A Cross-Layer Optimization Mechanism for Multimedia Traffic over IEEE 802.16 Networks

Dionysia-Katerina Triantafyllopoulou[+], Nikos Passas[+] and Alexandros Kaloxylos[*]

[+]Department of Informatics & Telecommunications
University of Athens, Greece

[*]Department of Telecommunication Science and Technology
University of Peloponnese, Greece

***Abstract* – In this paper, a cross-layer optimization mechanism for IEEE 802.16 metropolitan area networks is introduced. The proposed scheme utilizes information provided by the Physical and Medium Access Control layers, such as the signal quality, the packet loss rate and the mean delay, in order to control parameters at the Physical and Application layers and improve system's performance. The main idea is to combine the adaptive modulation capability of the Physical layer and the multi-rate data-encoding feature of modern multimedia applications, and attain improved end-user QoS. Simulations show that the proposed mechanism can assist legacy 802.16 systems accommodate to frequent channel and traffic changes, leading to considerably reduced packet loss rates, especially under heavy traffic conditions.**

***Index Terms* – adaptive modulation, cross-layer Optimization, IEEE 802.16, multi-rate applications.**

I. INTRODUCTION

During the last decade we are witnessing a growing interest in the area of wireless communication systems and networks. Commercial demand for advanced applications and services for people on the move keeps increasing and so do the supporting wireless technologies. One of the main issues that have emerged is the need for high-speed communication systems able to handle a wide range of services with different traffic characteristics and Quality of Service (QoS) requirements. In the field of metropolitan area wireless broadband networks the IEEE 802.16 standard [1] is considered today as the most promising technology.

One of the main features of modern communication systems is the parameterized operation at different layers of the protocol stack. This feature aims at providing them with the capability of adapting to the rapidly changing traffic, channel and system conditions and finally enjoying improved performance. An interesting research problem is the combination of individual adaptation mechanisms into a cross-layer design that can maximize their effectiveness.

In this paper, we introduce a cross-layer optimization mechanism for IEEE 802.16 networks, that interacts with the Physical (PHY), Medium Access Control (MAC), and Application layers, in order to take advantage of both the adaptive modulation capability of the PHY layer and the multi-rate feature of modern multimedia standards at the Application layer. Our aim is to provide an effective combination of the aforementioned adaptation capabilities and achieve efficient exploitation of the available optimization features. The proposed mechanism is also applicable to systems enhanced with the recent amendment IEEE 802.16e [2], which are more demanding for adaptive techniques due to frequent channel quality changes as a result of the inherent nodes' mobility.

The rest of the paper is organized as follows. Section II provides an overview of the architectural framework used for the design of the proposed cross-layer optimizer. Section III contains details on the control information exchanged between the various network elements. Section IV describes in detail the proposed cross-layer optimizer. Section V presents the simulation model used for the evaluation of the optimizer's performance and the derived simulation results. Finally, Section VI contains conclusions and plans for future work.

II. ARCHITECTURAL FRAMEWORK

In the recent bibliography ([3]-[6]) most cross-layer optimization proposals refer to mechanisms between the PHY and MAC layers. For example, [3] proposes a cross-layer protocol for downlink channel quality information feedback and uplink sounding as well as a channel aware scheduling algorithm. The authors in [4] propose a scheduling algorithm at the MAC layer where each connection employs adaptive modulation and coding at the PHY layer. In [5], a dynamic programming solution for computing optimal power, channel coding, and source coding policies for delay constrained traffic over wireless channels is proposed. Reference [6] introduces a cross-layer design, which optimizes the target packet error rate of adaptive modulation and coding schemes at the PHY, in order to minimize the packet loss rate and maximize the average throughput, when combined with a finite-length queue at the data link layer.

The cross-layer optimization mechanism for multimedia traffic over IEEE 802.16 networks described in this paper is based on the architectural framework introduced in [7], which consists of N layers and a cross-layer optimizer (Fig. 1). According to that, the optimization process is performed in three steps:

i. **Layer abstraction:** Computes an abstraction of layer-specific parameters that are processed by the optimizer. This process aims at reducing the overall data processing and communication overhead while maintaining consistency.
ii. **Optimization:** Determines the values of layer parameters that optimize a specific objective function.
iii. **Layer reconfiguration:** Distributes the optimal values of the abstracted parameters to the corresponding layers that in turn translate them back into layer-specific information.

The rate at which the above steps are repeated depends on the variance of the channel conditions and the applications' requirements.

In general, the different parameters can be divided in four main types:

a. Directly Tunable (DT) parameters that can be directly set as a result of the optimization process.
b. Indirectly Tunable (IT) parameters that may be modified as a result of the setting of the Directly Tunable parameters.
c. Descriptive (D) parameters that can only be read but not tuned by the optimizer.
d. Abstracted (A) parameters that are abstractions of the previous types of parameters.

More details about this framework can be found in [7]. The cross-layer optimization mechanism proposed in this paper works with three layers, namely the PHY, MAC and Application.

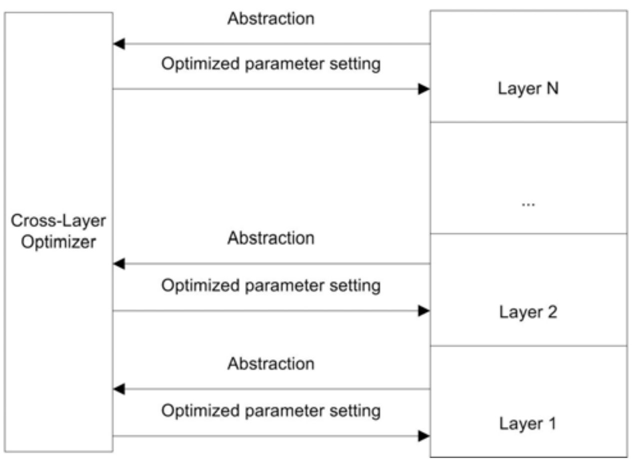


Fig. 1. General Cross-Layer Architecture

III. INFORMATION FLOW

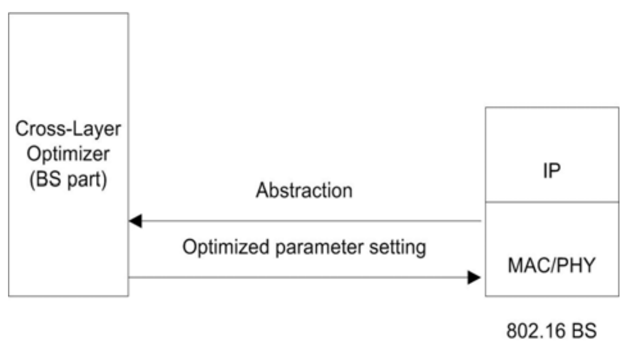
The proposed cross-layer optimizer for IEEE 802.16 networks is split into two parts, namely the *BS part* and the *SS part*, residing at the Base Station (BS) and the Subscriber Stations (SSs) of an IEEE 802.16 network respectively (Fig. 2). The BS part accepts an abstraction of layer-specific information, regarding the channel conditions and QoS parameters of active connections, provided by the BS PHY and MAC layers (Step 1: Layer Abstraction). According to this information, a specific decision algorithm determines the most suitable modulation and/or traffic rate of each SS, separately for each direction (uplink or downlink) (Step 2: Optimization). Finally, the BS part informs the corresponding layers of the required modifications (Step 3: Layer Reconfiguration) (Fig. 2a). If the decision of the BS part involves traffic rate changes, it communicates with the SS part through the SS MAC layer, which instructs the SS Application layer accordingly (Fig. 2b). The SS part may either accept the BS part's suggestions or refine them, based on its better knowledge of the status of active connections. In this paper we consider the SS part as a passive module that only instructs the SS Application layer based on the BS part's suggestions, but we plan to extend its functionality in the future. More details on the specific operation of the BS and SS parts are provided in the next section.

A. Channel Measurements

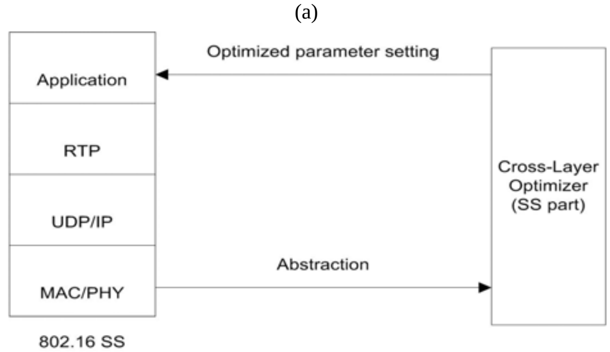
For downlink transmissions, the BS needs information regarding the quality of the signal received by the SSs. One solution would be to let the BS assume that the quality of the downlink signal is similar to the quality of the signal transmitted by each SS on the uplink. However, this is not a valid assumption at all times. A more safe option is for the SS to send channel measurements to the BS using standard IEEE 802.16 signalling either periodically, through the Channel Quality Information Channel (CQICH), or on demand, through the Channel measurement Report Request and Response (REP-REQ, REP-RSP) messages ([1], [2]).

The Channel Quality Information Channel (CQICH) is allocated by the BS as soon as the BS and the SS are informed of each other's modulation and coding capabilities and is used for periodic CINR (Carrier to Interference plus Noise Ratio) reports from the SS to the BS. If the SS decides that the last effective CINR report is no longer appropriate for the duration remaining until the next periodic CQI transmission, it may send an unsolicited REP-RSP message.

Alternatively, the BS may send a REP-REQ message in order to request either the RSSI (Received Signal Strength Indicator) and CINR channel measurements by the SS, or the results of previously scheduled measurements. The SS responds with the REP-RSP message that contains the corresponding values. Additionally, if regulation mandates so, the SS may send REP-RSP messages in an unsolicited manner in order to inform the BS on the detection of specific signals, or when other interference is detected exceeding a threshold value. The BS shall provide transmission opportunities for sending an unsolicited REP-RSP frequently enough to meet regulatory requirements. More details can be found in [1] and [2].



(a)



(b)

Fig. 2. BS (a) and SS (b) part functionality

B. *Traffic Rate Adjustments*

The operation of the proposed optimization mechanism described in the next section, relies on the information sent by the BS part to the SS part, regarding the recommended traffic rate adjustments. This recommendation is in the form of a target overall mean traffic rate either for the downlink or the uplink. None of the existing management messages of the standard 802.16 can serve for that purpose, thus we have to define a new one using one of the reserved message types. The message is referred to as RATE-ADJ-REQ and has the following syntax:

RATE-ADJ-REQ message

Syntax	Size
RATE-ADJ-REQ_Message_Format() {	
Management Message Type = 67	8 bits
Direction	1 bit
Total Rate Recommended	32 bits
}	

Message type 67 is determined as “reserved” in both [1] and [2], meaning that it is left unused for future purposes. The “Direction” parameter declares uplink or downlink direction. The “Total Rate Recommended” parameter contains the recommendation of the BS part in kb/s. Considering that this refers to the total transmission rate for a SS, the size of 32 bits allows rates up to approximately 4.3 Gb/s per direction.

IV. THE PROPOSED MECHANISM

The main functionality of the cross-layer optimizer resides at the BS part. Its mission is to collect information on the traffic, channel conditions and QoS parameters of active connections (both uplink and downlink), and instruct proper adjustments of the *modulation* and/or the (receiving or transmitting) *traffic rate* of each SS, aiming at optimizing QoS and system’s throughput. The counterpart of the cross-layer optimizer rests at the SS part and its mission is to accept requests for traffic rate changes from the BS and communicate with the Application layer in order to instruct proper application data rate adjustments (for multimedia traffic this usually means data encoding changes). Figures Fig. 3 and Fig. 4 show the signaling messages exchanged in the uplink and downlink case respectively. As can be observed, and explained later in this section, the downlink case requires two more messages ((1) and (7) in Fig. 4) compared to the uplink case.

The BS part decision algorithm relies on the values of two major QoS parameters, i.e., the packet loss rate and the mean delay. The *packet loss rate* is the sum of i) the *packet error rate* (i.e., the percentage of packets that are lost due to channel errors) and ii) the *packet timeout rate* (i.e., the percentage of packets that are lost due to expiration). To calculate these rates, the BS part has to maintain up-to-date information on channel conditions in both directions, as well as traffic and QoS status of active connections.

The packet error rate is estimated based on the channel conditions. Channel conditions on the uplink (RSSI and CINR) are known from the PHY layer of the BS. Channel conditions on the downlink can be assumed similar to the uplink conditions or can be obtained by either the received REP-RSP messages or through the CQICH, as described earlier in section III (arrow (1) in Fig. 4). Packet timeout rate and mean delay for all active connections in both directions can be provided by the BS MAC layer (arrow (1) in Fig. 3 and arrow (2) in Fig. 4).

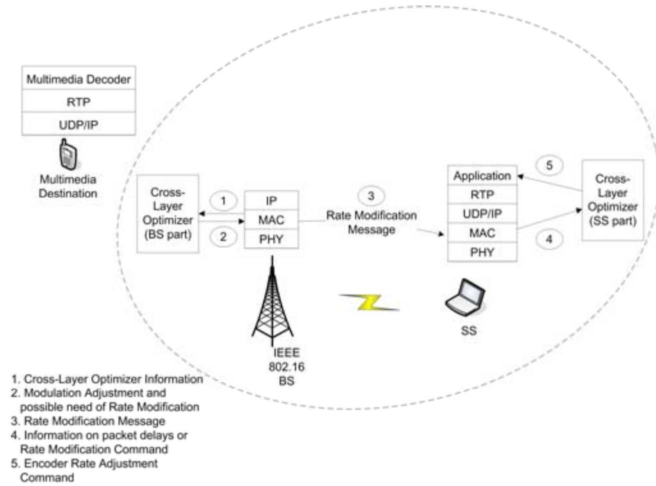


Fig. 3. Cross-Layer Optimization mechanism: Uplink

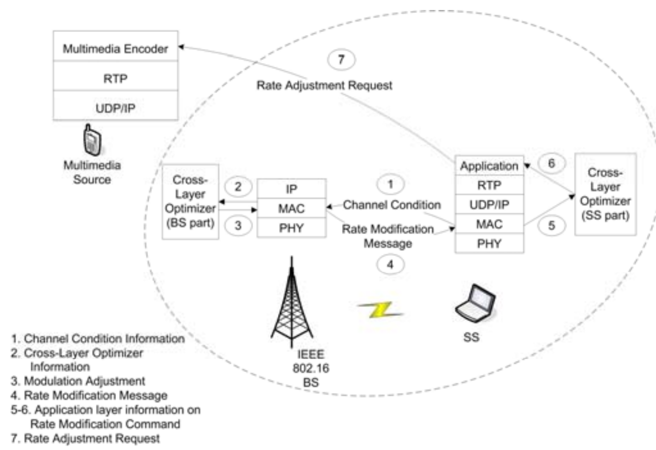


Fig. 4. Cross-Layer Optimization mechanism: Downlink

The decision algorithm of the BS part is shown at the flow chart of Fig. 5. More specifically, if at some point a SS faces unacceptable packet loss rates, the actions to be taken by the BS part depend on the nature of these losses:

1) *In case most of the losses are due to poor channel conditions (packet errors)*, the BS part instructs the MAC layer for a degradation of the modulation, in order to achieve higher channel error resilience and increase robustness against interference. Thus, the BS part selects the highest modulation that will restore the loss rate to acceptable values and instructs the MAC layer accordingly (arrow (2) in Fig. 3 and arrow (3) in Fig. 4). The error rate induced by the candidate modulation can be predicted using a BER vs. CINR curve for IEEE 802.16 PHY, such as the curve in [8]. The MAC layer sends the required primitives to the PHY for the modulation change and informs the SS through the DL-MAP (for downlink changes) and UL-MAP (for uplink changes) fields of the MAC time frame.

2) *In case most of the losses are the result of packet timeouts (unacceptable delays)*, the action to be performed depends on the contribution of these timeouts to the overall packet losses:

a. *If the loss rate is caused almost exclusively by packet timeouts*, the BS part can safely conclude that the channel is very slow and unable to satisfy the transmission speed requirements. In this case, the BS part should instruct a modulation upgrade in order to increase the transmission speed and reduce the losses caused by timeouts. Again, the error rate induced by the candidate modulation can be predicted using a BER vs. CINR curve, such as the curve in [8].

b. *In different case, when a significant percentage of packet losses are caused by errors due to the poor channel conditions*, that do not allow a modulation upgrade, the BS part instructs the SS part for a traffic rate reduction in order to moderate timeouts.

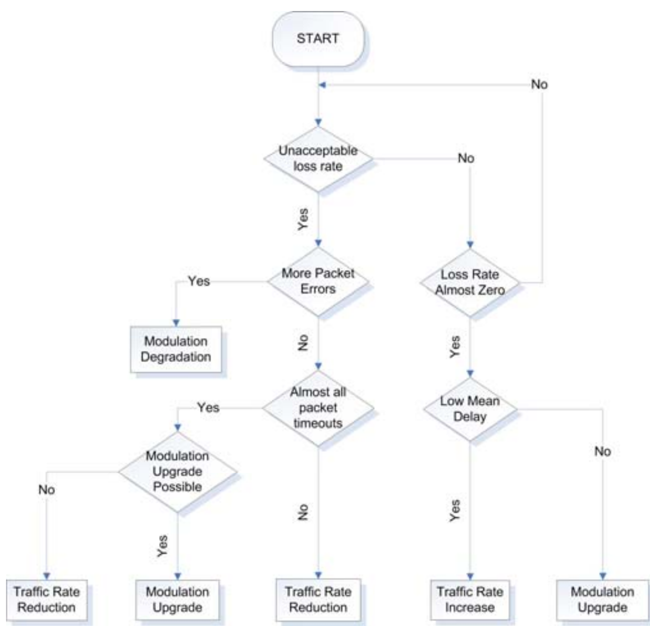


Fig. 5. Decision algorithm flow chart

To perform efficiently under all conditions, the cross-layer optimizer has to take proper actions also when the conditions are improved. Thus, when the loss rate decreases significantly, the BS part may decide to either switch to a higher modulation and increase the available bandwidth, or instruct the SS part to increase the traffic rate and improve QoS. The specific action depends on the mean delay experienced by the active connections of the SS:

1) *If the mean delay is relatively low compared to the delay bound*, the BS part may instruct for a traffic rate increase to improve the service provided to the user. However, this traffic rate increase should be performed carefully, as it may lead to considerable delay bound violations. One approach is to allow a traffic rate increase only when the current mean delay does not exceed a specific threshold. In any case, the percentage of this increase should be relevant to the difference between the current value of the mean delay and the delay bound.

2) *If the mean delay is close to the delay bound*, the BS part may instruct for a modulation upgrade to increase transmission speed and reduce delays.

The communication between BS and SS parts for a traffic rate reduction/increase is performed using the specially defined MAC management message RATE-ADJ-REQ (arrow (3) in Fig. 3 and arrow (4) in Fig. 4) described in section III, that contains the target overall mean traffic rate in either direction. The SS MAC layer transfers the received rate modification request to the SS part of the cross-layer optimizer (arrow (4) in Fig. 3 and arrow (5) in Fig. 4) that is responsible for the communication with the Application layer. The SS part decides on the connections that should be affected and sends proper cross-layer messages to the Application layer (arrow (5) in Fig. 3 and arrow (6) in Fig. 4). In the uplink direction, the Application layer can perform the proper adjustments (e.g., change the data encoding rates) to produce the required overall traffic rate. For example, in the case of a voice connection, the Adaptive Multi-Rate (AMR) voice encoder can switch to the desired maximum encoding rate ([9], [10]). In the downlink direction, the SS has to notify the distant traffic sources for the necessary rate adjustments (arrow (7) in Fig. 4), using whatever feedback channel is available. In the case of an AMR voice connection, the Codec Mode Request (CMR) signal is used [11], which suggests a new encoding mode to the sender (i.e., a new maximum encoding rate). Since most voice sessions are bi-directional, this signal is piggybacked into the data frames in the reverse direction (in-band signaling).

From the above discussion it is clear that a key issue for the performance of the proposed mechanism is the definition of the thresholds of the packet loss rate and mean delay that activate modulation and traffic rate changes. Moreover, due to the fact that the communication between the BS and the traffic source may be relatively slow, especially in the downlink case, the operation of the traffic rate adjustment should be performed less frequently than the channel adaptation because it might not be able to follow frequent condition changes, leading to inefficiency. Thus, the decision algorithm should activate traffic rate changes only when it is expected that these will remain valid for a relatively long period of time.

We have to note that the proposed mechanism introduces some extra complexity, mainly at the BS, and limited signaling overhead for channel quality information and traffic rate modifications. On the other hand, the benefit in terms of QoS and system's capacity is such that justifies its use, as shown in the next section.

V. SIMULATION MODEL AND RESULTS

To measure the performance of the proposed cross-layer optimization mechanism against a legacy IEEE 802.16 system employing no adaptation, a simulation model was constructed in C++. The main components of the simulator are the following (Fig. 6):

Traffic Generator: For each connection it generates one AMR variable-length traffic frame every 20ms, starting at a random instance within the first second of a simulation run. Each connection's maximum rate depends on the feedback received from the Optimizer.

Scheduler: It receives input traffic from the Traffic Generator and information on the SSs' modulation changes from the Optimizer, decides on the structure of each time frame and provides the Optimizer with the necessary QoS status information. In our simulations we used the "Frame Registry Tree Scheduler" (FRTS) described in [12].

Channel Modeler: The Channel Modeler emulates the system's channel conditions. Bit errors are randomly produced for each connection, according to its modulation scheme information provided by the Scheduler, with a mean rate according to the BER vs. CINR curve included in [8].

Optimizer: It follows the operation of the proposed cross-layer optimizer described in the previous section. Based on the information received from the Channel Modeler and the Scheduler, it decides on the optimal modulation scheme of each SS and the traffic-encoding rate of each connection, and informs the Scheduler and the Traffic Generator accordingly.

For the non-optimized system, the Optimizer remained idle without sending modulation adjustment instructions to the Scheduler or traffic rate changes to the Traffic Generator.

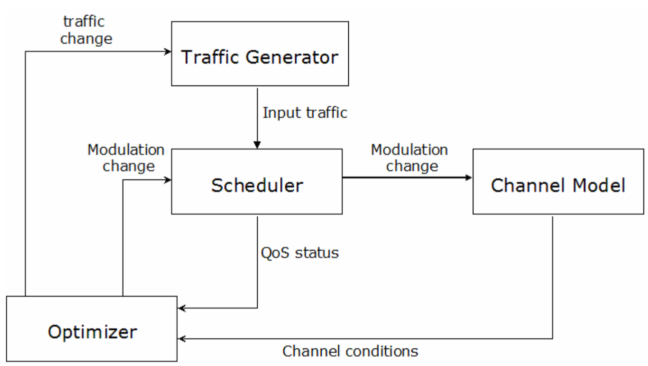


Fig. 6. Simulation Outline

The simulation scenario considered an increasing number of SSs, each one with one downlink AMR voice connection, and measured the system's performance with and without the use of the optimizer in terms of packet loss rate and mean delay. To focus on the optimization process, we considered the BS as the traffic source for all connections. The available modulation schemes were QPSK, 16-QAM and 64-QAM. The time frame length was set to 1ms and the maximum transmission speed to 120 Mb/s (when the modulation was equal to 64-QAM). In order to achieve lower processing complexity in the simulation model, a percentage equal to 15% of this bandwidth was reserved for the above connections, while the rest was assumed dedicated to other kinds of traffic. In all simulations, the initial states and channel conditions were the same for both systems (optimized and non-optimized).

Fig. 7 depicts the packet loss, packet error and packet timeout rates for an increasing number of subscribers, with and without the use of the cross-layer optimizer. Assuming that the total packet loss rate for AMR traffic should not exceed the value of $7x10^{-3}$ [13], we observe that the non-optimized system can accommodate up to approximately 228 voice connections. On the contrary, the cross-layer optimizer manages to keep the packet loss rate below the threshold for more than 250 connections. The exact number of the supported connections in the optimized system depends mainly on the channel conditions that affect the modulation adjustments.

The main reason for the considerably improved performance of the proposed cross-layer optimization mechanism, especially under heavy traffic conditions, is its ability to reduce the rate of packet timeouts by instructing the SSs to switch to a higher modulation schemes when possible, thus increasing the system's bandwidth. The inevitable increase of packet errors remains within reasonable values leading to an impressively low overall packet loss rate, especially for large numbers of subscribers. Its adaptation capability enables the connections to overcome severe channel error conditions and bandwidth restrictions. As expected, the mean delay was almost the same with and without the use of the optimizer in all cases, since this is affected mainly from the deadline-based scheduler (FRTS) [12], that schedules outgoing traffic as close to its deadline as possible.

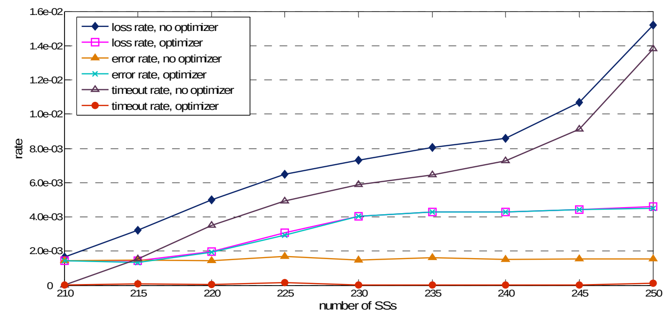


Fig. 7. Packet loss rate with and without the optimization mechanism

To reveal the effectiveness of the optimizer, we illustrate in Fig. 8 the correlation between the packet error and timeout rates (8a), the modulation changes (8b) and the AMR maximum rate adjustments (8c) during a time interval. For presentation purposes, we used simulation parameters that resulted in frequent modulation and maximum encoding rate changes, in order to illustrate the optimizer's operation in a relatively small interval of time. According to these figures, the SS starts operating in QPSK and produces AMR packets at a maximum encoding rate of 12.2 kb/s. However, this modulation proves to be unable to cope with the connection's transmission speed requirements and after 1.2 seconds packets start violating their delay bound and get rejected by the scheduler. The optimizer decides to increase the modulation to 64-QAM. However, the conditions remain unfavorable, resulting in more packet timeouts and packet errors. To handle the packet timeouts, the optimizer decides to gradually reduce the maximum encoding rate to 5.15 kb/s. After a short interval, the optimizer reduces the modulation as well, to avoid further packet errors. At around 2 seconds of operation, packets start dropping again due to timeouts, and the optimizer further reduces the maximum encoding rate to 4.75 kb/s. Nevertheless, the channel is very slow and packets continue to drop, thus, the optimizer instructs a modulation increase from QPSK to 64-QAM. After sometime, packet errors increase and force the optimizer to switch the modulation to 16-QAM. Finally, at 2.8 seconds the optimizer decides that the loss and error rates are low enough to allow an increase of the maximum encoding rate to 12.2 kb/s.

As already mentioned, key parameters in the proposed mechanism are the thresholds of packet losses and mean delays that activate the modulation and traffic rate changes. These thresholds depend on various factors (such as the number of SSs, their QoS requirements, the scheduler's performance, the channel conditions and the overall system load) and can strongly affect system's performance. In our simulation scenarios we performed manual adjustments of these thresholds at the beginning of every simulation run, in order to reveal the effectiveness of the proposed mechanism. Our future work includes an algorithm for dynamic thresholds settings.

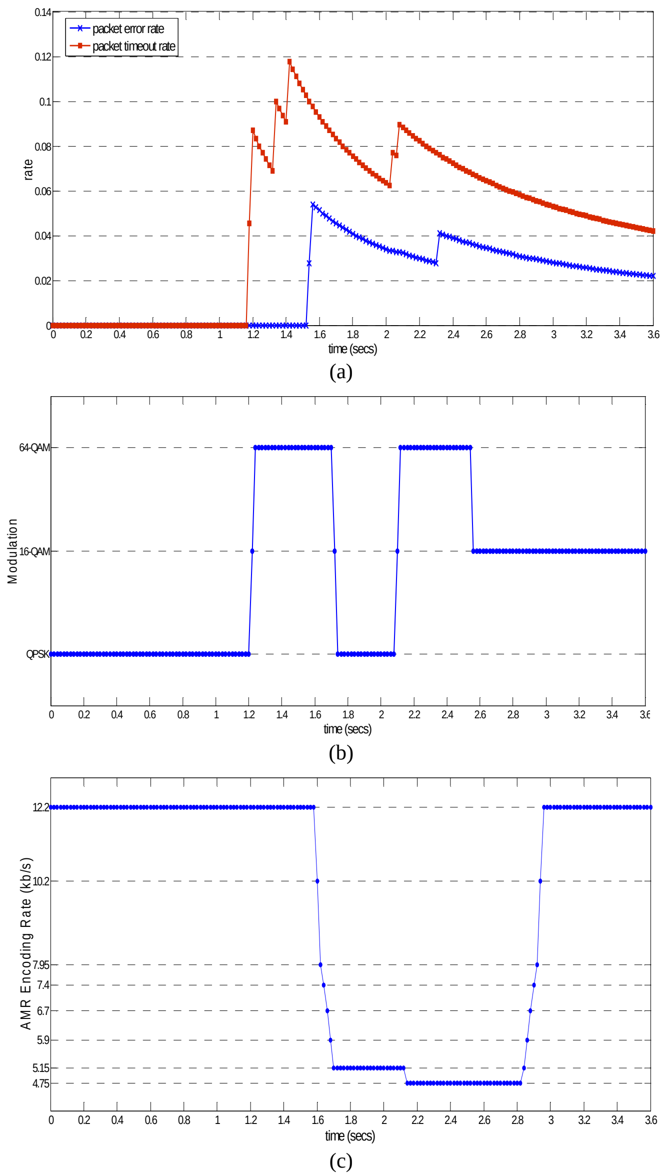


Fig. 8. Correlation between error and loss rates (a), modulation changes (b) and maximum encoding rate changes (c)

VI. CONCLUSIONS

In this paper we introduced a cross-layer optimization mechanism for IEEE 802.16 networks that aims at providing improved performance to multimedia applications by utilizing information provided by the PHY and MAC layers, in order to uniformly control modulation and data encoding rates. Its operation is based mainly on the rates and kind of packet losses and the values of mean delays that activate different actions.

With the aid of a detailed simulation model, we showed that the use of the proposed optimization mechanism leads to an efficient exploitation of the adaptation capabilities, resulting in considerably reduced packet loss rates and improved system's capacity, especially under heavy traffic conditions.

Our future plans include more extensive simulations with multiple kinds of traffic, the design of the algorithm for the dynamic thresholds setting, the use of a more active SS part able to refine or even change the BS part decisions, and the extension of the mechanism in order to control more parameters, such as the Forward Error Correction (FEC) coding and the Automatic Repeat reQuest (ARQ) window.

REFERENCES

[1] IEEE Std 802.16, "IEEE Standard for Local and metropolitan area networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems", October 2004.

[2] IEEE Std 802.16e, "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1", February 2006.

[3] T. Kwon, H. Lee, S. Choi, J. Kim, D.-H. Cho, S. Cho, S. Yun, W.-H. Park and K. Kim, "Design and Implementation of a Simulator Based on a Cross-Layer Protocol between MAC and PHY Layers in a WiBro Compatible IEEE 802.16e OFDMA System", IEEE Communications Magazine, vol. 43, no. 12, December 2005.

[4] Q. Liu, X. Wang and G. Giannakis, "A Cross-Layer Scheduling Algorithm with QoS Support in Wireless Networks", IEEE Transactions on Vehicular Technology, vol. 55, no. 3, May 2006.

[5] T. Holiday and A. Goldsmith, "Optimal Power Control and Source Channel Coding for Delay Constrained Traffic Wireless Channels", IEEE ICC 2003, Anchorage, AK, May 11-15, 2003.

[6] Q. Liu, S. Zhou and G. Giannakis, "Queuing with Adaptive Modulation and Coding over Wireless Links: Cross-Layer Analysis and Design", IEEE Transactions on Wireless Communications, vol. 4, no. 3, May 2005.

[7] S. Khan, Y. Peng, and E. Steinbach, "Application-Driven Cross-Layer Optimization for Video Streaming over Wireless Networks", IEEE Communications Magazine, vol.44, no.1, January 2006.

[8] D. Poulin, P. Piggin, R. Yaniv, and D. Andelman, "Correction to Rx SNR, Rx sensitivity, and Tx Relative Constellation Error for OFDM and OFDMA systems", IEEE 802.16 working group contribution, document number IEEE C802.16maint-05/112r8, September 2005.

[9] 3GPP; Technical Specification Group Services and System Aspects; Mandatory speech CODEC speech processing functions; AMR speech CODEC; General Description (3GPP TS 26.071 V6.00 (2004-12))

[10] 3GPP; Technical Specification Group Services and System Aspects; Mandatory speech codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec frame structure (3GPP TS 26.101 V6.00 (2004-09))

[11] RFC 3267, "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", June 2002.

[12] S. Xergias, N. Passas and L. Merakos, "Flexible Resource Allocation in IEEE 802.16 Wireless Metropolitan Area Networks", The 14th IEEE Workshop on Local and Metropolitan Area Networks, 2005, Sept. 2005.

[13] M. D. Yacoub, "Foundations of Mobile Radio Engineering", CRC Press, Boca Raton, FL, USA, 1993.