INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS
Int. J. Commun. Syst. 2002; **00**:1–15 *Prepared using* dacauth.cls *[Version: 2000/05/31 v1.0]*

A slotted multicode CDMA protocol for efficient integration of voice and data in cellular networks

H. Zhang and P. Gburzynski*

Department of Computing Science, University of Alberta
Edmonton, Alberta, Canada T6G 2E8

SUMMARY

This paper presents a novel radio channel structure based on slotted CDMA technology intended for carrying traffic with diverse bandwidth/QoS requirements in mobile environments, e.g., Personal Communication Systems (PCS). The essence of our approach is a combination of flexible slotting with allocation of multiple codes to high-bandwidth mobiles. As demonstrated by our performance studies, the proposed scheme efficiently integrates multiple traffic classes into a unified CDMA system. It is highly flexible and incurs low overheads for a wide range of realistic traffic conditions. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: Wireless protocols, cellular networks, CDMA, bandwidth allocation

1. INTRODUCTION

To support services within the bit rate ranges offered by the third-generation mobile systems, the scarce resource of the radio channel must be shared in a highly flexible and efficient manner. Two types of solutions have been proposed for such systems, one class based on the variable spreading gain technology (VSG-CDMA) [3, 5], and the other on allocating multiple codes to a single session (MC-CDMA) [1, 2, 4].

Few of the MC-CDMA protocols proposed in the literature address the issues of synchronization and power control, which are critical in CDMA communication. One would naturally expect that with a proper assistance of the MAC scheme, code synchronization and power control can be realized with less hardware complexity, and the system can achieve a higher overall efficiency, compared to a solution in which those issues are handled outside the MAC layer.

In this paper, we introduce an MC-CDMA protocol dubbed BRICS.† As two integral components of our solution, we suggest a quick code acquisition system and an air interface in which the uplink signaling channel is based on code-domain minislots. We also propose a

*Correspondence to: Department of Computing Science, University of Alberta
Edmonton, Alberta, Canada T6G 2E8
†Which stands for *Base Rate Incremental Coded Service*.

Received 17 October 2001
Accepted 3 February 2002

method to explicitly exploit silent periods in voice activity, which approach is different from the statistical (implicit) bandwidth reuse, e.g., in [7]. As indicated by our performance studies, BRICS efficiently accommodates multiple traffic classes with different bandwidth requirements and QoS expectations.

2. PROTOCOL DESCRIPTION

We consider a network with a single base station (BS) and a variable number of mobile stations (MS). The handoff part of the overall communication system implementing our protocol is beyond the scope of our present study.

Since the performance of a mobile network is limited by the capacity and flexibility of the link from MS to BS, only the structure of the *uplink* is discussed in detail. The protocol assumes the same prerequisites as WISPER [1], with a similar structure of the mobile transmitter and receiver. The carrier modulation is binary PSK.

All transmissions are carried out at the fixed *basic* rate R_b. A single mobile can transmit $1 \leq m \leq M$ packets simultaneously, using different spreading codes C_i, $(i = 1, \ldots, m)$ derived from a primary pseudo-noise (PN) code by a subcode concatenation scheme [2]. The primary code is assigned to the mobile when the station is admitted to the system. The transmission power P_t expedited by a mobile must increase along with the transmission rate m to provide the same signal-to-interference ratio (S/I) for each of the m parallel channels.

2.1. Code acquisition

A quick and reliable code acquisition scheme is essential for the correct operation of BRICS. As each code channel uses a different spreading code, it needs a separate correlator for code acquisition. Theoretically, if all m channels experience the same delay, they can all share a single code acquisition circuit. There are reasons, however, why it seems more reasonable to provide a separate circuit for each code channel—e.g., to take advantage of the multipath gain, as in RAKE receiver.

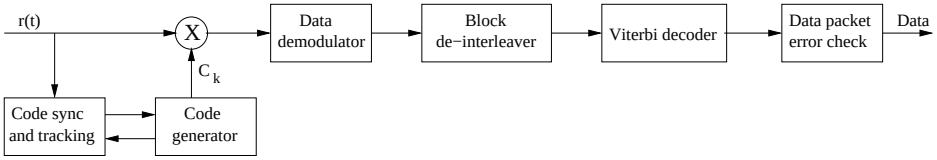


Figure 1. BRICS receiver

The high level layout of a single path receiver is shown in Figure 1. We propose a parallel code acquisition system [9, 10] built around a matched filter and utilizing the maximum likelihood strategy. The acquisition procedure consists of two phases: searching and verification, with the search block built of N parallel matched filters—as illustrated in Figure 2. A detailed description of such filters and correlators using them (depicted in Figure 3) can be found in [11].

Fast code acquisition at the base is more critical because of the compact organization of the uplink frame. To facilitate it, the mobile transmits a fixed length unmodulated PN sequence

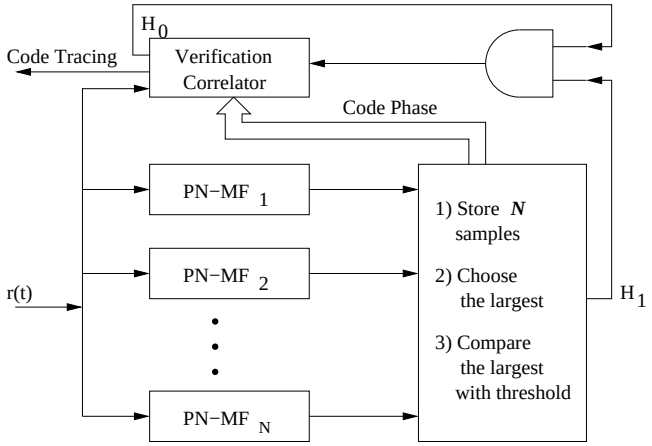
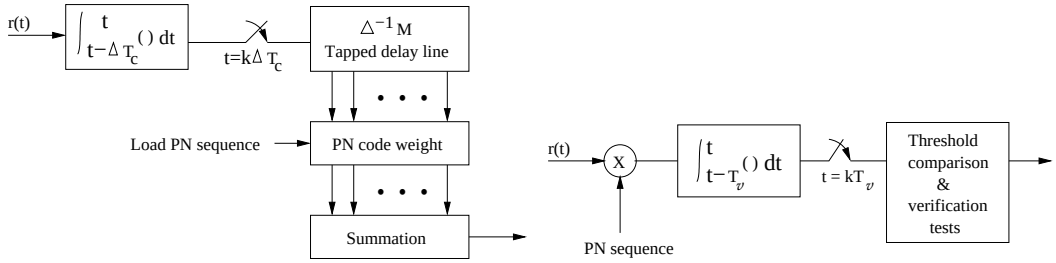


Figure 2. A PN code acquisition system based on parallel matched filters



(a) PN code matched filter (b) Verification correlator

Figure 3. Matched filter and correlator in a PN code acquisition system

(the acquisition preamble). We assume that the code is *long*, and partial-code correlation is applied in the matched filter. As the uplink frame is slotted, transmissions are constrained to start at slot boundaries, which simplifies the acquisition process [8].

The uncertain region for code acquisition at the base station is determined by the maximum distance D between the base and the mobile plus the uncertainty δ of the signal processing time, i.e.,

$$L = \frac{\frac{2D}{c} + \delta}{T_c} \tag{1}$$

where L is the duration of the uncertain region expressed as the number of PN chips, c is the speed of light, and T_c is the PN chip duration.

Assume that the code synchronization procedure starts at time zero. Using N parallel passive non-coherent PN matched filters (PN-MFs, see Figure 2), the uncertain region L is divided into N subsequences, each of length $K = L/N$. Each PN-MF is loaded with one of the N subsequences. The number of search cells on each delay line in the matched filter is K/Δ with

the delay of ΔT_c between successive taps (the standard recommended value of Δ is 1/2). In KT_c seconds, NK/Δ cells are searched, with each cell corresponding to one of the possible NK/Δ phases in the uncertain region. The largest sample and the corresponding code phase from each of the N parallel PN-MFs are stored and compared. If the sample exceeds a threshold γ_1, a tentative *hit* is assumed, and the corresponding phase is used to initiate the correlator and start the verification process. The searching continues until a true *hit* is declared in the verification process, or the preamble runs out, whichever happens first. In the former case, code tracking is started; otherwise, the code synchronization is presumed lost.

Every LT_c seconds, the N PN-MFs are reset with a new portion of the PN code shifted by LT_c seconds. If a certain specified number of all tests exceed the threshold γ_2, code acquisition is assumed and the code tracking system takes over the code synchronization. Otherwise, a false alarm is declared. If a new tentative hit was found already, say at time T_h, the verification process is immediately restarted with the phase shift of $t - T_h$, where t is the current time. Otherwise, the verification is aborted and postponed until the next hit. The verification is also aborted and reset if within the current search interval of KT_c seconds a new sample is found that is bigger than the one that triggered the previous hit. After that interval, however, the correlator is never reset unless it detects a false alarm.

An analytical estimate of the acquisition probability for a single matched filter is given in [9]. Since we employ N parallel filters, which match to multiple segments of the PN code, the acquisition probability in our case is considerably higher and of order $1-(1-P)^N$, where P is the probability for a single filter. This extrapolation may be slightly overoptimistic, however, because the N samples are not strictly independent. On the other hand, one can suggest a few directions for improving the accuracy and increasing the speed of the code acquisition algorithm even further. As subsequent slots transmitted with the same key will tend to be located closely in time domain, the uncertainty interval for code acquisition can be centered around the exact phase found for the previous slot. Also, this interval can be set much tighter than prescribed by Formula 1, based on how late the current slot follows the last one for which the code was successfully acquired. Even if the mobile moves at a high speed, it cannot move too far between two consecutive slots transmitted in two nearby frames. Notably, this will tend to be the case for high-rate sessions, for which the quality of code acquisition is especially important—from the viewpoint of the overall error rate and effective bandwidth.

2.2. Channel structure

Every uplink frame is partitioned into a number of logical channels spanning two dimensions, i.e., time and code. The frame can be envisioned as consisting of multiple layers of slots resembling a brick wall, as in Figure 4, that coincide in time but are separated by different codes (along the vertical axis). The amount of power assigned to a channel is illustrated as the height of the corresponding brick.

Except for the *RA* slots, which must all occur at the very beginning of the frame, the height of the remaining slots may vary across the code dimension. This is different, e.g., from [1], where different slots scheduled at the same time have identical properties. On the other hand, the *RA* slots play in our protocol a similar role to the contention slots in [1].

The remaining portion of the frame is built of four slot types. The granularity of bandwidth allocation is determined by the size of the standard (basic) slot, *TA*, whose duration is selected to accommodate exactly one active voice session. The second slot type, *TS*, is used to build

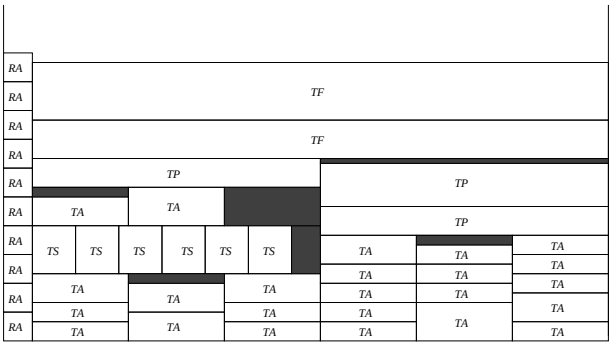


Figure 4. BRICS frame structure

signaling channels, needed by the mobiles admitted to the system to provide the base with feedback regarding their dynamic bandwidth requirements and received power. These slots are allocated from the beginning of the frame. They cannot appear too close to the end, because the base must be able to process the information contained in them before announcing the layout of the next frame. As they are shorter than the *TA* slots, the last signaling slot may be followed by an unusable gap.

A slot of the third type spans the entire frame space and is intended for high-bandwidth sessions. We call it a *flat slot* (or a flat channel) and denote by *TF*. To increase the flexibility of flat allocation, we admit *partial flat slots*, denoted by *TP*, spanning the width occupied by several *TA* slots, but less than the whole frame. As the allocation of partial flat slots is considerably more complex than for the remaining slot types, it makes sense to impose (possibly drastic) restrictions on their size and/or position within the frame. In our virtual implementation of the protocol (Section 3), we have assumed that a partial flat slot is half the size of the (full) flat slot, and that it must be aligned at a half-frame boundary.

2.3. Medium access

When a mobile wants to initiate a session, it randomly selects one of the *RA* channels and sends a request to the base station using a modified ALOHA protocol. The access request packet transmitted in the *RA* channel consists of the following components: mobile ID, service type, resource requirements, delivery deadline, and transmitted power level. The exact specification of resource requirements depends on the service type and may include transmission rate, message length, or be empty. For example, the characteristics of a voice session are always the same and completely determined by the service type.

If the base station correctly receives the access request packet, it will assign to the mobile an uplink signaling channel in the next frame. Depending on the available bandwidth, the mobile may be also immediately assigned a traffic channel.

As part of the control packet announcing the layout of the forthcoming frame, the base broadcasts to the mobiles the current *contention permission status*, indicating which traffic classes are allowed to compete for bandwidth. With this mechanism, the base is able to inhibit lower priority sessions when bandwidth becomes scarce. By thresholding the perceived noise level in the *RA* channels, the base may also selectively restrict contention to high priority

classes if that level appears to be too high.

On the mobile's end, we propose a combination of p-persistent behavior (which is a standard approach in ALOHA systems) with adjustments of transmitter power. Following an unsuccessful attempt, indicated by the lack of a response from the base in the next downlink control message, the mobile will reduce the probability of transmission p and, at the same time, increase the power for a subsequent attempt. This way it will become less aggressive with its requests, while improving their chances for a successful reception by the base. Also, by using more power, stations that have failed are given priority over new contenders: they are more likely to win considering a given level of multiple access interference (MAI) from other codes, and more likely to be captured if other contenders (transmitting at lower power) happen to pick the same code.

With our generic scheme, the mobile issues its first attempt with probability $p_0 = 1$ using power $P_0 = P_a$, where P_a denotes the initial (starting) power level. Following a failure, the station executes $P_{i+1} = \min(P_i \times \delta_P, P_{max})$, where $\delta_P > 1$ is the power increment factor and P_{max} is the maximum power, and $p_{i+1} = \max(p_i \times \delta_p, p_{min})$, where $\delta_p < 1$ is the probability decrement factor and $p_{min} > 0$ is the minimum probability of transmission.

2.4. Downlink signaling

As the base station enjoys exclusive access to the downlink channel, there is no need for the *RA* slots in a downlink frame. Also, instead of using multiple signaling channels, the base station transmits a single *control packet* aligned at the end of the frame and occupying an equivalent of a (partial) flat channel. With this solution, no bandwidth is wasted on slot boundaries, and the control messages can be of variable length, e.g., depending on the number of slots/codes assigned to a given mobile. All mobiles tune in to the control packet (using a predefined code) and identify its relevant fragments (individual control messages) specifying the locations of their slots, the positions and keys of their uplink signaling channels, and power adjustment. The number of predefined signaling codes is small, e.g., 2-4; so this part of the control message occupies a trivially small amount of space. With the power levels quantized into a reasonable number of discrete states, the size of a single control message is going to be small and so is the total amount of bandwidth needed for downlink signaling.

2.5. Bandwidth allocation

To provide a satisfactory reliability of transmission for a given service, the system must ensure that the bit error rate (BER) does not exceed the service-specific maximum [13, 14]. The BER specification can be mapped to the bit energy to noise spectral density ratio E_b/N_0 [15]. In contrast to [15], an allocated channel in BRICS is always active; thus, we have the following admission constraint:

$$\sum_{k=1}^{K} \alpha_k < 1, \qquad \alpha_k = \frac{(E_b/N_0)_k}{W/R_b + (E_b/N_0)_k}$$

where K is the number of simultaneous code channels in the time slot, $(E_b/N_0)_k$ is the E_b/N_0 requirement for k-th code channel, W is the total spread bandwidth and R_b is the base rate. Consequently, the minimum power assignment is $P_i = \alpha_i(\eta + P)$, for $i = 1, \ldots, K$, where η is the background noise, and the minimum total received power is $P = \left(\sum_{k=1}^{K} \alpha_k \eta\right) \Big/ \left(1 - \sum_{k=1}^{K} \alpha_k\right)$.

While building the structure of the next uplink frame, the base keeps an allocation table indexed by time slots located at the time boundaries of *TA* channels—see Figure 4. Each entry in that table stores the height of the brick wall across the slot (i.e., the sum of all α_k falling in the slot) and the list of requests accommodated at that location. A new *TA* channel is allocated at the location with the smallest height, which results in the update of a single entry in the allocation table, while the addition of a new *TF* channel simply raises the height of the entire wall by the same amount.

Formally, the optimal assignment of channels to the frame is NP-hard, being trivially equivalent to the bin packing problem. However the maximum error incurred by the simple greedy approach is bounded by $\max \alpha_k$, which seems quite acceptable. We postulate that high priority sessions be restricted to basic (*TA*) and full flat (*TF*) channels, whose allocation is straightforward. Following this "rigid" allocation phase, the leftover frame space can be partitioned among the more flexible lower priority sessions. This way, instead of trying to solve a computationally difficult task of partitioning the remaining portion of bandwidth into a number of rigid chunks, we reverse the problem and allocate whatever chunks come out handy to flexible sessions.

The bandwidth scheduler operates in cycles. Every cycle starts with the reception of all *RA* slots through which new mobiles register their sessions with the base. These new requests are appended at the end of the respective queues. Then the status of the sessions in progress is updated based on the received contents of the signaling channels *TS*. Having received the last signaling slot, the base is ready to process the request queues, allocate bandwidth, and announce the layout of the next frame.

3. A SAMPLE CONFIGURATION OF THE PROTOCOL

We assume that our cellular system offers four types of service listed in the decreasing order of priority: voice, video conferencing, file transfer, and SMS.

3.1. The radio channel

The basic rate R_b of our radio channel is 500 kbps and the transmission rate needed to sustain a voice session in its active phase is 22 kbps. The total length of a single uplink frame is $l_f = t_g+p+l_{ra}+N_{ta}\times(t_g+p+l_{ta})$, where t_g is the guard time (8 bits = 16 μs), p is the acquisition preamble length (32 bits or 64 μs), l_{ra} is the length of the contention slot *RA* (96 bits or 192 μs), l_{ta} is the length of the basic slot *TA* (384 bits or 768 μs–ATM payload size), and N_{ta} is the number of *TA* slots in a single layer of the frame (20). All these numbers add up to $l_f = 8616$ bits or 17.24 ms. The payload length of one flat slot *TF* is $l_{tf} = N_{ta}\times(t_g+p+l_{ta})-t_g-p = 8440$ bits.

The total amount of bandwidth available within a frame is determined by the basic rate R_b, the total spread bandwidth of the channel W, and the E_b/N_0 requirements of the individual sessions. In our model, we assume $W = 20$ MHz as the target bandwidth, although we consider some other values for comparison.

3.2. Contention resolution

Contention to the *RA* channels is resolved according to the generic strategy described in Section 2.3, using 10 codes, with $P_a = 80\,\text{mW}$, $P_{max} = 200\,\text{mW}$ (this is also the maximum transmitter power of every mobile in the system), $\delta_P = 1\,\text{dB}$, $\delta_p = 0.25$, and $p_{min} = 0.1$. These values imply three levels of persistence $(1, 0.25, 0.1)$ and five power levels. As determined by simulation, they result in the effective throughput of $3.0-4.0$ new requests per frame, assuming $E_b/N_0 = 6\,\text{dB}$ and some reasonable contribution from the capture effect.

3.3. Signaling channels

The total length occupied by a single signaling slot *TS* is 72 bits, with the payload restricted to 32 bits. These bits are partitioned into a 6-bit bandwidth specification, 10 bits for received power indication, and 8 bits left for extensions. All signaling slots occupy up to two layers (codes) of the first half of the frame space. The number of signaling channels per one layer is limited by 58, and the maximum total number of signaling channels is 116. The E_b/N_0 ratio for signaling channels is 6 dB.

3.4. Voice traffic

Our variant of the "on-off" voice model is similar to the one considered in [12], with the following parameters: *source rate* = 22 kbps, *mean call duration* = 3 minutes, *mean talkspurt length* = 1 second, *mean silence length* = 1.35 seconds, *deadline* = 10 seconds, *threshold for disabling voice requests* = 10%, *QoS* (E_b/N_0) = 5 dB. The arrival process of new calls is Poisson.

When a voice session enters the silent state, its traffic channel is temporarily released, but the connection sustains itself through the signaling channel. When the session gets back to the talkspurt mode, its traffic channel is reassigned in the next frame.

The bandwidth temporarily released by a voice session that has entered the silent phase can be assigned to lower priority sessions, but it cannot be used to accommodate a new voice session. This way, the bandwidth scheduler makes sure that all admitted voice sessions can always be accommodated in their active states.

The access contention permission flag for voice traffic is cleared if the population of unserviced voice requests exceeds 10% of all requests in the voice queue. The amount of bandwidth assigned to an active voice session is always the same and equal to one basic slot *TA* at $E_b/N_0 = 5\,\text{dB}$.

3.5. Video traffic

This traffic is described by a DAR(1) (Discrete Autoregressive) model [6] with the following parameters: *mean source rate* = 128 kbps, *variance* = 5536, *correlation* = 0.98, *mean call duration* = 30 minutes, *call deadline* = 10 seconds, *threshold for disabling video requests* = 10%. $E_b/N_0 = 5\,\text{dB}$.

Every admitted video session is guaranteed a certain minimum amount of bandwidth b_{min}. Any bandwidth requested in excess of the minimum is scheduled in a fair manner using the equal degradation approach, with the service grade $G = \min(1, B_a/B_r)$, where B_a is the total maximum bandwidth available for teleconferencing service, after the higher-priority voice sessions have been accounted for, and B_r is the total extra bandwidth requested by the

admitted video sessions. The amount of extra bandwidth assigned to a video mobile is equal to $\min(G, b_r)$, where b_r is the extra bandwidth requested by the station.

While admitting a new video session, the bandwidth scheduler assumes that the session can start with the minimum bandwidth b_{min}, and admits the session if that much bandwidth is available. For this calculation, all active video sessions are counted with their minimum bandwidth b_{min} and all voice sessions are assumed to be active. The assignment part of bandwidth scheduling for video sessions is trivial: it consists in adding one or more *TF* slots to the brick wall, appropriately updating its height everywhere.

As implemented at the mobile's end, a video session is buffered with the lag of 4 frames (or 69 ms), which allows the station to better tailor the current rate to the size of the *TF* slot and be more flexible with the bandwidth received from the base. As an admitted video session is guaranteed a minimum bandwidth in every frame, no per-packet deadlines need to be specified in uplink signaling messages.

3.6. File transfers and SMS

File transfers have no inherent delay requirements, and they can use any rate physically available to the mobile. Our primary objective in handling file transfers is to maintain a reasonable degree of fairness while avoiding unnecessary fragmentation.

File transfers occur in bursts, with burst duration and inter-burst periods being exponentially distributed. During a burst, a new file for transmission is generated at exponentially distributed intervals. The length of every file is exponentially distributed as well. The numerical parameters assumed in our model are as follows: *mean burst duration* = 30 minutes, *mean inter-burst interval* = 30 minutes, *mean interval between files in a burst* = 36 seconds, *mean file length* = 104 KB, *threshold for disabling file transfer requests* = 10%. $E_b/N_0 = 6$ dB.

File transfer session receive bandwidth with the *preferred granularity* of one *TF* slot, in a Round-Robin fashion. If a mobile needs less than a full *TF* slot to complete its request, it is assigned a partial flat slot *TP* or a basic slots *TA*. The procedure continues until all bandwidth has been assigned or all mobiles have been satisfied. If no more *TF* channels can be allocated, the preferred granularity is downgraded to *TP* and then to *TA*.

An SMS message is treated exactly as a file to transfer, except that the transfer is scheduled with the lowest possible priority. The numerical parameters are as follows: *mean message length* = 6250 bytes, *mean interval time* = 11 seconds, *deadline* = 2 hours, *threshold for disabling SMS requests* = 20%, $E_b/N_0 = 6$ dB. Both the interarrival time and message length are exponentially distributed.

4. SELECTED SIMULATION RESULTS

The performance of our virtual implementation of BRICS has been investigated by simulation and compared to the performance of three other CDMA protocols: WISPER [1], VSG-CDMA [3], and S-CDMA (described below). All protocols were implemented in the same virtual radio environment in which the sole criterion of a successful reception was the bit energy to noise density ratio E_b/N_0 (affected by a steady level of background noise) perceived at the receiver. We claim that this kind of environment is fair for a comparison of these

particular protocols because they all operate within essentially the same set of prerequisites: power control (and possibly other feedback from the base station) aimed at keeping E_b/N_0 in line with their assumed performance criteria. Consequently, more subtle properties of the radio channel, e.g., errors incurred by non-uniform noise and fading, would affect all protocols to the same degree. The original performance models of WISPER and VSG-CDMA [1, 3] were similar or simpler than our model.

4.1. Other protocols

Our virtual implementation of WISPER assumes exactly the same channel parameters as in BRICS, including the basic rate (500 kbps) and the numerical attributes of the frame (gap, preamble, request slot and regular slot). Consequently, the resulting frame consists of 20 384-byte slots, and its back to back time duration is 17.24 ms.

VSG-CDMA is a variable rate protocol, and its framing parameters cannot be directly related to those of BRICS (because the entire frame in VSG-CDMA is essentially a single slot). The frame (slot) length is equal to the duration of the probability distribution cycle, which in our model was set to 50 ms. Protocol independent elements (e.g., the power range available to the mobiles, fading properties of the channel, background noise, Eb/N_0 requirements of the receiver at the base) were identical in all models.

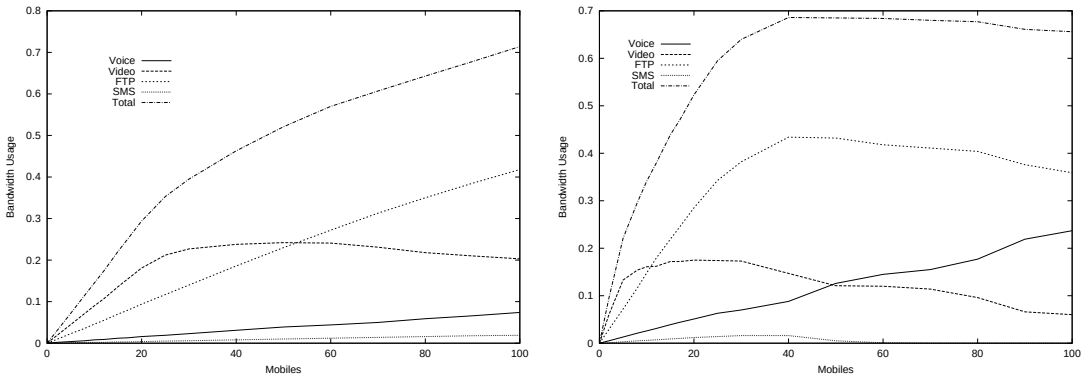
S-CDMA is a simple protocol without admission control, which we include in our study to illustrate the importance of admission control in CDMA. Time is divided into equal length slots, with each slot carrying 10 packets (ATM payloads). Each mobile has its own pre-assigned CDMA code. A packet transmitted by a mobile is lost if its received E_b/N_0 ratio is below the acceptable level. The base station provides a simple power feedback to the mobiles. When the base senses that the received E_b/N_0 ratio is over a certain threshold P_h, it commands the mobile to reduce its transmission power by a prescribed factor δ_s. Similarly, if the perceived E_b/N_0 is below another threshold P_l (but the packet can still be received), the base will request the mobile to raise its power by the same factor δ_s. For as long as the bit energy to noise density remains between P_l and P_h, the mobile is allowed to retain its current power. After a silent period, during which there has been no feedback from the base, the mobile uses its last power level as the initial value. After a failed transmission attempt, the mobile increases its power by δ_s and tries again.

4.2. Bandwidth utilization

Figure 5 illustrates how the link bandwidth in BRICS is shared among the four traffic classes described in Section 3 for two values of the total spread bandwidth W. A single point on the X-axis, which indicates the system load, corresponds to the number of mobiles engaged in sessions of a given type. This number is always the same for all four traffic classes, e.g., at $X = 80$ there are exactly 80 voice mobiles, 80 video mobiles, and so on.

Consider part (a) of Figure 5 first. With the voice traffic having the highest priority, the total amount of bandwidth occupied by voice sessions increases linearly with the increasing number of voice mobiles (the system capacity is much higher than 100 voice sessions). The part of bandwidth assigned to video sessions (which have the second highest priority and the largest bandwidth requirements) increases linearly until about 25 mobiles, when the system becomes saturated with video. From then on, the video sessions yield bandwidth to voice sessions.

Note that although file transfers and SMS sessions have lower priorities than video sessions,



(a) $W = 20\,\mathrm{MHz}$ (b) $W = 5\,\mathrm{MHz}$

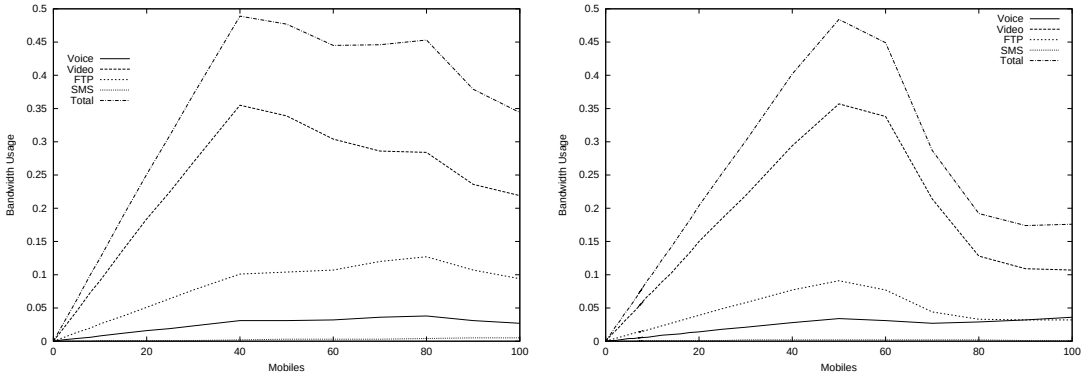
Figure 5. Bandwidth utilization in BRICS

their assigned bandwidth tends to increase until the very end of the investigated range of traffic conditions. There are two reasons for this behavior. First, there exist silent periods in voice sessions that cannot be reused to set up new voice or video connections, but are freely available to UBR/ABR traffic classes, like file and SMS transfers. Moreover, because of the inherent variability in video load (combined with intermediate buffering—Section 3.5), an equivalent of silent periods also occurs in video sessions. Second, the bandwidth requirements of a video session are stringent and must be granted in multiples of TF channels. Consequently, there are bound to exist chunks of bandwidth unusable by video sessions, but available to the much less picky file and SMS transfers.

The QoS tradeoffs in BRICS are somewhat better visible in a network with smaller total spread bandwidth W, which leaves less room for the low priority traffic to sneak in. One can clearly see in part (b) of Figure 5 how all three lower priority classes yield to voice.

The performance of WISPER and VSG-CDMA under identical offered load is shown in Figure 6. Notably, WISPER achieves lower maximum channel utilization than BRICS, and its total bandwidth utilization tends to drop when the system becomes saturated. This trend is followed by all traffic classes (except SMS, whose bandwidth is too small to be significant), which means that the prioritization of the four traffic types does not fulfill its purpose very well. For example, an increase in the number of admitted video sessions results in a drop in the portion of bandwidth effectively available to voice sessions. There are two reasons for this behavior: the rigidity of bandwidth allocation, which requires that a given time slot be filled with traffic of the same class, and the relatively poor performance of the contention resolution part of the protocol (whose impact was neglected in the original analysis of WISPER presented in [1]).

A similar but even more pronounced trend can be observed in VSG-CDMA. At first sight this is surprising because the method of adjusting the transmission probability in VSG-CDMA is intended to admit the right amount of traffic to the system, even under heavy contention. As it turns out, a realistic implementation of this protocol exhibits instabilities. Under light



(a) WISPER (b) VSG-CDMA

Figure 6. Bandwidth utilization at $W = 20\,\mathrm{MHz}$

load, the network operates without transmission control, and all ready stations transmit with probability 1. When the load becomes heavy, the system has the tendency to oscillate between two states. Having detected a congestion in one frame, the base turns on the controlled mode. Consequently, in the next frame all active users transmit with some probability p, which solely depends on the number of active mobiles. That frame tends to be underloaded, which forces the network back to the uncontrolled mode, which in turn causes congestion, and so on. These oscillations do not necessarily happen on a frame-by-frame basis, but they are clearly perceptible and so is their impact on the overall performance of the scheme.

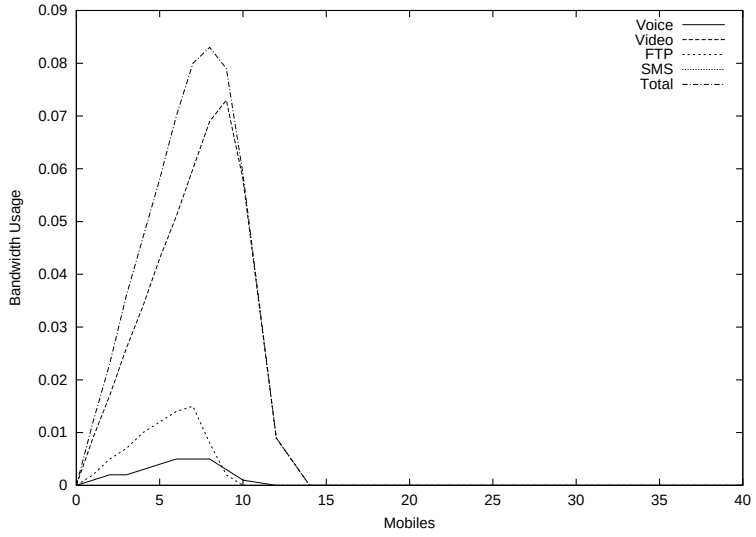
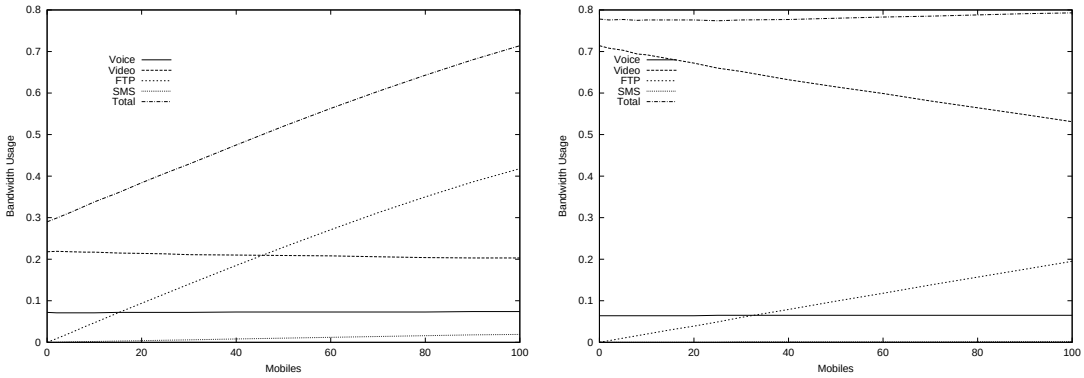


Figure 7. Bandwidth utilization in S-CDMA ($W = 20\,\mathrm{MHz}$)

The importance of admission control is well illustrated in Figure 7 obtained for S-CDMA. Although the protocol achieves a relatively high maximum bandwidth at about 7 mobiles in each traffic class, its performance breaks down completely and abruptly when the offered load exceeds the saturation point of the network. This happens around 10 mobiles, and at 14 mobiles the network practically ceases to deliver any traffic at all.

4.3. Quality of service

To see how the QoS received by high priority sessions in BRICS is affected by the presence or absence of other traffic types, we carried out a series of experiments in which the offered voice and video load remained steady, while the contribution of the remaining two traffic classes varied. The combined load of voice and video was set at a high level—to make the drop rate (and any deviations thereof) clearly visible, while the file transfer and SMS load increased in proportion to the number of mobiles in the network. Figure 8 compares the stability of the high-priority service in BRICS to that in WISPER.



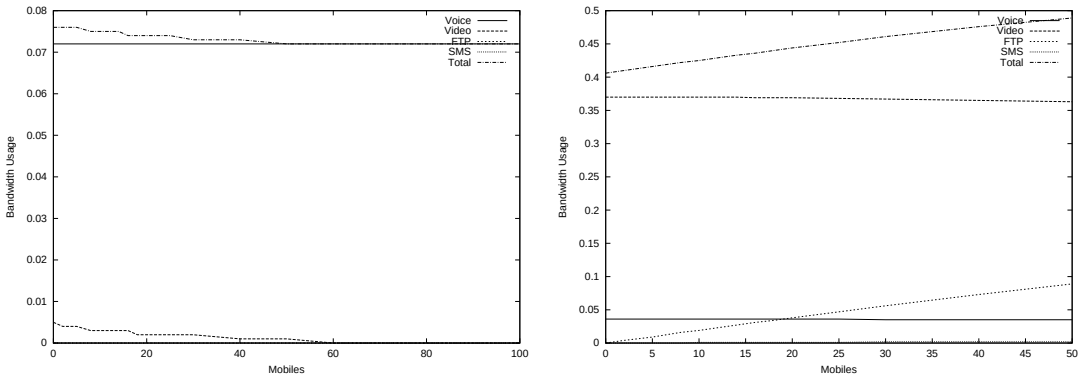
(a) BRICS (b) WISPER

Figure 8. Bandwidth utilization under fixed voice and video load ($W = 20\,\text{MHz}$)

While the amount of bandwidth used by voice and video sessions in BRICS is affected to little extent by the low priority sessions, there is a visible drop in the video bandwidth and a slight, hardly perceptible, increase in the voice bandwidth. These phenomena have their source in the varying level of contention on the RA channel, as the intensity of file transfers and SMS traffic becomes higher. With some bandwidth being reserved by video sessions, a few voice requests may remain queued for a while at the base before being blocked—until their short deadlines expire. The increasing number of file transfer and SMS requests contribute to the contention on the RA channel. According to the contention permission scheme, if the interference level in the RA slot is over the threshold, lower priority requests are denied access until the higher priority access queues at the base are emptied. This way, the queued voice requests are responsible for lowering the bandwidth used by video sessions under heavy contention from low priority traffic, as well as for marginally increasing the bandwidth used by the voice sessions themselves. The impact of this phenomenon can be somewhat influenced by adjusting the deadline of voice

sessions and the contention threshold for the RA channel. Note, however, that the reduction in the video throughput does not match the increase in voice throughput. This is because different contention opportunities translate into different bandwidth utilization patterns for the two traffic types. In particular, a sizable fraction of the bandwidth reserved for a voice session goes to its silent periods, which part is reused by file and SMS transfers.

Note that the video throughput starts higher in WISPER than in BRICS. This is because the priority scheme of WISPER favors heavy bandwidth sessions. This also explains why the SMS traffic, whose bandwidth is very low, can hardly make it through (the SMS curve is completely indistinguishable from the X-axis). Although WISPER is able to maintain a steady QoS for the voice sessions regardless of the presence of other traffic types, the bandwidth used by voice traffic drops significantly under heavy file transfers.



(a) Heavy load

(b) Reduced load

Figure 9. Bandwidth utilization in VSG-CDMA under fixed voice and video load ($W = 20\,\text{MHz}$)

Despite their differences, BRICS and WISPER exhibit comparable performance patterns, which is more than can be said for VSG-CDMA and S-CDMA. Figure 9 illustrates the case for VSG-CDMA. In part (a), the voice service remains steady throughout the entire range of traffic conditions, but the video bandwidth drops to zero and the remaining two traffic classes receive practically no service at all. This is because the network is completely saturated under the same load that is quite acceptable to BRICS and WISPER, and the heavy access contention makes it impossible for the lower priority traffic classes to force their way through. Notably, under reduced load, the protocol is able to cater to the four traffic classes according to their expectations. Part (b) shows the performance of VSG-CDMA when the offered load from voice and video is cut by half.

5. CONCLUSIONS

We have presented a CDMA protocol aimed at accommodating traffic classes with different QoS requirements. By using the brick wall approach to partitioning the uplink frame among the

multiple mobiles, the protocol is flexible with bandwidth allocation, yet the complexity of its bandwidth scheduler seems to be reasonably low. We have demonstrated that our protocol well caters to traffic classes with diverse QoS requirements and, in particular, efficiently accommodates data traffic without compromising the quality of service for voice and video. This property makes it a good candidate for future mobile networks, in which non-voice traffic will constitute a considerably more significant component than it does today.

At first sight, our approach of admitting only as many voice sessions as can be sustained simultaneously in their active phases may seem restrictive and run against the commonly accepted policy that relies on statistical multiplexing to offer more voice bandwidth to the users. Although one can only guess about the load patterns of future PCS networks, it is rather obvious that the contribution of traditional voice sessions to those patterns will tend to decrease. Consequently, with the increasing spread bandwidth of those networks, it will be pointless to try to accommodate as many voice sessions as physically possible, and the focus will shift towards efficient coexistence of voice with other session types. In this context, one should not worry about the "holes" caused by inactive voice sessions. They will be naturally reused by less picky (but no less important) asynchronous transactions.

REFERENCES

1. Akyildiz IF, Levine DA, Inwhee J. A slotted CDMA protocol with BER scheduling for wireless multimedia networks. *IEEE/ACM Transactions on Networking* 1999; **7**(2):146–159.
2. Chih-Lin I, Gitlin RD. Multi-code CDMA Wireless Personal Communication Networks. In *Proceedings of ICC'95*, Seattle, WA, June 1995; 1060–1064.
3. Chih-Lin I, Sabnani KK. Variable spreading gain CDMA with adaptive control for true packet switching wireless network. In *Proceeding of ICC'95*, Seattle, WA, June 1995; 1060–1064.
4. Choi S, Shin KG. An uplink CDMA system architecture with diverse QoS guarantees for heterogeneous traffic. *IEEE/ACM Transactions on Networking* 1999; **7**(5):616–628.
5. Fantacci R, Nannicini S. Multiple access protocol for integration of variable bit rate multimedia traffic in UMTS/IMT-2000 based on wideband CDMA. *IEEE Journal on Selected Areas in Communications* 2000; **18**(8):1441–1454.
6. Heyman DP, Lakshman TV. Source models for VBR broadcast-video traffic. *IEEE/ACM Transactions on Networking* 1996; **4**(1):40–48.
7. Kumar S, Nanda S. High data-rate packet communications for cellular networks using CDMA: algorithms and performance. *IEEE Journal on Selected Areas in Communications* 1999; **17**(3):472–492.
8. Madhow U, Pursley MB. Acquisition in direct-sequence spread-spectrum communication networks: an asymptotic analysis. *IEEE Transactions on Information Theory* 1993; **39**(3):903–912.
9. Madhow U, Pursley MB. Mathematical modeling and performance analysis for a two-stage acquisition scheme for direct-sequence spread-spectrum CDMA. *IEEE Transactions on Communications* 1995; **43**(9):2511–2520.
10. Park HR, Kang BJ. On the performance of a maximum-likelihood code-acquisition technique for preamble search in a CDMA reverse link. *IEEE Transactions on Vehicular Technology* 1998; **47**(1):65–74.
11. Polydoros A, Weber CL. A unified approach to serial search spread-spectrum code acquisition - Part II: a matched-filter receiver. *IEEE Transactions on Communications* 1994; **32**(5):550–560.
12. Qiu X, Li VOK, Ju JH. A multiple access scheme for multimedia traffic in wireless ATM. *Mobile Networks and Applications* 1996; **1**(3):259–272.
13. Sampath A, Kumar PS, Holtzman JM. Power control and resource management for a multimedia wireless CDMA system. In *Proceedings of PIMRC'95*, Toronto, Canada, Sep. 1995; 955–959.
14. Viterbi AM, Viterbi AJ. Erlang capacity of a power controlled cellular CDMA system. *IEEE Journal on Selected Areas in Communications* 1993; **11** (6):892–900.
15. Yun LC, Messerschmitt DG. Power control for variable QOS on a CDMA channel. In *Proceedings of MILCOM'94*, Boston, MA, Oct. 1994; 178–182.