Integrated voice, and data over OFDMA-based cellular networks

G.A.F.M. Khalaf
Electronics, Communications and Computer Dept. Faculty of Eng., Helwan, Cairo, Egypt

We consider the problem of integrating real, and non real time applications over an OFDMA-based cellular network under Rayleigh fading channels. A differentiated power allocation system is presented in conjunction with a dynamic bit assignment scheme. The bit assignment is based on a number of SNR-switching threshold levels that are predetermined by quantizing the overall spectrum of received SNR such that a predefined target BER is achieved. A queueing model that takes into account the effects of the sliding window flow control as well as wireless channel outages is developed, and validated. Simulation results have demonstrated that our system is able to provide an acceptable QoS levels to the real, and non real time individual applications.

يقدم هذا البحث نموذج لإرسال الإشارات الصوتية و المعلوماتية بإستخدام تكنولوجيا الموجات الحاملة المتعامدة في شبكات الاتصالات الخلوية. و قد تبين إمكانية تنفيذ هذا ألنظام عن طريق الضبط المستمر لقدرة الإرسال بما يحقق درجة الإستقبال الجيد للإشارات المرسلة. أظهرت النتائج قدرة النظام المقترح على الحفاظ على جودة الأداء رغم التغيرات ألتى تطرء على وسط ألتراسل.

Keywords: OFDM cellular networks, Differentiated services over OFDM, Adaptive bit, Power allocation

1. Introduction

Basically, Orthogonal Frequency Division Multiple Access, (OFDMA) operates by transmitting data through a set of orthogonal, parallel subcarriers at a lower data rate, effectively transforming a frequency selective fading channel into a collection of flat fading subcarriers. To improve both performance, and rate of communication links, transmitter design should matches the intended propagation channel. The resulting channel-adaptive transmission adjusts parameters such as power levels, constellation sizes, coding schemes, and modulation types. In this process, the decision for switching to an alternative (power/modulation) state is based upon the channel state information that is assumed to be in disposal of the transmitter [1, 2].

In this paper, we consider an OFDMA-based cellular network serving a real, and non real time applications. The real time application, e.g., Internet Protocol (IP) telephony, and video conferencing, use the User Datagram Protocol (UDP) protocol while the non real time data applications use Transmission Control Protocol (TCP). Being a window flow controlled protocol behaves very differently from UDP. Therefore, the traffic generated by UDP and TCP connections behave differently, and have different Quality-of-Service (QoS) requirements. In order to carry different applications, which demand different ranges of QoS requirements, the embedment of QoS provisioning in the Medium Access Control (MAC) is needed. For example, a QoS framework in the MAC layer has been integrated with the transmission system in the IEEE 802.16 standard [3]. The analysis of the queueing delay performance for the IEEE 802.16 was conducted in [4, 5]. A vacation queueing model was adopted in [6], to analyze the queueing performance of OFDMA system. As for the delay-sensitive, bandwidth-intensive, and loss-tolerant multimedia applications, the two classes are differentiated primarily by the queuenig delay in [7]. Concerning the transmission requirements, throughput is considered as a critical performance metric to the capability of a network in support of real time traffic. From a certain classic point of view, maximum capacity can be achieved by maximum-rate subcarrier allocation [8]. However, it is not fair. The Rhoo-Cieffi subcarrier allocation algorithm [9] achieves proportional fairness amongst the users. However, the overall capacity achieved is much lower. In general, joint subcarrier, and power allocations in OFDMA is a complex problem [10]. Usually, the problem is simplified by separating subcarrier allocation and power allocation [9].

Our main focus in this research is on power allocation, and modulation adaptation. This seems to be a reasonable approach regarding the multiclass service communications. The system we are proposing is intended to maximize the overall throughput for the real time applications without profound consequences on the non real time applications.

This paper is organized as follows, Section 2 provides details of the Rayleigh fading channel, and also describes basic TCP, and UDP traffic characteristics. The approach we have proposed for integrating the real, and non real time traffic is detailed in section 3. RF channel outages is modeled in section 4. In section 5, we develop a queueing model, which takes into account both the characteristics of the RF channel as well as the impact of the flow control dynamics. The findings in this paper are validated using simulations in section 6, and the implications of the results obtained are summarized in section 7.

2. Modeling approach, and preliminaries

2.1. Channel model

The model of this paper considers the downlink of an OFDMA-based cellular network serving real time UDP, and non real time TCP users. A wireless multipath fading channel is assumed constant over every OFDMA symbol transmission time, and noise is assumed to be independent, identically distributed (i.i.d), and AWGN with variance σ^2 for all carriers and all users. The normalized channel gain seen by k^{th} carrier of user n in a symbol period is,

$$H_{n,k} = \frac{|\, a_{n,k}\,|^2}{\sigma^2} \ , \tag{1}$$

We assume static subcarrier assignment where in subcarriers are divided evenly among users, and a fixed set of subcarriers is allocated to each user. For a Rayleigh fading channel, the fading amplitude α is distributed according to,

$$p(\alpha) = \frac{2\alpha}{\sigma_\alpha^2}\exp(\frac{-\alpha^2}{\sigma_\alpha^2}), \quad \alpha \geq 0\ , \tag{2}$$

where σ_α^2 is the mean square value of α. In this case, the instantaneous Signal-to-Noise Ratio (SNR) per symbol, $\gamma_{n,k} = E_{n,k}\ H_{n,k}$ is distributed according to an exponential distribution [5],

$$p(\gamma,\bar{\gamma}) = \frac{1}{\bar{\gamma}}\exp(\frac{-\gamma}{\bar{\gamma}}), \tag{3}$$

where $E_{n,k}$ is the power per symbol, and $\bar{\gamma}$ is the mean of α.

Our performance evaluation of the digital communication over a fading channel will generally be a function of the average SNR per symbol, γ. For instance, the bit-error rate (*BER*) of an M-QAM, (M-symbols= $2^{b_{n,k}}$, $b_{n,k}$ bits per symbol for the n^{th} user's; k^{th} subcarrier), modulation can be well approximated by [11],

$$BER(M,\gamma) \leq 0.1 \quad \exp\left(\frac{-3\gamma}{2(M^2-1)}\right), \tag{4}$$

and for a given *BER* value, rearranging eq. (4) gives the maximum number of bits per symbols, $b_{n,k}$ to be transmitted for the n^{th} user's; k^{th} subcarrier,

$$b_{n,k} = \log_2 (1 + \frac{\gamma_{n,k}}{\Gamma})\ , \tag{5}$$

where $\Gamma = -\ln(5BER)/1.5$, and finally, the maximum achievable rate for user n is given by,

$$\mu(t) = \sum_{k\,\in\,\Omega_n} b_{n,k}\ , \tag{6}$$

where Ω_n denotes the set of subcarriers assigned to user n in a symbol period.

2.2. The non-real time TCP traffic

A sender-based one way traffic scenario is considered where the mobile users act as TCP sinks. A bulk TCP connections are assumed (i.e., senders always have data to transmit, and can transmit as many packets as their transmission windows allow). The TCP sinks can accept packets out of sequence but deliver them only in sequence to the users and, they generate immediate acknowledgements (ACKs). Similar to that in all currently available TCP implementations, the basic window adaptation procedure consists of the slow-start, and the congestion avoidance phases where the evolution of the sender's congestion window W(t), and the slow-start threshold $W_{th}(t)$ at time t are triggered by TCP ACKs, and time outs (Tos) [12]. In our modeling, fast retransmission, and fast recovery procedures are implemented subsequent to a packet loss as in TCP Reno [13], which is the de facto standard for TCP implementation in today's internet. The model of this paper assumes a network link with maximum speed C packets/sec., and data buffer space of L_d packets. TCP flows arrive at the link according to a Poisson process with rate λ_d, whereas the packet lengths are i.i.d from a common distribution with mean $\mu_d^{-1} < \infty$. To insure stability, the load of the system must satisfy $\rho_d = \dfrac{\lambda_d}{\mu_d} < C$, with ρ_d is the traffic intensity. We also assume that flows have the same round trip time, RTT, and a maximum window size of $W_{\max.}$ packets, and RTT puts a limit on the user's transmission rate, $r \le \dfrac{W_{\max.}}{RTT}$.

2.3. The real time UDP traffic

In this paper, we design a system to guarantee QoS to the UDP telephony, and video conferencing. For these applications, the end-to-end delay for packets should not exceed (say) 150 m sec. If a packet exceeds this limit, it will be dropped. For satisfactory performance, the fraction of dropped packets for an application should be less than (say) 0.2 % (based on subjective evaluation/acceptance for such an application). For the non-window controlled UDP traffic, we assume a fluid flow with mean rate λ_v, and buffer size of L_v packets. The state of the UDP traffic is assumed to be independent of the behavior of the TCP connections. However, the shared bottleneck will be the total transmission power budget constraint as will be discussed later.

3. Differentiated bit, and power allocation

Assuming that the transmitter has knowledge of the normalized channel gain (eq.. (1)) seen by subcarriers of all users. This information is assumed to be updated periodically with the help of a feedback channel. In the following, we describe details of our proposed bit, and power allocation systems.

Consider the OFDM system over the fading channel described in section 2.1. A target BER is set for the UDP-voice traffic, BER_v. Assume a fixed M-QAM modulation is being employed throughout the subcarriers allocated to the UDP traffic. A power level $P_v(\gamma)P$ is allocated to voice subcarriers just to meet the BER_v requirement, where P is the peak power constraint. More specifically, we use the water-filling scheme [14],

$$P_{n,k}^{v}(\gamma) = P - \frac{1}{\gamma_{n,k}}. \tag{7}$$

The remaining available power is dynamically assigned to the non real time TCP traffic. Now, in contrast with the fixed bits (i.e., modulation) to be allocated to voice.

We start the bit allocation for data by first dividing the whole spectrum of received SNR values into several disjoint regions. The region boundaries, γ_i, i= 1, 2,... can be determined by inverting eq. (4) to obtain,

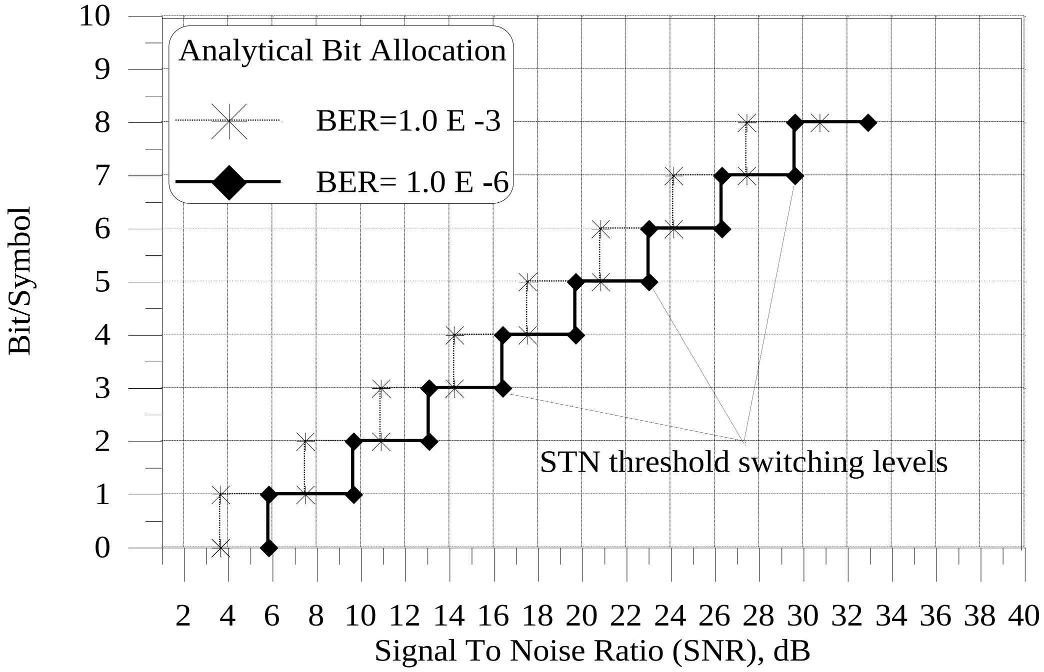

Fig. 1. Analytical bit allocation using quantized SNR.

$$\gamma_i = \frac{2}{3}(1-2^{2n})\ln(10BER_d), \tag{8}$$

then substituting eqs. (8 into 5), gives the number of bits to be allocated to the data at a given channel state. As can be seen fig. 1, the bit allocation process adapts to the channel conditions. That is, if the actual subcarrier's SNR falls in a particular region, a corresponding bits/symbol is chosen.

At this point, we can determine the expected number of bits loaded on to a given subcarrier. Based on the exponential distribution given by eq. (3). The expected number of bits, $\overline{b}_{n,k}$, loaded on to a given subcarrier, k is,

$$\overline{b}_{n,k} = \sum_i b_{n,k} \int_{\gamma_i}^{\gamma_{i+1}} p(\gamma)d\gamma . \tag{9}$$

hence, the total OFDMA data throughput is determined using eq. (6).

4. Outage probability

So fare, system behavior is analyzed by focusing on SNR. Differentiation amongst the two-class users is achieved in terms of the (constraint) available power. In order to grant distinct transmission qualities for users, emphasis should be placed on the transmission conditions. For instance, no voice is sent when the required power violates the constraint, $P_v(\gamma) < P$. In this case, voice transmission suffers an outage. The outage probability, $p_v^{out}(\gamma)$, is equivalent to the following definition,

$$p_v^{out}(\gamma) = p(\gamma_v \le \gamma_{thr}) \qquad = \int_0^{\gamma_{thr}} p(\gamma)\,d\gamma , \tag{10}$$

where γ_{thr} is the minimum SNR required to sustain the (fixed) bit allocation needed for voice. On the other hand, since voice traffic has to maintain a target BER_v, then γ_{thr} can easily be obtained from eq. (3),

$$p_v^{out}(\gamma) = 1 - \exp(\frac{-Y_{thr}}{\bar{Y}}). \quad (11)$$

5. The delay performance

In this section, we propose a model for the steady state delay performance of voice, and data traffics. As described in section 2, voice packets arrive at a FIFO queue having length L_v packets, and mean rate λ_v packets/sec. The service time for packets is approximated as being exponentially distributed with mean μ_v^{-1}. Therefore, the evolution of the packet queue can be modeled as an M/M/1/ L_v finite queueing model. The performance of this model can be obtained by classical techniques [15]. The steady state distribution $\{p_i\}$, i=0, 1,..., L_v of the number of packets in the buffer is given by,

$$p_i = \frac{1-\rho}{1-\rho^{L_v+1}}\rho^i, \quad (12)$$

where $\rho = \dfrac{\lambda}{\mu}$.

Then, the expected value of the queue length can be obtained by,

$$N_q = \sum_{i=1}^{L_v}(i-1)p_i$$

$$= \frac{\rho}{1-\rho} - \frac{\rho(1+L_v\rho^{L_v})}{(1-\rho^{L_v+1})}. \quad (13)$$

Using Little's formula [16], the mean delay time, T_v, is given by,

$$T_v = \frac{N_q}{\lambda_v}. \quad (14)$$

However, in a wireless fading channel, transmission outages often occurs. To count for this difficulty, we define a virtual transmission rate $\bar{\lambda}_v$,

$$\bar{\lambda}_v = \lambda_v(1-p_v^{out}), \quad (15)$$

where p_v^{out} is given by eq. (11). Substituting eqs. (13, 15 into 14) gives,

$$T_v = \frac{1}{\lambda_v(1-p_v^{out})} \times$$

$$(\frac{\rho}{1-\rho} - \frac{\rho(1+L_v\rho^{L_v})}{(1-\rho^{L_v+1})}). \quad (16)$$

Now, concerning the data traffic, the virtual TCP packet arrival rate can be approximated by,

$$\bar{\lambda}_d \cong \frac{\lambda_d}{(1-p_v^{out}-p_v^W)}. \quad (17)$$

where p_v^{out} counts for the extra (retransmission) traffic, and p_v^W counts for the limitation imposed on TCP user's transmission rate by the oscillatory dynamics of the sliding window. For a relatively small p_v^W, the relationship between the window, W, and p_v^W simplifies to [17],

$$W \cong \sqrt{\frac{K}{p_v^W}}. \quad (18)$$

where K is a constant ≈ 1.

Having obtained $\bar{\lambda}_d$, assuming Poissonian extra traffic components due to p_v^{out}, and p_v^W. Then the TCP packet delay, T_v, can, similarly, be obtained using eq. (14).

6. Verification by simulation

Computer simulation is performed in this section to verify the analysis presented in this paper. A FORTRAN program is created, and developed for this purpose. A Rayleigh fading channel model is adopted in the simulation. The square of the envelop of the propagation channel from the transmitter to the receiver- and vice-versa is given by,

$$\alpha^2 = K\,d^{-4}.\,X(\sigma^2)$$

where d being the transmitter to receiver distance, X is a lognormal random variable whose associated zero-mean Gaussian random variable has a variance equal to σ^2. In the OFDMA-wireless system, a TCP Reno is modeled following the approach detailed in [18]. Although OFDMA systems typically have a relatively larger number of subcarriers, we consider an identical 144- subcarrier setting for each of the UDP, TCP traffics for comparison purpose. The parameters chosen for simulation model are summarized in table 1.

Fig. 2 shows a typical sample path of the TCP window evolution.

Table 1
Simulation system parameters

Spread of lognormal (σ),dB	6
Normalized power budget	1
Path loss exponent	4
TCP segment size (byte)	500
TCP ACK packet size (byte)	52
Maximum TCP window	8, 16
RTT (m sec.)	10
Capacity (packet/sec.)	4000
BER_v	1E-3
BER_d	1E-6

Due to TCP's inherent loss recovery mechanism, its throughput may drop significantly even at high SNR values. That is, at higher SNR, the impact of reduced radio frame error/loss is masked by the smaller number of successfully transmitted segments. The impact of TCP conservativeness can be seen through the fast variations in the TCP's transmission queue shown in fig. 3. In our simulation, unlike the standard where the same modulation scheme is employed across all subcarriers, our allocation scheme uses same modulation type within a predetermined range of SNR. The region boundaries are pre-calculated using eq. 8, and the bit allocations during simulation runs are shown in fig. 4. Fig. 5 shows the corresponding distribution for each individual allocation state. This way, the expected throughput of the system (during simulation) can be obtained using eqs. (6, 9).

To insure a fixed, continuous throughput for the UDP voice traffic, we have differentiated amongst the two application types in terms of the power allocations. In the simulation, we have assumed a target BER for voice = 1E-3, and for data 1E-5.

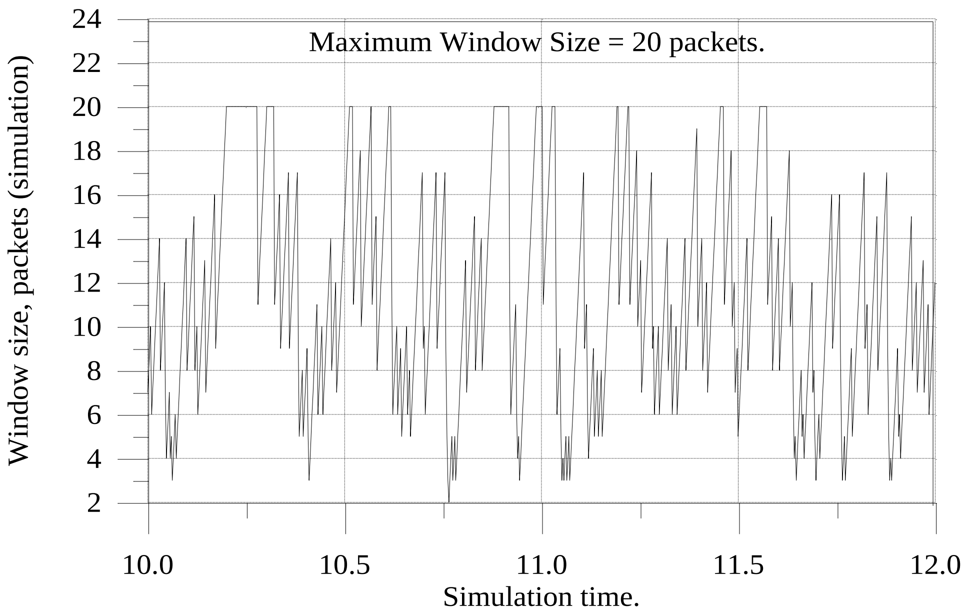

Fig. 2. Analytical bit allocation using quantized SNR.

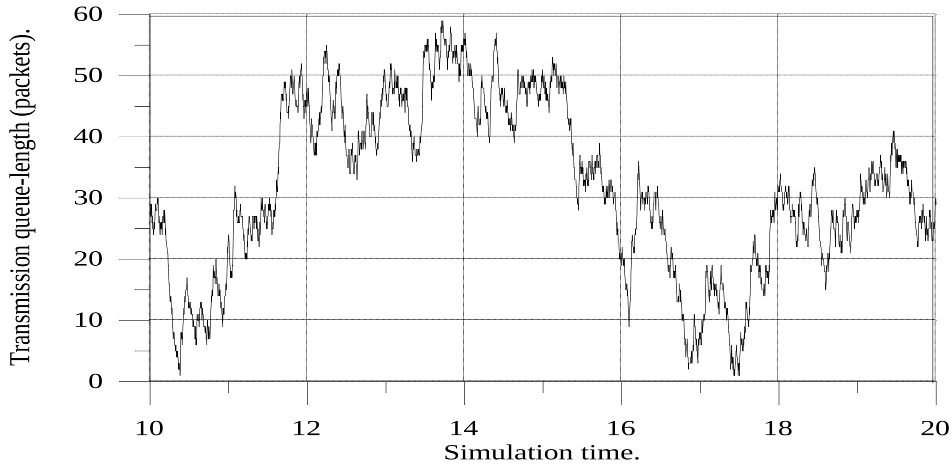

Fig. 3. Number of packets in TCP's transmission queue.

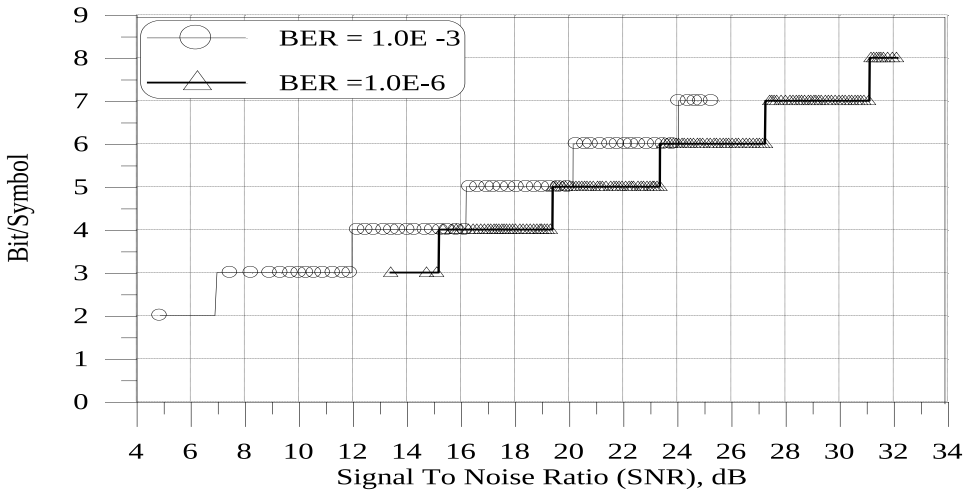

Fig. 4. Bit allocation via simulation.

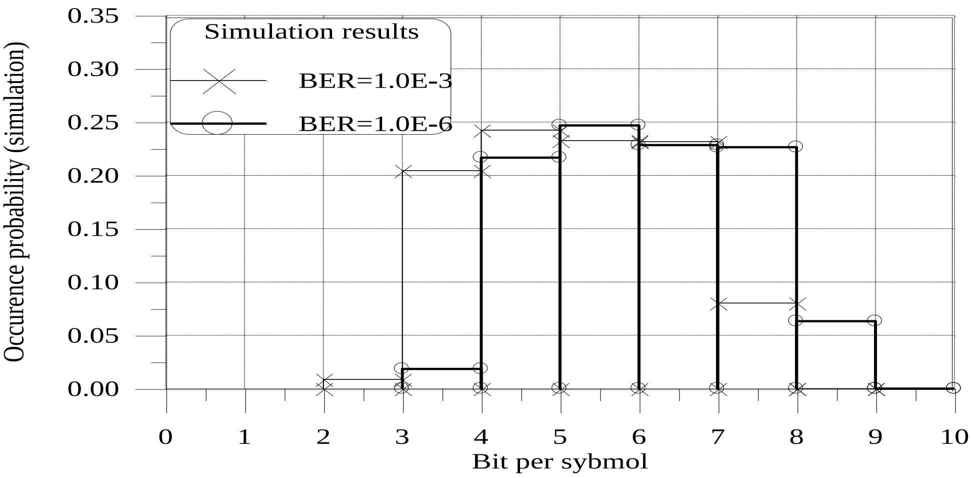

Fig. 5. Distribution of the bit allocation states.

Fig. 6, shows the expected total throughput for voice traffic at different average SNR levels. In simulation, we have changed the overall average SNR by changing the total power budget, P in eq. (7). The resulting TCP throughput with, and without power differentiation is also shown for comparison purpose. As can be seen, in all cases, voice traffic is guaranteed a fixed throughput. This has actually been achieved by compensating the variations in the channel fading envelope, α, in terms of the power allocation process. However, as expected, this has led to a corresponding reduction in the throughput of the TCP traffic. As can be seen, one way of squeezing the reduction in the TCP throughput is to use a reasonably high SNR power budget. Assuming an OFDMA system whose transmission rate is (say) 64 symbol/sec. Therefore, the actual transmission rate for voice is about 64 K b/sec. As mentioned earlier in this paper, concerning the varying nature of the transmission channel conditions. Voice transmission is not attempted when the power constraint is violated. Fig. 7, shows the outage probability at different average SNR values obtained analytically, as well as by simulation. As can be seen, the higher the average SNR, the lower the outage probability. This in itself provides us with an additional mean for service differentiation. That is, a service differentiation by setting different bounds on the outages of different application classes, i.e., $\gamma_{\rm thr}^{\rm v} \neq \gamma_{\rm thr}^{\rm d}$.

Finally, we are in position to investigate the validity of the assumptions we have made while modeling the delay performance in section 5. Fig. 8, presents the mean delay time of voice traffic as function of the traffic intensity. This time delay includes the waiting, and propagation time delays. As can be seen, the analytical delay eq. (16) agrees well with the actual (simulation) results as shown in fig. 8. However, at higher traffic intensity values the discrepancies are significant. This is attributed to the effects of channel outages, which often limits the voice transmission attempts. Along the same line, fig. 9 shows the delay performance of the TCP data traffic. In this case, the delay includes the transmission, and retransmission queueing delay in addition to the delay imposed by the transmission rate limitations due to the TCP's sliding window. The analytical delay shown is obtained by substituting eqs. (18, and 17 into 14) with window size of 20 packets, p_d^{out} =0.1. As can be seen, the simulation results are reasonably close to the analytical one.

6. Summary, and conclusions

In this paper, we have proposed an efficient, and practically implementable system for integrated voice, and data transmission over a common OFDMA technology. We have considered providing differentiated end-to-end QoS requirements This we have achieved through a dynamic power allocation scheme over a static subcarrier assignment in conjunction with a dynamic bit/symbol allocation scheme. The bit allocation scheme chooses the modulation type according to a pre-quantized SNR levels. It has been demonstrated that in order to exploit the OFDMA system capacity at its best, the wireless system designer has to make proper choice of the total OFDMA power budget in order to optimize radio resource assignment. An analytical model is presented to predict the expected time delay for the two service classes. Simulation results have indicated that our system is able to provide an acceptable (throughput, and delay) QoS framework for real, and non real time individual applications.

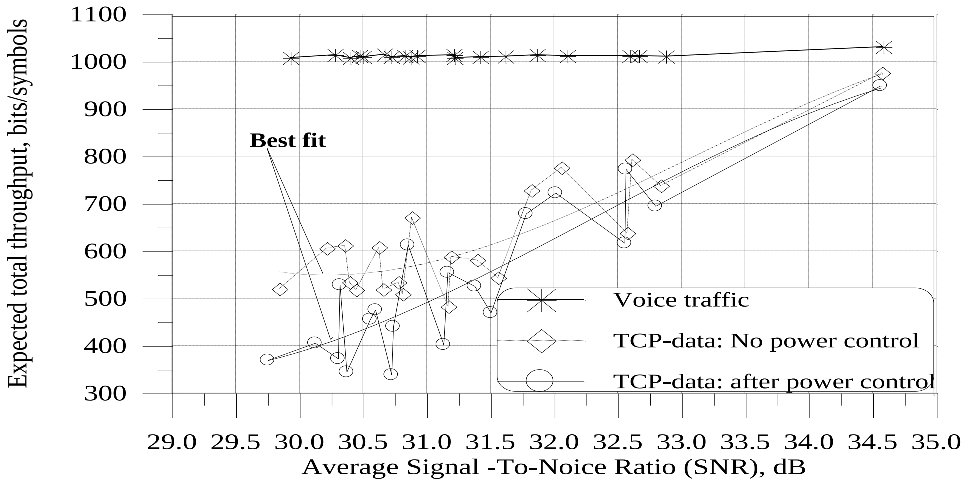

Fig. 6. Performance of (voice/data) power differentiation system.

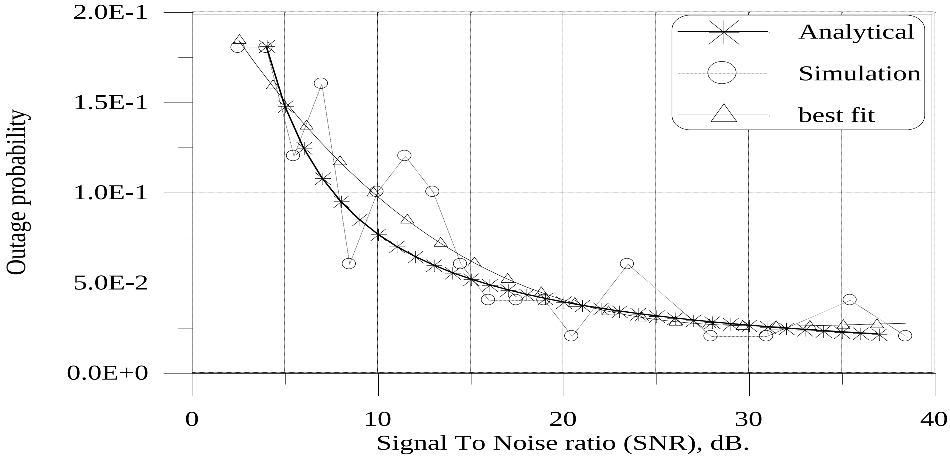

Fig. 7. Outage probability: analytic, and simulation.

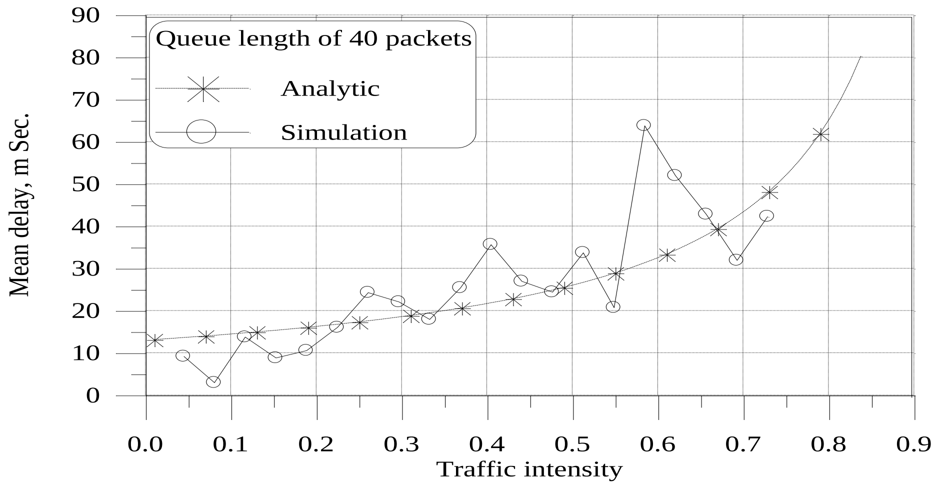

Fig. 8. Mean delay performance of voice traffic.

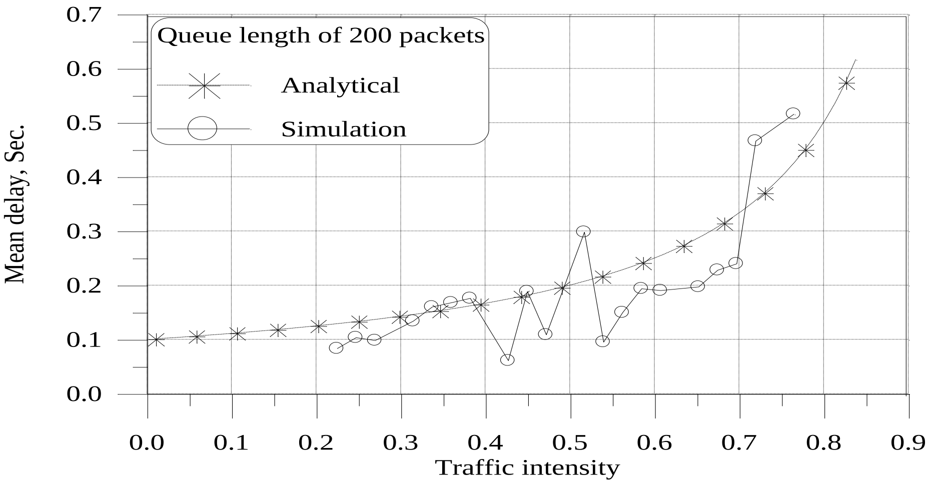

Fig. 9. Mean delay performance of TCP-data traffic.

References

[1] G.E. Oien et al., "Impact of Channel Prediction on Adaptive Coded Modulation Performance in Rayleigh Fading", IEEE Trans. on Veh. Technol., Vol. 53 (3), May pp. 758-769 (2004).

[2] Goldsmith and S.G. Chua, "Variable-Rate Variable Power for M-QAM for Fading Channel", IEEE Trans Communi., Vol. 45, pp.1218-1230 (1997).

[3] K. Wongthavarawat and A. Ganz, "IEEE 802. 16 Based Last Mile Broadband Wireless Military Networks with Quality of Service Support", IEEE MILCOM 03, pp. 779-784 (2003).

[4] R. Iyengar, P. Iyer and B. Sikdar, "Delay Analysis of 802.16 Based Last Mile Wireless Networks", in Proc. IEEE GLOBECOM’05, Nov. pp. 3123-3127 (2005).

[5] D. Niyato and E. Hossain, "Queue-Aware Uplink Bandwidth Allocation for Polling Services in 802.16 Broadband Wireless Networks", Proc. IEEE GLOBECOM’05, Nov. pp. 3702-3706 (2005).

[6] D. Niyato and E. Hossain, "Queueing Analysis of OFDM/TDMA Systems", in Proc. IEEE GLOBECOM’05, Nov. pp. 3712-3716 (2005).

[7] R.L. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation", IEEE Trans. Inform. Theory, Vol. 37, pp. 114-131 (1991).

[8] M. Ergen, S. Coleri and P. Varaiya, "QoS Aware Adaptive Resource Allocation Techniques for Fair Scheduling in OFDMA Based Broadband Wireless Access Systems", IEEE Trans. On Broadcasting, Vol. 49 (4), pp. 362-370 (2003).

[9] W. Rhee, and J. M. Cioffi, "Increase in Capacity of Multiuser OFDM System Using Dynamic Subchannel Allocation", in Proc. of the 51 IEEE Vehicular Techn. Confer, Tokyo, Vol. 2, pp. 1085-1089, Spring (2000).

[10] C.Y. Wong, R.S. Cheng and K.B. Letaief, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation", IEEE Journal on Selected Areas in Communi., Vol. 17 (10), pp. 1747-1758 (1999).

[11] J. Jang and K.B. Lee, "Transmit Power Adaptation for Multiuser OFDM Systems", IEEE Iournal on Selected Areas in Communic., Vol. 21 (2), Feb. (2003).

[12] E. Hossain, D. Kim, V.K. Bhargava, "Analysis of TCP Performance Under Joint Rate and Power Adaptation in Cellular WCDMA Networks", IEEE Trans. On Wireless Communication, Vol. 3 (3), pp. 865-879 (2004).

[13] ComputingTCP’s Retransmission Timer, RFC 2988, Proposed Standard, Nov. (2000).

[14] D. Rajan, A. Sabharwal and B. Aazhang, "Delay Bounded Packet Scheduling of

Bursty Traffic Over Wireless Channels", IEEE Trans. on Inform. Theory, Vol. 50 (1), pp. 125-144 (2004).

[15] Kleinrock, Queueing Systems, New York: Wiley, Vol. 1 (1975).

[16] J.D.C. Little, "A Proof of the Queuing Formula L=λW", Oper. Res., Vol. 9, pp. 383-387 (1961).

[17] H. Balakrishnan, V.N. Padmanabhan, S. Seshan, R.H. Katz, "A Comparison of Mechanisms for Improving TCP Performance Over Wireless Links", IEEE/ACM Trans. On Networking, Vol. 5. (6), pp. 756-769 (1997).

[18] J. Padhye, V. Firoiu, D.F. Towsley, J.F. Kurose, "Modeling TCP Reno Performance: A Simple Model and its Empirical Validation", IEEE/ACM Trans. On Networking, Vol. 8. (2), pp. 133-145 (2000).

Received February 17, 2007
Accepted October 22, 2007