Analysis of Delay Constrained Communications over OFDM systems

Beatriz Soret, M. Carmen Aguayo-Torres, J. Tomás Entrambasaguas
Department of Ingeniería de Comunicaciones, University of Málaga
Campus de Teatinos, E-29071, Málaga, Spain.
Email: \{bsoret, aguayo, jtem\}@ic.uma.es

Abstract—Ergodic capacity is not a suitable information-theoretic measure for delay sensitive applications over fading channels. On the other hand, delay-limited capacity becomes zero for Rayleigh channels. In this case, the Capacity with Probabilistic Delay Constraint becomes useful. In [1], the maximum constant source rate that may be supported in a wireless system with a probability $\varepsilon$ of exceeding certain delay bound $D^t, C_{D^t, \varepsilon}$, was evaluated for a Rayleigh channel. This contribution extends that work to an OFDM system over a generic Rayleigh frequency selective channel. An analytical expression of the capacity $C_{D^t, \varepsilon}$ is obtained and figures on how much the maximum allowable rate decreases for higher time or frequency correlation in the channel response are shown. Finally, the expected delay violation is compared to simulations in order to validate our results.

Index Terms—Capacity, OFDM systems, probabilistic delay constraint, effective bandwidth function

I. INTRODUCTION

Broadband networks are expected to support various applications that can generate a mixture of heterogeneous traffic to the network. Within this scenario, the Quality of Service (QoS) requirements are different depending on the traffic class. QoS provisioning of real-time applications such as video or voice transmission involves delay as a critical parameter, thus the understanding of its behavior becomes a main issue.

Ergodic capacity is an average measurement that does not limit delay. On the other hand, delay-limited capacity imposes an strict delay constraint but it results in zero for Rayleigh channel due to channel inversion [2]. In Rayleigh channels, the concept of Capacity with Probabilistic Delay Constraint $C_{D^t, \varepsilon}$ becomes useful. $C_{D^t, \varepsilon}$ is defined as the maximum source rate that may be supported in a wireless system with a probability $\varepsilon$ of exceeding certain delay bound $D^t$. Reducing $\varepsilon$ ensures $D^t$ for a higher percentage of time and the specific case with assurance ($\varepsilon = 0$) represents delay-limited capacity.

Orthogonal Frequency Division Multiplexing (OFDM) is a well-known method to combat frequency selective fading by decomposing the wideband channel into a set of flat fading narrowband orthogonal subchannels that can be independently modulated. Correlation between subcarriers is closely related to the delay spread of the channel. Moreover, OFDM multicarrier nature allows the use of adaptive modulation [3] to enhance its performance significantly. In particular, constellation modulating each subcarrier can be selected suitably to the frequency and time varying conditions of the physical channel.

In [1], $C_{D^t, \varepsilon}$ was evaluated for an uncorrelated flat channel carrying a constant rate flow. This work extends those results to a frequency selective fading channel assuming that OFDM is employed as modulation scheme. The effective bandwidth theory [4] is used as basis for calculating $C_{D^t, \varepsilon}$.

As the first contribution of this paper, we obtain closed-form expressions of the channel effective bandwidth function (EBF) for an OFDM system that employs adaptive modulation over each subcarrier. This result makes it possible to obtain an analytical expression of the capacity with probabilistic delay constraint $C_{D^t, \varepsilon}$ of an OFDM system over a frequency selective channel with any time correlation function and power delay profile.

The remainder of the paper is organized as follows. Section II describes the system model. Section III presents the derivation of the EBF for a frequency selective Rayleigh channel. The Capacity with Probabilistic Delay Constraint is derived and evaluated in Section IV. Simulations of the system are compared to our analytical results in Section V. Finally, concluding remarks are given in Section VI.

II. SYSTEM MODEL

The system is modeled through a queueing system where the source process characterizes the incoming user traffic and the server represents the information transmitted to the wireless channel.

A. Modulation

Over frequency selective channels, OFDM decomposes the total bandwidth into $F$ equally spaced subcarriers, each of them relatively narrowband. An IFFT is used at the transmitter to modulate the comb of subcarriers, and a cyclic prefix is added to eliminate intersymbol interference. At the receiver, the dual process is performed. At each subcarrier, the perceived time variant channel response at instant $n$ is denoted by $H[n, f]$.

Adaptive modulation and OFDM can be easily combined by performing adaptive modulation in each OFDM subcarrier. Subcarrier $f$ can be transmitted at an instantaneous rate $c_f[n]$, which changes according to the time varying channel state. In particular, we consider a transmission adaptation scheme that follows a discrete rate policy with constant transmitted power [3]. At each subcarrier the range of received SNR is divided into $R$ consecutive regions, each of which is associated with a constellation size. Thus, $R$ different constellations are available, selecting a constellation of $R$ symbols within the
fading region \((\Gamma_{r-1}, \Gamma_{r})\), \(r = 0, 1, \ldots, R\) (defining \(\Gamma_{-1} = 0\)). Thus, \(c_f[n]\) can be expressed as a function of the envelope of the response of the channel, \(z_{n,f} = \|H[n, f]\|\):
\[
c_f[n] = f(z_{n,f}) = \log_2(R_r) = c_r,
\]
\[
u^{r-1} \leq z_{n,f} < \nu^r, \ r = 0..R, \ \text{with } \nu^r = \sqrt[2]{\frac{\Gamma_r}{\pi}}
\]

B. Channel model

The frequency selective channel is modeled as a set of \(L\) taps, each of them delayed \(i_f\) from the first one and Rayleigh distributed with a time varying response \(h_t[n]\) [5]:
\[
h[n, i] = \sum_{i=0}^{L-1} h_i[n] \delta[i - i_f]
\]

The channel perceived at each subcarrier \(f\) is obtained by means of the Fourier transform of (2) with respect to the delay parameter \(i_f\):
\[
H[n, f] = \hat{H}\{h[n, i]\}
\]

Correlation among channel response at different frequencies and time instants can be evaluated as the expectation \(E[\|H(n, f)\| \|H(m, g)\|]\). After some manipulations, this expectation can be written as:
\[
E[\|H(n, f)\| \|H(m, g)\|] = E[\|H(n, f)\| \|H(m, g)\|] \cdot \sum_{l=0}^{L-1} p_l e^{-j \frac{2\pi}{\pi} (f - g) \tau_l}
\]

Notice that the autocorrelation in (4) consists of two independent factors, corresponding to time and frequency correlation, respectively. The first term is the time-correlation function. In this work, it is assumed to be the one from Jake’s model [5]:
\[
E[\|H(n, f)\| \|H(m, g)\|] = J_0(2\pi f_D T_S (m - n))
\]

where \(J_0(\cdot)\) is the zeroth order Bessel function of the first kind, \(f_D\) is the maximum Doppler frequency in Hertz, \(T_S\) is the symbol period in seconds. Any other autocorrelation functions can be considered.

The second factor in (4) is the Fourier transform of the Power Delay Profile (PDP), \(p_l\), which describes the number \(L\) and delay \(\tau_l\) of the multipath components, as well as the average power associated with each multipath delay of the considered channel.

The coherence bandwidth \(B_c\) is a parameter commonly used to provide a measure of the range of frequencies over which the channel can be considered “flat”. It is closely related to the Power Delay Profile as it is defined as \(B_c \approx \frac{1}{2\pi f_D}\), where \(D\) is typically taken to be the rms delay spread in seconds.

C. Queuing model

Using a discrete-time fluid model, the incoming user traffic has an instantaneous rate \(\alpha[n]\). On the other hand, subcarrier \(f\) can be transmitted at an instantaneous rate \(c_f[n]\), so that the total instantaneous rate of the wireless channel is \(c[n] = \sum_{f=1}^{F} c_f[n]\), where \(F\) is the number of subcarriers. The processes \(\alpha[n]\) and \(c[n]\) are not necessarily white and represent the amount of bits per symbol generated by the source and transmitted by the server, respectively. In addition, the accumulated source rate \(A[n]\) is the amount of bits generated by the source from 0 to instant \(n-1\), \(A[n] = \sum_{m=0}^{n-1} \alpha[m]\), and likewise, \(C[n] = \sum_{m=0}^{n-1} c[m] = \sum_{m=0}^{n-1} \sum_{f=1}^{F} c_f[m]\) is the accumulated channel rate.

The queue size is assumed to be infinite and \(Q[n]\) denotes the length of the queue at time \(n\). The dynamics of the system is characterized by the discrete-time Lindley equation \(Q[n] = (A[n] - C[n])^+\), with \((x)^+ = \max(0, x)\). The asymptotic log-moment generating function of \(Q[n]\) is defined as [4]:
\[
\Lambda(u) = \lim_{n \to \infty} \frac{1}{n} \log E [e^{uQ[n]}]
\]

Since \(\alpha[n]\) and \(c[n]\) are independent of each other, \(\Lambda(u)\) may be decomposed into two terms, \(\Lambda(u) = \Lambda_A(u) + \Lambda_C(-u)\), where \(\Lambda_A(u)\) and \(\Lambda_C(u)\) are the log-moment generating functions of the accumulated source process \(A[n]\) and the accumulated channel process \(C[n]\), respectively.

If the source and channel processes are stationary and the steady state queue length exists, (i.e. given a threshold \(B\), \(\sup_n Pr\{Q[n] > B\} = Pr\{Q(\infty) > B\}\)), then the workload process \(Q[n]\) satisfies a Large Deviation Principle (LDP) and the following asymptotic behavior for the queue length exceeding \(B\) is satisfied [4]:
\[
Pr\{Q(\infty) > B\} \approx e^{-\theta B} \ \ \ \ B \to \infty
\]

where \(f(x) \bowtie g(x)\) means that \(\lim_{x \to \infty} f(x)/g(x) = 1\) and \(\theta\), known as the QoS exponent, is the solution to:
\[
\Lambda_A(u) + \Lambda_C(-u) \big|_{u=\theta} = 0
\]

Defining \(\alpha(u) = \Lambda(u)/u\) (the so-called effective bandwidth function EBF), the equation to obtain \(\theta\) can also be expressed as:
\[
\alpha(u) = \alpha_A(u) - \alpha_C(-u) \big|_{u=\theta} = 0
\]

where \(\alpha_A(u)\) and \(\alpha_C(u)\) are the EBF’s of the source process and the channel process, respectively. In related literature [6], \(\alpha_C(-u)\) is referred to as the effective capacity function but, since it is the effective bandwidth function of the channel process, we prefer to keep the term effective bandwidth function for all the processes involved in the analysis.

A more accurate approximation for small values of \(B\) includes the probability that the queue is not empty, \(\eta = Pr\{Q[n] > 0\}\), and, therefore, the following less conservative approximation for the tail probability of the queue is satisfied:
\[
Pr\{Q(\infty) > B\} \approx \eta \cdot e^{-\theta B}
\]

We refer to the delay of the bits leaving the queue system at symbol \(n\) as \(D[n]\). As in the queue length process, the steady state solution for the delay process exists. Moreover, the probability of exceeding \(D^*\), denoted throughout this paper target delay, can be written as follows:
\[
\varepsilon = Pr\{D(\infty) > D^*\} \approx \eta \cdot e^{-\varepsilon \cdot \alpha_A(\theta) D^*} = \eta \cdot e^{-\varepsilon \cdot \alpha_C(-\theta) D^*}
\]
where $\varepsilon$ is the probability of exceeding $D^i$ and it provides a measure of the percentile of the delay directly as $1 - \varepsilon$.

The expression in (10) constitutes the basis for the calculation of the Capacity with Probabilistic Delay Constraint and it is linked to the EBF of the arrival and the channel processes through the QoS exponent $\theta$.

III. EFFECTIVE BANDWIDTH FUNCTION OF OFDM SYSTEMS

To address the problem of calculating the EBF of the channel $\alpha_C(u)$, we divide the cumulative channel rate $C[n]$ in the interval $[0..n]$ into $[n/k]$ blocks of size $k$:

$$C[n] = \sum_{b=0}^{[n/k]-1} C_b[k]$$  (11)

where $C_b[k]$ is the accumulated channel rate in block $b$:

$$C_b[k] = \sum_{p=0}^{k-1} c[k \cdot b + p] = \sum_{p=0}^{k-1} \sum_{f=1}^{F} c_f[k \cdot b + p]$$  (12)

The time and frequency correlation between the elements in the block is considered but, with the proper selection of the block’s length, $k$, independence between elements of different blocks may be assumed. The choice of $k$ is closely related to the correlation of the channel. If the channel is strongly correlated, longer blocks have to be defined in order to assume independence. Whatever the value of $k$ is, there is a residual value of correlation between the last elements of one block and the first elements of next one. Nevertheless, this “border” correlation is negligible when the value of $k$ is large enough.

With a proper value of $k$ (large enough), $C[n]$ is the sum of independent random variables, the blocks $C_b[k]$, and the Central Limit Theorem can be applied. Furthermore, the EBF is defined as the limit when $n$ tends to infinite, so that $n$ and $k$ can be selected indeed as large as wanted. If $n$ is long enough, $C[n]$ is the sum of a sufficiently large number of independent random variables and, as stated by the Central Limit Theorem, it will be approximately normally distributed.

In the numerical evaluations and simulations conducted throughout this paper, the validity of this Gaussian approximation for $C[n]$ has been validated by testing for normality (Lilliefors test) on the result for the selected values of $k$.

Under these conditions (sufficiently long $k$ and $n$), the EBF of the resulting Gaussian distribution of $C[n]$ is computed as:

$$\alpha_C(u) = \lim_{n \to \infty} \frac{1}{n \cdot u} \log \mathbb{E} \left[ e^{u C[n]} \right] = m_k + \frac{u}{2} \sigma_k^2$$  (13)

where $m_k$ and $\sigma_k^2$ are the mean and the variance of a block $C_b[k]$ normalized with the block length.

The mean $m_k$ easily computed from the mean $m_c$ of the discrete random variable $c_f[n]$:

$$m_k = F \cdot m_c = F \cdot \sum_{r=0}^{R} p_r \cdot c_r$$  (14)

where the probability of using the $r$th constellation is given by:

$$p_r = \exp(-\frac{\Gamma_{r-1}}{\gamma}) - \exp(-\frac{\Gamma_r}{\gamma})$$  (15)

However, the evaluation of the variance of the blocks $\sigma_k^2$ is not straightforward. These two properties will be used ([7]):

$$\var\left( \sum_{p=0}^{k-1} X_p \right) = \sum_{p=0}^{k-1} \sum_{q=0}^{k-1} \text{cov}(X_p, X_q)$$

$$\text{cov}\left( \sum_{i=0}^{k-1} X_i, \sum_{j=0}^{k-1} Y_j \right) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \text{cov}(X_i, Y_j)$$  (16)

with autocovariance function

$$\text{cov}(X_n, X_{n+m}) = E[X_n X_{n+m}] - E[X_n]E[X_{n+m}]$$  (17)

Applying the expression above to obtain $\sigma_k^2$:

$$\sigma_k^2 = \var\left( \sum_{p=0}^{k-1} \sum_{f=1}^{F} c_f[p] \right)$$

$$= \sum_{p=0}^{k-1} \sum_{q=1}^{k-1} \sum_{f=1}^{F} \sum_{g=1}^{F} \left( E[c_f[p]c_g[q]] - m_c^2 \right)$$  (18)

The values of $m_c$ for the instantaneous service rate at subcarrier $f$ have just been described. Thus, the evaluation of the variance of the blocks comes down to evaluating $E[c_f[n]c_g[m]]$. To do it the instantaneous service rate $c_f[n]$ is expressed as a function of the envelope of the response of the channel $z_{n,f} = |H[n,f]|$, as in (1). The expectation is solved by splitting the integral in $z_{n,f}$ in several addends and the same is done for the integral in $z_{m,g}$. The expectation can then be written as:

$$E[c_f[n]c_g[m]] = \sum_{r=1}^{R} c_r \sum_{s=1}^{R} c_s \mathcal{F}_{m-n-g-f}(v^r, v^s)$$  (19)

with:

$$\mathcal{F}_{Dn}\Delta f(v^1, v^j) = F_{x}(v^i, v^j, \Delta n, \Delta f) + F_{x}(v^{i-1}, v^j, \Delta n, \Delta f) - F_{x}(v^{i-1}, v^j, \Delta n, \Delta f)$$

$$F_{x}(v^i, v^j, \Delta n, \Delta f)$$ is the bivariate cumulative distribution function (CDF) for Rayleigh distributed variables and can be found in [8] (pp. 172, eq. 6.7). The correlation parameter $\rho$ that appears in the bivariate CDF depends on the expectation $E[|H(n,f)||H(m,g)|]$ described in (4).

This section about the effective bandwidth function of a frequency selective channel concludes by presenting the numerical evaluation of the expression obtained for $\alpha_C(u)$. The result is based on the Central Limit Theorem. Therefore it is simply expressed in terms of the mean and variance of the process. To check the validity of the outcomes, the mean and the variance have been measured over a long realization of the instantaneous transmission rate process and compared to the analytical results, following the same method as described in [1].

Figure 1 shows the result of computing (13) with an average SNR of 10dB and 14dB and 16 subcarriers. The PDP
corresponds to a vehicular channel A as defined in [9], which leads to a coherence bandwidth of $B_c \approx 430$KHz. The length of the blocks $k$ is set to 10000. Notice that with a proper selection of $k$ (high enough), the Central Limit Theorem can be applied, there is no dependence of the value of $\alpha_C(u)$ with the block length $k$. At each subcarrier, four different uncoded constellations and no-transmission are allowed (R=5): NTX, BPSK, QPSK, 16QAM and 64QAM; the fading regions are selected such that the instantaneous Bit Error Rate is kept under $BER_T = 10^{-2}$ [3]. In this paper, these are the specified parameters for the constellations in all subsequent simulations. As expected, $\alpha_C(u)$ decreases with $u$, as a stringent QoS constraint implies a high value of the parameter $u$.

IV. CAPACITY WITH PROBABILISTIC DELAY CONSTRAINT

In this section, we derive the maximum allowable source rate that fulfills certain delay QoS requirements for a constant rate source in an OFDM system. We call this rate the capacity with probabilistic delay constraint, $C_{D^t,\varepsilon}$.

The source traffic arrives to the buffer at a constant rate, i.e. $a[n] = \lambda$. In a high load scenario, the probability that the buffer is not empty approaches one, i.e. $\eta \to 1$. Under this assumption, the delay constraint is worked out from (10):

$$-\frac{\log(\varepsilon)}{D^t} = \theta \cdot \alpha_A(\theta) \tag{21}$$

The EBF of the source is constant: $\alpha_A(u) = \lambda$. The QoS exponent is obtained by solving (8) for a constant source and the effective bandwidth of the channel in (13):

$$\lambda - \alpha_C(-\theta) = 0 \Rightarrow \theta(\lambda) \triangleq \theta(m_k, \sigma_k^2, \lambda) = \frac{2(m_k - \lambda)}{\sigma_k^2} \tag{22}$$

With (22) substituted into (21), the value of $\lambda$ that matches the equation will represent the maximum source rate that the OFDM system may support with a probability $\varepsilon$ of exceeding a delay bound $D^t$. That’s the capacity with probabilistic delay constraints $C_{D^t,\varepsilon}$ and is written in very simple terms as:

$$C_{D^t,\varepsilon} = \frac{m_k}{2} + \frac{1}{2} \sqrt{m_k^2 - 2\sigma_k^2 \left( -\frac{\log(\varepsilon)}{D^t} \right)} \tag{23}$$

Notice that we have obtained a result analogous to the one obtained in [1] with differences in the involved parameters: $m_k = F m_c$ is the average of the instantaneous OFDM rate and the variance $\sigma_k^2$ includes now not only the time-frequency but also the frequency-correlation of the channel via the expression in (18). Two foreseen limits can be checked in the capacity $C_{D^t,\varepsilon}$. For high $D^t$ values or $\varepsilon \to 1$, the QoS requirement relaxes and $C_{D^t,\varepsilon}$ approaches $m_k = F m_c$, which is the ergodic capacity of the channel. On the other hand, as the target delay $D^t$ or $\varepsilon$ become lower, the wireless channel tolerates lower traffic arrival rates in order to guarantee the delay constraints. Moreover, the dependence on the variance captures the influence of both time and frequency correlation of the channel. Thus, as correlation increases, the second addend in the square root increases in absolute value and $C_{D^t,\varepsilon}$ diminishes. Finally, it is worth pointing out that the capacity has been obtained for a constant rate source. Nevertheless, the procedure can be extended to any other source process.

Figure 2 (a) shows the evaluation of $C_{D^t,\varepsilon}$ for different values of the time correlation of the channel. The channel is again the vehicular channel A defined in [9], with a coherence bandwidth of approximately 430KHz. The autocorrelation function of the different echoes is a first-kind Bessel function, with values of the product $f_D \cdot T_S$ of 0.05 and 0.005. The average SNR is set to 10dB and 15dB and $F$ is set to 4. One can see that a higher value of time-correlation (smaller value of the product $f_D \cdot T_S$) implies a lower value of the capacity, as expected. Likewise, Figure 2 (b) presents the influence of the frequency correlation, by means of different values of the coherence bandwidth. In this case two different PDP are represented: the vehicular channel of the previous figure and a pedestrian channel A as defined in [9], which leads to a coherence bandwidth of $B_c \approx 430$KHz and $B_c \approx 3.4$MHz, respectively. There are 4 subcarriers and the average SNR is again set to 10dB and 15dB. The time-correlation is fixed with a product $f_D \cdot T_S$ of 0.05 in both cases. As expected, a smaller value of the coherence bandwidth (vehicular channel) corresponds to a higher value of the capacity, and the reduction is more noticeable for smaller values of the average SNR.

V. SIMULATION COMPARISON

A queueing system has been simulated to validate the results presented in this paper. Bits are sent to the buffer of queue length $Q[n]$ in the $n$th symbol. The bits in the queue are transmitted over the fading channel on a First In First Out basis and at a service rate $c[n]$. Certain values of $D^t$ and $\varepsilon$ are fixed and the Capacity with Probabilistic Delay Constraint is evaluated with (23). Then, the result is given in the simulator to the arrival process to generate source data at the (constant) rate $C_{D^t,\varepsilon}$. The simulation is run and the tail probability of exceeding the target delay is measured based on the measurements of the delay suffered by bits leaving the queue. Notice that the expected value of this tail probability, $P_T\{D(\infty) > D^t\}$, is $\varepsilon$.

Figure 3 shows the results of the simulations. $P_T\{D(\infty) > D^t\}$ is represented for a target delay of 60 symbols. Two different simulations are presented, with
VI. CONCLUSIONS

In this work, we analyze the Capacity with Probabilistic Delay Constraint $C_{D^t,\varepsilon}$ in an OFDM system, which expresses the maximum constant source rate that can be supported by a channel for a particular delay constraint. To derive $C_{D^t,\varepsilon}$, the analytical evaluation of the effective bandwidth function of a frequency selective Rayleigh channel is first addressed. We then derive a closed-form of the capacity for a constant rate source, and the numerical results confirm the expected qualitative behaviour: $C_{D^t,\varepsilon}$ decreases for higher time- or frequency-correlations of the channel response and for shorter allowed delays. Finally, simulations addressing different scenarios are conducted to validate our outcomes.

VII. ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Government and the European Union (project TEC2007-67289) and the Andalusian Government (project TIC-03226).

REFERENCES