CROSS-LAYER FRAME SYNCHRONIZATION FOR H.264 VIDEO OVER WIMAX

Maria G. Martini and Chaminda T. E. R. Hewage

WMN, MINT Centre
Faculty of CISM, Kingston University London
Penrhyn Road, KT1 2EE, UK

ABSTRACT

In the WIMAX (802.16e) standard several small medium access control (MAC) packets may be aggregated to form a MAC burst, forwarded to the physical (PHY) layer [1] The burst transmission is a mechanism to allow multiple MAC protocol data units (PDUs) belonging to the same video channel to be transmitted / received in an aggregated way, in order to enable power saving at the mobile station, by putting the transceiver in sleep mode during off-burst interval. As a consequence of it, at the receiver side individual packets should be isolated within a burst, i.e. frame synchronization should be performed. We propose here to forward soft values from the PHY to the MAC layer, and to use our previous results on frame synchronization based on soft values [2] [3]. In addition, we exploit here a-priori information on the prevalence of 1's or 0's in the video bitstream and in packet headers, by performing application-aware and channel-aware MAC layer frame synchronization. Results show that the proposed approach results in an evident gain, at the expense of a small additional complexity.

Index Terms— Wimax, burst mode, cross-layer design, frame synchronization, soft values, a-priori information.

1. INTRODUCTION

In the mobile WIMAX (802.16e) standard several small MAC packets may be aggregated to form a MAC burst, forwarded to the PHY layer [1]. The burst transmission is a mechanism to allow multiple MAC PDUs belonging to the same video channel to be transmitted/ received in an aggregated way. Burst transmission enables mobile subscriber station (MSS) to save power by putting transceiver in sleep mode during off-burst interval. At the receiver side, individual packets should be isolated within a burst, i.e. frame synchronization has to be performed.

The most widely used method for providing frame synchronization is to insert a fixed symbol pattern or sync word (SW) into the data stream. The receiver obtains frame synchronization by locating the position of the sync word in the received data stream. A first, intuitive approach for locating such position consists of correlating, the received signal with the expected SW, looking for the position where this correlation is maximum in the case of fixed length window, or comparing the correlation with a threshold.

The problem of frame synchronization has been widely studied in the literature in the case of equiprobable data symbols: the performance evaluation for frame synchronization with periodically embedded sync words searched through correlation in binary symmetric channels (BSC) has been studied in [4], where also synchronization sequences with good aperiodic autocorrelation properties have been identified. The problem of optimum frame synchronization has been afforded in [5] on additive white gaussian noise (AWGN): the optimal metric for AWGN channel has been identified for the considered case of fixed length frames and equiprobable data symbols. In [6] a performance evaluation through simulation of these metrics has been presented. Synchronization for unknown frame lengths is studied in [2, 7, 8].

In many practical situations the assumption of uniform data distribution is not realistic. For instance in discrete cosine transform (DCT) coded video there is always a linear relationship between the coding bit rate and the percentage of zeros among the quantized transform coefficients [9], which is directly related to the image content and is a measure of picture complexity. Furthermore, variable length codes are used and consecutive runs of 0's and 1's are possible. For instance, MPEG-2, H.263, and MPEG-4 are generally based on fixed tables of variable length codes (VLC). Context-based Adaptive Binary Arithmetic Coding (CABAC) is a normative part of the ITU-T, ISO/IEC standard H.264/AVC for video compression [10]. If in the bitstream 1's or 0's are prevailing, this information can be exploited in the detection metric [11] [8] [2] [3].

In this paper we propose to perform frame synchronization at WiMAX MAC layer through the insertion of SWs in the burst to separate individual packets. Although classically sync word detection would be performed through correlation, we propose here to forward soft values from the PHY layer to the MAC layer, in order to use the optimal metric derived in [2]. Furthermore, we study the statistical distribution of 1's and 0's in the H.264 stream in order to exploit it as a-priori

This work was partially supported by the European Commission under FP7 project INFSO-ICT-214625 OPTIMIX.

information in the metric in [3].

The paper is organized as follows. Section 2 provides a short review of the relevant characteristics of the WiMax MAC layer. Section 3 presents an analysis of the a-priori information that can be exploited about data originated from H.264 compression. The proposed frame synchronization strategy is presented in Section 4 and its performance is evaluated through simulation in Section 5.

2. THE WIMAX MAC LAYER - BURST MODE

In mobile WIMAX (802.16e) several small MAC packets may be aggregated to form a MAC burst, forwarded to the PHY layer [1]. The burst transmission is a mechanism to allow multiple MAC PDUs belonging to the same video channel to be transmitted/ received in an aggregated way. Burst transmission enables MSS to save power by putting transceiver in sleep mode during off-burst interval. Figure 1 shows the TDD frame structure for mobile WiMAX [1]. DL sub-frames begin with a frame control section that contains the downlink map (DL-MAP) for the current downlink frame as well as the uplink map (UL-MAP) for a frame in future. The DL-MAP informs all Subscribers Stations (SSs) of which part of the current frame they should listen to. The UL-MAP informs SSs of their transmission opportunities as a response to their dynamic bandwidth requests, or on the basis of prior service agreements. These are then followed by the transmission of the DL sub-frame and the UL sub-frame. The DL sub-frame is divided into a number of Time Division Multiplex (TDM) portions, the so-called burst profiles: all traffic associated to a particular burst profile is transmitted sequentially. Each burst can contain multiple concatenated fixed-size or variable- size packets or fragments of packets received from the higher layers.

At the receiver side, individual packets should be isolated within a burst. This is not trivial when transmission occurs over noisy channels that corrupt the data in the burst.

3. ANALYSIS OF THE H.264 VIDEO STREAM

As anticipated, in DCT coded video there is always a linear relationship between the coding bit rate and the percentage of zeros among the quantized transform coefficients [9], which is directly related to the image content and is a measure of picture complexity. The percentage of zeros monotonically increases with the quantization parameter. Furthermore, variable length codes are used to encode the transform coefficients and consecutive runs of 0's and 1's are possible.

This can results in a global statistic prevalence of 0's or 1's in the bitstream or in specific regions in the bitstream where either bit prevails. Such information can be exploited as a-priori information for frame synchronization.

In fact, although the goal of source coding is to provide a bistream of i.i.d. bits, this is not true in practical coding schemes. As an example, in MPEG-4 1's are prevailing in the part of the stream corresponding to encoded video and for that reason markers are selected with a prevalence of zeros. We have observed that this is similar in H.264 video (see e.g. [10] for the selection of variable length codewords). In particular 1's tend to prevail in I frames (where DCT coefficients are encoded) when slices are not considered. When slices are considered this is less evident due to a prevalence of zeros in the packetization overhead.

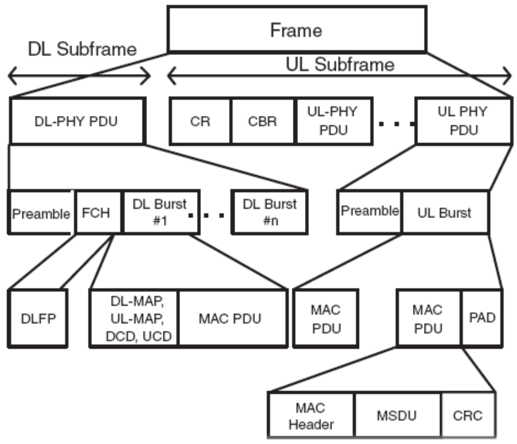

Fig. 1. TDD frame structure for mobile WiMAX [1].

We report in Figure 2 results for the case where slices are not considered and six different video sequences, with different characteristics in terms of motion and texture complexity, are considered. Results are obtained by encoding 300 frames for each sequence in CIF format with a quantization parameter $QP = 50$ and no slices. We can observe in particular that, for I frames, bits equal to 1 prevail in the stream and their percentage is 59.5 %. We have observed (data not reported due to lack of space) that the prevalence of 1's increases with the quantization parameter.

This result can be exploited, together with soft values, for improving frame synchronization at MAC layer.

4. FRAME SYNCHRONISATION WITH SOFT VALUES AND SOURCE A-PRIORI INFORMATION

We propose in this paper to exploit cross-layer information for improving frame synchronization at MAC layer. We focus in particular on the information below:

- soft values, or decision reliability information (DRI), from the physical layer;
- source a-priori information (SAI), in particular about the statistical prevalence of ones and zeros.

Figure 3 shows a schematic representation of our approach, where information from the upper layers and from the physical layer are exploited at MAC layer for the purpose of frame

Sequence	I frame 0s	I frame 1s	I frame 0s %	I frame 1s %	P frame 0s	P frame 1s	P frame 0s %	P frame 1s %
NEWS	3897	5623	40.93487395	59.06512605	58276	50956	53.35066647	46.64933353
FOREMAN	3052	4052	42.96171171	57.03828829	145857	147679	49.68964624	50.31035376
MOBILE	10482	14614	41.76761237	58.23238763	236049	276647	46.04073369	53.95926631
AKIYO	2866	4190	40.61791383	59.38208617	21144	13504	61.0251674	38.9748326
COASTGUARD	1680	3176	34.59637562	65.40362438	84025	86839	49.17653807	50.82346193
BRIDGE-FAR	1228	2452	33.36956522	66.63043478	14370	7158	66.75027871	33.24972129
Average	23205	34107	40.48890285	59.51109715	559721.0000	582783	48.99072563	51.00927437

Fig. 2. H.264 video statistics. Sequences in CIF format, 300 frames, QP=50.

synchronization. In particular, soft values or DRI improve detection of the synchronization words separating different packets, if an appropriate detection metric is exploited. Soft values have been used in the past in cross-layer approaches to support variable length codes (VLC) decoding, as recently in [12].

A-priori information about the source can contribute further to SW detection by selecting appropriately the SWs and the detection metric.

Note that in this work we focus on SAI, since source data represent the largest part of data at MAC layer, but available information about all the upper layers can be considered, by taking into account network headers' patterns and statistics. For instance, in case of transmission over RTP/UDP/IP the information about the relevant headers can be considered, substituted by information about RoHC header in case header compression is applied.

This approach can thus involve all the layers of the protocol stack. For signaling issues (transfer of the needed information across the protocol stack) we refer to [13].

We will show the separate and joint effect of the exploitation of the two types of information. We will perform the comparison: with the case where such information is not used, i.e. optimal metric when "hard" values are used (correlation); with the case with optimal metric when "soft" values are used, but no a-priori information.

We will use the notation in [2], in the following shortly reviewed and specialized to the case under investigation.

We consider payload symbols $d_i \in \{+1,-1\}$, with probabilities $Pr(d_i = 1) = p_1$, $Pr(d_1 = -1) = 1 - p_1$, SW symbols $c_i \in \{+1,-1\}$ and i.i.d. noise samples n_i. The demodulator output consists of a sequence of N symbols, the sampled matched filter outputs. Let this sequence (a random variable) be denoted by $\mathbf{R}$; the actual value of the received vector is the sequence $\mathbf{r} = (r_1, ..., r_N)$, and is composed of either SW symbols and noise or the sum of data symbols and noise. Transmission is over an additive white Gaussian noise channel (AWGN) whose samples are i.i.d with zero mean and variance $N_0/2$, with N_0 the one-sided power spectral density.

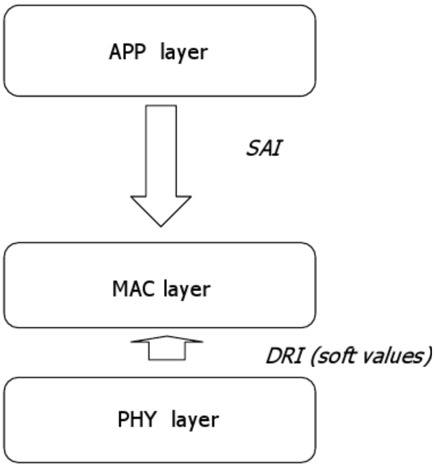

Fig. 3. Schematic representation of MAC layer frame synchronization based on cross-layer information.

In the case of binary modulation, the i^{th} transmitted bit $b_i \in \{0,1\}$ gives rise, after binary antipodal modulation, transmission through the AWGN channel, matched filter reception and perfect sampling, to a sample $r_i = (-1)^{b_i} + n_i$, where n_i are independent, identically distributed (i.i.d.) Gaussian random variables (r.v.'s), with zero mean and variance σ^2. This model is also valid for BPSK systems, for which the signal-to-noise ratio is $E_s/N_0 = 1/(2\sigma^2)$, where E_s is the energy per symbol.

We assume that a sync word composed of N symbols, $(c_1, ..., c_N)$, is aperiodically inserted in the data stream, composed of symbols d_i.

The acquisition algorithm we consider is as follows: starting from a position k, the synchronizer observes a vector of N subsequent samples; based on a suitable metric evaluated from this vector it decides if the SW is in position k; if not, it moves to position $k+1$, repeating the steps until the sync word is detected.

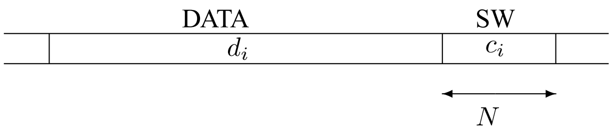

Fig. 4. Frame Structure.

We afford here the problem of deciding at each position k of the MAC burst whether a sync word is present or not.

As in our previous work, we assume that the statistical properties of the metric do not depend on the position in the bitstream. We thus avoid considering the effect of the aperiodic autocorrelation around SW's. We have evaluated in fact that, if the SW is properly chosen as a sequence with optimized aperiodic autocorrelation property (e.g. Barker sequences [4]) its symbols should mimic random data, with, moreover, the additional property that some configurations can be avoided.

As in [2] we study the problem through the statistical theory of hypothesis testing. After observing N subsequent samples, the synchronizer must choose between two possible situations

$$\begin{aligned} \mathcal{H}_0 &: \quad r_i = d_i + n_i, \quad i = 1, \ldots, N \\ \mathcal{H}_1 &: \quad r_i = c_i + n_i, \quad i = 1, \ldots, N \end{aligned} \tag{1}$$

the first hypothesis representing the case where there is no sync word, the second corresponding to the case the sync word is present. Decisions are indicated by $\mathcal{D}_0, \mathcal{D}_1$, corresponding to the "true" hypotheses $\mathcal{H}_0, \mathcal{H}_1$, respectively.

Differently from previous studies, we assume here as in [3] that the data symbols d_i are not necessarily equiprobable and that their probabilities of occurrence in the data stream are known.

4.1. Detection metrics

Assuming AWGN, soft values available, and knowledge of data distribution, based on the likelihood ratio test (LRT) [14] we derived the optimal detection metric in [3].

In the following we use capital letters to indicate random variables and bold for vectors and we indicate with $\mathbf{R} = (R_1, ..., R_N)$ the r.v. corresponding to the vector $\mathbf{r} = (r_1, ..., r_N)$ of received samples. In the aforementioned case we have:

$$\Lambda_p(\mathbf{r}) = \sum_{i=1}^{N} \log \left[Pr(d_i = c_i) + Pr(d_i \neq c_i) e^{-\frac{2 r_i c_i}{\sigma^2}} \right] \underset{\mathcal{D}_1}{\overset{\mathcal{D}_0}{\gtrless}} \lambda. \tag{2}$$

We thus decide $\mathcal{D}_1$ (the start code is present) if $\Lambda_p(\mathbf{r}) < \lambda$, $\mathcal{D}_0$ otherwise. The threshold is chosen according to the Neyman-Pearson criterion, i.e. by fixing the maximum tolerable probability of false alarm (emulation). We should note that the obtained metric, similar as the one for the equiprobable case, depends on channel conditions through σ^2 and thus the synchronizer requires the instantaneous knowledge of the signal to noise ratio in order to perform optimum detection. Furthermore, the evaluation of a non-linear function is required.

In case we have availability of soft values, but no a-priori information about the data distribution, we have [2]:

$$\Lambda(\mathbf{r}) = \sum_{i=1}^{N} \ln \left(1 + e^{-\frac{2 r_i c_i}{\sigma^2}} \right) \underset{\mathcal{D}_1}{\overset{\mathcal{D}_0}{\gtrless}} \lambda. \tag{3}$$

If we do not have availability of / we do not exploit soft values from the physical layer, we resort to correlation:

$$\Gamma(\mathbf{r}) = \sum_{i=1}^{N} c_i \text{sign}(r_i) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \lambda. \tag{4}$$

5. PERFORMANCE EVALUATION

We report in this section numerical results obtained with the exploitation of source a-priori information and soft values with the metric in (2), in comparison with results obtained with the metric obtained when a-priori knowledge about the data symbols is not exploited, i.e. in the hypothesis of equiprobable data symbols (LRTU) (3). Results will also be compared with the case where neither a-priori information nor soft values from the PHY layer are available/exploited.

Results are presented in terms of probability of emulation, P_{EM}, or false start code detection, and of probability of missed detection, P_{MD}.

The probability of emulation, P_{EM}, or false start code detection, of choosing hypothesis $\mathcal{H}_1$ when $\mathcal{H}_0$ is true is

$$P_{EM} = \Pr\left\{\mathcal{D}_1 | \mathcal{H}_0\right\}. \tag{5}$$

Note that here the false start code detection is due to the case where random data plus noise is interpreted as a SW. This can occur either in the case data symbols are coincident with the SW pattern or, due to noise, even if data symbols are different from the SW pattern.

The probability of missed detection, P_{MD}, of choosing $\mathcal{H}_0$ when $\mathcal{H}_1$ is true, is

$$P_{MD} = \Pr\left\{\mathcal{D}_0 | \mathcal{H}_1\right\} \tag{6}$$

and the probability of correct detection is

$$P_D = 1 - P_{MD}. \tag{7}$$

By analyzing such probabilities for different values of the threshold λ, we can draw a receiver operating characteristic (ROC) curve, reporting P_D versus P_{EM}.

In Figure 5, we report in the same plot different ROC curves, each related to a specific metric used for detection.

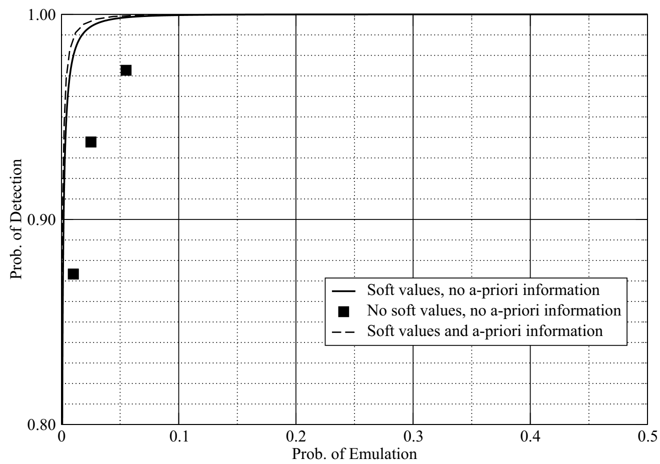

Fig. 5. Receiver Operating Characteristic (ROC) curves. I frames payload. Turyn [15] sync word of $N = 32$ bits. AWGN $\sigma^2 = 2dB; E_s/N_0 = -5dB$.

The considered SW is the Turyn [15] $N = 32$ SW, composed of 19 symbols equal to -1 and 13 symbols equal to 1. Transmission over AWGN with $E_s/N_0 = -5dB$ is assumed. We focus on I frames packets, with the statistics in Figure 2. We can observe that the exploitation of soft values results in a remarkable improvement in the performance of SW detection. An additional improvement can be achieved by exploiting a-priori information.

Ongoing work is investigating the impact of different a-priori information, e.g. local information about header patterns and statistics.

6. CONCLUSIONS

We have proposed a cross-layer approach for frame synchronization at MAC layer in video transmission over WiMax. We proposed to exploit soft values forwarded from the PHY to the MAC layer and a-priori information on the prevalence of 1's or 0's in the video bitstream and in packet headers, by performing application-aware and channel-aware MAC layer frame synchronization. Simulation results show that the proposed approach results in an evident gain, at the expense of a small additional signaling and computational complexity.

7. REFERENCES

[1] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMax*. Prentice-Hall, 2007.

[2] M. Chiani and M. G. Martini, "On sequential frame synchronization in AWGN channels," *IEEE Trans. Commun.*, vol. 54, no. 2, February 2006.

[3] M. G. Martini and M. Chiani, "Optimum metric for frame synchronization with gaussian noise and unequally distributed data symbols," in *Proc. IEEE Sig. Proc. Advances for Wireless Comms. (SPAWC)*, Perugia, Italy, June 2009.

[4] R. H. Barker, "Group synchronization of binary digital systems," in *Communication Theory*. London, England: W. Jackson, 1953.

[5] J. L. Massey, "Optimum frame synchronization," *IEEE Trans. on Commun.*, vol. 20, no. 4, pp. 115–119, 1972.

[6] P. T. Nielsen, "Some optimum and suboptimum frame synchronizers for binary data in Gaussian noise," *IEEE Trans. on Commun.*, vol. 21, no. 6, pp. 770–772, 1973.

[7] M. Chiani and M. G. Martini, "Optimum synchronization of frames with unknown, variable lengths on Gaussian channels," in *Proc. IEEE Global Telecomm. Conf.*, Dallas, TX, Nov. 2004, pp. 4087–4091.

[8] ——, "Practical frame synchronization for data with unknown distribution on AWGN channels," *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 456–458, May 2005.

[9] Z. He and S. Mitra, "A linear source model and a unified rate control algorithm for DCT video," *IEEE Transactions on Circuits and Systems for Video Tech.*, vol. 12, no. 11, pp. 970–982, Nov. 2002.

[10] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Transactions on Circuits and Systems for Video Tech.*, vol. 13, no. 7, pp. 620– 636, 2003.

[11] M. Chiani and M. G. Martini, "Frame synchronization for unequally distributed data symbols," in *Proc. of IEEE Globecom 2006*, San Francisco, Nov-Dec. 2006.

[12] U. Ali and P. Duhamel, "Joint protocol-channel decoding for robust aggregated packet recovery at WiMax MAC layer," in *Proc. IEEE Sig. Proc. Advances for Wireless Comms. (SPAWC)*, Perugia, Italy, June 2009.

[13] M. G. Martini, M. Mazzotti, C. Lamy-Bergot, J. Husko, and P. Amon, "Content adaptive network aware joint optimization of wireless video transmission," *IEEE Communications Magazine*, vol. 45, no. 3, pp. 1–10, 2007.

[14] H. L. Van Trees, *Detection, Estimation and Modulation Theory: Part I*. New York: Wiley, 1968.

[15] R. Turyn, "Sequences with small correlation," in *Error Correcting Codes*, H. B. Mann, Ed. New York: Wiley, 1968.