©1999 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The definitive version of this paper has been published in IEEE Wireless Communications and Networking Conference, 1999, 10.1109\/wcnc.1999.796713.

Teletraffic Models and Planning in Wireless IP Networks

Phuoc Tran-Gia and Kenji Leibnitz
University of Würzburg, Dept. of Computer Science
Am Hubland, 97074 Würzburg
Germany

Abstract—**In this paper we discuss teletraffic issues on transmitting IP traffic over a wireless connection. Due to the different characteristics of the transmission over wireless and wired networks concerning error probabilities it is necessary to consider additional measures to mitigate frame errors either by modifying the Transport Control Protocol (TCP) or introducing an error correcting Radio Link Layer (RLL). We will analyze a Radio Link Protocol operating with a Selective Repeat Automatic Repeat Request (SR-ARQ) and describe how different types of users operate with different activity factors and their impact on system capacity.**

I. Introduction

Over many years TCP/IP has been the most important protocol for computer networks. It's importance has even become greater due to the enormous growth of Internet traffic. The demand for data on the World Wide Web (WWW) or Electronic Mail (E-Mail) is further expected to increase over the next decade. It has become a necessity that a user can have access to such applications from anywhere he wants just like using a wireless phone.

While in principle TCP should work anywhere regardless of the underlying network architecture, it has been optimized for operating in wired networks with low bit error rates (BER) of approximately 10^{-8}. For each transmitted TCP segment a timer is set and when a timeout occurs, network congestion is assumed. In order to allow the network to recover from this congestion, TCP reduces its window size and thus the transmission rate. Problems occur when part of the transmission goes over a wireless link. Here, the BER is much higher and can be between 10^{-2} and 10^{-4}, which leads to lost packets being the major cause for timeouts instead of network congestion. However, TCP misinterprets these packet corruptions as congestion and reduces the window size. Additionally, due to the slow start algorithm the rate at which the window is increased will be restricted. All of this leads to a decrease in end-to-end throughput and latency since the window sizes will never be optimal.

In order to increase TCP performance for a heterogeneous wired and wireless environment, several approaches have been introduced. In [1] a method described as *Indirect-TCP* is presented that splits the TCP connection at the base station (BS) into a wired and wireless part. Errors occurring over the wireless link will be handled by the base station and have no impact on the window size of the wired TCP link. While this method improves the performance of TCP it also opposes to its concept of end-to-end flow control. A different approach is taken in [2]. Here a *snooping agent* is installed at the base station which caches the received TCP segments and autonomously handles the retransmission of those lost over the wireless link. This agent operates completely transparent to TCP and the influence of packet errors on the congestion control mechanism can be avoided. In [3] two enhanced versions of TCP are considered. In the *Last-Hop Acknowledgement* scheme (LHACK) the BS sends LHACK signals to the source and together with an end-to-end acknowledgement the source is able to distinguish between a network congestion and a packet corruption. This method leads to an increased network load, since two acknowledgements (ACK) are sent for each message. In [3] it is also shown that a better approach would be to distinguish packet errors by using a negative acknowledgement (NACK) in the TCP header option field for *explicit loss notification*. Another promising TCP enhancement would be to use a *Selective Acknowledgement TCP* [4] which shows an overall better performance when recovering from multiple packet losses in one transmission window than other TCP variants.

A completely different approach is to introduce a link layer protocol employing either *Forward Error Correction* (FEC) or more commonly *Automatic Repeat Request* (ARQ) or a combination of both. The advantage here is to operate in a layered structure of network protocols independently of the transport layer protocol without any further modification to TCP. The protocol stack for a code division multiple access (CDMA) cellular system is depicted in Fig. 1, cf. [5]. The *Radio Link Protocol* (RLP) given here operates with a NAK-based SR-ARQ mechanism and operates seamlessly in the layer above the cellular IS-95 [6] physical layer. In this paper we will analyze

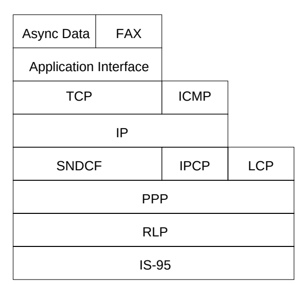

Fig. 1. Wireless CDMA protocol stack

the performance of the SR-ARQ in the RLP. This permits us to dimension the buffers on the forward and reverse link in order to maintain a desired Quality of Service (QoS). We will also see how the system parameters will influence the activity time of data bursts (which we will refer to as packet calls) and show the impacts that data users will have on an integrated voice/data cell.

The paper is organized as follows. In Section II we will describe in detail the SR-ARQ mechanism implemented in the current IS-707 Radio Link Protocol. Based on this description, we will investigate how RLP influences the transmission time of IP packets and thus the activity level of the data connection. The impacts on capacity of a CDMA cell will be shown in Section III. The paper is concluded in Section IV by giving an outlook on future work.

II. The IS-707 Radio Link Protocol (RLP)

In the IS-707 standard on data transmission the RLP is defined as a NAK-based Selective Repeat ARQ scheme. Its purpose is to achieve an acceptable frame error performance for the IP packets that will be transferred to higher layers. RLP uses the CDMA 20 *ms* frame structure. Since IS-95 was originally designed for voice transmission, it includes a variable bit rate vocoder, which produces frames with variable lengths. For data transmission this can be utilized by filling the space of voice packets which are not transmitted at full rate. Therefore, a CDMA frame can be used for voice, data and mixed voice/data operation. This permits a subdivision in primary traffic which is usually voice and secondary traffic (usually data). It should be noted that at full rate the payload of a frame is 171 *bit* which requires that an IP packet, which usually consists of several hundred bytes, is split up into many frames. Furthermore, it can be assumed that power control will maintain a nearly constant FER of approximately 1%.

For our analysis we will assume a user source model which corresponds to the one given in the proposal for the next generation CDMA system [7]. Two major classes of users exist: class A describes users with short messages, e.g. Telnet or E-Mail, and class B describes users performing WWW sessions. We can assume that the mobile unit will always be a client and that the server will be located in the wired network. Since WWW accesses are asymmetric, the traffic in the client-server direction will be less than the traffic being transmitted from the server to the client. Therefore, we will use class B for the description of the forward link and introduce another user category (class C) for describing the reverse link.

Users in classes B and C will behave in the following way, cf. Fig. 2. The user activates a WWW session with Poisson arrival rates λ_B. Each session will contain N_{pc} packet calls which is a geometric random variable with mean $E[N_{pc}] = 5$. The interarrival time will also be geometric with a mean of 120 *s*. Within each packet call N_p packets will be generated. The number of packets is Pareto-distributed with a mean of $E[N_p] = 25$. The IP packets will have a fixed size of 480 *bytes* on the forward link and 90 *bytes* on the reverse link. The interarrival time between two packets is geometric with a mean of 0.01 *s*. Thus, at the RLL each IP packet is split into K RLL frames, each with 171 *bit* payload.

A. Analysis of the Radio Link Protocol

Let the size of an IP packet be given as $I_F = 480$ *byte* and $I_R = 90$ *byte* for forward and reverse link, respectively. Assuming only primary channels, we can transport in each 20 *ms* RLL frame $X = 171$ *bit*. Therefore, an IP packet must be split into $K = I/X$ frames for transmission over the radio link. In our case, we have $K_F = 23$ and $K_R = 5$. We will now compute the transmission time for one IP packet over the wireless link. All IP packets are transmitted in sequence, however, if an error during the transmission occurs, the IP packets become interleaved. This again increases the duration of the transmission of a packet.

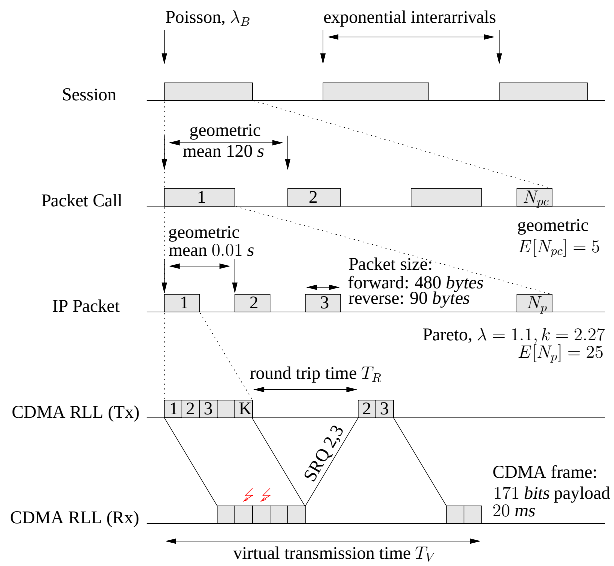

Fig. 2. CDMA user connection process

If we consider the virtual transmission time T_V of an IP packet, we need to take into account the number of round trips R and the number of frame transmissions Y, then

$$T_V = R\,T_R + Y\,T_Y.$$

We will assume that the time needed for the transmission of one frame is $T_Y = 20$ *ms* and due to the frame structure the round trip can be estimated as $T_R = 2\,T_Y$.

Each of the frames is transmitted Z times until it is successfully received depending on the frame error probability p_F in the wireless channel. We can assume that frame errors occur independently and that Z is geometrically distributed with distribution $z(i)$ and cumulative distribution function (CDF) $Z(i), i = 1, 2, \ldots$ according to

$$z(i) = p_F^{i-1}(1 - p_F) \qquad \text{and} \qquad Z(i) = 1 - p_F^i.$$

The random variable for the number of round trips R can then be expressed depending on Z as

$$R = \max\{\underbrace{Z, Z, \ldots, Z}_{K \text{ times}}\}.$$

The cumulative distribution function of R can be computed from the CDF of Z which leads to the distribution given by

$$R(i) = (1 - p_F^i)^K \qquad \text{and} \qquad r(i) = R(i) - R(i-1).$$

The number of frame transmissions Y can be computed as the K-fold sum of the random variable Z, which corresponds to the K-fold convolution of the distribution $z(i)$ with itself.

$$Y = \underbrace{Z + Z + \cdots + Z}_{K \text{ times}}$$

$$y(i) = z(i) \circledast z(i) \circledast \cdots \circledast z(i) \qquad i = 1, 2, \ldots$$

Figure 3 depicts the complementary distribution function of the virtual transmission time T_V of one IP packet. The time scale is in *ms*. We assumed that the frame error rate in this

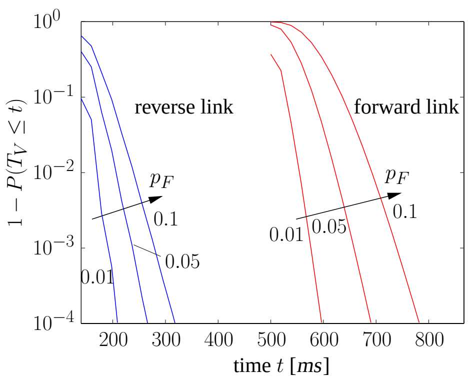

Fig. 3. Virtual transmission time for forward and reverse link

case is $p_F = 0.01, 0.05$, and 0.1 and that the number of frames required for one IP packet size is $K_R = 5$ and $K_F = 23$, which corresponds to the size of an IP packet on the reverse and forward link, respectively.

The mean and the coefficient of variation of the virtual transmission time T_V can be found in Fig. 4 and Fig. 5.

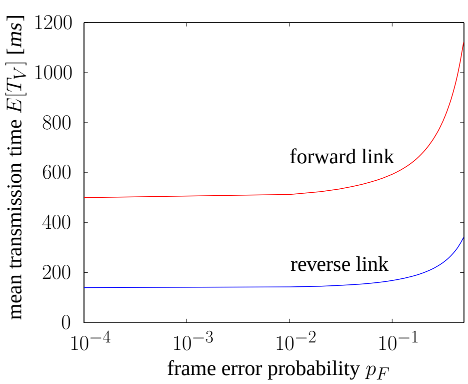

Fig. 4. Mean transmission time

B. Dimensioning Packet Buffer Size

With the knowledge of the transmission time of a single IP packet, we can dimension the buffer size needed to compensate for the high retransmission rate due to the bad link quality at an acceptable rate of blocking. For this we can assume our system to be an $M/G/n-0$ system, i.e., we have Poisson arrivals of IP packets, the service time is given by the distribution of T_V

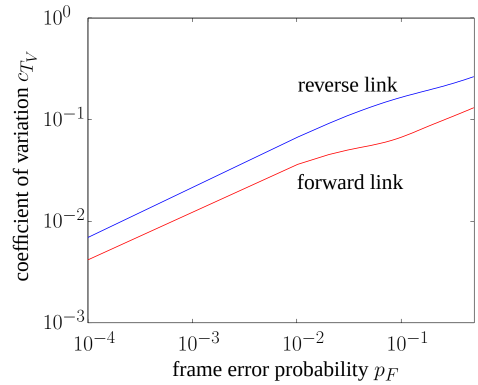

Fig. 5. Coefficient of variation of transmission time

and n is the number of buffer slots. On the forward link the mean number of packets transmitted during a 120 *s* packet call is 25. If the packet size on the forward link is on average 480 *bytes*, we have an arrival rate of:

$$\lambda_F = \frac{25 \cdot 480 \cdot 8 \textit{ bits}}{120 \textit{ s}} = 800 \textit{ bps}$$

The link utilization in this case is $\rho_F = \frac{800 \textit{ bps}}{9600 \textit{ bps}} = 8.33\%$.

On the reverse link, we can compute the same values and obtain $\lambda_R = 150$ *bps* and $\rho_R = 1.56\%$.

We can now compute the required buffer size for the forward and reverse links by using the Erlang-B formula with $a = \lambda E[T_V]$ and finding the minimum number of servers that satisfies our QoS requirements of a blocking probability $p_b < 10^{-3}$. The resulting curves are depicted in Fig. 6 for the forward and reverse link.

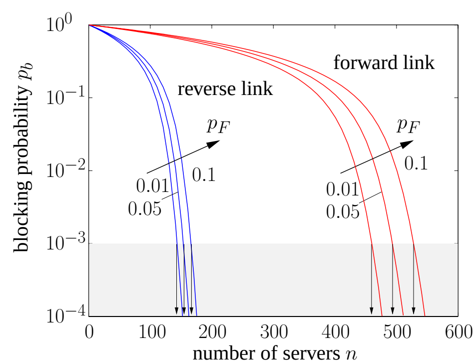

Fig. 6. Forward and reverse link buffer

III. CDMA Coverage in the Presence of IP Traffic

It is a well known fact that there is a tradeoff between coverage and capacity in a CDMA wireless network, [8], [9]. Unlike the conventional access mechanisms FDMA or TDMA, where coverage is purely determined by RF issues, due to soft capacity in CDMA also the coverage areas become elastic. Coverage and capacity are therefore extremely sensitive to the spatial customer distribution and corresponding time-dependent customer traffic intensity. In this section we will extend the work in [10] and investigate the impacts of a cell servicing both voice and data users. Both types of users will be assumed to be distributed according to a spatial Poisson process, which makes it possible to obtain a stochastic description of the quality of service (QoS) in the cell, which will be measured by the probability of outage.

A. *Voice and Data Activity Modeling*

We consider a cell currently supporting a number of k voice and data connections at an observed time instant. Let us consider that the percentage of data users is d. The time when the connection is active is being separated by idle phases. However, voice and data connections differ significantly when considering the burstiness of the user activity. Measurements of human conversations have yielded that the time when one party is active is only about 45% of the whole conversation time. To decrease the interference in these idle times, in IS-95 the transmission rate is being reduced from full rate to 1/8th rate. We can derive a voice activity factor for user j by modeling it as a Bernoulli random variable ν_j that is active with probability $\rho_v = 0.45$ and inactive with probability $1 - \rho_v$.

The activity in data connections is very asymmetric and depends on the direction of the examined link. As a client the MS will need only little bandwidth for applications as E-mail or WWW. Most traffic that will be generated is requests for data which will be transmitted on the forward link. For modeling purposes, we will assume that the data user i is active with probability $\rho_d = P(\psi_i = 1)$ using the Bernoulli random variables ψ_i. In the previous section we could see that the utilization of both links is below 10%. We therefore use $\rho_d = 0.1$ as the reverse link activity.

B. *Outage Model for a Fixed Number of Users*

The quality of a CDMA link is given by the signal-to-interference ratio (SIR). Due to fluctuations in interference induced by the other users in the cell, the SIR value becomes a random term. From [8] it is known that it can be well modeled by a log-normal random variable. Outage occurs when the SIR requirements can not be fulfilled.

Considering that the transmitted signal at the MS is attenuated by propagation loss $L(x)$ at distance x (e.g. Hata model [11]) and by shadow fading Z, the received signal S at the BTS can be given as: $S_{\text{trans}} = S + L(x) + Z$. The probability of outage can then be related to the probability of the transmitted power being larger than maximum transmit power, cf. [9] .

$$P_{\text{out}} = P(S + L(x) + Z > S_{\text{max}}) \tag{1}$$

The value of S depends on the number of users in the cell, leading to a relationship between coverage and capacity. In the following section we will extend previous work in [9], [10] to include the effects of data users on outage probability.

C. *Outage Probability for an Integrated Voice/Data System*

The probability that a user j has an acceptable link quality in a CDMA network can be measured with the signal-to-interference ratio $\hat{\epsilon}_j$ for this user. Let the cell be loaded with $(1-d)k$ voice and dk data users. The SIR for the user j can be expressed in terms of the received powers of the other users in the cell $\hat{S}_i$. Please note that for a linear power $\hat{X}$, X denotes its transformation to *dB*, i.e., $X = 10 \log \hat{X}$.

$$\hat{\epsilon}_j = \frac{\frac{\hat{S}_j}{R}}{\sum\limits_{i \neq j}^{(1-d)k} \frac{\nu_i \hat{S}_i}{W} + \sum\limits_{i \neq j}^{dk} \frac{\psi_i \hat{S}_i}{W} + N_0 + I}$$

The value $W = 1.25$ *Mhz* denotes the frequency bandwidth, $R = 9.6$ *kbps* is the data bitrate, N_0 is the background noise power density and I is the other cell interference.

We can assume that the power control mechanism is the same for both voice and data users. Therefore, the random variables $\hat{S}$ are modeled as i.i.d. random variables. From [8] it is known that the received SIR $\hat{\epsilon}$ can be approximated as log-normal random variable with mean $m_\epsilon = 7$ *dB* and standard deviation $\sigma_\epsilon = 2.5$ *dB*.

The first and second moment of the log-normal variable $\hat{\epsilon}$ can then be computed from m_ϵ and σ_ϵ by

$$m_{\hat{\epsilon}} = \exp\left(\frac{(\beta\sigma_\epsilon)^2}{2}\right)\exp(\beta m_\epsilon)$$
$$\delta_{\hat{\epsilon}} = \exp(2(\beta\sigma_\epsilon)^2)\exp(2\beta m_\epsilon)$$

where $\beta = \ln(10)/10$.

It can be shown that $\hat{S}$ is well approximated by a log-normal random variable, for which we only need to compute the first two moments:

$$m_{\hat{S}}(k,d) = \frac{W m_{\hat{\epsilon}}(N_0 + I)}{\frac{W}{R} - m_{\hat{\epsilon}}\zeta} \tag{2}$$

$$\delta_{\hat{S}}(k,d) = \frac{\delta_{\hat{\epsilon}}((W(N_0+I) + m_{\hat{S}}\zeta)^2 - m_{\hat{S}}^2\zeta)}{\left(\frac{W}{R}\right)^2 - \delta_{\hat{\epsilon}}((1-d)k\rho_v + dk\rho_d)}, \tag{3}$$

where $\zeta = (1-d)k\rho_v + dk\rho_d$. Examination of Eqn. (2) shows that k cannot exceed the value for which the denominator is zero. This value is defined as *pole capacity*. The maximum number of users that can be accepted to the system with mean SIR of $m_{\hat{\epsilon}}$ and a fraction of d data users is therefore

$$k_{\text{pole}} = \frac{W}{R(d\rho_d + (1-d)\rho_v)m_{\hat{\epsilon}}}.$$

Since $\hat{S}$ is log-normal, the variable S in *dB* is Gaussian. The mean and variance of S can be easily calculated in terms of $m_{\hat{S}}$ and $\delta_{\hat{S}}$.

$$m_S(k,d) = 20 \log m_{\hat{S}}(k,d) - 5 \log \delta_{\hat{S}}(k,d)$$
$$\sigma_S^2(k,d) = \frac{1}{\beta}(10 \log \delta_{\hat{S}}(k,d) - 20 \log m_{\hat{S}}(k,d))$$

Assuming that the variables S and Z are uncorrelated, we can hence rewrite Eqn. (1) using the properties of the Gaussian distribution.

$$P_{\text{out}}(x,k,d) = Q\left(\frac{S_{\max} - L(x) - m_S(k,d)}{\sqrt{\sigma_S^2(k,d) + \sigma_Z^2}}\right)$$

It is now possible to give an outage probability term unconditioned of the number of users k. In [10], [12], a spatial homogeneous Poisson process is used to characterize the relationship between the number and location of the users in the cell. Thus, the r.v. for the number of users distributed in a cell with area A is Poisson distributed with density λ, which translates to an expected number of users in the cell $\xi = \lambda A$. The probability to have k connections in a cell with radius x is Poisson distributed and this leads to

$$P_{\text{out}}(x,d) = \sum_{k=1}^{\infty} P_{\text{out}}(x,k,d)\,\frac{\xi^k}{k!}\,\exp(-\xi) \tag{4}$$

Fig. 7 shows the impact of the percentage d on the outage probability as function of the cell load. The distance of the observed user is fixed at $x = 2$ *km*. It can be seen that the higher the percentage of data users in the cell is, the cell capacity will become higher for the same outage probability. This is due to the fact that the activity phase of data users on the reverse link is much lower than that for voice users.

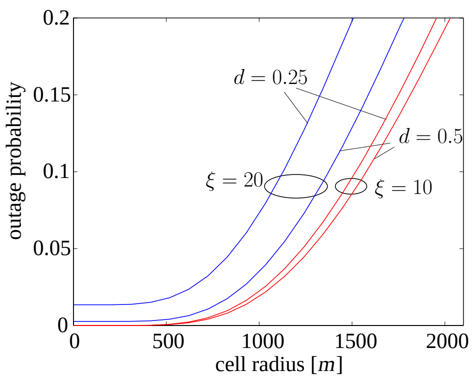

Fig. 7. The impacts of the number of data users on outage probability

The Eqn. (4) can be solved for x in order to obtain a tradeoff between coverage and capacity, cf. Fig. 8. In this case the maximum outage probability is given as $P_{\text{out}} = 0.01, 0.05$, and 0.1. The fraction of data users is here $d = 0.25$. It can be seen that the stricter the outage requirements the smaller the coverage areas get, which reach zero at k_{pole}.

IV. Conclusion and Outlook

In this paper we presented some issues concerning teletraffic modeling of wireless IP networks. The performance of the radio link protocol is influenced by the frame error rate on the radio channel by increasing the virtual transmission time of IP packets. This has an impact on the performance of the TCP connection and the buffer size needed until complete reception of IP packets. It was also shown that the different activities of voice and data users influences the capacity and coverage of the integrated system when employing CDMA technology.

Further work includes the consideration of real-time traffic and their implications on the system, e.g. power control for variable QoS (bandwidth, delay, reliability) [13], [14]. This is in view of third generation system (IMT-2000, UMTS) of special interest, where higher data rates are supported. Finally, further data protocols which have been specifically designed for wireless links have been introduced, (WAP: Wireless Application Protocol). The performance of such systems has yet to be investigated.

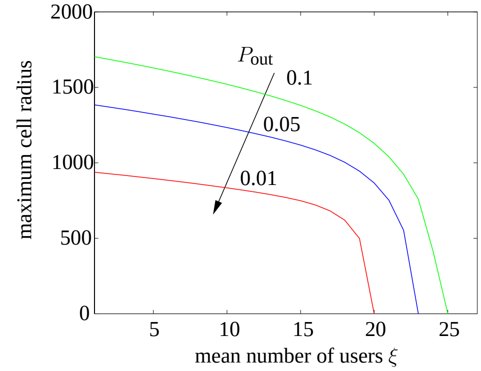

Fig. 8. Coverage and capacity tradeoff for varying data users

References

[1] A. Bakre and B.R. Badrinath, "I-TCP: Indirect TCP for mobile hosts," in *Proc. 15th Int. Conf. on Distributed Computing Systems*, May 1995.

[2] H. Balakrishnan, V.N. Padmanabhan, S. Seshan, and R.H. Katz, "A comparison of mechanisms for improving TCP performance over wireless links," *IEEE/ACM Trans. on Networking*, Dec. 1997.

[3] A.C.F. Chan, D.H.K. Tsang, and S. Gupta, "TCP over wireless links," in *Proc. of IEEE VTC97*, Phoenix, AZ, May 1997.

[4] K. Fall and S. Floyd, "Simulation-based comparisons of Tahoe, Renoe, and SACK TCP," *Computer Comm. Review*, 1996.

[5] TIA/EIA/IS-707, "Data service options for wideband spread spectrum systems," Nov. 1997.

[6] TIA/EIA/IS-95, "Mobile station – base station compatability standard for dual–mode wideband spread spectrum cellular systems," July 1995.

[7] ITU, "The cdma2000 RTT candidate submssion," TR45.5, June 1998.

[8] A.M. Viterbi and A.J. Viterbi, "Erlang capacity of a power controlled CDMA system," *IEEE JSAC*, vol. 11, no. 6, Aug. 1993.

[9] V.V. Veeravalli, A. Sendonaris, and N. Jain, "CDMA coverage, capacity and pole capacity," in *Proc. of IEEE VTC97*, Phoenix, AZ, May 1997.

[10] P. Tran-Gia, N. Jain, and K. Leibnitz, "Code division multiple access wireless network planning considering clustered spatial customer traffic," in *Proc. of NETWORKS98*, Sorrento, Italy, Oct. 1998.

[11] T.S. Rappaport, *Wireless Communications – Principles & Practice*, Prentice Hall, Upper Saddle River, NJ, 1996.

[12] U. Ehrenberger and K. Leibnitz, "Impact of clustered traffic distributions in cdma radio network planning," in *Proc. of the 16th International Teletraffic Congress*, Edinburgh, UK, June 1999.

[13] L.C. Yun and D.G. Messerschmitt, "Power control for variable QoS on a CDMA channel," in *Proc. IEEE MILCOM*, Fort Monmouth, NJ, Oct. 1994.

[14] M. Naghshineh and A.S. Acampora, "Design and control of micro-cellular networks with QoS provisioning for real-time traffic," in *Proc. of IEEE ICUPC94*, San Diego, CA, Sep. 1994.