

Space-Time Diversity-Enhanced QoS Provisioning for Real-Time Service over MC-DS-CDMA Based Wireless Networks

Xi Zhang and Jia Tang

Networking and Information Systems Laboratory

Department of Electrical Engineering

Texas A&M University, College Station, TX 77843, USA

Email: {xizhang, jtang}@ee.tamu.edu

Abstract—In order to support the Quality of Service (QoS) requirements for real-time traffic over broadband wireless networks, advanced techniques such as Space-Time Diversity (STD) and Multicarrier Direct-Sequence Code Division Multiple Access (MC-DS-CDMA) are implemented at the physical layer. However, the employment of such techniques evidently affects the QoS provisioning algorithms at the Medium Access Control (MAC) layer. In this paper, we propose a space-time infrastructure and develop a set of cross-layer real-time QoS-provisioning algorithms for admission control, scheduling, and subchannel-allocations. We analytically map the parameters characterizing the space-time diversity onto the admission-control region guaranteeing the real-time QoS. Our analytical analyses show that the proposed algorithms can effectively support real-time QoS provisioning. Also presented are numerical solutions and simulation results showing that the space-time diversity can significantly improve the QoS provisioning for real-time services over wireless networks.

Index Terms—QoS, Space-Time Diversity, Real-Time Service, Wireless Networks, Admission-Control, Cross-Layer Design.

I. INTRODUCTION

The explosive development for wireless network services such as the wireless Internet access, mobile computing, and wireless communications motivate an unprecedented revolution in the wireless broadband access [1]. This presents great challenges in designing the wireless networks since the wireless channel has a significant impact on supporting the real-time Quality of Service (QoS) requirements for different wireless mobile users.

A number of interesting techniques are developed at the physical layer to overcome the impact of wireless channels. Among them, the Space-Time (ST) processing is one of the most significant breakthroughs in wireless communications [2]–[5]. Besides, direct-sequence code division multiple access (DS-SS) integrating orthogonal frequency division multiplexing (OFDM), called multicarrier (MC) DS-SS, emerges as a promising technique for the next-generation wireless communication systems [5][6]. Clearly, employment of such integrated design – combining space-time processing and OFDM-SS – can achieve the integrated diversities from spatial, temporal, frequency, and code domains, which will result in significant improvements in supporting different QoS requirements over wireless networks.

While there have been a large body of literature on how to apply both space-time processing and MC-DS-SS in improving system throughputs at the physical layer, the problems on how to efficiently employ the unique nature of such architectures for designing higher layer protocols and supporting different real-time QoS requirements receives much less attention. Consequently, it becomes increasingly

important to develop the cross-layer scheme to integrate the QoS provisioning algorithms/protocols at higher network layers with the space-time infrastructures implemented at the physical layer.

In this paper, we propose the cross-layer QoS-provisioning scheme and algorithms for supporting QoS requirements of real-time services over space-time MC-DS-SS-based wireless networks. The proposed algorithms include admission control, scheduling, and subchannel-allocation. We derive the analytical analysis to decide the admission region as the function of space-time diversity parameters. Also, we conduct extensive simulations to evaluate the performance of the proposed algorithms. Both analytical analysis and simulations show that the proposed algorithms can efficiently support QoS requirements of real-time services. Moreover, the space-time infrastructure can significantly improve the performance of QoS provisioning in wireless networks.

The paper is organized as follows. Section II describes our proposed space-time MC-DS-SS system architecture. Section III discusses the QoS requirements of real-time services. Section IV maps the space-time diversity onto the admission control for real-time QoS and proposes QoS provisioning algorithms. Section V presents the numerical and simulation results. The paper concludes with section VI.

II. PHYSICAL-LAYER SYSTEM MODEL

We consider the BPSK modulation-based downlink in a packet-cellular wireless network with N antennas at the basestation (BS) and M antennas at each real-time mobile user. Let K denote the total number of the *admitted* real-time users and U the total number of subcarriers or subchannels,¹ which are to be assigned to the total K users. Define the index-set of all K admitted users by $\Omega \triangleq \{1, 2, \dots, K\}$ and the index-set of all U subcarriers by $\Lambda \triangleq \{1, 2, \dots, U\}$. The U subcarrier frequencies are denoted by $\{f_1, f_2, \dots, f_U\}$.

A. Downlink Basestation Transmitter Model

The basestation-transmitter structure of our proposed space-time MC-DS-SS system is shown in Fig. 1. We employ the synchronous transmissions over downlink channels, where the basestation synchronously transmits signals to all mobile users. To simplify the presentation, Fig. 1 only includes and illustrates the transmission downlink for the k th user, where $k \in \Omega$.

As shown in Fig. 1, using the Serial-to-Parallel (S/P) converter, a block of $U_k \cdot N$ BPSK symbols each with bit duration

This work was supported in part by the U.S. National Science Foundation CAREER Award under Grant ECS-0348694.

¹ We use the terms “subchannel” and “subcarrier” interchangeably in the following discussions.

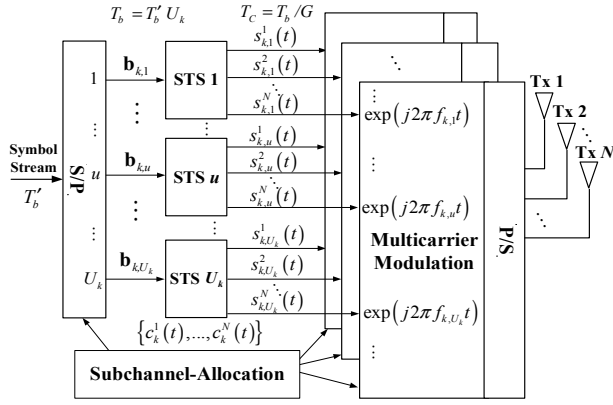


Fig. 1. Downlink block diagram of basestation transmitter for the k th mobile user.

of T_b' is converted to U_k parallel sub-streams. Each of U_k sub-streams consists of N bits, which are denoted by the vector $\mathbf{b}_{k,u} = (b_{k,u}^1 \ b_{k,u}^2 \ \dots \ b_{k,u}^N)^T$, where $(\bullet)^T$ represents the transpose of (\bullet) and $u \in \{1, 2, \dots, U_k\}$. The bit duration T_b after S/P conversion becomes $T_b = T_b' U_k$. The value U_k ($U_k \leq U$) is determined by our proposed subchannel-allocation algorithms, which will be described in Section III with more details. Then, each $\mathbf{b}_{k,u}$ is space-time spreaded (STS) [3] using the spreading code given by

$$\mathbf{c}_k = (c_k^0 \ c_k^1 \ \dots \ c_k^{G-1})^T \quad (1)$$

where $c_k^g \in \{\pm 1\}$, $g \in \{0, 1, \dots, G-1\}$ and G denotes the spreading gain of the code. The chip duration T_C of the spreading code satisfies $T_C = T_b/G = T_b' U_k/G$. In terms of Eq. (1), the waveform expression of the spreading code $c_k(t)$ can be characterized by

$$c_k(t) = \frac{1}{\sqrt{G}} \sum_{g=0}^{G-1} c_k^g p(t - gT_C), \quad 0 \leq t < T_b \quad (2)$$

where $p(t)$ is a normalized rectangular chip waveform which has the finite duration $[0, T_C)$. In this paper, we focus on a specific subset of the general STS schemes investigated in [3], by which the N -bit data is coded, spreaded, and allocated to N transmit antennas, and then transmitted by N time intervals. This kind of STS schemes can achieve the maximal transmit diversity without demanding extra spreading codes and thus be considered as *attractive* schemes [5]. Denote the corresponding space-time block coding square matrix of the u th sub-stream by

$$\mathbf{B}_{k,u} = \begin{pmatrix} b_{k,u}^{1,1} & b_{k,u}^{1,2} & \dots & b_{k,u}^{1,N} \\ b_{k,u}^{2,1} & b_{k,u}^{2,2} & \dots & b_{k,u}^{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k,u}^{N,1} & b_{k,u}^{N,2} & \dots & b_{k,u}^{N,N} \end{pmatrix} \begin{array}{l} \rightarrow \text{space: } N \text{ antennas} \\ \downarrow \text{time: } N \text{ intervals} \end{array} \quad (3)$$

where the rows and columns of matrix $\mathbf{B}_{k,u}$ consist of $\mathbf{b}_{k,u}$ with different signs and sequences according to the orthogonal design rules [4]. The spreading code vector $\mathbf{c}_k(t)$ used for STS can be expressed as

$$\mathbf{c}_k(t) = (c_k^1(t) \ c_k^2(t) \ \dots \ c_k^N(t)) \quad (4)$$

where $c_k^i(t)$, $i \in \{1, 2, \dots, N\}$ has the finite duration $[0, NT_b)$, which is given by

$$c_k^i(t) = \begin{cases} c_k(t - iT_b + T_b), & \text{if } (i-1)T_b \leq t < iT_b \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Using Eqs. (3) and (4), STS can be expressed as

$$\mathbf{s}_{k,u}(t) \triangleq (s_{k,u}^1(t) \ s_{k,u}^2(t) \ \dots \ s_{k,u}^N(t)) = \mathbf{c}_k(t) \mathbf{B}_{k,u}. \quad (6)$$

Following STS, the U_k data streams of each antenna are transmitted simultaneously by modulating U_k different subcarriers, which can be implemented by the operation of IFFT. The frequency spacing Δ between any of the adjacent subcarriers $\{f_1, f_2, \dots, f_U\}$ satisfies $\Delta = 1/T_C$, guaranteeing the orthogonal subcarrier condition. The selection of which U_k out of U subcarriers are used to transmit data is also determined by our proposed subchannel-allocation algorithms. Denote the U_k subcarrier central frequencies assigned to the k th mobile user by $\{f_{k,u} | u=1, 2, \dots, U_k\}$, the transmitted signal $x_{k,n}(t)$ from the n th transmit antenna of the basestation to the k th mobile user within a block-interval $[0, NT_b)$ can be expressed as

$$x_{k,n}(t) = \sqrt{\frac{P}{NU}} \sum_{u=1}^{U_k} \sum_{i=1}^N b_{k,u}^{i,n} c_k^i(t) \exp(j2\pi f_{k,u} t) \quad (7)$$

where P denotes the maximum transmission power for the k th mobile user, which can be achieved by letting $U_k = U$; the coefficient $\sqrt{P/(NU)}$ indicates that the maximum transmission power is independent of the total numbers of transmit antennas and subcarriers; $b_{k,u}^{i,n}$ is given by Eq. (3) and $c_k^i(t)$ is given by Eq. (5), respectively. Clearly, the larger the number of subcarriers U_k assigned to the k th mobile user, the higher bandwidth the k th mobile user can receive. The bandwidth R_k allocated for the k th real-time user can be expressed as

$$R_k = \frac{U_k}{T_b} \quad (8)$$

where the unit of R_k is bit per second (*bps*).

B. Downlink Wireless Channel Model

We assume that the wireless Rayleigh fading channel is frequency-selective, but the delay-spreads T_m of the channel satisfy $T_m \ll T_C$ such that each subchannel conforms to the flat fading. In addition, the channel is assumed to be *quasi-static*, i.e., the fading coefficients are invariant over a block-interval NT_b but vary from one block to another. Thus, during each block-interval NT_b , the fading coefficient of the u th subcarrier $h_{k,u}^{n,m}(t)$, between n th transmit antenna of the basestation and the m th receive antenna of the k th user, can be denoted by $h_{k,u}^{n,m}[i]$, $i=1, 2, \dots$, where i is the discrete time index for the i th block-interval. The time-varying channel can be modeled by an auto-regressive (AR) process [7] as follows

$$h_{k,u}^{n,m}[i] = \alpha_k h_{k,u}^{n,m}[i-1] + v_{k,u}^{n,m}[i] \quad (9)$$

where α_k is determined by the k th mobile user's Doppler velocity and $v_{k,u}^{n,m}[i]$ is a zero-mean independent identical distributed (i.i.d.) complex-Gaussian variable. In Section II and Section III we focus on the discussion within a block-interval $[0, NT_b)$. Therefore, we drop the time index i for convenience. Assume that $\{h_{k,u}^{n,m} | \forall n, m, k, u\}$ are independent² identically distributed (i.i.d.) complex-Gaussian variables with zero-mean and variance of σ^2 per dimension, i.e., $\sigma = E(|h_{k,u}^{n,m}|^2)$.

² Theoretically, the coefficients between different subcarriers are not independent. But the assumption will be valid if we employ frequency-interleaving operation [5][7]. In this paper, we omit the frequency-interleaving, while it is implied, for simplifying the presentation.

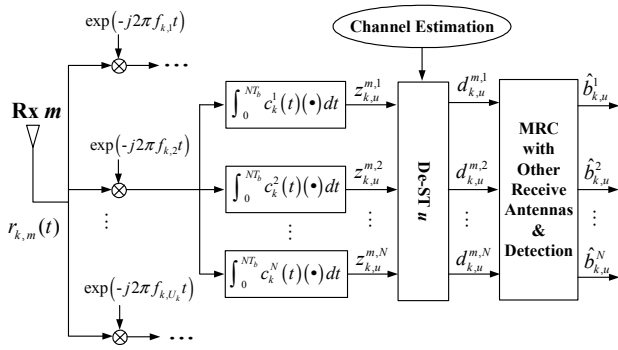


Fig. 2. Downlink block diagram of m th receive antenna at k th mobile user.

Assuming perfect power control, the received signal $r_m(t)$ received at the m th receive antenna of the k th mobile user is given by

$$r_{k,m}(t) = \sqrt{\frac{P}{NU}} \sum_{l=1}^K \sum_{n=1}^N \sum_{u=1}^{U_k} \sum_{i=1}^N h_{k,u}^{n,m} b_{l,u}^{i,n} c_l^i(t) \exp(j2\pi f_{l,u}t) + w_{k,m}(t) \quad (10)$$

where $w_{k,m}(t)$ denotes the complex Additive White Gaussian Noise (AWGN) at the m th receive antenna with zero-mean and double-sided power spectral density of $N_0/2$.

C. Downlink Mobile Receiver Model

The schematic for the m th receive antenna at the k th real-time mobile user of our proposed space-time MC-DS-CDMA system is shown in Fig. 2, where we focus on the decoding scheme within a block-interval $[0, NT_b)$. In addition, we assume that the downlink channel information can be perfectly estimated by the receiver. The basestation can also obtain the channel information by receiver's feedback.

Performing the inverse operation of the transmitter, the received signal $r_m(t)$ at the m th antenna is split to U_k sub-streams by demodulating U_k subcarriers $\{\exp(-j2\pi f_{k,u}t)\}$, where $u \in \{1, 2, \dots, U_k\}$. Then, each sub-stream correlates with the k th mobile user's referenced waveforms $\{c_l^i(t)\}$, where $i \in \{1, 2, \dots, N\}$ during $[0, NT_b)$ to obtain correlation outputs $\mathbf{z}_{k,u}^m = (z_{k,u}^{m,1} \ z_{k,u}^{m,2} \ \dots \ z_{k,u}^{m,N})^T$. Then, the space-time decoding (De-ST) is employed to obtain N decision variables $\mathbf{d}_{k,u}^m = (d_{k,u}^{m,1} \ d_{k,u}^{m,2} \ \dots \ d_{k,u}^{m,N})^T$, corresponding to the original transmitted N bits expressed by $\mathbf{b}_{k,u} = (b_{k,u}^1 \ b_{k,u}^2 \ \dots \ b_{k,u}^N)^T$. Following space-time decoding, all decision variables $\{\mathbf{d}_{k,u}^m | m=1, 2, \dots, M\}$ obtained from M receive antennas are combined together as follows:

$$\mathbf{d}_{k,u} = \sum_{m=1}^M \mathbf{d}_{k,u}^m, \quad k \in \Omega, u \in \Lambda \quad (11)$$

which represents the procedure of Maximum Ratio Combining (MRC). Based upon decision variables given in Eq. (11), the receiver makes the decisions of the transmitted bits by

$$\hat{\mathbf{b}}_{k,u} = \text{sgn}[\text{Re}(\mathbf{d}_{k,u})] \quad (12)$$

where $\text{sgn}(\bullet)$ is the signum function and $\text{Re}(\bullet)$ denotes the real part of (\bullet) . Denote the index-set of real-time mobile users allocated in the u th subchannel by Φ_u . We show in [8] that the

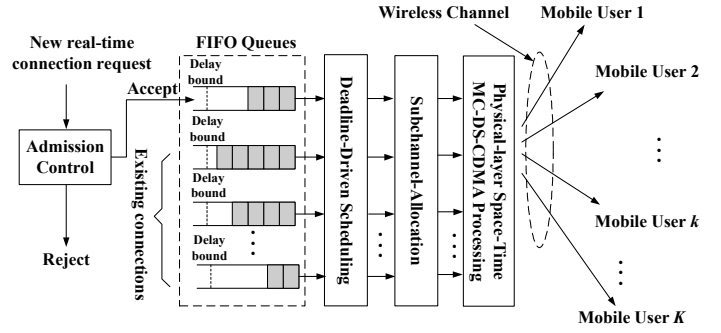


Fig. 3. The QoS provisioning architecture at the MAC layer in basestation.

SINR of decoding signals for the k th user at the u th subchannel can be expressed as follows:

$$\text{SINR}_{k,u} = \frac{\left(\frac{PT_b}{NU}\right) \left(\sum_{n=1}^N \sum_{m=1}^M |h_{k,u}^{n,m}|^2\right)}{\left(\frac{PT_b}{NU}\right) \left(\sum_{n=1}^N \sum_{m=1}^M |h_{k,u}^{n,m}|^2\right) \sum_{l \in \Phi_u, l \neq k} \xi_{k,l}^2 + N_0} \quad (13)$$

where $\xi_{k,l}$ denotes the orthogonal factor between the k th and the l th user.

III. QoS REQUIREMENTS FOR REAL-TIME SERVICE

The QoS provisioning architecture of our space-time MC-DS-CDMA basestation is shown in Fig. 3. We mainly focus on homogeneous real-time traffic in this paper, where all mobile users have the same QoS requirements for the same type of real-time service, such as video or audio. As shown in Fig. 3, the real-time service is provided by connection-oriented communications. In order to guarantee the QoS requirements for real-time mobile users, there are two requirements to be considered as follows.

A. Delay Upper-Bound Requirement

The real-time traffic requires the bounded delays. Once a real-time packet violates its delay bound, it is considered as useless and will be dropped. The delay-bound QoS corresponds to the transmission bandwidth requirement at the physical-layer. Let $\bar{\lambda}$ denote the average rate and λ_{\max} the peak rate of the real-time traffic, respectively. In terms of Eq. (8), if we expect to provide a *queuing-delay-free* transmission, the number U_k of subcarriers allocated to k th user need to satisfy

$$R_k = \frac{U_k}{T_b} \geq \lambda_{\max} \Rightarrow U_k \geq U_{\text{goal}} \triangleq \lceil \lambda_{\max} T_b \rceil, \quad \forall k \in \Omega \quad (14)$$

where $\lceil x \rceil$ represents the smallest integer larger than or equal to x and U_{goal} denotes the minimum number of subcarriers that can guarantee a *queuing-delay-free* transmission. Without loss of generality, we assume that the total number U of subcarriers satisfies $U_{\text{goal}} \leq U$. On one hand, if we assign the number $U_k < U_{\text{goal}}$ subcarriers to the k th user, the queue size may buildup and queuing-delay occurs. On the other hand, if we assign $U_k > U_{\text{goal}}$ subcarriers to the k th user, the bandwidth resources are wasted since the bandwidth exceeding λ_{\max} is not necessary for continuous media (CM) traffic. Therefore, each time when we execute subchannel-allocation algorithm, the *target* number of subcarriers allocated to each user is set to U_{goal} . However, due to the unreliable time-varying wireless

channel, the instantaneous number of subcarriers allocated to the k th user may not be able to reach U_{goal} , i.e., $U_k \leq U_{\text{goal}}$, which is because that our subchannel-allocation algorithm allocates subcarriers to users based on the *instantaneous* SINR (see Section III-B for details). In order to guarantee a specific bounded-delay QoS requirement, the average number of subcarriers \bar{U}_k allocated to each admitted user need to be larger than a subcarrier-threshold denoted by U_{th} , which is not necessary an integer. In fact, U_{th} corresponds to the minimum average service bandwidth requirement \bar{R}_{\min} to guarantee the specific delay-bound QoS, which is given by

$$\bar{R}_{\min} = \frac{U_{th}}{T_b}. \quad (15)$$

Obviously, we need $\bar{R}_{\min} > \bar{\lambda}$ to prevent the queuing-delay from blowing up. In summary, the average number of subcarriers \bar{U}_k allocated to k th admitted user needs to satisfy

$$U_{th} < \bar{U}_k \leq U_{\text{goal}}, \forall k \in \Omega \quad (16)$$

to guarantee the delay-bound QoS requirement. Due to the limited capacity of wireless channel, the number of admitted real-time connections need to be restricted in order to ensure that $\bar{U}_k > U_{th}, \forall k \in \Omega$. Thus, we need to employ the admission test before admitting any new real-time connection. A new real-time connection is admitted only when its bandwidth requirement can be satisfied without violating other existing connections. Otherwise, it will be rejected. Even though we limit the total number of admitted users, it is still impossible to ensure the zero probability that any QoS violation will occur because of the low reliability of time-varying fading channel. To minimize the probability of delay-bound violations, we adopt the deadline-driven scheduling scheme.

B. Loss-Rate Upper-Bound Requirement

While real-time services can tolerate a certain level of packet losses, it still needs a loss-rate upper-bound. If the packet loss-rate is higher than its loss-rate upper-bound, the real-time service is unacceptable. The packet loss-rate corresponds to the bit-error rate (BER) at the physical layer. In order to upper-bound the BER, the SINR of each user at each subcarrier needs to be higher than a SINR-threshold denoted by γ . As described earlier in Section II-C, we assume that the basestation knows the downlink channel information. Consequently, it can pre-compute the instantaneous SINR of decoding signals using Eq. (13). In order to guarantee the loss-rate upper-bound QoS requirement, our proposed subchannel-allocation algorithm (see Section IV-C for details) allocates subcarriers to users based upon the *instantaneous* SINR, which can be expressed as

$$\text{SINR}_{k,u} \geq \gamma, \forall k \in \Phi_u, \forall u \in \Lambda. \quad (17)$$

In the next section, we will describe our admission control, deadline-driven scheduling, and subchannel-allocation algorithms, respectively.

IV. SPACE-TIME DIVERSITY-ENHANCED QoS PROVISIONING ALGORITHMS

A. Admission Control

Let us reconsider Eq. (13) in more details. As mentioned in Section II-B, $\{h_{k,u}^{n,m} | \forall n, m, k, u\}$ are i.i.d. complex-Gaussian

variables with zero-mean and variance of $\sigma^2/2$ per dimension. Therefore, if we define the random part of Eq. (13) by the random variable $X_{k,u}$ as follows

$$X_{k,u} \triangleq \sum_{n=1}^N \sum_{m=1}^M |h_{k,u}^{n,m}|^2 \quad (18)$$

where $X_{k,u}$ satisfies χ^2 distribution with freedom of $2NM$. Due to the homogeneity of all real-time users, we remove the subscript k from $X_{k,u}$ in the following discussions. The probability density function (pdf) of X_u is determined by

$$f_{X_u}(x) = \begin{cases} x^{L-1} e^{-\frac{x}{\sigma}} / \sigma^L \Gamma(L), & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (19)$$

where $L = NM$ and $\Gamma(\bullet)$ denotes the Gamma function. For simplicity, we assume that the orthogonal factor $\xi_{k,l}$ between the k th and the l th users satisfies the following condition:

$$\xi_{k,l}^2 = \begin{cases} 1, & \text{if } k = l \\ \rho, & \text{if } k \neq l \end{cases} \quad (20)$$

where ρ is a positive constant. Then, Eq. (13) can be simplified as

$$Z_u \triangleq \text{SINR}_{k,u} = \frac{\left(\frac{PT_b}{NU}\right) X_u}{\left(\frac{PT_b}{NU}\right) (\mathcal{K}_u - 1) \rho X_u + N_0} \quad (21)$$

where we define Z_u instead of $\text{SINR}_{k,u}$ for simplicity. Let the cardinality $\mathcal{K}_u \triangleq \|\Phi_u\|$ denote the number of mobile users within the u th subcarrier. Clearly, $(\mathcal{K}_u - 1)$ represents the number of co-subchannel interferences at the u th subcarrier.

Given \mathcal{K}_u users within the u th subcarrier, the probability $\phi(\mathcal{K}_u)$ that the k th user satisfies its SINR threshold can be expressed as

$$\begin{aligned} \phi(\mathcal{K}_u) &\triangleq \Pr\{Z_{k,u} \geq \gamma \mid \|\Phi_u\| = \mathcal{K}_u\} \\ &= \begin{cases} \Pr\left\{X_u \geq \frac{\gamma N_0}{\left(\frac{PT_b}{NU}\right) [1 - (\mathcal{K}_u - 1) \gamma \rho]}\right\}, & \text{if } \mathcal{K}_u < \frac{1}{\gamma \rho} + 1 \\ 0, & \text{otherwise} \end{cases} \quad (22) \end{aligned}$$

where we define $\phi(\mathcal{K}_u)$ instead of $\Pr\{Z_{k,u} \geq \gamma \mid \|\Phi_u\| = \mathcal{K}_u\}$ for simplicity. Equation (22) provides us with an upper-bound of the number of co-subchannel users for a given SINR-threshold γ . When $\mathcal{K}_u < 1/(\gamma \rho) + 1$ is satisfied, the probability $\phi(\mathcal{K}_u)$ can be calculated by

$$\phi(\mathcal{K}_u) = \int_{\gamma N_0 / \left\{ \left(\frac{PT_b}{NU}\right) [1 - (\mathcal{K}_u - 1) \gamma \rho] \right\}}^{+\infty} \frac{x^{L-1} e^{-\frac{x}{\sigma}}}{\sigma^L \Gamma(L)} dx. \quad (23)$$

Our proposed subchannel-allocation algorithm assigns U_k ($U_k \leq U_{\text{goal}}$) subcarriers to the k th user. Therefore, given \mathcal{K} users within *each* subcarrier³, i.e., $\mathcal{K}_u = \mathcal{K}, \forall u \in \Lambda$, the

³ Due to the assumption that the fading coefficients between different subchannels are statistically independent, the average number of users allocated to each subcarrier is the same.

probability $\psi(\zeta, \mathcal{K})$ that the k th user is assigned ζ subcarriers follows the Binomial distribution as follows:

$$\psi(\zeta, \mathcal{K}) \triangleq \Pr\{U_k = \zeta \mid \|\Phi_u\| = \mathcal{K}, \forall u \in \Lambda\}$$

$$= \begin{cases} \binom{U}{\zeta} [\phi(\mathcal{K})]^\zeta [1 - \phi(\mathcal{K})]^{U-\zeta}, & \text{if } \zeta < U_{\text{goal}} \\ \sum_{i=U_{\text{goal}}}^U \binom{U}{i} [\phi(\mathcal{K})]^i [1 - \phi(\mathcal{K})]^{U-i}, & \text{if } \zeta = U_{\text{goal}}. \end{cases} \quad (24)$$

Thus, given \mathcal{K} users within each subcarrier, the average number $\bar{U}(\mathcal{K})$ of subcarriers assigned to the k th user can be expressed as

$$\bar{U}(\mathcal{K}) = \sum_{\zeta=1}^{U_{\text{goal}}} \zeta \psi(\zeta, \mathcal{K}). \quad (25)$$

The corresponding average bandwidth $\bar{R}(\mathcal{K})$ allocated to the k th user can be calculated by

$$\bar{R}(\mathcal{K}) = \frac{\bar{U}(\mathcal{K})}{T_b}. \quad (26)$$

As discussed in Section III, in order to ensure the delay-bound QoS requirement, we assume that the average number $\bar{U}(\mathcal{K})$ of subcarriers assigned to each user is larger than a subcarrier-threshold U_{th} , i.e., $\bar{U}(\mathcal{K}) > U_{th}$, or equivalently, $\bar{R}(\mathcal{K}) > \bar{R}_{min}$. Since every admitted user should satisfy $\bar{U}(\mathcal{K}) > U_{th}$, the total number K of admitted users can be approximated by

$$K \approx \frac{\mathcal{K}U}{U_{th}} \Rightarrow \mathcal{K} \approx \frac{KU_{th}}{U}. \quad (27)$$

By substituting Eq. (27) into Eq. (26), we obtain the one-to-one mapping-table $(K, \bar{R}(KU_{th}/U))$ between the number K of admitted users and the average achievable bandwidth $\bar{R}(KU_{th}/U)$ per user.

Suppose that the number of admitted real-time connections is i when the basestation receives a new real-time connection request. The basestation need only to lookup the mapping-table to decide whether the average achievable bandwidth $\bar{R}((i+1)U_{th}/U)$ can still satisfy the delay-bound QoS requirement $\bar{R}((i+1)U_{th}/U) > \bar{R}_{min}$. If so, the new connection request will be admitted. Otherwise, it will be rejected. The mapping table described above can be calculated off-line and stored at the basestation, without costing run-time CPU resources.

B. Deadline-Driven Scheduling

The wireless channel has a significant impact on supporting various QoS requirements of different users. Even though we employ admission control scheme to limit the total number of admitted real-time users, we still cannot eliminate the probability of delay-bound violations. In fact, we can only *statistically* guarantee the bounded delay. To further minimize the probability of delay-bound violations, we adopt the *deadline-driven scheduling* algorithm.

Note that in our CDMA system, data for all K users are transmitted *simultaneously*. The scheduling is to adjust the sequence of executing the subchannel-allocation. As will be shown in Section IV-C, according to our subchannel-allocation algorithm, the user which is allocated earlier has

```

01. // The allocation sequence is based on deadline-driven scheduling
02. for (j := 1 to K) {
03.   for (u := 1 to U) {
04.     if (U_j < U_goal) {
05.       SINR testing using Eq. (28);
06.       if (Eq. (28) is satisfied) { U_j := U_j + 1; }
07.     }
08.     else break; // user's maximum bandwidth requirement is satisfied
09.   }
10. }

```

Fig. 4. Pseudocode of subchannel-allocation algorithm.

advantage to occupy more bandwidth than the later allocated ones. This is due to the fact that SINR criterion described in Eq. (17) is easier to be satisfied by the earlier allocated users. As the result, the later allocated users have higher probability of violating delay-bounds. If we employ a RR based scheduling scheme allocating all K users, the later allocated users are more likely to deteriorate the QoS performance of the whole system. To solve this problem, we adopt deadline-driven scheduling, where the user with the earliest deadline will be allocated first (EDF).

Let Q_{max} denote the maximum queue size corresponding to the specific delay-bound. Each time when we execute scheduling algorithm, the current queue size Q_k of each real-time users are measured. Then, all admitted users are sorted by $D_k = Q_{max} - Q_k$. If $D_k < 0$ (i.e., delay-bound violated), the corresponding packet is useless and will be dropped from the k th user's queue. For those users whose $D_k \geq 0$, the user which has the smallest D_k will be allocated first.

C. Subchannel-Allocation Algorithm

Let the j th user ($j \in \Phi_u$) be the candidate which attempts to transmit bits using the u th subcarrier. Since the basestation knows the information and the statistical characteristics about the channel, we can pre-compute the SINR of decoding signals for each user at the u th subcarrier using Eq. (13). The j th user can be assigned to the u th subcarrier if and only if

$$\begin{cases} \text{SINR}_{j,u} \geq \gamma \\ \text{SINR}_{k,u} \geq \gamma, \forall k \in \Phi_u \end{cases} \quad (28)$$

where

$$\text{SINR}_{j,u} = \frac{\left(\frac{PT_b}{NU}\right) \left(\sum_{n=1}^N \sum_{m=1}^M |h_{j,u}^{n,m}|^2\right)}{\left(\frac{PT_b}{NU}\right) \left(\sum_{n=1}^N \sum_{m=1}^M |h_{j,u}^{n,m}|^2\right) \sum_{k \in \Phi_u} \xi_{j,k}^2 + N_0} \quad (29)$$

and

$$\text{SINR}_{k,u} = \frac{\left(\frac{PT_b}{NU}\right) \left(\sum_{n=1}^N \sum_{m=1}^M |h_{k,u}^{n,m}|^2\right)}{\left(\frac{PT_b}{NU}\right) \left(\sum_{n=1}^N \sum_{m=1}^M |h_{k,u}^{n,m}|^2\right) \sum_{l \in \Phi_u, l \neq k} \xi_{k,l}^2 + N_0} \quad (30)$$

If Eq. (28) is satisfied, the j th user is qualified to be assigned to the u th subcarrier. Otherwise, if any of the SINR in Eq. (28) is lower than the threshold γ , the j th user cannot be assigned to the u th subcarrier.

Based on testing procedure given by Eq. (28), we present our subchannel-allocation algorithm by the pseudocode in Fig. 4. All K

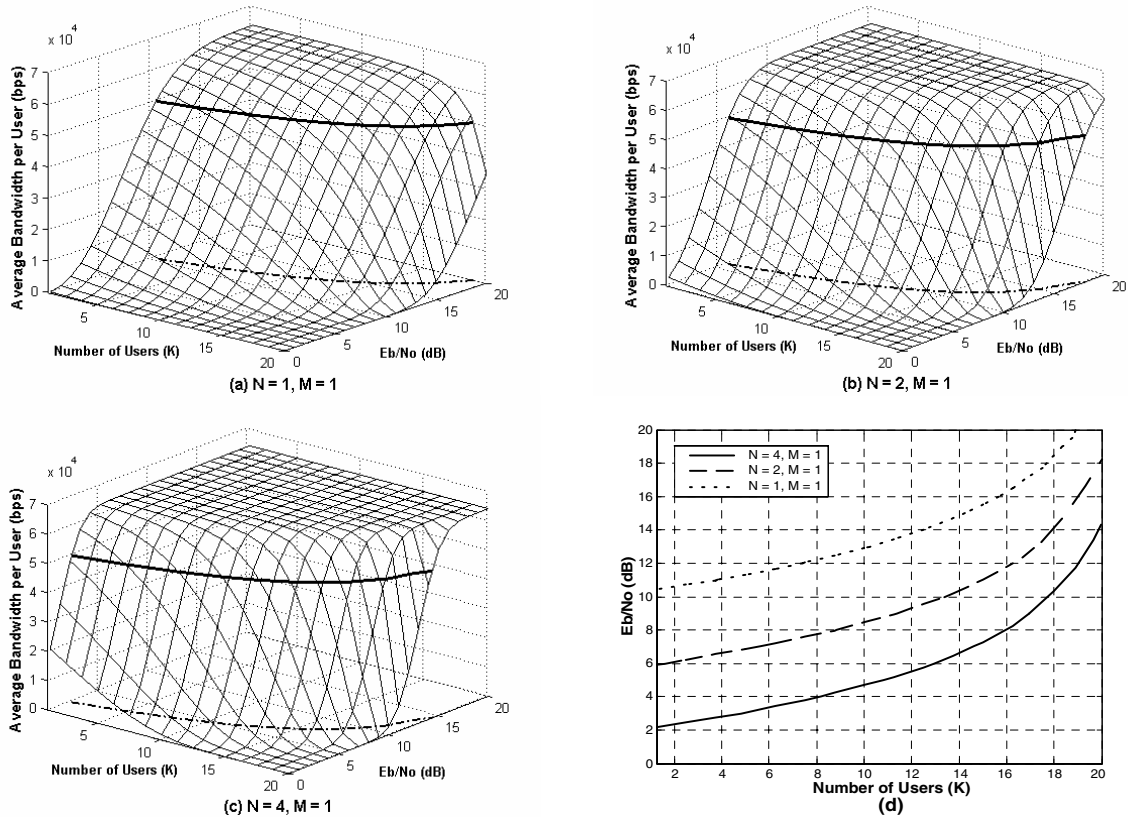


Fig. 5. Average bandwidth and the admission region versus the user number K and the SNR per bit E_b/N_0 with different number of transmit antennas.

users sequentially execute the SINR test (see Fig. 4, step-05). The sequence is determined by our deadline-driven scheduling scheme. If Eq. (28) is satisfied, the j th user is assigned one more subcarrier (step-06). Once the user obtains the number of subcarriers equal to U_{goal} , it achieves its maximum bandwidth requirement. We do not allocate any more subcarrier to it (step-08) and start searching bandwidth for the next user. The procedure repeats itself until all users are allocated to our system.

V. NUMERICAL AND SIMULATION RESULTS

We investigate the performance of our proposed system infrastructure and QoS provisioning algorithms by numerical analyses and simulations. The system chip duration is set to $T_C = 7.8125 \mu\text{s}$ and spreading gain G set to 16. Thus, the bit duration is $T_b = T_C G = 125 \mu\text{s}$ and bit rate per subcarrier is 8 Kbps. We set the total number of subcarriers equal to $U = 10$. Therefore, the total bandwidth of the system is $U/T_C = 1.28 \text{ MHz}$. The packet size is set to 1024 bits. The square of the orthogonal factor ρ between different users is set to 0.0125 and Doppler frequency is 20 Hz for all users. The SINR-threshold is set to $\gamma = 7$ dB, which corresponds to the BER lower than 10^{-3} . Since we mainly focus on transmit diversity in this paper, the number of receive antenna at each mobile user is set to $M = 1$ for all experiments. The cases of $M > 1$ will be presented in the full version of this paper. Also, we omit the guard time in our performance analyses for simplicity.

In simulations, we employ a simple traffic model used in [9] where each real-time user's traffic is modeled as a three-state Markov chain. In the first state, the packets are generated at the rate of 32 Kbps, which corresponds to the bandwidth of 4 subcarriers; In the second state, at the rate of 48 Kbps (6

subcarriers) and the third state at the rate of 64 Kbps (8 subcarriers). Real-time traffic will reside in each of the three state with probability 25%, 50%, and 25%, respectively. We set $U_{\text{goal}} = 8$ and $\bar{R}_{\text{min}} = 50$ Kbps to guarantee the upper-bounded delay. The corresponding U_{th} can be calculated by $U_{\text{th}} = \bar{R}_{\text{min}} T_b = 6.25$.

Using admission control strategy described in Section IV-A, Fig. 5 (a)-(c) plot the numerical solutions of average bandwidth per user $\bar{R}(KU_{\text{th}}/U)$ (see Eqs. (26) and (27)) versus two independent parameters the number K of users and the SNR per bit E_b/N_0 with different number of transmit antennas, where $E_b = PT_b\sigma/(NU)$. The contour-lines of the three figures are drawn at the service bandwidth equal to $\bar{R}_{\text{min}} = 50$ Kbps, indicating that the average bandwidth needs to be larger than 50 Kbps. The projections of the contour-lines onto the two-dimensional-space spanned by user number K and SNR per bit E_b/N_0 are drawn by the dotted lines, which lower-bound the admission regions of the systems. We can see from Fig. 5 (a)-(c) that the average bandwidth per user increases when SNR increases and K decreases, which is expected since the higher SNR represents the larger channel capacity and the smaller K represents the lower interferences. When the number of transmit antennas increases, the area in SNR- K spanned plane corresponding to curved surface above the 50 Kbps contour-lines increases significantly, indicating that the average bandwidth per user becomes larger and larger. The maximum bandwidth per user is always equal to 64 Kbps, which is determined by the parameter $U_{\text{goal}} = 8$ (64 Kbps).

Fig. 5 (d) plots the projections of the contour-lines onto the two-dimensional-space spanned by the user number K and SNR per bit E_b/N_0 with different antenna combinations. The

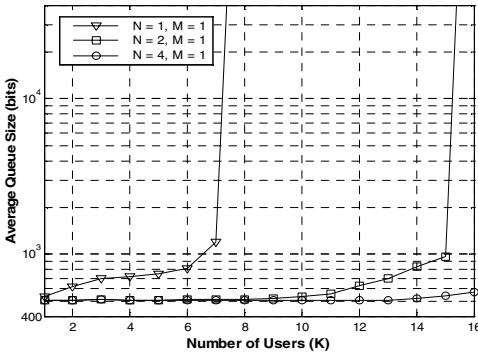


Fig. 6 Average queue size versus the number K of users.

areas above the projection-lines represent the admission regions. We can see from Fig. 5 (d) that the admission region significantly increases as the number N increases. Thus, Fig. 5 shows that the space-time diversity can significantly enhance the QoS provisioning for real-time services.

Fig. 6 plots the simulated average queue size versus the number K of users by employing our proposed QoS provisioning algorithms with the different number of transmit antennas. The SNR per bit E_b/N_0 is equal to 12 dB in simulations. We can see from Fig. 6 that the average queue size increases as the number K of users increases. When K reaches to a specific number K_0 , the average queue size blows up, indicating that the delay cannot be bounded. Clearly, the admitted number of users needs to be smaller than K_0 . For example, when the number of transmit antenna $N=1$, the number of admitted users should be smaller than $K_0=8$. When the number of transmit antenna $N=2$, the number of admitted users should be smaller than $K_0=16$. Agreeing with the conclusion drawn from Fig. 5, Fig. 6 shows that the space-time diversity can provide significant improvement for real-time QoS provisioning.

Furthermore, the simulated admission region shown in Fig. 6 agrees well with the numerical results drawn in Fig. 5 (d). In Fig. 5 (d) we can find that when the SNR per bit $E_b/N_0=12$ dB and the number of transmit antennas $N=1$, the number of users $K=8$ is out of the admission region. When the number of transmit antennas $N=2$, the number of users $K=16$ is within the admission boundary, which is close to our simulated results where the admission region is $K \leq 15$. When the number of transmit antennas $N=4$, both simulation and numerical results show that when the number of users $K \leq 16$, they all can be admitted, verifying the validity of our numerical analysis.

Finally, Fig. 7 plots the simulated probability of delay-bound violation versus user number K by employing different scheduling scheme. The SNR per bit E_b/N_0 is equal to 12 dB and the delay-bound $Q_{\max}=1200$ bits in simulations. We can see from Fig. 7 that the probability of delay-bound violation increases as K increases. If the number of admitted users K goes out of the admission region, the probability of delay-bound violation approaches to 1, which is consistent with the results shown in Fig. 6. Also, we observe from Fig. 7 that the deadline-driven scheduling scheme significantly outperforms Round-Robin scheduling scheme. For instance, the probability of delay-bound violation is reduced by more than 50% (from 0.43 to 0.19) by using deadline-driven scheduling

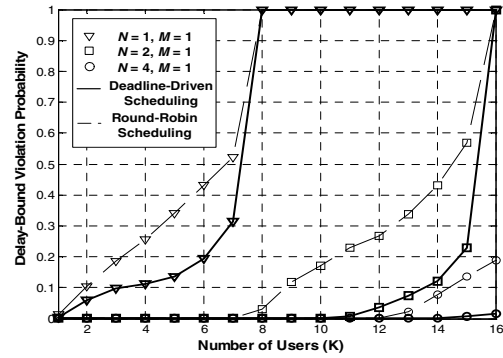


Fig. 7. The probability of delay-bound violation versus the number K of users.

instead of Round-Robin scheduling algorithm when $N=1$ and $K=6$. Moreover, the delay-bound violation probability is decreased by more than 60% (from 0.43 to 0.13) by applying deadline-driven scheduling when $N=2$ and $K=14$. Fig. 7 also illustrates the advantages of using space-time infrastructure at the physical-layer since Fig. 7 shows that the more the number of transmit antennas, the lower the delay-bound violation probability the system can achieve.

VI. CONCLUSION

We proposed the QoS provisioning algorithms including admission control, scheduling, and subchannel-allocation for real-time service over space-time MC-DS-CDMA-based wireless networks. We also derived the analytical analysis to find the admission region of our space-time system. Extensive simulations are conducted to verify the validity of our analytical analysis and to evaluate the performance of the proposed QoS provisioning algorithms. Both analytical analysis and simulation results show that the proposed algorithms can support more efficient real-time QoS guarantee. Thus, the space-time infrastructure can significantly improve the performance of the QoS provisioning in the wireless networks.

REFERENCES

- [1] Theodore Rappaport, A. Annamalai, R. Buehrer, and W. Tranter, "Wireless Communications: Past Events and a Future Perspective", *IEEE Comm. Magazine*, May, 2002, pp. 148-161.
- [2] D. Gesbert, M. Shafiq, D. Shiu, P. Smith, and A. Naguib, "From Theory to Practice: An Overview of MIMO Space-Time Coded Wireless Systems", *IEEE J. Select Areas Comm.*, vol. 21, no. 3, April 2003, pp. 281-302.
- [3] B. Hochwald, T. Marzetta, and C. Papadias, "A Transmitter Diversity Scheme for Wideband CDMA Systems Based on Space-Time Spreading", *IEEE J. Select Areas Comm.*, Vol. 19, Jan. 2001, pp. 48-60.
- [4] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-Time Block Codes from Orthogonal Designs", *IEEE Trans. Inform. Theory*, Vol. 45, No. 5, Jul. 1999, pp. 1456-1467.
- [5] L-L Yang and L. Hanzo, "Space-Time Spreading Assisted Broadband MC-CDMA", *IEEE 55th VTC*, Vol. 4, 2002, pp. 1881-1885.
- [6] ———, "Performance of Generalized Multicarrier DS-CDMA Over Nakagami- m Fading Channels", *IEEE Trans. Comm.*, Vol. 50, No. 6, Jun. 2002, pp. 956-966.
- [7] D. Kalofonos, M. Stojanovic, and J. Proakis, "Performance of Adaptive MC-CDMA Detectors in Rapidly Fading Rayleigh Channels", *IEEE Trans. Wireless Comm.*, Vol. 2, No. 2, Mar. 2003, pp. 229-239.
- [8] J. Tang and X. Zhang, "Subchannel-Allocation Algorithms and Performance Analysis for Space-Time OFDM-CDMA Based Systems in Wireless Networks", Technical Report, EE Dept. Texas A&M University.
- [9] S. Choi and K. G. Shin, "An Uplink CDMA System Architecture with Diverse QoS Guarantees for Heterogeneous Traffic", *IEEE Trans. Networking*, Vol. 7, No. 5, Oct. 1999, pp. 616-628.