Analysis of TCP Performance Under Joint Rate and Power Adaptation in Cellular WCDMA Networks

Ekram Hossain, *Member, IEEE*, Dong In Kim, *Senior Member, IEEE*, and Vijay K. Bhargava, *Fellow, IEEE*

***Abstract*—To improve the spectral efficiency while meeting the radio link level quality of service requirements such as the bit-error-rate (BER) requirements for the different wireless services, transmission rate and power corresponding to the different mobile users can be dynamically varied in a cellular wideband code-division multiple-access (WCDMA) network depending on the variations in channel interference and fading conditions. This paper models and analyzes the performance of transmission control protocol (TCP) under joint rate and power adaptation with constrained BER requirements for downlink data transmission in a cellular variable spreading factor (VSF) WCDMA network. The aim of this multilayer modeling of the WCDMA radio interface is to better understand the interlayer protocol interactions and identify suitable transport and radio link layer mechanisms to improve TCP performance in a wide-area cellular WCDMA network.**

***Index Terms*—Dynamic rate and power adaptation, link layer and transport layer protocol interaction, multicell multirate wideband code-division multiple-access (WCDMA) systems.**

I. Introduction

THE Third Generation Partnership Project (3GPP) wideband code-division multiple-access (WCDMA) system, which is being developed as one of the IMT-2000 standards, is expected to provide high-speed packet data services including wireless Internet access. Since the transport layer protocol performance is one of the most critical issues in data networking over noisy wireless links, the performance of transmission control protocol (TCP), which is the flagship protocol in today's Internet, would be crucial in such an environment. TCP is a connection-oriented transport layer protocol which guarantees reliable in-sequence delivery of packets and is generally more suited for delay-insensitive applications.

The performance of TCP in a wireless network depends on the service provided by the underlying radio link control (RLC)/medium access control (MAC) protocol. In a WCDMA system, the transmission rate and the power corresponding to the different mobile users can be dynamically varied depending on the variations in channel interference and fading conditions to improve the wireless channel utilization while meeting the lower layer (e.g., RLC/MAC layer, physical (PHY) layer) quality of service [e.g., bit-error rate (BER)] requirements. For example, if a mobile experiences larger interference, the rate assignment for this mobile is to be constrained to lower rates compared to that for the mobile experiencing better channel conditions. This is similar to the dynamic link adaptation concept proposed for narrowband systems [1].

The impact of dynamic radio link adaptation on TCP performance for forward link data transmission in a cellular WCDMA system is modeled and analyzed in this paper. The dynamic link adaptation is assumed to be achieved by joint rate and power adaptation under constrained BER requirements where the lower layer BER requirements for all the active TCP connections are assumed to be the same. Note that, even with a single BER constraint, radio link level automatic repeat-request (ARQ) techniques can be combined with channel coding to achieve multiple frame-error-rate (FER) constraints corresponding to the different TCP connections. Two joint rate and power allocation algorithms, namely, depth-first allocation (DFA) and breadth-first allocation (BFA), are considered for dynamically adjusting the transmission rate and power corresponding to the different TCP connections. The DFA and the BFA algorithms use exhaustive and round-robin principles, respectively, for dynamic resource allocation. The data transmission rate in the forward link physical layer of a WCDMA system can be controlled by using the variable spreading factor (VSF) method or the multicode method [2]. The VSF method of dynamic rate adaptation is considered in this paper.

The organization of the rest of the paper is as follows. The motivation and background of this work are presented in Section II. Section III describes the multilayer (TCP and the RLC protocol) system model analyzed in this paper. Starting with the signal-to-interference ratio (SIR) modeling for downlink data transmission in a multicell VSF WCDMA system, Section IV presents the dynamic link adaptation algorithms for both the uniform and the nonuniform traffic scenarios. In Section V, the simulation model is described. The simulation results are presented in Section VI. In Section VII, conclusions are stated.

II. Background, Motivation and, Methodology

Dynamic link adaptation for optimizing the radio link level performance and the issue of improving TCP performance in wireless networks have been addressed in literature as separate problems. Transmission protocol stack performance in a cellular

Manuscript received March 6, 2002; revised October 17, 2002 and February 16, 2003; accepted February 25, 2003. The editor coordinating the review of this paper and approving it for publication is J. Evans. This work was supported in part by a Research Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada and in part by the PRG Research Grant from Simon Fraser University.

E. Hossain is with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada (e-mail: ekram@ee.umanitoba.ca).

D. I. Kim is with the School of Engineering Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: dikim@sfu.ca).

V. K. Bhargava is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vijayb@ece.ubc.ca).

Digital Object Identifier 10.1109/TWC.2004.826308

wireless network would be optimized if the radio link/PHY level and the transport level protocol performances are jointly optimized, and a multilayer modeling is necessary to explore the issues of interlayer protocol interaction (e.g., implications of link adaptation on higher layer protocol performance) and identify suitable radio link and transport protocol mechanisms.

The problem of rate and power adaptation for uplink packet data transmission in a single-cell multirate CDMA system was addressed in [3] and the optimal centralized adaptive rate and power control strategy to maximize the total average weighted throughput was determined. With the objective of maximizing the throughput and with the constraints on maximum power and/or minimum rate, the problem of jointly controlling the user power and data rates for uplink data transmission in a single-cell CDMA system was formulated as a constrained optimization problem in [4]–[6]. The problem of joint rate and power adaptation under constrained BER for downlink data transmission in a multicell and multirate CDMA system was addressed in [7].

Some of the earlier works on wireless TCP (e.g., [8], [9]) investigated the performance of a single circuit-based TCP connection in a constant spreading factor DS-CDMA environment. In [10], the authors presented an approximate throughput analysis for a single connection TCP with highly persistent link layer in a time-division multiple-access (TDMA) system. A comparative performance evaluation of several TCP versions in a TDMA environment was carried out in [11]. The performances of several end-to-end, split-connection, and link layer schemes designed to improve TCP performance were evaluated in [12]. In [13], the throughput performance of TCP was analyzed over 3G wireless links in the presence of varying bandwidth and delay. Throughput and energy performance of circuit-switched TCP connections in a wideband CDMA air interface in the absence of link level retransmissions were investigated in [14].

The system dynamics representing the interactions between the higher layer and the lower layer protocols in an interference-limited cellular WCDMA environment is fairly complex because of the existence of numerous parameters and nonlinear nature of the protocol state machines at the different layers. For a particular system load, radio link level rate adaptation and scheduling, persistence in the retransmission of the failed radio frames, the desired SIR and the physical level channel coding mechanisms impact the radio link level and, hence, the higher layer protocol performance. Again, under a particular radio link level configuration, transport layer mechanisms such as the congestion and flow control and adaptive retransmission timeout (TO) determine the achieved transport layer protocol throughput.

This paper attempts to meaningfully evaluate some aspects of TCP performance for downlink (i.e., forward link) data transmission in such an environment. In particular, for a cellular VSF WCDMA system we do the following:

1) present a multilayer model to investigate system dynamics and interlayer protocol interaction for downlink transmission under multiple concurrent TCP connections;
2) analyze the impact of the different radio link level rate adaptation mechanisms on TCP throughput behavior and identify suitable mechanisms and protocol parameters;
3) analyze how the "conservativeness–aggressiveness" in TCPs congestion control behavior affect the TCP throughput performance;
4) analyze the impact of channel fading and shadowing on TCP throughput performance.

An event-driven simulator is developed to imitate a multilayer model of the WCDMA radio interface. It supports the TCP congestion and error control algorithms including slow-start, congestion avoidance, fast retransmit, and fast recovery. The RLC protocol in this multilayer model consists of a radio link quality-based dynamic rate adaptation mechanism and a semireliable hybrid-ARQ type of error control method. The dynamic rate adaptation is assumed to be achieved by either DFA or BFA method to be described later in this paper.

The radio link quality of the mobile users in a WCDMA network depends on the spatial distribution (and, hence, the micromobility patterns) of the users along with the short-term and the long-term channel fading conditions. To investigate TCP performance using the multilayer modeling approach, both the random and the directional micromobility models are considered. The nature of the long-term fading (i.e., shadowing) is assumed to be uncorrelated and correlated, respectively, for the random and directional mobility cases. Since the short-term fading changes more rapidly compared to the long-term fading, it is assumed to be independent over successive link adaptation intervals. Multipath channel fading with equal path gain and unequal path gain, are assumed for the random and the directional micromobility model, respectively.

III. Multilayer System Model

A list of the key mathematical notations used in the system model is provided in Table I.

A. TCP Model

A sender-based one way traffic scenario is considered where the mobile stations (MSs) act as TCP sinks. To study the steady-state behavior of TCP, bulk data TCP connections are assumed (i.e., senders always have data to transmit and can transmit as many packets as their transmission windows allow). The TCP sinks can accept packets out of sequence but deliver them only in sequence to the user and they generate immediate acknowledgment (ACKs).

Similar to that in all currently available TCP implementations, the basic window adaptation procedure consists of the slow-start and the congestion avoidance phases where the evolution of the sender's congestion window $W(t)$ and the slow-start threshold $W_{\rm th}(t)$ at time t are triggered by TCP ACKs and TOs [11]. Fast retransmit and fast recovery procedures are implemented subsequent to a packet loss (as in TCP Reno[1]). The RTO (Retransmission Time Out) value at the sending TCP host is updated based on the recommendations in [15]. After a TO, the sender also updates the RTO value using exponential backoff. The backoff continues until an ACK is received for a packet transmitted exactly once.

For the TCP ACKs, an ideal scenario is assumed where the fixed sending hosts successfully receive the ACK packets gener-

[1]In today's Internet, TCP Reno is the de facto standard for TCP implementation.

TABLE I
LIST OF KEY NOTATIONS

g_j	No. of mobiles in cell j ($j = 0, \cdots, J$)
G	No. of concurrent TCP connections in a cell (for uniform traffic load)
$\bar{G}$	Average no. of concurrent TCP connections per cell
$\mu_i^{(j)}$	Effective interference factor for ith mobile in cell j
$m_i^{(j)}$	Transmission rate allocated to ith mobile in cell j ($0 \le m_i^{(j)} \le \varphi$)
p_b	Prob. of bit error
p_c	Prob. of correct transmission of a radio frame
$P_{b,i}^{(j)}$	Power allocated to ith mobile in cell j corr. to the basic rate v_1
γ_j	Rate capacity resource in cell j
$\eta_{j'/j}(i)$	Intercell interference factor corr. to transmission to ith mobile in cell j
P_c	Pilot signal transmission power
$P_{B,j}$	Total power budget available at the jth BS
β_j	Ratio of $P_{B,j}$ and P_c
$\beta_{j'/j}$	Ratio of $P_{B,j'}$ and $P_{B,j}$
$(SIR)_o$	Target SIR
$(SIR)_{o,d}^{(j)}$	Achieved SIR at the mobiles in cell j
W_{max}	Maximum TCP window size
$D_{internet}$	Internet delay
$\lambda_i^{(j)}$	Average throughput for ith TCP connection in cell j
F	TCP throughput fairness

ated by the MSs corresponding to each successfully transmitted TCP packet (i.e., no ACK packet is lost in the air interface and the wired part of the network), and that the ACK packets undergo no extra queuing delay in addition to the propagation delay.

As a result of TCPs congestion control algorithm, a connection's throughput varies as the inverse of the connection's round trip time (RTT) [16]. In a wide-area wireless network, the inherent TCP bias against flows with long RTT may be mitigated through improved traffic management techniques (e.g., by suitable combination of packet marking and dropping policies as in the differentiated services Internet Protocol (IP) networks). Since the focus of this paper is on the radio link level dynamic resource allocation and its impact on the TCP performance and system dynamics under multiple concurrent TCP connections, packets corresponding to all the TCP connections are assumed to experience the same internet[2] delay (D_{internet}).

B. Radio Link Layer Model

A transport protocol data unit is typically forwarded to the radio link level protocol (with the necessary intermediate protocol overheads appended to it), where it is segmented into several radio frames for transmission in the air interface. For example, in 3G1X/UMTS system TCP packets from packet data service node/serving GPRS support node are tunneled to radio network controller (RNC) which fragments these packets into radio frames and performs local transmissions–retransmissions through the base stations (BSs) using RLP/RLC protocol. The RLC protocol for downlink transmission to the mobiles has two components—a mechanism for the dynamic selection of the number of radio frames to be transmitted during a frame-time (i.e., rate adaptation) and a mechanism for error control.

1) Dynamic Rate Adaptation: With variable rate frame transmission, the number of frames transmitted during a frame-time varies according to the transmission rate used during that frame-time. Let us assume that the transmission rates can be selected from the set of rates $\{r_0, r_1, r_2, \ldots, r_\varphi\}$ and $r_m = mr_1$ ($m = 0, 1, \ldots, \varphi$) so that the normalized value of r_m with respect to the *basic* rate r_1 is m. Therefore, if the frame length corresponding to the basic rate r_1 is M bits, the frame length for rate r_m is mM bits. For this, a VSF WCDMA system can be used where the basic gain is given by N chips per bit, and for rate r_m, the spreading gain is reduced to N/m chips per bit. Note that the noncontiguous rates $r_m = 2^{m-1}r_1$ are typically used for the VSF method ([2]), in which case, the rate granularity becomes very coarse at higher rates which may cause reduced network throughput compared to the case of the contiguous rate as considered herein.

The radio link level transmission rate corresponding to a TCP connection is determined by a dynamic rate adaptation algorithm and is based on the system load (in terms of the number of concurrent TCP connections) and the channel interference and fading conditions corresponding to the mobile TCP sinks. The specific dynamic rate adaptation methods used in this paper will be described later.

2) Radio Link Level Error Control: We consider an ACK/negative acknowledgment-based selective repeat (SR)-hybrid-ARQ protocol enhanced for radio link level variable rate transmission. To illustrate, the flow diagram of the SR-ARQ protocol under variable rate transmission is depicted in Fig. 1 for transmission to a single mobile user, assuming that the radio link level acknowledgment delay is one frame-time and that the data rate r_m can be selected from the set $\{r_1, r_2, r_4\}$. Note that, in this error control method, forward error correction (FEC) coding is applied to each of the radio frames transmitted during a frame-time so that each of them can be decoded individually at the mobile receiver. Only the frames transmitted in error are retransmitted. Under a heavy traffic condition, when the newly generated frames are always available, the variable rate transmission can be fully exploited from frame-time to frame-time.

If power adaptation during each frame-time compensates the path loss and the attenuation due to shadowing (which is assumed to be fixed during a frame-time), bit errors will be random due to independent multipath fading over a bit time. Therefore, with independent bit errors over a frame-time and bit error probability of p_b, for a n-bit radio frame with t-bit error correction capability, the probability of correct reception of a frame p_c can be expressed as

$$p_c = \sum_{e=0}^{t} \binom{n}{e} p_b^e (1 - p_b)^{n-e}. \tag{1}$$

[2]The term *internet* is used in a generic sense here.

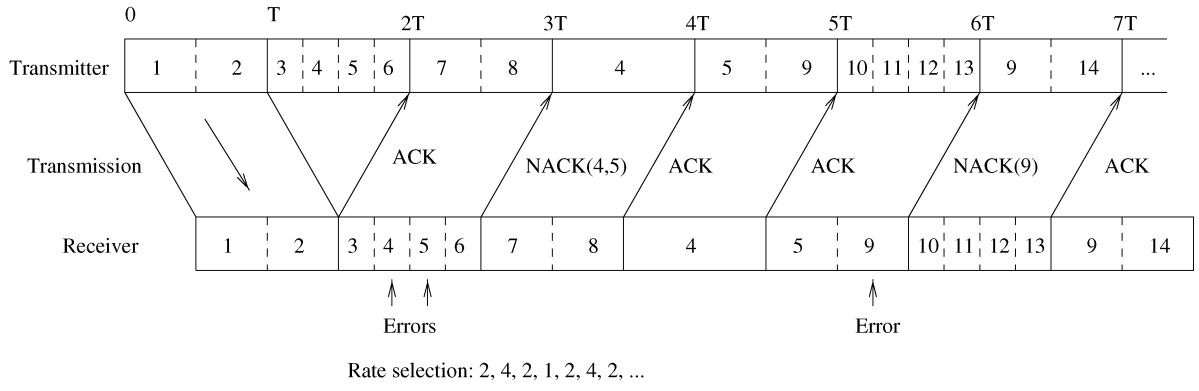


Fig. 1. SR-ARQ-based error control under variable rate transmission at the radio link level.

Note that the radio frame size n depends on t, the channel coding scheme, and the number of information bits per frame (M) corresponding to the basic transmission rate. For example, with $M = 120$, $t = 1$ and Bose, Chaudhuri, and Hocquenghem (BCH) coding, $n = 127$ [17] (i.e., (127, 120, 1) BCH code is used).

Even if all the frames transmitted during a frame-time are in error, the radio link level retransmission continues and the persistence in retransmission of the failed radio frames is controlled by the parameter $R_{\max}$. When $R_{\max}$ transmission failures occur, where all the radio frames transmitted during a frame-time are in error, all the frames for the corresponding upper layer protocol unit (e.g., TCP segment[3]) are discarded from the radio link level buffer. Therefore, for the radio frames corresponding to a TCP segment to be purged out from the link level buffers, it requires $R_{\max}$ transmission failures to occur while those frames are being transmitted. Since the local error recovery fails, an end-to-end error recovery will be initiated by the sending host after TO occurs.

Note that, the radio link level persistence here refers to the persistence in retransmitting radio frames corresponding to a TCP segment. Even though the radio frames corresponding to a TCP segment is discarded, radio frames corresponding to a different TCP segment for the same TCP connection may be transmitted during the successive frame-times. For the different TCP connections, the persistence in accessing the channel is determined by the dynamic rate adaptation algorithm only.

IV. Dynamic Link Adaptation

A. Modeling of Downlink SIR

We assume that the number of MSs in cell j is g_j $(j = 0, 1, \ldots, J)$, which are uniformly distributed in the cell and that the number of TCP connection per MS is one. To model the SIR at the mobiles under variable rate transmission using VSF, a sufficiently general model can be developed taking the channel fading and interference conditions into consideration under the assumption that the achieved SIRs are the same for all the mobiles in cell j (tagged cell).

The radio propagation is modeled by path loss, shadowing, and multipath fading. The attenuation of the signal from the BS j' (denoted by $(\mathrm{BS})_{j'}$) to the ith MS in cell j at a distance $r_{i,j'}$ is given by $L_{j'}(j,i) = r_{i,j'}^{-\delta} \cdot 10^{\xi_{i,j'}/10}$, where $\xi_{i,j'}$ is a Gaussian random variable with mean 0 and variance σ^2. The signal is also affected by an L-path Rayleigh fading channel with path gain $a_{i,j',l}$ $(l = 1, 2, \ldots, L)$, where $\sum_{l=1}^{L} a_{i,j',l}^2 = \zeta_i^{(j')}$ ($\zeta_i^{(j')}$ is a chi-square random variable with $2L$ degrees of freedom [18]). We define the intercell interference factor $(\eta_{j'/j})$ corresponding to the ith mobile in cell j as follows:

$$\eta_{j'/j}(i) \triangleq \frac{\zeta_i^{(j')} L_{j'}(j,i)}{\zeta_i^{(j)} L_j(j,i)}. \tag{2}$$

The effective interference at the MSs depends on the transmission powers used at the BS transmitters in the different cells. Suppose that $P_{b,i}^{(j)}$ is the transmission power corresponding to the ith MS in cell j when the basic rate r_1 is used for downlink transmission to that MS. This is referred to as the *basic* power allocation in this paper. Then, the *effective* power allocation for rate r_m would be $mP_{b,i}$. Therefore, the total transmission power $P_{B,j}$ at the jth BS is

$$P_{B,j} = \sum_{i=1}^{g_j} m_i^{(j)} P_{b,i}^{(j)} + P_c \tag{3}$$

where P_c is the pilot-signal transmission power.

Under the assumption that the same downlink quality is achieved for all MSs in cell j (tagged cell), the effective interference experienced by the ith MS in cell j can be modeled in terms of the in-cell orthogonality factor ν ($\nu = 1$ corresponds to perfectly orthogonal in-cell mobiles), the intercell interference factor $\eta_{j'/j}(i)$ and the total transmission power used at the different BS transmitters as follows [7]:

$$\mu_i^{(j)} = \left[(1-\nu) + \sum_{j' \neq j} \beta_{j'/j} \cdot \eta_{j'/j}(i) \right] \tag{4}$$

where $\mu_i^{(j)}$ is the effective interference factor corresponding to the ith MS in cell j and $\beta_{j'/j} = P_{B,j'}/P_{B,j}$. The achieved downlink SIR $(\mathrm{SIR})_{o,d}^{(j)}$ corresponding to all MSs in cell j can then be formulated as [7]

$$(\mathrm{SIR})_{o,d}^{(j)} = \frac{N}{\sum_{i=1}^{g_j} m_i^{(j)} \mu_i^{(j)}} \left(1 - \frac{P_c}{P_{B,j}} \right) \tag{5}$$

[3]TCP segment here refers to the TCP packet with the other necessary protocol overheads appended to it.

where we have used $(\mathrm{SIR})_{o,d}^{(j)} = \rho_j N/P_{B,j}$ for which $\rho_j = P_{b,i}^{(j)}/\mu_i^{(j)}$ $(i = 1, \ldots, g_j)$ is kept the same for all MSs in cell j, and N is the maximum spreading gain.

B. Joint Rate and Power Allocation

1) Uniform Traffic: For a given traffic load of G simultaneous TCP connections per cell with one connection per MS (i.e., $g_j = G$ and $P_{B,j} = P_B \ \forall\ j$), the transmission rates $m_i^{(j)}$ $(i = 1, 2, \ldots, G)$ can be dynamically selected based on the estimated $\{\mu_i^{(j)}\}_{i=1}^{G}$ (under the assumption $P_{B,j} = P_B \ \forall\ j$) in such a way that the weighted *sum interference* (i.e., $\sum_{i=1}^{G} m_i^{(j)}\mu_i^{(j)}$) does not exceed γ, the *a priori* bound in (6) below, where

$$\sum_{i=1}^{G} m_i^{(j)}\mu_i^{(j)} \le \frac{N}{(\mathrm{SIR})_o}\left(1 - \frac{P_c}{P_B}\right) \triangleq \gamma \tag{6}$$

where $(\mathrm{SIR})_o$ is the desired SIR and $(\mathrm{SIR})_o \le (\mathrm{SIR})_{o,d}^{(j)}$. We refer to γ as the *rate capacity resource*. Although for estimating the effective interference factors the total power allocations at the BS transmitters are assumed to be based on the distribution of the number of mobiles at the different cells, dynamic rate adaptation based on (6) results in close to the *optimal*[4] rate adaptation [7].

The power allocation at the basic rate corresponding to the different MSs is determined as follows [7]:

$$P_{b,i}^{(j)} = \frac{P_B - P_c}{\sum_{i=1}^{g_j} m_i^{(j)}\mu_i^{(j)}} \cdot \mu_i^{(j)}, \qquad i = 1, \ldots, G. \tag{7}$$

Based on the above formulation, the following two variations for joint rate and power allocation, namely, the DFA and the BFA algorithms, are presented which allocate the radio resources among TCP connections in an exhaustive and a round-robin fashion, respectively. Here, the connections are assumed to be logically arranged in the increasing order of the corresponding interference factors.

DFA: In this strategy, the rate allocation is made such that the sum-rate throughput in the tagged cell, namely, $\sum_{i=1}^{G} m_i^{(j)}$, is maximized while satisfying the constraint in (6). The rate allocation strategy is stated on the next page.

In this case, the maximum possible transmission rate is allocated to the connection with the lowest interference factor and then if rate capacity resource is available, the maximum possible transmission rate is allocated to the connection with the next lowest interference factor. In this way, transmission rates are assigned to the connections in a "depth-first" fashion starting from the one for which the corresponding interference factor is the lowest. Although this "greedy" mechanism will maximize the radio link level throughput, the throughput fairness may degrade.

BFA: In this case, the radio link level transmission rate corresponding to each TCP connection is first set to be inversely proportional to the corresponding interference factor. Then the rate allocations corresponding to all TCP connections are incremented in a round-robin fashion as far as the sum interference is less than the rate capacity resource γ in (6). The algorithm is stated on the next page.

Based on the rate allocations, the basic power allocation (and, hence, the total power allocation) can be determined using (7). Since the power allocation depends on the rate allocation, this can be interpreted as joint rate and power allocation.

The constrained total maximum transmission power at a BS would limit the rate capacity resource γ in (6) and, hence, the rate allocations corresponding to the TCP connections. For a limited power budget and a desired SIR $(\mathrm{SIR})_o$, the system load G can be controlled, for example, by the RNC to achieve the desired radio link level throughput.

2) Nonuniform Traffic: Similar to the uniform traffic case, the effective interference factors are estimated based on the total power allocations determined from the distribution of traffic load across the cells as follows:

$$P_{B_j} = \left[\frac{g_j}{\frac{1}{J+1}\sum_{j=0}^{J} g_j}\right] P_B \tag{8}$$

for $j = 0, 1, \ldots, J$. Based on the effective interference factors $\mu_i^{(j)}$, $i = 1, \ldots, g_j$, the rate allocations can be performed using the same algorithms as used in the case of uniform traffic distribution. In this case, P_B in (6) would be replaced by $P_{B,j}$ and α_j (for BFA) would be given by

$$\alpha_j = \frac{N}{g_j}\left(1 - \frac{P_c}{P_{B,j}}\right)(\mathrm{SIR})_o^{-1} = \frac{\gamma_j}{g_j}. \tag{9}$$

The power allocation is performed using (7) to have the same downlink quality corresponding to all the TCP connections.

V. Simulation Model and Assumptions

A. Random (Directional) Mobility With Uncorrelated (Correlated) Shadowing

For the random mobility case, the location of each mobile at the beginning of each frame-time is chosen randomly

```
DFA scheme
  (i). Initialize i to 1 and
     weighted_sum_interference to
     0.0.
  (ii). Allocate the maximum possible
     rate m (m ≤ φ) to the ith con-
     nection (in the list) with in-
     terference factor μ_i^(j) such that
     weighted_sum_interference + mμ_i^(j) ≤ γ
     and update weighted_sum_interference to
     weighted_sum_interference + mμ_i^(j).
  (iii). Increment i and repeat step (ii)
     for the next candidate connection
     (in the list).
```

[4]The optimality criterion is the maximization of the *sum-rate throughput* $\sum_{j=0}^{J}\sum_{i=1}^{g_j} m_i^{(j)}$ under constrained SIR.

BFA scheme
(i). As long as the inequality in (6) is satisfied, allocate rates corresponding to the all TCP connections according to $m_i^{(j)} = \min(\langle(\alpha)/(\mu_i^{(j)})\rangle, \varphi)$ $k = 1, \ldots, G$, where $\langle x\rangle$ is the integer nearest to x, $\alpha = (N/G)(1-(P_c/P_B))(\mathrm{SIR})_o^{-1} = \gamma/G$ and $P_{B,j} = P_B$.
(ii). Update the rate allocation incrementally in a round-robin fashion over all the TCP connections such that the sum interference does not exceed γ.

inside the target cell (in a hexagonal cell layout)[5] and the effect of shadowing at different locations is assumed to be uncorrelated.

For the directional random walk model [19], the mobile users travel at a constant speed v from a starting point to a destination point in a series of statistically independent discrete steps and, in this case, the effect of shadowing at the different locations is assumed to be correlated. For each step, the angular deviation (θ) of the travel direction from the *principal direction*[6] has the probability density function $f(\theta)$ given by

$$f(\theta) = \begin{cases} \frac{A_\theta}{2\left[1+A_\theta^2\theta^2\right]tan^{-1}(A_\theta\pi)}, & -\pi \le \theta \le \pi \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

The parameter A_θ controls how close the travel direction is to the principal direction. If a mobile user travels in a forward direction with probability 0.95, the corresponding value for A_θ is 4.2 [19].

The correlated shadowing is modeled as a Gaussian white noise process, filtered through a first degree low-pass filter as follows [20]:

$$\omega_{k+1(\mathrm{dB})} = c \times \omega_{k(\mathrm{dB})} + (1-c) \times u_k \quad (11)$$

where $\omega_{k(\mathrm{dB})}$ is the mean envelope level (in decibels) that is experienced at location k, c is the correlation coefficient given by $c = \varepsilon_d^{vT_s/d}$, and u_k is a zero-mean Gaussian random variable with variance $\tilde{\sigma}^2$. Here, $\tilde{\sigma}^2 = ((1+c)/(1-c))\sigma^2$ with σ^2 being the variance of log-normal shadowing. The parameter ε_d is the correlation between two points separated by distance d, T_s is the sampling interval (which is assumed to be equal to the frame-time T in this paper).

B. Fading Channel Model and Calculation of BER

Multipath fading is assumed to be independently varying. In the random mobility case, an L-path (e.g., $L = 4$) Rayleigh fading channel with uncorrelated scattering and *equal* average path power is considered. Multipath fading with *unequal* average path power is considered for the directional mobility with correlated shadowing case and the parameters are based on the vehicular-B model [21] for macro-cell. With the vehicular-B model, the values for the tapped-delay-line parameters corresponding to the L $(= 6)$ multipaths are $\{-2.5, 0.0, -12.8, -10.0, -25.2, -16.0\}$ dB.

Based on the allocated transmission rates during each frame-time corresponding to the different TCP connections, the achieved $(\mathrm{SIR})_{o,d}^{(j)}$ is calculated using (5) and then depending on the multipath fading channel model, the bit-error probability p_b is calculated.

1) Multipath Rayleigh Fading Channel With Equal Path Gain (Channel Model-A): For an L-path Rayleigh fading channel with uncorrelated scattering and equal average path power, assuming maximal-ratio combining (MRC) with independent L branches, p_b is given by [17]

$$p_b = \left[\frac{1}{2}(1-\kappa)\right]^L \sum_{l=0}^{L-1} \binom{L-1+l}{l} \left[\frac{1}{2}(1+\kappa)\right]^l \quad (12)$$

where $\kappa = \sqrt{(\mathrm{SIR})_{o,i,l}^{(j)}/[1+(\mathrm{SIR})_{o,i,l}^{(j)}]}$ and $(\mathrm{SIR})_{o,i,l}^{(j)} = (\mathrm{SIR})_{o,d}^{(j)}/L$.

2) Multipath Rayleigh Fading Channel With Unequal Path Gain (Channel Model-B): For maximal-ratio combining with independent L branches, p_b can be expressed as [17]

$$p_b = \frac{1}{2}\sum_{l=1}^{L} \pi_l \left[1 - \sqrt{\frac{\kappa_l}{1+\kappa_l}}\right] \quad (13)$$

where $\pi_l = \prod_{\substack{l'=1 \\ l' \ne l}}^{L} \kappa_l/(\kappa_l - \kappa_{l'})$ with $\kappa_l = (\mathrm{SIR})_{o,i,l}^{(j)}$ and $(\mathrm{SIR})_{o,i,l}^{(j)} = \sigma_l^2 (\mathrm{SIR})_{o,d}^{(j)}$ with the constraint $\sum_{l=1}^{L} \sigma_l^2 = 1$ (normalized) for $\sigma_l^2 = \mathbf{E}\{a_{i,j,l}^2\}$ (derived from the ETSI vehicular B-model [21]).

C. Scheduling and Buffering

Per-destination queuing with first-in first-out scheduling (within each queue) is assumed for the TCP segments at the BS. The TCP segments corresponding to the different connections are assumed to arrive at the BS in their original transmission order.[7] The buffer size at the BS is assumed to be sufficiently large so that there is no buffer overflow loss. Due to limited wireless channel bandwidth, at the TCP sender, the maximum transmission window size for a flow (for which the transmission pipe would be full) would be rather small and with a buffer size of the order of this window size there will be no buffer overflow.

The radio link level frames corresponding to a TCP segment are buffered in the corresponding RLC/MAC level queue (Fig. 2). From the ith RLC/MAC layer queue, m_i frames are transmitted during each frame-time, where m_i $(0 \le m_i \le \varphi)$ is determined using a dynamic rate adaptation procedure and a mobile specific channelization code and the cell specific scrambling code are used for the downlink transmission. To exploit the variable rate transmission fully from frame-time to frame-time, as long as a TCP segment queue is not empty, a sufficient number of link layer frames are assumed to be available in the corresponding RLC/MAC queue. Therefore, radio frames corresponding to more than one TCP segment may

[5]The random-way point mobility model in $ns2$ is similar to this model except that, in the former case, the user speed is not constant and that at each location the user is assumed to stay for a nonzero interval of time.

[6]At any point on the travel path, the line joining the point to the destination defines the principal direction.

[7]The impact of packet reordering on the end-to-end performance is not considered in this paper.

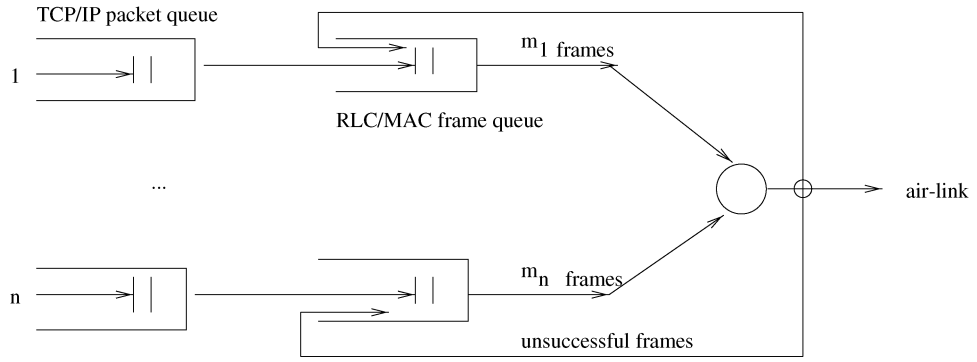


Fig. 2. Queueing of TCP/IP packets and the corresponding radio frames at the BS for downlink transmission.

be enqueued at the same time in the same RLC/MAC queue. Nonzero rates are allocated only to those TCP connections for which the corresponding RLC/MAC layer queues are nonempty.

D. Performance Metrics

Two performance metrics are considered in this paper—average per-connection TCP throughput (λ) and TCP throughput fairness (F). TCP throughput for the ith TCP connection in cell j ($\lambda_i^{(j)}$) is measured as the amount of successfully transmitted TCP data per second for this connection.

The TCP throughput fairness F among $\lambda_i^{(j)}$ for all the TCP connections in $(J+1)$ cells is defined as follows:[8]

$$F = \frac{\left[\sum_{j=0}^{J}\sum_{i=1}^{g_j}\lambda_i^{(j)}\right]^2}{\left[\left(\sum_{j=0}^{J} g_j\right)\sum_{j=0}^{J}\sum_{i=1}^{g_j}\lambda_i^{(j)2}\right]}.$$

The fairness index F measures the global fairness in average TCP throughput and $F = 1.0$ corresponds to the ideal case of perfect throughput fairness. Note that, even for TCP connections with similar RTT throughput, unfairness may be induced by wireless channel errors and the dynamics of the radio link level variable rate allocation.

E. Simulation Methodology

For the random mobility with uncorrelated shadowing case, locations of the users in the corresponding cell are generated randomly at the beginning of each frame-time (which is assumed to be equal to the rate adaptation interval). In the case of directional mobility model with correlated shadowing, the initial locations of the users are generated randomly within the corresponding cells and also the destination locations are assumed to be within the respective cells. The successive locations are generated by using the probability distribution for the angular displacement of the users [as in (10)] and it is ensured that the locations are within the same cells. Note that, the correlation between shadowing in the two successive locations is affected by the mobile speed and the measurement interval.

[8]This is similar to the fairness function used in [22] to quantify the fairness in a shared resource system with n users: $F = (\sum_{i=1}^{n} x_i)^2/(n\sum_{i=1}^{n} x_i^2)$ (where x_i is the ith user's throughput).

In each case, simulations are performed to obtain the maximum radio link level transmission rate corresponding to each TCP connection during each frame-time.[9] For this, the intercell interference factor $\eta_{j'/j}(i)$, corresponding to the ith mobile in cell j, is calculated using (2) to evaluate the effective interference factor $\mu_i^{(j)}$ in (4). In this case, $L_{j'}(j,i) > L_j(j,i), \forall j'$ (i.e., the BS in the "tagged cell" is selected as the serving BS). The values of $L_{j'}(j,i)$ and $L_j(j,i)$, which account for the long-term fading, are assumed to be constant over a frame-time T.

The values of $\zeta_i^{(j')}$ and $\zeta_i^{(j)}$, which account for the short-term fading in (2), are assumed to be constant only over a fraction of the frame-time Δ, where $T = n_s\Delta$. Therefore, the value of $\eta_{j'/j}(i)$ over a frame-time is calculated by using the average of the n_s independent values of $\zeta_i^{(j')}/\zeta_i^{(j)}$. The value of n_s is assumed to be 16 in this paper.

An event-driven simulator that supports the TCP congestion and error control and the radio link level dynamic rate adaptation algorithms is used to analyze the system dynamics under multiple concurrent TCP connections in a wide-area cellular WCDMA scenario. The different events are Time_Frame, Fixed_Host_Generates_Packet, Fixed_Host_Receives_Ack, BS_Receives_Packet, and MS_Sends_Ack. The different TCP mechanisms such as slow-start, congestion avoidance, fast retransmit, and fast recovery are implemented in the Fixed_Host_Receives_Ack event. The Time_Frame event is executed every frame-time and while executing this event the routine to calculate the rate allocations corresponding to the different TCP connections is invoked. The transmission status of a frame (success/failure) is determined by a Bernoulli Trial with parameter p_c as in (1), where p_b for a fading model is determined based on $(\mathrm{SIR})_{o,d}^{(j)}$ in (5).

All the simulations are run until the 95% confidence interval is less than 5% of the corresponding sample mean. In addition to the average per-connection TCP throughput and TCP throughput fairness, for each simulation scenario, the total number of duplicate acknowledgment (DA), the total number of dropped TCP segment (DP), and the total number of TCP TO are also measured.

To observe the impact of the "conservativeness" of TCP congestion control on the end-to-end performance, the behavior of TCP is simulated without fast recovery (e.g., as in TCP Tahoe).

[9]The radio link level transmission rate may fall below this maximum value in the case of soft handoff due to increased $\eta_{j'/j}(i)$.

Since the sender may transmit a large packet burst during fast recovery, this has the effect of preventing packet bursts during recovering from a packet loss. The TCP behavior is also simulated without fast retransmit in which case packet loss is detected using TOs only (as in TCP OldTahoe, the predecessor of TCP Reno/NewReno).

F. Simulation Parameters

Simulation results are obtained under varying traffic load (i.e., number of concurrent TCP connections) distribution across a three-cell system (i.e., for $J = 2$) in a hexagonal cell layout. Note that, increasing the number of cells has basically the same effect (i.e., increasing interference) as that due to increasing system load. With one bulk-data TCP connection per MS, the average number of concurrent TCP connections per cell ($\bar{G}$) is assumed to be 10, 15, and 20 with $\{g_0, g_1, g_2\} = \{8, 10, 12\}$, $\{14, 15, 16\}$, and $\{18, 20, 22\}$, respectively.

The different simulation parameters at the TCP and the RLC/MAC level are TCP segment size (MSS), TCP ACK packet size, TCP maximum window size ($W_{\max}$), TCP fast retransmit threshold (K), TCP timer backoff parameter (Q), initial RTO value (RTO_{initial}), wired-network (or internet) delay (D_{internet}), frame-length in bits (M), frame-length in time (T), radio link level persistence parameter $R_{\max}$, and the FEC parameter (t). The parameter D_{internet} denotes the go-trip-delay experienced by the TCP packets in the wired part of the network and, hence, affects the RTT value of the corresponding TCP connection. An ideal fixed network behavior is assumed where there are no delay jitter and no packet loss for the fixed network part. The values of some of the parameters used for obtaining the simulation results presented in this paper are listed in Table II.

With $M = 127$ b, assuming that BCH codes are used, for $t = 1, 2, 3,$ and 4, the number of information bits transmitted per frame-time would be 120, 113, 106, and 99, respectively [17]. Therefore, for $\text{MSS} = 1000$ B, with (127, 120, 1), (127, 113, 2), (127, 106, 3), and (127, 99, 4) BCH codes used as the radio frames, the number of radio frames corresponding to one TCP segment would be 67, 71, 75, and 81, respectively.[10] For an ACK packet size 52 B (40 + timestamp) and reverse link transmission rate of 16 Kb/s, the ACK transmission time (t_{ACK}) would be 26 ms. With $D_{\text{internet}} = 100$ ms and frame-time of 10 ms, the latency of transmission of a TCP packet would be, therefore, of the order of several hundred milliseconds. The *ping* latencies in a 3G1X system were observed to be of the same order [13].

TABLE II
SIMULATION PARAMETERS

Parameter	Value
Radio frame-length, T (ms)	10
Maxm. spreading gain, N	128
Path-loss exponent, δ	4.0
Var. of shadow fading, σ^2 (dB)	8
In-cell orthogonality factor, ν	0.2
Ratio of control signal power and total transmission power, P_c/P_B	0.2
Correlation for separation d, ε_d	0.1
Directional mobility parameter, A_θ	4.2
Mobile speed, v (km/hr)	20, 50, 80
Maxm. transmission rate, φ (normalized)	8
Radio frame-size, M $(bits)$	127
No. of correctable single bit errors/frame, t	1, 2
Normalized average power budget for a BS transmitter, P_B	1
TCP fast-retransmit threshold, K	3
TCP retransmission timer backoff parameter, Q	2.0
TCP segment size, MSS $(bytes)$	1000
TCP ACK packet size $(bytes)$	52
RTO_{min} (s)	1
RTO_{max} (s)	60
$RTO_{initial}$ (s)	2.5
Maximum TCP window size, W_{max}	8
Go-trip internet delay, $D_{internet}$ (ms)	100

VI. Simulation Results and Discussions

A. Radio Link Level Reliability and TCP Throughput

1) Impact of Variations in Target SIR: Fig. 3 shows normalized radio link level transmission rate for a tagged TCP connection during successive frame-times (during a 5-s simulation run) for both the DFA- and the BFA-based rate adaptation. Note that, the normalization is made with respect to the basic (or minimum) transmission rate for which the average number of transmitted frame per frame-time is one. As is evident, with DFA-based rate adaptation, the rate allocation to a tagged TCP connection is more bursty with the presence of frame periods when nonzero transmission rate is allocated followed by frame periods when no rate is allocated at all. With BFA-based rate adaptation, it is more probable that there is a sustained transmission rate allocated to a tagged TCP connection during successive frame-times and, depending on the radio link condition, higher transmission rate is allocated occasionally to the connection. With DFA/BFA-based rate adaptation, as the value of the target SIR increases, variance among the rates allocated to a tagged connection during successive frame-times increases/decreases.

Typical variations in average per-connection TCP throughput (λ) under DFA- and BFA-based link for different target SIR (i.e., $(\text{SIR})_o$) and radio link level FEC parameter t are shown in Fig. 4 for channel model-A with random user mobility (and, hence, uncorrelated shadowing).

The radio link/PHY level target SIR significantly impacts the TCP throughput performance for both the rate adaptation schemes. Due to the TCPs inherent end-to-end error recovery mechanism, too much "conservativeness" in terms of the target SIR and, hence, radio link level reliability may not be desirable from the perspective of end-to-end throughput (as is evident from Fig. 4). Although the BER (and, hence, the FER) decreases with increasing $(\text{SIR})_o$, the average per-connection TCP throughput (λ) may drop significantly at higher values of

[10]For example, $\lceil (1000 \times 8)/120 \rceil = 67$.

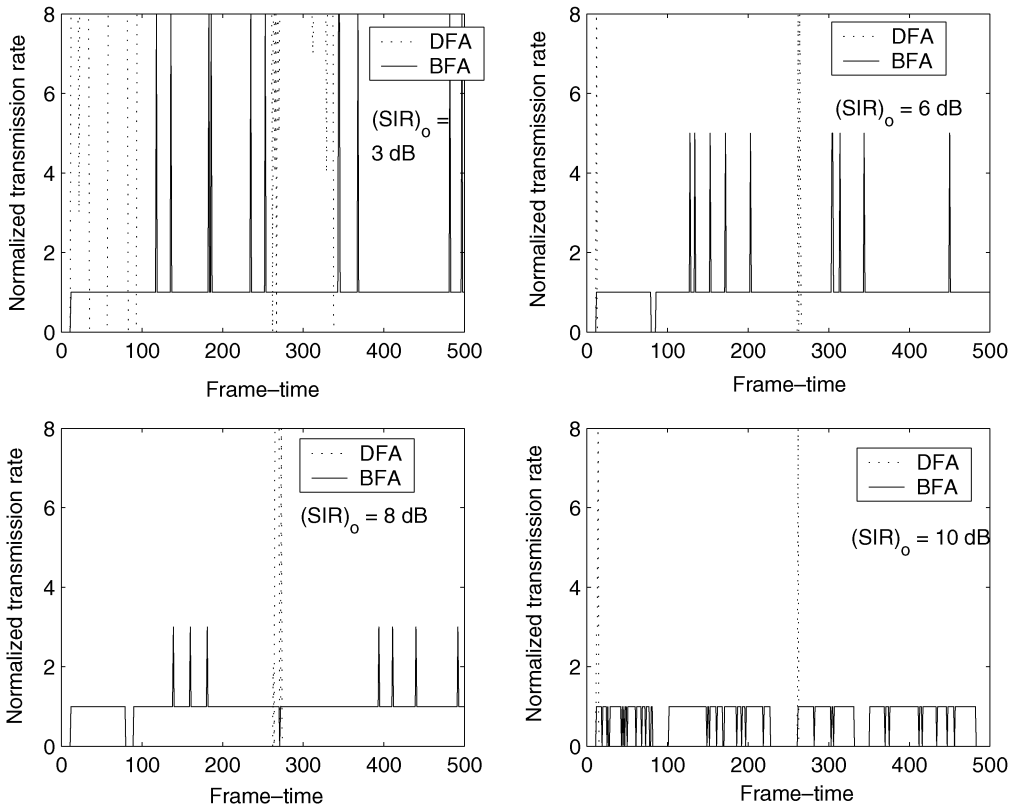


Fig. 3. Normalized radio link level transmission rate for a tagged TCP connection under DFA- and BFA-based link adaptation (for channel model-A with random user mobility).

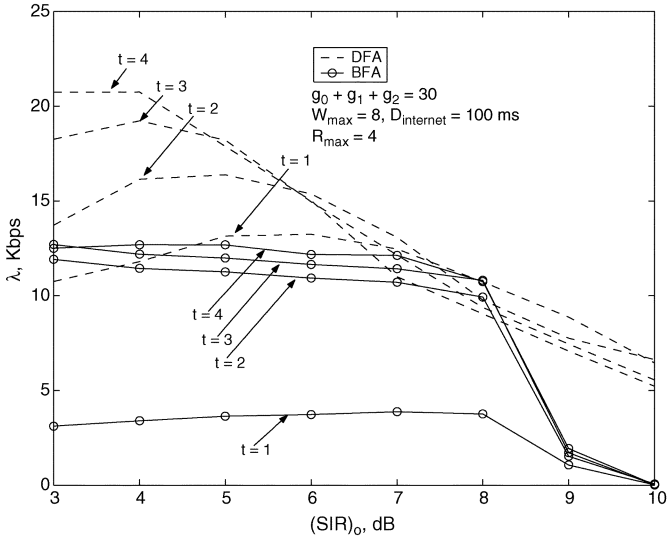


Fig. 4. TCP throughput comparison for radio link adaptation based on DFA and BFA for different target SIR and FEC coding (for channel model-A with random user mobility).

$(\mathrm{SIR})_o$ due to reduced radio link level throughput [as per (6) in which the upper bound of the rate capacity resource in a cell becomes increasingly tightened with increasing $(\mathrm{SIR})_o$]. That is, at higher values of $(\mathrm{SIR})_o$, the impact of reduced FER is masked by the smaller number of successfully transmitted radio frames, as a consequence of which, the number of TCP TOs increases. This results in a reduced transmission rate at the sending host and, hence, reduced end-to-end throughput. Again, for relatively smaller values of $(\mathrm{SIR})_o$ (e.g., 3 dB), although the radio frame transmission rate increases, increased FER causes more radio link level frame retransmissions resulting in reduced radio link level (and, hence, transport level) throughput. Therefore, a concave behavior of the TCP throughput curve as a function of the target SIR is observed which results due to the competing behavior of the diminishing number of radio link level frame errors and the rate capacity resource (γ) in a cell.

TABLE III
NUMBER OF DAS, TCP TOS, AND LOST TCP SEGMENTS FOR DIFFERENT VALUES OF t (IN A 3600-S SIMULATION RUN WITH $g_0 + g_1 + g_2 = 30$, $W_{\max} = 8$, $D_{\text{internet}} = 100$ ms, AND $R_{\max} = 4$ UNDER DFA AND CHANNEL MODEL-A WITH RANDOM USER MOBILITY)

$(SIR)_o$, dB	t = 2			t = 3			t = 4		
	DA	TO	DP	DA	TO	DP	DA	TO	DP
3	8618	1864	1833	12480	2040	1575	12384	1803	1251
4	9186	1892	2021	11823	1949	1547	11656	1836	1389
5	11406	2080	1892	12577	2100	1635	11077	1940	1635
6	12013	2177	1929	10333	2012	1836	9608	1897	1851
7	9696	2071	2143	8380	2158	2531	9264	2223	2572
8	9444	2305	2431	7423	1818	2007	7728	2159	2478
9	5127	1778	2297	6407	2126	2672	5326	1897	2482
10	5245	1906	2301	6234	2071	2359	5406	2080	2462

The fact that the degradation in TCP throughput performance at higher values of $(\mathrm{SIR})_o$ is primarily caused due to the increased number of TCP TOs (resulting from reduced radio link level throughput) can be explained through some of the transport level statistics obtained for the DFA-based rate adaptation scheme as presented in Table III. As $(\mathrm{SIR})_o$ increases, the number of DA decreases while the number of TCP TOs does

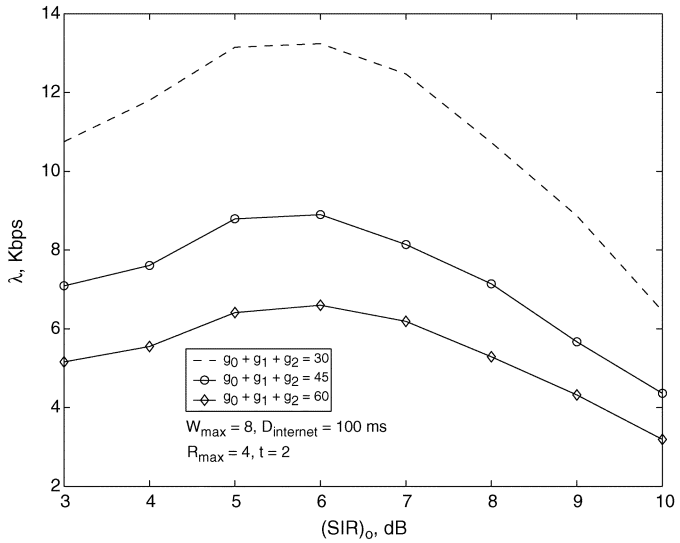


Fig. 5. Impact of network load on TCP throughput for DFA-based link adaptation (for channel model-A with random user mobility).

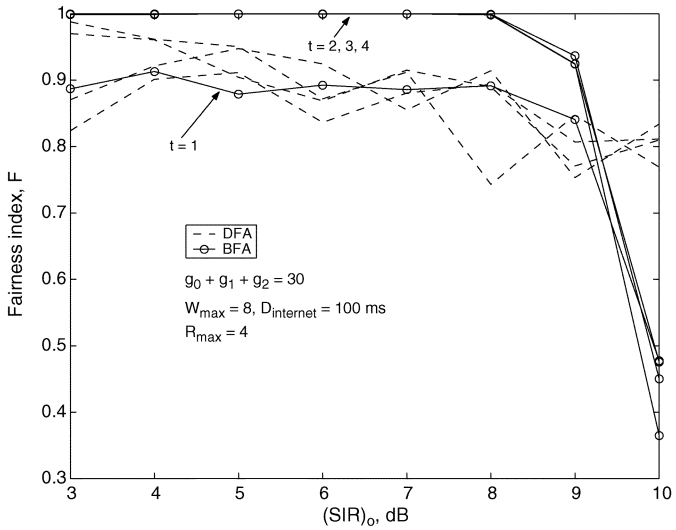


Fig. 6. Comparison among TCP throughput fairness under DFA- and BFA-based link adaptation for different target SIR and radio link level FEC parameter t (for channel model-A with random user mobility).

not change much compared to the cases with smaller $(\mathrm{SIR})_o$. A reduced number of DA along with relatively higher number of TCP TOs imply that the end-to-end error recovery is mostly triggered by the TCP TOs and the TCP sender stays in the slow-start phase most of the time. Note that, for relatively high $(\mathrm{SIR})_o$, the number of lost TCP packets (DP) is generally larger than that for lower $(\mathrm{SIR})_o$ (Table III).

Since TCP throughput variation is not monotonic with $(\mathrm{SIR})_o$, an optimum desired $(\mathrm{SIR})_o$ can be found out which maximizes the average per-connection TCP throughput. This is also evident from the typical variations in TCP throughput with network load (in terms of the number of concurrent TCP connections) for dynamic rate adaptation using DFA, as shown in Fig. 5.

It is to be noted that, for delay-sensitive data (e.g., interactive video) flows, a higher value of $(\mathrm{SIR})_o$ may be more desirable to ensure better reliability at the radio link level since there is generally no end-to-end recovery for this type of flows. However, this will cause reduction in the average radio link level throughput. Therefore, for real-time transport protocols (e.g., user datagram protocol), the selection of $(\mathrm{SIR})_o$ would be based on the desired throughput-reliability tradeoff.

Under similar network conditions, TCP throughput performance is observed to be generally better with DFA-based dynamic rate adaptation. In particular, when the network load and/or the radio link/PHY level target SIR is high, DFA offers remarkably higher end-to-end throughput compared to that due to BFA. Also, with BFA-based rate adaptation, the end-to-end throughput decreases significantly when the value of the radio link level FEC parameter is small. Therefore, channel-gain based "aggressive" dynamic rate adaptation would be desirable over less reliable radio links when the number of concurrent TCP connections is relatively large. Since the variance in the frame transmission delay corresponding to a connection is higher for rate adaptation using DFA,[11] it can be said that, the increased transmission delay variance at the radio link level (e.g., due to more aggressive rate adaptation scheme) may not necessarily deteriorate the per-connection average TCP throughput.

The TCP throughput fairness is observed to be better with BFA-based rate adaptation compared to that for DFA-based rate adaptation (Fig. 6) except when $(\mathrm{SIR})_o$ is high and/or the FEC parameter t is small. In the later cases, the average per-connection TCP throughput deteriorates significantly for rate adaptation using BFA. The reduced throughput fairness with DFA-based rate adaptation in the former cases is due to the more aggressive rate allocation among TCP connections with better channel conditions, as a consequence of which, the number of dropped TCP packets over a time period and the achieved TCP throughput becomes unevenly distributed over all the TCP connections. However, the fairness index with DFA is observed to be around or above 0.8 over the entire range of values of $(\mathrm{SIR})_o$ considered in this paper. The throughput fairness generally deteriorates with increasing traffic load.

The above observations hold true for the different TCP flow/congestion control mechanisms (i.e., for all TCP Reno, TCP Tahoe, and TCP OldTahoe) considered in this paper.

2) Impact of the Radio Link Level FEC Parameter: The effectiveness of radio link level FEC coding in improving the end-to-end throughput performance depends on the value of the radio link/PHY level target SIR and the dynamic rate adaptation scheme used. As is evident from Fig. 4, for DFA-based rate adaptation with lower $(\mathrm{SIR})_o$, TCP throughput can be significantly improved by employing stronger error correcting code at the radio link level. For link adaptation based on BFA, the end-to-end throughput suffers when the value of the radio link level FEC parameter is small.

When $(\mathrm{SIR})_o$ is relatively large (e.g., 8 dB), improved error correction coding at the radio link level does not help much in improving the end-to-end throughput. It may even result in reduced per-connection average TCP throughput (Fig. 4). This is due to the reduced radio link level average transmission rate re-

[11]This is due to the increased variance among rates allocated to a connection during successive frame-times.

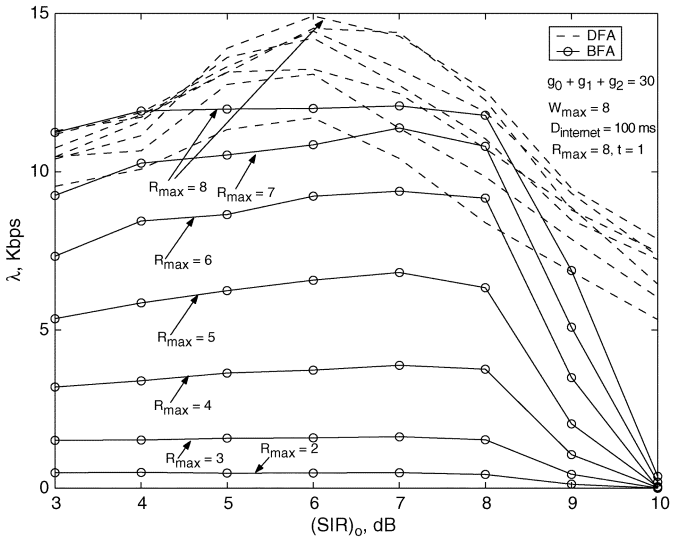


Fig. 7. Impact of radio link level persistence parameter $R_{\max}$ on average per-connection TCP throughput (for channel model-A with random user mobility).

sulting from high $(\mathrm{SIR})_o$. Also, as the value of the FEC parameter t increases, the number of radio frames per TCP segment also increases. Therefore, under the proposed model, the impact of $(\mathrm{SIR})_o$ on the end-to-end throughput is more pronounced than that of radio link level FEC parameter t.

3) Impact of Radio Link Level Persistence Parameter: In our model, the radio link level persistence refers to the persistence in retransmitting a TCP segment rather than persistence in accessing the channel by the radio frames corresponding to a TCP connection. Even though all the radio frames corresponding to a TCP segment are discarded after $R_{\max}$ "black-out" periods during which all the transmitted radio frames corresponding to that segment are lost, radio frames corresponding to other TCP segments for the same TCP connection may be transmitted during the successive frame-times. Note that radio link level dynamic rate adaptation considers only those TCP connections for which the TCP/IP packet queues are nonempty.

As $R_{\max}$ increases, the total number of TCP TOs and the number of dropped TCP segments decrease. Therefore, TCP throughput performance improves as the value of the radio link level persistence parameter $R_{\max}$ increases (Fig. 7). However, the effectiveness of increased persistence in improving end-to-end throughput performance depends on the nature of the radio link level rate adaptation mechanism. Improvement in average per-connection TCP throughput is observed to be much higher in the case of BFA-based link adaptation compared to that for the DFA-based link adaptation. This is due to the fact that since in the case of DFA the distribution of rate allocation is more extreme (i.e., only one or a few number of TCP connections are allocated nonzero transmission rate during a frame-time), for a tagged TCP connection, the radio frames corresponding to a TCP segment are transmitted faster, and therefore, higher values of $R_{\max}$ do not have any impact on the local error recovery. However, for BFA-based link adaptation, rate allocations are more uniform and the radio link level queues for the different TCP connections are emptied at a slower rate. Therefore, an increased value of $R_{\max}$ enables better local error recovery (to avoid end-to-end error recovery).

Simulation statistics reveal that for both the DFA- and the BFA-based link adaptation, the number of DA, TCP TOs, and dropped TCP packets (DP) generally decrease as $R_{\max}$ is increased (Table IV). Note that with $R_{\max} = 5$ for the BFA case, the number of DA decreases abruptly when $(\mathrm{SIR})_o = 10$ dB (last two rows in Table IV). In this situation, the number of successfully transmitted TCP segments is too low to generate a good number of DA. This scenario becomes somewhat better as $R_{\max}$ increases.

Interestingly, for the same value of $R_{\max}$, even though the number of TCP TOs and dropped TCP packets are larger for link adaptation based on DFA, the average per-connection TCP throughput is also higher in this case (Table IV and Fig. 7). Since for DFA-based link adaptation only a few among all the TCP connections use the channel resources aggressively (over a time period), the number of dropped TCP packets and the number of TCP TOs corresponding to other connections increase. But since the average transmission rate is high (due to channel state dependent more aggressive rate allocation), the average per-connection throughput is high. For a particular $(\mathrm{SIR})_o$, the TCP throughput fairness generally improves with increasing $R_{\max}$ and/or t.

B. Impact of User Speed on TCP Throughput

In this paper, with channel model-B, user speed (v) impacts the nature of the correlated long-term fading (i.e., shadowing) while the impact of multipath fading under different user speed is considered only in an average sense. As v increases, the value of low-pass filter gain a [in (11)] increases and the value of $\tilde{\sigma}^2$ decreases, and consequently, the shadowing becomes more random. Note that for channel model-A, the impact of shadowing at the different user locations is assumed to be uncorrelated.

Given that all the mobiles are at the same speed and they require the same SIR, the average per-connection TCP throughput is observed to deteriorate for both DFA- and BFA-based link adaptation as user speed v increases. With $g_0 + g_1 + g_2 = 30$, typical throughput results for $v = 20, 50,$ and 80 km/hr under the BFA-based link adaptation are shown in Fig. 8. For both the DFA- and BFA-based link adaptation simulation, statistics have revealed that the number of TCP TOs and the number of dropped TCP segments observed over a simulation period increase as shadowing becomes more uncorrelated. This results in reduced end-to-end throughput. The above observation is complementary to the observation in [14] that, in Rayleigh multipath fading environment, better performance is achieved as Doppler frequency decreases. Again, as user mobility increases the power control error may become more intense which may further deteriorate the end-to-end throughput performance.

Note that, in situations where the SIR requirements are different for the different TCP connections while the target frame error requirements are the same, and each mobile adapts its required SIR to the channel condition, the low velocity mobiles may require lower SIR than high velocity mobiles for the same target FER. Therefore, the above observation may still hold in these cases.

TABLE IV
NUMBER OF DAS, TCP TOS, AND LOST TCP SEGMENTS FOR DIFFERENT VALUES OF $R_{\max}$ (IN A 3600-s SIMULATION RUN WITH $g_0 + g_1 + g_2 = 30$, $W_{\max} = 8$, $D_{\text{internet}} = 100$ ms, AND $t = 1$ UNDER DFA AND BFA FOR CHANNEL MODEL-A WITH RANDOM USER MOBILITY)

$(SIR)_o$,	$R_{max} = 5$						$R_{max} = 8$					
	DFA			BFA			DFA			BFA		
dB	DA	TO	DP	DA	TO	DP	DA	TO	DP	DA	TO	DP
3	6175	1735	1903	14293	1663	1151	5570	1759	1889	4955	574	183
4	5696	1815	2047	13457	1509	937	6749	1854	2120	3017	356	69
5	9884	2011	2122	12845	1468	878	9207	1836	1996	2550	307	73
6	10060	1943	1893	12008	1358	757	11009	2067	1909	1907	234	44
7	8851	1922	1930	11599	1291	656	10047	2024	1916	1281	151	25
8	7483	1877	2044	11824	1390	760	11556	2306	2188	1493	180	41
9	5614	1957	2410	13554	2235	2410	7357	2141	2565	10715	1187	630
10	4511	1783	2155	287	1786	2226	4646	1805	2384	3866	2228	3065

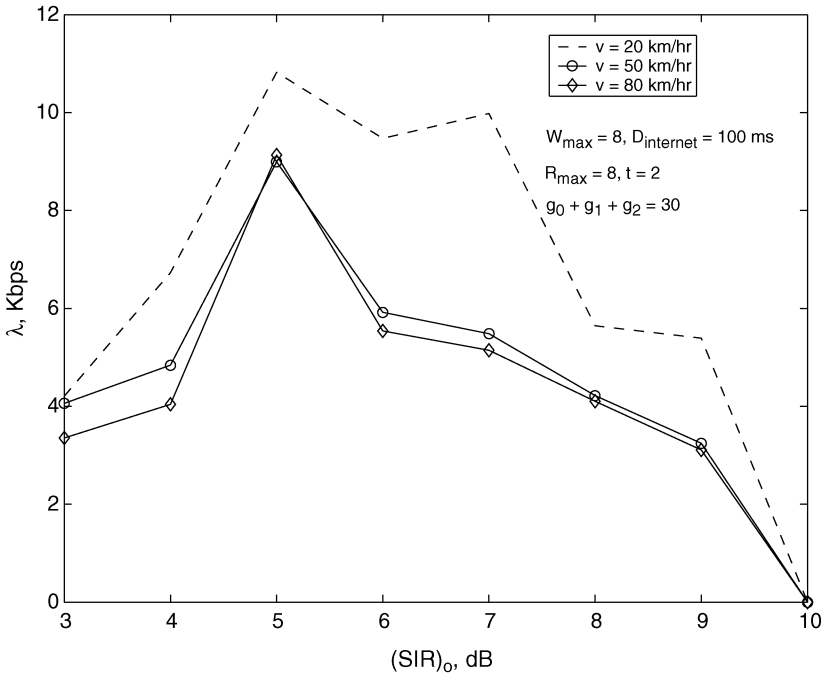


Fig. 8. Impact of user speed on average per-connection TCP throughput for link adaptation based on BFA (for channel model-B with directional user mobility).

C. TCP Parameters and Mechanisms

1) Maximum TCP Window Size, Internet Delay, and Radio Link Adaptation: Under multiple concurrent TCP connections, for a particular link adaptation scheme, the maximum TCP window size ($W_{\max}$) and the internet delay (D_{internet}), which are assumed here to be the same for all concurrent TCP connections, impact the per-connection average TCP throughput. The relationship among maximum transmission window size, internet delay, radio link adaptation, and TCP throughput can be explained as follows. For a tagged connection, if t_{TCP} is the transmission delay (from BS to MS) of a TCP/IP packet and t_{ACK} is the corresponding ACK delay (from MS to BS), then the minimum value of the maximum TCP transmission window size for which the "packet pipe" corresponding to the tagged connection would be full,[12] can be approximated as follows:

$$W_{\max}^{(\min)} \approx 1 + \frac{2 \times D_{\text{internet}}}{t_{\text{TCP}}} + \frac{t_{\text{ACK}}}{t_{\text{TCP}}}. \tag{14}$$

[12]For this value of the transmission window size, the BS transmitter would always have a packet for transmission corresponding to the tagged connection.

Note that, the value of t_{TCP} here depends on the nature of the radio link adaptation mechanism and the physical layer channel reliability parameter such as $(\text{SIR})_o$. As t_{TCP} increases, $W_{\max}^{(\min)}$ decreases. If the maximum transmission window size $W_{\max}$ is smaller than $W_{\max}^{(\min)}$ (which implies that the value of the time-varying transmission window size $W(t)$ for the tagged connection is always smaller than $W_{\max}^{(\min)}$), TCP throughput would be lower since the "packet pipe" would not be full (i.e., the network would be under utilized). Again, if $W_{\max}$ is greater than $W_{\max}^{(\min)}$ (which implies that the value of the time-varying transmission window size $W(t)$ for the tagged connection is always greater than or equal to $W_{\max}^{(\min)}$), it may cause more TCP TOs and, hence, more frequent invocation of the TCP slow-start phase, and consequently, the packet transmission rate corresponding to the tagged connection may decrease. This causes degradation in the achieved average throughput performance.

For a particular value of the maximum transmission window size, as the number of simultaneous TCP connections increases, t_{TCP} increases with a consequent decrease in $W_{\max}^{(\min)}$. There-

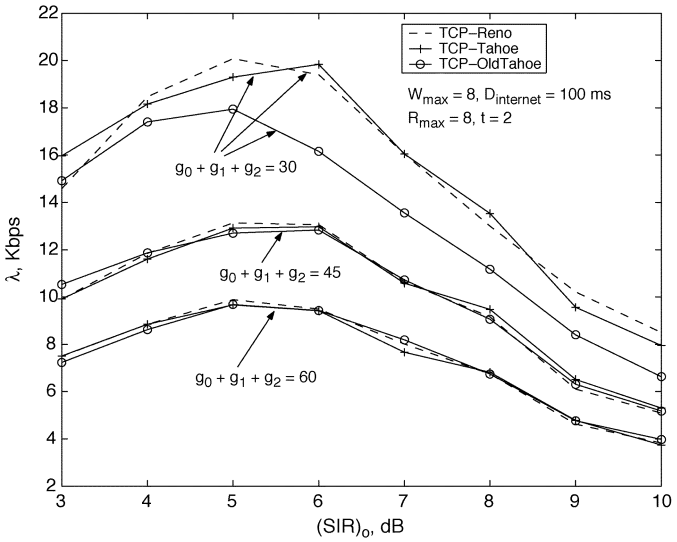


Fig. 9. Performance comparison among TCP Reno, TCP Tahoe, and TCP OldTahoe for link adaptation based on DFA (for channel model-A with random user mobility).

fore, the time-varying transmission window size may exceed $W_{\max}^{(\min)}$ more often (and, hence, more frequent invocation of the TCP congestion control), as a consequent of which the average per-connection TCP throughput may deteriorate. As D_{internet} increases, $W_{\max}^{(\min)}$ also increases, and therefore, $W_{\max}$ will need to be increased so that the transmission rate is high enough to keep the transmission "packet pipe" full. Therefore, for a particular value of $W_{\max}$, the per-connection average throughput may decrease as D_{internet} increases.

2) Conservativeness in TCPs Congestion Control Behavior and Network Load: Since interference limits the performance of a cellular WCDMA network, the impact of increased interference in the WCDMA air link is analogous to that of congestion in a network path. Therefore, in a wide-area WCDMA network, the invocation of the TCP congestion control mechanism in response to TCP segment loss may not be deleterious from the perspective of average per-connection TCP throughput. However, when the radio link is more reliable, a more conservative transport layer error recovery mechanism (as in TCP Tahoe and TCP OldTahoe) may deteriorate the per-connection average TCP throughput.

Typical simulation results on the comparative performance among the three TCP variants are illustrated in Fig. 9 for link adaptation using DFA. Although at lower network load TCP Reno may perform better than TCP Tahoe and TCP OldTahoe, almost similar throughput performances are achieved at higher network load conditions. Simulation statistics on the transport level system dynamics reveal that, at higher network load, almost all of the end-to-end error recoveries are triggered by TCP TOs and for all three TCP variants, the number of TCP TOs and the number of dropped TCP segments observed over a simulation period do not differ significantly.

With BFA-based radio link adaptation, the differences among the end-to-end throughputs for all the TCP variants diminish even when the number of concurrent TCP connections is relatively smaller (compared to the DFA case). Therefore, apart from the network load, the radio link level dynamic rate adaptation also impacts the effectiveness of the "conservativeness–aggressiveness" in TCPs congestion control behavior in improving end-to-end throughput.

Under light network load conditions (e.g., when $g_0 + g_1 + g_2 \leq 30$), TCP Tahoe and TCP OldTahoe may even perform better than TCP Reno when the target SIR is small (Fig. 9). This is due to the reduced instantaneous network load resulting from the less aggressive end-to-end error recovery at the TCP sender after a packet loss. As the instantaneous traffic load decreases, the radio link level throughput increases and, hence, the per-connection average TCP throughput improves. However, for large target SIR, the number of dropped TCP/IP packets increases due to reduced average radio link level frame transmission rate. Therefore, the more aggressive recovery mechanism of TCP Reno results in higher per-connection average TCP throughput compared to that in TCP Tahoe and TCP OldTahoe when $(\text{SIR})_o$ is relatively high (Fig. 9).

VII. Conclusion

In this paper, a multilayer model has been presented to investigate the impacts of the different radio link/PHY level parameters and adaptive radio link level mechanisms on the end-to-end throughput and fairness performances in cellular WCDMA networks. Based on this model, performance of TCP has been evaluated under dynamic radio link adaptation achieved through joint rate and power allocation for downlink data transmission in cellular multirate WCDMA networks. End-to-end system dynamics with multiple concurrent TCP connections in the presence of two interference-based dynamic link adaptation schemes, namely, DFA and BFA, has been analyzed based on the transport level and the radio link level simulation traces. Link adaptation using DFA is more "aggressive" and nonuniform in nature while link adaptation using BFA entails a more uniform (or more resource fair) dynamic rate allocation at the radio link level. Effectiveness of the "conservativeness–aggressiveness" in TCPs congestion control behavior in improving end-to-end throughput under the above mentioned dynamic link adaptation schemes has been also analyzed for different traffic load (in terms of the number of concurrent TCP connections) conditions.

The following provides a summary of the key results.

1) The end-to-end TCP throughput and TCP throughput fairness in a cellular WCDMA network depend largely on the radio link adaptation mechanisms and the relevant radio link/PHY level parameters. Under the presented dynamic link adaptation schemes, too much conservative a value for the target SIR (at the radio link/PHY level) impairs the average TCP throughput performance. Better throughput fairness is achieved for a more uniform dynamic link adaptation (e.g., by using BFA) at the radio link level. However, significantly better TCP throughput is achieved with a more aggressive dynamic link adaptation scheme (e.g., by using DFA) when the network load and/or the value of the target SIR is high and/or the radio link is less reliable.

2) Increased radio link level persistence on retransmitting the failed radio frames is more effective in increasing the end-to-end throughput when a more uniform rate allocation scheme is used at the radio link level.
3) In the presented multilayer system model, increasing the target SIR may reduce the effectiveness of increased FEC coding at the radio link level to improve the end-to-end throughput performance. Therefore, the target SIR level and the FEC coding parameter t can be jointly optimized in the lower layer protocol design to achieve high TCP throughput.
4) The effectiveness of the "conservativeness–aggressiveness" in TCPs congestion control behavior in improving end-to-end throughput in a cellular multirate WCDMA network largely depends on the network load. When the number of concurrent TCP connections is relatively high, similar throughput performances are observed for TCP Reno, TCP Tahoe, and TCP OldTahoe.
5) Per-connection average TCP throughput under dynamic rate and power adaptation improves as shadowing becomes more correlated.

ACKNOWLEDGMENT

The authors would like to thank all the reviewers for their constructive comments and suggestions.

REFERENCES

[1] R. V. Nobelen, N. Seshadri, J. Whitehead, and S. Timiri, "An adaptive radio link protocol with enhanced data rates for GSM evolution," *IEEE Pers. Commun.*, vol. 6, pp. 54–64, Feb. 1999.
[2] M. O. Sunay and J. Kähtävä, "Provision of variable data rates in third generation wideband DS CDMA systems," in *Proc. IEEE WCNC'99*, pp. 505–509.
[3] S. A. Jafar and A. Goldsmith, "Optimal rate and power adaptation for multirate CDMA," in *Proc. IEEE VTC 2000 (Fall)*, pp. 994–1000.
[4] A. Sampath, P. S. Kumar, and J. M. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," in *Proc. PIMRC'95*, pp. 21–25.
[5] S.-J. Oh and K. M. Wasserman, "Adaptive resource allocation in power constrained CDMA mobile networks," in *Proc. IEEE WCNC'99*, pp. 510–514.
[6] D. Ayyagari and A. Ephremides, "Power control based admission algorithms for maximizing throughput in DS-CDMA networks with multimedia traffic," in *Proc. IEEE WCNC'99*, pp. 631–635.
[7] D. I. Kim, E. Hossain, and V. K. Bhargava, "Downlink joint rate and power allocation in cellular multi-rate WCDMA systems," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 69–80, Jan. 2003.
[8] A. Chockalingam and G. Bao, "Performance of TCP/RLP protocol stack on correlated fading DS-CDMA wireless links," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 28–33, Jan. 2000.
[9] Y. Bai, G. Wu, and A. T. Ogielski, "TCP/RLP coordination and interprotocol signaling for wireless internet," in *Proc. IEEE VTC'99*, pp. 1945–1951.
[10] N. M. Chasker, T. V. Lakshman, and U. Madhow, "TCP over wireless with link level error control: analysis and design methodology," *IEEE/ACM Trans. Networking*, vol. 7, pp. 605–615, Oct. 1999.
[11] A. Kumar, "Comparative performance analysis of versions of TCP in a local network with a lossy link," *IEEE/ACM Trans. Networking*, vol. 6, pp. 485–498, Aug. 1998.
[12] H. Balakrishnan, S. Seshan, and R. Katz, "A comparison of mechanisms for improving TCP performance over wireless links," *IEEE/ACM Trans. Networking*, vol. 5, pp. 756–769, Dec. 1997.
[13] M. C. Chan and R. Ramjee, "TCP/IP performance over 3G wireless links with rate and delay variation," in *Proc. ACM MobiCom 2002*, Atlanta, GA, Sept. 23–26.
[14] M. Zorzi, M. Rossi, and G. Mazzini, "Throughput and energy performance of TCP on a wideband CDMA air interface," *Wireless Commun. Mobile Computing*, vol. 2, no. 1, pp. 71–84, Feb. 2002.
[15] *Computing TCP's retransmission timer*, RFC 2988, Proposed Standard, Nov. 2000.
[16] S. Floyd and K. Fall, "Promoting the use of end-to-end congestion control in the internet," *IEEE/ACM Trans. Networking*, vol. 7, pp. 458–472, Aug. 1999.
[17] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.
[18] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
[19] T. K. Kim and C. Leung, "Generalized paging schemes for cellular communications systems," in *Proc. IEEE PACRIM'99*, pp. 217–220.
[20] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Lett.*, vol. 27, no. 23, pp. 2145–2146, Nov. 1991.
[21] *Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000*, Rec. ITU-R M.1225, 1997.
[22] D. Chiu and R. Jain, "Analysis of increase and decrease algorithms for congestion avoidance in computer networks," *Comput. Networks ISDN Syst.*, vol. 17, pp. 1–14, June 1989.

Ekram Hossain (S'98–M'01) received the B.Sc. and M.Sc. degrees in computer science and engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1995 and 1997, respectively. He received the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2000. He was a University of Victoria Fellow.

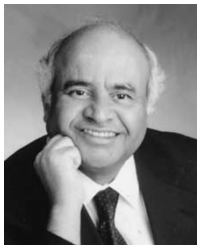
He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Manitoba, Winnipeg, MB, Canada. His research interests include radio link control and transport layer protocol design issues for the next-generation wireless data networks.

Currently, Dr. Hossain serves as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the *Journal of Communications and Networks (JCN)*. He served as a Technical Program Committee Member for the IEEE WCNC 2004, IEEE GLOBECOM 2003, and IEEE VTC 2003 (Fall).

Dong In Kim (S'89–M'91–SM'02) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1980 and 1984, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California (USC), Los Angeles, in 1987 and 1990, respectively.

From 1984 to 1985, he was with the Korea Telecom Research Center, Seoul, Korea, as a Researcher. During 1986–1988, he was a Korean Government Graduate Fellow in the Department of Electrical Engineering, USC. From 1991 to 2002, he was with the University of Seoul, Seoul, Korea, leading the Wireless Communications Research Group. Since 2002, he has been with Simon Fraser University, Burnaby, BC, Canada, where he is an Associate Professor of the School of Engineering Science. He was a Visiting Professor at University of Victoria, Victoria, BC, Canada, during 1999–2000. He has performed research in the areas of packet radio networks and spread-spectrum systems since 1988. His current research interests include spread-spectrum systems, cellular mobile communications, indoor wireless communications, and wireless multimedia networks.

Dr. Kim has served as an Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: WIRELESS COMMUNICATIONS SERIES and also as a Division Editor for the *Journal of Communications and Networks*. Currently he serves as an Editor for Spread Spectrum Transmission and Access for the IEEE TRANSACTIONS ON COMMUNICATIONS and as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.



Vijay K. Bhargava (S'70–M'74–SM'82–F'92) received the B.Sc., M.Sc., and Ph.D. degrees from Queen's University, Kingston, ON, Canada in 1970, 1972, and 1974, respectively.

Currently, he is a Professor and Chair of Electrical and Computer Engineering at the University of British Columbia, Vancouver, BC, Canada, and holds a Canada Research Chair in Wireless Communications. He is a coauthor of the book *Digital Communications by Satellite* (New York: Wiley, 1981) and coeditor of *Reed-Solomon Codes and Their Applications* (New York: IEEE Press). His research interests are in multimedia wireless communications.

Dr. Bhargava is a Fellow of the B.C. Advanced Systems Institute, Engineering Institute of Canada (EIC), the Canadian Academy of Engineering, and the Royal Society of Canada. He is a recipient of the IEEE Centennial Medal (1984), IEEE Canada's McNaughton Gold Medal (1995), the IEEE Haraden Pratt Award (1999), the IEEE Third Millennium Medal (2000), and the IEEE Graduate Teaching Award (2002). He is very active in the IEEE and was nominated by the IEEE Board of Directors for the office of IEEE President-Elect in the year 2002 election. Currently, he serves on the Board of the IEEE Information Theory and Communications Societies. He is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is a Past President of the IEEE Information Theory Society.