QoS-Guaranteed Admission Control for OFDMA-based Systems

Ramkumar, Venkata; Anggorojati, Bayu; Lucian, Stefan P. Andrei; Prasad, Neeli R.; Prasad, Ramjee

Published in:
Globecom. I E E E Global Telecommunications Conference

DOI (link to publication from Publisher):
10.1109/GLOCOMW.2010.5700392

Publication date:
2010

Document Version
Early version, also known as pre-print

QoS-Guaranteed Admission Control for OFDMA-based Systems

Ramkumar M. V, Bayu Anggorati, Andrei Lucian Stefan, Neeli R. Prasad, Ramjee Prasad

Center for Teleinfrastruktur (CTIF), Aalborg University, Aalborg, Denmark
{rk, ba, andrei, np, prasad} @es.aau.dk

This paper proposes a novel admission control (AC) algorithm for guaranteed quality of service (QoS) to all users. The proposed solution provides better utilization of system capacity using adaptive modulation (AM). A scheduler based on a per user priority function is also given in this paper. The AC is given by allocating the exact number of slots for each user that will meet its QoS. For every new user the number of slots required to meet its service requirements is estimated based on its channel quality information (CQI), packet arrival rate and buffer length. Using the average QoS achieved, the satisfaction index (SI) and priority is calculated for every user, which is used as key input for the scheduler. Further, the resource allocation in time and frequency for Orthogonal Frequency Division Multiplexing (OFDMA) systems is also discussed based on user satisfaction and number of slots required for each user. Finally, the proposed design is validated with OFDMA systems, but can be extended to any wireless system.

***Index Terms*— Admission control, Cross-layer, OFDMA, RRM**

I. Introduction

Next generation wireless networks support various kinds of services with different QoS requirements. These services are provided through the wireless medium, which is dynamic due to fading and the mobility of the user. The main challenge for an operator is to efficiently provide a variety of services in dynamic propagation conditions. The efficiency can be defined in different ways by considering various parameters. Here, we consider efficient usage of system resources by the better utilization of the bandwidth along with the QoS and fairness to all the existing users irrespective of the system conditions like load and channel conditions like fading etc.

The radio resource management (RRM) [1] [2] is the key feature that needs to be considered in next generation wireless systems. The main features of RRM are handoff and AC, and AC is the first thing that needs to be addressed in RRM, as even in case of handoff from one cell to another, the user has to be admitted into the cell before performing handoff. The user admitted in the cell can affect the system conditions and it also affects the QoS for the existing users. A wrong admission congests the network and also degrades the QoS of existing users.

In this work the admission of a new user is done by considering these issues by using a cross layer approach with adaptive modulation and coding (AMC) in the physical layer and a scheduler in the MAC layer. AMC is used in the physical layer for the efficient bandwidth utilization for a given error performance. The scheduler plays an important role in the provision of QoS. Here the design of the scheduler is based on the priority. Each user is given a priority, which is calculated based on the level of user satisfaction. In this work, the radio Resource Allocation (RA) is done on the DL OFDMA, by which a better radio link capacity is achieved exploiting multiuser diversity with a dynamic allocation in both time and frequency dimensions.

In [4], Qingwen proposed a cross layer scheduling algorithm with QoS guaranteed for the user. In this paper we extend the work in [4] for an OFDMA based system. The RA is done in time and frequency; hence the diversity is achieved in two domains. Also we use a multiple-input and multiple-output (MIMO) system, and hence the spatial diversity is also achieved. The diversities in temporal, spectral and spatial domains result in support for diverse QoS guarantees, which are not considered in [4].

In [5] Jia Tang proposed a cross layer RA for OFDMA systems, but did not consider scheduling, which is one of the important features for QoS. Also in [6], the author proposed utility based cross layer AC for OFDMA systems, but the RA and scheduling were not considered. Also [5] and [6] are not for OFDMA based systems.

The proposed framework is not related to any specific standard and can be used in any next generation air interface technology based on OFDMA.

The paper is organized as follows. The system model is described in Section II and the priority based scheduling based on user satisfaction is described in Section III. The radio RA for the OFDMA frame is explained in Section IV and the AC algorithm in Section V. The results based on simulations in MATLAB are given in Section V and Section VII concludes the paper and proposes the follow up work.

II. System model

Figure 1 explains the cross layer system model used in this work. The objective of this model is to perform the AC algorithm for an OFDMA-based system, by taking into account the effects of the queue in the scheduler, AMC in choosing the right mode based on the received SNR and packet error rate (PER), RA based on CQI and priorities, and MIMO with 2x2 antennas.

The AC algorithm is triggered when a new user sends a request. The new user can also be due to handoff, which needs to be admitted into the cell based on the resources available. The user sends a request with its service requirements and with the Doppler frequency and fading index. Based on the type of service, QoS and channel parameters the AC algorithm estimates the number of slots needed by the user to meet its requirements, which is explained in section V. If the estimated number of slots is available in the system then the user is admitted.

The priority based scheduler receives the QoS achieved by the user, based on which it calculates the level of user satisfaction - SI. The user with the least SI gets the highest priority in the RA. Hence the user priorities are sent to RA by the scheduler. RA is done based on the priorities from the scheduler and from the preferred slot input from the user.

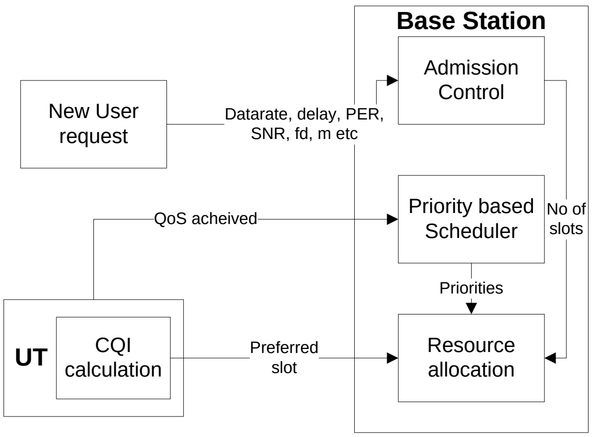

Fig. 1. System model for AC.

The preferred slot is calculated from the CQI, which is calculated by each user in each subcarrier. Then the CQI per slot is calculated, which is sent to the base station. CQI can also be calculated by the base station by using the channel estimation for every user, which will reduce the overhead sent by user. In this work we assume the base station has perfect knowledge about the channel.

III. Priority Based Scheduler

The main function of the priority based scheduler is to schedule the users based on the estimated priority, so that the user with highest priority is scheduled first, by which fairness will be achieved for each user. The amount of resources to be allocated to each user is estimated by the AC algorithm based on the QoS. The scheduler calculates the priorities for each user and sends these to RA. The priority is calculated for each user based on the achieved QoS by the user. For each user the SI is calculated, which gives the level of user satisfaction with the throughput and delay achieved with respect to the desired values. The desired values of QoS are assumed to be dependent on the type of service.

Here we consider four types of service classes, which are shown in Table I. Class 1 users, have equal weights for rate and delay. Class 2 is for real time users like video applications, Class 3 users have more weight for the rate and Class 4 users are the best effort users. The weights for each class are used in calculating the priority function.

TABLE I
SERVICE CLASSES AND SI COEFFICIENTS

Class	Rate	Delay	ω_u^{rt}	ω_u^{nrt}	notes
1	200kbps	50ms	1	1	High rate and low delay
2	-	50ms	1	0	Low delay
3	200kbps	-	0	1	High rate
4	-	-	0	0	Best effort

The SI [7] is represented as a function of delay $\Gamma_u(t)$ (Eq. 1) or as a function of rate $\psi_u(t)$ (Eq. 2). In either case, the lower the SI, the higher is the priority that a user will be assigned.

$$\Gamma_u(t) = \begin{cases} \dfrac{T(u)-\Delta T(u)}{\omega_u(t)} & if \quad \omega_u(t) < T(u)-\Delta T(u) \\ 1 & otherwise \end{cases} \tag{1}$$

where $\Delta T(u)$ is a safety margin.

The delay component SI is expressed referring to the head of line delay (HOL), ω_u, and the maximum delay for service *u*, $T(u)$, while the rate component is expressed in terms of the average rate measured, η_u, and the desired data rate, $\hat{\eta}_u$.

$$\psi_u(t) = \frac{\eta_u}{\hat{\eta}_u - \Delta\hat{\eta}_u} \tag{2}$$

where $\Delta\hat{\eta}_u$ is the margin coefficient.

The priority function given by Eq. 5 has two components ϕ_u^{rt} and ϕ_u^{nrt}. The expressions for these two components are given below (Eq. 3 and 4):

$$\phi_u^{rt} = \left\{ \begin{array}{ll} R_u \dfrac{1}{\Gamma_u(t)} & \Gamma_u(t) \geq 1; R_u \neq 0 \\ 1 & 0 < \Gamma_u(t) < 1; R_u \neq 0 \\ 0 & R_u \neq 0` \end{array} \right\}$$

$$\phi_u^{nrt} = \begin{cases} R_u \dfrac{1}{\psi_u(t)} & \psi_u(t) \geq 1; R_u \neq 0 \\ 1 & 0 < \psi_u(t) < 1; R_u \neq 0 \\ 0 & R_u \neq 0 \end{cases} \tag{4}$$

In both components, $R_u = R_F / R_{\max}$ is the a-dimensional coefficient expressing the sustainable data rate on the current frame F determined by the AMC normalized to the maximum PHY data rate. The sustainable data rate on frame F is based on the mean received SNR and target PER and the direct proportionality to the priority function is another step towards achieving fairness in the presented system. R_u makes sure that the user with better channel quality gets more priority than the user will lower channel quality.

Using the above equations, the calculation of priority function ϕ_u is shown below (Eq. 5):

$$\phi_u = \omega_u^{rt}\phi_u^{rt} + \omega_u^{nrt}\phi_u^{nrt} \tag{5}$$

where the weight coefficients ω_u^{rt} and ω_u^{nrt} are determined based on the service type described in Table I.

IV. Resource Allocation

The RA is done on the OFDMA grid of subcarriers spanning the time-frequency domain. On the time-axis, the OFDM symbols are represented while on the frequency axis a certain number of subcarriers are represented depending on the FFT length used [8]. For the two axes we define two partition factors - Λ_t and Λ_f. In OFDMA, the minimum allocation unit is represented by a slot. Thus, one OFDMA frame is composed of $\Lambda = \Lambda_t \cdot \Lambda_f$ slots.

The OFDMA parameters are presented in Table II. The PHY layer parameters are chosen according to the IEEE 802.16 standard and payload mapping is done according to adjacent subcarrier permutation, both defined in [9].

TABLE II
PHY layer parameters

Parameter	Value
FFT points	128
Pilot sub	12
Left guard sub	10
Right guard sub	19
Data sub	108
OFDM symbols	48
CP length	11.43 µs
OFDM symbol time	102.86 µs
Λ_t	12
Λ_f	16
Slot dimension	4*6

The allocation of slots to the users is based on the CQI calculated at the user side. Each user calculates the CQI for each subcarrier and averages this information to find the CQI in each slot before sending this information to the base station. By receiving the CQI calculation, the base station determines the preferred slot for each user and assigns a priority $\Phi_u(t, \Lambda_i, \Lambda_j)$ to this best slot. This information is spread to the neighbor slots to calculate the priorities of the user in other slots, which is useful in case two or more users denote the same preferred slot. The spreading is based on the Euclidean distance (Eq. 6) between the preferred slot and all other slots:

$$l_{\hat{i}\hat{j},ij} = \sqrt{(\lambda_i - \hat{\lambda}_i)^2 + (\lambda_j - \hat{\lambda}_j)^2} \qquad (6)$$

where i $\in$ [1; Λ_t] and j $\in$ [1; Λ_f].

Taking into consideration the spreading factor, the priorities that each slot is assuring are given by (Eq. 7):

$$\Phi_u(t, \Lambda_i, \Lambda_j) = (1 - \alpha_\lambda \cdot l_{\hat{i}\hat{j},ij}) \qquad (7)$$

where α_λ is the coefficient denoting the loss by assigning the user to a slot other than the one considered the preferred one. By using the spreading function we are assuring that even if a service does not receive the best slot for its request, slots are assigned as to maximize the priority to a value close to the priority value calculated in the scheduler for that particular service. Hence the goal of the RA is to maximize the sum of the priorities of all slots.

V. Admission Control

For each new request the admission control calculates the number of slots required to meet the QoS requested by the user. For this we use Markov analysis to predict the number of slots. Each state of the Markov Chain (U, C), is represented by a pair where index *U* indicates the number of data packets waiting in the queue and *C* indicates the number of packets transmitted in the OFDMA time frame. If the buffer length is K then U can take a value between 0 and K. The number of packets transmitted depends on the number of slots allocated to the user and the AMC mode used by the user. Hence

$$Cn = bRn, \qquad n = 1, 2, \ldots\ldots, N \qquad (8)$$

where b is the number of slots allocated to the user and R_n depends on the AMC mode. C can take any value in $\{C_0, \ldots C_N\}$ where N is the number of modes in AMC. The modes used in AMC are shown below in Table III.

TABLE III
AMC parameters

Transmission mode	Modulation	R_n	R_u
0	No transmission	0	0
1	4-QAM	2	2/6
2	16-QAM	4	4/6
3	64-QAM	6	1

The number of packets waiting in the queue depends on the arrival process A and on the service process and the buffer length allocated to the user. The service process status depends on the CQI. The service process can be in one of the transmission modes reported with a pre-defined probability, see (4) in [10], derived from the average SNR, the Doppler Frequency and the Nakagami fading index *m*. As the number of packets transmitted depends on the number of slots and AMC mode (Eq. 8), the probability of the service process changing from one state to another depends on the transition probability of a user changing from one AMC mode to another mode, by assuming that the number of slots allotted to the user is fixed.

In [11], SNR was divided in adjacent regions based on the desired BER. Then the transition probabilities between the various SNR regions were determined based on the Level Crossing Rate (LCR) of the channel fading distribution.

From the desired BER and the BER expression for Rectangular Quadrature Amplitude Modulation (QAM) modulation over the Nakagami fading channel, Eq.8.108 in [13], the AMC procedure determines a partition of the SNR in N +1 regions $\{(-\infty, \gamma_0], (\gamma_0, \gamma_1], (\gamma_1, \gamma_2], (\gamma_2, \infty]\}$. In this way, from the average SNR on the radio link, we determine the modulation to be used in order to satisfy $BERu < BER_{target}$. The transmission mode *i* is used when $\gamma_{av} \in (\gamma_{i-1}, \gamma_i]$ for $i >$ 0 and transmission mode 0 is used when $\gamma_{av} < \gamma_0$.

The arrival process is modelled with a Poisson distribution.

$$P(a) = \lambda^a e^{-\lambda} / a! \qquad (9)$$

where $a \geq 0$. $E\{A\} = \lambda$ ¸ is the packet arrival rate: The average number of packets arriving over one second and it depends on the traffic model.

The Packet Loss Rate (PLR) is composed of two factors: The PER due to channel fading statistics and the PER due to dropped packet because of the finite buffer length *K*. The total PLR can be expressed as

$$PLR = 1 - (1 - P_d)(1 - P_0) \qquad (10)$$

where P_0 is the PER due to channel fading and P_d is the dropping probability. P_d depends on the packet arrival rate *E{A}* and the buffer lengths.

From the steady state distribution of the two-dimensional Markov Chain $P(U=u,C=c)$, expected number of packets dropped from queue D can be expressed as in [10], from which P_d is calculated as shown below.

$$E\{D\} = \sum_{a\in A, u\in U, c\in C} \max\{0, a-K+\max\{0,u-c\}\} . P(A=a) . P(U=u, C=c) \quad (11)$$

$$P_d = E\{D\} / \lambda T_f \quad (12)$$

From P_d, the PLR is determined as in Eq 10 and throughput is calculated as

$$\eta_{prior} = \lambda(1-PLR) \quad (13)$$

The expected delay is derived as follows. Let *Nw* be the number of average packets waiting in the queue plus the number of average packets transmitted in one frame. This is expressed using steady probability $P(U=u,C=c)$.

Following Little's Theorem [12], we can derive the average

$$\tau_{prior} = N_w / E\{A\}(1-P_d) \quad (14)$$

Hence the throughput and delay for the new user can be estimated by assuming the user is allocated *b* slots. By increasing the number of slots allocated to the user, the throughput achieved by the user will increase and the delay will decrease. Hence the AC algorithm finds the right number of slots needed by the user for achieving the requested delay and throughput.

$$b + \sum_{j\in J} b_j \le N_d \quad (15)$$

J is the total number of users available in the system and b_j is the number of slots allocated to user *j*. If the total number of slots used in the system plus the number of slots required for the new user is less than the total number of slots N_d in the system, then the user is admitted.

VI. NUMERICAL RESULTS

The simulation setup used to measure the performance of the AC is as follows. We consider four types of users with different requirements as given in Table I. For each service the target PER is taken as 0.05 and the target SNR is 27dB. The base station can transmit to each user with a different modulation scheme in every frame, which is decided based on the received SNR and the PER requirement of the service. Here we use 3 modes of transmission, which are 4-QAM, 16-QAM and 64-QAM, based on the SNR, if the received SNR is below a certain threshold then mode 0 is used, where there is no transmission. 2x2 MIMO system is used, and between each antennas the Nakagami fading channel is simulated with fd = 10Hz and fading index m = 1.

The base station transmits to each user, in their preferred slots, which is measured by the user based on CQI. The OFDMA modulation is used for the transmission, where the frame duration T is 5ms. Each frame has 48 symbols and each symbol contains 128 subcarriers. Each slot is defined as 4 symbols of 8 subcarriers, which constitutes 32 subcarriers in each slot. The total number of slots that can be allocated to the users is 192.

The packet arrival rate and buffer length are the key parameters in the Markov analysis to predict the number of slots needed by the new user. The values used are shown in the table below.

TABLE IV
PACKET ARRIVAL PARAMETERS

Class 1 and 3		Class 2 and 4	
Buffer length	40	Buffer length	10
Avg. pkt arrival rate	6 pkts/frame	Avg. pkt arrival rate	1 pkt/frame

Figure 2 shows the average throughput of the users in class 1, 2 3, and 4 over time, without AC, where each user gets a fixed allocation of 5 slots in every frame. It can be seen clearly that for class 1 and 3 users the throughput is not satisfied and is around 120 kbps not fulfilling the required 200kbps and for class 2 and 4 users the average throughput is 80 kbps, which is above the required QoS. Hence the QoS is not guaranteed for class 1 and 3 users and for class 2 and 4 users the bandwidth is not utilized efficiently.

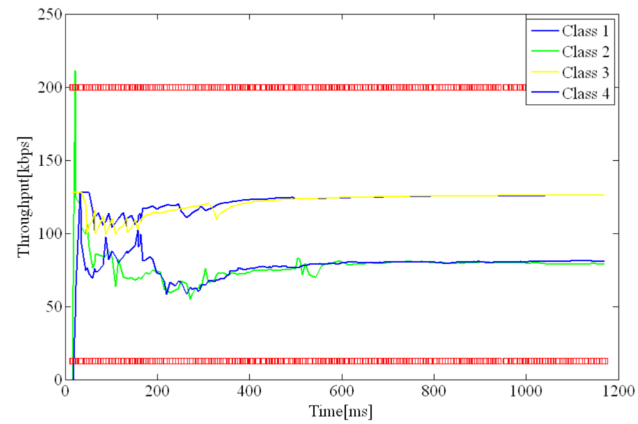

Fig. 2. Throughput with fixed allocation

Figure 3 shows the average throughput, when the system has only class 1 users. The average throughput of Class 1 users was plotted w.r.t the number of users in the system. It can be seen that as the load or number of users increases the average throughput of the existing users is maintained consistently.

Figure 4 shows the average throughput obtained by class 1 and 3 users over time. The system reaches its maximum load at the end of the simulation, where all slots are allocated to all users. When the system reaches maximum load, there are 12 users of class 1, 7 users of class 2, 9 users of class 3 and 7 users of class 4. In total 35 users in the system. Hence it can be observed that as the system reaches its maximum usage of slots the QoS for class 1 and 3 is maintained. For class 1 and 3 users the throughput is above the 200kbps consistently. The throughput achieved by class 2 and 4 users is shown in Figure 5 where it can be seen that the throughput achieved is consistently greater than 12.8 kbps. Hence, using the AC, the throughput is guaranteed for each class of user.

In Figure 6 the delays achieved by class 1 and 2 users are plotted, as both are delay sensitive users. It can be observed that the delay starts with zero, in the start of simulation where there are less users and as the simulation progresses the load increases to maximum. The average delay of each class is within the required limit of 50 ms.

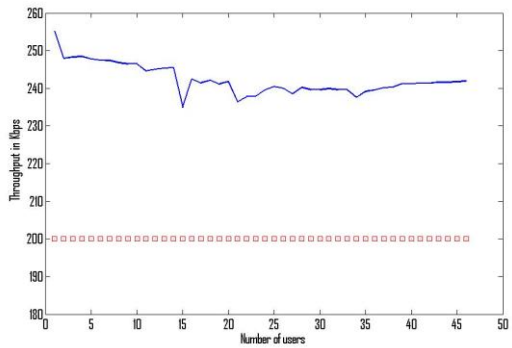

Fig. 3. Throughput with only class 1 users

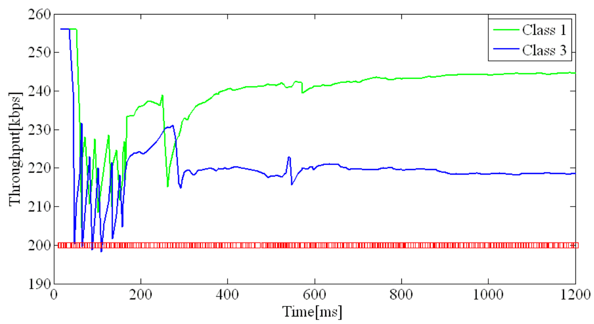

Fig. 4. Throughput for different class 1 and 3 users

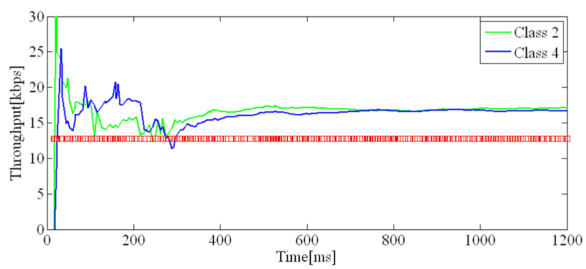

Fig. 5. Throughput for different class 2 and 4 users

VII. Conclusion

In this paper we proposed an AC algorithm based on schedulers in MAC, AMC and MIMO with RA based on priority and CQI by considering cross-layer effects in the admission of a new user. Using this the number of slots required to meet a user's QoS requirement is estimated, hence the QoS is guaranteed to each user irrespective of the system and channel conditions. By using the AMC the bandwidth is utilized efficiently in all the channel conditions and by achieving diversity in time and frequency using RA and spatial diversity using MIMO, this system can support users with varied QoS. Here we showed for four types of services the achieved differences in QoS, or the lack of fairness, will be taken care of by the scheduler by giving more priority to users with less satisfaction.

The main application of this is in RRM for next generation networks. The future work will be to extend for heterogeneous wireless networks, where the AC should select the most suitable cell and radio access technology for the user that can cater the best QoS to the user while maximizing the overall

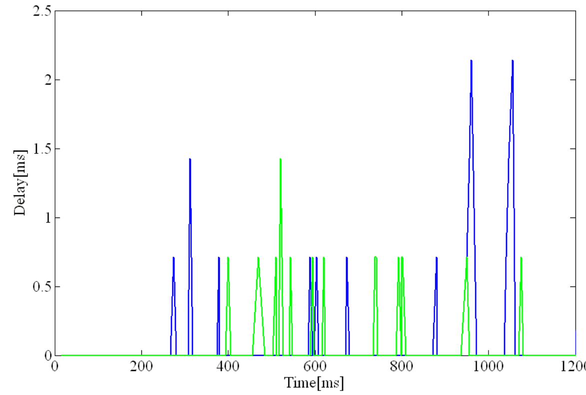

Fig. 6. Delay profile for classes 1 and 2

capacity of the heterogeneous network. Also this work can be used in self optimization of the network, by taking into account the cross layer dependencies, like dynamic buffer allocation etc.

Acknowledgements

This work was supported in part by the European Commission, in the context of the project FUTON "Fibre Optic Networks for Distributed, Extendible Heterogeneous Radio Architectures and Service Provisioning", grant agreement FP7 ICT-2007-215533.

References

[1] M.V. Ramkumar, "Management of Distributed and Extendible Heterogeneous Radio Architectures," NAEC2009, Lake Garda, Italy, October 8-11, 2009

[2] Carlos Santiago at al, Next Generation Radio over Fibre Network Management for a Distributed Antenna System

[3] M.V. Ramkumar et al, "FUZZY-LOGIC BASED CALL ADMISSION CONTROL FOR A HETEROGENEOUS RADIO ENVIRONMENT" 12th International Symposium on Wireless Personal Multimedia Communications (WPMC 2009), September 7 – 10, 2009

[4] Qingwen Liu et al, "Cross-Layer Scheduling With Prescribed QoS Guarantees in Adaptive Wireless Networks*," IEEE Journal on selected areas*, vol. 23, No 5, May 2005

[5] Jia Tang et al, "Cross layer design of dynamic resource allocation with diverse guarantees for MIMO-OFDM wireless network," ICC 08, 19-23, May 2008, Pages 205-212

[6] Hu Nan et al, "Integrated cross-layer design of utility-based connection admission control in packet switched OFDM wireless networks,"2007, IWCLD, 20-21 Sept 2007, Pages(s) 152-156

[7] Qingwen Liu et al, "A Cross-Layer Scheduling algorithm with QoS Support in Wireless Networks," Transactions on Vehicular technology, Vol 55 No 3 May 2006

[8] F. Meucci et al, "Joint Resource Allocation and Admission Control Mechanism for an OFDMA-Based System," in Proc. of WPMC '08 Lapland, Finland, September 8-11, 2008

[9] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," *IEEE Std* 802.16-2004 (Revision of IEEE Std 802.16-2001), pp. 1–857, 2004.

[10] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer modelling of adaptive wireless links for QoS support in heterogeneous wired-wireless networks," *Wireless. Networking.* vol. 12, no. 4, pp. 427–437, 2006.

[11] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," Communications, IEEE Transactions on, vol. 47, no. 11, pp.1688–1692, Nov 1999

[12] R. B. Cooper, *Introduction to Queuing Theory*, 2nd ed. Elsevier North-Holland, 1981

[13] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels, 2nd Edition*. Wiley-IEEE Press, 2005