RESOURCES ALLOCATION OPTIMIZATION FOR SCALABLE MULTIMEDIA DATA SUBJECT TO QUALITY OF SERVICE CONTRAINTS

Heykel Houas, Cleo Baras, Inbar Fijalkow

ETIS laboratory - UMR 8051- ENSEA University of Cergy Pontoise-CNRS
6, Avenue du ponceau,F-95014 Cergy-Pontoise Cedex, France
Email: {heykel.houas, cleo.baras, inbar.fijalkow}@ensea.fr

ABSTRACT

Scalable multimedia data transmission are subject to specific constraints such as the Quality of Service (QoS) of sensitivity classes and the transmission rate (yielding a maximum size of each frame to send). Many scalable source decoders are used to discarding data than processing an erroneous stream. This featuring class structure is helpful to define a strategy that determines the maximum number of classes to send and delivers the suitable protection and transmission scheme (coding rates and modulation) to apply in accordance with the transmission constraints. It leads to the possible truncation of frame parts transmitted with an Unequal Error Protection (UEP) scheme for severe channel conditions. In a MPEG-4 speech frames context, we compare our approach to other methods using equal error and existing UEP schemes. It results in a significant improvement of the Peak SNR (PSNR) quality in poor channel transmission conditions.

1. INTRODUCTION

Scalable encoding technics make easier the multimedia content delivery over heterogeneous networks with diverse end-user devices: as far as the representation of the multimedia data is concerned, they integrate into an unique bitstream several description of the same data at various rates and various perceived qualities; as far as the transmission of these data is concerned, they structure the data in frames, each split into classes of importance (or sub-frames) according to their sensitivity to erroneous bits. In this paper, we aim at profiting from this hierarchical structure to design a transmission strategy that ensures the Quality of Service (QoS) of the received data.

Most of transmission strategies proposed by the State-Of-The-Art for scalable data [1, 2] use an Unequal Error Protection (UEP) scheme [3, 4] in order to protect in a selective manner all classes of a frame with respect to their error sensitivity: coding rates and modulation for each class are selected to find the best tradeoff between the information rate over the channel and the perceived quality of the decoded data. Nevertheless, this method reaches its limit for severe channel conditions, where data are too much corrupted to be correctly decoded. The following question could therefore be tackled: is it better transmitting all enhancement layers and venturing to decoded roughly corrupted data than sending only a part of the frame with a better error protection? The tradeoff (on which the transmission strategy is based) should take into account the number of classes to send, as in [5] in a specific system. Thus, the transmission strategy design refers to the optimization of resources allocation with the constraints laid down by the application layer: the fixed high data rate (*i.e.* symbol rate) and the QoS to achieve.

In order to solve this optimization problem, we provide a strategy which estimates the maximum number of classes to send (with respect to the channel state) while ensuring the constraints of transmission rate and QoS. For poor channel conditions, this method appeals to a sub-frames truncation procedure combined with the matched modulation and error protection scheme on the frame to transmit. In the one hand, the UEP method ([6] and [3]) adapts the encoding process to the required QoS starting from the most important classes to the least ones. In the other hand, throwing a part of the frame away allows to fulfill the TU size constraint. This approach is called **Flexible Transmission** (FT). To load the FT process, the Cross-Layer strategy [8] provides us the required information from the application layer about the sub-frames sizes, the transmission data rate, the QoS requirements to the downer layers (network, physical) and the current channel state information from the physical layer to the network one.

The paper is organized as follows. Section 2 describes the communication system. Section 3 formulates the optimization problem and proposes a solution while section 4 draws up its efficiency through an example (MPEG-4 audio stream for telephony) in competition with current used transmissions, that is full Equal Error Protection (EEP)/UEP transmissions of the frame. Finally section 5 will conclude on the contribution of this method.

2. SYSTEM DESCRIPTION

2.1. Source coding parameters

Thanks to the scalable encoding process, each TU can be split into I sub-frames (or sensitivity classes) denoted by $\{C_i\}_{i=1..I}$ with various importance degrees, named source significant information. This parameter represents the susceptibility to errors for each class C_i. In other words, these degrees of importance feature the weight of each sub-frame on the perceived quality of the decoded media. Therefore, the higher the significance class is, the more the transmission of this class is sensitive to transmission errors. For each data-class C_i with $1 \leq i \leq I$, this level of importance can be featured by a BER to achieve, that will be denoted by B_i. It is assumed that BER requirements of each class (i.e. B_i) are previously defined according to some perceptual quality measures. Finally, P_i is defined as the proportion of the class C_i within a frame (so that $\sum_{i=1}^{I} P_i = 1$) and N is the frame length before channel coding (in bits).

This work is part of the FP6/IST project M-Pipe and is co-funded by the european Commission.

2.2. Transmission model

At the emitter, source data are encoded through an UEP scheme [4] that can be achieved using rate-compatible channel encoder [6]. In this paper, we use *Rate Compatible Punctured Convolutional* (RCPC) codes are used, but it could easily be extended to any other rate-compatible encoding process (such as turbo codes or LDPC codes [7]). The data are punctured with different coding rates, depending on the importance degree of each sub-frame C_i with $1 \leq i \leq I$. The mother convolutional coding rate is denoted by $R_1 = \frac{1}{N_1}$, whereas punctured coding rates are:

$$R_i = \frac{P}{P+l},\ 1 \leq l \leq (P-1)N_1,\ 1 \leq i \leq I, \qquad (1)$$

where P is the puncturing period. We will choose as an example a short-constraint-length L=9 as used in UMTS standard [**?**]. If we select $R_1 = \frac{1}{3}$, enumerator polynoms are:

$$G_1 = [557]_8, G_2 = [663]_8 \text{ and } G_3 = [711]_8. \qquad (2)$$

Three modulation schemes will be considered (BPSK, QPSK, 16-QAM) in order to transmit the data over an Additive White Gaussian Noise (AWGN) channel with constant symbol energy. We will refer to the term "mode m" to refer to a specific choice of a modulation.

At the receiver, a soft output Maximum A Posteriori (MAP) demapper is considered and soft values are used as the input of a classical Viterbi decoder. Three observations can even now be set:

- we do not need to transmit coding rates information to the receiver. Indeed, as we will see in section 3.3, the algorithm only depends on the $\frac{E_s}{N_o}$, applied on the physical layer and estimated by the receiver. By processing the algorithm at the receiver, coding rates information can therefore be recomputed at the emitter.
- $\frac{E_s}{N_o}$, required by the algorithm and estimated by the receiver, is supposed to be transmitted to the emitter using a perfect feedback channel.
- it is assumed that the transmission of the bitstream header is error-free.

The maximum symbol rate, denoted by S_{max}, is application-dependent and is defined at the physical layer. It refers to the physical payload.

3. PROBLEM AND SOLUTION

3.1. Problem Setting

The framework is the transmission of scalable multimedia data under transmission rate and QoS constraints. Several transmission strategies [3, 4] overcome the maximum bitstream rate in order to keep the QoS for each data class while other respect a maximum data rate constraint without ensuring a QoS for each class. In both cases, reliability or availability is sacrificed in the communication chain. Our aim is to solve the following problem: "How may QoS be hold back without overcoming S_{max}?". Considering the previously specified system, we propose a solution suitable for the scalable data transmission. This method is based on the following assumption: it is up to the physical layer to select which part of the frame is to be sent while respecting the transmission rate and QoS required from the application layer. It leads to an optimization problem, consisting in in the maximization of the number of classes to transmit within a frame and still:

- ensuring the QoS constraints,
- respecting the specified frame maximum size on the channel,
- favouring classes with higher importance.

This purpose can be fulfilled by making coding rates and the modulation vary from a sub-frame to another.

This problem can be stated as follows: for a given a $\frac{E_s}{N_o}$ (where E_s is the symbol energy and N_o the noise variance),

- look for $I_{flex} = \arg \max_{1 \leq k \leq I} \sum_{i=1}^{k} P_i N$, where I_{flex} is the adapted number of classes to send, and the matched coding rates R_i and m_i for $1 \leq m_i \leq Card(m_i)$ $1 \leq i \leq I_{flex}$, where $Card(m_i)$ is the number of available modes.
- under the following constraints:

  $\mathcal{A}_i$: $BER_i \leq B_i$, for $1 \leq i \leq I_{flex}$.

  $\mathcal{B}$: $S_{TU} \leq S_{max}$, where $S_{TU} = N \sum_{i=1}^{I_{flex}} \frac{P_i}{N_{m_i} R_i}$ is the number of symbols to transmit, N_{m_i} is the mode m_i symbol size for the class C_i,

  $\mathcal{C}$: class C_i is sent if the (i-1) previous ones are sent (transmission priority).

  where it is reminded that S_{Max} and B_i are defined by the application layer.

This is equivalent to source block data rate optimization under fixed channel rate, QoS and priority transmission constraints. Note that $\frac{E_s}{N_o}$ is our invariant parameter for the classes transmission over the channel (unlike $\frac{E_b}{No}$ that changes for each class) since energy per symbol is set to be constant when switching the modulation block. Thus, it will be used in order to establish performance curves of the methods we will compare in the next section.

3.2. Description of the algorithm

The algorithm we propose can be divided into three steps:

Step 1: **Initialization**

- $S_{Max} = \frac{N}{RN_m}$ inferred from the transmission rate (physical layer), N_m denoting the mode m symbol size,
- the most important class index to process: $j = 1$,
- the least important class index: $k = I$,
- $R_{m,i}$ the code rate applied on C_i with mode m
- η_m the spectral efficiency of the selected mode m,

Given a $\frac{E_s}{N_o}$,

Step 2: **Coding gain optimization**. Processing C_j,

- Scanning each mode m, $S_{TU} = S_{TU} + \frac{NP_j}{R_{m,j}N_m} - \frac{NP_j}{R_{m-1,j}N_{m-1}}$
  - ∗ While the constraint $\mathcal{B}$ is ensured or while the constraint $\mathcal{A}_j$ is not satisfied,
    1. decrease $R_{m,j}$ in order to reach $\mathcal{A}_j$ constraint,
    2. update S_{TU} according to the updated $R_{m,j}$ value.

- $*$ If $\mathcal{A}_j$ is reached for mode m,
  1. memorize the configuration $R_{m,j}^{-1}\eta_m$,
  2. select the minimum among this one and the previous one that ensures $\mathcal{A}_j$ condition.

– if $\mathcal{A}_j$ condition is reached for one mode m
  - $*$ if j=k then $I_{flex} = j = k$. Stop the algorithm,
  - $*$ else increment j so that j+1 becomes the most important class index and loop to step 2.

  Otherwise, go to step 3.

Step 3: **Truncation process**.

– While ($S_{TU} \geq S_{max}$),
  1. delete C_k and update S_{TU},
  2. decrement k so that k-1 becomes the new least important class index.

– If the removed class corresponds to the most important class that is under process during step 2 (ie k=j-1), stop the algorithm.
  Otherwise return to step 2.

First of all, the minimum selection process allows to run the least $R_m^{-1}\eta_m$ configuration in order to minimize S_{TU} consumption. As a matter of fact, the truncation level applied on the TU depends on the $\frac{S_{TU}}{S_{max}}$ ratio and on the data classes size we truncate. Truncation task can be invoked several times.

We finally get the set of classes $(C_1, ..., C_{I_{flex}})$ that can be transmitted and the associated coding rate $(R_1, ..., R_{I_{flex}})$.

3.3. Interpretation and discussion

- *Reception system complexity reduction*
  Information data rate optimization is performed at the physical layer: source coder sends a maximum and fixed bitstream to the physical layer that triggers the flexible transmission algorithm. This layer then gives the I_{flex} number of classes to the MAC layer that is supposed to make the decision about the sub-frames to truncate. Hence lower layers manage the information level within a frame that is to be transmitted over the channel. Therefore, messages exchange (between the source and the channel units) reduction results from this strategy.

- *Requirements in the communication system*
  On the one hand, the transmitter needs to access the channel state information (within the feedback information) in order to derive the flexible transmission. On the other hand, the receiver needs to know the truncation level applied to each received frame (number of deleted classes), code rates applied to the transmitted classes in order to decode correctly the received sequence. As feedback load consumes resources that would be otherwise used for data, we design our algorithm in both receiver and transmitter sides. It avoids to waste transmission rate for non information bits that are employed to describe the frame structure (in order to decode correctly the received TUs): truncation flags, I_{flex}, coding rates to apply to each class.

  Perfect feedback information received by the transmitter is assumed. It will ensure the optimality of the derived rate codes set given by our method. In addition, classes size is assumed to be constant so it will not affect the performance accuracy of the $(R_1, ..., R_I)$ set (if an important class C_i size increases, a part of its may not be coded with the same R_i (the tail) but with R_{i+1} which can be weaker than R_i).

  One critical question dealing with realtime applications is about the period of code rates calculation (and the number of classes to send to be maximized): for short frames, it is relevant not to run our algorithm for every TU, especially if the transmission rate is high (as it enlarges feedback information). It depends on the channel state time variation. We assume that propagation channel is time-invariant or if the case arises time-variant with not too short coherence time. Moreover, we consider that classes proportion does not change quickly.

4. APPLICATION TO MPEG-4 AUDIO DATA

4.1. CELP Audio scalable structure

An adaptation of the proposed algorithm to the transmission of scalable speech data is proposed to evaluate the algorithm performance. The MPEG-4 standard [9] includes a CELP coder, which can encode speech signals sampled at 8 kHz into 12-kbps bit-rate bitstream and provides Bit-Rate Scalability (BRS). By an appropriate choice of the encoder parameters, the bitstream frames are slit into a 4-classes structure. The first class, namely the base layer, is generated with a core CELP encoder operating at 6 kbps: it represents the speech signal thanks to a production model based on an *excitation signal* (encoded with 98 bits per frame) passed through an *auto-regressive filter* (represented by 22 bits per frame). The BRS tool provides the 3 remaining classes. Each class adds a 2 kbps information to the core bitstream and progressively reduces coding artefacts thanks to an improved encoding of the *excitation signal* (with 40 bits per frame).

The whole scalable bitstream is finally constructed as the succession of frames, each containing $N = 240$ bits and each divided into a 120-bit base layer and three 40-bit enhancement layers.

4.2. Test plan

4.2.1. Standard link parameters

The puncturing period P of the RCPC codes is chosen equal to 8. The code rates list $\mathcal{L}$ can then be obtained from equations (1) and (2):

$$\mathcal{L} = \{\frac{8}{9}, \frac{8}{10}, \frac{8}{12}, \frac{8}{14}, \frac{8}{16}, \frac{8}{18}, \frac{9}{20}, \frac{8}{22}, \frac{8}{24}\}.$$

The maximum symbol transmission rate is set to: $S_{max} = 360$ symbols per frame.

4.2.2. Audio scalable data parameters

According to the chosen scalable coder structure, the class proportions are the following: $P_1 = \frac{1}{2}$, $P_2 = P_3 = P_4 = \frac{1}{6}$. We assume that the QoS to reach for our application can be described by the following BER_i of each class:

$$B_1 = 10^{-4}, B_2 = 2.10^{-3}, B_3 = 6.10^{-3}, B_4 = 2.10^{-2}.$$

4.2.3. Performance criteria

Evaluating the efficiency of our flexible transmission strategy requires to determine if the QoS of each class (that is the B_i) is ensured but also to measure the perceived quality of the decoded speech signal. The latter criterion can be estimated by a Peak-Signal-To-Noise Ratio (PSNR) measure, defined as:

$$\text{PSNR} = \frac{1}{M}\sum_{m=0}^{M-1}\left[\frac{\max\limits_{n=0..N-1} x(mM+n)^2}{\frac{1}{N_a}\sum\limits_{n=0}^{N_a-1}\left(y(mM+n)-x(mM+n)\right)^2}\right],$$

where $x(n)$ is the original speech signal (from which the scalable bitstream is computed), $y(n)$ is the decoded speech signal (after its transmission on the considered channel), M is the number on analysis windows and N the length of each analysis windows. A sequence speech signal is sampled at 8 kHz and each with a duration of about 40-seconds, is used to measure the average PSNR value, in order to obtain more reliable results. N_a is chosen equal to 128 samples.

The performance of our flexible transmission will also be compared to those of three other strategies of protection against errors: an EEP scheme and UEP schemes. All these schemes use RCPC on audio scalable data, but differs in the choices of coding rates. Their parameters are the following:

- As far as the EEP scheme is concerned, QPSK and coding rates are chosen with respect to S_{max} and are set to: $R_i = \frac{8}{24}$ for $1 \leq i \leq 4$.
- The three specific UEP schemes are referred to as UEP_1, UEP_2 and UEP_3. The coding rates that define each UEP_i scheme are chosen among $\mathcal{L}^4$ in manner in the sense of PSNR performance, on three different regions of the $\frac{E_s}{N_o}$ scale. For instance, the coding rates of UEP_1 are determined in such a way that the first class of the speech data is favored with the strongest coding rate of $\mathcal{L}$, while the 3 remaining classes are encoded with the weakest ones coding rates and the highest level modulation. So each UEP_i favors the i most important classes (i=1,2,3). The derived coding rates are presented in table 1.

Table 1. UEP configurations list

Scheme	Class 1	Class 2	Class 3	Class 4
UEP_1	$R_1 = 8/18$	$R_2 = 8/9$	$R_3 = 8/9$	$R_4 = 8/9$
	BPSK	QPSK	QPSK	QPSK
UEP_2	$R_1 = 8/16$	$R_2 = 8/24$	$R_3 = 8/12$	$R_4 = 8/12$
	BPSK	QPSK	QPSK	QPSK
UEP_3	$R_1 = 8/16$	$R_2 = 8/22$	$R_3 = 8/16$	$R_4 = 8/9$
	BPSK	QPSK	QPSK	QPSK

4.3. Experimental results

According to the chosen constraints for QoS and maximum transmission unit size (that is S_{Max} and BER_i for each class $i = 1..I_{flex}$), our flexible transmission strategy can display with respect to $\frac{E_s}{N_o}$:

- the maximum number of classes I_{flex} to transmit
- the appropriate coding rates and modulations.

The obtained results, using the algorithm presented in section 3, are highlighted in figure 1. Four different configurations are available and delimit four regions in the $\frac{E_s}{N_o}$ scale: for example, in severe channel with $\frac{E_s}{N_o} \leq -1.8$ dB, the first class C_1 is only sent, while both classes C_1, C_2 and C_3 are transmitted for the $-1 \leq \frac{E_s}{N_o} \leq -0.27$ dB range. These four configurations are then used to simulate the flexible transmission of the speech coded signal over a physical layer with a varying $\frac{E_s}{N_o}$ and evaluate the efficiency of our strategy on the perceived quality of the decoded signal.

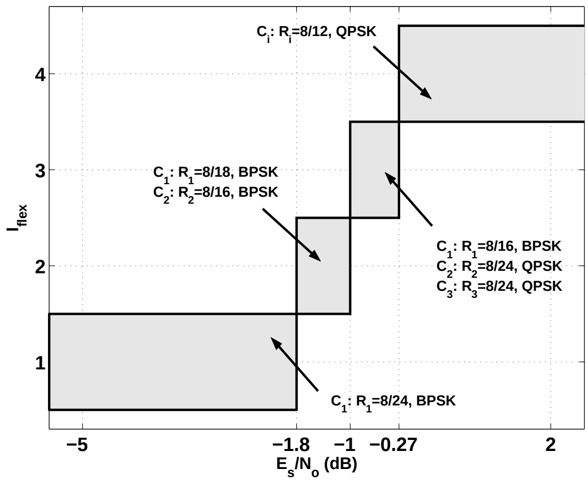

Fig. 1. A view of the solution to the optimization problem

Evaluating the perceived quality of the received speech signal with respect to $\frac{E_s}{N_o}$ is the purpose of figure 2. The performance of the different transmission schemes are here presented: the PSNR obtained thanks to our flexible transmission (FT) strategy but also those resulting from the transmission of the coded signal according to the EEP and the three UEP schemes. The figure also provides the maximum reachable PSNRs, denoted by the horizontal lines Q_1, Q_2, Q_3 and Q_4. The maximum reachable PSNR Q_i sets the perceived quality of the received signal when the channel is free from error and the number of decoded classes is i. Therefore, for any transmission strategy, the obtained PSNR should tend to these bounding values when $\frac{E_s}{N_o}$ increase. The results presented in figure 2 lead to the following conclusions:

- We observe that our proposed scheme has better performance for all $\frac{E_s}{N_o}$ values than the other schemes. The FT curve is prompter than the other ones to reach the maximum reachable PSNR Q_1, Q_2 and Q_3, when $\frac{E_s}{N_o} \leq 1$ dB: for instance, the FT method reaches the Q_1 line 1 dB sooner than the other (that is, more particulary, the EEP scheme at $\frac{E_s}{N_o} = -1$ dB). For $\frac{E_s}{N_o} \geq 0$ dB, the FT curve matches the EEP one, since EEP is known to reach the best transmission results for good channel conditions.
- The performance gap widens with the decrease of $\frac{E_s}{N_o}$. Indeed, for $\frac{E_s}{N_o} = -3$ dB (poor channel conditions), a 8 dB PSNR gain is noticed for FT compared to the best of the EEP and the UEPs scheme, that is UEP_1. In the meanwhile, for medium $\frac{E_s}{N_o}$ range ($-1.8 \leq \frac{E_s}{N_o} \leq -0.27$ dB), the contribution for FT is smaller but still remains significant: for instance, 0.8 dB for $\frac{E_s}{N_o} = -1$ dB.

  It means that the worst the channel state is, the higher PSNR improvement is achieved in relation to the other strategies (EEP, UEPs), fixing a minimum QoS.
- For poor physical channel conditions, it must be noted that

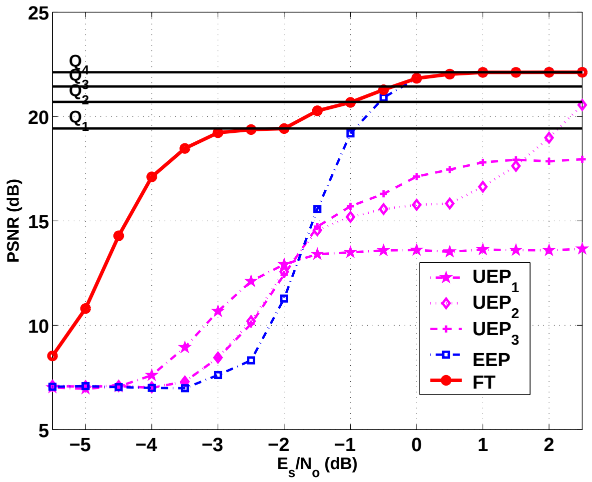

Fig. 2. Average PSNR with respect to $\frac{E_s}{N_o}$ for different transmission strategy: the EEP scheme, two UEP schemes and our FT strategy, and a free-from-error transmission (yielding the maximum achievable quality Q_1, Q_2, Q_3 and Q_4).

the EEP and usual UEP methods don't achieve the required QoS. Indeed as it focuses on the protection of the most i important classes for UEP$_i$, neglecting the other classes. Consequently, it significantly degrades the PSNR of the frame. In this case of low-to-moderate $\frac{E_s}{N_o}$, the proposed method achieves an important improvement with respect to the existing methods (about 5 dB in PSNR for a $\frac{E_s}{N_o}$ of -1.5 dB).

The quality improvement is obvious above all for FT scheme.

5. CONCLUSION

In this paper, we proposed a new algorithm for the transmission of scalable data with QoS and transmission load constraints. This so-called flexible transmission scheme discards some enhancements layers data from transmission in order to better protect the transmitted layers of higher importance. It process needs to maximize the number of enhancements layers to send under QoS and load requirements. It allows to improve the perceived QoS in terms of PSNR for poor transmission channel conditions, when comparing to several existing schemes. The proposed strategy adapts the resources allocation to the (heterogeneous) radio link, with modulation, code rates parameters. It tends therefore to minimize retransmissions of data (i.e., Acknowledgement ReQuest (ARQ)).

The proposed resources allocation method is analyzed in this paper with a speech scalable coder (a MPEG-4 CELP encoder) and can be applied without loss of generality to other scalable data coders (audio or video for instance).

Future work will spot on comparing our method to optimal UEP scheme and on the impact of the transmission rate on the modulation choice when we derive our method, extending its design to the case of a more realistic propagation channel on the physical layer.

6. ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to Dr. Charly Poulliat for many useful discussions.

7. REFERENCES

[1] Q. Zhang, G. Wang, Z. Xiong, J. Zhou and W. Zhu, "Error Robust Scalable Audio Streaming over Wireless IP Networks," *IEEE Trans. on Multimedia*, vol. 6, no. 6, pp. 897-909, dec. 2004.

[2] M. M. Ghandi and M. Ghanbari, " Robust video transmission with an SNR scalable H.256 codec," Proceeding 7^{th} IEEE Int. Conf. on HSNMC, June/July 2004, pp. 932-940.

[3] C. Poulliat, D. Declercq, I. Fijalkow, "Optimization of LDPC codes for UEP channels and applications to scalable image transmission", in revision in *IEEE Trans. on Communications*, 2005.

[4] J. Hagenauer and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia", Proceeding of IEEE, vol.87, pp. 1764-1777, Oct. 199.

[5] D. Krishnaswamy and M. van der Schaar, "Adaptive Modulated Scalable Video Transmission over Wireless Networks with a Game Theoretic Approach," Proc. of IEEE MMSP. Workshop, sept. 2004.

[6] J. Hagenauer, "Rate compatible punctured convolutional (RCPC) codes and their applications", *IEEE Trans. on Communications*, Vol. 36, pp. 389-399, April 1988.

[7] A.S. Barbulescu and S. S. Pietrobon, *Rate Compatible Turbo Codes*, IEEE Electronic Letters, vol.31, no. 7, PP. 535 Mar. 1995.

[8] Z. H. Haas, "Design methodologies for adaptive and multimedia networks", Guest Editorial, *IEEE Communications Magazine*, Vol.39, No.11, pp. 106-107, november 2001. *IEEE Trans. on Communications*, Vol. 36, pp. 389-399, April 1988.

[9] F. Pereira and T. Ebrahimi, *The MPEG-4 book*, IMSC Press Multimedia Series, Prentice Hall PTR, 2002.