QoS Architecture over Heterogeneous Wireless Access Networks

Zoran S. Bojkovic[1], Bojan M. Bakmaz[2]

Abstract – **This article discussed the process of establishing end-to-end quality of service QoS and focused on the architecture over heterogeneous wireless access networks. The cross-layer QoS architecture for video delivery over multimedia wireless channel is presented and analyzed. There are several important blocks like QoS interaction between coding and transmission modules, QoS mapping and adaptation, as well as source rate constraint. This architecture enables to perform QoS mapping between statistical QoS guaranties at the network level to a corresponding priority class with different quality requirements.**

Keywords – **Quality of service QoS, Wireless access networks, rate constraint, transmission module, cross-layer QoS.**

I. Introduction

Quality of service QoS is a topic that attains a lot of attention in the wireless world. On the other hand, the future networking environment will be strongly characterized mainly by the heterogeneity of networks, especially regarding the network access part, although having IP as the common denominator. The end-to-end part will first traverse one access network that may be a high-speed wired segment (e.g., digital subscriber line DSL), a wireless local area network LAN, a wireless wide area network WAN (e.g., Universal Mobile Telecommunication System UMTS), or even a satellite one [1]. These segments will be supported by an IP-based core network, which will at least supply end devices with Internet Protocol IP connectivity, and include a gateway to the Internet backbone [2]. In heterogeneous, overlopping networks, a handover to a more suitable access point offer more capabilities needed to enable additional services [3]. For example, when passing by a wireless hot spot, one can perform a handover to this access point for a short period of time to facilitate some demanding service, such as download of bulk data or video conferencing. In many cases the availability of resources at the potential access point is not known before handover is performed. For QoS this means that the resources at the new access point should be allocated before attaching to the new network. This is often called anticipated or planed handover. This kind of handover offers two advantages. First, it reduces handover latency, because most signaling to set up resources in the new path is carried out in advance. Second, it avoids unsuccessful handovers or unnecessary periods of QoS degradation because handovers should only be performed if the resources are actually available.

QoS signaling architecture integrates resource management with mobility and location management. Resource management has to take care of admission control and release of requested resources. To ensure QoS on the data path, there are several techniques such as differentiated services or integrated services. Differentiated services (DiffServ) are a recent approach defined by the Internet Engineering Task Force (IETF). Instead of manipulating per-flow state at each router in a network, QoS preferences or guarantees at the network edges [4,5]. This requires the marking of packets in a special field of the IP header. Resource manager approach fulfills the main requirements of future mobile networks. It is flexible regarding heterogeneous networks with different QoS capabilities and mobility models. Mobility management protocols like Mobile IP ensure that a mobile device is reachable by a home address, although the local IP address may change during handover.

Traffic generated by the different services will not only increase traffic loads on the networks, but will also require different QoS requirements (call loss rate, delay and jitter) different streams (e.g., video, voice, data). Delivering multiple QoS to different types of traffic while maintaining high utilization of the bandwidth is the objective of efficient traffic management, which encompasses technologies like call admission control, policing, scheduling, buffer management and congestion control.

The article is organized as follows. We present important problems in mapping QoS parameters. We address cross-layer QoS architecture issues over wireless channel. We then introduce and explain substream rate constraint in the priority transmission module.

II. Important Issues

With the development of third generation (3G) and fourth generation (4G) wireless standards, new broadband applications can be offered to mobile users [6]. In addition to delivering high bit rate applications, 3G and 4G systems are also expected to provide multiple QoS guarantees to different types of user applications. For example, the packet switched connection in the UMTS provides four different services differentiated by delay sensitivity conversational, streaming, interactive and background classes.

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

[1]Zoran S. Bojkovic is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia and Montenegro, E-mail: zsbojkovic@yahoo.com

[2]Bojan M. Bakmaz is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia and Montenegro, E-mail: bakmaz78@eunet.yu

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.

Wireless networks have time-varying and nonstationary links due to:

- fading effects coming from path loss, large scale fading and small scale fading,
- roaming between heterogeneous mobile networks (e.g., from wireless LAN to wireless WAN),
- the variation in mobile speed, average received power and surrounding environments.

Consequently, the quality of wireless link varies, which can be measured by the variation of the signal-to-noise ratio (SNR) or the bit error rate (BER). These variations result in time-varying available transmission bandwidth at the link layer. This is also called the channel service rate. Since the buffer size at the link layer is typically finite, the time-varying channel service rate can induce buffer overflow and therefore packet loss, due to the bit rate mismatch between the transmitting packet and the channel service rate. At the application layer, due to variation in arrival time of video packets, some packets may become useless during playback if its arrival time exceeds certain threshold [4].

To coordinate effective adaptation of QoS parameters at video application layer and QoS mechanism are required. Generally speaking, a good cross-layer QoS mapping and adaptation mechanism that offers a good compromise between the video quality requirement and the available transmission resource is a challenging task. Namely, at the priority transmission layer, QoS is expressed in terms of probability of buffer overflow and/or the probability of delay violation at the link layer. On the other hand, at the video application layer, QoS is measured objectively by the mean squared over (MSE) and/or the peak signal-to-noise ratio (PSNR).

III. Additional Remarks

There are some important problems in mapping QoS parameters. These problems can be summarize as follows:

- A QoS-based adaptation model, which shows how QoS parameters of both priority transmission systems and video applications should be adjusted based on time-varying wireless channel.
- A coordination mechanism between priority transmission system and video applications, which provides interaction between two layers.
- A resource allocation within the priority transmission system, which provides soft QoS guarantees based on time-varying wireless channel.

In an attempt to take into account the problem statement, we will use QoS mapping architecture that will address cross-layer QoS issues for video delivery over wireless networks. Details for each important building blocks include: (a) the derivation of the rate constraint of a priority of transmission system; (b) the development of a QoS mapping mechanism that optimally maps video classes to statistical QoS guaranties of a priority transmission system and (c) the QoS interaction procedure between video applications and the priority transmission system to provide the best tradeoff between the video application quality and the transmission capability under time-varying wireless channel. To address these problems, we need a QoS mapping architecture that address cross-layer QoS issues for video delivery over wireless networks.

IV. Cross-layer QoS Architecture over Wireless Channel

This architecture considers an end-to-end delivery system for a video source from the sender to the receiver, which includes source video encoding module, cross-layer QoS mapping and adaptation module, link layer packet transmission module, wireless channel which is time varying and nonstationary, adaptive wireless channel modeling module, and decoder output at the receiver. Cross-layer QoS architecture for video delivery over multimedia wireless channel is shown in Fig. 1.

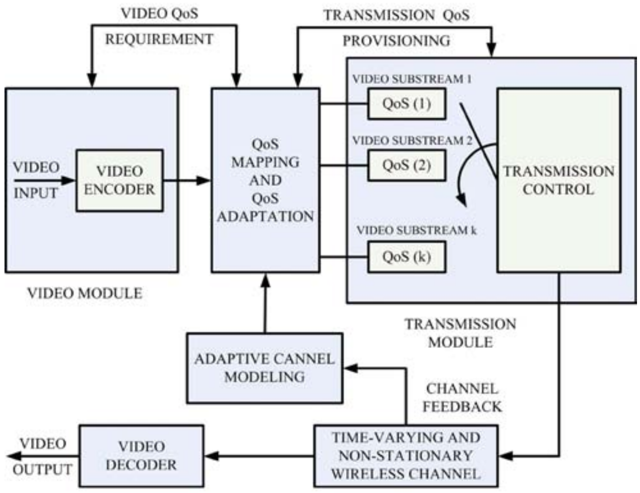

Fig. 1. Cross-layer QoS architecture for video delivery over multimedia wireless channel

The wireless channel is modeled at the link layer, since modeling of this layer is more amenable for analysis and simulations of the QoS provisioning system (e.g., delay bound or packet loss rate). The wireless link is expected to be fading, time-varying and nonstationary, which will provide a time-varying available transmission bandwidth for video service. The characteristics of the wireless channel can be modeled by a discrete-time Markov model, where each state represents the available transmission rate under current channel conditions. The channel modeling process is performed by the adaptive channel modeling module.

The QoS mapping and adaptation module is the key component to achieve cross-layer QoS mapping in the video delivery architecture. Unlike the adaptive channel modeling module and link-layer transmission module, the QoS mapping and adaptation module is application specific. It is designed to optimally match video application layer QoS and the underlying link-layer QoS. At the video application layer, each video packet is characterized based on its loss and delay properties, which contribute to the end-to-end video quality and service. Then, these video packets are classified and optimally mapped to the classes of link transmission module under the rate constraint. The video application layer QoS and link-layer QoS are allowed to interact with each other and adopt along with the wireless channel condition. The objective of these interaction and adaptation is to find a satisfactory QoS tradeoff so that each end user's video service can be supported with available transmission resources.

In the link-layer transmission control module, we employ a class-based buffering and scheduling mechanism to achieve differentiated services. A strict priority scheduling policy is employed to serve packets among the classes. That is, packets in a higher priority queue will always be sent first: packets in the lover priority queue will be sent only if there is no packet in the higher priority queues. Also, packets within the same class queue are served in a firs-in-first-out (FIFO) manner. Based on the class-based buffering and strict priority scheduling mechanism, we expect that each QoS priority class will have some sort of statistical QoS guarantees in terms of probability of packet loss and packet delay.

V. Substream Rate Constraint in the Priority Transmission Module

The rate constraint specifies the maximum input data rate to a particular buffer class that can be transmitted with certain statistical QoS guarantee. It will be used as the basis to allocate the channel bandwidth for data transmission. Since the wireless channel is expected to be fading, time-varying and nonstationary we will start with characterizing time-varying nonstationary wireless channel rate.

Suppose that the service rate for the time-varying wireless channel can be modeled by a first-order L-state Markov model, within each small time interval "g". Denote $X_C(u)$ as the state of the channel at time "u" and $X_C(u) \in \{1,..., L\}$. Each state $X_C(u)=i$ corresponds to a channel link condition, which can be characterized by an achievable channel transmission rate r_i. The channel transmission rate at state i can be computed as

$$r_i = R \cdot \log_2(1+\gamma_i) \tag{1}$$

in bits per second. Here, R is the transmission bandwidth in Hz and γ_i is the SNR value of the wireless channel condition at state i (physical layer parameter). For the L-state discrete-time Markov chain, denote p_{ij} as the state transition probability from state "i" (at time u-1) to state "j" (at time u) with a transition time interval of 1 time unit and 1<g. It means that $p_{ij}=P\{X_C(u)=j/\ X_C\ (u-1)=i\}$. Then, the L-state Markov chain can be completely characterized by the L×L state transition matrix which can be written in the form [8]

$$P_{transition} = \begin{pmatrix} p_{11} & \cdots & p_{1L} \\ \cdots & & \cdots \\ p_{L1} & \cdots & p_{LL} \end{pmatrix} \tag{2}$$

We can calculate the state probability for Markov model within the time interval g, which we denote in the form $[p_1, p_2, ..., p_L]$. Therefore, the expected link-layer transmission rate $r_{channel}$ during this time interval g is

$$r_{channel} = \sum_{i=1}^{L} r_i \cdot p_i \tag{3}$$

where r_i is the achievable link layer transmission rate given by Eq. (1). At the end of each time interval g, the state transition matrix in Eq. (2) will be updated by the adaptive channel modeling module to reflect the nonstationay nature of the wireless environment.

Fig. 2 shows a queueing system for time-varying source rate and channel service rate. The amount of data generated by the source from time "0" to "t" is a random variable of the form

$$A(t) = \int_0^t \alpha(u)du \tag{4}$$

where $\alpha(u)$ is the source data generated rate. The amount of data A(t) will be stored in the buffer of size B^{max} awaiting for transmission. On the other hand, the accumulated channel service from time "0" to "t" is of the form

$$S(t) = \int_0^t \alpha^{(c)}(u)du \tag{5}$$

where $\alpha^{(c)}(u)$ is the channel service rate at time "u". The time-varying channel service rate has been modeled by a L-state discrete time Markov chain, where $\alpha^{(c)}(u) \in \{r_1, r_2, ..., r_L\}$.

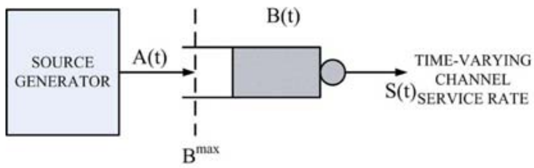

Fig. 2. A queueing system for time-varying source rate and channel service rate

The stochastic behavior of the accumulated channel service S(t) can be described by the concept of effective capacity, which can be written in the form [9]

$$\mu(\theta) = -\frac{\Lambda^{(c)}(\theta)}{\theta} \tag{6}$$

where $\Lambda^{(c)}(\theta)$ is the asymptotic log-moment generating function of S(t), defined as

$$\Lambda^{(c)}(\theta) = \lim_{t\to\infty} \frac{E[-\theta\ S(t)]}{t}$$

where θ is the QoS exponent of transmission service. It is called the QoS exponent to the effective capacity μ(θ). The parameter θ is related to the statistical QoS guarantee, e.g., packet loss probability of the time-varying channel. The statistical QoS guarantee in terms of packet loss probability can be derived as a function of θ in the form

$$P\{B(t) > B^{\max} | \ \theta\} \approx \xi \cdot e^{-\theta \cdot B^{\max}} \tag{7}$$

Here, B(t) represents the buffer occupancy at time t, $B^{\max}$ is the maximum buffer size, ξ is the probability that the buffer is not empty and $\xi \cdot e^{-\theta \cdot B^{\max}}$ is the approximate packet loss probability guarantee. As it can be seen, the effective channel capacity is related with the statistical QoS guarantee through the QoS exponent θ. Effective capacity μ(θ) in Eq. (6) imposes a limit for maximum amount of data that can be transmitted over time-varying channel with statistical QoS guarantee in (7). The statistical QoS guarantee required by the may mismatch with the statistical QoS guarantee provided by the channel [10]. In particular, if the source generating rate corresponding to the effective capacity of the QoS exponent "k" of the generated source, is grater than the effective channel capacity, i.e., μ(k)>μ(θ), part of the source rate would be expected to be cut-off or shaped. Here, θ represents QoS exponent of transmission service.

The maximum source rate that can be transmitted when there is a mismatch between the QoS exponents corresponding to the source and channel

$$\mu(k) = \begin{cases} \mu(\theta), \ 0 \le k \le \theta \\ \mu(\theta)\dfrac{\theta}{k} + \dfrac{k-\theta}{k} e_{S(t)}(k-\theta), \ k > \theta \end{cases} \tag{8}$$

where k>0 is the QoS exponent corresponding to the packet loss probability required by the source generation rate, while μ(k) is the source generation rate with QoS exponent k. At the same time

$$e_{S(t)}(k-\theta) = \frac{\Lambda^{(c)}(\theta - k)}{k-\theta}$$

can be viewed as the effective bandwidth with the QoS exponent k-θ.

The rate constraints for multiple priority classes are dependent on each other. Channel occupation by higher priority classes, i.e., rate constraints affect the rate constraints of lower priority classes, the lower opportunity of channel resource usage from lower priority ones.

VI. Concluding Remarks

The necessity to support high-capacity bursty traffic in extremely unpredictable wireless channels has posed a great challenge to all existing air link technologies based on time-division multiple access (TDMA) or code-division multiple access (CDMA) alike. Many research initiatives have been underway to investigate the issues in which type of multiple access technologies could be most suitable for wireless applications. A new wave of worldwide research is on the way to search for next generation multiple access technologies, which should effectively address all the constraints and problems existing in current TDMA and CDMA technologies, such as poor bandwidth efficiency, strictly interference-limited capacity, difficulties in performing rate-matching algorithms, and complexity in implementing fast adaptive equalizers. As we know, the access link remains the primary bottleneck in the network due to a number of existing economical and technological factors. As the available access bandwidth increases and more flexible operations management, administration and provisioning tools become available, new user applications requiring broadband connectivity become more attractive and economically viable. In some cases, these applications are bandwidth-intensive, such as video on demand. In other cases, broadband access allows less efficient use of the line in order to gain efficiencies in other portion of the network. Voice over IP is an example of the latter. A QoS mapping architecture that address cross-layer QoS issues for delivery over wireless networks include: rate constraint of a priority transmission system, the development of a QoS mapping mechanism that optimally maps video classes to statistical QoS guarantees of a priority transmission system and QoS interaction procedure between applications.

References

[1] K. R. Rao, Z. S. Bojković, D. A. Milovanović, *Introduction to Multimedia Communications: Applications, Middleware, Networking*, Wiley, 2005.

[2] K. R. Rao, Z. S. Bojković, D. A. Milovanović, *Multimedia Communication Systems: Techniques, Standards and Networks*, Prentice-Hall, 2002.

[3] Z. Bojković, B. Bakmaz, "Quality of Service and Security as Frameworks toward Next-Generation Wireless Networks", *Proceedings of the 6th WSEAS International Conference on Automation and Information ICAI'05*, Buenos Aires, Argentina, March 2005., http:/www.wseas.org

[4] K. R. Rao, Z. S. Bojković, *Packet Video Communications over ATM Networks*, Prentice-Hall, 2000.

[5] Z. Bojković, B. Bakmaz, "Quality of Service in DiffServ IP-based Networks", *WSEAS Transactions on Communications*, issue 1, vol. 4, pp. 30-35, January 2005.

[6] A. Jamalipour, T. Wada, T. Yamazato, "A Tutorial on Multiple Access Technologies for Beyond 3G Mobile Networks", *IEEE Commun. Mag.*, vol. 42, no. 2, pp. 110-117, Feb. 2005.

[7] D. Wu, R. Negi, "Effective Capacity: A Wireless Link Model for Support of Quality of Service", *IEEE Transactions on Wireless Commun.*, vol. 2, pp. 630-643, July 2003.

[8] H. Wang, N. Moayeri, "Finite-state Markov Channel-a Useful Model for Radio Communication Channels", *IEEE Trans. Vehicular Tech.*, vol. 44, pp. 163-171, Feb. 1995.

[9] C. S. Chang, J. A. Thomas, "Effective Bandwidth in High Speed Digital Networks", *IEEE Journal on Selected Areas in Commun.*, vol. 13, pp. 1091-1100, Aug. 1995.

[10] Z. Bojković, "Challenges in Multimedia Across Wireless Broadband", *Invited plenary lecture presented at the 6th WSEAS International Conference on Automation and Information ICAI'05*, Buenos Aires, Argentina, March 2005.