Capacity of Multi-service Cellular Networks with Transmission-Rate Control: A Queueing Analysis

Eitan Altman
INRIA, BP93, 2004 Route des Lucioles,
06902 Sophia-Antipolis, France
also CESIMO, Facultad de Ingeniería,
Univ. de Los Andes, Mérida, Venezuela
altman@sophia.inria.fr

ABSTRACT

In this paper we compute the uplink capacity of power-control CDMA mobile networks with an idealized power control, that contain best-effort type applications, i.e. applications whose transmission rate can be controlled. An arriving best-effort call is assumed to have a fixed amount of traffic to send, so the transmission rate assigned to it determines the duration of the call. We allow for multi-services (so that mobile stations have different quality of service requirements). Unlike some previous published work where soft blocking was considered (and the system was thus allowed to operate beyond capacity), we assume that a call admission mechanism is implemented in order to prevent a new call to arrive when the system is already saturated. This guarantees the quality of service of ongoing calls. Our first result is that slowing the transmission rates in the case of a single cell with homogeneous quality of service characteristics increases capacity. This suggests that there is a limit capacity that can be approached when slowing down the transmission rates. We identify this limit and show that it has the following property: as long as the arrival rate of information is below some level, blocking probability can become arbitrarily small by sufficiently slowing down the transmission rates. We then extend the results to the general heterogeneous and multi-cell case.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Queueing theory; C.2.1 [**Computer Communication Networks**]: Network Architecutre and Design—*Wireless communication*

General Terms

Performance, Theory

Keywords

Best-effort, Erlang capacity, CDMA, multi-cell, multi-class, queueing theory

1. INTRODUCTION

The traditional way to define capacity has been to ask "how many calls can a system handle". Various definitions have been used. The Erlang capacity which has been used in telephony networks is a probabilistic definition, it specifies the arrival rate of calls that the system can allow so that the probability of blocking of an arrival is lower than some threshold. Another version of this definition has been introduced in [17] for the wireless context where the capacity is taken to be the rate of calls that the system can allow so that the probability that the quality of service is not attained is sufficiently small; here calls are not blocked when exceeding the limit of the system to provide the required quality of service. In the above definitions, the transmission rate used by a call is a fixed constant which may be class dependent.

Third generation wireless networks allow for multimedia applications and new services are proposed, in particular file transfers, Internet browsing and electronic mail. These non-interactive applications are less sensitive to the assigned throughput. We could consider them as part of a best-effort service in which the transmission rate can be assigned by the base-station For a given rate of arrival of best-effort sessions, or calls, the capacity of the system will depend on the assigned throughput. However, we assume that the total volume V_s of traffic created by an application s does not depend on the assigned traffic. The duration that this session will be present and will occupy network resources is $V_s/R(s)$ where $R(s)$ is its assigned transmission rate. We assume that the power control is such that the energy per bit of the best-effort application s does not depend on its transmission rate.

Adding this new flexibility of transmission rate assignment, the capacity of best-effort applications can now be defined as the number of sessions that the network can handle assuming the "best possible" assignment of transmission rates. We shall also consider the case of a system with both best-effort calls as well as Real-Time (RT) applications whose throughput is fixed and is not controlled.

For simplicity, we do not take into considerations the effects of imperfect power control that have already been considered in many previous papers, see e.g. [2, 17]: we assume that power control is instantaneous and we ignore saturation phenomena that impose in practice a maximum on the transmitted power of a mobile. We indicate however (Remark 5) that effects of non-ideal power control can be included in our model by a simple transformation of the problem's parameters. We restrict in the beginning to a single cell for which we obtain an explicit expression for the capacity of best-effort traffic. We further study combination of best-effort with real-time (non best-effort) call classes. We finally discuss the extensions to the multi-cell case.

The paper is structured as follows. We begin by presenting at section 2 known concepts of capacity as well as a new concept adapted for best-effort classes. We also show in this section that the capacity of a system with homogeneous quality of service increases when throughputs are slowed down. We then compute the best-effort capacity at Section 3 which is related to Erlang capacity with very low input transmission rates. We proceed in Section 4 to study a system containing both real-time as well as best-effort call classes. We discuss the multi-cell case in Section 5. We end with a section that presents some perspectives and conclusions.

2. DEFINITION OF CAPACITIES

We begin by introducing in Subsection 2.1 capacity notions for a fixed transmission rate assignment, with a fixed number of mobiles of each class. Based on that, we introduce in Subsection 2.2 an extension of the well known Erlang capacity to the multiclass situation and show how it is computed. In Subsection 2.3, we study the influence of the transmission rates of best effort connections on the Erlang capacity for the case of homogeneous quality of service.

Note that lowering transmission rates has two conflicting impacts: on the one hand, less resources are needed at a given time for handling the connections whose transmission rates were slowed down. On the other hand (and in contrast to real time connections), the transmission duration of such calls becomes longer, since the amount of information to be transmitted does not depend on the transmission rate. We show that the overall result of these competing effects is still that slowing the transmission rates increases capacity. This suggests that there is a limit capacity that can be approached when slowing down the transmission rates.

We identify this limit in Section 2.4 for the single-class case, and show that it has the following property: as long as the arrival rate of information is below some level, blocking probability can become arbitrarily small by sufficiently slowing down the transmission rates (this is similar to the behavior of the well known Shannon capacity in information theory, in which we can get an arbitrarily low error probability if we use sufficiently long codes, as long as we transmit at a rate below the capacity). The multi-class case is then discussed in the following section.

2.1 The case of a fixed number of mobiles

Consider uplink power control of a multi-service CDMA system. Consider an arbitrary sector within some arbitrary cell. We consider a set $K = \{1, ..., k\}$ of best-effort service classes. Example of best-effort applications are file transfers, voice mails, fax.

Let $M(s)$ be the number of ongoing calls of class s which are active, and let $\mathbf{M} = (M(1), ..., M(k))$ be the vector of number of active mobiles.

We assume that when a fixed vector $\mathbf{M}$ is given, the following standard equation [9] is used to determine the power $P(s)$ that should be received at the base station from mobile $s \in K$

$$\frac{P(s)}{N + I_{own} + I_{other} - P(s)} = \tilde{\Delta}(s), \quad s = 1, ..., k \qquad (1)$$

where N is background noise, I_{own} is the total power received from mobiles within the considered sector, and I_{other} is the total power received from mobiles within other sectors and other cells. $\tilde{\Delta}(s)$ is the target ratio of the received power from mobile of class s to the total interference energy received at the base station, and is given by

$$\tilde{\Delta}(s) = \frac{E(s)}{W N_o} R(s).$$

Here, $E(s)$ is the energy corresponding to a transmitted bit of type s, N_o is the thermal noise density, W is the spread-spectrum bandwidth and $R(s)$ is the transmission rate (in bits/s) of class s service.

Remark 1. Note that we implicitly assume that the target value $(E(s)/N_o)$ does not depend on the transmission rate $R(s)$. This is a standard assumption in the literature, see e.g. [7, 8]. In practice, however, it may depend on $R(s)$, see e.g. [4, p. 151, 222, 239]. But as we see from [4, Fig. 10.4, p. 222], it is close to a constant throughout long range of bit rates. For example, between 16Kbps and 256Kbps, the maximum variation around the median value is less than 20%. We thus propose to take for the value of $(E(s)/N_o)$ its average or median value over the range of interest. However, if the exact dependence is available analytically, it can be included into our model.

Remark 2. Note that the required quality of service reflected through $\tilde{\Delta}(s)$ is not only a function of the application but also depends on the transport layer. For example, to achieve a reliable file transfer, typically TCP/IP protocol is used at the transport layer which can support packet losses (due to errors or to congestion) of a few percent by resorting to retransmission of lost packets. The same application may thus be transmitted using different levels of power per bit (which would correspond to different classes in our modeling) depending on the transport protocols used.

We have

$$I_{own} = \sum_{j=1}^{k} M(j) P(j). \qquad (2)$$

To model inter-cell interference we make the standard simplifying assumption [9] that

$$I_{other} = i \times I_{own} \qquad (3)$$

for some given constant i which is obtained from measurements.

We shall find it more useful to rewrite (1) as

$$\frac{P(s)}{N + I_{own} + I_{other}} = \Delta(s), \qquad (4)$$

where

$$\Delta(s) = \frac{\tilde{\Delta}(s)}{1 + \tilde{\Delta}(s)} \Longleftrightarrow \tilde{\Delta}(s) = \frac{\Delta(s)}{1 - \Delta(s)},$$

$s = 1, ..., k$. Solving the set of k equations (4) yields

$$P(s) = \frac{N\Delta(s)}{1 - (1+i)\sum_{j=1}^{k} M(j)\Delta(j)}. \tag{5}$$

(The solution is in particular simple to obtain, since by multiplying the left side of (4) by $M(s)$ and summing over s we get a single equation with the single unknown $\sum_{j=1}^{k} M(j)P(j)$. This then provides immediately the values of all the $P(s)$.)

The pole capacity of the system can be defined as the polyhedron M^* of vectors $\mathbf{M}$ that make the denominator of (5) vanish. It is thus given by

$$M^* = \{\mathbf{M} : 1 = (1+i)\sum_{j=1}^{k} M(j)\Delta(j)\}.$$

We say that $\mathbf{M}_1 < \mathbf{M}_2$ (in the Pareto sense) if $M_1(j) \leq M_2(j)$ for all $j = 1, ..., k$, with strict inequality for at least one j. It is easily seen that the solution $P(s)$ of (5) is finite if and only if $\mathbf{M} < m$ for some $m \in M^*$.

We may slightly change the above definition so as to take into account that $M(j)$ are in practice integer numbers ($\mathbf{M}$ belongs to $\mathbb{N}^k$).

Definition 1. Let $\mathcal{M}$ be the finite subset of $\mathbb{N}^k$ for which $1 > (1+i)\sum_{j=1}^{k} M(j)\Delta(j)$, and let

$$\eta = \max_{m \in \mathcal{M}} (1+i)\sum_{j=1}^{k} m(j)\Delta(j). \tag{6}$$

We define the Integer Capacity $\mathcal{M}_B$ of the system as the boundary of $\mathcal{M}$ for which any additional call would result in an infinite power assignment in (5), or equivalently the set of $\mathbf{M}$'s for which

$$\eta = (1+i)\sum_{j=1}^{k} M(j)\Delta(j).$$

Remark 3. We note that the power corresponding to one service is finite in (5) if and only if it is finite for all services. Thus if a call admission control is not used to avoid exceeding the Integer Capacity of the system then the result is harmful not only for the call accepted beyond capacity but also for all other ongoing calls.

We conclude with another useful definition.

Definition 2. The blocking set $\mathcal{M}_B^j$ of class $j \in K$ is defined as the subset of $\mathcal{M}$ for which another call from class j cannot be accepted, i.e. $m \in \mathcal{M}_B^j$ if and only if

$$m \in \mathcal{M} \text{ and } m + e_j \notin \mathcal{M},$$

where e_j is the unit vector in direction j.

2.2 Random number of mobiles: a single cell

We consider the case of a single isolated cell consisting of a single sector (i.e. $i = 0$). We now make standard statistical assumptions [17]: calls of class s arrive according to a Poisson process with intensity $\lambda(s)$, and their duration is exponentially distributed with parameter $\mu(s)$. Denote $\lambda = \lambda(s)$. Let $\rho(s) := \lambda(s)/\mu(s)$ be the load of class s calls. At this point we assume that all calls are always active.

Remark 4. If this were not the case and a call of class s were active with probability p_s, our model could still be useful if we focus on the process of arrival of active periods of calls, and assume that it can be modeled as an M/M/S/S system. Note that in that case, the blocking events will correspond to blocking of an active period rather than of the whole session. Alternatively, one may introduce an activity factor; we mention a method for transforming a problem with activity factor to ours in Remark 5. An example of a best-effort application with activity and inactivity periods is the HTTP1.1 [10].

We note that the vector of number of calls is an irreducible ergodic finite Markov chain whose state space is given by $\mathcal{M}$. Let $\pi_\rho(\mathbf{M})$ denote the steady state probability of this Markov chain.

Extending the definition of Erlang capacity to our multiservice case we have:

Definition 3. Define the Erlang capacity $EC(\epsilon)$ as the set of vectors $\rho = (\rho(1), ..., \rho(k))$ such that the corresponding blocking probability $P_B(\rho)$ is smaller than a given ϵ.

The Erlang capacity is thus a set instead of a single constant.

THEOREM 1. *The steady state probabilities of the Markov chain are given by*

$$\pi_\rho(\mathbf{M}) = \frac{1}{G_\rho}\prod_{s=1}^{k} \frac{\rho(s)^{M(s)}}{M(s)!}, \quad \mathbf{M} \in \mathcal{M},$$
$$\text{where } G_\rho = \sum_{m \in \mathcal{M}} \prod_{s=1}^{k} \frac{\rho(s)^{m(s)}}{m(s)!}. \tag{7}$$

The probability $P_B^s(\rho)$ that an arriving call of class s is blocked and the average global blocking probability $P_B(\rho)$ are given by

$$P_B^s(\rho) = \sum_{m \in \mathcal{M}_B^s} \pi_\rho(m), \quad P_B(\rho) = \sum_{s=1}^{k} \frac{\lambda(s)}{\lambda} P_B^s.$$

PROOF. The statements of the theorem are very similar to [8] and references therein (see also [11]) and the proof can thus directly be obtained from those references. For completeness we sketch the basic idea. The Markov chain has the same structure as that used in multiclass loss systems [6, 14] without power control, in which there is a set K of classes of calls with the same distribution of arrivals and durations of calls as in our case, in which there is a total given bandwidth of η units, and in which class $i \in K$ requires an amount $\Delta(i)$ of bandwidth. We thus interpret the target ratios $\Delta(i)$ as bandwidth requirement in the equivalent loss model and obtain the steady state probabilities. Since the arrival processes are Poisson, the distribution upon arrival equals to the steady state distribution (the so called PASTA property, see e.g. [18]), which provides the formulas for the blocking probabilities. $\square$

Remark 5. In [8] (and references therein) it is shown in fact that the general form of the steady state probabilities in Theorem 1 extends also to the case of non-perfect power control and to the case where the transmission of a mobile may be interrupted by silence periods. To handle the latter case in our framework, we may proceed as in [8] (see also [17]) and introduce the activity factor $\alpha(s)$ (between 0 and 1). Then $M(j)$ in (2) is replaced by $\alpha(j)M(j)$. The impact on the capacity would be to multiply in (6) each $m(j)$ by $\alpha(j)$. We see from the capacity definition (Definition 1) that the capacity (and hence blocking probabilities) of a system with activity factors $\alpha(j), j = 1, ..., k$, is equivalent to the capacity of the original model we analyze (without activity factors) but where each $\Delta(j)$ is replaced by $\alpha(j)\Delta(j)$. Also, if we use the approach of [8] we may incorporate the imperfect power control in our model by further multiplying the $\Delta(j)$'s by some constants that depend on the standard deviation of the received signal to interference ratio. We conclude that in spite of the simplicity of our model, it can be used in fact to study also non-perfect power control and non constant transmission rates. A more precise way to include both the activity factor and imperfect power control can be done as in [17] but it does not lead to analytical expressions.

2.3 The case of homogeneous quality of service

In order to illustrate the impact of transmission rates on capacity, we introduce the special case in which all classes have a common value of $\Delta = \Delta(s)$, $s = 1, ..., k$. The arrival process of calls of class s is Poisson with parameter $\lambda(s)$ as before, yet the call duration distribution of a class s call may have a general distribution G_s with mean $\mu(s)^{-1}$. The system evolution is equivalent to that of a single class with arrival rate of $\lambda = \sum_{j=1}^{k} \lambda(j)$ and where the distribution of the duration of an arrival, G, is according to G_s with probability $\lambda(s)$. Thus the expected duration of the arrival is

$$\mu^{-1} = \sum_{j=1}^{k} \frac{\lambda(j)}{\lambda} \mu(j)^{-1}.$$

We can thus define the load of the system as

$$\rho = \frac{\lambda}{\mu} = \sum_{j=1}^{k} \rho(j).$$

The integer capacity of the system is given by

$$\mathcal{M}_B = \max\{m \in N : m\Delta < 1\} = \lceil \Delta^{-1} \rceil - 1 \qquad (8)$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to x. Due to the general distribution of the duration of calls, the process of number of calls in the system, taking values in $\mathcal{M} = \{1, ..., \mathcal{M}_B\}$ is no more a Markov chain. Still we have the following:

Theorem 2. *The steady state probabilities of the Markov chain are given by*

$$\pi_\rho(M) = \frac{\rho^M/M!}{\sum_{m\in\mathcal{M}} \rho^m/m!}, \qquad M \in \mathcal{M}.$$

The probability that an arriving call of class s is blocked does not depend on s and is given by $P_B(\rho) = \pi_\rho(\mathcal{M}_B)$.

Proof. Similar to the proof of Theorem 1. The process of number of calls has the same structure as that used in loss systems [6, 14] without power control, in which there is a total given bandwidth of $\mathcal{M}_B$ and in which each call requires an amount of one unit bandwidth. In this equivalent system, the steady state distribution is known to be insensitive to the call duration distribution (it only depends on its expectation), see [15]. Since the arrival processes are Poisson, the distribution upon arrival equals to the steady state distribution (the so called PASTA property, see e.g. [18]), which provides the formulas for the blocking probabilities. □

Remark 6. One can even further relax the statistical assumptions under which Theorem 2 holds, by allowing the call durations to be general stationary ergodic. This means that the duration of successive calls need not be independent. The Theorem then follows from the insensitivity result of [3]. Dependence between durations of successive calls may be useful especially when a call represents in fact an active period rather than a whole connection. For example, a single HTML application may contain several successive file transfers whose durations may be correlated.

We shall assume below that Δ^{-1} (or equivalently $\tilde{\Delta}^{-1}$) is an integer. We show the impact of the assigned throughput. Assume that all classes are best-effort. Keeping the same value $E(s)$ of energy per bit for all classes, we slow the transmission rate of all classes by dividing them by a constant $a > 1$. In other words, we replace $R(s)$ by $R^a(s) = R(s)/a$. As a consequence $\tilde{\Delta}$ as well as the call durations are divided by a. Hence with the slower transmission rates we get a new load $\rho_a = \rho a$. From (8) it then follows that the new integer capacity is

$$\mathcal{M}_{B,a} = \lceil a/\Delta \rceil - 1.$$

Note that for integer values of Δ^{-1} and a we have

$$\mathcal{M}_{B,a} = \lceil \frac{a}{\Delta} \rceil - 1 = \lceil \frac{a + \tilde{\Delta}}{\tilde{\Delta}} \rceil - 1 = \frac{a}{\tilde{\Delta}} = a\mathcal{M}_B. \qquad (9)$$

The steady state probabilities for the new system are

$$\pi_{\rho,a}(M) = \frac{(a\rho)^M/M!}{\sum_{m=1}^{\mathcal{M}_{B,a}} (a\rho)^m/m!}, \qquad M = 0, ..., \mathcal{M}_{B,a}.$$

Next we show the influence of a on the blocking probability and the Erlang capacity.

Theorem 3. *As a increases, the blocking probability decreases.*

Proof. Let X be the state (global number of calls) of the initial system and Y the state of the new one. Define

$$Z = \max(0, Y - \mathcal{M}_{B,a} + \mathcal{M}_B).$$

Thus Z takes values in $\{0, 1, ..., \mathcal{M}_B\}$. For a random variable V defined on $\mathcal{M}$ we define $r_V(0) = 0$, and

$$r_V(m) = \frac{P(V = m - 1)}{P(V = m)}, \quad m = 1, ..., \mathcal{M}_B.$$

We have for $m > 0$

$$\begin{aligned} r_X(m) &= \frac{m}{\rho}, \\ r_Z(m) &= \frac{m + \mathcal{M}_{B,a} - \mathcal{M}_B}{a\rho} = \frac{m + \lceil a/\Delta \rceil - \Delta^{-1}}{a\rho} \\ &\geq \frac{m + (a-1)/\Delta}{a\rho} \end{aligned}$$

Thus

$$r_X - r_Z(m) \le \frac{(a-1)(m-\Delta^{-1})}{a\rho} \le 0.$$

It then follows (see [13] and references therein) that $Z \le_{st} X$ or equivalently $E[f(Z)] \le E[f(X)]$ for any nondecreasing function f. In particular, by taking f to be the indicator $f(x) = 1\{x = \mathcal{M}_{B,a}\}$ we conclude that

$$P(X = \mathcal{M}_B) \ge P(Z = \mathcal{M}_B) = P(Y = \mathcal{M}_{B,a})$$

from which we conclude that the blocking at the original system indeed has greater probability than in the system a. $\square$

If we consider the homogeneous case as an equivalent system with a single class (as in the first paragraph of the Subsection), then the fact that the blocking probability decreases with a implies that the Erlang capacity of the one-dimensional system increases (we use the fact that for fixed a, the blocking probability increases with the input rate, see [13, 12]).

2.4 Best-effort Capacity

We now define the capacity in systems with best-effort applications. To motivate our definition, we go back to the homogeneous model of Subsection 2.3. We showed there that the blocking probability decreases as the transmission rate decreases. We now show through a simple calculation that if $\rho < \mathcal{M}_B$ then the blocking probability tends to zero as a tends to infinity. This will then motivate us to introduce a definition of capacity which, in contrast to the Erlang capacity, does not depend on a given parameter ϵ. It will be more related to the Shannon capacity concept.

THEOREM 4. *Consider the homogeneous system introduced in Subsection 2.3, and let $\rho < \mathcal{M}_B$. Then the blocking probability tends to zero as a tends to infinity.*

PROOF. Choose any $\epsilon > 0$, and choose

$$m \ge \max(1/\epsilon, \mathcal{M}_B).$$

Denote

$$n_0(\epsilon) = m/(\mathcal{M}_B - \rho).$$

We have for any integer $a > n_0$.

$$\begin{aligned}
P_B^a &= \frac{\rho_a^{\mathcal{M}_{B,a}}/\mathcal{M}_{B,a}!}{\sum_{j=0}^{\mathcal{M}_{B,a}} \rho_a^j/j!} \le \frac{(a\rho)^{\mathcal{M}_{B,a}}/(\mathcal{M}_{B,a})!}{\sum_{j=\mathcal{M}_{B,a}-m}^{\mathcal{M}_{B,a}} (a\rho)^j/j!} \\
&= \Big(\frac{\mathcal{M}_{B,a}(\mathcal{M}_{B,a}-1)(\mathcal{M}_{B,a}-2)\cdots(\mathcal{M}_{B,a}-m)}{(a\rho)^m} \\
&\quad + \cdots + \frac{\mathcal{M}_{B,a}(\mathcal{M}_{B,a}-1)(\mathcal{M}_{B,a}-2)}{(a\rho)^3} \\
&\quad + \frac{\mathcal{M}_{B,a}(\mathcal{M}_{B,a}-1)}{(a\rho)^2} + \frac{\mathcal{M}_{B,a}}{a\rho} + 1\Big)^{-1} \\
&= \Big(\frac{\mathcal{M}_B(\mathcal{M}_B - \frac{1}{a})(\mathcal{M}_B - \frac{2}{a})\cdots(\mathcal{M}_B - \frac{m}{a})}{\rho^m} + \cdots \\
&\quad + \frac{\mathcal{M}_B(\mathcal{M}_B - \frac{1}{a})(\mathcal{M}_B - \frac{2}{a})}{\rho^3} \\
&\quad + \frac{\mathcal{M}_B(\mathcal{M}_B - \frac{1}{a})}{\rho^2} + \frac{\mathcal{M}_B}{\rho} + 1\Big)^{-1} < \frac{1}{m} \le \epsilon.
\end{aligned}$$

The last equality follows from (9). The inequality before the last follows since for any $0 \le j \le m$ and $a > n_0(\epsilon)$ we have $\mathcal{M}_B - j/a \ge \rho$. This establishes the proof. $\square$

Remark 7. Note that the above proof shows that for any given ϵ, if we slow down the transmission rates by a factor larger than $n_0(\epsilon)$ then the blocking probability is smaller than the given ϵ. An alternative simpler proof that does not give a bound on rate of convergence is as follows. Let $Y(a)$ be a Poisson random variable with parameter $a\rho$. Then

$$P_B^a = \frac{P(Y(a) = \mathcal{M}_{B,a})}{P(Y(a) \le \mathcal{M}_{B,a})} = \frac{P\Big(Y(a)/a = \mathcal{M}_B\Big)}{P\Big(Y(a)/a \le \mathcal{M}_B\Big)} \tag{10}$$

(where we use (9)). Let $Y_s, s = 1, ..., a$ be i.i.d. Poisson random variables with parameter ρ. Then $Y(a)$ has the same distribution as $\sum_{s=1}^a Y_s$. Due to the strong law of Large numbers, $Y(a)/a$ converges P-almost surely to its expectation, ρ, as $a \to \infty$. Since $\rho < \mathcal{M}_B$, this implies that the enumerator of (10) converges to zero and the denominator to 1 as $a \to \infty$, which establishes the proof.

In order to define the best-effort capacity, we need some more definitions. Consider a system where all k classes are best-effort classes, where the amount of data that a call of class s has to transmit has an exponentially distributed size with parameter $\zeta(s)$ (its expected size is $1/\zeta(s)$) and the arrival rates of calls of the classes are given by the vector $\lambda = (\lambda(1), ..., \lambda(k))$. Let the assigned transmission rate of class s calls be $R(s)$, and denote $\mathbf{R} = (R(1), ..., R(k))$. Then the transmission time of class s is an exponentially distributed random variable with parameter $\mu(s) = R(s)\zeta(s)$ (the expected call duration is thus $(\zeta(s)R(s))^{-1}$). Define the utilization density $\nu(s) = \lambda(s)/\zeta(s)$, and define the vector $\nu = (\nu(1), ..., \nu(k))$. Define

$$\delta(s) = \frac{E(s)}{WN_o}$$

so that

$$\tilde{\Delta}(s) = R(s)\delta(s) \text{ and } \Delta(s) = \frac{R(s)\delta(s)}{1 + R(s)\delta(s)},$$

and let $\delta = (\delta(1), ..., \delta(k))$.

Definition 4. Consider the system described above with a given δ. Define the BE (best-effort) capacity as the supremum of the set of vectors $\nu = (\nu(1), ..., \nu(k))$ for which for any $\epsilon > 0$, there exists a vector $\mathbf{R} = (R(1), ..., R(k))$ of transmission rates such that $P_B^s(\nu, \mathbf{R}) < \epsilon$ for all $s = 1, ..., k$.[1]

Numerical Example:
Consider a single class homogeneous system as described in Section 2.3, that handles best-effort sources and offers them a high-speed connection, i.e. a large transmission rate of $R = 160KB/sec$ (i.e. 1.28 Mbps). Let $\Delta = 0.199$ so that the system is dimensioned such that its integer capacity is 5, i.e. no more than 5 calls can be simultaneously handled. Assume

[1]The supremum is taken in the Pareto sense, i.e. a vector $\mathbf{n}$ belongs to a supremum set satisfying a property, if there is no other vector larger (in the Pareto sense) than $\mathbf{n}$ which satisfies that property, and if for any $\epsilon' > 0$ there exists a vector ν satisfying the property such that $\nu(s) \ge n(s) - \epsilon'$ for all $s = 1, ..., k$.

that the average amount of information (e.g. the average file size) of a connection is 10KB (the average size of files transfered on the Internet is known to be between 8-12 KB, see [16] and references therein). The average duration needed for handling a session is $\mu^{-1} = \frac{10KB}{160KB/sec} = 62.4msec$, and $\mu = 16.03$.

Assume that we wish that the blocking probability be inferior to 1%. Using Theorem 2 we see that the Erlang capacity of the system is $\rho = 1.361$, which means that for having at most 1% of losses the rate of arrival of sessions should be limited to $\lambda = \rho\mu = 21.8$ calls per second.

Table 1 shows the gain by slowing the transmission rates by a factor of a. For each a it gives the Erlang Capacity EC(1%) as well as the rate λ of arriving calls that the system can handle without exceeding 1% of blocking. In particular, we see that we double the capacity by slowing the transmission rates by a factor of around five. This could indicate that among connections that have the same volume of information to transmit, connections that are five time slower use half of the effective amount of resources than the others. Hence if we were to assign prices per volume of information transmitted as well as of the speed transmitted, the slower connections could be priced the half per transmitted volume than the original ones. The Best-effort capacity for this problem is $\rho = 5$.

3. COMPUTING THE BEST EFFORT CAPACITY

Focusing on the single class, we showed in the previous Section that as long as the arrival rate of information is below some level, blocking probability can become arbitrarily small by sufficiently slowing down the transmission rates. In this section we establish the same result for the multi-class case.

Theorem 5. *The BE Capacity of the system (with Poisson arrivals with rate vector λ and with sizes of calls exponentially distributed with (vector) parameter ζ) is given by the set of ν satisfying*

$$\sum_{s=1}^{k} \nu(s)\delta(s) = 1. \tag{11}$$

The proof is based on two parts.

Lemma 1. *If ν satisfies $\sum_{s=1}^{k} \nu(s)\delta(s) > 1$ then the blocking probabilities $P_B^s(\mathbf{R})$ for classes $s = 1, ..., k$ satisfy for any $\mathbf{R}$*

$$\sum_s P_B^s \delta(s)\nu(s) > \sum_s \delta(s)\nu(s) - 1 > 0.$$

Proof. Assume that for some ν, $\sum_{s=1}^{k} \nu(s)\delta(s) > 1$. Choose an arbitrary $\mathbf{R}$. Define the following random processes, taken to be right continuous with left limits:
$M_t(s)$:= the number of s-type calls at time t,
$A_t(s)$:= the number of arrivals of s-type calls till time t,
$D_t(s)$:= the number of s-type calls that ended till time t,
$B_t(s)$:= the number of s-type calls that have been blocked by time t.

Then we have

$$M_t(s) = M_0(s) + A_t(s) - D_t(s) - B_t(s). \tag{12}$$

Let $\mu(s) = R(s)\zeta(s)$. Note that at time t, the departure rate of class s has a stochastic intensity of $M_t(s)\mu(s)$. The loss rate of class s calls at time t is given by $\lambda(s)P_B^s(t)$ where $P_B^s(t)$ is the probability that an arriving call of type s is blocked at time t, i.e. that $\mathbf{M}_t \in \mathcal{M}_B^s$. Differentiating (12) and taking expectations we thus obtain

$$\frac{dE[M_t(s)]}{dt} = \lambda(s) - \mu(s)E[M_t(s)] - \lambda(s)P_B^s(t). \tag{13}$$

The process $\mathbf{M}_t$ being a finite irreducible Markov chain, converges to a steady state distribution, for which the expected time derivative vanishes above. Multiplying (13) by $\delta(s)/\zeta(s)$, taking the sum over the k classes and omitting t from the notation (to indicate that we are at steady state), we obtain

$$\begin{aligned}\sum_{s=1}^{k} P_B^s \delta(s)\nu(s) &= \sum_{s=1}^{k} \Big(\delta(s)\nu(s) - \tilde{\Delta}(s)E[M(s)]\Big) \\ &> \sum_{s=1}^{k} \delta(s)\nu(s) - 1\end{aligned}$$

which establishes the proof. (We used the fact that

$$\frac{\delta(s)\mu(s)}{\zeta(s)} = \tilde{\Delta}(s) > \Delta(s)$$

and that $\sum_s \Delta(s)M(s)$ cannot exceed 1.) □

The above Lemma gives a lower bound on the blocking probability when $\sum_{s=1}^{k} \nu(s)\delta(s) > 1$. For the homogeneous case this gives, in particular,

$$P_B > \frac{\sum_{s=1}^{k} \delta(s)\nu(s) - 1}{\sum_{s=1}^{k} \delta(s)\nu(s)}.$$

Next, we make the following observation on the way $\Delta(s)$ scales with a. We have for $a \geq 1$,

$$\begin{aligned}\frac{\tilde{\Delta}(s)}{a} &\geq \Delta_a(s) = \frac{\tilde{\Delta}(s)}{a + \tilde{\Delta}(s)} = \left(1 + \frac{a}{\tilde{\Delta}(s)}\right)^{-1} \\ &\geq \frac{1}{a}\left(1 + \frac{1}{\tilde{\Delta}(s)}\right)^{-1} = \frac{\Delta(s)}{a}.\end{aligned} \tag{14}$$

Lemma 2. *Consider the system with fixed vectors δ and ν, and with an arbitrary given rate vector $\mathbf{R}$ with positive components. Assume that $\sum_{s=1}^{k} \nu(s)\delta(s) < 1$. Then the limit as $a \to \infty$ of the blocking probability when $\mathbf{R}$ is divided by a is zero.*

Proof. Let Y_s be independent Poisson random variables with parameters $\rho(s)$, $s = 1, ..., k$, and let $Z = \sum_{j=1}^{k} \Delta(j)Y_j$. Then the steady state probabilities (7) can be rewritten as

$$\pi_\rho(\mathbf{M}) = \frac{\prod_{j=1}^{k} P(Y_s = M_s)}{P(Z \leq \eta)} \qquad \mathbf{M} \in \mathcal{M},$$

where η is given in (6). Then the blocking probability satisfies

$$P_B^s(\rho) \leq \frac{P(Z \geq 1 - \Delta(s))}{P(Z < 1)} = \frac{P(Z \geq 1 - \Delta(s))}{1 - P(Z \geq 1)}. \tag{15}$$

Next we use Chebycheff bound for Z; for any positive real numbers α and β we have:

$$P(Z \geq \alpha) \leq \frac{E[\exp(\beta Z)]}{\exp(\beta\alpha)}. \tag{16}$$

Slowing factor a	1	2	3	4	5	6	20
Δ	0.199	0.110	0.0764	0.0583	0.0473	0.0398	0.0122
$\tilde{\Delta}$	0.248	0.124	0.0827	0.0620	0.0497	0.0414	0.0124
$\mathcal{M}_{B,a}$	5	9	13	17	21	25	81
Erlang Cap. EC(1%)	1.361	1.891	2.202	2.413	2.568	2.688	3.315
Arrival rate λ	21.8	30.3	35.29	38.67	41.15	43.08	53.13
Average call duration in msec	62.4	124.8	187.2	249.6	312	374	1248
Gain in %	0	43.7	67.4	83.4	95	104	144

Table 1: Gain in Erlang capacity by slowing transmission rates by a factor of a

We recall that the PGF of the Poisson random variable Y_s is given by $E[\xi^{Y_s}] = \exp(\rho(s)(\xi - 1))$. It then follows that

$$\begin{aligned} E[\exp(\beta Z)] &= \prod_{s=1}^{k} E\left[\exp\Big(\beta\Delta(s)Y_s\Big)\right] \\ &= \prod_{s=1}^{k} \exp\Big(\rho(s)(\exp[\beta\Delta(s)] - 1)\Big). \end{aligned}$$

Choose an arbitrarily small $\epsilon > 0$ and let $\beta > 0$ be such that

$$\exp[\beta\Delta(s)] - 1 \le \beta\Delta(s)(1 + \epsilon).$$

Then we obtain for (16):

$$P(Z \ge \alpha) \le \exp\Big(\beta\Big[(1+\epsilon)\sum_{s=1}^{k}\rho(s)\Delta(s) - \alpha\Big]\Big).$$

Thus

$$P_B(\rho) \le \frac{\exp\Big(\beta\Big[(1+\epsilon)\sum_{s=1}^{k}\rho(s)\Delta(s) - 1 + \Delta(s)\Big]\Big)}{1 - \exp\Big(\beta\Big[(1+\epsilon)\sum_{s=1}^{k}\rho(s)\Delta(s) - 1\Big]\Big)} \tag{17}$$

We now divide all transmission rates by $a > 1$.

Then the new bound derived from (17) for the blocking probabilities is obtained by defining $Z^a = \sum_{j=1}^{k}\Delta(j)Y_j(a)$ where $Y_j(a)$ have Poisson distributions with parameters $a\rho(s)$. We obtain

$$\begin{aligned} & P_{B,a}^{s}(\rho) \\ \le\; & \frac{\exp\Big(\beta\Big[a\{(1+\epsilon)\sum_{s=1}^{k}a\rho(s)\Delta_a(s) - 1\} + \Delta_a(s)\Big]\Big)}{1 - \exp\Big(\beta a\Big[(1+\epsilon)\sum_{s=1}^{k}a\rho(s)\Delta_a(s) - 1\Big]\Big)} \\ \le\; & \frac{\exp\Big(\beta\Big[a\{(1+\epsilon)\sum_{s=1}^{k}\rho(s)\tilde{\Delta}(s) - 1\} + \Delta(s)\Big]\Big)}{1 - \exp\Big(\beta a\Big[(1+\epsilon)\sum_{s=1}^{k}\rho(s)\tilde{\Delta}(s) - 1\Big]\Big)} \end{aligned} \tag{18}$$

where the last inequality follows from (14). Condition (11) implies that

$$(1+\epsilon)\sum_{s=1}^{k}\rho(s)\tilde{\Delta}(s) - 1 < 0$$

for all ϵ sufficiently small, which implies that $P_{B,a}^{s}$ tends to zero as $a \to \infty$. $\square$

Remark 8. Note that the above proof provides also a bound on the rate of convergence of the BE to 0 as $a \to \infty$. A more direct proof that does not provide a rate of convergence can be proposed by extending the approach in Remark 7.

4. COMBINED REAL-TIME AND BEST EFFORT APPLICATIONS

In this section we examine the situation in which we can slow down the transmission time of only some classes (of best-effort traffic); other classes that may correspond to real time applications transmit at a fixed rate[2].

More precisely, we consider ℓ BE classes enumerated by $1, ..., \ell$ whose transmission we may slow down and a set of real-time classes: $\ell + 1, ..., k$ with fixed transmission rate.

For each parameter a one can use (7) for computing blocking probabilities and use it then to compute the capacity. However, pursuing our approach from previous sections we proceed to compute the limiting behavior of the system as the throughput assigned to best-effort classes is slowed down. We shall show that slowing the throughputs of best-effort traffic improves their performance.

Unlike the case of best-effort traffic only, we cannot expect the global blocking probabilities to vanish as transmission rates of best-effort traffic are slowed down for any values of $\delta(s)$ and $\zeta(s)$, as long as there is positive probability of arrivals of real-time traffic. Yet we shall show that for a given set of parameters of best-effort calls, their blocking probabilities can be made arbitrarily small by slowing sufficiently their transmission rate.

THEOREM 6. *If $\sum_{j=1}^{\ell}\nu(j)\delta(j) \ge 1$ then at steady state, all RT calls are blocked, and the system is thus equivalent to one with no RT traffic.*
(1) Assume that $\sum_{j=1}^{\ell}\nu(j)\delta(j) < 1$. Define

$$\begin{aligned} \tilde{\mathcal{M}} &= \Big\{(m(\ell+1), ..., m(k)) : \\ & \qquad \sum_{j=1}^{\ell}\nu(j)\delta(j) + \sum_{j=\ell+1}^{k}\Delta(j)m(j) < 1\Big\}, \\ \tilde{\mathcal{M}}_B^s &= \Big\{(m(\ell+1), ..., m(k)) \in \tilde{\mathcal{M}} : \\ & \qquad \sum_{j=1}^{\ell}\nu(j)\delta(j) + \sum_{j=\ell+1}^{k}\Delta(j)m(j) \ge 1 - \Delta(s)\Big\}. \end{aligned}$$

[2]In practice also video and voice applications may be transmitted with a lower throughput using various compression mechanisms. We do not treat this possibility here, we already assume that if different possible throughputs are available for the real-time traffic then the ones used correspond to the required quality of these applications (of course larger compression rates result in lower quality). Note that slowing the throughput of real-time applications by a factor of a does not result in a longer call duration (unless the call has such a bad quality that speakers have to repeat entire phrases. This is not an interesting case from a system design point of view).

(2) Assume that for all $s = \ell+1, ..., k$ and for all $\mathbf{m} \in \tilde{\mathcal{M}}_B^s$ we have

$$\sum_{j=1}^{\ell} \nu(j)\delta(j) + \sum_{j=\ell+1}^{k} \Delta(j)m(j) > 1 - \Delta(s). \qquad (19)$$

Then the limiting steady-state and blocking probability of a real-time class are given by Theorem 1 where $\mathcal{M}$ is replaced by $\tilde{\mathcal{M}}$, $\mathcal{M}_B^s$ by $\tilde{\mathcal{M}}_B^s$ and where we restrict summations to classes $\ell+1, ..., k$. The limiting blocking probabilities of all BE classes are zero as $a \to \infty$.

PROOF. Let Y_s be independent Poisson random variables with parameters $a\rho(s)$ for $s = 1$, and with parameter $\rho(s), s = \ell+1, ...k$. Denote

$$Z(a) = \sum_{s=1}^{\ell} \Delta_a(s) Y_s + \sum_{s=\ell+1}^{k} \Delta(s) Y_s.$$

Then the blocking probability $P_{B,a}^s$ of a RT class s when transmission rates of BE calls are slowed down by an integer $a > 1$ is given by

$$P_{B,a}^s = \frac{P(1-\Delta(s) \le Z(a) < 1)}{P(Z(a) < 1)}. \qquad (20)$$

Define $\tilde{Y}_s^r$ to be independent Poisson variables with parameter $\rho(s)$, $r = 1, ..., a$. Then Y_s has the same distribution as $\sum_{r=1}^{a} Y_s^r, s = 1, ..., \ell$. Now, the strong law of large numbers implies that $\sum_{r=1}^{a} Y_s^r/a$ converges in distribution to the constant $\rho(s)$, $s = 1, ..., \ell$. Since $\lim_{a\to\infty} a\Delta_a(s)/\tilde{\Delta}(s) = 1$, $Z(a)$ converges weakly to Z as $a \to \infty$, where

$$Z := \sum_{s=1}^{\ell} \delta(s)\nu(s) + \sum_{s=\ell+1}^{k} \Delta(s) Y_s.$$

Define the sets

$$A_1 = [1-\Delta(s), 1), \ A_2 = [0, 1).$$

(20) implies that

$$P(Z = 1) = P(Z = 1 - \Delta(s)) = 0.$$

Moreover, $P(Z = 0) = 0$. It then follows by Portmanteau's Theorem [1, p. 11],

$$\lim_{a\to\infty} P(Z(a) \in A_1) = P(Z \in A_1),$$

$$\lim_{a\to\infty} P(Z(a) \in A_2) = P(Z \in A_2).$$

Then the enumerator of (20) converges to $P(Z \in A_1)$ and the denominator of (20) converges to $P(Z \in A_2)$. Thus, defining the event

$$\mathcal{A} := \Big\{1 - \sum_{j=1}^{\ell} \delta(j)\nu(j) - \Delta(s) \le \sum_{j=\ell+1}^{k} \Delta(j) Y_j < 1 - \sum_{j=1}^{\ell} \delta(j)\nu(j))\Big\}$$

we have

$$\lim_{a\to\infty} P_{B,a}^s = \frac{P(\mathcal{A})}{P\Big(\sum_{j=\ell+1}^{k} \Delta(j) Y_j < 1 - \sum_{j=1}^{\ell} \delta(j)\nu(j)\Big)} = \frac{P(\mathbf{Y} \in \tilde{\mathcal{M}}_B^s)}{P(\mathbf{Y} \in \tilde{\mathcal{M}})}.$$

The last expression coincides with blocking probabilities of equivalent loss systems from [6, 14] with the $k-\ell$ classes that correspond to RT traffic. Thus this expression can be identified with the probabilities stated in the theorem.

We see that if $\sum_{s=1}^{\ell} \nu(s)\delta(s) \ge 1$ then all real-traffic is blocked at steady-state. The system is thus equivalent to one with no RT traffic, and we can use Lemma 1 to show that BE traffic will suffer a positive loss rate.

Assume now that $\sum_{s=1}^{\ell} \nu(s)\delta(s) < 1$. There is some $\epsilon > 0$ such that

$$\lim_{a\to\infty} P(A_3) = 0$$

where A_3 is the event given by

$$\left\{(m_{\ell+1}, ..., m_k) : \sum_{s=\ell+1}^{k} \Delta(s)m(s) > 1 + \epsilon - \sum_{s=1}^{\ell} \nu(s)\delta(s)\right\}.$$

Hence by taking large a, the probability of A_3 can be made arbitrarily small. Moreover, due to the strong law of large numbers, also

$$P\left(\sum_{s=1}^{\ell} (\tilde{\Delta}(s)/a) M_s > \sum_{s=1}^{\ell} \nu(s)\delta(s) + \epsilon\right)$$

can be made arbitrarily small. Hence the blocking probability of BE converges indeed to 0 as $a \to \infty$ since for BE traffic to be blocked at the system a we need to have

$$\sum_{s=1}^{\ell} \Delta_a(s) M_s + \sum_{s=\ell+1}^{k} \Delta(s) M_s \ge 1 - \Delta_a(s), \quad s = 1, ..., \ell,$$

and since $\lim_{a\to\infty} a\Delta_a(s)/\tilde{\Delta}(s) = 1$. □

5. THE MULTI-CELL CASE

In this section we introduce a method for approximating the blocking probabilities (and thus obtaining the Erlang capacity) for fixed transmission rates for the multi-cell case. Our method is based on a mean field approach and on fixed point arguments; we study in particular the existence and uniqueness of the fixed point. We then present the corresponding BE capacity.

5.1 Blocking probability for the multi-cell case

Our approach is inspired by the approximation used for computing the pole capacity [9] for the multi-cell case, which we already mentioned at (5). We assume a symmetric system of cells. In our stochastic framework of Poisson arrivals of calls and exponential call duration, it is not reasonable to expect that (3) holds at each moment. Instead, we assume that it holds in expectation: $E[I_{other}] = i \times E[I_{own}]$. The mean field approximation amounts on further assuming that the *instantaneous interference from other cells is replaced by its average*: $I_{other} = i \times E[I_{own}]$. We then get instead of (5) the relation for the (random) power of type s call:

$$P(s) = \frac{N\Delta(s)}{1 - \sum_{j=1}^{k} M(j)\Delta(j) - Q} \qquad (21)$$

where $Q = i\sum_{j=1}^{k} E[M(j)]\Delta(j)$ (the randomness comes since here, M is a random variable).

For each fixed value of Q (possibly different than the value $i\sum_{j=1}^{k} E[M(j)]\Delta(j)$), we can obtain the probability distribution of $M(s), s = 1, ..., k$ (under the assumption that calls of class s are blocked whenever the denominator of (21) would vanish or become negative if the call were accepted). More precisely, define $\mathcal{M}(q)$ as the set of $\mathbf{M}$ for which the assigned power according to (21) (with a general parameter q replacing Q) is finite, and let $\mathcal{M}_B^s(q)$ be the blocking set of class s, i.e.

$$\begin{aligned}\mathcal{M}(q) &= \Big\{(m(1),...,m(k)) : \sum_{j=1}^{k} \Delta(j)m(j) < 1-q\Big\},\\ \mathcal{M}_B^s(q) &= \Big\{(m(1),...,m(k)) \in \mathcal{M} :\\ &\qquad \sum_{j=1}^{k} \Delta(j)m(j) \geq 1-q-\Delta(s)\Big\}.\end{aligned}$$

Using the same arguments as those used to derive Theorem 1, we conclude that the steady state probabilities $\pi_\rho(\mathbf{M}, q)$ (for the given parameter q) of $\mathbf{M}$ is given by (7), where $\mathcal{M}$ is replaced by $\mathcal{M}(q)$; the blocking probabilities are also obtained as in Theorem 1. Denote by E_q the expectation operator that corresponds to the probability measure $\pi_\rho(\mathbf{M}, q)$.

Define

$$F(q) = i \sum_{j=1}^{k} E_q[M(j)]\Delta(j).$$

We can characterize Q as the solution of the fixed point equation:

$$q = F(q). \tag{22}$$

Note that $F(q)$ is in fact piecewize constant in q, and has thus discontinuities. This implies that (22) need not have a solution. However, the set of values of i for which a solution to (22) does not exist has Lebesgue measure zero. In other words, a slight change in the value of i will yield a solution.

$F(q)$ is nonincreasing in q which implies uniqueness of the solution to (22). Indeed, let

$$X(q) = \sum_{s=1}^{k} \Delta(s)M(s)$$

in system q. Define $r_q(0) = 0$, and

$$r_q(m) = P_q(X = m-1)/P_q(X = m), \qquad m \in \mathcal{M}(q).$$

The for $q1 < q2$ we have $r_{q1}(m) = r_{q2}(m)$ for all $m \in \mathcal{M}(q2)$. It then follows from the point 1 after Theorem 3 in [13] that $E_{q1}[X] \geq E_{q2}[X]$ which establishes the monotonicity.

5.2 Best-effort capacity for the multi-cell case

Using the above approach, one can now establish the following using similar steps as in Section 3:

THEOREM 7. *The BE Capacity of the multi-cell system (with Poisson arrivals with rate vector λ and with sizes of calls exponentially distributed with (vector) parameter ζ) is given by the set of ν satisfying*

$$(1+i)\sum_{s=1}^{k} \nu(s)\delta(s) = 1. \tag{23}$$

6. CONCLUDING COMMENTS AND PERSPECTIVES

We have studied in this paper the capacity of CDMA systems that handles best-effort traffic whose transmission rate can be determined by the network. We assumed perfect power control which allowed us to obtain explicit expressions for the capacity of the network. It was shown that capacity is in fact approached by slowing transmission rates of best-effort traffic.

We indicated how non-ideal power control can be integrated into our model. In practice, however, close loop power control is typically not implemented for packet transmissions. One reason for that is that the time it takes to transmit a packet may be too short for a feedback control to converge, given that close loop power control is updated around 1500 times per second. However, our findings suggest that the system capacity can be improved by slowing transmission rates. This would make the transmission duration of best-effort traffic longer, which might make closed-loop power control more appropriate. Still, one could possibly restrict our approach to those best-effort applications that are sufficiently long such as fax, long file transfers (in particular video on demand in which a whole video file is transferred), voice-mail, etc.

Throughout our paper, best-effort calls of a given class were assumed to have pre-determined transmission rates (and the question was how to determine them). For large file transfer applications, as mentioned in the previous paragraph, this modeling assumption is quite realistic. Some best effort applications do not have a constant transmission rate, see e.g. http transfers (which contain silence periods). If the instantaneous transmission rates do not depend on the state of the system, the analysis could be handled within our model using a proper transformation, as mentioned in Remark 5. Yet, in other networking contexts, one further allows the instantaneous transmission rates of best-effort applications to depend on the system's state, see for example the ABR (Available Bit Rate) class in ATM networks, or the TCP congestion control in the Internet. We could also consider this additional feature in wireless networks offering integrated services, in order to better use the resources: at low congestion periods we could allow for larger throughputs of best-effort classes which would reduce the duration of such calls. Even within the duration of a call one could consider varying the throughput (especially to avoid dropping of calls). We shall pursue these research directions in the future.

We followed the standard modeling assumption on the arrivals of sessions (see also [11, 8, 17]), assuming that they follow Poisson processes. This implicitly implies that we have an infinite source of connections. An alternative modeling assumption can be to assume a finite population of sources of connections, which would give rise to different blocking probabilities and different expressions for the capacity.

The BE capacity defined here was established by a scaling of the system in which transmission rates were slowed down by a factor a, and consequently, the energy per bit was unchanged, Consequently, the integer-capacity of the system (as opposed to the Erlang capacity) grew linearly in a. Arrival rate of calls, as well as the amount of information to be transmitted were not affected by this scaling. We should mention that an equivalent scaling has been studied in the context of loss systems (without power control) in which the call durations were not changed, the capacity increased by a factor a as well as the arrival rates, see [6, 5]. By taking the limit as a grows to infinity, the trajectories of the system has been shown in [5] to converge to some fluid model.

7. ACKNOLEDGMENT

We wish to thank Mr. Jean-Marc Kelif and Dr. Zwi Altman from France Telecom R&D for many useful discussions.

8. REFERENCES

[1] P. Billingsley. *Convergence of Probability Measures*. Wiley, 1968.

[2] A. Chockalingam, P. Dietrich, L. B. Milstein, and R. R. Rao. Performance of closed-loop power control in DS-CDMA cellular systems. *Trans. on Vehicular Technology*, 47(3):774–789, 1998.

[3] P. Franken, D. Konig, U. Arndt, and V. Schmidt. *Queues and Point Processes*. Akademie-Verlag, Berlin, 1981.

[4] H. Holma and A. Toskala. *WCDMA for UMTS*. Revised Edition, J. Wiley & Sons, 2001.

[5] P. J. Hunt and T. G. Kurtz. Large loss systems. *Stochastic Processes and their Applications*, 53:363–378, 1994.

[6] F. P. Kelly. Loss networks. *The Annals of Applied Probability*, 1(3):319–378, 1991.

[7] S. L. Kim, Z. Rosberg, and J. Zander. Combined power control and transmission selection in cellular networks. In *Proceedings of IEEE Vehicular Technology Conference*, Fall 1999.

[8] I. Koo, J. Ahn, H. A. Lee, and K. Kim. Analysis of Erlang capacity for the multimedia DS-CDMA systems. *IEICE Trans. Fundamentals*, E82-A(5):849–855, May 1999.

[9] J. Laiho and A. Wacker. Radio network planning process and methods for WCDMA. *Ann. Telecoomun.*, 56, 2000.

[10] B. Mah. An empirical model of http network traffic. In *Proceedings of INFOCOM '97*, Kobe, Japan, April 1997.

[11] M. Meo and E. Viterbo. Performance of wideband CDMA systems supporting multimedia traffic. *IEEE Comm. Lett.*, 5:251–253, June 2001.

[12] P. Nain. Qualitative properties of the Erlang blocking model with heterogeneous user requirements. *Queueing Systems*, 6:189–206, 1990.

[13] K. Ross and D. Yao. Monotonicity properties for the stochastic knapsack. *IEEE Transactions on Information Theory*, 36:1173–1179, 1990.

[14] K. W. Ross. *Multiservice Loss Models for Broadband Telecommunication Networks*. Springer-Verlag, Berlin, 1995.

[15] B. A. Sevasatyanov. Limit theorems for Markov processes and their application to telephone loss systems. *Theory Probab. Appl.*, 2:104–112, 1957.

[16] B. Sikdar, S. Kalyanaraman, and K. S. Vastola. An integrated model for the latency and steady-state throughput of tcp connections. *Performance Evaluation*, 46:139–154, 2001.

[17] A. M. Viterbi and A. J. Viterbi. Erlang capacity of a power controlled CDMA system. *IEEE Journal of Selected Areas in Communications*, pages 892–900, 1993.

[18] R. W. Wolff. *Stochastic Modeling and the Theory of Queues*. Prentice Hall, 1989.