Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2006, Article ID 62657, Pages 1–14
DOI 10.1155/WCN/2006/62657

Cross-Layer Quality-of-Service Analysis and Call Admission Control in the Uplink of CDMA Cellular Networks

Chun Nie,[1,2] Yong Huat Chew,[1] and David Tung Chong Wong[1]

[1] *Institute for Infocomm Research, Agency for Science, Technology, and Research, Singapore 119613*
[2] *Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576*

Received 26 September 2005; Revised 16 March 2006; Accepted 26 May 2006

This paper addresses cross-layer quality-of-service (QoS) provisioning in the uplink of CDMA cellular mobile networks. Each mobile can take up to four UMTS traffic classes in our model. At the data link layer and the network layer, the QoS performances are defined in terms of signal-to-interference-plus-noise ratio and outage probability, and packet loss rate and delay, respectively. A call admission control scheme which fulfills these QoS metrics is developed to maximize the system capacity. The novelty of this paper is that the effect of the lengthening of the on-periods of non-real-time traffic classes is investigated by using the Go-Back-N automatic retransmission request mechanism with finite buffer size and limited number of retransmissions in the event of transmission errors. Simulation results for a specific example demonstrate the reasonableness of the analytical formulation.

1. INTRODUCTION

The currently deployed universal mobile telecommunications system (UMTS) network is characterized by its ability to support multimedia communications with different bit rates and quality-of-service (QoS) requirements. Four traffic classes, conversational, streaming, interactive, and background, are defined in the UMTS QoS architecture together with their respective QoS requirements [1]. Code division multiple access (CDMA) is the multiple access technology used to support the transmissions of multiclass services. In this paper, voice, video, web-browsing, and data are used as typical applications of these four traffic classes. Their QoS performances in the uplink are investigated and their QoS metrics are formulated at both the data link layer and the packet level of the network layer.

In the literature, QoS provisioning in CDMA networks has attracted a lot of research interests. At the data link layer, Gilhousen et al. [2] studied the outage probability for a single class on/off source in CDMA networks. Wong et al. [3–5] extended the analysis of outage probability from a single class sources to on/off multiclass sources, variable bit rate multiclass sources, and video multiclass sources. Recently, the outage probabilities of multiclass multiconnection services are investigated in [6]. At the network layer, packet loss rate and delay performances are studied for CDMA systems [7, 8]. However, [7, 8] do not provide an analytical platform which can be directly applied to the QoS provisioning of practical systems. For example, only voice and data services in single-cell systems are considered in [7] and traffic sources are simply modeled as exponential-on/exponential-off and Poisson arrivals. Reference [8] investigated packet loss rate and delay performances in CDMA networks for voice, video, and data services. However, analytical QoS formulation is given only for voice services, while video and data services are only obtained through computer simulations.

The main contribution of this paper is an analytical formulation for the QoS performances of all of the four traffic classes jointly at both the data link and network layers. We adopt more realistic traffic models for both real-time (RT) and non-real-time (NRT) traffic than those in the literature. The effect of the lengthening of the on-periods of the NRT services is analyzed under Go-Back-N (GBN) automatic retransmission request (ARQ) scheme. The QoS attributes are formulated in terms of the signal-to-interference-plus-noise ratio (SINR) and outage probability at the data link layer, and the average delay and packet loss rate at the network layer. A QoS-based call admission control (CAC) scheme is also proposed. The maximum system capacity satisfying all QoS requirements at both the data link and network layers is computed analytically.

The subsequent sections of this paper are organized as follows. Section 2 develops a system model that describes a cellular mobile network and establishes appropriate traffic models for the four traffic classes. In Section 3, an efficient power control method is designed and the outage probabilities at the data link layer are formulated accordingly. Section 4 deals with the packet level QoS performances. Section 5 presents analytical and simulation results to verify the reasonableness of the analysis. Section 6 develops a CAC scheme with cross-layer QoS satisfactions. Finally, Section 7 concludes this paper.

2. SYSTEM MODEL

A cellular mobile system with multiple square cells is considered. This model is commonly adopted and referred to as the Manhattan model [9]. A base station (BS) is located at the center of each cell to serve a number of mobiles. Each mobile supports multiconnection to transmit multiclass services. The type of traffic classes is denoted by an index k, where $k = 1$ for voice, $k = 2$ for video, $k = 3$ for web-browsing, and $k = 4$ for data, respectively. In order to evaluate the QoS performances, appropriate traffic models are defined. Voice and video are, respectively, modeled as an exponential-on/exponential-off process and a two-dimensional discrete-state continuous-time Markov chain, as shown in Figures 1(a) and 1(b).

In Figure 1(a), a voice service is modeled as a two-state on/off birth-death process. In Figure 1(b), a video service ($k = 2$) is a variable bit rate source and is described by the Sen's model [10]. Each video service can be decomposed into one high-bit-rate (HBR) and M low-bit-rate (LBR) minisources. Hereafter, one HBR mini-source ($k = 2h$) and M LBR minisources ($k = 2l$) will be used to replace a video service. The activity factors, which are the probabilities that the process stays in the on state, for the voice, LBR video, and HBR video are, respectively, given by

$$p_k = \frac{\alpha_k}{\alpha_k + \beta_k}, \quad k \in \{1, 2l, 2h\}, \tag{1}$$

where $1/\beta_k$ and $1/\alpha_k$ are, respectively, the average on and off periods, and $k = 1$ for voice, $k = 2l$ for LBR video minisources, and $k = 2h$ for HBR video minisources, respectively.

The source traffic of web-browsing and data services are more accurately modeled as a Pareto-on/Pareto-off process [11]. Let us denote the on and off periods of web-browsing and data by $t_{\text{on},k}$ and $t_{\text{off},k}$, $k \in \{3, 4\}$, respectively. The probability density functions (pdf) of $t_{\text{on},k}$ and $t_{\text{off},k}$, $k \in \{3, 4\}$, denoted by $u_k(t_{\text{on},k})$ and $v_k(t_{\text{off},k})$, $k \in \{3, 4\}$, respectively, are given by [12]

$$u_k(t_{\text{on},k}) = c_{\text{on},k} {a_{\text{on},k}}^{c_{\text{on},k}} {t_{\text{on},k}}^{-c_{\text{on},k}-1}, \quad t_{\text{on},k} \geq a_{\text{on},k}, \tag{2}$$

$$v_k(t_{\text{off},k}) = c_{\text{off},k} {a_{\text{off},k}}^{c_{\text{off},k}} {t_{\text{off},k}}^{-c_{\text{off},k}-1}, \quad t_{\text{off},k} \geq a_{\text{off},k}. \tag{3}$$

In (2) and (3), $c_{\text{on},k}$ and $c_{\text{off},k}$ represent the shape parameters of the on and off periods, while $a_{\text{on},k}$ and $a_{\text{off},k}$ represent the corresponding location parameters for web-browsing ($k = 3$) and data ($k = 4$) services, respectively. The location and shape parameters are defined in [12].

For a Pareto-on/Pareto-off process, the activity factors of web-browsing and data traffic at their source can still be approximately defined by $p_k, k \in \{3, 4\}$, as

$$p_k = \frac{\bar{t}_{\text{on},k}}{\bar{t}_{\text{on},k} + \bar{t}_{\text{off},k}}, \quad k \in \{3, 4\}, \tag{4}$$

where $\bar{t}_{\text{on},k}$ and $\bar{t}_{\text{off},k}$ are the means of $t_{\text{on},k}$ and $t_{\text{off},k}$, respectively. The reasonableness of this assumption is verified through simulations in [13], at least for these parameters whose ranges are around the values specified in the 3GPP specification [1].

The assumptions and system parameters used are listed as follow.

(i) There exist N mobiles in each cell and they are uniformly located in the cell.

(ii) The area of a cell is denoted by A and the cellular network comprises of n square cells.

(iii) $n_{i,k}$ denotes the number of voice, video, web-browsing, and data streams of the ith ($1 \leq i \leq N$) mobile, for $k \in \{1, 2, 3, 4\}$, respectively.

(iv) G_k, γ_k^*, and BER_k^*, $k \in \{1, 2l, 2h, 3, 4\}$, denote the spreading gains, SINR, and bit-error-rate (BER) requirements for voice, LBR video, HBR video, web-browsing, and data services, respectively.

(v) $S_{i,k}$ and $l_{i,k}$, $k \in \{1, 2l, 2h, 3, 4\}$, $1 \leq i \leq N$, denote the received power and total number of active spreading codes used by voice, LBR video, HBR video, web-browsing, and data services of the ith mobile, respectively.

(vi) Perfect power control is implemented for each service/minisource to ensure that the desired received powers are achieved at the intracell BS.

(vii) All receivers have additive white Gaussian noise (AWGN) with power η.

(viii) $I_{\text{intercell}}$ is the intercell interference from all neighboring cells.

(ix) GBN ARQ has limited number of retransmissions for web-browsing and data services.

(x) Web-browsing and data services are equipped with finite buffer of buffer sizes B_3 and B_4, respectively, both in unit of packets.

Voice and video services carry RT traffic and thus are not very relevant to implement ARQ mechanism. Comparatively, web-browsing and data services carry NRT traffic and thus can initiate the GBN ARQ scheme in case of packet errors. Since GBN ARQ is a continuous retransmission scheme, web-browsing/data traffic observed in the channel is still an on/off process except that the on-period observed in the channels is lengthened as a result of retransmissions. This results in larger activity factors being observed in the channels than those in the sources.

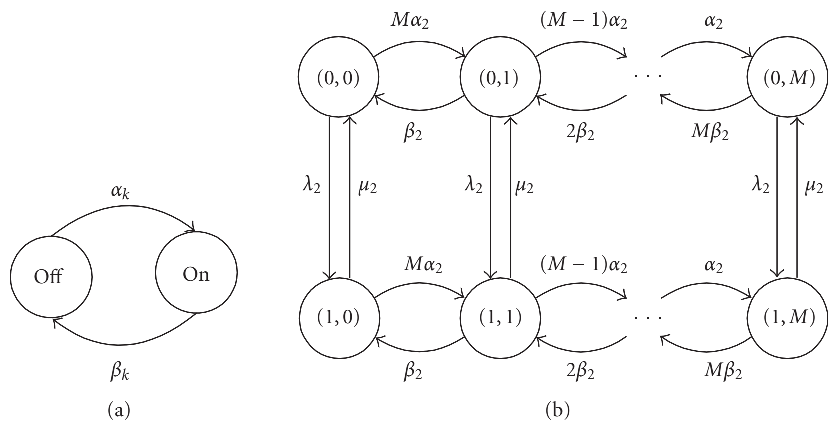


FIGURE 1: Traffic models: (a) 2-state Markov chain for a voice source, (b) 2-dimensional Markov chain for a video source.

Since each mobile experiences different amount of interference and retransmissions, the lengthened activity factors of each mobile can be different even for the same class of service. Let us denote the average on and off periods of web-browsing and data services in the CDMA channel as $\bar{t}_{\text{on},k,c}$ and $\bar{t}_{\text{off},k,c}$, $k \in \{3,4\}$, respectively, where the subscript c is used to represent the channel, obviously, $\bar{t}_{\text{on},k,c} > \bar{t}_{\text{on},k}$ and $\bar{t}_{\text{off},k,c} < \bar{t}_{\text{off},k}$. Let $p_{i,k,c}$, $1 \le i \le N$, $k \in \{1, 2l, 2h, 3, 4\}$, denote the lengthened activity factors of voice, LBR video, HBR video, web-browsing, and data services of the ith user in the channel, respectively. $p_{i,k,c} = p_k$ for $k = 1, 2l, 2h$ as there is no retransmission scheme and $p_{i,k,c} > p_k$ for $k = 3, 4$ as these services use GBN ARQ scheme.

3. POWER CONTROL ALGORITHM AND QoS ANALYSIS AT DATA LINK LAYER

System capacity and QoS performance metrics in CDMA networks are associated with the multiple access interference (MAI) contributed from the interfering mobiles. MAI includes both intracell and intercell interference resulting from mobiles within and outside the reference cell. SINR is a function of the received powers, spreading gains and number of active spreading codes, and is an important attribute at the data link layer. It is necessary that the average SINR of each service should be maintained at a required level. Denote set $\mathbf{V}$ as $\{1, 2l, 2h, 3, 4\}$, $\mathbf{V}'$ as $\{1, 2h, 3, 4\}$, $n_{i,2l} = Mn_{i,2}$, and $n_{i,2h} = n_{i,2}$ $(1 \le i \le N)$ for the ith mobile, the average SINR of the kth service stream can be expressed as [6]

$$\frac{S_{i,k}G_k}{\left\{\sum_{j=1;j\neq i}^{N}\sum_{k\in\mathbf{V}} p_{i,k,c}n_{i,k}S_{i,k} + \bar{I}_{\text{intercell}} + \eta\right\}} = \gamma_k^*, \qquad (5)$$

where $k \in \mathbf{V}$ and $i \in \{1, 2, \ldots, N\}$. $\bar{I}_{\text{intercell}}$ denotes the mean of the intercell interference. Our path loss model includes only path attenuation and lognormal shadowing which has been widely adopted [2–6]. Rayleigh and Ricean fading are ignored. The total intercell interference is approximated by a Gaussian distribution if the number of mobiles is sufficiently large [2–6], with mean and variance given by

$$\bar{I}_{\text{intercell}} \le \left[\sum_{i=1}^{N}\sum_{k\in\mathbf{V}} p_{i,k,c}n_{i,k}S_{i,k}\right]\iint f\left(\frac{r_m}{r_d}\right)\frac{dA}{A},$$

$$\begin{aligned}\operatorname{Var}\left[I_{\text{intercell}}\right] &\le \sum_{i=1}^{N}\left\{\sum_{k\in\mathbf{V}'} S_{i,k}^2 n_{i,k}\iint\left[p_{i,k,c}\,g\left(\frac{r_m}{r_d}\right) - p_{i,k,c}^2 f^2\left(\frac{r_m}{r_d}\right)\right]\frac{dA}{A}\right.\\ &\quad + S_{i,2l}^2 n_{i,2}\iint\left[Mp_{i,2l,c}\left[1 + (M-1)p_{i,2l,c}\right]\right.\\ &\quad \left.\left.\times g\left(\frac{r_m}{r_d}\right) - \left(Mp_{i,2l,c}\right)^2 f^2\left(\frac{r_m}{r_d}\right)\right]\frac{dA}{A}\right\},\end{aligned} \qquad (6)$$

where

$$\begin{aligned}f\left(\frac{r_m}{r_d}\right) &= \left(\frac{r_m}{r_d}\right)^4 e^{(\sigma\ln 10/10)^2}\left[1 - Q\left(\frac{40\log\left(r_m/r_d\right)}{\sqrt{2\sigma^2}} - \sqrt{2\sigma^2}\frac{\ln 10}{10}\right)\right],\\ g\left(\frac{r_m}{r_d}\right) &= \left(\frac{r_m}{r_d}\right)^8 e^{(\sigma\ln 10/5)^2}\left[1 - Q\left(\frac{40\log\left(r_m/r_d\right)}{\sqrt{2\sigma^2}} - \sqrt{2\sigma^2}\frac{\ln 10}{5}\right)\right].\end{aligned} \qquad (7)$$

In (7), σ^2 is variance of the lognormal shadowing, r_m and r_d denote the distance between a mobile and its own intracell BS, and the distance between the mobile and the intercell BS, respectively. Let

$$\begin{aligned}\Gamma_i &= \frac{\sum_{k\in\mathbf{V}} p_{i,k,c}n_{i,k}\gamma_k^*}{G_k},\\ \epsilon &= \frac{1 - \sum_{i=1}^{N}\Gamma_i\left[1 + \iint f\left(r_m/r_d\right)dA/A\right]}{\left(1 + \Gamma_i\right)}.\end{aligned} \qquad (8)$$

According to the formulation that is presented in [6], the following power level is derived:

$$S_{i,j} = \frac{\eta\gamma_j^*}{[\epsilon(1+\Gamma_i)G_j]}, \quad 1 \leq i \leq N, \; j = \{1, 2l, 2h, 3, 4\}. \quad (9)$$

The data link layer QoS performance is analyzed in terms of the outage probability, which refers to the probability that the achieved SINR is below the SINR requirement or the achieved BER is above the BER requirement. Within the ith mobile, the outage probabilities for voice, LBR video, HBR video, web-browsing, and data services are formulated as $P_{\text{out},i,k}$, $1 \leq i \leq N$, $k \in \{1, 2l, 2h, 3, 4\}$, and given by [6]

$$P_{\text{out},i,k} = \sum_{N}^{\rightarrow}\prod_{V}^{\rightarrow} \times Q\Big(\frac{\delta_{i,k} - \mu_i}{\sigma_i}\Big), \quad (10)$$

where $\sigma_i^2 = \text{Var}[I_{\text{intercell}}]$, $\mu_i = \sum_{j=1;j\neq i}^{N}\sum_{k\in V}(l_{j,k}S_{j,k}) + \overline{I}_{\text{intercell}}$, $\delta_{i,k} = S_{i,k}G_k/\gamma_k^* - \eta$, $Q(x) = \int_x^\infty e^{-t^2/2}dt/\sqrt{2\pi}$, and the notation

$$\begin{aligned}\sum_{\mathbf{N}}^{\rightarrow}\prod_{\mathbf{V}}^{\rightarrow} = &\sum_{l_{1,1}=0}^{n_{1,2}}\sum_{l_{1,2l}=0}^{Mn_{1,2}}\sum_{l_{1,2h}=0}^{n_{1,2}}\sum_{l_{1,3}=0}^{n_{1,3}}\sum_{l_{1,4}=0}^{n_{1,4}} \\ &\cdots\sum_{\substack{l_{j,1}=0\\ j\neq i}}^{n_{j,1}}\sum_{\substack{l_{j,2l}=0\\ j\neq i}}^{Mn_{j,2}}\sum_{\substack{l_{j,2h}=0\\ j\neq i}}^{n_{j,2}}\sum_{\substack{l_{j,3}=0\\ j\neq i}}^{n_{j,3}}\sum_{\substack{l_{j,4}=0\\ j\neq i}}^{n_{j,4}} \\ &\cdots\sum_{l_{N,1}=0}^{n_{N,1}}\sum_{l_{N,2l}=0}^{Mn_{N,2}}\sum_{l_{N,2h}=0}^{n_{N,2}}\sum_{l_{N,3}=0}^{n_{N,3}}\sum_{l_{N,4}=0}^{n_{N,4}} \\ &\times\prod_{\substack{j=1\\ j\neq i}}^{N}\prod_{k\in V}\binom{n_{j,k}}{l_{j,k}}(p_{i,k,c})^{l_{j,k}}(1-P_{i,k,c})^{n_{j,k}-l_{j,k}}.\end{aligned} \quad (11)$$

Compared to the results in [6], the main contribution here is to calculate the outage probabilities in the environment with the GBN-ARQ scheme. The computation of the lengthened activity factors will be discussed in the next section.

4. PACKET LEVEL QoS ANALYSIS AT THE NETWORK LAYER

In this section, our aim is to formulate the packet level QoS performance in the uplink of CDMA systems in terms of packet loss rates and average delays. Packet level QoS at the network layer is directly associated with the outage probability. If outage occurs, the packets are assumed erroneous due to excessive bit errors. For RT voice and video traffic, these packets are discarded and treated as packet loss. For NRT web-browsing and data traffic, GBN ARQ mechanism is implemented to retransmit the erroneous packets which also result in longer packet delays. In previous works [7, 14, 15], infinite buffer is considered and thus is not realistic. In the following, we will investigate and provide the analytical platform on the effect of a finite buffer on the packet loss rate and the average delay of a Pareto-on/Pareto-off distributed NRT traffic for CDMA systems.

4.1. *Go-Back-N ARQ*

Compared to the stop-and-wait ARQ, GBN is more efficient and easy to implement. Furthermore, it guarantees that the received packets are in sequence as compared to the selective repeat ARQ. Figure 2(a) gives a good illustration on the mechanism of GBN ARQ. At the source, the mobile has a finite buffer to accommodate the newly arrived packets. When the first and subsequent few packets arrived, they are queued in the buffer and at the same time transmitted over the channel. Upon reception, BS decodes the packet and sends an acknowledgment (ACK if correctly decoded and NACK if is in error) back to the mobile. Only if ACK is received, the mobile will remove that packet from the buffer. In case if NACK is received, both the particular packet and all its subsequent packets are retransmitted sequentially. BS will ensure that NACK is not sent for more than a given maximum number. In the process of retransmission, new packets continue to arrive and are queued in the buffer, as shown in Figure 2(b). There are two situations where packets will be lost.

(a) Since the buffer size is finite, when there are many retransmissions, buffer will overflow and newly arrived packet will be dropped.
(b) A packet has been retransmitted for the allowable maximum number of times.

Assuming that $k = \{3, 4\}$ represents web-browsing and data services, respectively, the system parameters and assumptions are defined as follows.

(1) A finite buffer with a size of B_k packets, $k \in \{3, 4\}$, is used by a sender.
(2) Each on-period contains ℓ_k packets of the same size, where the total length of the ℓ_k packets is a random variable which follows a pdf that is defined in (2) or (3). Packets are generated continuously during the on-period with a fixed time duration, T_k, $k \in \{3, 4\}$.
(3) When a packet is transmitted from a mobile to the BS, the mobile waits for an acknowledgment within a time interval of T_k'. The packet will be removed from the buffer upon the receipt of ACK. The ratio of T_k' to T_k is assumed to be an integer, s_k, (e.g., $s_k = T_k'/T_k = 2$ in the example shown in Figure 2(a)) and $B_k \geq s_k$ holds.
(4) Packet error probability is defined as $p_{e,k}$, $k \in \{3, 4\}$.
(5) ACK and NACK are always received correctly.
(6) Let the maximum number of retransmissions be $M_{\text{re},k}$, $k \in \{3, 4\}$.

Next, the following variables, which are useful for our analysis, are defined. For simplicity and ease of notations, the subscript k, which is used to differentiate between the two NRT services, will not be shown in the next few subsections. For

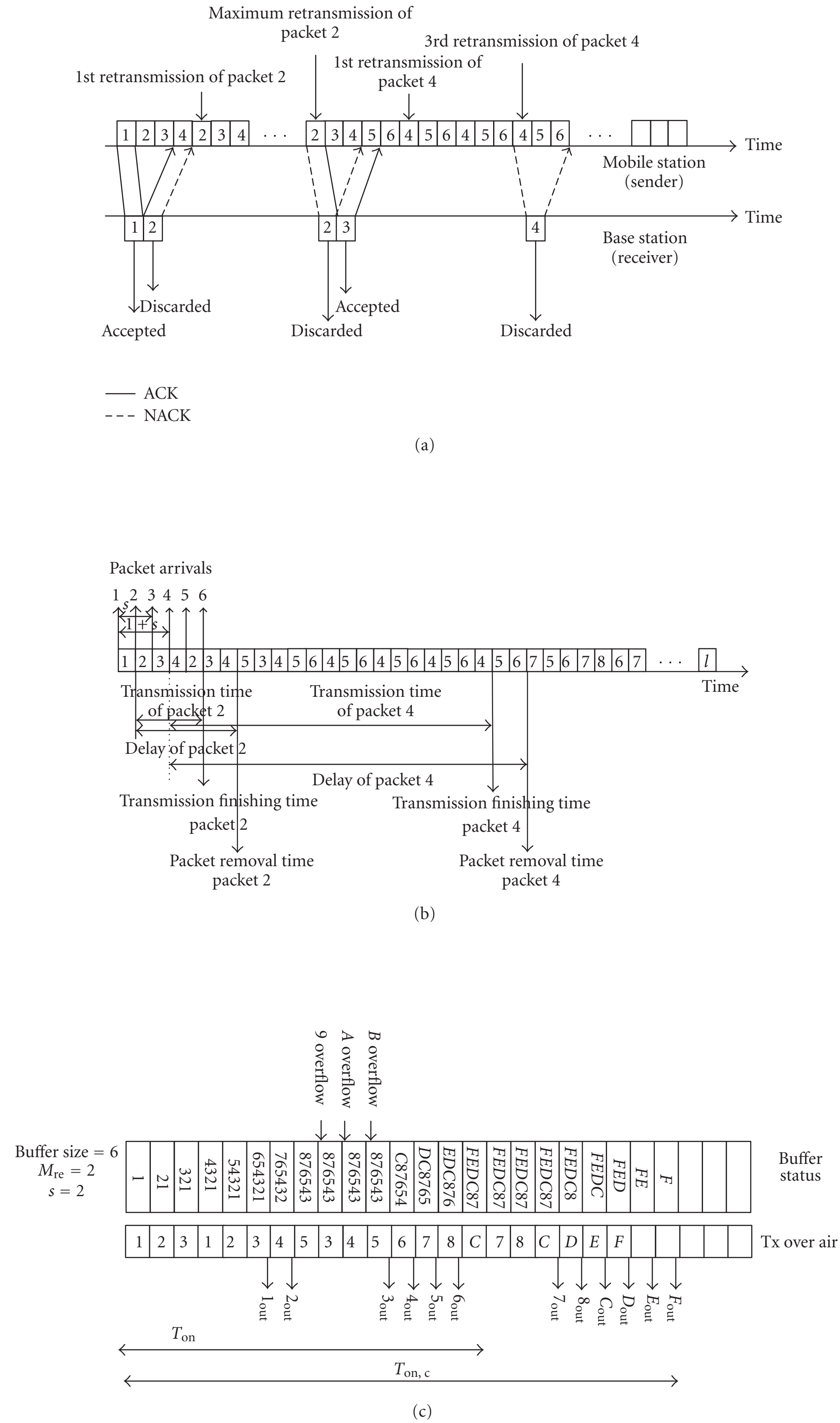


Figure 2: (a) GBN ARQ mechanism, (b) definition of packet transmission and removal time in Go-Back-N ARQ, and (c) lengthening effect of the on-period: an example.

example, M_{re} would mean $M_{\mathrm{re},k}$, T would mean T_k, and so on.

(1) v is the index used to represent packet sequence appearing in the source, $v = 1, \ldots, \ell$.
(2) $t_{\mathrm{in},v}$ denotes the initial transmission time of the vth packet at the mobile at time, $t_{\mathrm{arr},1} = 0$.
(3) $t_{fn,v}$ denotes the finishing time of the vth packet at the mobile.
(4) $t_{\mathrm{rm},v}$ denotes the time when the vth packet is removed from the buffer of the mobile. From definition, this will only happen if ACK is received, and hence $t_{\mathrm{rm},v} = t_{fn,v} + sT$.
(5) $T_{\mathrm{tr},v} = t_{fn,v} - t_{\mathrm{in},v}$ is the transmission time before the packet is successfully transmitted.
(6) m_v denotes the number of retransmissions for the vth packet such that $m_v \le M_{\mathrm{re}}$.

The definitions of these variables can also be found in Figures 2(b) and 2(c). There are a few interesting relationships which can be derived if the buffer size is infinite:

$$t_{\mathrm{in},v} = \begin{cases} (v-1)T, & v \le s+1, \\ \left\{(v-1) + \sum_{q=1}^{v-s-1} m_q(1+s)\right\}T, & v > s+1, \end{cases}$$
$$T_{\mathrm{tr},v} = \begin{cases} \left\{1 + (1+s)\sum_{q=1}^{v} m_q\right\}T, & v \le s, \\ \left\{1 + (1+s)\sum_{q=v-s}^{v} m_q\right\}T, & v > s. \end{cases} \tag{12}$$

The example shown in Figure 2(c) is used for illustration. Take $t_{\mathrm{in},1} = 0$ (referenced, $v = 1$), then $t_{\mathrm{in},2} = T$, $t_{\mathrm{in},3} = 2T$ ($v = s+1$), packet 1 is not retransmitted, hence $m_1 = 0$, therefore $t_{\mathrm{in},4} = 3 + 0 = 3$, $t_{\mathrm{in},5} = 7$ since $m_2 = 1$ and $m_3 = 0$, and so on. In the following, based on the above definition, we are going to derive a few results for finite buffer size.

4.2. The number of overflowed packets

Assume there are ℓ packets in an observed on-period. When the ℓth packet arrives at the buffer, we assume χ packets have been removed from the buffer and ω ($\omega \le \chi$) packets are correctly received. The finite buffer can store a maximum of B packets, therefore, $N_{\mathrm{of}}(\ell) = \max(\ell - \chi - B, 0)$ denotes the number of overflowed packets (if any) and $\chi - \omega$ is the number of unsuccessful packets which have attempted to retransmit for M_{re} times. This is illustrated using Figure 2(c). In this example, when $\ell = 15$th packet arrival, $\chi = 6$ packets (1 to 6) have been removed from the buffer. All of these packets have been correctly received eventually, and hence $\omega = 6$. This means that $l - \chi - B = 3$ packets (9, A, and B) are lost. Note that after the ℓth packet, no packets will arrive and hence there will not be any packet overflow.

Using the relationships that $t_{\mathrm{rm},\chi} \le (\ell-1)T$ and $t_{\mathrm{rm},\chi+1} \ge (\ell-1)T$, together with the fact that m_q ranges from zero to M_{re}, the range of χ can be found to be

$$\chi \le \ell - 1 - s, \qquad \chi \ge \max\left\{\frac{\ell - 1 - s}{1 + (1+s)M_{\mathrm{re}}} - 1, 0\right\}. \tag{13}$$

Similarly, based on the fact that $\chi - \omega$ packets have been retransmitted for M_{re} times, we can obtain

$$\chi - \frac{\ell - 1 - s - \chi}{(1+s)M_{\mathrm{re}}} \le \omega \le \chi. \tag{14}$$

Although packet i is transmitted for $\sum_{q=i-s}^{i}(1+m_q)$ times, the first $\sum_{q=i-s}^{i-1}(1+m_q)$ is due to the erroneous transmissions of its previous packets and only the final $1 + m_i$ transmissions will determine whether it will be successfully transmitted. Hence, out of $n_{\mathrm{tr}} \le \chi + (l-1-s-\chi)/(1+s)$ transmissions associated to the χ packets only ω packets are successfully received. The probability that ω packets are correctly received out of all the χ removed packets when the ℓth packet arrive is given by $C_\omega^\chi \cdot (1 - p_e^{M_{\mathrm{re}}+1})^\omega (p_e^{M_{\mathrm{re}}+1})^{\chi-\omega}$, where C_ω^χ is the binomial coefficient. The probability that there are ω correct transmissions in all the n_{tr} transmissions is given by $C_\omega^{n_{\mathrm{tr}}} \cdot (1-p_e)^\omega p_e^{n_{\mathrm{tr}}-\omega}$. Averaging over all possible retransmission and overflow scenarios, the average overflowed packets conditional on ℓ are given by

$$\overline{N}_{\mathrm{of}}(\ell) = \frac{\sum_{\chi=\chi_{\min}}^{\chi_{\max}} \sum_{\omega=\omega_{\min}}^{\chi} C_\omega^{n_{\mathrm{tr}}}(1-p_e)^\omega p_e^{n_{\mathrm{tr}}-\omega} C_\omega^\chi (1-p_e^{M_{\mathrm{re}}+1})^\omega (p_e^{M_{\mathrm{re}}+1})^{\chi-\omega} \max(\ell - \chi - B, 0)}{\sum_{\chi=\chi_{\min}}^{\chi_{\max}} \sum_{\omega=\omega_{\min}}^{\chi} C_\omega^{n_{\mathrm{tr}}}(1-p_e)^\omega p_e^{n_{\mathrm{tr}}-\omega} C_\omega^\chi (1-p_e^{M_{\mathrm{re}}+1})^\omega (p_e^{M_{\mathrm{re}}+1})^{\chi-\omega}}, \tag{15}$$

where $\omega_{\min} = \chi - (l-1-s-\chi)/\lceil(1+s)M_{\mathrm{re}}\rceil$, $\chi_{\min} = \max\{(l-1-s)/\lceil 1 + (1+s)M_{\mathrm{re}}\rceil - 1, 0\}$, $\chi_{\max} = l - 1 - s$, and $n_{\mathrm{tr}} \le \chi + (l-1-s-\chi)/(1+s)$ can be derived using (13) and (14). In (15), the denominator is the normalization factor.

4.3. The lengthened activity factor

Under the assumption of a small retransmission probability, the lengthened activity factor in the GBN ARQ system, $p_{\mathrm{on},c}$, can still be approximated by

$$p_{\text{on},c} = \frac{\bar{t}_{\text{on},c}}{\bar{t}_{\text{on},c} + \bar{t}_{\text{off},c}}, \tag{16}$$

where $\bar{t}_{\text{on}} + \bar{t}_{\text{off}} = \bar{t}_{\text{on},c} + \bar{t}_{\text{off},c}$. We first illustrate how $\bar{t}_{\text{on},c}$ can be obtained.

The lengthened on-period is given by $t_{fn,\ell}$, that is, the time when it completed the transmission of the ℓth packet. Another variable $k(\ell)$ is defined, where $k(\ell) \leq \ell$ is the number of packets transmitted over the channel. In the case when there are overflowed packets, $k(\ell)$ will exclude these packets. For the example shown in Figure 2, since there is 3 overflowed packets, $k(\ell = 15) = 12$. Mathematically, the on-period is given by

$$t_{\text{on},c|\ell} = t_{fn,k} - t_{\text{in},1} = \left[k(\ell) + (1+s)\sum_{q=1}^{k(\ell)} m_q\right]T. \tag{17}$$

All retransmissions will follow the same statistics. Taking the expectation of (17) with respect to $k(\ell)$ and m, we have

$$\bar{t}_{\text{on},c|l} = E[k(l)]T + (1+s)E[m]E[k(l)]T. \tag{18}$$

Using the packet error probability (outage probability) p_e, the number of retransmissions m is a random variable with probability given by

$$\Pr(m=\rho) = \begin{cases} (1-p_e)p_e^{\rho-1}, & \rho < M_{\text{re}}, \\ (1-p_e)p_e^{M_{\text{re}}} + p_e^{M_{\text{re}}+1} = p_e^{M_{\text{re}}}, & \rho = M_{\text{re}}, \end{cases} \tag{19}$$

and its mean is given by

$$E[m] = \frac{p_e - p_e^{M_{\text{re}}+1}}{1-p_e}. \tag{20}$$

Since $k(\ell) = \ell - N_{\text{of}}(\ell)$, average over all retransmission and overflow scenarios,

$$E[k(\ell)] = \bar{k}(\ell) = \ell - \overline{N}_{\text{of}}(\ell). \tag{21}$$

As the on-period is Pareto distributed, the probability that an on-period has ℓ packets, denoted by $p(\ell)$, is approximately given by

$$p(\ell) = \Pr\{t = \ell T\} = \int_{\ell T}^{(\ell+1)T} c_{\text{on}} a^{c_{\text{on}}} t^{-c_{\text{on}}-1} dt, \quad t \geq a_{\text{on}}. \tag{22}$$

Based on (15), (20), and (21), the mean of the lengthened on-period of a web-browsing/data service in the GBN ARQ system given in (18) can be formulated by

$$\bar{t}_{\text{on},c} = \sum_{\ell=a_{\text{on}}/T}^{\infty} \left\{ p(\ell) \times \left[1 + \frac{(p_e - p_e^{M_{\text{re}}+1})(1+s)}{1-p_e}\right] \times [\ell - \overline{N}_{\text{of}}(\ell)] \times T \right\}, \tag{23}$$

where a_{on} is the minimum length of Pareto on-period and a_{on}/T means the minimum number packets in each Pareto on-period.

4.4. *Total packet loss*

Packet losses result from both finite buffer overflow and retransmissions exceeding the maximum limit. The conditional average packet loss conditioned on ℓ is given by

$$\overline{N}_{\text{loss}}(\ell) = [\ell - \overline{N}_{\text{of}}(\ell)]p_e^{M_{\text{re}}+1} + \overline{N}_{\text{of}}(\ell). \tag{24}$$

Then, the mean of the packet loss rate over time is the probabilistic summation of all possible instantaneous packet loss rates based on (22) and (24), and thus is given by

$$P_{\text{loss}} = \sum_{\ell=a_{\text{on}}/T}^{\infty} \left[\frac{p(\ell)\overline{N}_{\text{loss}}(\ell)}{\ell}\right]. \tag{25}$$

4.5. *Average buffer length and delay*

The retransmissions are assumed to be minimal so that each new on-period arrives with an empty buffer. If an on-period contains ℓ packets, the buffer length shows the following behaviors: (a) increase by one if a retransmission is made, (b) no change if a transmission or retransmission is successfully, (c) the number of packets in the buffer may reach a maximum value and stay at this state until the ℓth packet arrives, and (d) the number of packets in the buffer then decreases from the maximum value to zero. Figure 2(c) shows the buffer length from $t = 0$ to $23T$ which is given by [01234566666666666654321] and illustrates this behavior. The buffer is empty after the last packet in the buffer is removed until the arrival of next on-period. In each on/off cycle, the buffer length varies similarly.

Assume when the ξth packet arrives, the buffer is getting full, $\xi \leq \ell$. If there is no overflow, the buffer length conditioned on ℓ can be described by the following function:

$$Q_{\text{length}}(t \mid l) = \begin{cases} \left\lceil \dfrac{t}{T} \right\rceil - q, & \ell - 1 \geq t_{\text{rm},q+1} > t \geq t_{\text{rm},q}, \\ \ell - \chi, & t_{\text{rm},\chi+1} > t \geq \ell - 1, \\ \ell - \chi - p, & t_{\text{rm},\chi+p+1} > t \geq t_{\text{rm},\chi+p},\ \ell - \chi - 1 \geq p \geq 1, \\ 0, & t_{\text{on},c} + t_{\text{off},c} > t \geq t_{\text{rm},\ell}, \end{cases} \tag{26}$$

where $\lceil x \rceil$ is the smallest integer greater than x. χ is the index of the last removed packet when packet ℓ arrived and defined as $t_{\text{rm},0} = 0$. On the other hand, if there are $N_{\text{of}}(\ell)$ overflow packets, then

$$Q_{\text{length}}(t \mid l) = \begin{cases} \left\lceil \dfrac{t}{T} \right\rceil - q, & \xi - 1 > t_{\text{rm},q+1} > t \geq t_{\text{rm},q}, \\ B, & t_{\text{rm},\chi+N_{\text{of}}(\ell)+1} \geq t \geq \xi - 1, \\ B - q, & t_{\text{rm},\chi+N_{\text{of}}(\ell)+q+1} > t > t_{\text{rm},\chi+N_{\text{of}}(\ell)+q}, \\ & \qquad B - 1 \geq q \geq 1, \\ 0, & t_{\text{on}} + t_{\text{off}} > t \geq t_{\text{rm},\chi+N_{\text{of}}(\ell)+B}. \end{cases} \tag{27}$$

These expressions can be verified by looking at the queue length at time t, conditioned by l, in the example, where $T_{\mathrm{rm},1} = 5$, $T_{\mathrm{rm},2} = 7$, $T_{\mathrm{rm},3} = 11$, $T_{\mathrm{rm},4} = 12, \dots$, and $T_{\mathrm{rm},7} = 18, \dots$, as shown in Figure 2(c).

However, there are many possible retransmissions and packet overflow scenarios (ensemble space) that need to be considered for a given $t_{\mathrm{on},c}$ and $t_{\mathrm{off},c}$, denoted by $t_{\mathrm{on},c}(\ell)$ and $t_{\mathrm{off},c}(\ell)$. We approximate the ensemble average of $Q_{\mathrm{length}}(t \mid \ell)$ under all of these scenarios by $\hat{Q}_{\mathrm{length}}(t \mid \ell)$. In $\hat{Q}_{\mathrm{length}}(t \mid \ell)$, the transition time of each incremental increase in queue length as in (27) is replaced by its statistical average, which is determined by the retransmission and overflow statistics. For example, in $\hat{Q}_{\mathrm{length}}(t \mid \ell)$, $\overline{N}_{\mathrm{of}}$, $\overline{\xi}$, and $\overline{\xi}$ are used to replace N_{of}, ξ, and χ, respectively. The value of $\overline{\xi}$ is estimated using the average number of retransmissions as below:

$$\overline{\xi} - \frac{\overline{\xi} - s}{1 + E[m](1+s)} = B \Longrightarrow \overline{\xi} = \frac{B\{1 + E[m](1+s)\} - s}{E[m](1+s)}, \tag{28}$$

and $\overline{\chi}$ is estimated by

$$\overline{\chi} = \frac{\sum_{\chi=\chi_{\min}}^{\chi_{\max}} \sum_{\omega=\omega_{\min}}^{\chi} C_{\omega}^{n_{\mathrm{tr}}} (1-p_e)^{\omega} p_e^{n_{\mathrm{tr}}-\omega} C_{\omega}^{\chi} (1-p_e^{M_{\mathrm{re}}+1})^{\omega} (p_e^{M_{\mathrm{re}}+1})^{\chi-\omega} \chi}{\sum_{\chi=\chi_{\min}}^{\chi_{\max}} \sum_{\omega=\omega_{\min}}^{\chi} C_{\omega}^{n_{\mathrm{tr}}} (1-p_e)^{\omega} p_e^{n_{tr}-\omega} C_{\omega}^{\chi} (1-p_e^{M_{\mathrm{re}}+1})^{\omega} (p_e^{M_{\mathrm{re}}+1})^{\chi-\omega}}. \tag{29}$$

The average queue length conditioned on ℓ is given by

$$\overline{Q}_{\mathrm{length}}(\ell) = \frac{\int \hat{Q}_{\mathrm{length}}(t \mid \ell)\, dt}{t_{\mathrm{on},c}(\ell) + t_{\mathrm{off},c}(\ell)}. \tag{30}$$

Furthermore, if the on-period has ℓ packets, the arrival rate is assumed to be

$$\overline{\lambda}(\ell) = \frac{[\ell - \overline{N}_{\mathrm{of}}(\ell)]}{[t_{\mathrm{on},c}(\ell) + t_{\mathrm{off},c}(\ell)]}. \tag{31}$$

Since ℓ is random variable, we want to determine the average packet delay over time, denoted as D. Based on (22) and (30)-(31), D is given by

$$D = \frac{\left\{ \sum_{\ell=a_{\mathrm{on}}/T}^{\infty} p(\ell) \overline{Q}_{\mathrm{length}}(\ell) \right\}}{\left\{ \sum_{\ell=a_{\mathrm{on}}/T}^{\infty} p(\ell) \overline{\lambda}(\ell) \right\}}. \tag{32}$$

In the discussion given above, one traffic class is considered, and the outage probability is assumed known. In the following, a more practical situation is considered. The fact that multiclass services are present and the performance metrics are interdependent, the computation becomes more complicated. In general, the computation needs to be performed iteratively.

4.6. *Lengthened activity factor of non-real-time service*

In order to facilitate further analysis, let us denote the parameter set vector $[T_k, T'_k, B_k, c_k, a_k, b_k, Q\{(\delta_{i,k} - \mu_i)/\sigma_i\}, M_k]$ for the ith mobile as $\overrightarrow{U_{i,k}}$, $1 \le i \le N$, $k \in \{3,4\}$, respectively. Among the vector elements of $\overrightarrow{U_{i,k}}$, $1 \le i \le N$, $k \in \{3,4\}$, $Q\{(\delta_{i,k} - \mu_i)/\sigma_i\}$, which is shown in (10), represents the instantaneously outage probabilities of the web-browsing and data services for the ith mobile, respectively. The average lengthened activity factors of web-browsing and data services within the ith mobile are supposed to be the summation of all probabilistic activity factors over a long time. Let $\mathrm{AfFun}(\overrightarrow{U_{i,k}})$ denote instantaneous lengthened activity factor using (16) with respect to the parameter set $\overrightarrow{U_{i,k}}$. Thus, the lengthened activity factors of web-browsing and data are given by

$$p_{i,k,c} = \sum_{\vec{N}} \prod_{\vec{V}} \times \mathrm{AfFun}(\overrightarrow{U_{i,k}}). \tag{33}$$

It is shown in (5), (9), (10), and (33) that the QoS performances are intertwined across both the data link and network layers. That is, the outage probabilities, lengthened activity factors, packet loss rates, and delays are interrelated with each other. Therefore, an iteration process is developed to obtain the stable outage probabilities ($P_{\mathrm{out},i,k}$, $1 \le i \le N$, $k \in \mathbf{V}$) and the stable lengthened activity factors ($p_{i,k,c}$, $1 \le i \le N$, $k \in \{3,4\}$), satisfying (5), (9), (10), and (33). The steps of the iteration are given as follows.

(1) Set initial $p_{i,k,c}$ to be $p_{i,k,c} = p_k$, $1 \le i \le N$, $k \in \mathbf{V}$.
(2) Calculate $S_{i,k}$, $P_{\mathrm{out},i,k}$, $1 \le i \le N$, $k \in \mathbf{V}$, according to (9) and (10).
(3) Based on (33), the new $p_{i,k,c}$, $k \in \{3,4\}$, are calculated.
(4) With the new $p_{i,k,c}$, $k \in \{3,4\}$, iterate steps 2 and 3 until $p_{i,k,c}$ and $P_{\mathrm{out},i,k}$ converge.
(5) If convergence occurs, the stable values of $P_{\mathrm{out},i,k}$, $1 \le i \le N$, $k \in \mathbf{V}$, and $p_{i,k,c}$, $1 \le i \le N$, $k \in \{3,4\}$, are obtained. If it does not converge, it means that there is no feasible solution jointly satisfying (5), (9), (10), and (33).

4.7. *Packet level QoS performance at the network layer*

Based on the above analytical work of the lengthened activity factors, the packet loss rate and delay performances of the

Table 1: System parameters.

Parameter type	Value	Parameter type	Value
Shadowing mean μ	0	Number of cells, n	9
Shadowing variance σ^2	$\sigma = 6$ dB	Thermal noise power η	-103.2 dBm (4.8×10^{-14} Watt)
Path loss constant	4		

Table 2: Traffic parameter.

Traffic parameter type	Real-time services		Non-real-time service	
	Voice	Video	Web-browsing	Data
Average on-period (second)	1	0.418 (LBR) 1.5 (HBR)	1.6	2.937
Average off-period (second)	1.5	0.663 (LBR) 1.5 (HBR)	12	25.643
Activity factor (source traffic)	0.4	0.3867 (LBR) 0.5 (HBR)	0.1176	0.1028
Average rate (kbps)	24	122.3	14.1	22.8
Channel rate (kbps)	60	30 (LBR) 60 (HBR)	120	240
Spreading gain	64	128 (LBR) 64 (HBR)	32	16
Number of spreading codes	1	8 (LBR) 1 (HBR)	1	1
Buffer size (number of packets)	0	0	200	400
Convolutional rate	1/2	1/2	1/2	1/2

four classes are formulated. Within the ith mobile, let the packet loss rates and delays for voice, LBR video, HBR video, web-browsing, and data services be denoted by $P_{\text{loss},i,k}$ and $D_{i,k}$, $1 \leq i \leq N$, $k \in \{1, 2l, 2h, 3, 4\}$, respectively.

As voice and video are NRT delay-sensitive services, no ARQ mechanism is implemented in their packet transmissions. Thus, their packet loss rates are just equal to their outage probability, which is given by

$$P_{\text{loss},i,k} = P_{\text{out},i,k}, \quad k \in \{1, 2l, 2h\}, \tag{34}$$

and their average delays are simply their packet transmission time, which is given by

$$D_{i,k} = T_k, \quad k \in \{1, 2l, 2h\}. \tag{35}$$

On the other hand, the lengthened activity factors, average packet loss rates, and average delays of web-browsing and data are based on both their instantaneous outage probabilities and the GBN ARQ mechanism. Let us denote the average packet loss rates and average delay as $P_{\text{loss},i,k}$ and $D_{i,k}$, $1 \leq i \leq N$, $k \in \{3, 4\}$, respectively, which are the average values over the time. Let $\text{PlossFun}(\overrightarrow{U_{i,k}})$ and $\text{DelayFun}(\overrightarrow{U_{i,k}})$ denote instantaneous packet loss rate and delay using (25) and (32), respectively, with respect to the parameter set $\overrightarrow{U_{i,k}}$. Therefore, the average packet loss rates of web-browsing and data services are given by

$$P_{\text{loss},i,k} = \sum_{\vec{N}} \prod_{\vec{V}} \times \text{PlossFun}\left(\overrightarrow{U_{i,k}}\right), \tag{36}$$

and the average delays of web-browsing and data services are given by

$$D_{i,k} = \sum_{\vec{N}} \prod_{\vec{V}} \times \text{DelayFun}\left(\overrightarrow{U_{i,k}}\right), \tag{37}$$

where $1 \leq i \leq N$, $k \in \{3, 4\}$, respectively.

5. NUMERICAL RESULTS

In our analytical model, each mobile can support multiconnection multiclass traffic. In order to demonstrate the reasonableness of our analytical formulation presented in previous sections, numerical results are presented in this section. A cellular mobile network with n square cells is considered. We assume that the number of mobiles with heterogeneous classes is identical in each cell and all mobiles are uniformly distributed. We simulate the network model with SMPL simulation kernel, a type of discrete event simulator [16]. System parameters and traffic parameters are shown in Tables 1 and 2.

Each mobile in our analysis supports up to four diverse classes simultaneously. Suppose that all mobiles in each cell can be divided into four groups including different classes. The class distribution and group size are given in Table 3. In practice, with 4 different traffic classes, there can be up to 15 different combinations and similar analytical approach can be applied. We vary the number of users in Group 1 and fix the number of users in all the other groups. The numerical results are plotted in Figures 3–14.

Firstly, we can clearly observe that all analytical results show better agreements when the systems are in light and medium loads (less than 1.3 Mbps) than when they are in

Table 3: Number of services in each mobile user.

Group index	Group 1	Group 2	Group 3	Group 4
Number of mobiles	$N_1 = 5 \sim 23$	$N_2 = 2$	$N_3 = \mathbf{2}$	$N_4 = 5$
Classes in each mobile	1 voice	1 video	1 voice + 1 video	1 web + 1 data

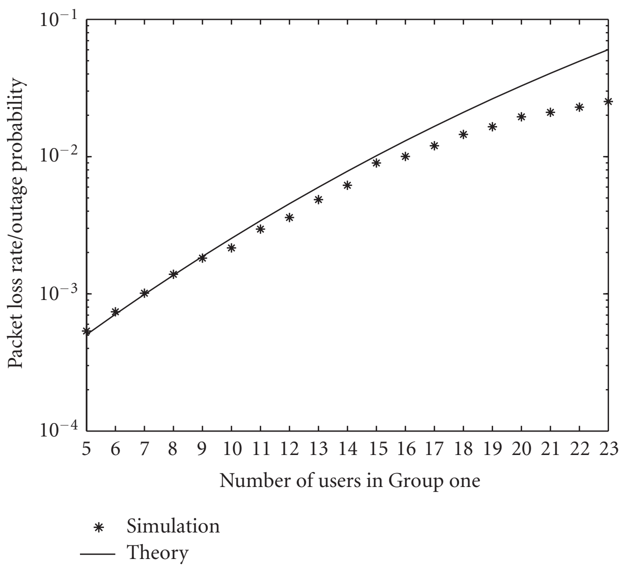


Figure 3: Packet loss rate/outage probability of voice services (Group 1).

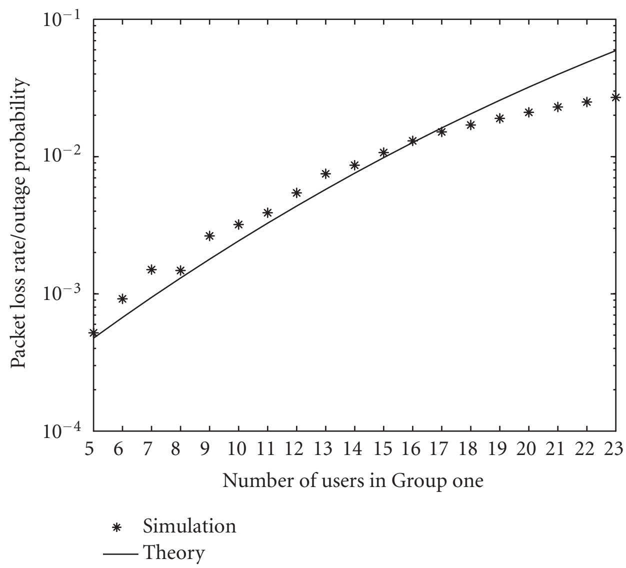


Figure 4: Packet loss rate/outage probability of video services (Group 2).

heavy load. The deviation during heavy load, that is, when there are more mobiles in the system, can be explained as follows. The outage becomes more severe and thus retransmissions occur more frequently during heavy load. Our GBN ARQ analysis is accurate assuming the retransmissions occur less frequently and the packet error rate is low. If a lot of retransmissions happen under high load, the on-periods of web-browsing or data services in the CDMA channel may overlap, which influences the computation of their lengthened activity factors, outage probabilities, packet loss rates, and delays. As all classes in CDMA systems are intertwined with each other, the QoS metrics therefore deviate from simulation results. Therefore, our analytical formulation is more suitable for light and medium loads when the throughput of the system is below or around 1.3 Mbps. On the other hand, under higher load, the packet loss rates and delay performances have already exceeded their specific requirements. For example, the packet loss rates requirements of these classes should be less than either 10^{-2} for voice and video or 10^{-3} for web-browsing and data, which are defined in [1].

Secondly, we also have some comments on the complexity of the analysis. Our final analytical expressions are relatively complex. This is due to the fact that we jointly consider more realistic traffic models, GBN ARQ, multicell network, and four traffic classes in order to approximate the real network. These factors complicate the analysis. Despite this, the analysis still takes much shorter time to work out the results than using simulation. For example, it takes more than 24 hours to obtain the simulation results, while the analytical results can be computed in less than one hour. Therefore, the analytical solution proves to be much more efficient in estimating the QoS performances.

6. CALL ADMISSION CONTROL METHOD AND ADMISSION REGION

In previous literatures, CAC is analyzed with many approaches in [17, 18]. But these works are not totally QoS-based and do not address cross-layer CAC in CDMA networks. Our contribution is that the analytical formulation in this paper leads to the determination of the cross-layer admission region (AR) in the uplink of a CDMA system. A QoS-based CAC scheme is given here. If the outage probability, packet loss rate, and delay requirements are defined as δ_{out}, δ_{loss}, and δ_d, the AR at the packet level in the uplink of CDMA systems, denoted by $\mathbf{R}$, is given by

$$\mathbf{R} = \{(1, 2, 3, \ldots, i, \ldots, N) \mid P_{\text{loss},i,k} \leq \delta_{\text{loss}}, D_{i,k} \leq \delta_d, P_{\text{out},i,k} \leq \delta_{\text{out}}, \text{SINR}_{I,K} = \gamma_K^*\}, \quad (38)$$

where $1 \leq i \leq N$, $k \in \mathbf{V}$.

Figure 15 shows the CAC scheme in the uplink of CDMA systems. This CAC scheme admits or rejects call admission

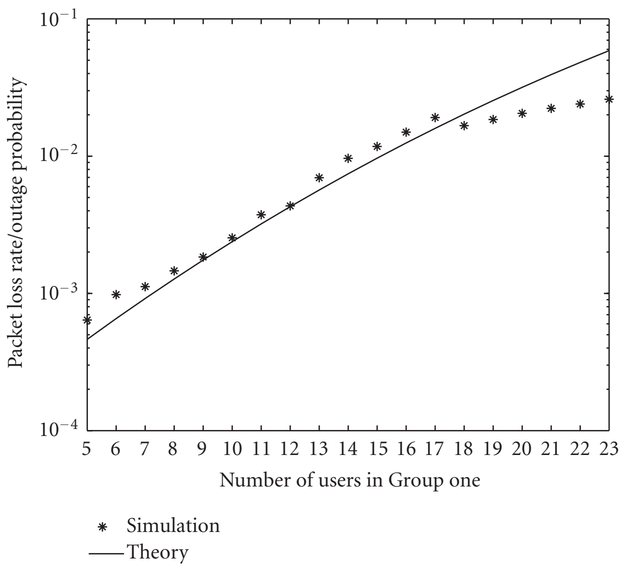


Figure 5: Packet loss rate/outage probability of voice services (Group 3).

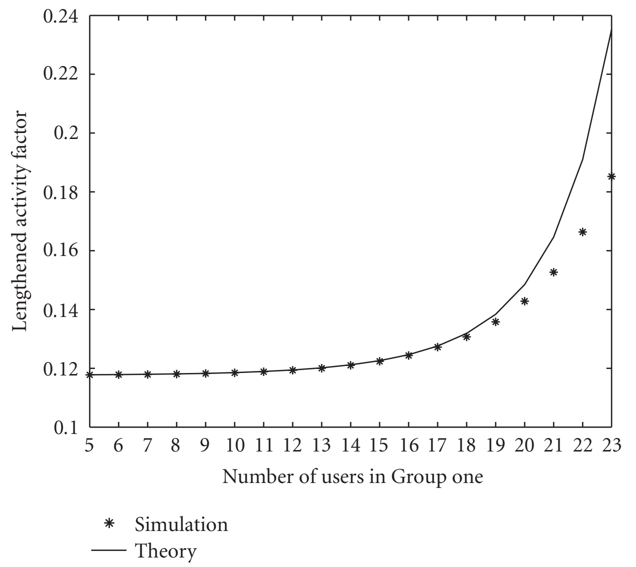


Figure 7: Lengthened activity factor of web-browsing services (Group 4).

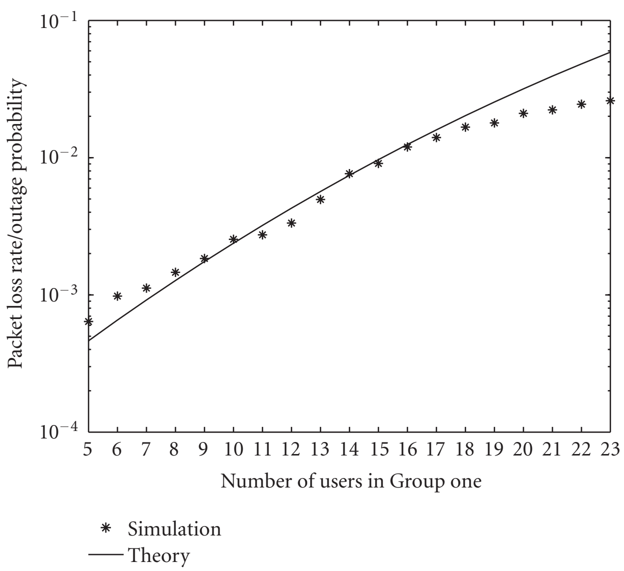


Figure 6: Packet loss rate/outage probability of video services (Group 3).

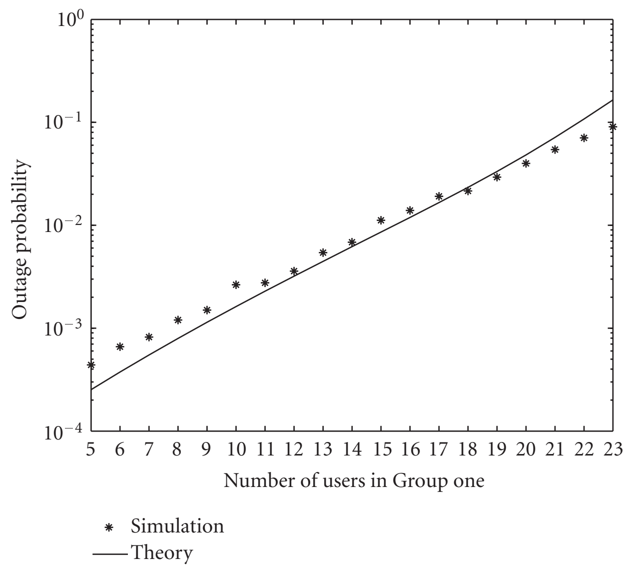


Figure 8: Outage probability of web-browsing services (Group 4).

requests based on the satisfaction of average SINR requirements and outage probability performance at the data link layer and packet level QoS performances including packet loss rate and delay at the network layer. In Figure 15, when a specific set of mobile requests to be admitted into the network, the CAC process is initiated. The CAC first obtains the power levels for all mobiles. If positive power solutions are available, the SINR requirements of these mobiles are satisfied at the data link layer. Otherwise, the CAC rejects this set of mobiles directly due to their unsatisfactory average SINR. With the positive power solutions, the CAC computes the outage probabilities of all mobiles and the lengthened activity factors of NRT services. Iterations are performed to make both the outage probabilities and the lengthened activity factors converge. If the iterations cannot reach convergence, feasible solutions are not available and thus this combination of mobile users should be rejected by CAC. If the iterations converge, the stable outage probabilities for all services and the lengthened activity factors for NRT services are obtained. Next, the packet loss rate and average delay of each service are calculated. If the obtained outage probability, packet loss rate, and delay requirements are simultane-

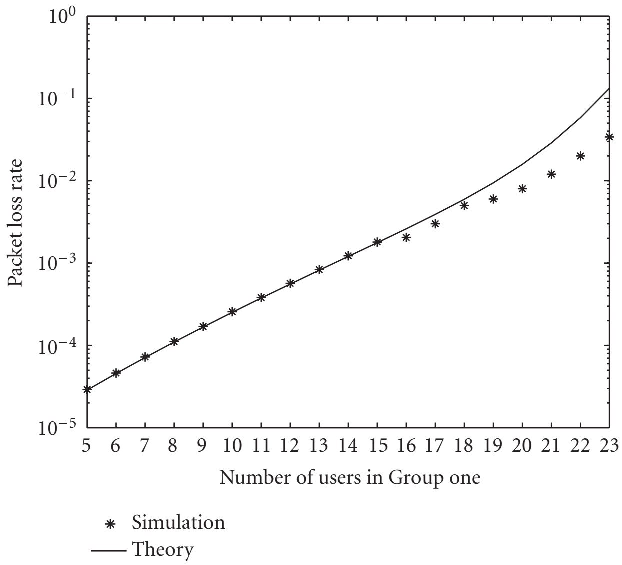


Figure 9: Packet loss rate of web-browsing services (Group 4).

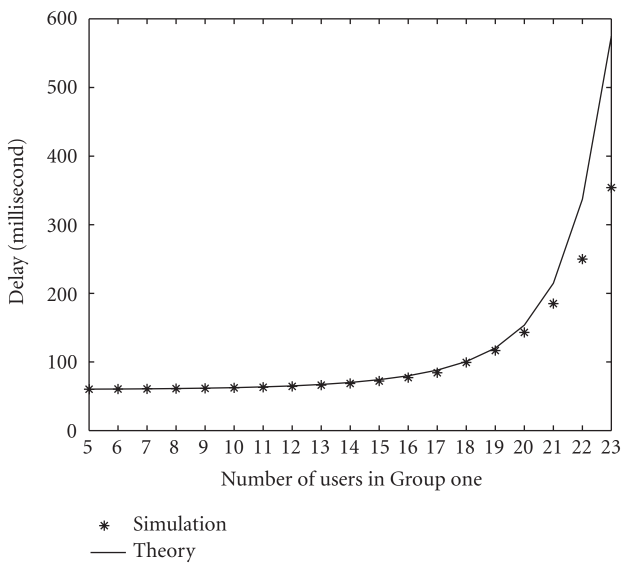


Figure 10: Average delay of web-browsing services (Group 4).

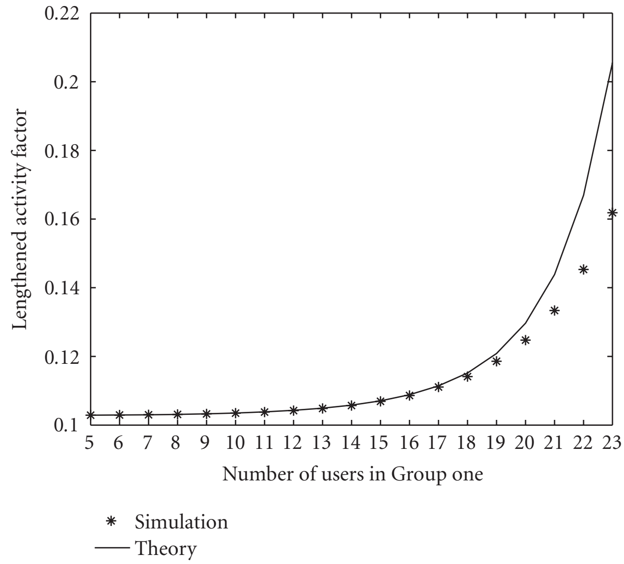


Figure 11: Lengthened activity factor of data services (Group 4).

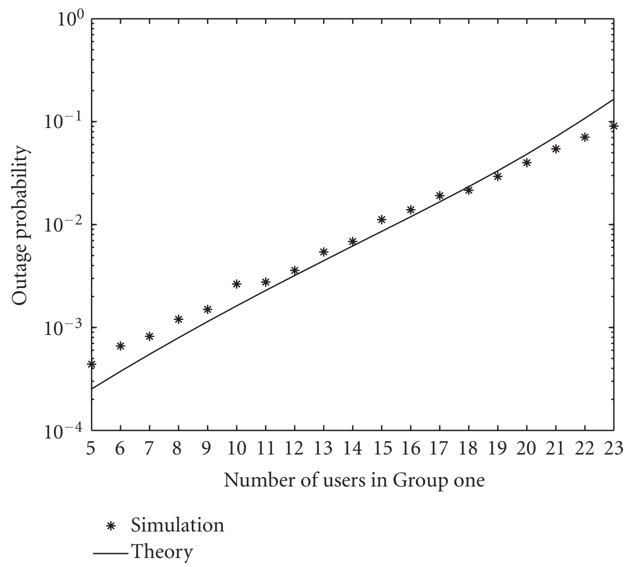


Figure 12: Outage probability of data services (Group 4).

ously satisfied, this set of mobiles can be assured of QoS requirements and thus can be admitted into the network by the CAC scheme. Otherwise, this set of mobiles should be rejected by the CAC scheme. Compared to existing CAC methods in [17, 18], the main advantage of this CAC scheme is that it is totally based on the cross-layer QoS satisfaction of all admitted mobiles in terms of specific SINR, outage probability, packet loss rate, and delay requirements. That is, the QoS requirements of all admitted mobiles are completely satisfied at both the data link layer and packet level of the network layer, and the system capacity is thus maximized. Using the given parameters in Table 1, an example of a 3-dimensional feasible AR is shown in terms of the number of mobiles in Figure 16.

In a realistic CDMA system, the BS can utilize dedicated control channels to do fast power control and guarantee that each traffic stream is received with the desired power level. Based on the global information gathered from the network, the CAC can find out admission region with our analytical model in advance and save as a table at the BS. During operation, CAC at the BS can simply look up the table to make CAC decisions.

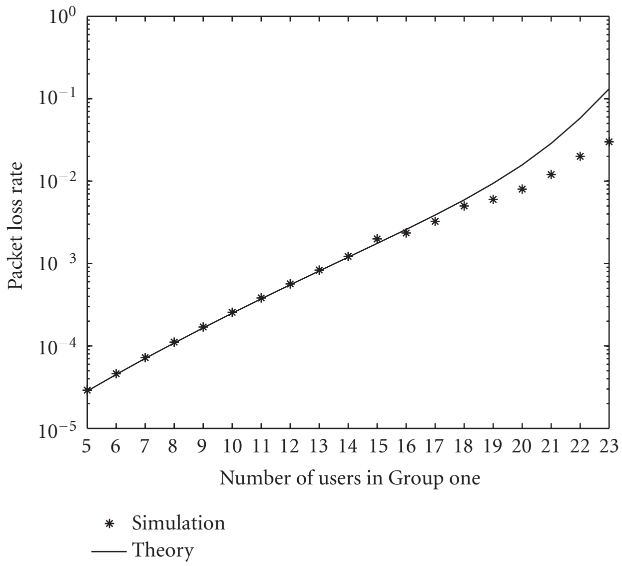


Figure 13: Packet loss rate of data services (Group 4).

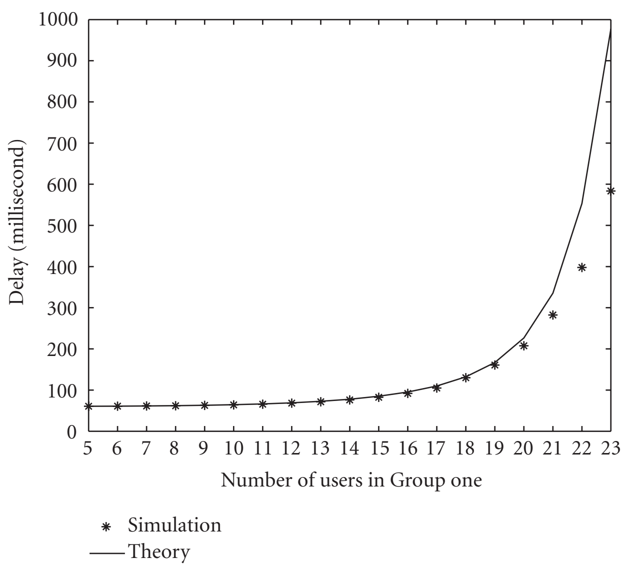


Figure 14: Average delay of data services (Group 4).

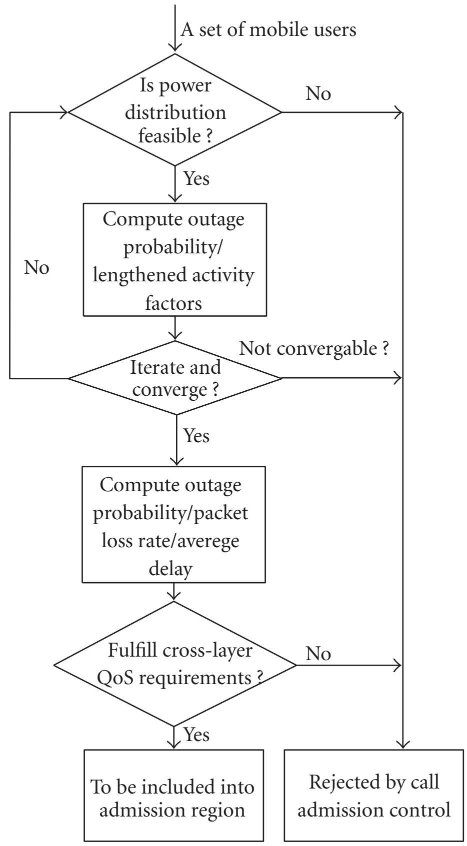


Figure 15: Call admission control procedures.

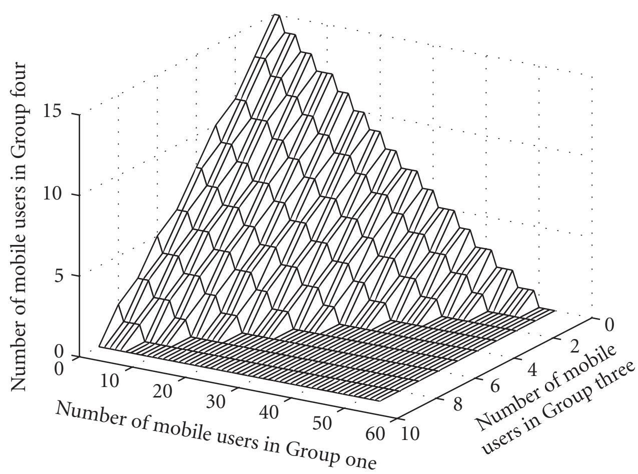


Figure 16: Admission region with three groups of users (assume $N_2 = 0$).

7. CONCLUSION

We have presented an approximate analytical framework for the cross-layer QoS in CDMA networks. Four classes of services are served within the same mobile and GBN ARQ with finite buffer size and limited retransmissions is implemented for NRT traffic with Pareto-on/Pareto-off sources for the first time. In our analysis, the coupling of packet-level QoS at the network layer and data-link-layer QoS is investigated. The numerical results show that our analytical approach can approximate the simulation results quite well up to medium traffic load. Based on the cross-layer QoS constraints, a CAC method is proposed to maximize the system capacity and leads to the determination of admission region in the uplink of CDMA systems. Our analytical work can be further combined with the call level analysis of QoS performances to provide a joint capacity evaluation at both call and packet levels in CDMA networks.

REFERENCES

[1] 3GPP TS 23.107, "QoS Concept and Architecture," v5.7.0, December 2002.

[2] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver Jr., and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, pp. 303–312, 1991.

[3] D. T. C. Wong, J. W. Mark, K. C. Chua, J. Yao, and Y. H. Chew, "Performance analysis of multiclass services in the uplink of wideband CDMA," in *Proceedings of the 8th International Conference on Communication Systems (ICCS '02)*, vol. 2, pp. 692–696, Singapore, November 2002.

[4] D. T. C. Wong, J. W. Mark, K. C. Chua, and B. Kannan, "Performance analysis of variable bit rate multiclass services in the uplink of wideband CDMA," in *Proceedings of IEEE International Conference on Communications*, vol. 1, pp. 363–367, Anchorage, Alaska, USA, 2003.

[5] D. T. C. Wong, J. W. Mark, and K. C. Chua, "Performance evaluation of video services in a multirate DS-CDMA system," in *Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '03)*, vol. 2, pp. 1490–1495, Beijing, China, September 2003.

[6] C. Nie, D. T. C. Wong, and Y. H. Chew, "Outage analysis for multi-connection multiclass services in the uplink of wideband CDMA cellular mobile networks," in *Proceedings of the 3rd IFIP-TC6 Networking Conference*, pp. 1426–1432, Athens, Greece, May 2004.

[7] J. B. Kim and M. L. Honig, "Resource allocation for multiple classes of DS-CDMA traffic," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 506–519, 2000.

[8] R. Fantacci and S. Nannicini, "Multiple access protocol for integration of variable bit rate multimedia traffic in UMTS/IMT-2000 based on wideband CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 8, pp. 1441–1454, 2000.

[9] 3GPP TR 25.942, "RF system scenarios," v.5.3.0, June 2004.

[10] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Transactions on Communications*, vol. 36, no. 7, pp. 834–844, 1988.

[11] A. Jamalipour, *The Wireless Mobile Internet: Architectures, Protocols and Services*, John Wiley & Sons, Chichester, West Sussex, UK, 2003.

[12] M. Evans, *Statistical Distributions*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1998.

[13] C. Nie, "Packet level quality of service analysis of multiclass services in a WCDMA mobile network," M.S. thesis, National University of Singapore, Kent Ridge, Singapore, 2004.

[14] Y. Hayashida, N. Sugimachi, M. Komatsu, and Y. Yoshida, "Go-back-N system with limited retransmissions," in *Proceedings of the 8th IEEE Annual International Phoenix Conference on Computers and Communications*, pp. 183–187, Scottsdale, Ariz, USA, March 1989.

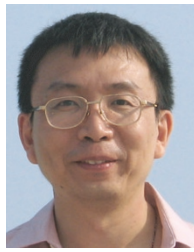
[15] Y. Hayashida, S. Fujii, and Y. Tezuka, "Delay performance of a continuous ARQ system with copy-transmissions," in *Proceedings of the 7th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '88)*, pp. 714–721, New Orleans, La, USA, 1988.

[16] M. H. MacDougall, *Simulating Computer Systems: Techniques and Tools*, MIT Press, Cambridge, Mass, USA, 1987.

[17] B.-J. Hwang and J.-S. Wu, "Call admission control for a multiple rate CDMA system," *IEICE Transactions on Communications*, vol. E84-B, no. 11, pp. 2932–2945, 2001.

[18] L. Zhang, X. Cheng, and K. R. Subramanian, "Admission control for integrated video-conferencing/voice/data services in broadband CDMA networks," *Computer Communications*, vol. 23, no. 5, pp. 499–510, 2000.

Chun Nie received the B.Eng. degree from Northwestern Polytechnic University, China, and the M.Eng. degree from the National University of Singapore, Singapore, in 2000 and 2005, respectively, all in electrical engineering. He is currently working towards his Ph.D. degree at the Department of Electrical and Computer Engineering, University of North Carolina, Charlotte, NC, USA. His research interests include medium access control, quality-of-service, cross-layer protocol design, and resource management in wireless networks.



Yong Huat Chew received the B.Eng., M.Eng., and Ph.D. degrees in electrical engineering from the National University of Singapore (NUS), Singapore. He has been with the Institute for Infocomm Research (formerly also known as Centre for Wireless Communications, NUS, and Institute for Communications Research), an institute under the Agency for Science, Technology, and Research, where he is currently a Senior Scientist, since 1996. He is also an adjunct Associate Professor in the Department of Electrical and Computer Engineering, National University of Singapore. His research interests are in technologies related to high spectrally efficient wireless communication systems and radio resource management.

David Tung Chong Wong received the B.Eng. and M.Eng. degrees from the National University of Singapore (NUS) in 1992 and 1994, respectively, and the Ph.D. degree from the University of Waterloo, Canada, in 1999, all in electrical engineering. He is with the Institute for Infocomm Research, Singapore (formerly Centre for Wireless Communications, NUS, and Institute for Communications Research, NUS) first as a Research Engineer and currently as a Scientist, since 1994. His research interests are in communications networks, 3 G/4 G, and ultra-wideband wireless mobile multimedia networks. His areas of research are in the medium access control, resource allocation with quality-of-service constraints, traffic policing with heterogeneous traffic, and cross-layer design. He is a Senior Member of the IEEE. He was on the Technical Program Committee of the IEEE WCNC 2003, IEEE WCNC 2005, IEEE GLOBECOM 2005, and IEEE ICCS 2006.