Performance of Space-Division Multiple-Access (SDMA) With Scheduling

Hujun Yin, *Member, IEEE*, and Hui Liu, *Member, IEEE*

***Abstract*—Efficient exploitation of spatial diversity is fundamentally important to resource critical wireless applications [Tsoulos (1999)]. In this paper, we first study the performance of intelligent scheduling for space-division multiple-access (SDMA) wireless networks [Suard (1998), Farsakh (1998)]. Based on the existing scheme, we propose a new medium access protocol (MAC) for multimedia SDMA/time-division multiple-access (TDMA) packet networks [Xu (1994), Ward (1993)]. The improved protocol performs scheduling based on users' spatial characteristics and quality-of-service parameters to achieve throughput multiplication and packet delay reduction. Performance of SDMA with scheduling is evaluated under mixed audio and data traffic patterns and results show that significant improvement in network performance can be achieved under the new protocol.**

***Index Terms*—Antenna array, packet networks, scheduling, space-division multiple-access.**

I. Introduction

PACKET-SWITCHED networks are seen as the generic network architecture for future wireless systems that must support mixed traffic in an efficient manner. In the presence of multimedia traffic and environmental changes, it is required that a medium access control (MAC) can effectively adapt to changing channel conditions and traffic requirements. Many innovative MAC protocols have been proposed in the past decade [6]–[8]. Most schemes, however, inherit some architectural features of wire-line systems and treat "layers" as uncoupled entities [9], [10]. For example, data link/media-access protocols have been frequently designed and analyzed without sufficient regard to design details at the physical layer.

One of the key resources in wireless communication is the spatial diversity provided by antenna arrays. While inherently rich, the spatial resource is *highly irregular* and, thus, is difficult to deal with at the MAC level. For this reason exploitation of the spatial diversity has been limited to the physical layer using techniques such as spatial beamforming. However, if the MAC protocol is designed without adaptability to the channel conditions, it has to be designed for the *worst*-case scenario and as a result, there will be too much signal level margin and a significant waste of network resources.

Manuscript received August 21, 2000; revised January 18, 2002; accepted January 18, 2002. The editor coordinating the review of this paper and approving it for publication is L. J. Cimini. This work was supported in part by the Office of Naval Research (NOR) under Grant N00014-97-1-0475 and the National Science Foundation (NSF) CAREER Program under Grant MIP-9703074. This paper was presented in part at the 33rd Asilomar Conference on Signals, Systems, and Computers, 1999.

H. Yin is with ViVATO Inc., Spokane, WA 99216 USA (e-mail: hujun_yin@vivato.net).

H. Liu is with the Electrical Engineering Department, University of Washington, Seattle, WA 98195 USA (e-mail: hliu@ee.washington.edu).

Digital Object Identifier 10.1109/TWC.2002.804188

The need for coordinated MAC and physical layers in a system architecture is pronounced in antenna array networks. Only limited results are available to date, one of which is the so-called space-division multiple-access (SDMA) scheme that attempts to multiply the throughput of a wireless local area network (WLAN) by incorporating spatial channels [4], [5], [11], [12]. The SDMA scheme is particularly suitable for fixed wireless networks where the spatial characteristics are relatively stable. For a system with M antenna elements, the underlying idea is to divide each time-division multiple-access (TDMA) slot into M *space slots* so that the total number of noninterfering slots is increased by M fold.

Though intuitively promising, an obvious flaw of this scheme is that spatial channels are rarely orthogonal in practice. If multiple terminals are assigned to one time slot without considering their spatial characteristics, the one with an unfavorable spatial configuration will experience significant throughput disadvantages. Since the effectiveness of spatial separation depends on the basestation array responses (often referred to as the spatial signatures) of all co-slot terminals, the instantaneous signal-to-noise ratio (SNR) of beamforming outputs can vary dramatically. This problem is investigated by Ward and Compton [5] where *collision* due to unresolvable packets (when the arrival angles of multiple packets are within a threshold θ_r) is accounted for. Although the analysis is somewhat simplified, since in the presence of multipath the capability of reception algorithms may not be a sole function of the arrival angles, the study reveals the limitations of the basic SDMA scheme.

A fundamental solution to the above problem is a "channel-aware" MAC that controls the traffic based on the spatial characteristics of the terminals. In SDMA in particular, the performance of the system can be enhanced with spatial signature based scheduling (e.g., assigning the "most orthogonal" terminals to the same time slot to increase the traffic throughput). Such a MAC treatment allows a system to exploit the spatial diversity in an efficient manner with fixed-complexity physical layer processing. Several dynamic scheduling algorithms are proposed and studied in [13]. All of them assume a static homogeneous traffic with terminals transmitting packets at all times. It is shown that under steady traffic, the scheduling algorithm provides a minimum performance gap between the average and the worst case throughput among the terminals, indicating a maximum capacity improvement in worst case limited applications.

In most real scenarios, however, the network traffic is bursty and heterogeneous due to mixture of voice, video and data, each with different quality-of-service (QoS) parameters. In this case, the MAC protocol has to take the traffic dynamics and QoS requirements into account. Based on the existing scheduling algo-

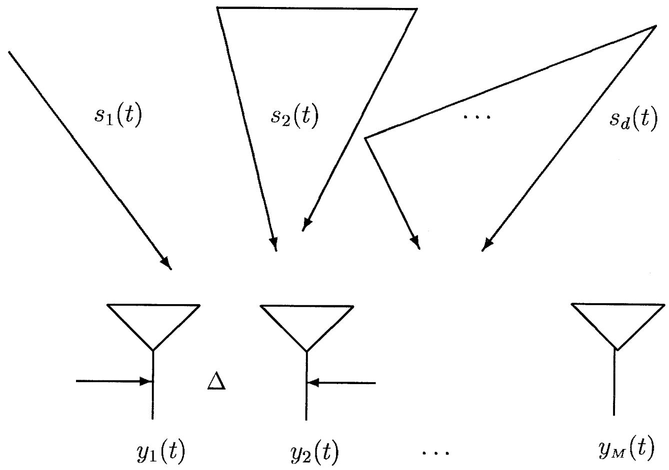


Fig. 1. A typical array system.

rithm, we propose an enhanced scheme for multimedia packet networks. In this protocol, not only the terminals' spatial signatures, but also their QoS parameters are considered in order to optimized the system performance. Simulation shows that a substantial improvement can be obtained than the best effort scheme in which all traffics are treated homogeneously.

The remainder of this paper is organized as follows. In Section II, we provide an overview of slotted antenna array packet networks and evaluate the performance of the SDMA scheduling algorithm. In Section III, we propose a new scheduling protocol for multimedia networks. Simulation studies of the protocol are presented in Section IV. In Section V, we conclude the paper by highlighting our contributions.

II. Antenna Array Packet Networks

A. Spatial Beamforming

Antenna array processing has been an active topic in wireless communications for more than a decade. Through controlling its beam pattern, an antenna array can improve some key operation parameters such as the signal-to-interference-plus-noise-ratio (SINR) over single antenna wireless system. Fig. 1 depicts a typical antenna array system where the spatial diversity is exploited for multiple access communications. In the absence of noise, the response of an M-element antenna array to a narrowband source $s(t)$ can be written as

$$\mathbf{y}(t) \stackrel{\text{def}}{=} [y_1(t)\ y_2(t) \dots y_M(t)]^T = \mathbf{a}s(t)$$

where $(\cdot)^T$ denotes transposition, and $\mathbf{a} = [a(1)\ a(2)\ \dots\ a(M)]^T$ is the array response vector that captures the spatial characteristics (direction-of-arrivals, number of multipath reflections, and attenuation) associated with the terminal. In some literature the array response vector is referred to as the *spatial signature*.

When d terminals communicate simultaneously with the basestation, the total output of the antenna array is given by

$$\mathbf{y}(t) = \sum_{k=1}^{d} \mathbf{a}_k s_k(t) + \mathbf{n}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

$$\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_d],\ \mathbf{s}(t) = [s_1(t), \dots, s_d(t)]^T$$

$$\mathbf{n}(t) = [n_1(t),\ n_2(t),\ \dots, n_M(t)]^T \tag{2}$$

where $\mathbf{n}(t)$ is an additive noise vector and $\mathbf{A}$ is defined as the *array manifold* whose columns are the spatial signatures. To facilitate our presentation in the remainder of this paper, we assume 1) all signals have unit power and $\|\mathbf{a}_k\| = 1$ (i.e., perfect power control) and 2) the noise is i.i.d. with strength σ_n^2. Under these assumptions the covariance matrix of the array output has the form of

$$\mathbf{R}_{\mathbf{yy}} = E\{\mathbf{y}(t)\mathbf{y}^H(t)\} = \mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I}. \tag{3}$$

In most practical situations, spatial signatures of different terminals are different, allowing the basestation to decouple superimposed signals through spatial separation. To retrieve individual signals, e.g., $s_k(t)$, from the antenna outputs, outputs of the antenna array are weighted and summed with a set of weight coefficients

$$\hat{s}_i(t) = \sum_{m=1}^{M} w_i^*(m) y_m(t)$$

with $\{w_i(m)\}$ so designed to constructively combine the signal of interest and destructively combine the interference-plus-noise. This processing is referred to as *spatial beamforming* [14]. The optimum weight vector $\mathbf{w}_k = [w_k(1) \dots w_k(M)]^T$ that minimizes the mean squared error (MSE) of the signal estimate is given by

$$\mathbf{w}_k = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{a}_k$$

in which case the output MSE and SINR, respectively, of the kth signal estimate are

$$MSE_k = 1 - \mathbf{a}_i^H \mathbf{R}_{\mathbf{yy}}^{-1} \mathbf{a}_i; \tag{4}$$

$$SINR_k = \frac{\|\mathbf{a}_k^H \mathbf{R}_{\mathbf{yy}}^{-1} \mathbf{a}_k\|^2}{\sum_{j \neq k}^{K} \|\mathbf{a}_j^H \mathbf{R}_{\mathbf{yy}}^{-1} \mathbf{a}_k\|^2 + \|\mathbf{a}_k^H \mathbf{R}_{\mathbf{yy}}^{-2} \mathbf{a}_k\| \sigma_n^2}. \tag{5}$$

It can be shown that through spatial beamforming, up to M co-channel terminals can be perfectly separated in the absence of noise. The capability of simultaneous communications with multiple terminals using an antenna array makes SMDA very attractive to resource critical wireless applications.

B. SDMA/TDMA Packet Networks

Before we review the concept of SDMA/TDMA, let us describe the packet radio network under consideration. The slotted packet network consists of K radio terminals actively communicating with a basestation. Information is transmitted in blocks of symbols called *packets*. The time axis is divided into time *frames* of fixed length. Each frame is evenly divided into L *time slots* through which multiple packets can be transmitted. Each terminal transmits packets in one out of the L slots at a fixed data rate.

Assume that the basestation has total control of the network traffic. When a terminal has packets to transmit, it places an admission-request packet through a reserved request slot, from which the basestation obtains the spatial signature and traffic information of the terminal. The basestation then assigns certain time slots to the terminal depending on the request and the traffic situation. The focus of this paper is on how to schedule

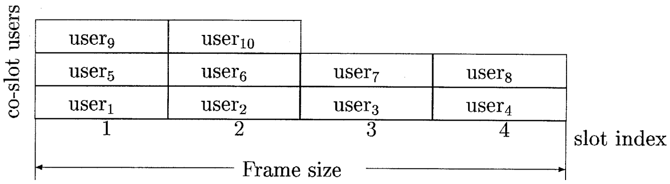


Fig. 2. SDMA without scheduling.

packet transmissions after the terminals are admitted. We do not emphasize access initiation in this work.

In most practical situations $K \gg M$. SDMA/TDMA accommodates these terminals in both the space and the time domain. Fig. 2 illustrates how this objective is achieved in a 4-slot, 10-terminal SDMA/TDMA network [4], [5]. Built upon the slotted TDMA network, the SDMA expands the capacity by allowing multiple (up to M) terminals in each slot, and within each slot spatial beamforming is performed to acquire packets from multiple terminals. In essence, SDMA adds another dimension to the spectrum resource by expanding each slot into M space slots (vertical axis), translating into an M-fold increase of system throughput in an ideal scenario.

In reality however, the ability of capturing multiple packets depends critically on the spatial configuration of the co-slot terminals. Intuitively, by assigning "most-orthogonal" terminals to the same time slot, the worst case and average performance of the system can be improved. This motivates us to design efficient scheduling algorithms that separates "aligned" terminals so that the spatial resource is efficiently exploited. Note that the scheduling criteria depends heavily on the system configuration. Thus there is no single "optimal" solution for all the systems. In the following, we focus on a TDMA-based single-cell system with perfect power control. We try to develop scheduling algorithms to accommodate packets with fixed size into the time slots to improve the system performance.

C. Scheduling Algorithms

Assume that the transmission data rate of each terminal is fixed. Then the terminal has to operate above certain SINR threshold $\mathrm{SINR}_{\min}$ to avoid excessive packet errors. The threshold depends on the coding and modulation scheme and the packet PER requirement. In the scheduling process, a terminal is allocated to a time slot as long as the resulting SINR after beamforming is above the threshold $\mathrm{SINR}_{\min}$

$$\mathrm{SINR} \geq \mathrm{SINR}_{\min}. \tag{6}$$

Shad *et al.* [13] studied several scheduling algorithms to improve the frame capacity. Among all algorithms studied, the Best Fit algorithm has the best performance.

The best-fit scheduling algorithm first pick a terminal randomly from the terminal pool (to ensure the fairness among terminals). The algorithm tests the current unallocated terminal by calculating the SINRs for all co-slot terminals in the current time slot as if the candidate terminal were to be added. The SINR margins of co-slot terminals are calculated by subtracting $\mathrm{SINR}_{\min}$ from the SINR they experienced. Assume there are d terminals in current slot, The SINR margin of the ith terminal in this slot is: $\mathrm{SINR}_{\mathrm{margin}}(i) = \mathrm{SINR}(i) - \mathrm{SINR}_{\min}$. The the SINR margin of this slot is $\mathrm{SINR}_{\mathrm{margin}} = \min\{\mathrm{SINR}_{\mathrm{margin}}(i)\}$. The algorithm would test all the time slots and insert the terminal into the time slot with the largest SINR margin provide that the margin is greater than zero. Otherwise, the terminal is ignored. The algorithm then randomly picks the next terminal. The process continues until all the terminals are tested.

The best-fit scheme implicitly assumes that the traffic is steady, i.e., all terminals have packets to transmit at all times. In real applications the number of terminals changes every time a terminal becomes active/inactive. Under these scenarios it may not be reasonable to perform a full-scale scheduling on all terminals every time when there is a change in network traffic. While optimization for variable traffic patterns is an important topic that requires further investigation, one straightforward strategy is partial best-fit scheduling, i.e., allocating only the newly activated terminals based on their spatial signatures. We shall refer to this scheme as "partial-fit" SDMA. Let $\{m_k\}_{k=1}^K$ be the K mobile terminals in the system, and $\mathbf{m}_l^{f-1} \equiv \{m_{l1} \ldots m_{ld_l}\}$ be the terminal set allocated to the lth time slot in the $f-1$th frame. Then, in the fth frame, the "partial-fit" algorithm updates the terminal allocation as follows:

1 Let $\mathbf{m}_l^f = \mathbf{m}_l^{f-1}$ for $l = 1, \ldots, L$.
2 Remove all the expired terminals. If terminal $m_{lk} \in \mathbf{m}_l^{f-1}$ finishes it's transmission, let $\mathbf{m}_l^f = \mathbf{m}_1^f - m_{lk}$, where subtraction removes the terminal from the set.
3 Add new terminals. If $m_{f1}, \ldots, m_{fn}$ are newly arrived terminals, then they are added to the existing terminal profile $\mathbf{m}_l^f, \quad l = 1, \ldots, L$ according to the "best-fit" algorithm described above.

Performance of the above scheduling algorithms is studied in next section.

D. Performance of Scheduling SDMA

Let $\{\mathbf{a}_k\}_{k=1}^K$ be the spatial signatures of the K terminals in the system and

$$\mathbf{A}_l = [\mathbf{a}_{l1} \ldots \mathbf{a}_{ld_l}]$$

denote the array manifold of the co-slot terminals in the lth slot. The basestation performs spatial beamforming on the received co-slot signals as described in Section II-A. According to Poor and Verdu [15], the residual interference and noise of the MMSE beamforming output has a Gaussian-like characteristic, allowing us to calculate the bit-error rate (BER) based on the output SINR in (5)

$$P_E = Q(\mathrm{SINR}). \tag{7}$$

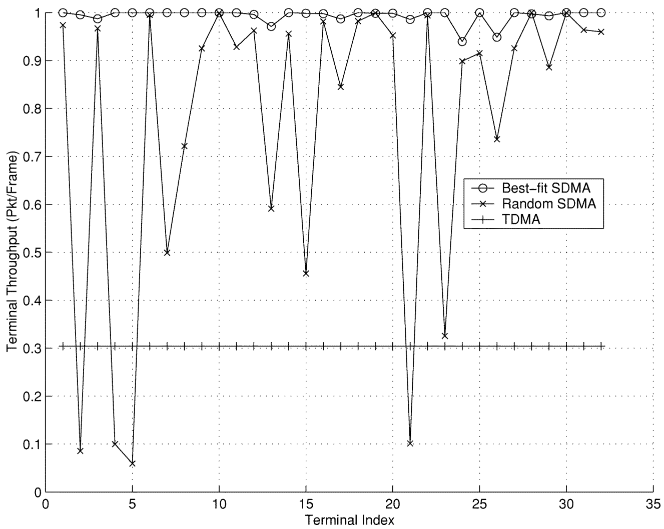


Fig. 3. Throughput improvement of SDMA due to SS-based allocation.

Assuming the error correction capability of a coded packet of N bits is t, i.e., a maximum number of t bits in error is correctable, the *packet success probability* is

$$\begin{aligned} P_k =& P(\mathrm{SINR}_k) \\ =& \sum_{n=0}^{t} \binom{N}{n} Q(\mathrm{SINR}_k)^n \left(1 - Q(\mathrm{SINR}_k)\right)^{N-n}. \end{aligned} \quad (8)$$

Where SINR_k is the beamforming output SINR of the kth terminal. The *throughput* of each terminal in SDMA can be defined accordingly.

Definition 1: The terminal frame throughput S_k is defined as the packet throughput of the kth terminal in a frame. Let L be the total number of time slots in a frame. Note that in a given time slot, the kth terminal can transmit at most one packet. Denote P_{kl} the packet success probability of the kth terminal in the lth slot. If the terminal dose not transmit in this slot, $P_{kl} = 0$. Then, S_k is the summation of all P_{kl}

$$S_k = \sum_{l=1}^{L} P_{kl}. \quad (9)$$

In cases where the frame length is variable, the average number of packets can be successfully transmitted in a time slot is a better performance measure. The average slot throughput is defined as

$$T = \frac{1}{L} \sum_{k=1}^{K} S_k. \quad (10)$$

Fig. 3 illustrates the performance gain of SDMA by comparing the terminal frame throughput of pure TDMA (single antenna), random SDMA (terminals allocated to slots without regarding to their spatial signatures), and best-fit SDMA/TDMA. We consider a TDMA-based single-cell system, where the basestation is equipped with four antennas. Each TDMA frame has ten time slots, and each time slot accommodates packets with fixed length of 256 b. There are 32 terminals in the system. All terminals are under perfect power control and the spatial signatures are generated as independent Gaussian vectors (normalized to unit power). The SNR is set at 20 dB and SINR threshold is 10 dB. Clearly, SDMA/TDMA is fundamentally better than pure TDMA in that each terminal has four times the chance to transmit successfully. However, the worst case performance of random SDMA could be even worse than pure TDMA due to high possibility of collisions of co-channel terminals. With simple MAC layer planning, the system throughput of best-fit SDMA/TDMA increases substantially across the board. More importantly, the variance of the terminals' throughput becomes very small, one can practically use the average throughput as the design parameter for the system capacity. This is particularly important to wireless applications whose capacity is usually worst case limited.

III. Scheduling Protocol for Multimedia Traffic

In the presence of multimedia traffic, different traffic class may have different QoS requirements. For example, the voice traffic is sensitive to delay, but can tolerate certain degree of packet loss. However, the data traffic must not have any missing packet, but can endure longer delay. Therefore, it is beneficial for the MAC protocol to take the specific QoS parameters of each traffic class into account. In the following, we describe a scheduling protocol that significantly increases the performance of a multimedia SDMA network by incorporating the following QoS parameters:

- SINR;
- Packet timeout;
- Packet loss rate.

We consider a single-cell network with one base station and N mobile terminals. Fixed-length packets arrive at the mobiles according to bursty random processes. The packets are buffered at the mobile terminals until they are transmitted to the base station.

When a mobile terminal has packets to transmit, it sends an access packet in the access slot using ALOHA random access. After the basestation successfully receives an access packet from a mobile terminal, it obtains information from the mobile terminal which includes the QoS class of the packets and the spatial signature of the terminal. Based on received information, the basestation then grants certain time slots to the terminal according to a desired packet transmission policy. The transmit permission is broadcasted in the following downlink frame.

The mobile terminal listens to the downlink frame, when it hears the permission, it begins to transmit in the next uplink frame. Each time a mobile terminal transmits a packet, it also includes a Piggybacking Request to indicate whether it has more packets backlogged. The basestation checks the Piggybacking Request, and updates the mobile terminal information accordingly.

Now our focus is on how to design a packet scheduling policy so that the mobile terminals share the limited communications bandwidth in an efficient manner: maximizing throughput and minimizing packet loss. To accommodate the different QoS requirement, a *packet scheduler* is added to the protocol to schedule packet transmissions in such a way that the spatial resource can be used efficiently. The ensuing sections describe

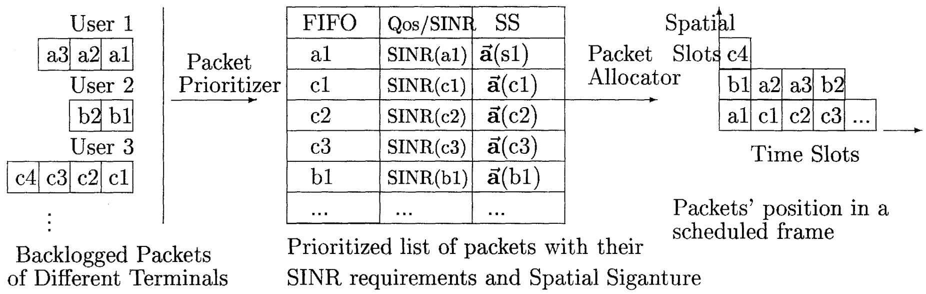


Fig. 4. Packets assigning procedure.

how the scheduler prioritizes and organizes the transmission of packets by taking these criteria into consideration.

The main objective of the scheduler is to maximize throughput and minimize packet loss. This is achieved in two steps:

1) determine packet priorities;
2) allocate prioritized packets into SDMA/TDMA slots.

These tasks are performed by a *packet prioritizer* and a *packet allocator*, respectively [16]. In particular, the packet prioritizer minimizes the packet loss, while the packet allocator maximizes the throughput. The basic protocol is illustrated in Fig. 4.

A. Packet Prioritizer

There are many ways to define packet priorities [16]–[18]. Here, we consider the case where each terminal only supports one type of traffic (voice, data, or video). However, terminals can be easily configured to support heterogeneous traffics by maintaining multiple first-in–first-out (FIFO) queues for different traffics. Since the packets in each queue are of the same type, the priority values can be determined based on their time out requirements.

Assume that the kth terminal has B_k backlogged packets in its buffer, with timeout values $[t_1, t_2, \ldots, t_{B_k}]$. Ideally, if every packet in the queue is allowed to transmit at a minimal constant rate so that it can finish transmission before timeout, the rate of the ith packet is $r_k(i) = 1/t_i$. The aggregate transmission rate at current frame is thus $r_k = \sum_{i=1}^{B_k} r_k(i)$. It indicates how many slots are required by this terminal at current frame. We can use the aggregate transmission rate as the measurement of priority.

Definition 2: The priority of a packet is defined as the aggregate transmission rate of this packet and the remaining packets backlogged after it. So the priority of the ith packet in kth terminal queue can be calculated as

$$P_k(i) = \sum_{j=i}^{B_k} \frac{1}{t_j}. \tag{11}$$

If a packet cannot be transmitted before its time out, it is simply discarded. $P_k(i)$ is a monotonic decreasing function of i. Its value reflects how urgent one packet must be transmitted, because a higher value indicates a higher probability of packet loss. The value also indicates the average transmission rate this terminal should have. If the value is greater than one, it means more than one slot should be assigned to this terminal to avoid possible packet loss. So the priority definition adopted here has the following properties: 1) the priorities of each queues are calculated independently such that it has much lower complexity; 2) the priority based on timeout value gives advantage to the more time sensitive packets (voice, video); 3) the priority values indicate how many slots should be allocated to a terminal in the current frame; and 4) the priority values are calculated based on all buffered packets so that the longer queues which subject to higher packet loss rate can have more advantage. The priority defined here is not optimal because the optimal solution would require to consider all the existing queues jointly. However, given the desirable properties described above, it works well as we will show in the simulations.

After prioritization, the packets are put into a FIFO (an ordered list $\mathbf{L}_p$ according to their priority values). Notice that the frame length is L. So one terminal can be assigned up to L time slots. So one terminal can have at most L priority values, and it can appear in the list for up to L times. As will be explained next, the list $\mathbf{L}_p$ is used by the packet allocator to determine if a terminal should be granted transmission privileges for the next frame, and if so, the number of packets that will be allowed to send. After each frame, the packet prioritizer updates the priority list for the next frame.

B. Multimedia Best-Fit Packet Allocator

After prioritization, each ordered packet enters the buffer with an associated SINR/data rate requirement and a spatial signature. The objective of the packet allocator is to arrange packets that have the highest priority into SDMA/TDMA frames so that the throughput in the next frame is maximized. Toward this end, we modify the best-fit algorithm to accommodate the heterogeneous traffic. First of all, the order of t the backlogged packets are determined by the prioritizer. In addition, we allow one terminal to have more than one slot to transmit packets within each frame. The list $\mathbf{L}_p$ implicitly contains information of the number of packets that the packet allocator should allow a terminal to send. Furthermore, packets belong to different traffic classes may be assigned to one time slot, as long as their own SINR requirements are all satisfied. Taking these factors into consideration, a QoS-aware SDMA scheduling algorithm is summarized as follows.

Multimedia Best-Fit SDMA/TDMA:

- The algorithm picks packets in the order given by list $\mathbf{L}_p$. Each packet is associated with its SINR requirement determined by its QoS class.
- The scheduling starts with the first packet in $\mathbf{L}_p$. The algorithm tries to assign the current packet into one of the time

slots. The algorithm tests the current unallocated packet by calculating the SINR for all co-slot packets in current time slot as if the candidate packet were to be added. The SINR margins of co-slot packets are calculated by subtracting their SINR requirements from the SINR they experienced. Assume there are d terminals in current slot, The SINR margin of the ith terminal in this slot is: $\mathrm{SINR}_{\mathrm{margin}}(i) = \mathrm{SINR}(i) - \mathrm{SINR}_{\mathrm{req}}(i)$. The SINR margin of this slot is $\mathrm{SINR}_{\mathrm{margin}} = min\{\mathrm{SINR}_{\mathrm{margin}}(i)\}$. The algorithm would test all the time slots and insert the current packet into the time slot with largest SINR margin provide that the margin is greater than zero. Otherwise, the packet is ignored.
- The algorithm moves to next packet and continues to search for appropriate slot to insert. The process continues until all packets are tested.

The packet prioritizer orders the packets according to their priority. The packet allocator then searches and allocates packets according to their priority order. The throughput is maximized for the given packet priority order.

IV. Performance Evaluation

Our analysis so far assumes that all terminals have packets to transmit at all times, which is not always true in practical wireless networks [19]. In this section, we present some performance results for a network with more realistic traffic models. The SDMA/TDMA schemes (random SDMA, best-fit SDMA, partial-fit SDMA and multimedia best-fit SDMA) are studied under homogeneous bursty traffic and mixed voice and data traffic.

A. Homogeneous Bursty Traffic

For a network with homogeneous bursty traffic [16], we assume

1) There are a fixed number of terminals in the system each with its own spatial signature.
2) The packets arrive at each terminal according to a bursty traffic model. In this model, the terminal has two alternating states: ON and OFF. During the ON state, packets arrive at the terminal with Poisson distribution. No packet arrives in the OFF period. The length of the ON and OFF periods obey certain distributions.
3) The spatial signature of each terminal varies slowly during the ON period, so that all the packets arrived during one period are transmitted with the same spatial signature.

The above case presents a realistic WLAN type of scenario most suitable for SDMA. In practice, once a terminal becomes ON, it submits a request to the scheduler and the scheduler marks this terminal as active and estimates its spatial signature. It either randomly assigns one slot to this terminal (random SDMA) or allocates a slot based on its spatial signature (best-fit SDMA).

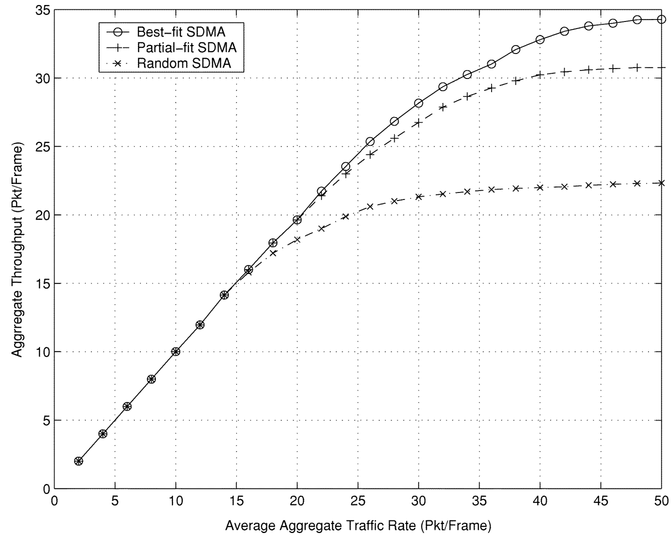


Fig. 5. Throughput with bursty traffic.

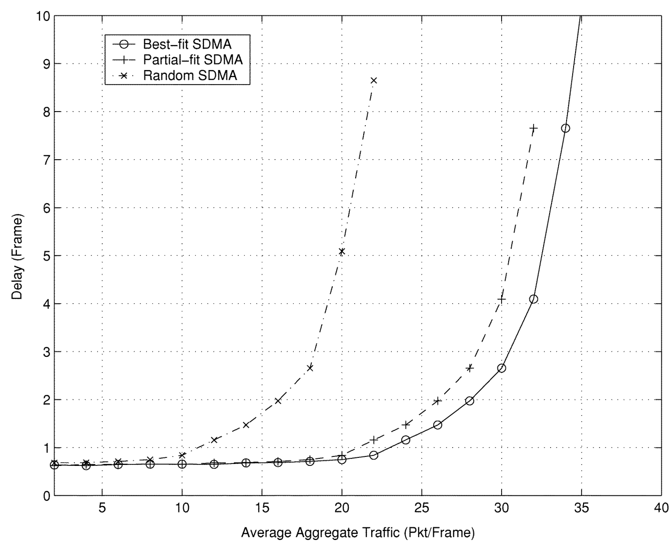


Fig. 6. Delay with bursty traffic.

The terminal, after receiving the scheduling information, begins to transmit the backlogged packets at the given slot.

We simulate a basestation with four antennas serving 160 terminals. There are ten time slots per frame. In each slot, the packet length is 256 b. The SNR is set at 20 dB, and the scheduling SINR threshold is 10 dB. The length of the OFF period is exponentially distributed with mean time 100 frames. The ON period has fixed length 1/2 frame. The average aggregate traffic rate λ is given by the equation at the bottom of the page.

Figs. 5 and 6, respectively, show the throughput and delay of the three SDMA schemes as functions of the aggregate traffic rate λ. For low traffic, the rise in throughput is approximately linear with respect to the traffic rate for all schemes. In this region, there are essentially no collisions due to failure in

$$\lambda = \frac{\text{Traffic Rate during ON period} \times \text{Number of Users} \times \text{ON Period}}{\text{ON Period} + \text{OFF Period}}$$

TABLE I
VOICE TRAFFIC PARAMETERS

Patterns	Mean Value (s)
Main Talkspurt	1.000
Main Gap	1.350
Minispurt	0.275
Minigap	0.050

TABLE II
SIMULATION PARAMETERS

Parameter		Value
Frame Size		16 msec
Slots per Frame N_p		10
Packet Length (bits)		256
Number of Antennas		4
SNR		20dB
Voice Traffic	Average Rate	16 Kbps
	Packet Timeout	2 frames
	SINR Requirement	9db
	Packet Loss	1 percent
Data Traffic	Average Rate	15Kbps
	Packet Timeout	200 frames
	SINR Requirement	12db
	Packet Loss	10^{-6}

beamforming acquisition. As the traffic increases, the effect of scheduling becomes more and more evident. The throughput of best-fit SDMA saturates at around 35 packet/frame when the traffic rate is greater than 50 packet/frame, showing a throughput gain of over 50% than that of random SDMA (without scheduling). The improvement in delay characteristics due to scheduling is as significant as that in throughput. As evident from the figure, there is only a small performance gap (about 10%) between the best-fit SDMA (full scheduling) scheme and the partial-fit SDMA (partial scheduling) scheme for the bursty traffic. This is particularly encouraging for practical bursty data applications because of its reduced system overhead.

B. Multimedia Traffic

We now evaluate the performance of the proposed protocol by considering an SDMA system with two types of traffics: voice and data. Each terminal only generates one type of traffic, and there are the same number of voice terminals and data terminals in the system. The voice traffic model is based on the three-state Markov model in [20]. In the model, it is assumed that a speech source generates patterns of talkspurts and gaps as the result of the talking, pausing, and listening behaviors in a conversation. All spurts and gaps have independent exponentially distributed durations. The values for the speech activity model is listed in Table I. The data packet length and packet arrival interval are modeled as exponentially distributed. The mean data unit size is 30 kB. The simulation parameters are listed in Table II.

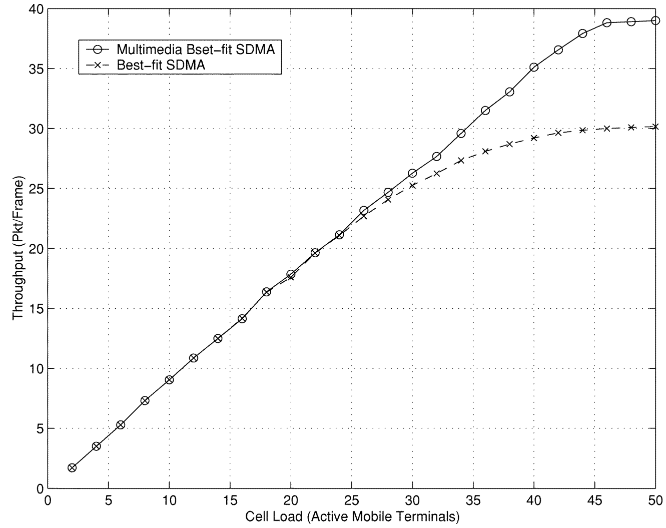


Fig. 7. Throughput per frame vs. cell loads.

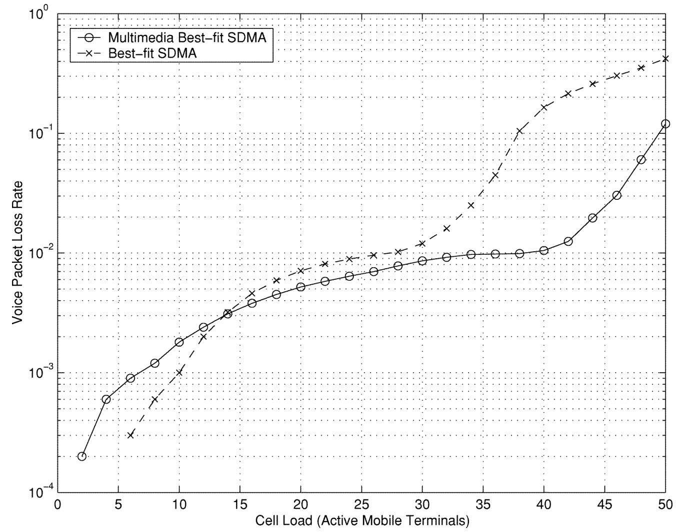


Fig. 8. Packet loss rate of voice traffic.

Simulation results are provided to compare the throughput, delay, and packet loss of the improved protocol (multimedia best-fit SDMA) with that of best-fit SDMA. In the later case, the QoS requirements are worst case limited, e.g., the most strict SINR requirement of all traffic classes must be applied to all classes. In the case studied, the data traffic is assumed to have the highest SINR requirement at 12 dB, which is applied to both data and voice traffic in the simulation.

Fig. 7 compares the throughput characteristics of the two protocols. We observe about 30% of increase of throughput due to QoS-based prioritization. The is because that by taking the QoS into consideration, we are able to pack more voice packets into a frame with an acceptable packet loss penalty. Fig. 8 shows that the new protocol can keep the voice packet loss rate at an acceptable level even when there are a large number active terminals in the system. Assume that the voice packet loss must be kept below 1%, then the QoS-enhanced multimedia SDMA can accommodate up to 40 active terminals, which is 50% than the regular best-fit SDMA. This is because we cannot only pack more voice packets into a frame, but also we give higher priority

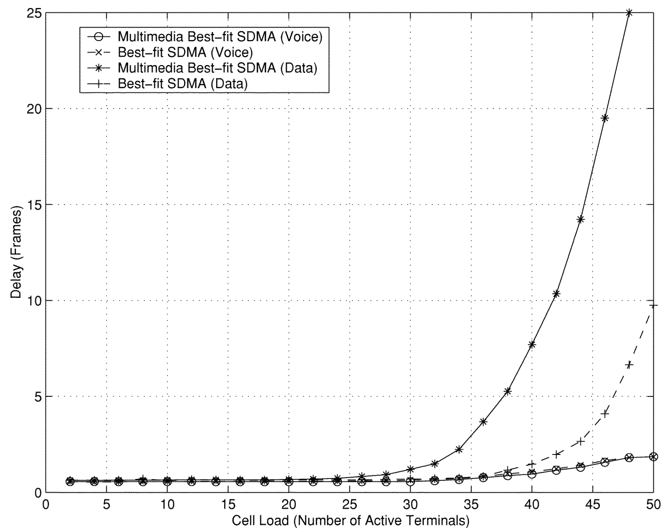


Fig. 9. Delay characteristics of different traffics.

to the voice packets based on their QoS requirements. The increased ability to handle voice traffic is paid by the increased delay of data traffic. Fig. 9 shows that for the multimedia SDMA protocol, the data traffic experiences longer delay. However, because the data traffic is relatively insensitive to delay, it will not cause severe performance degradation. This demonstrates that the new protocol is able to recognize the delay characteristics of different traffic and favors the more delay constrained traffic like voice traffic.

V. Conclusion

In this paper, we have investigated the efficiency of an SDMA/TDMA scheduling protocol. An improved scheduling scheme (multimedia best-fit SDMA) has been proposed for SDMA/TDMA with multimedia traffic. The protocol 1) prioritizes the incoming packets based on their QoS/timeout requirements and 2) judiciously selects co-channel terminals based on their spatial signatures to enable significant improvement in the system throughput-delay characteristics. We also showed the significant performance improvement of the new protocol by simulations.

References

[1] G. V. Tsoulos, "Smart antennas for mobile communication systems: Benefits and channenges," *Electron. Commun. Eng. J.*, vol. 11, Apr. 1999.

[2] B. Suard, G. Xu, H. Liu, and T. Kailath, "Uplink channel capacity of space–division-multiple-access schemes," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1469–1476, July 1998.

[3] C. Farsakh and J. A. Nossek, "On the mobile radio capacity increase through SDMA," in *Accessing, Transmission, Networking Proceedings*: 1998 International Zurich Seminar on Broadband Communications, 1998, pp. 293–297.

[4] G. Xu and S. Q. Li, "Throughput multiplication of wireless LAN's for multimedia services: SDMA protocol design," in *IEEE Proc. GLOBECOM'94*, vol. 3, 1994, pp. 1326–1332.

[5] J. Ward and R. T. Compton, "High throughput slotted ALOHA packet radio networks with adaptive arrays," *IEEE Trans. Commun.*, vol. 41, pp. 460–470, Mar. 1993.

[6] S. Nanda, D. J. Goodman, and U. Timor, "Performance of PRMA: A packet voice protocol for cellular systems," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 584–598, Aug. 1991.

[7] M. J. Karol, Z. Liu, and K. Y. Eng, "Distributed-queueing request update multiple access(DQRUMA) for wireless packet(ATM) networks," in *IEEE Proc. ICC'95*, 1992, pp. 1224–1231.

[8] A. E. Brand and A. H. Aghvami, "Multidimensional PRMA with prioritized bayesian broadcast – A MAC strategy for multiservice traffic over UMTS," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 1148–1161, Nov. 1998.

[9] J. Geier, *Wireless LANs*. New York: Macmillan, 1999.

[10] S. Kumar and D. R. Vaman, "An access protocol for supporting multiple classes of service in a local wireless environment," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 288–302, May 1996.

[11] J.-P. M. Linnartz, "Imperfect sector antenna diversity in slotted ALOHA mobile network," *IEEE Trans. Commun.*, vol. 44, pp. 1322–1328, Oct. 1996.

[12] S. Ghez, S. Verdu, and S. C. Schwartz, "Optimal decentralized control in the random access multipacket channel," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 1153–1163, Nov. 1989.

[13] F. Shad, T. D. Todd, V. Kezys, and J. Litva, "Indoor SDMA capacity using a smart antenna base station," in *IEEE Proc. ICUPC'97*, 1997, pp. 868–872.

[14] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[15] H. V. Poor and S. Verdu, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inform. Theory*, vol. 43, pp. 858–871, May 1997.

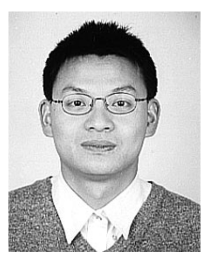
[16] I. F. Akyildiz, D. A. Levine, and I. Joe, "A slotted CDMA protocol with BER scheduling for wireless multimedia networks," *IEEE/ACM Trans. Networking*, vol. 2, pp. 146–158, Apr. 1999.

[17] S. Choi and K. G. Shin, "A cellular wireless local area network with QoS guarantees for heterogeneous traffic," in *IEEE Proc. Infocom'97*, 1997, pp. 1030–1037.

[18] D. S. Lee and C. H. Chou, "Joint control of slot assignments and traffic policing in wireless packet switched networks," in *IEEE Proc.VTC'01*, 2001, pp. 556–560.

[19] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling of PCS networks," *IEEE Trans. Commun.*, vol. 47, pp. 1062–1072, July 1999.

[20] D.J. Goodman and S. X. Wei, "Efficiency of packet reservation multiple access," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 170–176, Feb. 1991.



Hujun Yin (S'99–M'01) received the B.S. degree, in 1995, the M.S.degree from Shanghai Jiaotong University, Shanghai, China, in 1998, and the Ph.D. degree from University of Washington, Seattle, in 2001, all in electrical engineering.

He is currently with ViVATO, Inc., Spokane, WA. His research interests include broadband wireless networks, antenna diversity, and digital signal processing.

Hui Liu (S'99–M'01) received the B.S. degree from Fudan University, Shanghai, China, in 1988, the M.S. degree from Portland State University, Portland, OR, in 1992, and the Ph.D degree from the University of Texas, Austin, in 1995, all in electrical engineering.

From June to December 1996, he served as the Director of Engineering at Cwill Telecommunications, Inc., Austin, TX, where he played a key role in the design and commercialization of a smart antenna S-CDMA wireless local loop system. From September 1995 to July 1998, he held the position of Assistant Professor in the Department of Electrical Engineering at the University of Virginia, Charlottesville. He is currently with the Department of Electrical Engineering at the University of Washington, Seattle. His research interests include broadband wireless networks, array signal processing, digital signal processing and very large scale integration applications, and multimedia signal processing. He has published more than 25 journal articles and has six awarded or pending patents. He is the author of "*Signal Processing Applications in CDMA Communications*" (Norwood, MA; Artech House, 2000).

Dr. Liu's activities for the IEEE Communications Society include memberships on several technical committees and serving as an Editor for the IEEE Transactions in Communications. He is a recipient of the 1997 National Science Foundation (NSF) CAREER Award and the 2000 Office of Naval Research (ONR) Young Investigator Award.