rates are expected to reach as high as 20 Mb/s. Because wireless systems have very scarce bandwidth of available frequency spectrum, the limited resources have to be used efficiently to provide satisfactory services to the users.

Multimedia information sources can exhibit highly bursty traffic rates. Packetized transmission over wireless links makes it possible to achieve a high statistical multiplexing gain. 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 rate characteristics. Due to the heterogeneous nature of multimedia traffic flows, the traditional voice-based medium access control (MAC) protocols do not perform well in a multimedia environment. A flexible MAC protocol that can efficiently accommodate multimedia traffic is required. One important MAC issue is the packet scheduling which determines the order of packet transmissions. Most packet scheduling strategies, such as first-in-first-out (FIFO), round-robin, and generalized processor sharing (GPS) [1], were originally proposed for wireline networks. Random access protocols have been widely used in the past for wireless communications [2–4]. More recently, a MAC protocol with bit error rate (BER) scheduling called WISPER is proposed in [5] for code-division multiple access (CDMA) communications, 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 the 3G system proposals, mobile terminals use random access for sending the transmission requests, to which short data bursts can be appended. For other data transmissions, the base station assigns dedicated channels to the users when the resources are sufficient. The order of channel assignments depends on the time moments when the requests are received and the priorities associated with the traffic classes. The users keep the channels as long as they have packets to transmit. As there is no specific packet scheduling, efficient statistical multiplexing cannot be achieved at the packet level. For high resource utilization, MAC should take the current packet

 0163-6804/02/$17.00 © 2002 IEEE

flow loads and the users' QoS requirements into account. A MAC protocol with packet scheduling needs to be further developed to achieve efficient packet-level statistical multiplexing under QoS constraints.

In this article we propose a MAC protocol with fair packet loss sharing (FPLS) scheduling for 4G wireless multimedia communications. The MAC protocol exploits both time-division and code-division multiplexing for efficient resource utilization. FPLS is a QoS requirement based packet scheduling algorithm. The objectives of the scheduling are to provide QoS guarantees in terms of transmission delay and accuracy and to maximize the system resource utilization. In a wireless environment, a packet is expected to be delivered to the destination within a required time frame and with certain accuracy. Any violation of these two requirements will cause the packet to be useless and therefore be discarded. Since 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 loads 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 the users is essential. The main features of the FPLS scheduler are that:

- The packet losses are shared fairly among all the users according to each and every user's QoS requirements.
- Both the number of users supported by the system (with QoS provisioning) and the resource utilization are maximized.

With FPLS, the bandwidth is shared among all the users in such a way that, when the QoS requirements are guaranteed for one user, they are also 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.

SYSTEM MODEL

We consider a hybrid time-division/code-division multiple access (TD/CDMA) wireless system with packetized transmission [6]. The system operates in the time division duplex (TDD) mode, as TDD can best accommodate asymmetric traffic between the uplink and downlink. 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, and each packet requires a time slot for transmission. Packet transmission from and to mobile users is synchronized in time.

The QoS parameters considered are transmission delay (over the wireless link) and BER requirements. The delay requirement can be represented by the *life span* of each packet, which is a set of frames from the moment the packet is generated to the moment the delay bound is reached. The *time-out value* of a packet is the difference between the delay bound and the total accumulated queuing delay up to the current frame. 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 that due to buffer overflow and exceeding the delay bound will be referred to as *packet loss probability* (PLP).

Consider a single-cell system where there is no intercell interference. The maximum number of packets to be transmitted in a time slot from sources of the same type of traffic can be determined in order to achieve the required TBER value. The maximum number decreases with a more stringent TBER requirement. Let $N_{\max}$ denote the maximum number of packets requiring the least stringent TBER, and $P_{\min}$ denote the corresponding required received signal power level for each packet. All other required power levels for more stringent BERs can be represented in terms of the minimum power level. For example, the required received power level for the ith packet can be represented as $p_iP_{\min}$, where p_i (≥ 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 $p_iP_{\min}$. Transmission of a packet with $P_{\min}$ is referred to as one code slot, and transmission of a packet with $p_iP_{\min}$ 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 over the wireless medium.

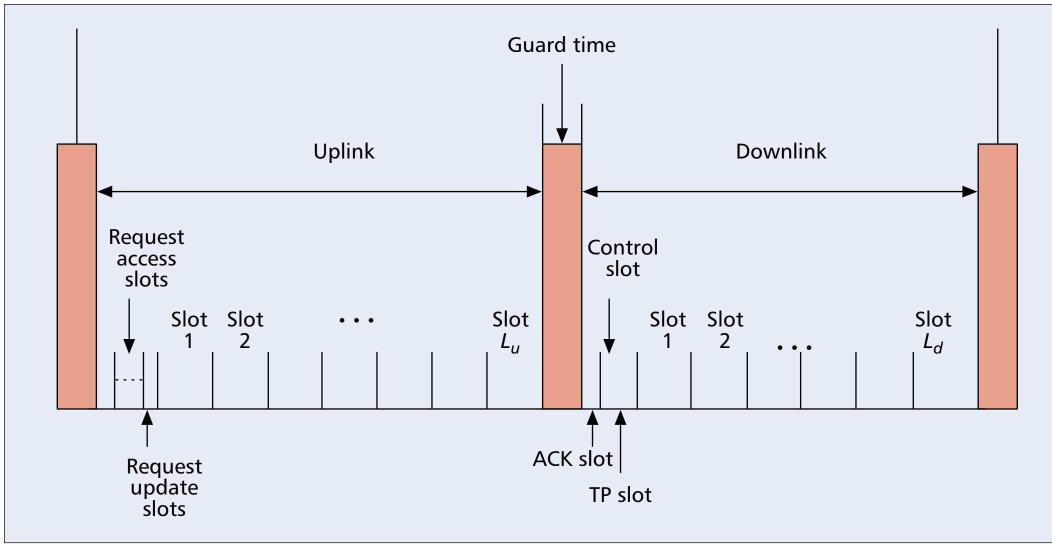
In summary, the TBER requirements are to be guaranteed by properly arranging simultaneous packet transmissions and controlling their received power levels, while the delay and PLP requirements are to be guaranteed by proper packet scheduling. The decision on packet transmission in each time slot for both uplink and downlink is made at the base station and is broadcast to the mobile users.

MAC PROTOCOL

For multimedia communications, the MAC protocol has to be able to accommodate packets with different QoS requirements. Figure 1 illustrates the TDD TD/CDMA medium access for

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.

When a mobile terminal 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 terminal can send its request in the same request access slot if different users use different codes.



■ **Figure 1.** *Time slots for the uplink and downlink in each frame.*

multirate packet transmission. In the uplink transmission, there are several request access mini-slots of a constant duration in the beginning of the frame. When there are active video traffic sources, some of the request access slots are reserved as request update slots. Since each video traffic source 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 update slots are assigned to active video users in order to avoid packet transmission collision. The request slots are followed by a number of packet transmission slots of a constant duration longer than the request slot 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) subslot to acknowledge the requests from the terminals that have been successfully received and a transmission permission (TP) subslot to broadcast the packet scheduling result for the uplink transmission in the next frame.

For the request access slots, a direct sequence (DS)-CDMA with slotted ALOHA random access protocol is used. Dedicated codes are used for the requests. The base station broadcasts these codes to the users. When a mobile terminal 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 terminal can send its request in the same request access slot if different users use different codes. 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, the request process repeats 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 terminal in which time slot(s) and how many packets are going to be transmitted in each allocated time 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 result for the next frame is sent in the control slot of the current frame.

Voice traffic can be modeled by the two-state on-off model [7]. In the on state, packets are generated at a constant rate. The user need only send a request when the voice state switches from off to on. The frame duration is determined in such a way that each voice source generates one packet in each frame during a talk spurt. As a result, the base station automatically allocates the resources for each on state voice user to transmit one packet in each frame. The allocated time slot may vary from frame to frame depending on packet scheduling for that frame. The terminal listens to the broadcasting in the TP slot in each frame to know the time slot for transmission. When the talk spurt is over, the resources reserved for the user for one packet will be wasted, which informs the base station that the user has gone into a silent period (off state). The process repeats when the next talk spurt starts. The protocol is similar to packet reservation multiple access (PRMA) [2] in terms of packet reservation after the transmission of the first packet in each talk spurt. However, the contention occurs only in the request slots. Since a request slot is much shorter than a packet slot, less bandwidth is wasted if a collision occurs. Furthermore, the reserved time-slot for transmission is not fixed. This allows efficient resource allocation. For the bursty data traffic, the user sends a request when a data burst arrives at the terminal. In the request, the number of data packets will be included. The request for video traffic will be sent to the base station in a way integrating those of the voice and data traffic. In the beginning of the transmission, the video user will use a request slot to send its first request in the same way a data user does. It informs the base station how many packets have arrived in the terminal buffer. When the request is received by the base station, a request slot is converted to a request update slot and a PN code is assigned to the user. Using the request update slot, the user sends only the changes of the number of packets arrived in the previous frame from the frame before. Based on the information in the first request slot and then the request update slots, the base station will allocate resources to the user and broadcasts the allocation information in the TP slot. Upon completion of the call, the user sends a termination request in the request update slot to the base station to indicate the end of the transmission. Thus, by avoiding unnecessary random requests, collisions in contending for the request slots can be greatly reduced. In the following, we focus on the uplink transmission. The downlink transmission is controlled by the base station and the decision process can be carried out in a similar way.

THE FPLS SCHEDULER

The design of a packet scheduler 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 fair sharing of the available resources among all users. Since 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 admitted users, overallocating bandwidth to one user will cause failure to satisfy the 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

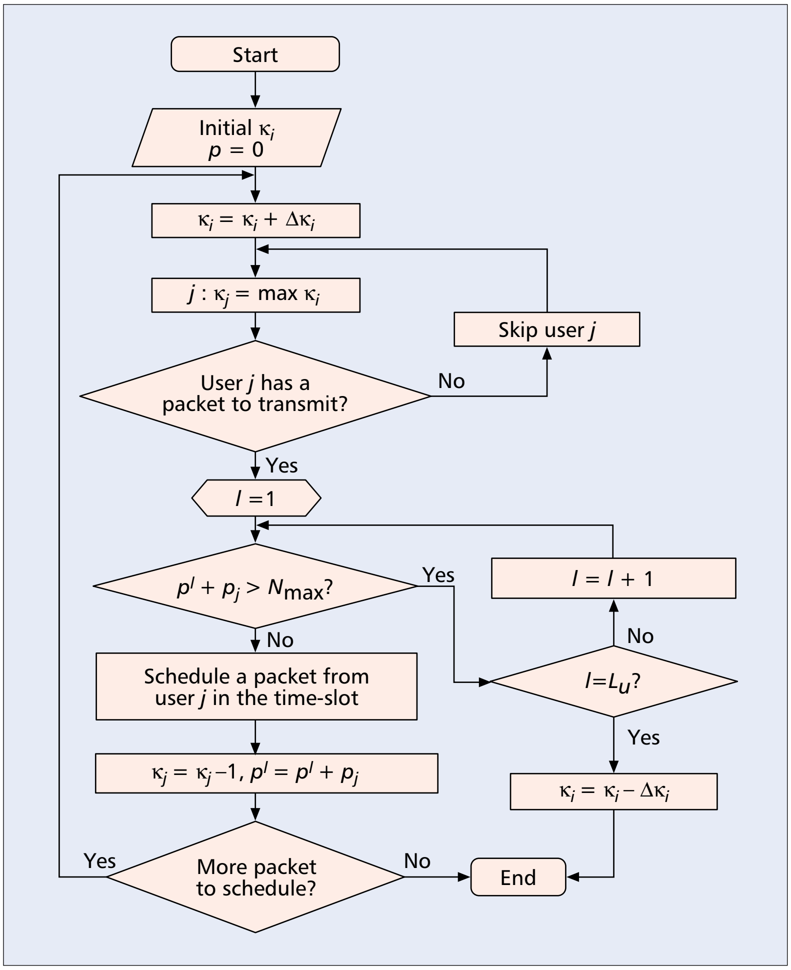
Here we discuss how to calculate the number of packets to be dropped for each user in the current frame for a given fixed total capacity. To focus on the PLP requirement, we first assume that all packets have the same TBER requirements and therefore represent the system capacity (total resources), denoted 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 timeout 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, packet loss happens only during scheduling when the number of MUPs exceeds system capacity. To guarantee 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 users who can tolerate some degree of packet loss. Packets from a user with more strict delay requirements 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 order of sequentially increased timeout values, starting with packets having a timeout value equal to two.

Consider a radio cell with N users in service. In order to determine the number of MUPs to be dropped for each user based on the PLP requirements of all 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 users. Let $P_L^{(i)}$ (> 0) denote the PLP upper bound required by user i, $1 \leq i \leq N$. 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 lost MUPs divided by the average number of generated packets in each frame. When the user number N is maximized, the packet loss of all users will be at their limit (i.e., $\hat{P}_L^{(i)} = P_L^{(i)}$). If one more user is admitted to the system, the QoS requirements of all 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 users will reach their limits at the same time. The choice of $\hat{P}_M^{(i)}(R)$ is described in [8]. 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 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 that deals with how to pack 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 2 illustrates a heuristic algorithm developed for this problem, where i is the user index, $1 \leq i \leq N$; l is the time slot index in each frame and L_u the total number of time slots in each frame for the uplink; p_j is the packet size of user j; p^l is the total size of the scheduled packets in time slot l, and $\mathbf{p} = (p_1, p_2, \ldots, p_{L_u})$ 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. Packets with the same timeout value will be scheduled according to the number of lost packets for each user calculated using the FPLS method. Since the packet sizes

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 users who can tolerate some degree of packet loss.



■ **Figure 2.** *The proposed FPLS bin-packing packet scheduling algorithm for each frame.*

are different, the total system capacity C is unknown before 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 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 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 users. 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 overscheduled if $\kappa_i < 0$ and underscheduled 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 users. If user i has the largest priority index value, 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 anymore MUPs 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.

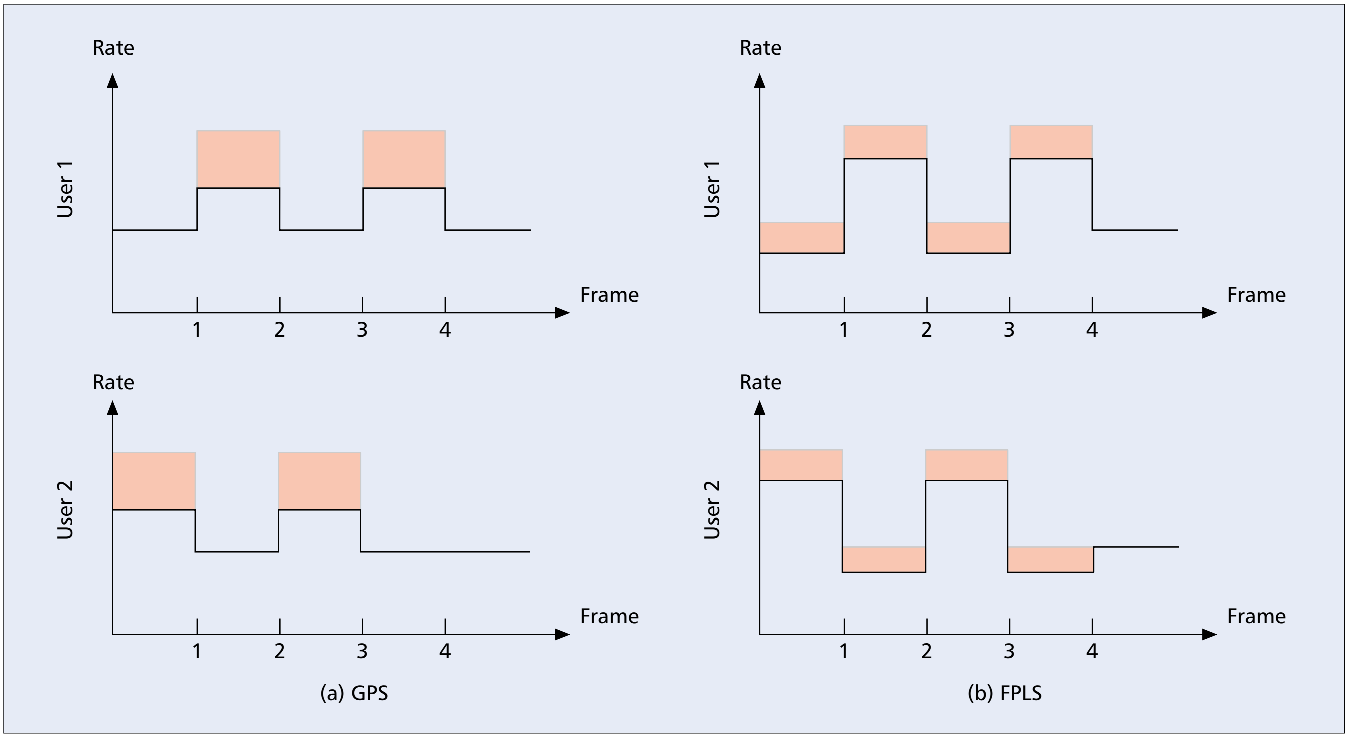
Performance Evaluation

Comparison with WISPER and GPS — The performance of the proposed MAC protocol is demonstrated in comparison with a MAC protocol using other packet scheduling schemes. Here we consider the comparison with:

- WISPER [5], which is a MAC protocol with packet scheduling for multimedia traffic in a system model similar to that considered here
- The MAC protocol using discretized GPS, which is a well-known work conserving protocol for bandwidth allocation in wireline systems

The three MAC protocols are compared via computer simulation using the same system model and guaranteeing the same QoS requirements for the same traffic flows. Although many methods have been reported so far 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 [9]. The effective bandwidth is the amount of resources required for a given performance objective. For real-time traffic without buffering, the effective bandwidth for user i can be obtained according to [7]. However, for non-real-time traffic with delay and PLP requirements, the calculation of effective bandwidth is very complex. In [9], 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 MAC protocols 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 packet timeout 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 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. 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 users share the packet loss depending on their PLP requirements. The difference between GPS and FPLS is illustrated in Fig. 3 for two users as an example, where packet loss happens whenever the allocated rate is below the input traffic rate. Using GPS, there will be little (or no) packet loss for a nonbursty traffic flow even though the user may tolerate packet loss to some degree. In FPLS, a nonbursty traffic flow will experience packet



■ **Figure 3.** *Illustration of resource allocation using GPS and FPLS. Solid lines represent the allocated rates. Shaded areas represent the differences between required rates and allocated rates.*

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 by 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. This gives a fair share of system resources to all users. In this way, FPLS is expected to achieve higher resource utilization than GPS.

The Simulation Environment — Consider the uplink transmission of three traffic types, voice, video, and data, each having its own transmission delay and accuracy requirements. Voice traffic is simulated by the on-off model. During the on state, one packet is generated in each frame of 10 ms, which is equivalent to a rate of 100 packets/s. Video traffic has a variable rate that varies among four rates (0, 4, 8, 12), in packets per frame, with a probability of (1/6, 1/3, 1/3, 1/6), respectively. We also use the same rate characteristics for data traffic. Both real-time (voice and video) and non-real-time (voice and data) transmissions are considered. There are eight time slots in each frame for the uplink information packet transmission. The maximum number of code slots (N_{max}) per time slot is chosen to be 22 for real-time traffic and 30 for non-real-time traffic. Table 1 gives the simulation parameter values. The three MAC protocols to be compared differ from each other mainly in packet scheduling. Each simulation is carried out for 10,000 time frames. For simplicity, the resource overhead necessary for signaling and control in MAC is not considered.

Real-Time Traffic — Figure 4a shows the maximum numbers of voice and video users that can be supported by the system with QoS

Parameter	Real-time traffic		Non-real-time traffic	
	Voice	Video	Voice	Data
Timeout value (frame)	1	1	2	20
Required received power (code slot)	1	1.9	1	3.2
Required PLP upper bound	10^{-2}	10^{-2}	10^{-2}	10^{-5}
Average talk spurt length (frame)	10		10	
Average silent period (frame)	15		15	
Average rate (packet/frame)		6		6
Peak rate (packet/frame)		12		12

■ **Table 1.** *System parameters used in the simulations.*

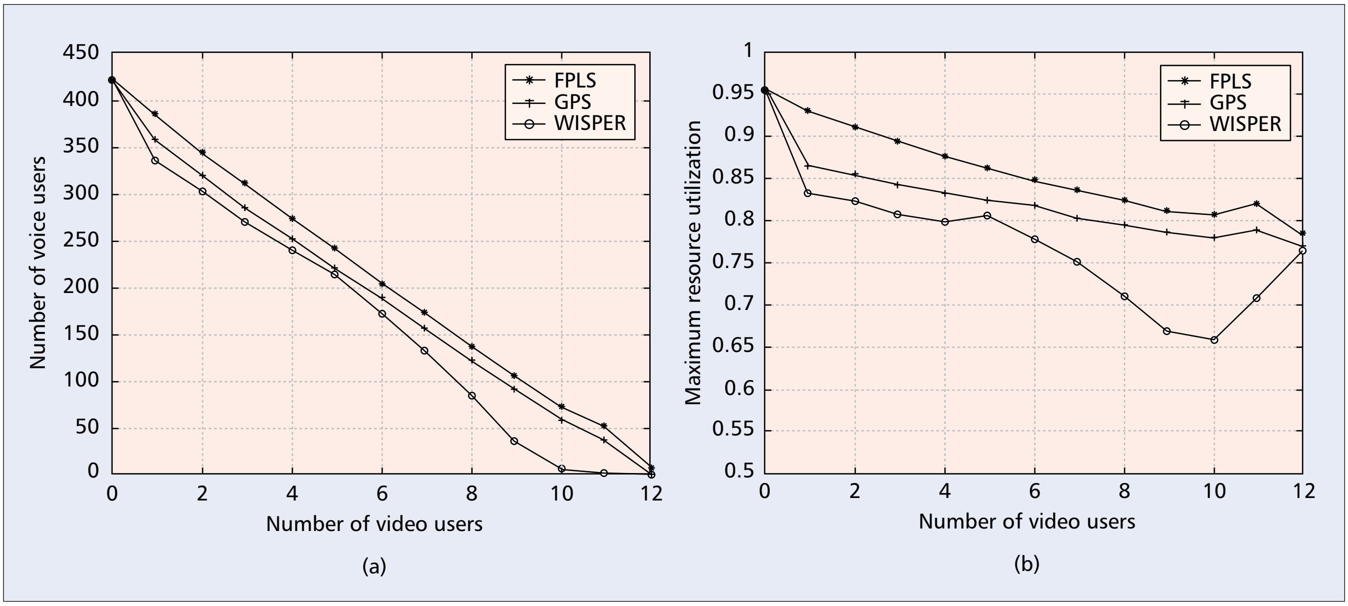


Figure 4. *Comparison of MAC protocols using FPLS, WISPER, and GPS, respectively, with real-time traffic: a) maximum user numbers; b) resource utilization efficiency.*

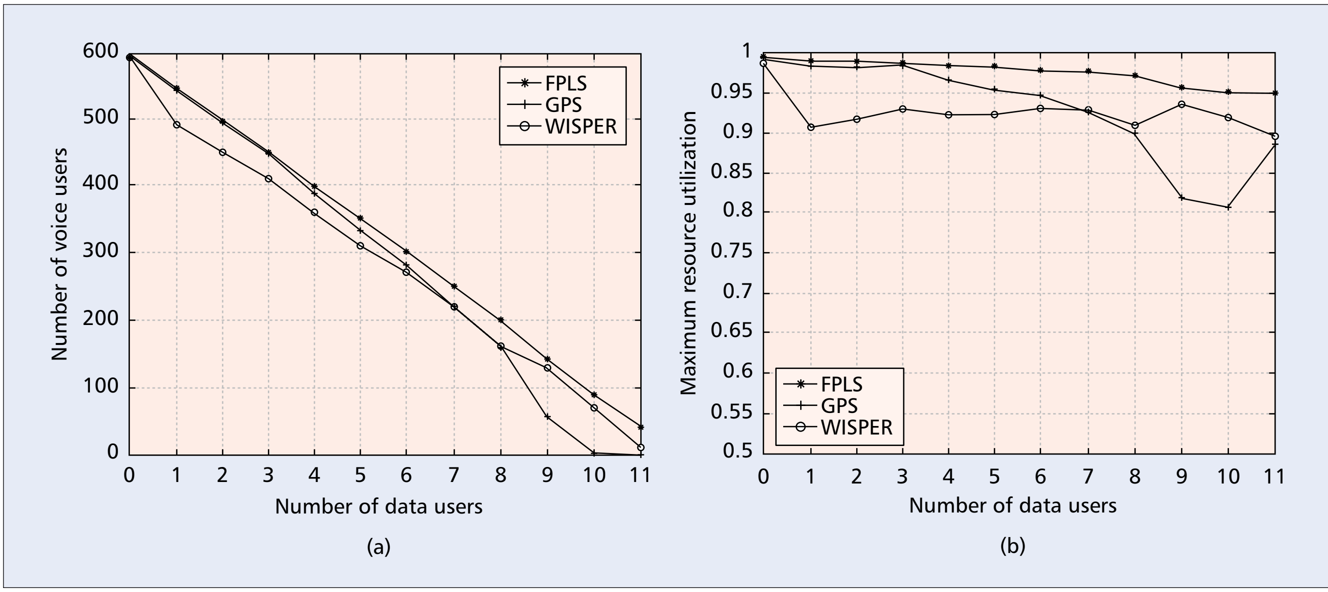
satisfaction. FPLS outperforms GPS and WISPER in most situations. When there is only one traffic type (voice or video) in the system, the total number of lost packets is the same for all three MAC protocols; therefore, there is no performance difference among the scheduling algorithms. However, if there exists a mixture of the two traffic types, FPLS performs better because of more effective statistical multiplexing. Since all arrived packets are MUPs, 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. Figure 4b shows the corresponding resource utilization efficiency with guaranteed QoS requirements for all users, where efficiency is defined as the average percentage of the total resources being used all the time. FPLS provides a higher resource utilization to accommodate more users than GPS and WISPER. It is expected that, with an increase of system capacity (e.g., L_u and/or $N_{\max}$), the performance improvement of FPLS over 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 flows (as compared to the real-time traffic case) are expected to increase the resource utilization efficiency. Figure 5a shows the maximum numbers of the voice and data users that can be supported by the system using FPLS, GPS, and WISPER, respectively. When the maximum number of 11 data users is reached, FPLS can still provide service for about 40 voice users while GPS and WISPER can barely support any voice users. Figure 5b shows the corresponding resource utilization efficiency. FPLS clearly outperforms both GPS and WISPER. In particular, when the difference between traffic loads is large, WISPER gives a higher priority to traffic with high loads, thus causing high packet loss for users with lower traffic loads. In WISPER, since 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 of in FPLS where all QoS requirements are used to determine the order for transmission; thus, resource utilization is greatly improved. As the number of traffic types and/or difference in PLP requirements increase, further performance improvement achieved by FPLS over WISPER is expected. Using GPS scheduling, resource utilization efficiency fluctuates when the number of data users is close to the maximum value. This is because the large number of data users dominates resource usage, and 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. As a result, the QoS for data users can be higher than required, which translates to reduced resource utilization efficiency.

CONCLUSION

We propose a MAC protocol with an FPLS scheduler for a hybrid TD/CDMA wireless communications system. Statistical multiplexing in both the time and code domains is achieved. Transmission accuracy over the wireless link is guaranteed by proper received power allocation, while packet loss probability and delay requirements are guaranteed by proper packet scheduling. For high resource utilization, the MAC protocol allocates a minimum amount of resources to each user for QoS provisioning by:

- Letting each user have a fair share in packet loss
- Assigning a minimum required received power level



■ **Figure 5.** *Comparison of MAC protocols using FPLS, WISPER, and GPS, respectively, with non-real-time traffic: a) maximum user numbers; b) resource utilization efficiency.*

- 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. 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.

Acknowledgments

This work was supported by a research grant from Communications Information Technology Ontario (CITO).

References

[1] 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 Trans. Net.*, vol. 1, no. 3, June 1993, pp. 344–57.

[2] D. J. Goodman *et al.*,"Packet Reservation Multiple Access for Local Wireless Communications," *IEEE Trans. Commun.*, vol. COM-37, no. 8, Aug. 1989, pp. 885–90.

[3] M. J. Karol, Z. Liu, and K. Y. Eng, "Distributed-Queuing Request Update Multiple Access (DQRUMA) for Wireless Packet (ATM) Networks," *Proc. IEEE ICC '95*, vol. 1, Seattle, WA, June 1995, pp. 1224–31.

[4] A. E. Brand and A. H. Aghvami, "Multidimensional PRMA with Prioritized Bayesian Broadcast — A MAC Strategy for Multiservice Traffic over UMTS," *IEEE Trans. Vehic. Tech.*, vol. VT-47, no. 4, Nov. 1998, pp. 1148–61.

[5] I. F. Akyildiz, D. A. Levine, and I. Joe, "A Slotted CDMA Protocol with BER Scheduling for Wireless Multimedia Networks," *IEEE/ACM Trans. Net.*, vol. 7, no. 2, Apr. 1999, pp. 146–58.

[6] V. Huang and W. Zhuang, "Optimal Resource Management in Packet-Switching TDD CDMA Systems," *IEEE Pers. Commun.*, Dec. 2000, pp. 26–31.

[7] M. Schwartz, *Broadband Integrated Networks*, Prentice Hall, 1996.

[8] V. Huang and W. Zhuang, "Fair Packet Loss Sharing (FPLS) Bandwidth Allocation in Wireless Multimedia CDMA Communications," *Proc. Int'l. Conf. 3G Wireless and Beyond*, San Francisco, CA, May–June 2001, pp. 198–203.

[9] A. Elwalid and D. Mitra, "Design of Generalized Processor Sharing Schedulers which Statistically Multiplex Heterogeneous QoS Classes," *Proc. IEEE INFOCOM '99*, vol. 3, 1999, pp. 1220–30.

Biographies

VINCENT HUANG [S] (huang@bbcr.uwaterloo.ca) received an M.Sc. degree in electrical engineering in 1994 from Chalmers University of Technology, Gothenburg, Sweden. He is currently working toward a 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 Chalmers University of Technology. His current research interests are MAC protocols and resource management for multimedia wireless communications.

WEIHUA ZHUANG [SM] (wzhuang@bbcr.uwaterloo.ca) received her 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 an associate 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 in 2001.