# Cross-Layer Analysis of Wireless TCP/ARQ Systems over Correlated Channels

Yi Wu, Zhisheng Niu, and Junli Zheng

*Abstract:* **In this paper, we present a cross-layer analysis of wireless TCP systems over correlated channels. The effects of error correlation on the behavior of link retransmission strategy and the end-to-end throughput of TCP layer are investigated. Based on the cross-layer analysis, an efficient refinement of link layer protocol is proposed by consciously utilizing the information of channel correlations, which leads to the performance improvement of wireless TCP systems.**

*Index Terms:* **Correlated channels, cross-layer, SR ARQ, wireless TCP.**

## I. INTRODUCTION

With the development of 3G and beyond, the increasing trend of IP-based wireless data service is undeniable [1], [2]. As the vital component of the transport layer in Internet protocol (IP) suite, transmission control protocol (TCP) is intended to provide connection-oriented reliable service over unreliable networks. Because of the original design and development in wired networks, TCP has the well-known weakness on wireless links with significant errors. Therefore, the link layer (LL) protocols such as ARQ and FEC are needed to protect TCP from the error-prone wireless channels.

Actually, because of the acknowledged fading characteristics of wireless mobile environment, the dependence between errors in the transmission process absolutely affects the system performance. Related work has been presented in [3] by Zorzi, where the impact of error burstiness on the throughput performance of wireless TCP without the cooperation of radio link protocol is studied. It is observed that the TCP performance depends significantly on the channel error correlation and the link layer design is supposed to be important for the system performance.

Now research works have begun to focus on performance enhancing mechanisms at all layers of the network in order to deliver high performance at the end-user level. Residing between the physical layer and the transport layer, the link layer plays an important role in protocol stacks of the wireless networks. A better understanding of link layer behavior on fading channel would significantly contribute to the design and evaluation of the higher layer protocols such as TCP.

There have been several papers on performance evaluation of TCP over radio link protocols (RLP) in wireless networks [4]–[12]. However, it is noted that most of the analytical models are based on the ideal ARQ schemes at the link layer. As the practical radio link protocol which has been employed by the 3G standard, the particular NAK-based multi-copy selective repeat (SR) ARQ scheme [2] needs more investigations on the statistic behavior over correlated fading channels and its interactions with TCP.

In this paper, we focus on the cross-layer analysis of all the three layers: The physical layer with correlated errors, the radio link layer with NAK-based multi-copy selective repeat (SR) ARQ scheme, and the transport layer with TCP NewReno. The effects of error correlation on the behavior of link layer strategy and the interactions between end-to-end TCP and local link layer protocols are investigated analytically. Having an insight into the behavior of link layer over fading channels, an improved retransmission scheme is proposed by making full use of the error correlation for configuration designs and optimizations of wireless link layer. The performance enhancement brought by the link layer refinement is also evaluated at the TCP layer.

This paper is organized as follows. In Section II, the related works are discussed. In Section III, the system models which include the correlated fading channel at physical layer, the multi-copy SR ARQ local retransmission at link layer and the end-to-end TCP NewReno at the transport layer, are presented. The analytical approach is introduced in Section IV. Section V provides the numerical results and corresponding discussions. In Section VI, an improved retransmission strategy of link layer is proposed and the enhancement of system performance is analyzed and validated by simulations. Finally, conclusions are presented in Section VII.

## II. RELATED WORKS

There have been a large volume of literatures on the interaction investigation between TCP and wireless radio link protocols. In this section, we present a brief review of the previous works and state the contribution of our work.

In [4] and [5], the effects of TCP behavior on data transmission over IS-2000 wireless links are studied based on computer simulation. Based on the simulation results, it is shown that the NAK-based non-reliable ARQ recovery scheme is crucial to mitigate the incompatibilities of TCP/IP with lossy radio links and improve the performance of wireless TCP.

In [6] and [7], the analytical models on TCP over ARQ schemes are presented. The transport layer and the radio link layer are considered in a combined model for the cross-layer study. However, as for ARQ, only the Go-Back-$N$ ARQ with fully persistent policy is analyzed. Moreover, the correlation effect of the wireless fading on the TCP/RLP performance is not studied and the simulation validation is not provided sufficiently.

---

Manuscript received October 28, 2003; approved for publication by Sudhir Dixit, Division II Editor, July 29, 2004.

The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing, P. R. China, email: wuyee00@mails.tsinghua.edu.cn, niuzhs@tsinghua.edu.cn, zhengjl@atm.mdc.tsinghua.edu.cn.

Part of this paper was presented at the IEEE VTC 2004 Spring, Milan, Italy, 2004.

1229-2370/05/$10.00 © 2005 KICS

In [8], an analytical model of the NAK-based multi-copy SR ARQ is developed for performance study of TCP over RLP in cdma2000 systems. However, this work does not consider the correlated fading effects and the channel errors are assumed to be independently and identically distributed (i.i.d.) in analysis, which is not feasible in wireless communication systems.

Reference [9] studies the throughput performance of transport control protocol/radio link protocol (TCP/RLP) stack on correlated fading wireless channels. The results show that because of significant burstiness in RLP frame errors of the highly correlated Rayleigh fading, longer persistence at the RLP layer to recover lost RLP frames is beneficial at low-link fading margins. However, since the TCP and RLP throughput performance are both obtained through simulations, the analysis of the link behavior over correlated channels needs to be improved.

In [10] and [11], the dependency of TCP and RLP performance on various RLP configuration parameters in the correlated fading channels are investigated based on model analysis and simulations. The results show that the RLP retransmission configuration (1,1,1,1,1,1) has better performance in highly correlated channels compared with the standard RLP retransmission configuration (1,2,3).

In [12], the interactions between ARQ mechanisms and TCP over wireless links are investigated in a wired-cum-wireless network scenario. This work focuses on the investigation of TCP/RLP performance with different ARQ persistence policies and RLP configurations. However, the TCP and RLP behaviors with varying channel correlation degrees are not considered and therefore, it does not reveal the correlation effect of the channel errors on the TCP/RLP performance.

In this paper, we investigate the interactions between TCP and radio link protocols in the presence of correlated channel errors. The contributions of this paper are listed as follows: i) We develop an analytical model for the NAK-based multi-copy SR ARQ schemes over correlated channels and get an insight into the effect of the error correlation on the TCP/RLP performance. ii) Based on the proposed analytical model, we propose an intuitive and efficient solution to improve the TCP/RLP performance in the highly correlated channels.

## III. MODEL DESCRIPTION

### A. Correlated Channel Model of the Physical Layer

The error process at the frame level of wireless channel has been modeled as a first-order, discrete-time, and two-state Markov chain, which is shown to be accurate for correlated fading [13] in spite of its simplicity. Let $G$ and $B$ denote *good* and *bad* states in which frame transmissions would be successful (without frame errors) and unsuccessful (with frame errors), respectively. Then, the one-step transition probability matrix of the Markov chain is given by

$$\Gamma = \begin{bmatrix} 1-\gamma_{GB} & \gamma_{GB} \\ \gamma_{BG} & 1-\gamma_{BG} \end{bmatrix}, \tag{1}$$

where $\gamma_{XY}$ denotes the probability that the current frame transmission is successful (if $Y = G$) or unsuccessful (if $Y = B$), given that the previous frame transmission was unsuccessful (if $X = B$) or successful (if $X = G$).

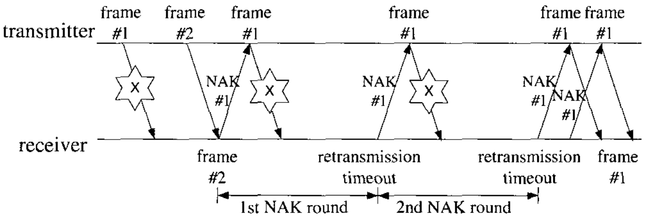


Fig. 1. Illustration of NAK rounds.

Denoting by $P_E$ the frame error rate of physical layer and by $f_dT$ the normalized Doppler bandwidth which describes the correlation of the fading process, the Markov parameters can be derived as [13]

$$\gamma_{GB} = Q(\theta, \rho\theta) - Q(\rho\theta, \theta), \tag{2}$$

$$\gamma_{BG} = \frac{1-P_E}{P_E}(Q(\theta, \rho\theta) - Q(\rho\theta, \theta)), \tag{3}$$

where

$$\theta = \sqrt{\frac{-2\ln(1-P_E)}{1-\rho^2}}. \tag{4}$$

Here $\rho = J_0(2\pi f_d T)$ is the Gaussian correlation coefficient of two successive frames with $T$ seconds interval over a fading channel with Doppler frequency $f_d$. $J_0(\cdot)$ is the first kind Bessel function of zero order, and $Q(\cdot,\cdot)$ is the Marcum Q function [14]. In this way, the Markov parameters are totally expressed in terms of the physical parameters $P_E$ and $f_dT$.

In our analysis, we apply this Markov model at frame level of the link layer. Although motivated by the behavior of wireless fading process, the error model considered here also applies to any environment where the frame loss process exhibits correlation, e.g., due to trellis coding.

### B. System Model of Radio Link Protocol

Link layer error recovery is usually categorized into two strategies: Automatic repeat request (ARQ) and forward error correction (FEC). In this paper, we consider the ARQ scheme only because it is always necessary to be the supplementary mechanism even when FEC is implemented. Here we concern the NAK-based multi-copy SR ARQ scheme which is widely applied by the radio link protocols of wireless systems [2]. The receiver does not acknowledge received data frames, but only requests the retransmission of the particular frames that were not received correctly. To prevent from conflicts with TCP retransmissions, the transceiver performs a partial link recovery through a limited number of frame retransmissions. In analysis, the basic unit is a NAK round as in Fig. 1, which is the interval that begins when the receiver supplies the last of one or more NAK frames for a missing data frame described by a NAK list entry and ends when the retransmission timer expires for that NAK list entry. For convenience of expression, we denote by $R$ the maximum number of NAK rounds limited by link layer, and by $N_i$ ($i = 1, \cdots, R$) the number of NAKs to be sent in the $i$-th NAK round which trigger $N_i$ multiple copy retransmissions for the lost data frame.

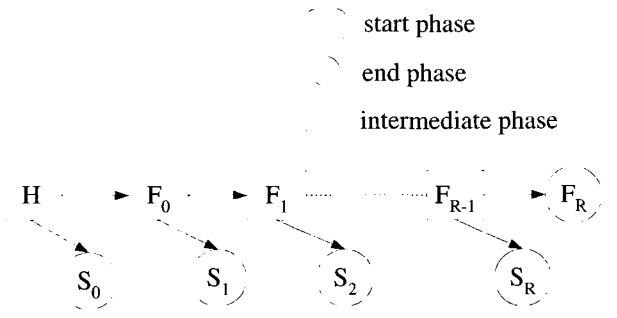


Fig. 2. The phase transition diagram of wireless link recovery process.

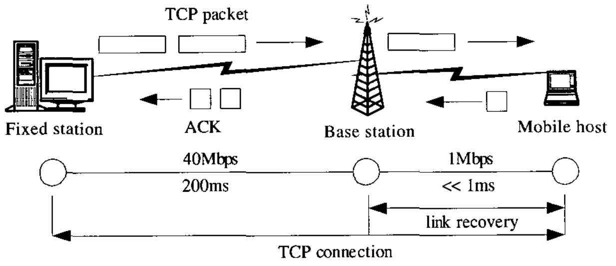


Fig. 3. Network topology of the end-to-end system.

Upon the receiver detects a frame loss, it will create a NAK list entry correspondingly, which contains the sequence number of the missing data frame, an associated NAK round counter and an associated retransmission timer. Subsequently, the receiver sends back $N_1$ consecutive NAKs requesting the multi-copy retransmission, increments the NAK round counter by 1 and starts the associated timer. When the timer expires and the data frame is not received successfully, the receiver sends $N_2$ consecutive NAKs, increments the NAK round counter by 1, initializes the timer and enters the next NAK round. The process continues until either the missing data is recovered or the NAK round counter exceeds $R$. If all the NAK rounds are unsuccessful in recovering the missing data, the link layer recovery has to give up, leaving the residual error for the upper layer protocols (e.g., TCP).

The system model for the wireless link recovery process is illustrated in Fig. 2 with the phase definition as follows:

$H$: The transmitter begins to supply the particular data frame;
$S_0$: The data frame is received successfully by the receiver;
$F_0$: The data frame is detected missing at the receiving side;
$S_k$: The recovery is successfully accomplished in the $k$-th NAK round, for $k = 1, 2, \ldots, R$;
$F_k$: The recovery process fails in the $k$-th NAK round, for $k = 1, 2, \ldots, R$.

### C. TCP End-to-End Connection

The TCP mechanism is far more sophisticated than that of ARQ at link layer, which uses different and complex types of error recovery strategies [15]. There have been some existing studies concerning the analytical model of TCP performance over lossy links as in [16]. However, plenty of assumptions and approximations have to be made in analysis resulting in the insufficient accuracy. Since our intention is to investigate the influence of link layer behavior on end-to-end TCP throughput, we prefer the simulation method to evaluate performance of the transport layer.

Fig. 3 shows the network topology considered in our work. We employ the TCP NewReno for the end-to-end connection and focus on the scenario where the TCP packets are sent from a fixed station to a mobile host passing through an intermediate base station, which is the most general scenario in practice. Nevertheless, it is observed that the analytical result is not confined by the direction of data flow. The wired part of the network is assumed to have a bandwidth of 40 Mbps with a long propagation delay of 200 ms. As the last hop of the network, the wireless link is assumed to be the bottleneck of data transfer with the limited bandwidth of 1 Mbps. The propagation delay of the local wireless link is assumed to be negligible compared with the frame transmission time. This is a representative situation where the interaction of link layer protocol and TCP is minimized because that the local recovery process could have a more efficient reaction and consume less time than the end-to-end error correction process of TCP.

## IV. ANALYTICAL APPROACH

Without loss of generality, we assume each link frame is carried by a single physical layer frame. For the local wireless link, the propagation delay and the transmission time of a NAK frame are assumed to be negligible compared with the data frame transmission duration. Also, we assume the backward channel to be error-free so that the NAKs will never be lost. This is reasonable in practice due to the small length of the NAK frames.

From the diagram of the phase transition in Fig. 2, it is reality that

$$P(F_k|H) = P(F_k|F_{k-1})P(F_{k-1}|H), \quad 0 < k \le R, \tag{5}$$

where $P(Y|X)$ denotes the transition probability from phase $X$ to phase $Y$.

Since the phase transition probability $P(S_0|H)$ and $P(F_0|H)$ are the successful and unsuccessful probabilities of the original data frame transmission, respectively, we get

$$P(S_0|H) = 1 - P_E, \quad P(F_0|H) = P_E. \tag{6}$$

If the initial transmission of the data frame fails, the link recovery process goes on for retransmissions as illustrated in Fig. 2. According to the NAK-based SR ARQ strategy, the detection of data frame loss is only performed when at least one of the subsequent data frame with a larger sequence number is received successfully. Thus, one can conclude that the first NAK round for any missing data frame must follow a successful transmission of some subsequent frame. Therefore, the initial channel state distribution at the instance of loss detection denoted by $\alpha$ is

$$\alpha = [1 \quad 0]. \tag{7}$$

Fig. 4 illustrates the retransmission process in detail. Since the frame errors are correlated, the sequence of the arbitrary lost frame within the error burst should be derived first. For the two-state Markov chain as in (1), it is well known that the length of the error burst $N$ has the geometric distribution with average value of $1/\gamma_{BG}$

$$P\{N = n\} = \gamma_{BG}(1 - \gamma_{BG})^{n-1}, \quad n = 1, 2, 3, \cdots. \tag{8}$$

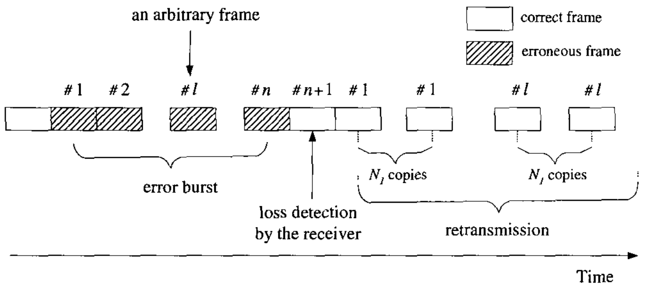


Fig. 4. Illustration of the error burst and the retransmission sequence at receiving side.

Then it is reasonable to assume that the particular lost frame is uniformly distributed within the $N$-frame error burst. Denoting by $L$ the sequence of the lost frame, its probability distribution is given by

$$P\{L=l\} = \sum_{n=l}^{\infty} \frac{1}{n} P\{N=n\} \tag{9}$$

$$= \frac{-\gamma_{BG}}{1-\gamma_{BG}} \ln \gamma_{BG} - \sum_{n=1}^{l-1} \frac{(1-\gamma_{BG})^{n-1}\gamma_{BG}}{n}. \tag{10}$$

Fixing the random variable $L$ as $l$, we have

$$P^{(l)}(F_1|F_0) = \alpha \Gamma^{(l-1)N_1} \Phi^{N_1} \vec{e}, \tag{11}$$

where $\vec{e}$ denotes the all 1's column vector and $\Phi$ denotes the transition probability matrix of an erroneous frame transmission, which represents that the correlated channel falls in the bad state

$$\Phi = \Gamma \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \gamma_{GB} \\ 0 & 1-\gamma_{BG} \end{bmatrix}. \tag{12}$$

In (11), $\Gamma^{(l-1)N_1}$ denotes the transition probability of the channel state during the retransmissions of the previous $l-1$ lost frames of the current error burst, and $\Phi^{N_1}$ represents the probability matrix for the event that all $N_1$ retransmitted copies of frame $l$ are corrupted.

Denote by $\tau_{LL}$ the retransmission timeout interval normalized by the frame transmission time. Using Bayes' formula, from (11) we have

$$P(F_1|F_0) = \sum_{l=1}^{\tau_{LL}} P\{L=l\} P^{(l)}(F_1|F_0), \tag{13}$$

where the summation is truncated at $\tau_{LL}$ because of the assumption that $\tau_{LL}$ is set sufficiently large compared to the transmission time of frame. Recalling (5), the residual error probability after the first retransmission is given by

$$\begin{aligned} P(F_1|H) &= P(F_1|F_0)P(F_0|H) \\ &= \sum_{l=1}^{\tau_{LL}} P\{L=l\} P^{(l)}(F_1|F_0) P_E. \end{aligned} \tag{14}$$

From the second NAK round, the retransmissions are all triggered by timeout events, and therefore the residual error probability $P(F_k|H)$ after $k$ NAK rounds is derived recursively by

$$P(F_k|H) = \beta \left[ \prod_{i=2}^{k} \Gamma^{\tau_{LL}} \Phi^{N_i} \right] \vec{e}\ P(F_1|H), \qquad k = 2, \cdots, R, \tag{15}$$

where $\beta = [0 \quad 1]$ represents the distribution of the channel states after the first failed retransmission ($F_1$), $\Gamma^{\tau_{LL}}$ represents the transition probability of the channel states during the timeout interval $\tau_{LL}$ in each NAK round, and $\Phi^{N_i}$ is the probability matrix that the $N_i$ retransmitted packets lost in the $i$-th NAK round.

Since the number of retransmission times at the link layer is limited by $R$, the residual frame error rate denoted by $P_{LL}$ being provided to TCP is given by

$$P_{LL} = P(F_R|H). \tag{16}$$

To be consistent with the general definition of *throughput* at TCP layer, we denote by $\eta_{LL}$ the normalized throughput of the link layer which is defined as the average number of data frames transmitted during a frame transmission time, independent of whether the delivery is successful or not.[1] As every failure of link recovery in the $i$-th NAK round results in the $(i+1)$-th retransmission of $N_{i+1}$ multiple retransmitted copies, the normalized throughput is derived as

$$\eta_{LL} = \frac{1}{1 + \Sigma_{k=1}^{R} N_k P(F_{k-1}|H)}. \tag{17}$$

Because the maximum number of local retransmission times has been strictly limited (typically set as small as 3), link layer retransmissions are indeed much faster than the end-to-end ones driven by TCP. Note that in [17], the simulation results have shown that the range of 3–10 RLP retransmissions offers the smallest performance degradation caused by TCP and RLP retransmission conflicts. Furthermore, when we focus on a heterogeneous network with the topology as in Fig. 3, where the delay time of the wired portion is much larger than that of the local wired link due to multiple routing and queueing, the impact of the local link recovery delay on the TCP end-to-end round trip time can be ignored. Consequently, the probability of the interference between TCP and local recovery process that may lead to the conflicting retransmissions is kept small enough. Therefore in this sense, the residual error probability and the normalized throughput provided by the link layer perform as the major factors that would influence the end-to-end TCP performance.

## V. NUMERICAL RESULTS AND DISCUSSIONS

To study the interactions among wireless fading channels, link layer protocols and TCP at transport layer, we examine numerical examples based on both analysis and simulation. In the simulation, we employ *ns-2* [19]. Additional source codes for wireless error module and radio link module of the NAK-based multi-copy SR ARQ schemes are developed and implemented. We consider the typical heterogeneous network as shown in Fig. 3 and focus on a downlink TCP connection. A TCP segment is supposed to be 1024 bytes in length, which is then segmented

---

[1] The so-called *goodput* defined as the normalized rate of successful data transmission at the receiver yields $\eta_{LL}(1-P_{LL})$.

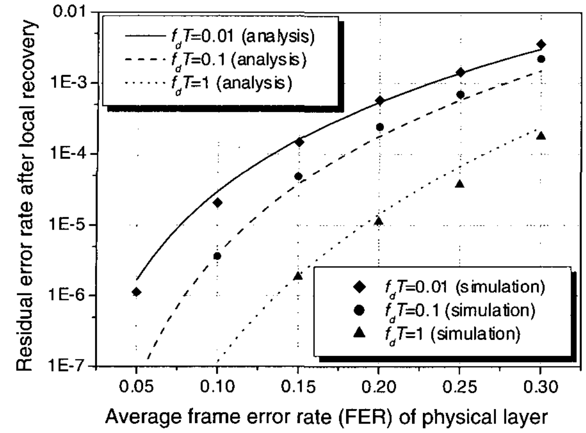


Fig. 5. Residual FER of radio link protocol over correlated channels ($P_{LL}$).

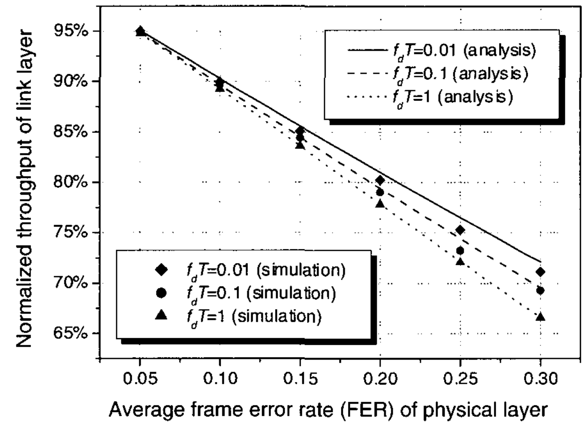


Fig. 6. Normalized throughput of radio link protocol over correlated channels ($\eta_{LL}$).

into 8 link frames equally (i.e., 128 bytes per frame). Suppose the wireless link has the capacity of 1 Mbps, the frame transmission time is determined by 1 ms. The other parameters of the link layer are set as $R = 3$, $\{N_i\} = \{1, 2, 3\}$, and $\tau_{LL} = 80$.

As for the correlated channels, we consider the flat Rayleigh fading channels with the error process characterized by a two-state Markov chain described in Section III. In particular, the degree of error correlation of the channel is decided by the maximum normalized Doppler shift $f_dT$ [13]. Here the values of $f_dT$ adopted are 0.01, 0.1, 1, descending with the degree of the channel correlation. Actually in the case of fast fading channel with $f_dT = 1$, the influence of the underlying physical channel is just the same with that of a memoryless channel with i.i.d. errors from the viewpoint of link layer.

### *A. Performance Issue of Radio Link Layer*

Numerical results of the theoretical analysis on link layer performance are exhibited in Figs. 5 and 6. Besides the analytical curve, simulation points are also given. Fig. 5 compares the residual frame error rate after local wireless link recovery over fading channels with different normalized Doppler bandwidth $f_dT$. The residual frame error rate of link layer increases with correlation degree of the error process. Explanatorily, the bursty characteristic of error process increases the probability of all retransmission trails failing in a deep fading period which leads to the degradation of reliability performance. Fig. 6 shows the normalized throughput performance of link layer over the correlated channels. The results clearly reveal that the data throughput of link layer benefits from correlation of the underlying error process.

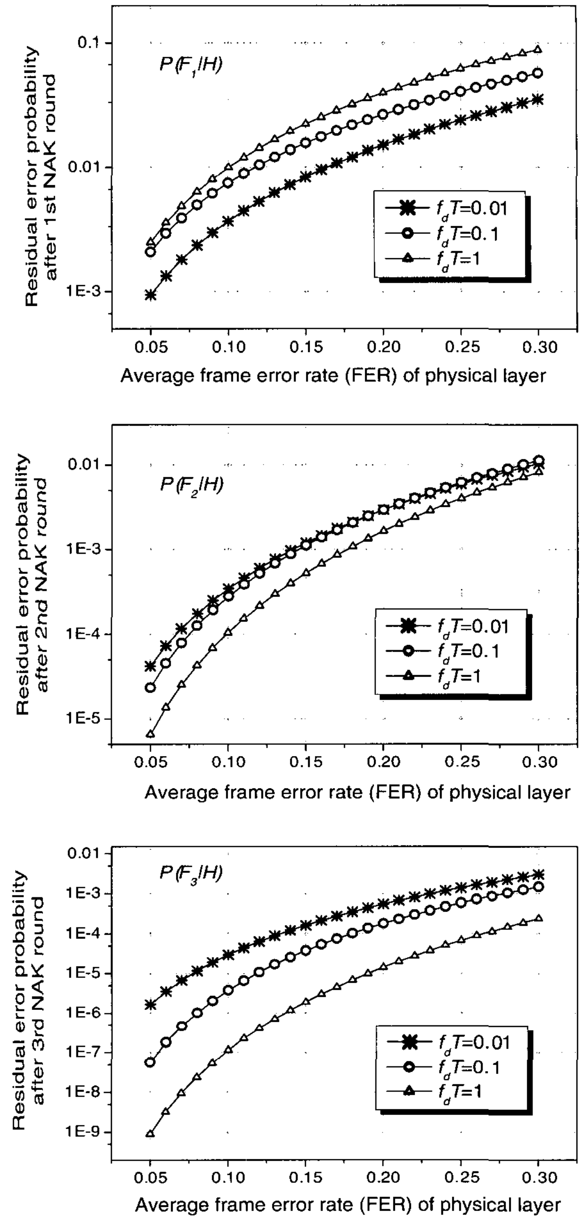


Fig. 7. Evolution of residual FER with increasing NAK round.

What makes the throughput performance of link layer benefit from the error correlation? The relationship between throughput and residual frame error rate of link layer in the environment of burst errors needs to be thought over. To gain an insight into the behavior of link layer protocol, the evolution of the residual error rate of link frames with increasing NAK rounds is illustrated in Fig. 7. Considering the NAK-based SR retransmission process, the first NAK round is not triggered at the receiver until that a successful delivery event of the future data frame occurs, i.e., the initial channel state at the beginning of the first NAK round is good as expressed in (7). Therefore, the retransmission of the first NAK round unconsciously utilizes the correlated in-

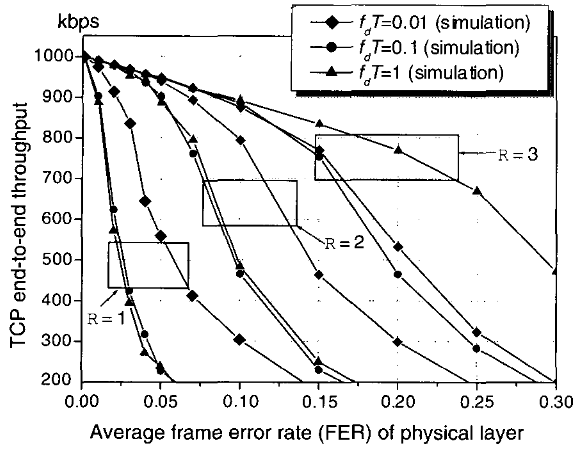


Fig. 8. TCP throughput versus average FER at different values of Doppler bandwidth $f_dT$ and maximum number of retransmission times $R$.

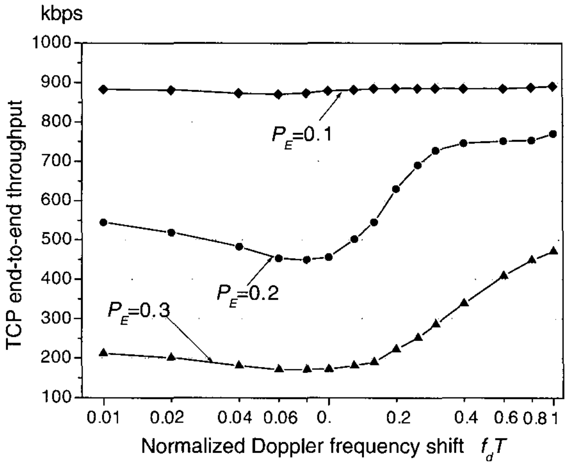


Fig. 9. TCP throughput vs. different normalized Doppler frequency shift ($R$=3).

formation of the physical channels and certainly benefits from the correlation of the error process. Unlike the first NAK round, the subsequent NAK rounds are triggered by timeout events of retransmission timer, so that the benefit from the initial state distribution does not exist anymore and the degradation influence of the burst error process dominates the residual error rate. This finally makes the residual error rate $P(F_3|H)$ smaller in the case of less correlated channels. Then recalling (17), the value of throughput $\eta_{L_L}$ directly depends on weighted aggregation of the residual frame error rates of all NAK rounds. As the dominating element with the largest value (above 10 times of $P(F_2|H)$), $P(F_1|H)$ absolutely determines the relative difference of link layer throughput over different fading channels.

### B. Performance Issue of End-to-End TCP

Next, the performance issue of end-to-end TCP connection at the user level is studied. The results show that with the cooperation of link layer, the TCP performance mainly depends on three factors: 1) The throughput limit of wireless link layer, 2) the residual error probability after local recovery, 3) the correlation degree of the residual errors.

In Fig. 8, the TCP throughput curves with different maximum number of retransmission times $R$ are plotted. The results show that when $R = 1$, the fast fading case $f_d = 1$ performs worst compared with the other slow fading cases. However, as the value of $R$ is increased (from 1 to 3), the performance difference of the residual RLP FER among the three fading cases changes greatly (please refer to Fig. 7). Therefore, dominated by the packet loss probability, the TCP throughput of the fast fading case $f_dT = 1$ becomes the best one in the three fading cases. Note that the same effect of $f_dT$ on the TCP throughput for different $R$ is also observed by [9] through simulations. Here our model analysis has clearly revealed that this interesting effect is caused by the different retransmission triggering mechanisms between the first retransmission and the latter retransmissions.

As shown in Fig. 8, with a low physical frame error rate there is almost no loss event and the TCP end-to-end throughput is chiefly limited by the throughput of wireless link layer which is presented to be near 1 Mbps in Fig. 6. When the channel condition turns worsen, the failure probability of local link recovery increases. Then the incremental losses of TCP segments frequently shrink the TCP congestion window and decline the performance sharply.

In the error-prone situation, the TCP throughput becomes under the joint effect of the other two factors: The residual error probability and the residual error correlation. The higher probability of residual errors leads to larger loss number of the packets and degrades the TCP performance as a result. Contrarily, with higher correlation degree of residual errors the lost frames tend to be clustered so that the halt frequency of TCP data transfer is reduced and the throughput performance of end-to-end connection is improved due to the TCP additional increase multiple decrease (AIMD) mechanism. As both of the two factors increase with the correlation degree of the error process at physical layer, the curves of TCP throughput vs. the normalized Doppler frequency shift $f_dT$ become nonlinear under the joint conflicting influences as shown in Fig. 9. Specifically, in the case of high correlated fading channel, e.g., $f_dT = 0.01$, the performance definitely benefits from the high degree of error correlation. However, with increasing $f_dT$, the error distribution of segmented link frames is randomized which will aggrandize the average error rate of the TCP segments and furthermore degrade the end-to-end performance by discontinuous failures of TCP transmission. In such a situation, the benefit from the lower residual error rate of link layer can not sufficiently compensate the performance degradation caused by error randomization, and consequently the throughput curve turns into a low trough near the area of $f_dT = 0.1$ as in Fig. 9. When the $f_dT$ continues increasing, the correlation among the errors gets further weakened and results in the comparatively low residual error rate of link layer as exhibited in the previous subsection, which greatly contributes to the improvement of TCP throughput and dominantly causes the distinct advantage of the end-to-end performance.

## VI. IMPROVEMENT ISSUE OF WIRELESS TCP: LINK LAYER APPROACH

Optimization of link layer protocol is one of the key methods to improve the performance of wireless TCP. Among all the proposals of TCP enhancement over wireless links, link layer

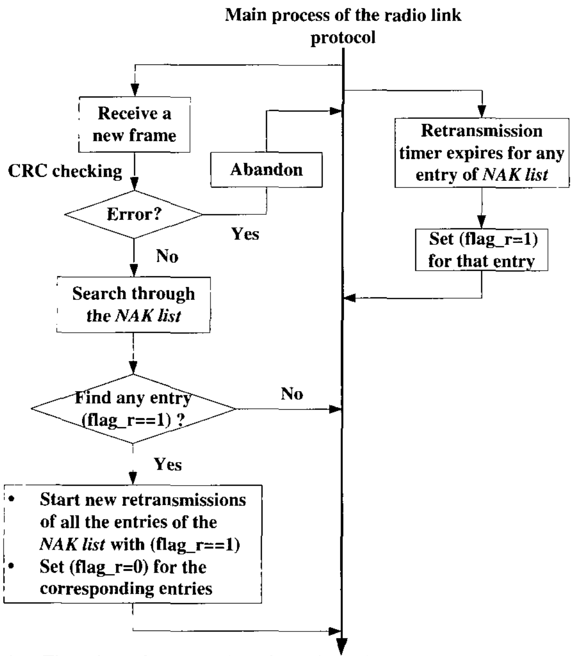


Fig. 10. Flowchart for added codes of the improved scheme at the receiver.

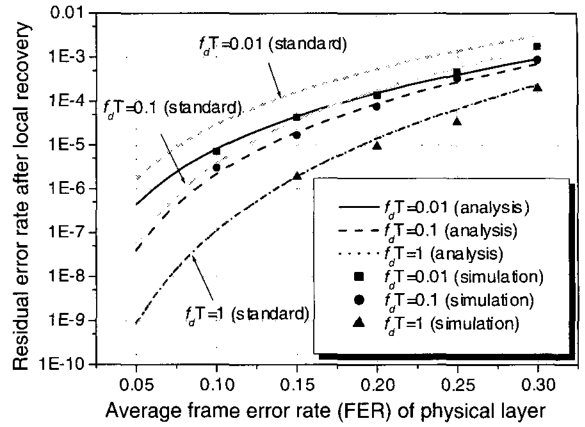


Fig. 11. Residual FER of the improved link retransmission scheme.

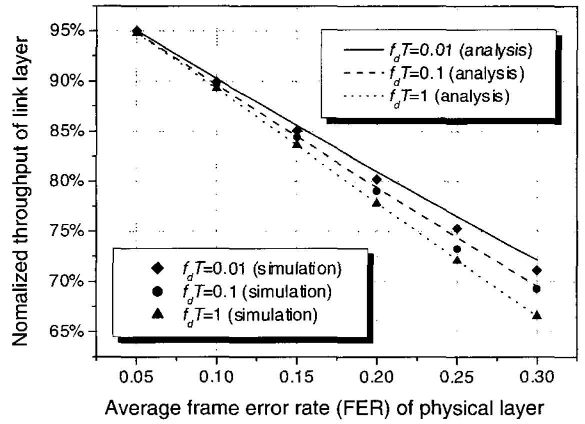


Fig. 12. Normalized throughput of improved link retransmission scheme.

approaches have the advantage over higher layer ones of providing a solution that can be locally employed, without requiring changes to higher layers, without the violation of the end-to-end semantics of TCP, and without incompatibility with IP security. They are thus transparent to the rest of the whole network, faster to recover than end-to-end solutions and efficient to exploit the varying information of the physical layer [20]. In this section, we discuss the issue of link layer refinement as an extension of our cross-layer analysis.

The observation and discussion on the interactions of the TCP layer, link layer and physical channels lead us to emphasize the key operation of link layer protocols. Results have shown that especially in the correlated channel with high loss rate, the major factor that limits the end-to-end TCP throughput is the residual error rate provided by the link layer protocol. To increase the number of retransmission times is one possible solution, but it will increase the probability of conflicts between the TCP retransmission and local link recovery. Here, along with our study on the link layer behavior, we propose a simple but efficient method to make the TCP system consciously use the information of channel correlation.

As we have mentioned, although the residual error rate of the link layer $P_{LL}$ is degraded by the correlation of error process, $P(F_1|H)$ keeps predominant with higher correlated error conditions in the first NAK round. This is due to the different triggering mechanism of the first NAK round which unconsciously utilizes the correlated information of the underlying physical channels. The advantage of the first NAK round over correlated channels gives us a heuristic suggestion: It would efficiently enhance the system performance to make all the NAK rounds "learn" from the first NAK round in order to consciously utilize the correlated information of wireless channels.

Particularly, to imitate the first NAK round behavior in our proposal, when the retransmission timeout occurs at receiver, the next NAK round would not start immediately as in the standard link protocol until a successful transmission of the forward channel is detected. In respect that the timeout event gives no information of the underlying channel which is always varying in wireless environment, we resort to the successful receiving of the data frame as the good state indication of the correlated channels to trigger the beginning of next NAK round. Practically, to avoid the unexpected situation that no data transmission exists on the communication pipe, a successful delivery of idle frame with no data [2] should also be accepted as the condition of retransmission triggering. The flowchart is shown in Fig. 10 to summarize the algorithm of the improved strategy.

As an extension of the analytical approach in Section III, we have the following derivations.

1) The failure probability of the original data frame is still determined by the steady state distribution of the correlated channel.

$$\widetilde{P}(F_0|H) = P_E. \tag{18}$$

2) Since every retransmission of the NAK round is triggered by a successful transmission on the forward channel, all the NAK rounds begin with the conditional distribution of channel state as $\alpha = [1 \quad 0]$. And similar to the analysis on the behavior of the first NAK round in Section III, we have

$$\widetilde{P}^{(l)}(F_k|F_{k-1}) = \alpha\Gamma^{(l-1)N_k}\Phi^{N_k}\vec{e}, \tag{19}$$

$$\widetilde{P}(F_k|H) = \sum_{l=1}^{\tau_{LL}} P\{L=l\}\widetilde{P}^{(l)}(F_k|F_{k-1})\widetilde{P}(F_{k-1}|H), \tag{20}$$

$$k = 1, \cdots, R.$$

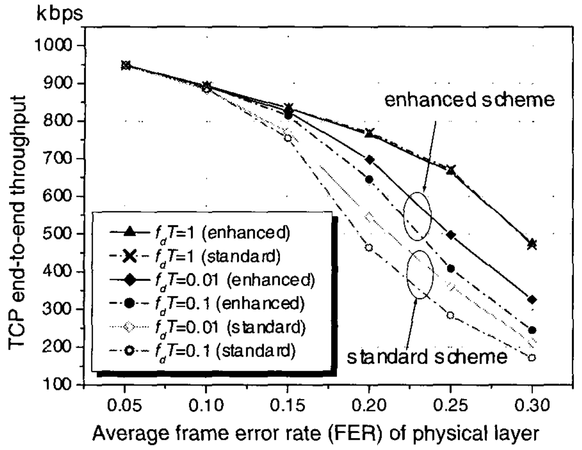


Fig. 13. Enhancement of the end-to-end TCP throughput over fading channels.

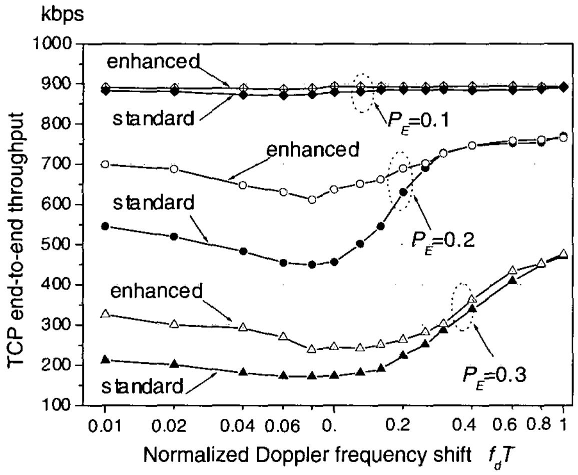


Fig. 14. TCP throughput vs. different normalized Doppler frequency shift.

3) The normalized throughput of link layer is determined by the residual error rate of the NAK rounds as shown in (17).

The numerical results of the proposed improved link retransmission scheme as well as the simulation results are presented in Figs. 11 and 12. To exhibit the performance improvement of the residual error probability of the proposed scheme, the curves of the standard multi-copy SR ARQ scheme are also plotted for comparison. The results show that this simple modification in retransmission strategy efficiently decreases the residual error probability of the link local recovery especially in high correlated channels. Simulation results in Figs. 13 and 14 show the enhancement of end-to-end TCP throughput caused by the link layer optimization over correlated channels. As the performance improvement results from the channel correlation in the channel condition with $f_dT = 1$ where it is very close to the random loss situation, the TCP throughput comes back to be almost the same with that of the TCP/ARQ system with standard radio link protocols. The analysis and results lead us to conclude that an optimized design of link layer could efficiently enhance the end-to-end performance of user level without interference with the higher layer.

Since our solution only modifies the triggering method of the RLP retransmission, it is independent with the retransmission configuration. In [10] and [11], an improved RLP retransmission configuration [1 1 1 1 1 1] has been proposed to obtain better performance in the highly correlated channels instead of the standard retransmission configuration [1 2 3]. We should point out that our proposed solution is orthogonal with the retransmission configuration enhancement in [10] and [11]. Furthermore, they could complement each other to achieve better performance. Fig. 15 presents the TCP performance advantage of the RLP retransmission configuration [1 1 1 1 1 1] over that of the standard retransmission configuration [1 2 3] in the highly correlated channels. The results show that when our solution is employed on the new retransmission configuration [1 1 1 1 1 1], further improvement of about 20% TCP throughput can be obtained.

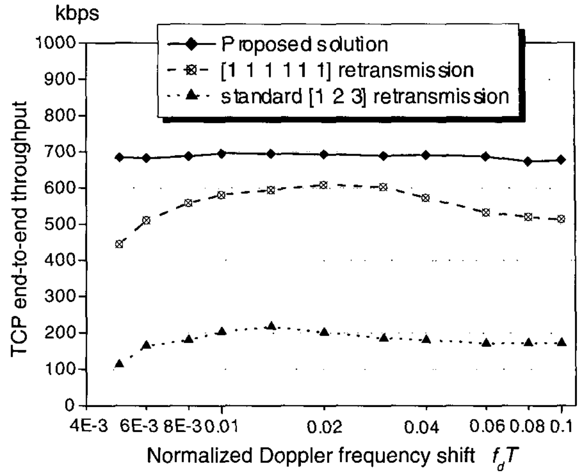


Fig. 15. TCP throughput comparison among different RLP schemes.

Although our improved strategy focuses on the information integration of link layer and physical layer, it is based on the cross-layer analysis of all the three layers of wireless networks. Note that there are three distinct advantages of the proposed scheme: Simplicity, practicality, and compatibility. Depending on the theoretical analysis, the algorithm keeps limited modification on the standard protocols only at the receiving side. There is no need to change the program of base stations and only an additional module at the receiver (e.g., mobile phones) is necessary. Also it is compatible and supplementary to be with the other high-layer solutions of wireless TCP improvement such as explicit congestion notification (ECN) [21], explicit loss notification (ELN) [22], TCP SACK [23], etc.

## VII. CONCLUSION

In this paper, we have addressed the cross-layer issue of the wireless TCP systems over two-state Markov channels. The interactions of wireless fading channels, link local recovery process and TCP are investigated. The analytical model is suggested to be helpful for performance evaluation and configuration optimization of radio link layer in wireless networks. Results show that the burstiness of the error process at physical layer caused by wireless fading significantly affects the system performance. Although the data throughput efficiency of link layer benefits from the channel correlation, the degradation of the residual error probability after local recovery due to the error burstiness dominates the behavior of link layer and finally degrades the end-to-end TCP throughput. Based on the analysis of link layer behavior and the interactions of the adjacent layers, an improved strategy of link layer is proposed to enhance the TCP/ARQ system performance by better utilization of the burstiness nature of error process. The research shows that an optimization of link layer setting with the advantages of simplicity, practicality and compatibility is significant and necessary, which is highly recommended. Conclusively, improvement of the end-to-end user performance is the final goal of our cross-layer issue, and therefore, we suggest that any evaluation or refinement of the wireless networks should depend on the viewpoint of all the network protocol stacks.

## ACKNOWLEDGMENTS

The authors would like to thank Mr. Tamio Saito of Fujitsu Research and Development Center, China, for his continuous encouragement and helpful advice. Also authors thank for valuable comments of the reviewers that greatly improve the presentation of this paper.

## REFERENCES

[1] H. Honkasalo *et al.*, "WCDMA and WLAN for 3G and beyond," *IEEE Wireless Commun.*, pp. 14–18, Apr. 2002.

[2] TIA/EIA/IS-707-A-2, "Data services options for spread spectrum systems: Radio link protocol Type 3," Addendum no. 2, Washington: Telecommunication Industry Association, Mar. 2001.

[3] M. Zorzi, A. Chockalingam, and R. R. Rao, "Throughput analysis of TCP on channels with memory," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1289–1300, July 2000.

[4] F. Khafizov and M. Yavuz, "Running TCP over IS-2000," in *Proc. IEEE ICC 2002*, vol. 5, May 2002, pp. 3444–3448.

[5] Y. Bai *et al.*, "Performance of TCP/IP over IS-2000 based CDMA radio links," in *Proc. IEEE VTC 2000* Fall, vol. 3, Sept. 2000, pp. 1036–1040.

[6] A.-F. Canton and T. Chahed, "End-to-end reliability in UMTS: TCP over ARQ," in *Proc. IEEE GLOBECOM 2001*, vol. 6, Nov. 2001, pp. 3473–3477.

[7] T. Chahed, A.-F. Canton, and S.-E. Elayoubi, "End-to-end TCP performance in W-CDMA/UMTS," in *Proc. IEEE ICC 2003*, May 2003.

[8] F. Khafizov and M. Yavuz, "Analytical model of RLP in IS-2000 CDMA networks," in *Proc. IEEE VTC 2002 Fall*, vol. 1, Oct. 2002, pp. 487–491.

[9] 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.

[10] F. Khan *et al.*, "TCP performance over cdma2000 RLP," in *Proc. IEEE VTC 2000 Spring*, vol. 1, May 2000, pp. 41–45.

[11] H. Lin and S. K. Das, "Performance study of TCP/RLP/MAC in next generation CDMA systems," in *Proc. IEEE PIMRC 2003*, Sept. 2003.

[12] F. Vacirca, A. De Vendictis, and A. Baiocchi, "Investigating interactions between ARQ mechanisms and TCP over wireless links," in *Proc. European Wireless*, Barcelona, Spain, Feb. 2004.

[13] M. Zorzi, R. R. Rao, and L. B. Milstein, "On the accuracy of a first-order Markov model for data transmission on fading channels," in *Proc. IEEE ICUPC'95*, Nov. 1995, pp. 211–215.

[14] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, New York: Wiley, 2000.

[15] W. Stevens, "TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms," *RFC 2001*, Jan. 1997.

[16] J. Padhye *et al.*, "Modeling TCP Reno performance: A simple model and its empirical validation," *IEEE/ACM Trans. Networking*, vol. 8, pp. 133–145, Apr. 2000.

[17] Z. Kostic, X. Qiu, and L. F. Chang, "Interactions between TCP and RLP protocols in a cellular system," in *Proc. IEEE VTC 2001*, May 2001, pp. 2244–2248.

[18] A. DeSimone, M. C. Chuah, and O. C. Yue, "Throughput performance of transport-layer protocols over wireless LANs," in *Proc. IEEE Globecom'93*, Nov. 1993, pp. 542–549.

[19] UCB/LBNL/VINT Network Simulator - ns (version 2), available at http://www.isi.edu/nsnam/.

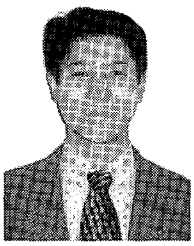
[20] G. Xylomenos and G. C. Polyzos, "Wireless link layer enhancements for TCP and UDP applications," in *Proc. IEEE IPDPS 2003*, 2003.

[21] R. Ramani and A. Karandikar, "Explicit congestion notification (ECN) in TCP over wireless network," in *Proc. IEEE Int. Conf. Pers. Wireless Commun. 2000*, 2000, pp. 495–499.

[22] H. Balakrishnan and R. H. Katz, "Explicit loss notification and wireless web performance," in *Proc. IEEE GLOBALCOM'98* Internet Mini-Conference, Nov. 1998.

[23] K. Fall and S. Floyd, "Simulation-based comparisons of Tahoe, Reno, and Sack TCP," *Computer Commun. Review*, July 1996.

**Yi Wu** received his B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, P. R. China, in 2000 and 2002, respectively. Currently, he is pursuing his Ph.D. degree in the Department of Electronic Engineering at Tsinghua University, Beijing, P. R. China. He is an IEEE student member. His major research interests include wireless TCP, cross-layer optimizations, and QoS issues in wireless multimedia communication systems.



**Zhisheng Niu** graduated from Northern Jiaotong University, Beijing, China, in 1985, and got his M.E. and D.E. degrees from Toyohashi University of Technology, Toyohashi, Japan, in 1989 and 1992, respectively. In 1994, he joined with Tsinghua University, Beijing, China, where he is now a full professor at the Department of Electronic Engineering. He is also an adjunction professor of Beijing Jiaotong University (formally Northern Jiaotong University). From April 1992 to March 1994, he was with Fujitsu Laboratories Ltd., Kawasaki, Japan. From October 1995 to February 1996, he was a visiting research fellow of the Communications Research Laboratory of the Ministry of Posts and Telecommunications of Japan. From February 1997 to February 1998, he was a visiting senior researcher of Central Research Laboratory, Hitachi Ltd. He also visited Saga University, Japan, and Polytechnic University, USA, in 2001 and 2002, respectively. He received the PAACS Friendship Award from the Institute of Electronics, Information, and Communication Engineers (IEICE) of Japan in 1991 and the Best Paper Award (1st prize) from the 6th Chinese Youth Conference on Communication Technology in 1999. He also received the Japanese Government Research Awards for Foreign Specialists from Science and Technology Agency (STA) of Japan and the Award for the Telecommunications Advancement Research Fellowship from Telecommunication Advancement Organization (TAO) of Japan in 1995. His current research interests include teletraffic theory, performance evaluation of high-speed broadband integrated networks, wireless ATM/IP networks, mobile Internet, and Stratospheric Communication Systems. He has published more than 100 journal and conference papers so far. Dr. Niu is a senior member of the IEEE and the Chair of Membership Development Committee of Asia-Pacific Board of IEEE Communication Society. He is also serving as the Chair of the Steering Committee of the Asia-Pacific Conference on Communication (APCC), and the Technical Program Chair of APCC2004. He is a senior member of the Chinese Institute of Electronics (CIE), a member of the IEICE, and the Beijing Representative of IEICE.

**Junli Zheng** graduated in Tsinghua University, Beijing, China, in 1961. Since 1961, he has been with the Department of Electronic Engineering, Tsinghua University, Beijing, China, where he now serves as a Professor. He is a member of China Institute of Circuits & Systems, and China Neural Network Committee. His current research interests include signal processing for communications, neural network, chaos theory, and their applications.