WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. (in press)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/wcm.515

Space-time diversity-enhanced QoS provisioning for real-time service over MC-DS-CDMA based wireless networks

Xi Zhang[1*,†], Jia Tang[1] and Hsiao-Hwa Chen[2]
[1]*Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, U.S.A.*
[2]*Department of Engineering Science, National Cheng Kung University, Taiwan, Republic of China*

Summary

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 STD 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 STD can significantly improve the QoS provisioning for real-time services over wireless networks. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: quality-of-service (QoS); space-time diversity (STD); real-time service; wireless networks; admission control; cross-layer design

1. 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-CDMA) integrating orthogonal frequency division multiplexing

*Correspondence to: Xi Zhang, Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, U.S.A.
†E-mail: xizhang@ece.tamu.edu

(OFDM), called multicarrier direct-sequence code division multiple access (MC-DS-CDMA), emerges as a promising technique for the next-generation wireless communication systems [5,6]. Clearly, employment of such integrated design combining ST processing 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 has been a large body of literature on how to apply both ST processing and MC-DS-CDMA 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 requirement 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 ST infrastructures implemented at the physical layer.

In this paper, we propose the cross-layer scheme for supporting QoS requirements of real-time services over ST MC-DS-CDMA-based wireless networks. We study the impact of the physical-layer ST MC-DS-CDMA infrastructure on QoS provisioning performance, which shows that efficiently utilizing the physical-layer resources can significantly improve the QoS performance of upper-protocol layers. The proposed algorithms include admission control, scheduling, and subchannel-allocation. We derive the analytical analysis to decide the admission region as the function of STD 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 ST infrastructure can significantly improve the performance of QoS provisioning in wireless networks.

The paper is organized as follows. Section 2 describes our proposed ST MC-DS-CDMA system architecture. Section 3 discusses the QoS requirements of real-time services. Section 4 maps the STD onto the admission control for real-time QoS and proposes QoS provisioning algorithms. Section 5 presents the numerical and simulation results. The paper concludes with Section 6.

2. Physical-Layer System Model

We consider the binary phase shift keying (BPSK) modulation-based downlink in a packet-cellular wireless network with N antennas at the base station (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,[‡] which are to be assigned to the total K users. Define the index-set of all the K admitted users by $\Omega \triangleq \{1, 2, \ldots, K\}$ and the index-set of all the U subcarriers by $\Lambda \triangleq \{1, 2, \ldots, U\}$. The U subcarrier frequencies are denoted by $\{f_1, f_2, \ldots, f_U\}$.

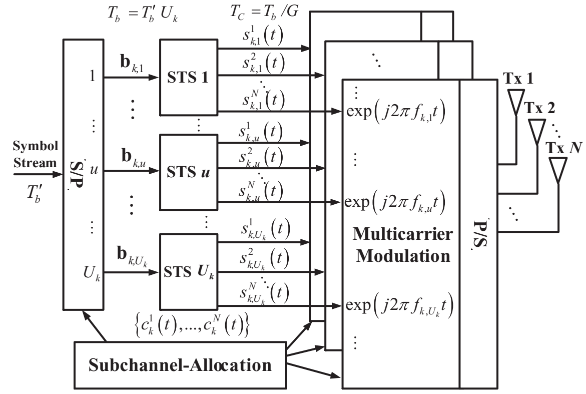

Fig. 1. Downlink block diagram of BS transmitter for the kth mobile user.

2.1. Downlink BS Transmitter Model

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

As shown in Figure 1, using the serial-to-parallel (S/P) converter, a block of $U_k \cdot N$ BPSK symbols each with bit duration 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 $\mathbf{b}_{k,u} = \left(b^1_{k,u}, b^2_{k,u}, \ldots, b^N_{k,u}\right)^{\mathrm{T}}$ where $(\cdot)^{\mathrm{T}}$ represents the transpose of $(\cdot)$ and $u \in \{1, 2, \ldots, U_k\}$. The bit duration T_{b} after S/P conversion becomes $T_{\mathrm{b}} = T'_{\mathrm{b}} U_k$. The value U_k $(U_k \leq U)$ is determined by our proposed subchannel-allocation algorithms, which will be described in Section 3 with more details. Then, each $\mathbf{b}_{k,u}$ is space-time spread (STS) [3] using the spreading

[‡] We use the terms subchannel and subcarrier interchangeably in the following discussions.

DOI: 10.1002/wcm

code given by

$$\mathbf{c}_k = \begin{pmatrix} c_k^0 & c_k^1 & \cdots & c_k^{G-1} \end{pmatrix}^{\mathrm{T}} \tag{1}$$

where $c_k^g \in \{\pm 1\}$, $g \in \{0, 1, \ldots, G-1\}$ 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 Equation (1), the waveform expression of the spreading code $c_k(t)$ can be characterized by

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

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 Reference [3], by which the N-bit data is coded, spread, and allocated to N transmit antennas, and then transmitted by N time intervals. These kinds 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 ST block coding square matrix of the uth substream by

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

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

$$\mathbf{c}_k(t) = \begin{pmatrix} c_k^1(t) & c_k^2(t) & \cdots & c_k^N(t) \end{pmatrix} \tag{4}$$

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

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

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

$$\begin{aligned} \mathbf{s}_{u,k}(t) &\triangleq \begin{pmatrix} s_{k,u}^1(t) & s_{k,u}^2(t) & \cdots & s_{k,u}^N(t) \end{pmatrix} \\ &= \mathbf{c}_k(t)\, \mathbf{B}_{u,k} \end{aligned} \tag{6}$$

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 inverse fast Fourier transform (IFFT). The frequency spacing between any of the adjacent subcarriers $\{f_1, f_2, \ldots, f_U\}$ satisfies $\Delta \triangleq 1/T_c$, 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 kth mobile user by $\{f_{k,u} \mid u = 1, 2, \ldots, U_k\}$, the transmitted signal $x_{k,n}(t)$ from the nth transmit antenna of the BS to the kth mobile user within a block-interval $[0, NT_b)$ can be expressed as

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

where P denotes the maximum transmission power for the kth 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 number of transmit antennas and subcarriers; $b_{k,u}^{i,n}$ is given by Equation (3) and $c_k^i(t)$ is given by Equation (5), respectively. Clearly, the larger the number of subcarriers U_k assigned to the kth mobile user, the higher bandwidth the kth mobile user can receive. The bandwidth R_k allocated for the kth real-time user can be expressed

$$R_k = \frac{U_k}{T_b} \tag{8}$$

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

Note that in our system, we do not consider adaptive modulation and coding (AMC) technique in the resource allocation. In addition, we focus on an uncoded system, where neither forward error correction (FEC) nor automatic repeat request (ARQ) is not considered.

DOI: 10.1002/wcm

2.2. 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*, that is, 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 uth subcarrier $h_{k,u}^{n,m}(t)$, between nth transmit antenna of the BS and the mth receive antenna of the kth user, can be denoted by $h_{k,u}^{n,m}[i]$, $i = 1, 2, \ldots$, where i is the discrete time index for the ith block-interval. The time-varying channel can be modeled by an auto-regressive (AR) process [7] as follows

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

where α_k is determined by the kth mobile user's Doppler velocity and $v_{k,u}^{n,m}[i]$ is a zero-mean independent identically distributed (i.i.d.) complex-Gaussian variable. In Sections 2 and 3 we focus on the discussion within a block-interval $[0, NT_b)$. Therefore, we drop the time index i for convenience. Assume that $\{h_{k,u}^{n,m} \mid \forall n, m, k, u\}$ are i.i.d.[§] complex-Gaussian variables with zero-mean and variance of $\sigma/2$ per dimension, that is, $\sigma = E\left(|h_{k,u}^{n,m}|^2\right)$.

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

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

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

2.3. Downlink Mobile Receiver Model

The schematic for the mth receive antenna at the kth real-time mobile user of our proposed ST MC-DS-CDMA system is shown in Figure 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 BS can also obtain the channel information by receiver's feedback.

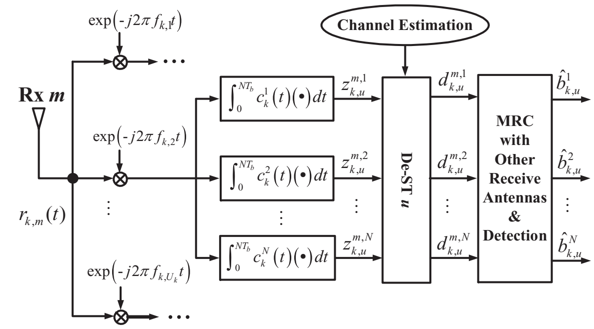

Fig. 2. Downlink block diagram of mth receive antenna at the kth mobile user.

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

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

which represents the procedure of maximum ratio combining (MRC). Based upon decision variables given in Equation (11), the receiver makes the decisions of the transmitted bits by

$$\hat{\mathbf{b}}_{k,u} = \mathrm{sgn}\left[\mathrm{Re}\left(\mathbf{d}_{k,u}\right)\right] \tag{12}$$

where $\mathrm{sgn}(\cdot)$ is the signum function and $\mathrm{Re}(\cdot)$ denotes the real part of $(\cdot)$. Denote the index-set of real-time mobile users allocated in the uth subchannel by Φ_u. We show in Reference [8] that the signal-to-noise-and-interference ratio (SINR) of decoding signals for the kth

[§] Theoretically, the coefficients between different subcarriers are not independent. However, this 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.

DOI: 10.1002/wcm

user at the uth subchannel can be expressed as follows:

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

where $\xi_{k,l}$ denotes the orthogonal factor between kth and the lth users, that is, co-channel interference, due to the possible software/hardware imperfectness.

3. QoS Requirements for Real-Time Service

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

3.1. 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 $\overline{\lambda}$ denote the average rate and $\lambda_{\max}$ the peak rate of the real-time traffic, respectively. In terms of Equation (8), if we expect to provide a *queuing-delay-free* transmission, the number U_k of subcarriers allocated to kth user need to satisfy

$$R_k = \frac{U_k}{T_\mathrm{b}} \geq \lambda_{\max} \Rightarrow U_k \geq U_\mathrm{goal} \triangleq \lceil \lambda_{\max} T_\mathrm{b} \rceil, \quad \forall k \in \Omega \tag{14}$$

where $\lceil x \rceil$ represents the ceiling function of 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 $U_\mathrm{goal} \leq U$ subcarriers to the kth user, the queue size may build up and queuing-delay occurs. On the other hand, if we assign $U_k > U_\mathrm{goal}$ subcarriers to the kth user, the bandwidth resources are wasted since the achievable bandwidth exceeding $\lambda_{\max}$ is not necessary for the 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 kth user may not be able to reach U_goal, that is, $U_k \leq U_\mathrm{goal}$, which is because our subchannel-allocation algorithm allocates subcarriers to users based on the *instantaneous* SINR. In order to guarantee a specific bounded-delay QoS requirement, the average number $\overline{U}_k$ of subcarriers allocated to each admitted user need to be larger than a subcarrier-threshold denoted by U_th, which is not necessarily an integer. In fact, U_th corresponds to the minimum average service bandwidth requirement $\overline{R}_\mathrm{min}$ to guarantee the specific delay-bound QoS, which is given by

$$\overline{R}_\mathrm{min} \triangleq \frac{U_\mathrm{th}}{T_\mathrm{b}} \tag{15}$$

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

$$U_\mathrm{th} < \overline{U}_k \leq U_\mathrm{goal}, \quad \forall k \in \Omega \tag{16}$$

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 $\overline{U}_k > U_\mathrm{th}, \forall k \in \Omega$. 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.

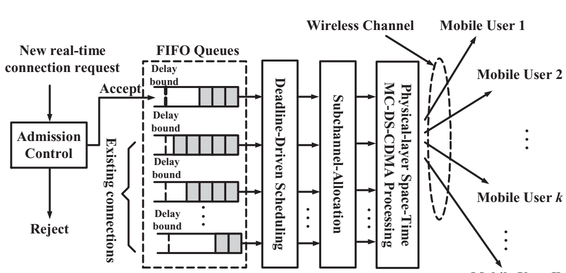

Fig. 3. The QoS provisioning architecture at the MAC layer in BS.

DOI: 10.1002/wcm

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.

3.2. 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 2.3, we assume that the BS knows the downlink channel information. Consequently, it can pre-compute the instantaneous SINR of decoding signals using Equation (13). In order to guarantee the loss-rate upper-bound QoS requirement, our proposed subchannel-allocation algorithm allocates subcarriers to users based upon the *instantaneous* SINR, which can be expressed as

$$\mathrm{SINR}_{k,u} \geq \gamma, \quad \forall k \in \Phi_u, \forall u \in \Lambda \tag{17}$$

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

4. Space-Time Diversity-Enhanced QoS Provisioning Algorithms

4.1. Admission Control

Let us reconsider Equation (13) in more details. As mentioned in Section 2.2, $\{h_{k,u}^{n,m} \mid \forall k, u, n, m\}$ are i.i.d. complex-Gaussian variables with zero-mean and variance of $\sigma/2$ per dimension. Therefore, if we define the random part of Equation (13) by the random variable $X_{k,u}$ as follows:

$$X_{k,u} \triangleq \sum_{n=1}^{N} \sum_{m=1}^{M} \left| h_{k,u}^{n,m} \right|^2 \tag{18}$$

where $X_{k,u}$ satisfies χ^2 distribution with the freedom of $2NM$. Due to the homogeneity of all the 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 Reference [9]

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

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

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

where $0 \leq \rho < 1$ is a constant. Then, Equation (13) can be written as

$$Z_u \triangleq \mathrm{SINR}_{k,u} = \frac{\left(\frac{PT_\mathrm{b}}{NU}\right) X_u}{\left(\frac{PT_\mathrm{b}}{NU}\right)(\kappa_u - 1)\rho X_u + N_0} \tag{21}$$

where we define Z_u instead of $\mathrm{SINR}_{k,u}$. Let the cardinality $\kappa_u \triangleq \|\Phi_u\|$ denote the number of mobile users within the uth subcarrier. Clearly, $(\kappa_u - 1)$ represents the number of co-subchannel interferences at the uth subcarrier.

Given κ_u users within the uth subcarrier, the probability $\phi(\kappa_u)$ that the kth user, $\forall k \in \Phi_u$, satisfies its SINR threshold can be expressed as

$$\phi(\kappa_u) \triangleq \Pr\{Z_{k,u} \geq \gamma \mid \|\Phi_u\| = \kappa_u\}$$
$$= \begin{cases} \Pr\left\{ X_u \geq \dfrac{\gamma N_0}{\left(\frac{PT_\mathrm{b}}{NU}\right)[1-(\kappa_u - 1)\gamma\rho]} \right\}, & \text{if } \kappa_u < \frac{1}{\gamma\rho} + 1 \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

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

$$\phi(\kappa_u) = \int_{\gamma N_0 / \left\{\left(\frac{PT_\mathrm{b}}{NU}\right)[1-(\kappa_u - 1)\gamma\rho]\right\}}^{+\infty} \frac{x^{L-1} e^{-\frac{x}{\sigma}}}{\sigma^L \Gamma(L)} dx \tag{23}$$

DOI: 10.1002/wcm

Our proposed subchannel-allocation algorithm assigns U_k ($U_k \leq U_{\text{goal}}$) subcarriers to the kth user. Therefore, given κ users within *each* subcarrier[‖], that is, $\kappa_u = \kappa, \forall u \in \Lambda$, the probability $\psi(\beta, \kappa)$ that the kth user, $\forall k \in \Phi_u$, is assigned β subcarriers follows the Binomial distribution as follows:

$$\psi(\beta,\kappa) \triangleq \Pr\{U_k = \beta \,\big|\, \|\Phi_u\| = \kappa,\ \forall u \in \Lambda\}$$
$$= \begin{cases} \binom{U}{\beta}[\phi(\kappa)]^{\beta}[1-\phi(\kappa)]^{U-\beta}, & \text{if } \beta < U_{\text{goal}} \\ \sum_{i=U_{\text{goal}}}^{U} \binom{U}{\beta}[\phi(\kappa)]^{i}[1-\phi(\kappa)]^{U-i}, & \text{if } \beta = U_{\text{goal}} \end{cases} \tag{24}$$

Thus, given κ users within each subcarrier, the average number $\overline{U}(\kappa)$ of subcarriers assigned to the kth user, $\forall k \in \Phi_u$, can be expressed as

$$\overline{U}(\kappa) = \sum_{\beta=1}^{U_{\text{goal}}} \beta\psi(\beta,\kappa) \tag{25}$$

The corresponding average bandwidth $\overline{R}(\kappa)$ allocated to the kth user, $\forall k \in \Phi_u$, can be calculated by

$$\overline{R}(\kappa) = \frac{\overline{U}(\kappa)}{T_{\text{b}}} \tag{26}$$

As discussed in Section 3, in order to ensure the delay-bound QoS requirement, we assume that the average number $\overline{U}(\kappa)$ of subcarriers assigned to each user is larger than a subcarrier-threshold U_{th}, that is, $\overline{U}(\kappa) > U_{\text{th}}$, or equivalently, $\overline{R}(\kappa) > \overline{R}_{\min}$. Since every admitted user should satisfy $\overline{U}(\kappa) > U_{\text{th}}$, the total number K of admitted users can be approximated by

$$K \approx \frac{\kappa U}{U_{\text{th}}} \Rightarrow \kappa \approx \frac{KU_{\text{th}}}{U} \tag{27}$$

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

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

4.2. 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 a *deadline-driven scheduling* scheme.

Note that in our CDMA system, data for all the K users are transmitted simultaneously. The scheduling is to adjust the sequence of executing the subchannel-allocation. As will be shown in Section 4.3, according to our subchannel-allocation algorithm, the user which is allocated earlier has advantage to occupy more bandwidth than the later allocated ones. This is due to the fact that SINR criterion described in Equation (17) is easier to be satisfied by the earlier allocated users. As a result, the later allocated users have higher probability of violating delay-bounds. If we employ a round-robin (RR) based scheduling scheme allocating all the 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) [10].

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 the 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 kth user's queue. For those users whose $D_k \geq 0$, the user which has the smallest D_k will be allocated first.

4.3. Subchannel-Allocation Algorithm

Let the jth user ($j \notin \Phi_u$) be the candidate which attempts to transmit bits using the uth subcarrier. Since the BS knows the information and statistical characteristics channel state information (CSI) about

[‖] 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.

DOI: 10.1002/wcm

the channel, we can pre-compute the SINR of decoding signals for each user at the uth subcarrier using Equation (13). The jth user can be assigned to the uth subcarrier if and only if

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

where

$$\mathrm{SINR}_{j,u} = \frac{\left(\frac{PT_\mathrm{b}}{NU}\right)\left(\sum_{n=1}^{N}\sum_{m=1}^{M}|h_{j,u}^{n,m}|^2\right)}{\left(\frac{PT_\mathrm{b}}{NU}\right)\left(\sum_{n=1}^{N}\sum_{m=1}^{M}|h_{j,u}^{n,m}|^2\right)\sum_{l\in\Phi_u}\xi_{j,l}^2 + N_0} \tag{29}$$

and

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

If Equation (28) is satisfied, the jth user is qualified to be assigned to the uth subcarrier. Otherwise, if any of the SINR in Equation (28) is lower than the threshold γ, the jth user cannot be assigned to the uth subcarrier.

Based on testing procedure given by Equation (28), we present our subchannel-allocation algorithm by the pseudocode in Figure 4. First, all the K users sequentially execute the SINR test (see Figure 4, step 05) by Equation (28). The sequence is determined by our deadline-driven scheduling algorithm. If Equation (28) is satisfied, the jth 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 the users are allocated within our system.

```
01. The allocation sequence is based on the deadline-driven scheduling
02.   for (j := 1 to K) {
03.      for (u := 1 to U) {
04.          if U_j < U_goal;
05.              SINR testing by 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. The pseudocode of the subchannel-allocation algorithm.

5. Numerical and Simulation Results

We investigate the system performance of our proposed system infrastructure and QoS provisioning algorithms by numerical analyses and simulations. The system chip duration T_c is set to $T_\mathrm{c} = 7.8125\,\mu$s and spreading gain G set to 16. Thus, the bit duration T_b is $T_\mathrm{b} = T_\mathrm{c}G = 125\,\mu$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 $U/T_\mathrm{c} = 1.28$ 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 the users. Note that the smaller the orthogonal factor, the larger the number of admitted mobile users, and the higher the throughput. However, the specific value of the orthogonal factor does not affect the performance trends of our proposed subchannel allocation. 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 M of receive antenna at each mobile user is set to $M = 1$ for all the performance analyses. Also, we omit the guard time in our performance analyses for simplicity.

In simulations, we employ a simple traffic model used in Reference [11] 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 four subcarriers. In the second state, at the rate of 48 Kbps (six subcarriers) and the third state at the rate of 64 Kbps (eight subcarriers). The real-time traffic may stay in one of the three states with probability 25, 50, and 25%, respectively. We set $U_\mathrm{goal} = 8$ and $\overline{R}_\mathrm{min} = 50$ Kbps to guarantee the upper-bounded delay. The corresponding U_th can be calculated by $U_\mathrm{th} = \overline{R}_\mathrm{min}T_\mathrm{b} = 6.25$.

Using admission control strategy described in Section 4.1, Figure 5(a–c) plot the numerical solutions of average bandwidth $\overline{R}(KU_\mathrm{th}/U)$ (see Equations (26) and (27)) versus two independent parameters the number K of users and the SNR per bit E_b/N_0 with different transmit antennas, where $E_\mathrm{b} = PT_\mathrm{b}\sigma/(NU)$. The contour-lines of the three figures are drawn at

DOI: 10.1002/wcm

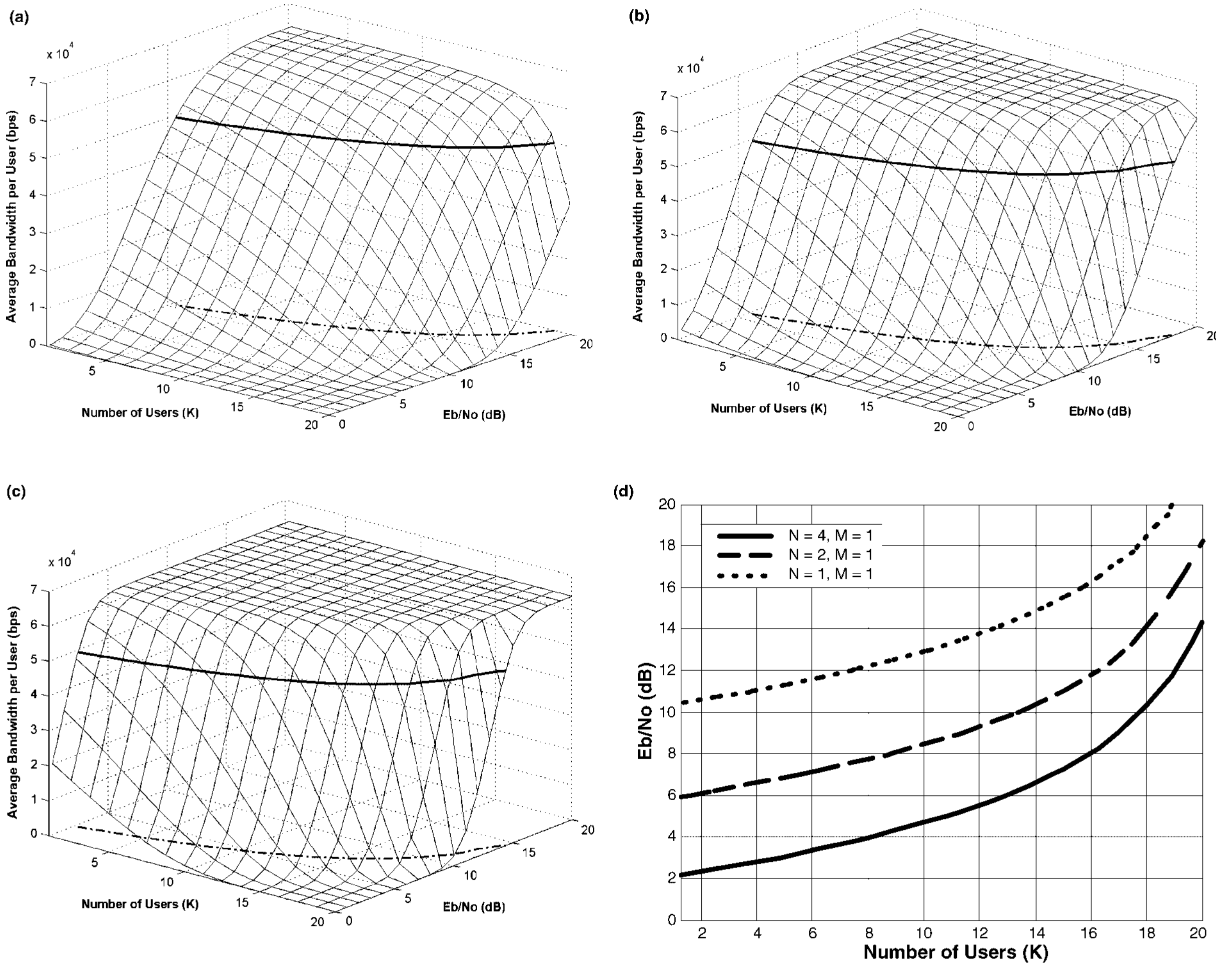

Fig. 5. Average bandwidth and the admission region versus the user number K and E_b/N_0. (a) N = 1, M = 1; (b) N = 2, M = 1; (c) N = 4, M = 1; (d) The number of users (K) which can be admitted.

the service rate equal to $\overline{R}_{\min} = 50$ 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 Figure 5(a–c) that the average bandwidth increases when SNR E_b/N_0 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).

Figure 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 number of antenna combinations. The areas above the projection-lines represent the admission regions. We can see that the admission region significantly increases as the number N increases. Thus, Figure 5 shows that the STD can significantly enhance the QoS provisioning for real-time services.

Figure 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 12 dB in simulations. We can see from Figure 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,

DOI: 10.1002/wcm

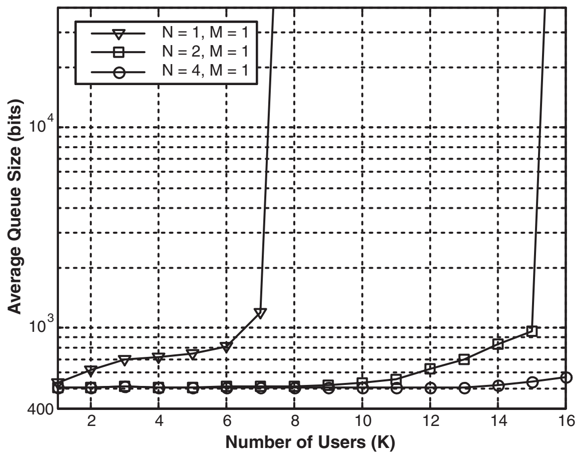

Fig. 6. Average queue size versus the number of K of users with $M = 1$.

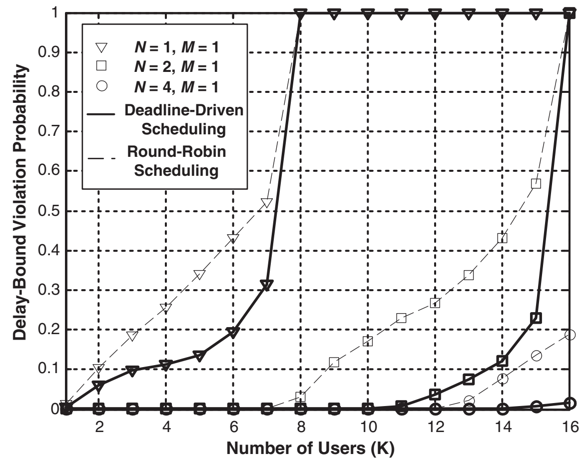

Fig. 7. The probability of delay-bound violation versus the number of K of users with $M = 1$.

the admitted number of users needs to be smaller than K_0. For example, when the number of transmit antennas $N = 1$, the number of admitted users should be smaller than $K_0 = 8$. When $N = 2$, the number of admitted users should be smaller than $K_0 = 16$. Agreeing with the conclusions drawn from Figure 5, Figure 6 shows that the STD can provide significant improvement for real-time QoS provisioning.

Furthermore, the simulated admission region shown in Figure 6 agrees well with the numerical results drawn in Figure 5(d). In Figure 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 analytical analyses.

Finally, Figure 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 12 dB and the delay-bound $Q_{\max} =$ 1200 bits in simulations. We can see from Figure 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 Figure 6. Also, we observe from Figure 7 that the deadline-driven scheduling scheme significantly outperforms RR scheduling algorithm. 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 instead of RR scheduling algorithm when $N = 1$ and $K = 6$. Moreover, the delay-bound violation probability is reduced by more than 60% (from 0.43 to 0.13) by applying deadline-driven scheduling when $N = 2$ and $K = 14$. Figure 7 also shows the advantages of using ST infrastructure at physical layer since Figure 7 shows that the more the number of transmit antennas, the lower the delay-bound violation probability the system can achieve.

6. Conclusions

We proposed the QoS provisioning algorithms including admission control, scheduling, and subchannel-allocation for real-time service over ST MC-DS-CDMA-based wireless networks. We also derived the analytical analysis to find the admission region of our ST 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. The ST infrastructure can significantly improve the performance of the QoS provisioning in the wireless networks.

Acknowledgements

The research performed by Xi Zhang and Jia Tang in this paper was supported in part by the U.S. National Science Foundation CAREER Award under Grant ECS-0348694 and the research performed by

DOI: 10.1002/wcm

Hsiao-Hwa Chen in this paper was supported in part by Taiwan NSC grant number NSC 95- 2213-E-110-008.

References

1. Rappaport T, Annamalai A, Buehrer R, Tranter W. Wireless communications: past events and a future perspective. *IEEE Communicatoins Magazine* 2002; **40**(5): 148–161.
2. Gesbert D, Shafi M, Shiu D, Smith P, Naguib A. From theory to practice: an overview of MIMO space-time coded wireless systems. *IEEE Journal on Selected Areas in Communications* 2003; **21**(3): 281–302.
3. Hochwald B, Marzetta T, Papadias C. A transmitter diversity scheme for wideband CDMA systems based on space-time spreading. *IEEE Journal on Selected Areas in Communications* 2001; **19**: 48–60.
4. Tarokh V, Jafarkhani H, Calderbank A. Space-time block codes from orthogonal designs. *IEEE Transactions on Information Theory* 1999; **45**(5): 1456–1467.
5. Yang L-L, Hanzo L. Space-time spreading assisted broadband MC-CDMA. *IEEE 55th Vehicular Technology Conference* 2002; **4**: 1881–1885.
6. Yang L-L, Hanzo L. Performance of generalized multicarrier DS-CDMA over Nakagami-m fading channels. *IEEE Transactions on Communications* 2002; **50**(6): 956–966.
7. Kalofonos D, Stojanovic M, Proakis J. Performance of adaptive MC-CDMA detectors in rapidly fading Rayleigh channels. *IEEE Transactions on Wireless Communications* 2003; **2**(2): 229–239.
8. Zhang X, Tang J, Chen H-H. Subchannel-allocation algorithms and performance analysis for space-time OFDM-CDMA based systems in wireless networks. *Technical Report*, Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, USA.
9. Simon MK, Alouini MS. *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. Wiley: New York, 2000.
10. Choi S, Shin KG. A unified wireless LAN architecture for real-time and non-real-time communication services. *IEEE Transactions on Networking* 2000; **8**(1): 44–59.
11. Choi S, Shin KG. An uplink CDMA system architecture with diverse QoS guarantees for heterogeneous traffic. *IEEE Transactions on Networking* 1999; **7**(5): 616–628.

Authors' Biographies

Xi Zhang (S'89-SM'98) received his B.S. and M.S. degrees from Xidian University, Xi'an, China, his M.S. degree from Lehigh University, Bethlehem, Pennsylvania, U.S.A., all in electrical engineering and computer science, and the Ph.D. degree in electrical engineering and computer science (Electrical Engineering—Systems) from The University of Michigan, Ann Arbor, Michigan, U.S.A.

He is currently an Assistant Professor and the Founding Director of the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, U.S.A. He was an Assistant Professor and the Founding Director of the Division of Computer Systems Engineering, Department of Electrical Engineering and Computer Science, Beijing Information Technology Engineering Institute, Beijing, China, from 1984 to 1989. He was a Research Fellow with the School of Electrical Engineering, University of Technology, Sydney, Australia, and the Department of Electrical and Computer Engineering, James Cook University, Queensland, Australia, under a Fellowship from the Chinese National Commission of Education. He worked as a Summer Intern with the Networks and Distributed Systems Research Department, AT&T Bell Laboratories, Murray Hills, NJ, U.S.A. and with AT&T Laboratories Research, Florham Park, NJ, U.S.A. in 1997. He has published more than 100 research papers in the areas of wireless networks and communications, mobile computing, cross-layer optimizations for QoS guarantees over mobile wireless networks, effective capacity and effective bandwidth theories for wireless networks, DS-CDMA, MIMO-OFDM and space-time coding, adaptive modulations and coding (AMC), wireless diversity techniques and resource allocations, wireless sensor and Ad Hoc networks, cognitive radio and cooperative communications/relay networks, vehicular Ad Hoc networks, multi-channel MAC protocols, wireless and wired network security, wireless and wired multicast networks, channel coding for mobile wireless multimedia multicast, network protocols design and modeling, statistical communications theory, information theory, random signal processing, and control theory and systems.

Professor Zhang received the U.S. National Science Foundation CAREER Award in 2004 for his research in the areas of mobile wireless and multicast networking and systems. He also received the TEES Select Young Faculty Award for Excellence in Research Performance from the Dwight Look College of Engineering at Texas A&M University, College Station, Texas, U.S.A. in 2006. He is currently serving as an Editor for the *IEEE Transactions on Wireless Communications*, an Associate Editor for the *IEEE Transactions on Vehicular Technology*, an Associate Editor for the *IEEE Communications Letters*, and an Editor for the *Wiley's Journal on Wireless Communications and Mobile Computing*, and is also serving as the Guest Editor for the *IEEE Wireless Communications Magazine* for the Special Issues on "Next Generation of CDMA versus OFDMA for 4G Wireless Applications". He has frequently served as the Panelist on the U.S. National Science Foundation (NSF) Research-Proposal Review Panels. He also served as the WiFi-Hotspots/WLAN and QoS Panelist at the IEEE QShine 2004. He is serving or has served as the Co-Chair for the IEEE Global Telecommunications (Globecom 2008) – Wireless Communications Symposium 2008 and the Co-Chair for the IEEE International Conference on Communications (ICC 2008) – Information and Network Security Symposium 2008, respectively, the Symposium Chair for the IEEE International Cross-Layer Optimized Wireless Networks Symposium 2006 and 2007, respectively, the TPC Chair for the IEEE International Wireless Communications and Mobile Computing Conference (IWCMC) 2006 and 2007, respectively, the Poster Chair for the IEEE INFOCOM 2008, the Student Travel Grants Committee Co-Chair for the IEEE INFOCOM 2007, the Panel Co-Chair for the IEEE 16th International Conference on Computer Communications and Networks (ICCCN) 2007, the Poster Chair for the IEEE QShine 2006 and the IEEE/ACM 10th International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM) 2007, and the

DOI: 10.1002/wcm

Publicity Chair for the IEEE WirelessCom 2005 and QShine 2007. He has served as the TPC members for more than 40 IEEE/ACM conferences, including the IEEE INFOCOM, IEEE Globecom, IEEE ICC, IEEE WCNC, IEEE VTC, IEEE/ACM QShine, IEEE WoWMoM, IEEE ICCCN, etc.

Professor Zhang is a Senior Member of the IEEE and a Member of the Association for Computing Machinery (ACM).

Jia Tang received his B.S. degree in Electrical Engineering from Xi'an Jiaotong University, Xi'an, China, and Ph.D. in Computer Engineering from Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, U.S.A., in 2001 and 2006, respectively.

His research interests include mobile wireless communications and networks, with emphasis on cross-layer design and optimizations, wireless quality-of-service (QoS) provisioning for mobile multimedia networks and wireless resource allocation.

Dr Tang received Fouraker Graduate Research Fellowship Award from Department of Electrical and Computer Engineering, Texas A&M University in 2005.

Hwiao-Hwa Chen (hshwchen@ieee.org) was the founding Director of Institute of Communications Engineering, National Sun Yat-Sen University, Taiwan. He has been a visiting Professor to Department of Electrical Engineering, University of Kaiserslautern, Germany, in 1999, the Institute of Applied Physics, Tsukuba University, Japan, in 2000, Institute of Experimental Mathematics, University of Essen, Germany in 2002 (under DFG Fellowship), Department of Information Engineering, The Chinese University of Hong Kong, 2003, and Department of Electronics Engineering, The City University of Hong Kong, 2006. His current research interests include wireless networking, MIMO systems, and next generation CDMA technologies for future wireless communications. He has authored and co-authored over 200 technical papers in major international journals and conferences, and five books in the areas of communications, including '*Next Generation Wireless Systems and Networks*' (512 pages) published by John Wiley & Sons. He served or is serving as International Steering Committee member, Symposium Chair and General Co-Chair of more than 50 international conferences, including IEEE VTC, IEEE ICC, IEEE Globecom, and, IEEE WCNC, etc. He served or is serving as the Editor, Associate Editor, Editorial Board member, or Guest Editor of many international journals, including *IEEE Communications Magazine*, *IEEE JSAC*, *IEEE Vehicular Technology Magazine*, *IEEE Wireless Communications Magazine*, *IEEE Network Magazine*, *IEEE Transactions on Wireless Communications*, *IEEE Communications Letters*, *Wiley's Wireless Communications and Mobile Computing (WCMC) Journal*, *Wiley's International Journal of Communication Systems*, etc. He is an adjunct Professor of Zhejiang University, China, and Shanghai Jiao Tung University, China. Currently, Professor Chen is serving as the Chair of Radio Communications Committee, IEEE Communications Society.

DOI: 10.1002/wcm