A Framework for Performance Evaluation and Optimization of an Emerging Multimedia DS-CDMA Network*

Vladimir Marbukh
National Institute of Standards and Technology
100 Bureau Drive, Stop 8920
Gaithersburg, MD 20899-8920, USA
1(301) 590-0932
marbukh@nist.gov

ABSTRACT
This paper proposes a framework for performance evaluation and optimization of an emerging multimedia, packet Direct-Sequence Code Division Multiple Access (DS-CDMA) network with a wide range of Quality of Service (QoS) requirements on losses and delays. The need for a new framework arises from inability of the traditional approach, based on the outage probability, to capture the queueing aspects of DS-CDMA network behavior in presence of delay tolerant traffic. Accounting for these aspects becomes essential for emerging multimedia DS-CDMA networks attempting to approach their capacity limits by using coding and spreading gain control, retransmissions, as well as transmission scheduling/power control. Since in a DS-CDMA network transmissions compete for simultaneous access to several resources, including wireless bandwidth and transmission power, the paper proposes to approximate the feasible QoS region for the network by the intersection of the feasible QoS regions for the corresponding single-resource systems. The feasible QoS region for a single-resource system is estimated by using $M/G/1$ conservation laws. Based on this "bottleneck resource" approximation, the paper estimates the admission region for the network and outlines the approach to the network management aimed at maximizing the admission region.

Keywords
Direct sequence code division multiple access, multimedia, quality of service, performance, control.

1. INTRODUCTION
1.1 Challenges
Ability to provide multimedia services with wide range of bit rates and Quality of Service (QoS) requirements on losses and delays is the main force driving evolution of the wireless systems. Among other multiple access schemes, Direct-Sequence Code Division Multiple Access (DS-CDMA) has received much attention as an underlying technology for future wireless networks [1]-[2]. The 2nd generation DS-CDMA networks have been designed to support primarily circuit-switched voice traffic. Performance of a DS-CDMA network in the circuit-switched mode, measured by the outage probability, has been successfully modeled by loss systems with losses occurring when the bit energy-to-interference ratio falls below certain threshold [3]-[4]. The need for a new framework for performance evaluation and optimization of the emerging multimedia, packet DS-CDMA networks arises from inability of the traditional approach, based on evaluation of the outage probability, to capture the queueing aspects of a DS-CDMA network behavior in presence of delay tolerant traffic. In a DS-CDMA network transmissions compete for the simultaneous access to several resources, including wireless bandwidth and transmission power. Delays occur when, as a result of fluctuation, the aggregated instantaneous incoming load exceeds the capacity of at least one resource. The fluctuations may be caused by burstiness of the multimedia traffic and by the wireless channel impairments. The aggregated delay is determined by the capacity of the resources and by the characteristics of the aggregated traffic. The delays for individual sessions depend on the traffic scheduling and/or resource sharing disciplines resolving competition for the resources. Accounting for the queueing aspects of the network behavior is essential for a multimedia DS-CDMA network attempting to approach its capacity limit by scheduling transmissions of delay tolerant traffic.

Qualitatively, importance of queueing behavior of a DS-CDMA network can be demonstrated by analysis of the shape of the admission region. The admission region represents all possible combinations of sessions the system can accommodate, given the QoS requirements and system resources. Consider a case of two classes of service: Real Time (RT), i.e., voice, and Non-Real Time (N-RT), i.e., data. If the sources within each class are identical, the admission region is $(N_1, N_2) \in \mathrm{A}$ where N_1 is the number of RT sessions and

* This work was supported by NIST ATP program

N_2 is the number of N-RT sessions. It is known [5] that for a single-resource system admission region A has a trapezoidal shape OABCD shown in Fig. 1, assuming that scheduling gives priority to the RT traffic over N-RT traffic and RT traffic is bursty. The qualitative reason for the trapezoidal shape of the admission region is that even serving the maximum number of bursty RT sources the system is able to "squeeze" certain amount of N-RT traffic between the bursts of RT traffic. The same system with First In First Out ($FIFO$) scheduling has triangular admission region OAD. Gain ABCD in the admission region is a result of correlation between RT and N-RT queues produced by traffic scheduling.

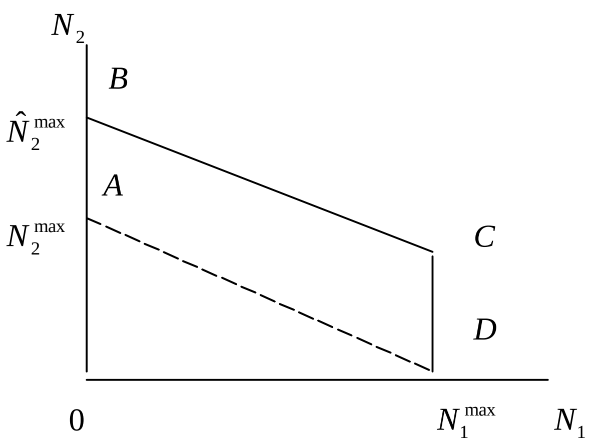

Figure 1. Typical admission region for a mixture of RT and N-RT traffic.

The same qualitative result on the shape of the admission region may be expected for a DS-CDMA network. Despite importance of the transmission scheduling for increasing throughput for delay tolerant sessions has been established [6]-[7], known models of a DS-CDMA network do not take into account correlation between different queues and, as a result, produce linear boundary of the admission region [8]-[9]. Known models of a DS-CDMA network have been intended to evaluate effect of specific parameters on specific performance metrics for specific traffic mixtures. For example, papers [7] and [9] address effect of dynamic processing gain control on average delay and loss for a mixture of data and voice traffic, papers [6] and [8] analyze effects of transmission scheduling/power control and dynamic processing gain control on throughput for delay tolerant traffic. Developing an accurate yet tractable performance model of a multimedia, packet DS-CDMA network for the purpose of evaluation of the combined effect of retransmissions, power control/transmission scheduling as well as processing and coding gains on the relevant performance metrics for various traffic mixtures remains an open problem. This paper is intended to contribute to the solution of this problem.

1.2 Outline of the Proposed Framework

This paper proposes a framework for performance evaluation and optimization of a quasi-stationary, multimedia, cellular DS-CDMA network. Quasi stationary assumption implies that the wireless channel properties and traffic patterns change slowly as compared to the time needed for the queues to closely approach the steady state. The framework uses the packet and burst level models. The packet level model describes the effect of the average traffic load, system parameters, and control actions on the average transmission power by nodes and interference at the base stations. The packet level description can be interpreted as an enhancement of the proposed in [10] characterization of congestion in a DS-CDMA network in terms of the average transmission powers and interference. We take this characterization a step further by relating the average transmission powers and interference to the traffic and system parameters as well as control actions, including dynamic spreading gain control, retransmission policy, and transmission scheduling/power control for delay tolerant sessions. At the burst level a DS-CDMA network is modeled as a pool of the network resources, including wireless bandwidth and transmission power, accessed by transmissions. Since the capacity of a DS-CDMA network is interference limited, the average transmission power and interference enter the burst level model through the resource allocation scheme. The burst level performance of a DS-CDMA network is evaluated under assumption that the performance is limited by the bottleneck resource. Given the session QoS requirements, an effective resource management scheme strives to eliminate the bottleneck to approach the network capacity. The burst level model assumes the fluid traffic model [11], and uses conservation laws [12] to assess the admission region.

Note that accurate and tractable performance model of a multimedia packet DS-CDMA network should take proper account not only of the average interference levels, but also of their fluctuations. Significant amount of work has been done on the average interference in a DS-CDMA network (see, for example, [6], [8] and [10]) since the average interference is directly related to the average resource utilization and throughput. However, delays in a queueing system depend not only on the average resource utilization, but also on the fluctuations in the resource utilization. Since under heavy traffic approximation the average delay in a $M/G/1$ system depends on the average as well as variability in the resource utilization, one may expect that performance of a DS-CDMA network depends not only on the average interference level, but also on its variability. This paper proposes simple approximation for the variability of the transmission powers and interference levels. This approximation in combination with $M/G/1$ conservation law can be used to estimate the feasible QoS and admission region for a multimedia packet DS-CDMA network. However, better description of power fluctuations is needed for accurate performance modeling of an emerging multimedia packet DS-CDMA network.

The paper uses the proposed framework to analyze some particular scenarios. The analysis yields two major results. First, on the example of a mixture of RT and N-RT traffic we demonstrate that the proposed framework captures the non-linear aspects of the queueing network behavior, essential for the network trying to approach its capacity limit by transmission scheduling/power control. Second, we show that the proposed framework produces results qualitatively as well as quantitatively consistent with those obtained by customized methods in various particular scenarios. These scenarios include optimal balance between spreading and retransmissions [9], as well as between transmission scheduling/power control and resource sharing [6], [8]. Our analysis suggests that the proposed framework may be used in the domain of a multimedia DS-CDMA network where the traditional approach, based on the outage probability, is insufficient.

The paper is organized as follows. In the rest of this section we describe the reverse link in a DS-CDMA network and fluid traffic model considered in this paper. Section 2 describes some possible control actions in a DS-CDMA network, including spreading gain control, packet retransmissions, and transmission scheduling/power control. Section 3 proposes the packet level model of a DS-CDMA network and briefly discusses the effect of the control actions on the transmission power and interference. Section 4 proposes burst level performance model of a DS-CDMA network, and characterizes the admission region for multimedia traffic as a function of the system parameters, control actions and burst level traffic characteristics. Section 4 also discusses the shape of the admission region and demonstrates that the proposed framework captures the non-linear effects of interference between queues resulted from transmission scheduling. Section 5 analyses effect of the control actions on the system performance under assumption that the network performance is determined by the bottleneck resource. Specifically, subsection 5.1 finds optimal balance between spreading and retransmissions. These results generalize earlier results [9]. Subsection 5.2 discusses optimal balance between transmission scheduling/power control and bandwidth sharing. These results generalize earlier results [6] and [8].

1.3 Multimedia Cellular DS-CDMA Network

We consider the reverse link (from mobiles to the base station) in a cellular DS-CDMA network with K cells and N users. Transmissions by mobile $n \in Cell(k)$ are decoded by base station k. We assume that each transmission is perceived as additive white noise by other transmissions; thus, we do not consider joint multi-user detection here. Note that results from [13] suggest that our approach can be extended to DS-CDMA networks with joint multi-user detection. The network supports S services. We refer to a session of class of service s initiated by user n as the session (n,s). Service $s=1$ is a circuit-switched mode of operation suitable for RT applications with very stringent delay requirements. Packet services $s=2,..,S$ are suitable for RT packet applications and N-RT applications with less stringent delay requirements than the circuit-switched service $s=1$. Since the number of services S is small comparatively to the possible variety of applications and their delay requirements, we assume that service s has much more stringent delay requirements than service s' if $s<s'$. We assume the time frame structure of operations. At the beginning of each frame the base station knows the amount of backlogged traffic from each session. Any circuit switched session of class $s=1$ backlogged at the beginning of a frame spreads its transmission over the entire frame with spreading factor W^*. We assume ideal power control able to maintain the same power p_k received at the base station k from all class $s=1$ transmissions within cell k. Retransmissions are not allowed for circuit-switched service $s=1$. Packet services $s=2,..,S$ allow transmission scheduling and packet retransmissions.

Since the circuit-switched service $s=1$ assumes constant spreading gain and no retransmissions, the QoS requirements for a circuit-switched session in cell k can be expressed as follows:

$$\Pr ob\{p_k / I_k \le c\} \le e \tag{1}$$

where the random interference experienced by this session at the base station k is I_k, minimum allowable signal-to-interference ratio is c, and the maximum tolerable outage probability is $e<<1$. We assume the following QoS requirements for a packet session (n,s), $s=2,..,S$:

$$L_{ns} \le L_s^{\max} \tag{2}$$

$$T_{ns} \le T_s^{\max} \tag{3}$$

where the bit error probability and average bit delay for session (n,s) are respectively L_{ns} and T_{ns}. We assume that delay tolerance for packet services are high enough and packet sizes are small enough to allow for replacing the average bit delays with the average packet delays in (3).

A DS-CDMA network uses two basic resources: the wireless bandwidth and the transmission power supplied by batteries carried by mobiles. The wireless bandwidth is represented by the chip rate F. The technology imposes the following peak transmission power constraints for transmitter n:

$$P_n^-(t) \stackrel{def}{=} \sum_s P_{ns}^-(t) \le \hat{P}_n^- \tag{4}$$

where the transmission power for a session (n,s) at moment t is $P_{ns}^-(t)$ and the peak transmission power for mobile n is $\hat{P}_n^-$. Constraints on the remaining battery life can be expressed in terms of the average transmission power by mobiles as follows:

$$\overline{P}_n^-(t) \stackrel{def}{=} \sum_s \overline{P}_{ns}^-(t) \le \widehat{\overline{P}}_n^- \tag{5}$$

1.4 Traffic Model

Despite our framework can be extended to a case of arbitrary fluid traffic sources with known effective bandwidths [11], due to space constraints, we consider a specific case of fluid traffic sources. We model all circuit-switched sessions of service $s = 1$ by independent on-off fluid traffic sources with the same peak bit rate $\hat{m}$, activity factor f, and exponentially distributed on and off periods with averages t_{on} and t_{off} respectively. The average bit rate of a circuit-switched session is $m = f\hat{m}$. We model a packet session of class $s = 2,..,S$ by a compound Poisson process $x_s(t) = \sum_{j=1}^{J(t)} Y_{sj}$ where burst sizes Y_{s1}, Y_{s2}, … are independent identically distributed random variables with distribution function B_s, and the number of bursts J_s is an independent Poisson process of rate h_s. The moments of the distribution B_s are $b_s^{(i)} = \int x^i dB_s(x)$. The average bit rate of a packet session of class $s \geq 2$ is $m_s = h_s b_s^{(1)}$.

2. CONTROL ACTIONS

2.1 Spreading Gain Control

The backlogged traffic generated by a packet session (n,s), $s \geq 2$ is transmitted at rate $C_{ns} = F/W_{ns}$ where F is the chip rate of the system and W_{ns} is the spreading gain. However, due to interference, only a fraction of transmitted packets is accepted at the base station as received correctly, i.e. without bit errors, or with an acceptable number of bit errors. Let G_{ns} be the probability that a packet generated by a session (n,s) is received without bit errors or with an acceptable number of bit errors. Success probability G_{ns} is an increasing function of the bit energy-to-interference ratio $(E_b/I_0)_{ns}$: $G_{ns} = G_{ns}[(E_b/I_0)_{ns}]$, where the bit energy-to-interference ratio is $(E_b/I_0)_{ns} = W_{ns} SIR_{ns}$. The Signal-to-Interference Ratio for a session (n,s), $n \in Cell(k)$ is $SIR_{ns} = x_n P_{ns}^{-}/I_{ns}$ where the transmission power is P_{ns}^{-}, the path gain from transmitter n to the base station k, $n \in Cell(k)$ is x_n, and the interference experienced by session (n,s) at the receiving base station k is I_{ns}. Following [7]-[9] we assume that a base station $k = 1,..,K$ instantaneously notifies all transmitters $n \in Cell(k)$ about current path gain x_n. Transmitter n maintains fixed bit energy-to-interference ratio for a packet session (n,s), $s = 2,..,S$ equal to $(E_b/I_0)_s = g_s$ by selecting the spreading gain

$$W_{ns} = g_s / SIR_{ns} \qquad (6)$$

where g_s is the system parameter aimed at optimizing the system performance. The spreading gain control scheme (6) is idealization since it assumes that the signal-to interference ratios are immediately known to the transmitter and the spreading gains can be adjusted immediately, even during transmission of a packet. However, this scheme is a good approximation in situations when the variability of the signal-to-interference ratios SIR_{ns} tend to be low or can be predicted or controlled.

In practical situations the spreading gains can take some positive integer values within feasible region $W_{ns} \geq W_{ns}^{\min}$ where the minimum spreading gain for session (n,s) is determined by the technological constraints, e.g., obviously, $W_{ns}^{\min} \geq 1$. Following [7]-[9] we assume that in a congested mode, with high spreading gains $W_{ns} >> 1$, the discrete nature of the spreading gain is negligible and (6) can be rewritten as follows:

$$W_{ns} = \max\{W_{ns}^{\min}, g_s / SIR_{ns}\} \qquad (7)$$

2.2 Retransmissions

Packets that are received at the base station with an unacceptable number of bit errors can be retransmitted. We assume that success of a packet transmission or lack of therein is acknowledged immediately, and retransmissions are scheduled on the $FIFO$ basis and have priority over new packets generated by the same session. Figure 2 presents the packet flow diagram for a session (n,s).

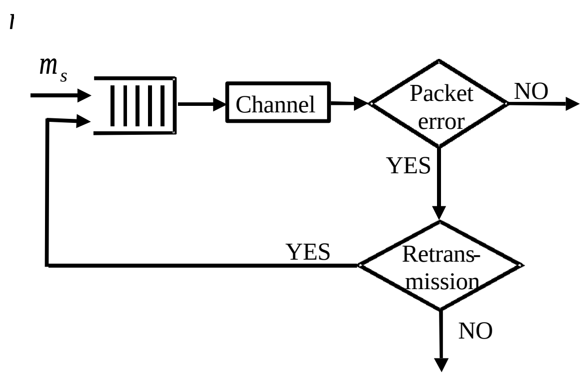

Figure 2. Packet flow diagram.

If a packet containing x bits is retransmitted $r-1$ times, then assuming that checking for packet correctness at the receiving base station, notification of the sending node and possible retransmissions are instantaneous, the total number of bits sent is $X = rx$. Since spreading gain control (6) maintains the same bit energy-to-interference ratio g_s for all packet sessions of service $s = 2,..,S$, the average number of retransmissions for a packet generated by session (n,s) is $\bar{r}_s - 1$ where $\bar{r}_s$ is independent of node n. It can be shown that $\bar{r}_s = \bar{r}(R_s, Q_s)$ is the following function of the maximum allowable number of a packet retransmissions $R_s - 1$ and fixed packet error probability $Q_s = 1 - G_s(g_s)$:

$$\bar{r} = \frac{1 - Q^R}{1 - Q} \tag{8}$$

Note that it is not difficult to take into account the processing and propagation delays associated with checking for packet correctness at the base station, sending back the acknowledgement and possible retransmissions.

2.3 Transmission Scheduling/Power Control

Since the capacity of a DS-CDMA network is interference limited, unscheduled packet transmissions by mobiles may result in high average transmission power violating QoS and power constraints. A natural way to reduce the average transmission power by mobiles is the following class of scheduling disciplines for packet transmissions:

(a) If some packet session of service $s = 2,..,S$ transmits, all backlogged packet sessions of higher priority services $s' = 2,..,s-1$ within the cell also transmit. This condition reflects much more stringent delay requirements for a packet service s_1 than for service s_2 if $s_1 < s_2$.

(b) A mobile $n \in Cell(k)$ can transmit packet traffic of service $s = 2,..,S$ only if the level of interference at the base station k from packet transmissions of this service within this cell does not exceed certain threshold $\hat{i}_s$. This condition allows the network to control the interference level from packet transmissions within a cell by selecting appropriate thresholds $\hat{i}_s$.

(c) A backlogged session of service $s = 2,..,S$ may transmit if the total level of interference at the base station does not exceed certain threshold $\hat{I}$. This condition is intended to prevent disrupting the circuit-switched transmissions.

(d) If conditions (a)-(c) allow transmissions by some but not all backlogged packet sessions, the conflict is resolved by some scheduling discipline, for example based on random selection.

For simplicity, in this paper we only consider a case of $S = 2$ services: circuit switched service $s = 1$ and packet service $s = 2$. Generalization to a case of arbitrary number of services S while is possible, requires some additional new ideas. We assume the power control policy for the packet service $s = 2$ allowing a mobile $n \in Cell(k)$ either transmit packet traffic with maximum available power, or not transmit at all:

$$P_n^- = \hat{P}_n^- - p_k / x_n \tag{9}$$

if mobile n also transmits circuit switched traffic, and

$$P_n^- = \hat{P}_n^- \tag{10}$$

otherwise. Results from [8] provide some justification for this power control policy. Note that condition (c) with threshold

$$\hat{I} \le p/c \tag{11}$$

ensures that packet transmissions do not disrupt the circuit switched transmissions.

3. RESOURCE UTILIZATION AND TRASNSMISSION POWER

3.1 Average Uplink Utilization

Average utilization of the uplink from mobile $n \in Cell(k)$ to the base station k by packet transmissions, i.e., the average portion of time mobile n transmits packet traffic, is

$$J_n = \bar{P}_n^- / \tilde{P}_n^- \tag{12}$$

where the average transmission power for a packet traffic by mobile n is $\bar{P}_n^-$, and the average transmission power by mobile n for a packet session during this session activity is $\tilde{P}_n^-$. Further in the paper we only consider the most practically interesting case when the level of interference $i_{n\to k}$ experienced by a mobile $n \in Cell(k)$ packet transmission lies in the following region:

$$W_n^{\min} \le \frac{i_{n\to k}}{\tilde{P}_n^- x_n} g \le \frac{F_n}{l} \tag{13}$$

where $l = g\bar{r}m$. Low bound in (13) ensures that the level of interference is high enough for the technological constraints $W_n \ge W_n^{\min}$ do not limit the network capacity. Upper bound in (13) is a necessary condition for stability, i.e., limited backlog, for the packet session. Assuming that (13) holds, the average transmission power $\bar{P}_{ns}^-$ for a session (n,s), $n \in Cell(k)$, is determined by the following formula:

$$\bar{P}_{ns}^- x_n = \begin{cases} f p_k & if \ s = 1 \\ r \bar{i}_{n\to k} & if \ s = 2 \end{cases} \tag{14}$$

where the "normalized" average utilization of the uplink $n \to k$ by packet transmission is $r = l/F$, and the average interference experienced by this transmission is $\bar{i}_{n\to k} = E[i_{n\to k}]$. Combining (12) with (14) we obtain

$$J_n = \frac{\bar{i}_{n\to k}}{\tilde{P}_n^- x_n} r \tag{15}$$

3.2 Average Resource Utilization

The condition that the average battery utilization at each node n is less than 1 imposes the following throughput constraints:

$$p_n \stackrel{def}{=} p_{n1} + p_{n2} < 1 \tag{16}$$

where the average utilization of the mobile $n \in Cell(k)$ battery by transmissions of class s is

$$p_{ns} = \bar{P}_{ns}^- / \hat{P}_{ns}^- \tag{17}$$

and the average transmission powers $\bar{P}_{ns}^-$ are given by (14).

The throughput region with respect to the bandwidth is determined by the condition that the average bandwidth utilization at each base station k is less than 1:

$$b_{(k)} \stackrel{def}{=} \sum_{n \in Cell(k)} b_n < 1 \tag{18}$$

where the average utilization of the wireless bandwidth at the base station k by packet traffic transmitted by mobile $n \in Cell(k)$ is b_n. To calculate b_n first notice that in a case of ideal packet traffic scheduling when no more than one packet session is active within any cell $b_n = J_n$ where J_n is given by (12). Consider a general case when more than one backlogged packet sessions are allowed to transmit simultaneously within the same cell. In this case the average fraction of time a given mobile n transmits packet traffic is still J_n given by (12). However, the power received at the base station from this given transmission is the following fraction

$$h_n = \frac{P_n^- x_n}{P_n^- x_n + \bar{i}_{n\to k}^*} \tag{19}$$

of the total power received at the base station from all packet transmissions within this cell. In (19) the average inner cell interference experienced by a mobile $n \in Cell(k)$ packet transmission from all packet transmissions within cell k is

$$\bar{i}_{n\to k}^* = \bar{i}_{n\to k} - \left(q_k + \bar{I}_k^{out} + \bar{I}_{1\to k}^{in}\right) \tag{20}$$

where the average thermal noise power at the base station k is q_k, the average outer cell interference is $\bar{I}_k^{out}$, and the average inner cell interference from circuit switched service $s = 1$ is $\bar{I}_{1\to k}^{in}$. Since DS-CDMA allocated wireless bandwidth proportionally to the received powers it is natural to assume that

$$b_n = h_n J_s \tag{21}$$

Combining (15), (19) and (21) we obtain

$$b_n = r \frac{\bar{i}_{n\to k}}{P_n^- x_n + \bar{i}_{n\to k}^*} \tag{22}$$

3.3 Average Transmission Power and Interference

Let

$$\tilde{I}_k = \frac{q_k + \bar{I}_k^{out} + \bar{I}_{1\to k}^{in}}{1 - N_{k2} r}, \tag{23}$$

$$\bar{I}_k = \min\left\{\tilde{I}_k, q_k + \bar{I}_k^{out} + \bar{I}_{1\to k}^{in} + \hat{i}\right\}, \tag{24}$$

$$\bar{I}_k^* = \bar{I}_k - (q_k + \bar{I}_k^{out} + \bar{I}_{1\to k}^{in}) \tag{25}$$

where the number of active packet sessions in cell k is N_{k2}. It is not difficult to derive the following approximations:

$$\bar{i}_{n\to k} \approx \bar{I}_k, \quad \bar{i}_{n\to k}^* \approx \bar{I}_k^* \;\text{ if }\; \bar{I}_k \le \hat{I}_k \tag{26}$$

and packet traffic within cell k is blocked if $\bar{I}_k > \hat{I}_k$. Note that accuracy of the approximations (26) improve as the number of mobiles in each cell increases and, as a result, relative contribution of each mobile into the level of interference at the base station decreases [10]. It is possible to proceed without approximation (26) at the cost of making formulas more cumbersome. Further we use approximation (26) and assume that condition $\bar{I}_k \le \hat{I}_k$ holds.

The average outer cell interference is

$$\bar{I}_k^{out} = \sum_{n \notin Cell(k)} x_{nk} \sum_{s=1,2} \bar{P}_{ns}^- \tag{27}$$

where the average transmission power $\bar{P}_{ns}^-$ is given by (14), and the path gain from node n to the base station k is x_{nk}. Note that $x_{nk} = x_n$ if $n \in Cell(k)$. Since the outage probability (1) for circuit switched service $e << 1$, the average inner cell interference from circuit-switched service $s = 1$ is approximately

$$\bar{I}_{1\to k}^{in} = f N_{k1} p_k \tag{28}$$

where the number of sessions of service s in cell k is N_{ks}.

Combining (23)-(24) and (28) we obtain

$$\bar{I}_k = \frac{q_k + fN_{k1} p_k + \bar{I}_k^{out}}{1 - N_{k2} r} \quad (29)$$

if

$$\frac{N_{k2} r}{1 - N_{k2} r} \left(q_k + fN_{k1} p_k + \bar{I}_k^{out} \right) < \hat{i}\,, \quad (30)$$

$$N_{k2} r \le 1 \quad (31)$$

and

$$\bar{I}_k = q_k + \bar{I}_k^{out} + \bar{I}_{1\to k}^{in} + \hat{i} \quad (32)$$

otherwise. To simplify formulas we further assume, unless stated otherwise, that each mobile n has two active sessions (n,s): circuit switched session $s=1$ and packet session $s=2$. In this case

$$\bar{I}_k^{out} = \sum_{m \ne k} (fp_m + r\bar{I}_m) \sum_{n \in Cell(m)} \left(x_{nk} / x_n \right) \quad (33)$$

Equations (29), (33) form a closed system of $2K$ fixed point equations for the average interference levels $\bar{I}_k$ and $\bar{I}_k^{out}$. After this system is solved average transmission powers by mobiles can be easily determined from (14). System (29), (33) can be used for studying effect of the average traffic rates, system parameters and various control actions on the average transmission power by mobiles and average level of interference experienced by the sessions at the base station. Note that in a case of a single-cell network $\bar{I}_k^{out} \equiv 0$, and (29) gives the following explicit result:

$$\bar{I} = \min\left\{ \frac{q + fN_1 p}{1 - N_2 r},\;\; q + fN_1 p + \hat{i} \right\} \quad (34)$$

3.4 Examples

Two extreme cases for the average transmission power by mobiles and interference levels at the base stations are complete bandwidth sharing and ideal packet transmission scheduling. A case of complete bandwidth sharing, when all packet sessions transmit "at will", is realized when

$$\hat{I} = \hat{i} = \infty \quad (35)$$

In this case, assuming that (31) holds, we have from (29), (33)

$$\bar{I}_k = \frac{q_k + fN_{k1} p_k + \bar{I}_k^{out}}{1 - N_{k2} r} \quad (36)$$

where $\bar{I}_k^{out}$ is the solution to the following system of K linear equations:

$$\bar{I}_k^{out} = \sum_{m \ne k} \left(fp_m + r\,\frac{q_k + fN_{m1} p_k + \bar{I}_m^{out}}{1 - N_{m2} r} \right) \sum_{n \in Cell(m)} \left(x_{nk} / x_n \right) \quad (37)$$

If power is abundant and wireless bandwidth is the limiting resource, (30) is a necessary condition for feasibility, i.e., the existence of finite, non-negative solution of the linear system (36). In this case the average bandwidth utilization (22) takes the following form:

$$b_n = \frac{r}{\tilde{P}_n^{-} x_n \dfrac{1 - N_{k2} r}{q_k + fN_{k1} p_k + \bar{I}_k^{out}} + N_{k2} r} \quad (38)$$

It is easy to see that the throughput condition with respect to the wireless bandwidth (18) holds if $N_{k2} r \le 1$, i.e., in a case of complete bandwidth sharing the transmission power is the limiting resource.

In a case of ideal transmission scheduling only one packet session is allowed to transmit within any cell k at any given moment even if more than one packet session is backlogged within this cell. The packet transmissions are synchronized according to some scheduling discipline. The idea behind ideal transmission scheduling is to reduce the interference [6], [8]. This case can be realized with sufficiently small $\hat{i}_k$. In this case we have from (29), (33)

$$\bar{I}_k = q_k + fN_{k1} p_k + \bar{I}_k^{out} \quad (39)$$

where $\bar{I}_k^{out}$ is the solution to the following system of K linear equations:

$$\bar{I}_k^{out} = \sum_{m \ne k} \left\{ fp_m + r\,(q_k + fN_{m1} p_k + \bar{I}_m^{out}) \right\} \sum_{n \in Cell(m)} \left(x_{nk} / x_n \right) \quad (40)$$

In this case the average bandwidth utilization (22) takes the following form:

$$b_n = r\,\frac{q_k + fN_{k1} p_k + \bar{I}_k^{out}}{\tilde{P}_n^{-} x_n} \quad (41)$$

Note that existence of the positive finite solution to the systems (37) and (40) can be investigated by using approach proposed in [10].

4. BURST LEVEL PERFORMANCE

In this section we consider the burst level performance and the admission region for an isolated cell assuming that outer cell interference is fixed. This conditional admission region can be used for decentralized adaptive admission control when each base station k adapts to slow variation in the outer cell interference by monitoring the outer cell interference and correspondingly adjusting the admission region for cell k.

Combining the conditional admission region with equations for the average outer cell interference one can derive the unconditional admission region for the entire network. We estimate the queueing delays under assumption that the variability in the signal-to-interference ratio for a packet transmission by a mobile n is low $SIR_n \approx E[SIR_n]$. This assumption can be justified for uncontrolled large-scale network by the law of large numbers, and for any network with controlled level of interference.

4.1 Losses and Delays for the Packet Service

The QoS requirements (2) impose the following constraint on the target bit energy-to-interference ratio g and maximum allowable number of retransmissions $R-1$ for the packet service $s=2$:

$$[Q(g)]^R \le L^{\max} \tag{42}$$

where the packet error probability is $Q(g)=1-G(g)$. We approximate the QoS region for packet sessions by using $M/G/1$ conservation laws for the average delays for packet traffic under heavy traffic approximation [12]. If the only limiting resource is the peak transmission power (1), the $M/G/1$ conservation law yields the following low boundary on the average delay for the packet traffic transmitted by mobile $n \in Cell(k)$:

$$T_n \ge \frac{u_{n1}+u_{n2}}{1-\bar{P}_n^- \big/ \tilde{P}_n^-} \tag{43}$$

where the average transmission power by mobile n for a packet session during this session activity $\tilde{P}_n^-$ can be approximated as follows:

$$\tilde{P}_n^- \approx \max\left\{\hat{P}_n^- - q\ p_k / x_n,\ 0\right\}, \tag{44}$$

the variability due to the circuit switched transmission is

$$u_{n1} = \frac{1}{2}\frac{t_{on}^2 + t_{off}^2}{t_{on}+t_{off}}\frac{p_k}{\hat{P}_n x_n}, \tag{45}$$

the variability due to the packet transmission is

$$u_{n2} = \frac{1}{2} m(\bar{r}\overline{W}_n)^2 b^{(2)}, \tag{46}$$

and the average spreading factor for packet traffic transmitted by mobile n is

$$\overline{W}_n = E[g/SIR_n] \approx g\,\frac{\bar{I}_k}{\tilde{P}_n^- x_n} \tag{47}$$

If the only limiting resource is the wireless bandwidth, the $M/G/1$ conservation law yields the following low bound on the on the aggregated average delay for the packet traffic transmitted by mobiles $n \in Cell(k)$:

$$\sum_{n\in Cell(k)} b_n T_n \ge \frac{b_{(k)}}{1-b_{(k)}}\left(V_{k1}+V_{k2}\right) \tag{48}$$

where the variability due to the aggregated circuit switched transmissions is

$$V_{k1} = \frac{1}{2}\frac{t_{on}^2 + t_{off}^2}{t_{on}+t_{off}}\frac{p_k}{\bar{I}_k} \tag{49}$$

the variability due to the packet transmissions is

$$V_{k2} = \frac{1}{2} m\bar{r} b^{(2)} \sum_{n\in Cell(k)} (\overline{W}_n)^2 \tag{50}$$

and the average spreading factor $\overline{W}_n$ is given by (48).

Combining (43) and (48) with (3) we obtain the following approximate admission region for packet services with respect to the QoS constraints (3):

$$\frac{u_{n1}+u_{n2}}{1-\bar{P}_n^- \big/ \tilde{P}_n^-} \le T^{\max} \tag{51}$$

$$\frac{V_{k1}+V_{k2}}{1-b_{(k)}} \le T^{\max} \tag{52}$$

4.2 Losses for the Circuit Switched Service

Given QoS requirement (1), the number N_1 of circuit-switched sessions a cell can sustain is determined by the following inequality:

$$\sum_{I}\sum_{n>1/c-I/p} \Pr ob\left(I|n\right)\frac{N_1!}{n!(N_1-n)!} f^m (1-f)^{N_1-n} < e \tag{53}$$

where the number of active circuit-switched sessions in the cell is n, the interference experienced by an active circuit-switched session at the base station is I, and conditional probability of I given n is $\Pr ob(I|n)$. Interference I can be divided into two parts: $I = I^{(1)} + I^{(2)}$ where the thermal noise plus outer cell interference at the base station is $I^{(1)}$, and the inner cell interference from packet service is $I^{(2)}$. For a quasi-stationary network interference $I^{(1)}$ can be assumed constant, while interference level $I^{(2)}$ is controlled by the scheduling discipline. If (11) holds, the admission region for circuit-switched sessions is

$$N_1 \le N_1^{\max} \tag{54}$$

where $N_1^{\max} = N_1^*$ is the maximum integer N_1 which satisfies the following inequality

$$\sum_{n>1/c-I^{(1)}/p} \frac{N_1!}{n!(N_1-n)!} f^n (1-f)^{N_1-n} < e \tag{55}$$

Since typically $e << 1$ and $N_1^{\max} >> 1$, we can use large deviation approximation to simplify (55). According to this approximation, $N_1^{\max}$ is the solution to the following equation (see [14]):

$$N_1^{\max}\left[u\log\left(\frac{u}{f}\right)+(1-u)\log\left(\frac{1-u}{1-f}\right)\right]=\log\left(\frac{1}{e}\right) \tag{56}$$

where $u = U/N_1^{\max}$, and $U = 1/c - I^{(1)}/p$. This approximation can be used when $e << 1$, $U >> 1$, $U \propto \log(1/e)$.

4.3 Admission Region

Admission region is determined by constraints (4)-(5), (51)-(53) and depends on the burst level traffic characteristics, system parameters, and control actions. The capacity region is the upper limit of the admission region over the control system parameters. It can be shown that in an extreme case of heavy bandwidth sharing, the admission region typically has almost linear boundary. In another extreme case of ideal transmission scheduling for packet services, the boundary of the admission region typically has piece-wise linear boundary. Different linear boundary pieces correspond to different bottleneck resources limiting the performance of the network. This is consistent with qualitative discussion in subsection 1.1 of the paper and with results obtained in [5].

In the rest of this subsection we demonstrate the essential non-linearity of the boundary of the admission region in a case of ideal transmission scheduling for packet services for a simple scenario. The scenario includes $N = N_1 + N_2$ nodes in a single cell where N_1 nodes transmit in circuit switched mode (service $s = 1$), and N_2 nodes transmit in packet mode (service $s = 2$). We assume that condition (11) holds, the packet service has very loose QoS requirements on delay, power constraints (4)-(5) are not critical, and all packet sessions in the cell are identical with respect to transmission power and path gains. Under this scenario the admission region is determined by throughput constraint with respect to the wireless bandwidth (18) and constraint on losses (42) and (54). It is easy to see that constraints (18) can be rewritten as follows:

$$r(\bar{I}^* + fpN_1)N_2 \le \bar{I}^* + \hat{P}x \tag{57}$$

and the admission region has approximately a trapezoidal shape OABCD shown on Fig. 1.

5. NETWORK MANAGEMENT

5.1 Spreading vs. Retransmissions

The optimal balance between spreading and retransmissions maximizes the admission region over target bit energy-to-interference ratio g subject to QoS and resource constraints. In a case of very loose delay requirements for packet services, the optimal balance minimizes the "effective" bit rate $l = g\bar{r}m$. Using (8) we can rewrite this optimization problem as follows:

$$\min_{g \ge 0}\left\{g\,\frac{1-Q^R}{1-Q}\right\} \tag{58}$$

subject to

$$Q^R \le L^{\max} \tag{59}$$

where $Q = Q(g)$ is the packet error probability for a packet session, and $R = R$ is the maximum number of packet transmissions allowed. In a case $R = 1$ the solution to (58)-(59) is given by the equation $Q(g) = L^{\max}$. In a case $R = \infty$, optimization problem (58)-(59) takes form of the following optimization problem which has been derived and analyzed in [9]:

$$\min_{g \ge 0}\left\{\frac{g}{1-Q(g)}\right\} \tag{60}$$

Note that $g_s^{opt}(R) \downarrow g_s^*$ and $l_s^{opt}(R) \downarrow l_s^*$ as $R \uparrow \infty$, where g_s^* and l_s^* are determined by solution of the optimization problem (60). Our analysis leads to the conclusion that the optimal maximum allowable number of packet retransmissions $R^{opt} - 1 = \infty$. It can be shown that in a case of finite delays the optimal maximum allowable number of packet retransmissions $R^{opt} - 1$ is also finite.

5.2 Transmission Scheduling vs. Bandwidth Sharing

Combining (17) and (22) with (14) and (26) we obtain

$$p_{n2} = r\,\frac{\bar{I}_k}{P_n^- x_n} \tag{61}$$

$$b_n = r\,\frac{\bar{I}_k}{P_n^- x + \bar{I}_k^*} \tag{62}$$

Assuming P_n^- constant, consider the differential effect of an increase in the bandwidth sharing among packet services on the average utilization of the resources:

$$\frac{\partial p_{n2}}{\partial \bar{I}_k^*} = \frac{r}{P_n^- x_n}\left(\frac{\partial \bar{I}_k}{\partial \bar{I}_k^*}\right) \tag{63}$$

$$\frac{\partial b_n}{\partial \bar{I}_k^*} = r\,\frac{\left(\hat{P}_n^- x_n + \bar{I}_k^*\right)\left(\partial \bar{I}_k / \partial \bar{I}_{ns}^*\right) - \bar{I}_k}{(\hat{P}_n^- x_n + \bar{I}_k^*)^2} \qquad (64)$$

$n \in Cell(k)$. Since $\partial \bar{I}_k / \partial \bar{I}_k^* > 0$, increase in the inner cell interference $\bar{I}_k^*$ always increases the packet session average activity factor: $\partial p_{n2} / \partial \bar{I}_k^* > 0$. However, increase in the inner cell interference $\bar{I}_k^*$ increases the average bandwidth utilization b_n if

$$\left(\hat{P}_n^- x_n + \bar{I}_k^*\right)\left(\partial \bar{I}_k / \partial \bar{I}_{ns}^*\right) > \bar{I}_k \qquad (65)$$

and reduces b_n otherwise. This conclusion has important implications for the network management. If the transmission power is the bottleneck, the system should schedule transmissions in order to reduce interference $\bar{I}_k^*$. When the bandwidth limitation is the bottleneck, transmission scheduling may improve or worsen performance. This conclusion is consistent with results [6] and [8]. Note that overhead required for scheduling can be easily taken into account.

6. CONCLUSION

This paper proposes a framework for burst level performance evaluation and optimization of a packet, multimedia DS-CDMA. The network is modeled as the multi-resource system under assumption that the burst level performance is limited by the bottleneck resource. Resources include wireless bandwidth and transmission power. Analysis of the proposed model yields two major results. First, the model captures queueing aspects of the network behavior, essential for the network trying to approach its capacity limit by transmission scheduling/power control. Second, the model produces results consistent with those obtained by customized methods.

The proposed framework can be extended to include effects of coding and modulation techniques, account for various wireless channel impairments and relax assumptions regarding traffic model and structure of services. Future research should be directed toward these extensions in order to incorporate specific system features and gain better understanding of the joint effect of multiple controlled parameters on multimedia DS-CDMA network performance with final goal to approach the network capacity limit.

7. REFERENCES

[1] T. Ojanpera and R. Prasad, "Overview of air interface multiple access for UMTS/IMT-2000," *IEEE Commun. Mag.*, vol. 36, no. 9, 1998, 82-95.

[2] F. Adachi, M. Sawahashi, and H. Suda, "Wideband DS-CDMA for next generation mobile communication systems," *IEEE Commun. Mag.*, vol. 36, no. 9, 1998, 56-69.

[3] K. Gilhousen, I. Jacobs, A. Padovani, A. Viterbi, L. Weaver, and C. Wheatley, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol*., vol. 40, 1991, 303-312.

[4] A.M. Viterbi and A.J. Viterbi, "Erlang capacity of a power controlled CDMA system," *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 892-900, 1993.

[5] A.W. Berger and W. Whitt, "Extending the effective bandwidth concept to networks with priority classes," *IEEE Comm. Magazine*, vol. 36, no.8, Aug. 1998, 78-83.

[6] S. Ramakrishna and J. Holtzman, " A scheme for throughput maximization in a dual-class CDMA system," *IEEE J. Select. Areas Commun*., vol. 16, no. 6, 1998, 830-844.

[7] M.L. Honig and J.B. Kim "Allocation of DS-CDMA parameters to achieve multiple rates and qualities of service," in *Proc. IEEE Globecom*, 1996.

[8] S.J. Oh and K.M. Wasserman, " Optimality of greedy power control and variable spreading gain in multi-class CDMA mobile networks," in Proc. Mobicom, 1999, 102-112.

[9] S.J. Oh and K.M. Wasserman, "Dynamic spreading gain control in multi-service CDMA networks," *IEEE J. Select. Areas Commun*., vol. 17, no. 5, 1999, 918-927.

[10] S.V. Hanly, "Congestion measures in DS-CDMA networks," *IEEE Trans. Commun*. Vol. 47, no. 3, 1999, 426-437.

[11] F.P. Kelly, "Notes on effective bandwidth," *Stochastic Networks*, F.P. Kelly, S. Zachary, I. Ziedins, Eds., Oxford: Clarendon Press, 1996, 141-68.

[12] L. Kleinrock, Queueing Systems, Volume II: Computer Applications, J. Wiley & Sons, 1976.

[13] D.N. Tse and S.V. Hanly, "Effective bandwidth in wireless networks with multi-user receivers," INFOCOM'98, San Francisco, CA, 1998, Paper IB-1.

[14] A. Elwalid, D. Mitra, and R.H. Wentworth, "A new approach to allocating buffers and bandwidth to heterogeneous, regulated traffic in an ATM node," *IEEE J. Select. Areas Commun*., vol. 13, no. 6, 1995, 1115-1127.