Cross-layer Quality of Service for video wireless multimedia delivery: some challenges and principles

ZORAN BOJKOVIC(1), DRAGORAD MILOVANOVIC(2)

(1)Faculty of Transport and Traffic Engineering, University of Belgrade,
Vojvode Stepe 305, 11000 Belgrade,

(2) Faculty of Electrical Engineering, University of Belgrade,
Bulevar Revolucije 73, 11120 Belgrade, Serbia

Abstract: *Providing cross-layer quality of service (QoS) for wireless multimedia delivery is challenging task which has attracted intensive research over the years. A fundamental problem in this area is how to map QoS criterion at different layers and optimum QoS across the layers. We present end-to-end QoS architecture for multimedia delivery as well as QoS metrics over wireless links together with channel model and QoS evolution. Then, we invoke a cross-layer design for multimedia scheduling at the data link layer, with each user employing adaptive modulation and coding (AMC) at the physical layer. The cross layer design provides low-complexity implementation and analysis as well as service isolation and scalability. Also, the QoS of a certain user in terms of its throughput, packet bit rate and average delay for a given bandwidth allocation is provided.*

Key-words: QoS cross-layer design, adaptive modulation and coding (AMC), medium access control (MAC)

1 Introduction

Wireless communications has emerged as one of the largest sectors of the telecommunications industry evolving from a business in the last decade to one of the most promising areas for growth in the 21st century. The rapid growth of wireless communications and access, together with the success of the Internet has brought a new era of wireless multimedia applications and services. Enormous recent developments have been undertaken by both academia and industry [1]. The high-speed multimedia service is one of the driving forces of the next wireless LAN generation. However due to wireless channel characteristics and lack of QoS support, the basic 802.11-based channel access procedure's merely sufficient to deliver non-real time traffic [2]. The delivery should be augmented by appropriate mechanisms to better consider different QoS requirements and adjust the medium access parameters to the multimedia data content characteristics [3]. Providing multimedia services to mobile and fixed users through wireless access, can be a reality with the development of two high-speed physical (PHY) layers IEEE802.11g (54 Mbps), and IEEE802.11n (100 Mbps) as well as the IEEE802.11e quality of service (QoS)-based medimum access control (MAC) layer [4, 5].

With the development of fourth-generation (4G) wireless standards, new broadband video applications can be offered to mobile users [6]. In addition to delivering high bit rate multimedia applications, 4Gsystems are also expected to provide multiple QoS guarantees to different types of user applications. An important issue in providing multiple QoS guarantees to video applications in wireless systems is dynamic QoS management for services with mobility support. A dynamic QoS management system allows video applications and the underlying prioritized transmission system to interact with each other in order to cope with service degradation and resource constraint in a time-varying wireless environment [7].

QoS metrics of a connection include data throughput, packet error/loss rate and delay performance. Usually, multimedia applications can be classified in two categories: QoS guaranteed and best efforts one. The service for a connection is considered to be *best-effort* when no guarantees on delay and throughput are provided. Nearly error-free reception of information is typically required for such services [8]. The first category includes voice, video/audio streaming, video/audio telephony and conferencing. Applications such as web-browsing, e-mail and file transfer protocol (FTP) belong to the second category.

Powerful forward error correction (FEC) coding or automatic-repeat-request (ARQ) protocols can reduce the loss rate, at the expense of increased bandwidth and delay [9]. However, allocating a fixed amount of bandwidth to each user may not be as efficient, because the queues may be empty from time to time due to the dynamic nature of the traffic. Scheduling plays an important role in QoS provision. A challenge to scheduler designs is predicting role on QoS provision. A challenge to scheduler designs is predicting all three aspects of QoS, namely, throughput, loss and delay. This motivates the design and performance evaluation of schedulers guaranteeing prescribed QoS with efficient resource utilization, over wireless fading links [10]. Efficient bandwidth utilization for a prescribed error performance at the physical layer can be accomplished with adaptive modulation and coding (AMC) schemes, that match transmission parameters to the wireless channel conditions [11]. However, most existing AMC designs are tailored for the physical layer. Their impact on, and interaction with, higher protocol layers remain largely unexplored.

To address these issues, we present QoS architecture that takes into account cross-layer QoS for wireless multimedia delivery, which include: key components for end-to-end QoS support, framework for end-to-end QoS support for video delivery in the case of network-centric cross-layer solutions, as well as general diagram for end-system centric QoS provisioning. The rest of the paper deals with QoS metrics over wireless links, channel modeling and QoS evaluation. The finale section concludes this paper.

2 Challenges to cross-layer optimization

In order to meet the challenges of wireless access, various Open System Interconnection (OSI) network layers must be considered together when designing the network. Traditional OSI layers are shown in Figure 1 [12].

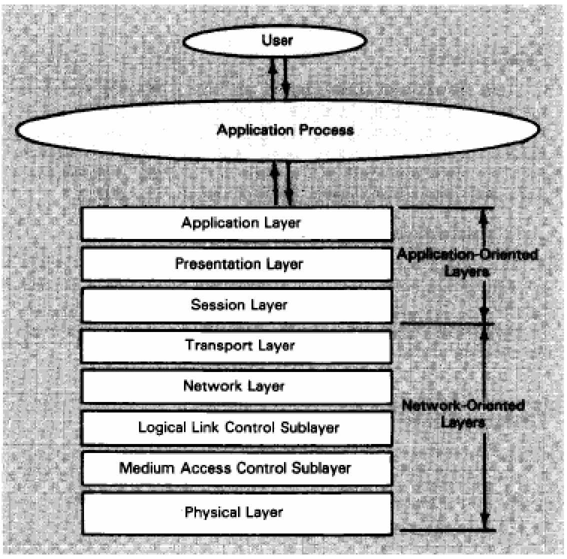

Figure 1. OSI layers.

Quality of Service (QoS) requirements that can and will vary according to application will force the network layer to account for the physical-layer design when optimizing network throughput. Different applications are better served by different optimizations. This leads to a design methodology that blurs the lines between layers and attempts to optimize across layer functionality.

Cross-layer network design is an important step when attempting to optimize new multimedia networks. However, it is still a step below what will be necessary to truly maximize the performance of future networks. True optimization will not only require cross-layer design, but also cross layer adaptability. Traditionally, networks have contained some ability to adapt. For example, many communication systems can adjust to changing channel conditions using signal processing methods, or, to changing traffic loads by adjusting routing tables. These adjustments have been isolated to a specific layer: cross-layer adaptability will allow all network functions to pass information between functions and adapt simultaneously. Such adaptability will be required to meet the demand of changing requirements along with changing network loads and channel conditions [13]. While the cross-layer design requires static optimization across network layers, adaptability requires dynamic optimization across layers.

Here are several challenges and research issues associated with the vision of cross-layer optimization. First, full network design and optimization is extremely complicated. This is particularly true when attempting real-time dynamic optimization. Some attempt must be made to determine design methodologies that encompass the incredible freedom offered to the designer when cross-layer optimization is possible.

A second problem involves the metrics to be used in the optimization. Network layers have their own isolated optimization criteria. For example, physical-layer design is primarily focused on minimizing the bit error rate while the medium Access Control (MAC) layer design is concerned with node throughput or channel availability.

A related issues arises in the context of dynamic optimization. In dynamic optimization, information is passed between the network layers. The system designer must choose the information to be passed. It must not be overly complicated for risk of creating large delays or computationally expensive optimization routines. However, it cannot be overlay simplistic for the risk of communicating too little information.

The design of such systems requires sophisticated modeling (simulation) procedures. Some possible options include:

- Combined simulation and semi-analytic approaches that simulate high-level functionality and use semi analytic simulation approaches to approximate lower level functionality.
- Combined simulation and hardware approaches that use hardware to perform lower-level functionality.
- Variable-granularity approaches that use a network simulator with coarse granularity for a majority of physical layer links and fine granularity for links of specific interest.
- Emulation and real-time processing involving all facets, from physical layer to application, simultaneously.

These hybrid approaches have to be firmly established. A final research issue in the area of dynamic network optimization concerns network control. Namely, when functionality across layers is allowed to adapt, it is important that something has control of the process.

To provide QoS for multimedia applications, the IEEE802.11 Working Group has currently defined a new supplement, to the existing legacy, 802.11 MAC sublayer, called IEEE802.11e [14]. It should be noted that even though emerging MAC standards provide QoS support, there are no QoS guarantees for multimedia applications and system-wide resource management is not always fair or efficient. This is due to the time-varying nature of the wireless channel and multimedia characteristics, and also the lack of cross-layer awareness of the application and MAC layers about each other.

3 End-to-end QoS for multimedia delivery

To support end-to-end quality of service (QoS) for wireless multimedia delivery, there are several fundamental challenges like [15]:

- QoS support encompasses a wide range of technological aspects
- different applications have very diverse QoS requirements in terms of data rates, delay bounds, and packet loss probability
- different types of networks inherently have different characteristics
- there is heterogeneity among end users
- QoS provisioning from networks
- multilayered scalable coding for applications
- network adaptive congestion/error/power control in end-systems.

Figure 2 illustrates key components for end-to-end QoS support. End users have different requirements in terms of latency, visual quality, processing capabilities, power and bandwidth. Thus, it is a challenge to design a delivery mechanism that not only achieves efficiency in network bandwidth but also meets the heterogeneous requirements of the end users.

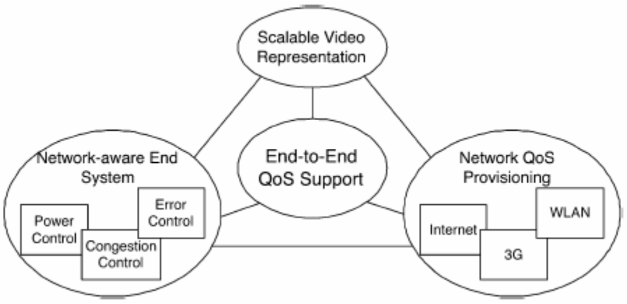

Figure 2. Key concepts for end-to-end QoS support [15].

There are two approaches in providing the end-to-end QoS support: the network-centric QoS provisioning and end-system centric QoS support. In network-centric QoS provisioning, routers/switches, or/and base stations/access points in the networks provide prioritized QoS support to satisfy data rate, delay bound and packet loss requirements by different applications. In the prioritized transmission, QoS is expressed in terms of probability of buffer overflow and/or the probability of delay violation at the link layer. At the video application layer, QoS is measured by the mean squared error (MSE) and/or peak-signal-to-noise ratio (PSNR). Thus, one of the key issues for end-to-end QoS provisioning using network-centric solution is the effective QoS mapping across different layer. More specifically one needs to consider how to model the varying network and coordinate effective adaptation of QoS parameters at video application layer and prioritized transmission system at link layer.

Different layers (e.g., application layer and link/network layer) have different metrics to measure quality of service, which brings challenge for end-to-end QoS provisioning. Figure 3 shows the general block diagram of end-to-end QoS support for video delivery in the network-centric cross-layer solution. This solution considers an end-to-end delivery system for a video source from the sender to the receiver, which includes source video encoding, cross-layer QoS mapping and adaptation, prioritized transmission control, adaptive network modeling, and video decoder/output modules. To support end-to-end QoS with network-centric approach, a dynamic QoS management system is needed in order for video applications to interact with underlying prioritized transmission system to handle service degradation and resource constraint in time-varying wireless Internet. To offer a good compromise between video quality and available transmission resource the key is how to provide an effective cross-layer QoS mapping and an efficient adaptation mechanisms.

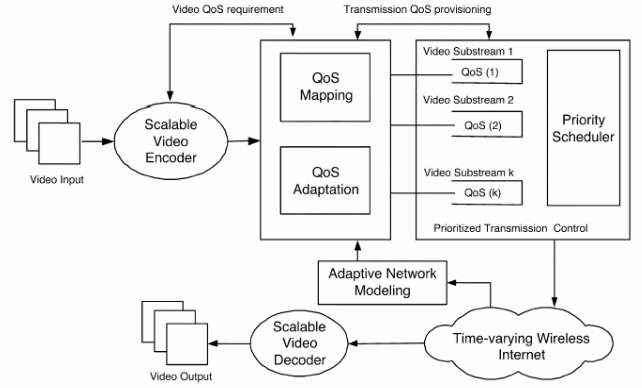

Figure 3. A framework of end-to-end QoS support for video delivery in the case of network-centric cross-layer solution [15].

The second type of approach to provide end-to-end QoS support is end system centric. The end systems employ various control techniques, which include congestion control, error control and power control to maximize the application layer video quality without any QoS support from the underlying network. The advantage of end system control is that there are minimum changes required in the core network. However, the main challenge is how to design efficient power/congestion/error control mechanisms.

To provide end-to-end QoS with end-system solution, the video applications should be adaptive to the variation of network congestion in wireless Internet. This adaptation consists of network adaptation and media adaptation. The network adaptation refers to haw many network resources (bandwidth and battery power) a video application should utilize for its video content, i.e., to design an adaptive media transport protocol for video delivery. The media adaptation controls the bit rate of the video stream based on the estimated available bandwidth and adjust error and power control behaviors according to the varying wireless Internet conditions.

General framework for end-to-end QoS provisioning for video over wireless Internet with end-system-centric solution is shown in Figure 4. To address network adaptation, an end-to-end video transport protocol is needed to handle congestion control in wireless Internet.

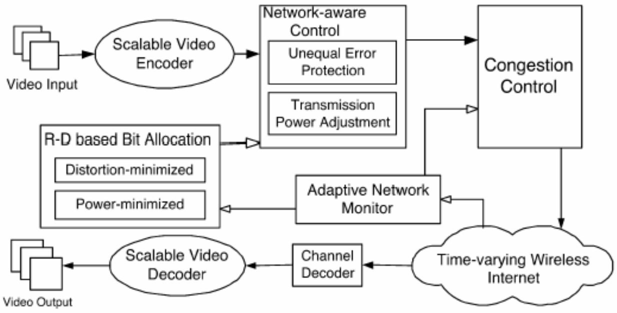

Figure 4. A general diagram for end-system centric QoS provisioning [15].

The Adaptive network monitor deals with probing and estimating the dynamic network conditions. The Congestion Control module adjusts sending rate based on the feedback information.

For media adaptation, considering that different parts of compressed, scalable video bitstream have different importance level, Network-aware Unequal Error protection (UEP) mobile protects different layers scalable video against congestive packet losses and erroneous losses according to their importance and network status. Network-aware Transmission Power Adjustment module adjusts the transmission power of the end-system to affect the wireless channel conditions. *R-D* based bit allocation module performs media adaptation control with two different targets: distortion minimization and power consumption minimization.

4 QoS metrics over wireless links

In order to obtain QoS metrics over wireless links, we will model the queuing the queuing arrival process and service process. Let A_t denote the number of packets arriving at time *t*. Assume that the process A_t is stationary with and is independent of the queue state as well as of the channel state. If A_t is *Poisson* distributed with parameter (packet/frame), it will be

$$P(A_t = a) = \lambda^a e^{-\lambda} / a!, \quad a \in A \qquad (1)$$

where $E\{A_t\} = \lambda$ and $A = \{0,1,...\infty\}$.

Adaptive modulation and coding (AMC) dictates a dynamic rather than deterministic, service process for the queue, capable of transmitting a variable number of packets per time unit (frame). Let C_t (packet/frame) denote this transmission capability (the number of packet that can be transmitted at time *t*). Corresponding to each transmission mode *n*, let C_t (packet/frame) denote the number of packets transmitted with AMC mode *n* per time unit. Then, we have

$$C_t \in C, C := \left\{ c_n : c_n = \frac{bR_n}{R_1}, \quad n = 0,1,...,N \right\} \qquad (2)$$

Where *b* is the bandwidth coefficient, i.e., the number of time slots reserved for the user. At the physical layer, the data are transmitted frame by frame through the wireless link. Each frame contains a fixed number of symbols N_s. Given a fixed symbol rate the frame duration (T_f seconds) is constant end represents the time unit. Using time-division multiplexing TDM, each frame is derived into $N_c + N_d$ time slots. For convenience, we let each time-slot contain a fixed number of N_b/R_1 symbols. As a result, each time slot can transmit exactly R_n/R_1 packets with transmission mode *n*. In particular, one time-slot can accommodate $R_1/R_1=1$ packet with mode n=1, $R_2/R_1=2$ packets with mode n=2 and so on. The packet and frame structures a re detailed in Figure 5. The N_c time slots contain control information and pilots. The N_d time slots convey data, which are scheduled to different users with time division multiple-access TDMA dynamically. Each user is allocated a certain number of time slots during each frame.

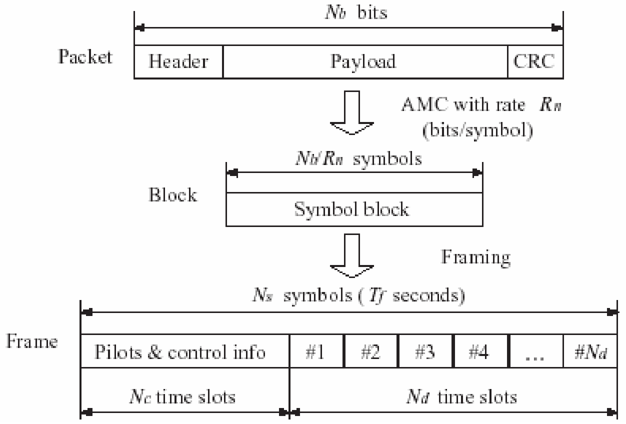

Figure 5. packet and frame structures with processing units at each layer [19].

As specified by (2), the AMC module yields a queue server with a total of N+1 $\{C_n\}_{n=0}^{N}$ states with the service process C_t representing the evaluation of server states. Since the AMC mode *n* is chosen the channel enters the state *n*, the service process C_t has to be modeled.

5 Channel model

One of the operating assumptions for flat fading channels is that for each user, the channel is frequency flat and remains invariant for frame, but is allowed to vary from frame to frame. This corresponds to a block fading model, which is suitable for slowly varying wireless channels [16]. Thus, AMC is adjusted on a frame-by-frame basis.

For flat fading channels, the channel quality can be captured by a single parameter, the received signal-to-noise ratio (SNR)γ. The channel varies from frame to frame. Thus, the general *Nakagami m* model can be adopted. It encompasses a large class of fading channels [17]. The received SNRγ per frame is a random variable with a *Gamma* probability density function

$$P_\gamma(\lambda) = m^m \gamma^{m-1} \exp\left(-\frac{m^\gamma}{\bar{\gamma}}\right) \tag{3}$$

where $\bar{\gamma} = E\{\gamma\}$ is the average received SNRn Γ(m) is the *Gamma* function $\Gamma(m) = \int_0^\infty t^{m-1}\exp(-t)dt$ while *m* is the *Nakagami* fading parameter ($m \geq 1/2$). Each user relives on AMC at the physical layer. The objective of AMC is to maximize the data rate by adjusting transmission parameters to channel variations, while maintaining a prescribed packet error rate P_0. Let *N* denote the total number of transmission modes available. Assuming constant power transmission, and partition the entire SNR range into N+1 nonoverlapping consecutive intervals, with boundary points denoted as $\{\gamma_n\}_{n=0}^{N+1}$ mode *n* chosen, when $\gamma \in \{\gamma_n, \gamma_{n+1}\}$. In order to avoid deep channel fades, no data are sent when $\gamma_0 \leq \gamma < \gamma_1$, which corresponds to the mode n=0 with rate R=0 [bits/symbol]. The design objective of AMC is to determine the boundary points, i.e., $\{\gamma_n\}_{n=0}^{N+1}$. Given P_0, $\bar{\gamma}$ and *m*, a threshold searching algorithm is developed in [19], so that the average packet error rate for each mode is exactly P_0. The *SNR* region $\{\gamma_n, \gamma_{n+1}\}$ corresponding to transmission mode *n* constitutes the channel states, we rely on a finite-state *Markov* chain (FSMC) model. The state transition matrix of such FSMC is

$$[P_c] = [P_{l,n}]_{(N+1)(N+1)} \tag{4}$$

which depends on the statistical channel parameters: average received SNR$\bar{\gamma}$, *Nakagami* fading parameter *m* and mobility-induced *Doppler* spread f_d. It can be also applied to more general transition matrices.

6 QoS evaluation

Let U_t, $U_t \in U := \{0,1,...,K\}$ denotes the queue state and (U_{t-1}, C_t) denotes the pair of queue and server states, whose variation is modeled as FSMC. The steady state distribution of exists and it is unique for the calculation of the steady-state distribution presented in the form

$$P(U = u, C = c) := \lim_{t\to\infty} P(U_{t-1} = u, C_t = c) \tag{5}$$

Based on the steady-state distribution $P(U=u, C=c)$, it becomes possible to evaluate the QoS for each user. Let P_d denote the packet dropping, i.e., overflow or blocking, probability upon the queue. Tacking into account $P(A_t=a)$ in (1) and $P(U=u, C=c)$ in (5), P_d can be computed. A packet is correctly received at the user side, only if it is not dropped from the queue with probability 1- P_d. it is correctly received through the wireless channel with probability 1-P_0. The packet loss rate is

$$\xi = 1-(1-P_d)(1-P_0) \qquad (6)$$

With the steady-state distribution , the average number of packet in the queue and in the transmission is (7)

$$N_{wl} = \sum_{u\in U,\, c\in C} uP(U=u, C=c) + + \sum_{u\in U,\, c\in C} \min\{u,c\} P(U=u, C=c)$$

The subscript *wl* stands for wireless link. In an attempts to find the coverage delay per packet through the wireless link, Little's Theorem [20] can be used to obtain

$$\tau = \frac{N_{wl}}{\lambda(1-P_d)} \qquad (8)$$

Thus, given the bandwidth coefficient *b*, target packet error rate P_0, *Doppler* spread f_d, average SNR $\overline{\gamma}$, *Nakagami* parameter *m*, buffer length *K* and data arrival rate λ, we can obtain QoS, i.e. ξ, τ and throughput $\eta = \lambda(1-\xi)$, over the wireless link in an analytical form.

7 Conclusion

We present some challenges and principles concerning cross-layer quality if service (QoS) for wireless multimedia delivery in adaptive networks. The QoS guaranteed multiuser scheduler at the medium access control (MAC) sublayer and is coupled with the adaptive modulation and coding (AMC) at the physical layer. We carried out the channel modeling induced by AMC in order to determine QoS in an analytic form through the wireless link. The parameters are packet loss rate, throughput and the average delay per packet.

Realistic integrated models for the delay and multimedia quality need to be developed. Also, the benefits in terms of multimedia quality of employing a cross-layer optimized framework for different multimedia applications with difficult delay sensitivities and loss tolerances still need to be quantified.

References:

[1] K.R.Rao, Z.S.Bojkovic, D.A.Milovanovic, *Introduction to multimedia communications: applications, middleware, networking*, Wiley 2005.

[2] K.R.Rao, Z.S.Bojkovic, D.A.Milovanovic, *Wireless multimedia communications*, CRC Press 2007.

[3] Z.Bojkovic, D.Milovanovic, H.264 video transmission over IEEE802.11 based wireless networks: QoS cross-layer optimization, *WSEAS Trans. Communications*, vol.5, issue 9, pp.1777-1782, Sept. 2006.

[4] IEEE802.11g 2003, *Part II Wireless LAN medium access control (MAC) and physical layer (PHY) specification band*, Supp. IEEE802.11, 2003.

[5] IEEE802.11e, *Wireless LAN MAC enhancement for quality of service (QoS)*, 802.11e Draft 80.0, 2004.

[6] B.A-Fernandez, Life after 3G mobile communications, *IEEE Commun. Mag.*, vol.39, no.8, Aug. 2001.

[7] A.Goldsmith, S.Wicker, Design challenges for energy-constrained ad hoc wireless networks, *IEEE Wireless Commun.*, vol.9, pp.8-27, Aug. 2002.

[8] S.Shakkottai, T.S.Rappaport, P.C.Karlson, Cross-layer design for wireless networks, *IEEE Commun. Magazine*, vol.41, no.10, pp.74-80, Oct. 2003.

[9] Q.Liu, S.Zhon, G.B.Giannaris, Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links, *IEEE Trans. Wireless Commun.*, vol.3, no.5, pp.1746-1755, Sept. 2004.

[10] H.Fattah, C.Leung, An overview of scheduling algorithms in wireless multimedia networks, *IEEE Wireless Commun.*, vol.9, no.5, pp.76-83, Oct. 2002.

[11] M.B.Pursley, J.M.Shea, Adaptive nonuniform phase-shift-key modulation for multimedia traffic in wireless networks, *IEEE Journ. on Select Areas in Commun.*, vol.18, no.8, pp.1394-1407, Aug. 2000.

[12] D.Wetteroth, *OSI Reference model for telecommunications*, McGraw Hill, 2001.

[13] Z.J.Hass, Design methodologies for adaptive and multimedia networks, *IEEE Commun. Magazine*, vol.39, no.11, pp.106-107, Nov. 2001.

[14] IEEE802.11e D8.0 Draft supp. to Part 11, *Wireless MAC and Physical Layer (PHY) specifications: MAC enhancements for QoS*, Nov. 2003.

[15] Q.Zhang, W.Zhu, Y-Q.Zhang, End-to-end QoS for video delivery over wireless Internet, *Proc. of the IEEE*, vol.93, no.1, pp.123-143, Jan. 2005.

[16] E.Malkamaki, H.Leib, Performance of truncated type-II hybrid ARQ schemes with noisy feedback over block fading channels, *IEEE Trans. Commun.*, vol.48, no.9, pp.1477-1487, Sept. 2000.

[17] G.L.Stuber, *Principles of mobile communication*, 2nd ed. Norwell, MA: Kluwer, 2001.

[18] M.S.Alonin, A.J.Goldsmith, Adaptive modulation over Nakagami fading channels, *Kluwer J. Wireless comun.*, vol.13, no.1-2, pp.119-143, May 2000.

[19] Q.Liu, S.Zhou, G.B.Giannakis, Cross-layer scheduling with prescribed QoS guarantees in

adaptive wireless networks, *IEEE Journal on SAC*, vol.23, no.5, pp.1056-1066, May 2005.

[20] D.Bersekas, R.Gallager, *Data networks*, 2nd ed., Upper Saddle River, NJ: Prentice-Hall, 1992.