www.ietdl.org

Published in IET Communications Received on 27th June 2007 Revised on 1st September 2008 doi: 10.1049/iet-com:20070285

Quality of service for multimedia traffic using cross-layer design

T. Quazi H.J. Xu F. Takawira

School of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, Durban 4041, Republic of South Africa E-mail: quazit@ukzn.ac.za

Abstract: Quality-of-service (QoS) guarantees are critical for the transmission of multimedia traffic over mobile wireless networks. Currently, wireless networks provide QoS guarantees using the legacy layered protocol architecture where each layer provides a separate, independent solution, with its own optimised adaptation and protection mechanisms. Cross-layer design has been proposed as a methodology to extend that paradigm in wireless links where there is interdependence between the layers and hence opportunity for information sharing. Recently, cross-layer adaptation mechanisms have been proposed which attempt to solve the QoS provisioning problem. However, most of these mechanisms only use the lower (physical and data link) layers and the possibility of using higher protocol layers remains unexplored. As a result, restrictions are placed on the system which introduces functional and efficiency limitations. Here, one such limitation is highlighted, namely the inability to insert more than one class of traffic in a physical layer frame. A physical and application layer cross-layer adaptation mechanism is then proposed, which overcomes this limitation. The performance results of the scheme show that the cross-layer mechanism can be efficiently applied for the purpose of providing QoS guarantees for multimedia traffic.

1 Introduction

Providing guarantees in the quality of service (QoS) has become essential for the operation of today's multimedia wireless networks. However, this poses quite a challenge due to the variable nature of wireless links and the diverse QoS requirements of different applications including voice, video and data. In the current paradigm of the layered protocol stack each layer of the communications stack works independently to provide a solution to these challenges. The benefits of dynamic adaptation to system and channel conditions have been accepted, but the true potential of optimised adaptation is lost if the layers operate in an individual fashion, ignoring possible interdependencies between them. Cross-layer design (CLD) provides mechanisms in which these interdependencies are exploited to achieve challenging objectives by providing QoS guarantees for multimedia traffic over wireless links.

One of the earliest contributions in the research area of CLD was to increase spectral efficiency by adapting certain parameters of the transmitted signal to match the wireless

channel conditions. Physical and data link layers were combined to allow the adaptive variation of transmission power, symbol rate, coding rate, constellation size or any combination of these parameters. The work presented in [1, 2] showed that efficient bandwidth utilisation for a prescribed bit error-rate (BER) performance at the physical layer can be accomplished using adaptive modulation and coding (AMC) algorithms. Recently, Liu et al. [3, 4] used an AMC scheme to provide QoS support in wireless networks. The objective of the AMC module in such a system is to maximise the data rate by adjusting the transmission modes, that is, varying the modulation and coding schemes to the channel variations while maintaining a prescribed packet error rate (PER) P_0 . The limitation, however, of such a method in the time-division multiplexing/ time-division multiple-access (TDM/TDMA) system is that once a target PER has been chosen for a frame (depending on the requirement of the type of traffic being transmitted), then only that type of traffic can be inserted into the frame. If P_0 is set for voice traffic then, although the PER target for voice will be met, the PER will not be low enough for video or data, if such traffic were to be inserted into the

frame. However, if the target has been set for data traffic, the AMC system will be inefficient for voice and video traffic. The efficiency of the resource allocation module of the system would be noticeably higher if it was able to insert more than one type of traffic in a frame and it had a finer control over the allocation of resources for the different types of traffic being transmitted. In order to achieve this, we propose a cross-layer system where the AMC module is combined with an adaptive coding mechanism at the application layer to guarantee QoS performance in terms of PER for voice, video and data traffic. The physical layer guarantees a fixed error target for a frame and the application layer coder applies additional error protection to meet the QoS packet error target depending on the type of traffic being inserted into the frame.

In related work, QoS provisioning using CLD has been proposed by combining the physical and data link layers of the protocol stack $[3, 4]$. In these schemes, multiuser scheduling at the medium access control (MAC) layer is employed in conjunction with the AMC module at the physical layer. In their system, each frame only contains a single class of traffic. In the concluding section of their paper [4], the authors note this limitation and suggest that multiple connections allowing heterogeneous traffic may lead to better performance. The work in [5] has a similar objective to the one proposed in this paper, namely to allow the transmission of different classes of traffic in a single unit of transmission; however, that work is a CDMA setting. Using CDMA, the transmitter has more channels and subsequently a higher degree of freedom during resource allocation, as compared with the schemes in [3, 4] and the one proposed in this paper. The 'effective capacity' model [6] has been employed in cross-layer resource allocation mechanisms for achieving statistical QoS guarantees in the transmission of heterogeneous traffic. This model was used in [5] and more recently in the work in $[7]$. The work in $[8, 9]$ studies the scheduling of heterogeneous traffic over wireless links. The objective of both the schemes is to determine the order in which the packets in the queue should be transmitted, thus allowing for only a single packet to be transmitted in a frame. Furthermore, the high computational cost of both the schemes is a disadvantage. The cross-layer mechanism in [10] exploits coupling between the physical and MAC to achieve multiuser diversity. However, as the scheme does not consider the type of traffic being transmitted, the opportunity for diversity across multiple classes of traffic with different QoS requirements is missed. Hence the aim of the proposed scheme is to exploit the opportunity for diversity across multiple classes of traffic.

The desired outcome of QoS provisioning schemes for multimedia traffic varies from one scheme to the next and is incorporated into the QoS metric emphasised by a particular scheme. A good review of this discussion is provided in [11] and includes example solutions to the problem. The work in $[5, 7-9]$ considers packet delay

violation as the main metric, and the work in $[12]$ uses handoff dropping probability and average allocated bandwidth. Other QoS metrics include packet loss rate through buffer overflow [7], throughput and fairness [13, 14]. These are MAC or higher-layer-based adaptive crosslayer schemes which assume that the physical layer is either error free or adapts to the channel to guarantee a single error target and is independent of the upper layers in its operation. However, the objective of the proposed scheme is to use both the physical and the application layer to explicitly solve the QoS provisioning problem with target PERs for various classes of traffic as the metric, and to the best of the authors' knowledge this has not been attempted by any previously proposed scheme.

The concept of combining the physical and the application layer for error protection has been applied in the area of video transmission over wireless links [15, 16]. The physical layer parameters are adapted to the different channel conditions, whereas an application layer forward error correction (FEC) provides an additional error control strategy. The novelty of the use of an application layer FEC for video transmission is that it can offer the flexibility of unequal or selective error protection. A multiple description source encoder splits a video stream into multiple bit streams or descriptions, and each sub-stream has a different priority. The priority is based on the level of fidelity with which each sub-stream describes the original stream. Thus, a higher level of FEC protection is then applied to a high priority bit stream than one with a lower priority.

The system proposed in this paper applies the unequal error protection concept mentioned above to the three classes of traffic to be transmitted over the wireless link. Data traffic has the highest priority in terms of QoS error protection and receives the highest level of protection while voice traffic, having lowest priority, receives the lowest level of protection. The remainder of the paper is organised as follows. The system model is described in Section 2 followed by Section 3 which describes the design of system algorithm. Performance evaluation is detailed in Section 4 and the paper is concluded with future work in Section 5.

2 System model

2.1 System description

The system under consideration has multiple users or nodes connected to a central unit which could be a base station or a backbone gateway for a wireless local area network (WLAN) or ad hoc network. Each user/node is connected to the central unit over the wireless channel using TDM/ TDMA. Only the downlink is considered in this system, with the assumption that the results for the uplink would not be dissimilar.

The wireless link between two nodes is shown in Fig. 1.

Figure 1 Wireless link between communicating nodes

This is a modification of the model used in other AMC schemes, as the ones presented in $[3, 4]$. The main differences between these schemes and the one proposed in this paper are as follows.

(a) The schemes in $\left[3, 4\right]$ only accommodate one class of traffic, whereas the proposed scheme accommodates multiple classes of traffic.

(b) The schemes in $[3, 4]$ use only the physical and data link layer, whereas the proposed CLD adds a third layer, in the form of the application layer, to cater for multiple classes of traffic.

In the proposed scheme, the AMC module at the physical and data link layer provides a certain level of error protection which is the same as in $[3, 4]$. The application layer module, using an adaptive application layer coder, provides additional error protecting to different classes of traffic by varying its coding parameters depending on the channel conditions and the class of traffic being transmitted. Thus the three layers combine to guarantee QoS (in terms of error performance) to various classes of traffic. A buffer (queue) is implemented at the central node for each user and operates in a first-in-first-out mode. The AMC module follows and precedes the buffer at the transmitter and receiver, respectively. The AMC controller is implemented at the central node (transmitter) and an AMC channel estimator at each user (receiver).

The traffic (voice/video/data) streams are encoded by the coder at the application layer to form code blocks. These code blocks are passed to the data link layer as packets which are then stored in the buffer waiting to be served by the AMC module of the physical layer. After the AMC module processing, the symbols are packaged into a frame which is transmitted over the wireless channel. The AMC module allows each user to have multiple transmission modes, each of which represents a pair of a specific modulation format and an FEC code. Such transmission modes are

implemented in systems operating the HIPERLAN/2 and IEEE 802.11a protocol standards. Channel estimation is done at the receiver and the results are fed back to the transmitter. This information is fed to both the AMC controller and the adaptive application layer coder, and they adjust their parameters accordingly. The channel state information (CSI) is used by the AMC controller to determine a modulation-coding pair (or mode) and by the application layer coder to determine a coding rate.

2.2 Packet structure at various layers

• The application layer: The voice, video or data packets are encoded into Reed-Solomon (RS) code blocks. The encoder takes k data symbols of L bits each and adds parity symbols to make an n symbol code word/block.

• The data link layer: Each packet contains a fixed number of bits (N_b) and this includes the packet header, the payload from the application layer coder and cyclic redundancy check code bits. Each packet is mapped to a symbol block containing $N_{\rm b}/R_n$ symbols where R_n is the rate for mode *n*.

• The physical layer: Each frame contains a fixed number of symbols (N_s) . The symbol rate is fixed and so the frame duration (T_f) remains constant. With TDM, the frame is divided into $N_c + N_d$ time slots, where N_c is the number of time slots with control information and pilots and N_d the number of time slots that convey the data being transmitted. Each time slot contains a fixed number of $N_{\rm b}/$ R_1 symbols [5]. Given a transmission mode *n*, each time slot will then contain R_n/R_1 packets. For example, a time slot will contain $R_1/R_1 = 1$ packet with mode $n = 1, R_2/$ $R_1 = 2$ packets with mode $n = 2$ and so on. The data time slots are scheduled to different users using TDMA dynamically. The aim of this work is to allow various classes of traffic to be transmitted to one user or many users in the same frame.

The data structures at the interacting layers are shown in Fig. 2.

2.3 Operating assumptions

The channel model assumed for this system is a block fading model, where the channel between users is frequency flat and invariant for the duration of one frame, but varies from frame to frame. Such a model is suitable for slowly varying wireless channels [4]. The adaptive schemes at the physical and application layer are then adjusted on a frame-by-frame basis. Perfect CSI is available at the receiver which is derived from training-based channel estimation. The feedback channel is assumed to be error free and instantaneous, and thus the determined signal-to-noise ratio (SNR) value is fed back to the transmitter without error and latency.

Figure 2 Data structures at the interacting layers

3 System design

3.1 Physical layer

The objective of the adaptive algorithm at the physical layer is to maximise the data rate, while maintaining the required BER performance at a target value. The transmission modes (no coding is used currently in the system) at the physical layer are arranged such that the rate increases as the mode index n increases. This is shown in Table 1.

Let N denote the total number of transmission modes available. Assuming constant power, the SNR range is partitioned into $N+1$ non-overlapping consecutive fading intervals with the boundary points denoted as $\{\gamma_n\}_{n=0}^{n=N+1}$.

Mode *n* and thus the constellation size M_n is chosen when $\gamma \in [\gamma_n, \gamma_{n+1})$ for $n = 1, 2, \ldots, N$.

To avoid deep fades, no payload bits are sent when the SNR is below a certain threshold γ_1). Table 1 shows the boundary points for the non-coded modulation scheme for a target BER value of 10⁻³ (appropriate for voice traffic) using M_n quadrature amplitude modulation (QAM) over an additive

white gaussian noise (AWGN) channel. The expressions to determine the boundary points are derived in [1]

$$
\gamma_1 = [\text{erfc}^{-1}(2BER_0)]^2
$$
\n
$$
\gamma_n = \frac{2}{3} K_0 (2^n - 1), \quad n = 0, 1, 2, 3, ..., N
$$
\n
$$
\gamma_{N+1} = +\infty
$$
\n(2)

where $erfc^{-1}$ (.) denotes the inverse complementary error function and $K_0 = -\ln(5BER_0)$. All of the thresholds or boundary points (except for γ_1) are calculated using (2), which is the inverse of the following BER approximation [1, eq. (28)]

$$
BER(M_n, \gamma) \cong 0.2 \exp\left(-\frac{3\gamma}{2(M_n - 1)}\right) \tag{3}
$$

The authors in [1] motivate the use of the approximation by the fact that both (3) and its inverse (2) are very simple functions which lead to closed-form analytical expressions and that further insights are unattainable with more complicated BER expressions. However, (3) is an upperbound for the BER only for $M_n \geq 4$ and thus γ_1 is chosen according to the exact BER performance of binary phase shift keying (BPSK). The inverse of (1) used to determine the BER performance of BPSK is given by

$$
BER = Q(\sqrt{2\gamma})
$$
 (4)

When the switching thresholds (modes) are chosen according to the boundaries, the physical layer will ensure a BER performance that is below the target BER.

3.2 Application layer

The BER performance target chosen at the physical layer was 10⁻³. This error target, although sufficient for voice applications, is not low enough for more demanding applications such as video and data. In order to accommodate these types of traffic in the system, additional error protection is required. The variable error protection is achieved in this CLD system using RS codes at the application layer.

The metric for error-rate performance in most applications, such as video, is the frame error rate. The analytical expression for frame error rate for general RS codes is as follows.

Table 1 Boundary points for adaptive non-coded modulation

Mode (n)							
M_n	BPSK	4PSK	8PSK	16QAM	32QAM	64QAM	128QAM
SNR, dB	6.79	10.25	13.93	17.24	20.39	23.47	26.52
R_n						ь	

The frame error probability (P_f) can be expressed as

$$
P_{\text{fe}} = \sum_{k=t+1}^{N} {N \choose k} (1 - P_{\text{s}})^{N-k} P_{\text{s}}^{k}
$$
 (5)

where N is the number of code symbols per frame, t the number of code symbols the code can correct, P_s the symbol error probability and k the number of data symbols.

Then, P_s can be expressed as

$$
P_{\rm s} = 1 - (1 - P_{\rm b})^L \tag{6}
$$

where $P_{\rm b}$ is the bit error probability and L the number of bits per code symbol.

The frame error rate was evaluated for $RS(127,k)$ and RS(63,k) codes with error-correcting capabilities of $t = 1$, 2, 3 and 4 over an AWGN channel within the range for each mode of the AMC module. The frame error-rate performance can be seen in Figs. 3 and 4. The *k* value, that is, the level of protection for each class of traffic, depends on the P_{fe} target and the SNR value fed back from the receiver. The boundaries in the SNR range for each k value were determined by setting P_{fe} targets for each class of traffic. The target P_{fe} for voice, video and data was 10^{-1} , 10^{-2} and 10^{-4} , respectively. These target values are similar to those used in the performance evaluation of the scheme in [5], but, in this case, they are used to design the SNR thresholds for the level of protection at the application layer for each class of traffic. The k values and the respective SNR threshold values guaranteeing the P_{fe} target for each traffic class were determined using the graphical data in Figs. 3 and 4 and in expressions (5) and (6). The base

station has separate queues for each active traffic session with the users with which it is communicating. Before the transmission of a frame, once it obtains the CSI information (i.e. the channel gain) from the receiver, the AMC module uses the value to select the appropriate mode, ensuring that BER rate is no higher than 10^{-3} . The channel gain information is also passed to the application layer coder and it uses the value to determine k in an RS (n, k) code for each class of traffic being transmitted. Thus, a different k value will be selected for voice, video and data traffic, and this value will be used by the RS encoder. The output code blocks, which could contain any of the three classes of traffic, are passed to the AMC module as packets. The AMC module then produces symbol blocks using the selected mode and they are inserted into the time slots of the frame to be transmitted. The process is reversed at the receiver. The AMC module processes the received frame and produces packets which could be either a voice, video or data packet. At the application layer, the packet header is examined to find out the k value that was used in the encoding process and this is then passed to the RS decoder. The decoder then decodes the packet and passes it to the appropriate application.

The two interacting layers (application and physical) are combined to guarantee that the target error rate for each class of traffic being transmitted is achieved. However, they operate quite independent of each other since the physical layer makes no distinction between the packets that it receives from the application layer. The packets are inserted into the frame which is modulated by a single AMC mode. Thereafter, the application layer encoder, given the channel gain, selects the k value depending on the class of traffic being transmitted and uses this to encode the packets.

Figure 3 Frame error-rate performance for RS(127,k)

www.ietdl.org

Figure 4 Frame error-rate performance for RS(63,k)

Owing to this independence, a frame can contain a mix of voice, video and data traffic. This is the ideal form of CLD in which inter-layer coupling is used to achieve a design objective without compromising the structure of the layers.

4 Performance analysis

The first task in the performance analysis was to determine the theoretical performance of the system using the threshold boundary points and analytical expressions for both the AMC module at the physical layer and the

variable error protection system at the application layer. The number of bits for each RS code symbol (L) used was 8 bits. Shortened $(255, k)$ RS codes were employed with a code word length (N) of 127 for voice traffic and 63 for video and data traffic. The performance of the three classes of traffic was determined separately using (3) – (6) and can be seen in Figs. 5–7 for voice, video and data, respectively.

The system was then simulated at certain SNR points over an AWGN channel using the communications toolbox in MATLAB. The message or packet length was N^*L and

Figure 5 Frame error-rate performance for voice traffic

Figure 6 Frame error-rate performance for video traffic

was modulated at the physical layer by the schemes shown in Table 1. Initially, the results of the simulations showed noticeable difference from the values expected from the theoretical evaluations for modes with $M_n \geq 4$. This was due to the analytical expression (3) which is only an approximation for modulations with $M_n \geq 4$. Owing to the sensitivity of the BER performance on the frame error-rate results, this difference resulted in the divergence from the theoretical result at the application layer. A compensation factor was introduced that shifted the curves at the physical layer to match the result of the analytical approximation with that derived from the simulation. The boundary points were adjusted accordingly. This resulted in a much

tighter fit of the simulation and theoretical results of the overall system, as can be seen in Figs. 5–7. The objective of these experiments was to prove that the design decisions made in the formulation of the scheme were valid for the purposes of meeting the targets of the CLD. This was critical as some approximations were made in the analytical expressions used, and in the process of determining the SNR thresholds. Thus the simulations were conducted to verify the analytical results.

The graphs in Figs. 5–7 show that the performance of the adaptive application and physical layer CLD system meets the target for all three classes of traffic across the channel

Figure 7 Frame error-rate performance for data traffic

SNR range. The frame error rate goes no higher than 10^{-1} , 10^{-2} and 10^{-4} for voice, video and data, respectively. Thus, having met this objective, the system will allow for the insertion of any of the three classes of traffic, in any combination, in a single physical layer frame without the danger of not meeting the frame error-rate QoS targets, thus fulfilling the purpose of the proposed CLD system. The overall system benefits include the increased efficiency of the resource allocation mechanism which will now have a fine-grain control of the available system resources (time slots) when scheduling the three classes of traffic.

Figs. 5–7 also show the effect of removing the adaptive application layer coder from the system which results in the P_{fe} for each class of traffic being much higher than the set target. This shows the ineffectiveness of only using the adaptive physical layer mechanism in meeting the error-rate QoS targets.

5 Conclusion

In this paper, we have emphasised the need for CLD in providing QoS guarantees in multimedia wireless networks. Current CLD resource allocation mechanisms use a static form of allocation, where a physical layer frame is designed to hold only one class of traffic. The proposed system facilitates a dynamic resource allocation where more than one or all three traffic classes can be transmitted simultaneously in one frame. This not only has functional uses in providing QoS in a multimedia wireless system, but also makes the resource allocation process more efficient. The predicted result of this will be increased data link layer throughput and lower average delays. However, proving this will be the aim of future work. We will try and improve the data throughput by optimising the length of the RS codes in the adaptive application layer coding process. In addition, we will investigate the use of alternative, nonideal channel estimation and feedback mechanisms for the CLD system.

6 References

[1] ALOUINI M.S., GOLDSMITH A.J.: 'Adaptive modulation over Nakagami fading channels', J. Wirel. Commun., 2000, 13, $(1-2)$, pp. $119-143$

[2] HOLE K.J., HOLM H., OIEN G.E.: 'Adaptive multidimensional coded modulation over flat fading channels', IEEE J. Sel. Areas Commun., 2000, 18, (7), pp. 1153– 1158

[3] LIU Q., ZHOU S., GIANNAKIS G.B.: 'Cross-layer scheduling with prescribed QoS guarantees in adaptive wireless networks', IEEE J. Sel. Areas Commun., 2005, 23, (5), pp. 1056– 1065

[4] LIU Q., WANG X., GIANNAKIS G.B.: 'A cross-layer scheduling algorithm with QoS support in wireless networks', IEEE Trans. Veh. Technol., 2006, 55, (3), pp. 839-847

[5] CI S., GUIZANI M., BENBRAHIM G.: 'A dynamic resource allocation scheme for delay-constrained multimedia services in CDMA 1xEV-DV forward link', IEEE J. Sel. Areas Commun., 2006 (Special Issue on Next Generation CDMA Technologies) 24 , (1), pp. $46-53$

[6] WU D., NEGI R.: 'Effective capacity: a wireless link model for support of quality of service', IEEE Trans. Wirel. Commun., 2003, 2, (4), pp. 630-643

[7] TANG J., ZHANG X.: 'Cross-layer-model based adaptive resource allocation for statistical QoS guarantees in mobile wireless networks', IEEE Trans. Wirel. Commun., 2007, 6, (12), pp. 4504– 4512

[8] HAN Z., LIU X., WANG J., LIU K.J.R.: 'Delay sensitive scheduling schemes for heterogeneous QoS over wireless networks', IEEE Trans. Wirel. Commun., 2007, 6, (2), pp. 423– 428

[9] JOHNSSON K.B., COX D.C.: 'An adaptive cross-layer scheduler for improved QoS support of multiclass data services on wireless systems', IEEE J. Sel. Areas Commun., 2005, 23, (2), pp. 334– 343

[10] QIUYAN X., HAMDI M.: 'Cross layer design for IEEE 802.11 WLANs: joint rate control and packet scheduling'. Proc. IEEE Conf. Local Computer Networks, 2005, 30th Anniversary, 15-17 November 2005, pp. 624-631

[11] VAN DER SCHAAR M., SAI SHANKAR N.: 'Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms', IEEE Wirel. Commun., 2005, 12, (4), pp. 50– 58

[12] YU F., WONG V., LEUNG V.: 'Efficient QoS provisioning for adaptive multimedia in mobile communication networks by reinforcement learning'. Proc. 1st Int. Conf. Broadband Networks, 2004 (BroadNets 2004), 25– 29 October 2004, pp. 579– 588

[13] CHOI Y., BAHK S.: 'WAF: wireless-adaptive fair scheduling for multimedia stream in time division multiplexed packet cellular systems'. Proc. 8th IEEE Int. Symp. Computers and Communication, 2003 (ISCC 2003), vol. 2, pp. 1085– 1090

[14] TANG J., ZHANG X.: 'Cross-layer design of dynamic resource allocation with diverse QoS guarantees for MIMO-OFDM wireless networks'. Proc. 6th IEEE Int. Symp. World of Wireless Mobile and Multimedia Networks, 13-16 June 2005, pp. 205-212

[15] CHAN Y., COSMAN P., MILSTEIN L.B.: 'A cross-layer diversity technique for multicarrier OFDM multimedia networks', IEEE Trans. Image Process, 2006, 15, (4), pp. 833–845

[16] KRISHNAMACHARI S., VANDERSCHAAR M., CHOI S., XU X.: 'Video streaming over wireless LANs: a cross-layer approach'. Proc. IEEE Packet Video '2003 (PV'03), Nantes, France, April 2003

Copyright of IET Communications is the property of Institution of Engineering & Technology and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.