A Demonstration Platform for Network Aware Joint Optimization of Wireless Video Transmission

M. G. Martini, M. Mazzotti, C. Lamy-Bergot, P. Amon, G. Panza,
J. Huusko, J. Peltola, G. Jeney, G. Feher, S. X. Ng

Abstract— **A software platform for demonstration and performance evaluation of network aware joint optimization of wireless video transmission is presented in this paper. This demonstrator has been realized in the framework of the IST PHOENIX project (www.ist-phoenix.org) and is based on the joint system optimization model developed by the project. After a description of the considered cross-layer approach and of the information to be exchanged among the system component blocks, the techniques considered for information exchange between the system blocks and the concept of "JSCC/D controllers" are introduced. The implementation of the demonstrator realized is then described. After a description of the envisaged scenarios for the system taken into account for performance evaluation, example simulation results obtained with such demonstration platform, confirming the validity of the described joint optimisation approach, are shown.**

I. Introduction

THE EVOLUTION of wireless telecommunication systems is towards a more integrated and global system, meeting the requirements of both users and industrial world by offering convergence of different technologies and making efficient use of existing and emerging technologies. To help meet this goal, an efficient allocation of the user and system resources is a must, that need to be flexible enough to be adapted to a variety of technologies at every level of the transmission chain, for instance via a co-operative optimization of the communication system components in the different layers.

Our approach, following the already known joint source channel coding and decoding (JSCC/D) approach, aims at developing strategies where the source coding, channel coding and modulation parameters are jointly determined to yield the best end-to-end system performance. The developed architecture can be then utilized in different practical contexts, whether at the radio link level, with e.g., orthogonal frequency division multiplexing (OFDM) or wide band code division multiple access (WCDMA), with multiple input/multiple output (MIMO) and adaptive antennas or at application level, with e.g. MPEG-4 or H.264/AVC.

This approach goes against the so called "separation theorem" derived from Shannon's theory [1] and follows the recent trend of modern popular applications such as audio/video streaming, for which it was shown that separation does not necessarily lead to the less complex solution [2], nor

M. G. Martini, M. Mazzotti are with CNIT/University of Bologna, Italy. C. Lamy–Bergot is with Thales Communications, Colombes, France. G. Panza is with CEFRIEL/Politecnico Milano, Milan, Italy. J. Huusko and Johannes Peltola are with VTT Electronics, Oulu, Finland. P. Amon is with Siemens AG, Munich, Germany. G. Jeney and G. Feher are with Budapest University of Technology and Economics, Hungary. S. X. Ng is with University of Southampton, U.K.

is always applicable, in particular when transmitting data with real-time constraints or operating on sources whose encoded data bit error sensitivity varies significantly. Recently, JSCC/D techniques that include a co-ordination between source and channel encoders were investigated, improving both encoding and decoding processes while keeping the overall complexity at an acceptable level [3] [4] [5].

In this paper, a quality driven approach for wireless video transmission relying on the joint source and channel coding paradigm is proposed. In particular, the management of the information to be exchanged is addressed and the logical units responsible for the system optimization, in the following referred to as joint source channel coding and decoding (JSCC/D) controllers, having a key role in the system, are analyzed. The demonstrator of the described system realized in the framework of the IST PHOENIX project is also presented and simulation results are given.

II. System Architecture

Figure 1 illustrates the overall system architecture developed in the framework of the PHOENIX project, from the transmitter side in the upper part of the figure to the receiver side in the lower part, and including the signalisation used for transmitting the JSCC/D control information in the system. Beside the traditional tasks performed at application level (source encoding, application processing such as ciphering), at network level (including RTP/UDPLite/IPv6 packetisation, impact of IPv6 wired network and Robust header Compression (RoHC)), medium access (including enhanced mechanisms for WiFi and UMTS) and radio access (channel encoding, interleaving, modulation), the architecture includes two controller units at physical and application layers. Those controllers are used for supervising the different (de)coders, (de) modulation, (de)compression modules and to adapt said modules parameters to changing conditions, through the sharing of information about the source, network and channel conditions and user requirements. For the controlling purpose, a signalling mechanism has been defined, as detailed in the following.

A. Side information exchanged in the system and methods for exchanging it

In particular, the information that are taken into account by the system for optimization are source significance information (SSI), *i.e.* the information on the sensitivity of the source bitstream to channel errors; channel state information (CSI); decoder reliability information (DRI), *i.e.* soft values output by the channel decoder; source a-priori information (SRI), *e.g.* statistical information on the source
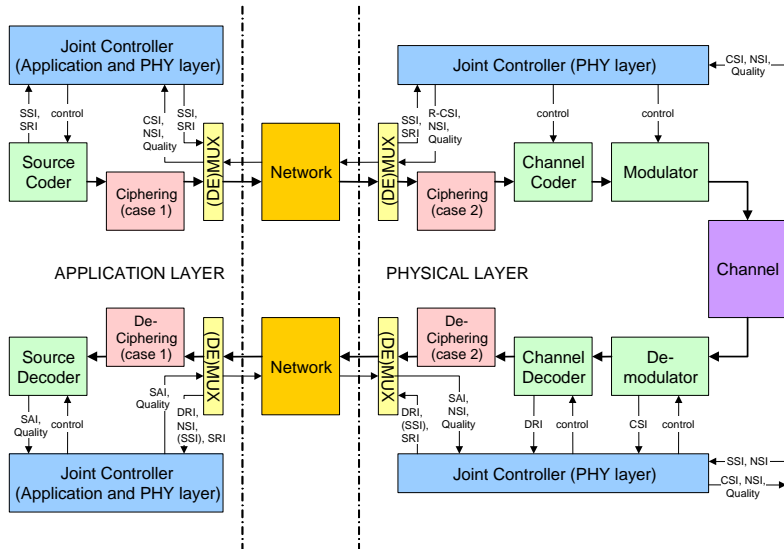
Fig. 1. PHOENIX system architecture for joint optimisation.

like importance of data elements; source a-posteriori information (SAI), *i.e.* information only available after source decoding; network state information (NSI), represented *e.g.* by packet loss rate and delay and finally the video quality measure, output from the source decoder and used as feedback information for system optimization. This last measure is critical as the target of the overall system optimisation is the maximisation of the received video quality. In practice, due to the fact this quality measure is regularly sent back to the transmitter side for action, the evaluation should be performed "on the fly" and without reference (or reduced reference) to the transmitted frame.

Naturally, when considering real systems, this control information needs to be transferred through the network and systems layers, in a timely and bandwidth efficient manner. The impact of the network and protocol layers is quite often neglected when studying joint source and channel coding and only minimal effort is made into finding solutions for providing efficient inter-layer signalling mechanisms for JSCC/D. Different mechanisms identified by the authors, that could allow information exchange transparently for the network layers (which we call *Network Transparency* concept), are summarized in Table I, together with the related overhead. Besides the possibilities listed in Table I, one should not forget that several transport protocols exist, each of which can carry the payload and also some control information. The Phoenix simulation platform supports *e.g.* UDP, UDPlite and datagram congestion control protocol (DCCP) protocols at transport layer level. The PHOENIX-framework supports also the cross-layer information exchange by introducing novel features for link layer and application layer. The PHOENIX IEEE 802.11 based medium access control layer (medium access control (MAC)) supports identification and transmission of CSI, NSI, DRI and SAI control messages. To support more realistic and sophisticated environments, there is the possibility to use the Universal Mobile Telecommunication System (UMTS) standard compliant simulator at the data link layer and the physical layer in the Phoenix framework. Finally, it should be noted that additional information is requested by the system for the set-up phase, where information on available options (*e.g.* available channel encoders and available channel coding rates, available modulators, ...) are exchanged, the session is negotiated and default parameters are set (*e.g.* authentication key, modules default settings). This is emulated in the demonstrator by use of the considered scenario (see Section II-C) precise definition.

B. Principle of JSCC/D controllers

The system controllers are represented by two distinct units, namely the "physical layer (PHY) controller" and the "application layer (APP) controller". The latter collects information from the network (NSI: packet loss rate, delay and delay jitter) and from the source (*e.g.* SSI) and has availability of reduced channel state information and of the quality metric of the previously decoded frame (or group of frames). According to this information, it produces controls for the source encoder block (*e.g.* quantization parameters, frame rate, error resilience tools to activate) and for the network. The physical layer (PHY) controller's task is to provide controls to the physical layer blocks, *i.e.* the channel encoder, modulator and interleaver. A more detailed description of the controllers is made in section III, where exemplifications of their behavior are also provided.

C. Envisaged scenarios

In order to emphasize the interest of the end-to-end optimization strategy developed by PHOENIX project for the transmission of multimedia data over an IP wireless link, different scenarios, in which this optimization would be in-

Control signal	Suitable Mechanism	Results
SSI	IPv6 hop-by-hop or solutions as in [6]	Overhead of few Kbits/s (depending on the source coding rate); high synchronization with the video data.
CSI	ICMPv6	Overhead of less than 10 Kbyte/sec for CSI update frequencies up to 10 ms; slight synchronization with the video data.
NSI	ICMPv6	Low overhead with suitable frequency of 100 ms (less than 1 Kbyte/sec).
DRI/SAI	IPv6 packets	Very high bandwidth consuming (even higher than the video data flow of a fixed multiplying factor). Probably better to send these control signals only when the wireless receiver is also the data traffic destination.
SRI	IPv6 hop-by-hop or Destination	Overhead of few Kbytes/sec (depending on the source coding rate) and high synchronization with the video data.

teresting for the end-user or the provider, have been identified. A first scenario consists of applications such as *video conference on the move*. In this case conversational mode, UMTS channel (or 4G), multicast, a mobile phone as the terminal device, no strict cost limitations, confidentiality, multi-user systems are the main characteristics. In the case of a stationary *video conference* (e.g. *from a cafe*) considered in the second scenario, the same parameters as in the first scenario are used, but a WLAN channel is taken into account instead of UMTS. Further scenarios have been identified, such as *video on demand* and *learning applications*, video calls, stationary or in the move, and *pushed video information* (e.g. live news). These scenarios are the reference sets that will be taken into account during the demonstration process and results derivation, in order to show the different optimization solutions and the advantages provided by the cross-layer design under analysis in the different application contexts.

III. THE DEMONSTRATION PLATFORM

In order to provide a realistic performance evaluation of the proposed approach, all the involved system layers blocks have been realistically implemented. Namely: application layer controller; source encoder/decoder (three possible codecs: MPEG-4, H.264/AVC and Scalable Video Coding in H.264/AVC Annex F), where soft-input source decoding is also allowed for H.264/AVC; cipher/decipher unit; content level unequal error protection (UEP) block, using rate compatible punctured convolutional (RCPC), alternative to UEP at PHY level, and leading to equal error protection (EEP) at PHY level when activated; real time transport protocol (RTP) header insertion/removal; transport protocol header (*e.g.* UDPLite, UDP, or DCCP) insertion/removal; IPv6 header insertion/removal; IPv6 mobility modelling; IPv6 network simulation; Robust Header Compression (RoHC); DLL header insertion/removal; Radio Link, including physical layer controller, channel encoder/decoder (convolutional, RCPC, low density parity check (LDPC) codes with soft and iterative decoding allowed), interleaver, modulator (OFDM, TCM, TTCM, STTC; soft and iterative demodulation allowed [11]) and channel. A more detailed description is provided in the following for the blocks whose impact is highlighted in the

results section, in particular for the controlling units.

A. Application layer (APP) controller

The application controller has been modelled as a finite state machine. At the beginning of each iteration cycle of the duration of one second (corresponding roughly to one or two group of pictures (GOP)), it decides the next operating state, which is defined by a fixed set of configuration parameters for the different blocks of the chain. According to these values, the application layer (APP) controller updates its configuration parameters. The choice of the new state is based on the history and feedback information coming from the blocks at the receiver side, relevant to the previous cycle.

The feedback information are:
• Quality: peak signal-to-noise ratio (PSNR) or other quality metric (e.g. based on structural distortion [7], or without reference to the original sequence [8]);
• Reduced CSI: average signal-to-noise ratio (SNR) in one controller step, channel coherence time;
• NSI: number of lost packets, average jitter, average round trip time (RTT).

The main configuration parameters set by the APP JSCC/D and modifiable at each simulation step are the video encoder frame rate, quantization parameters, and GOP size; Code rates $R_{c,i}$, where the index i is related to the ith sensitivity class, for content UEP if applied; average channel code rate, as a consequence of the choice of the source encoding parameters and the knowledge of the available bandwidth.

In order to reduce the dimension of the possible configurations, and to avoid switching continuously from one given set of parameters to another, which is not efficient in terms of compression for any source encoder, the demonstrator takes into account only a limited set of possibilities for these parameters. In particular the choice made considers frame rates of 30, 15 and 7.5 fps; spatial resolutions of QCIF, CIF; MPEG-4 quantization parameters (frame I, frame P) equal to (8,12) or (14,16); GOP length assuming values of 8, 15 and 30. Furthermore, some constraints on these values must be satisfied, thus the number of the controller states is reduced.
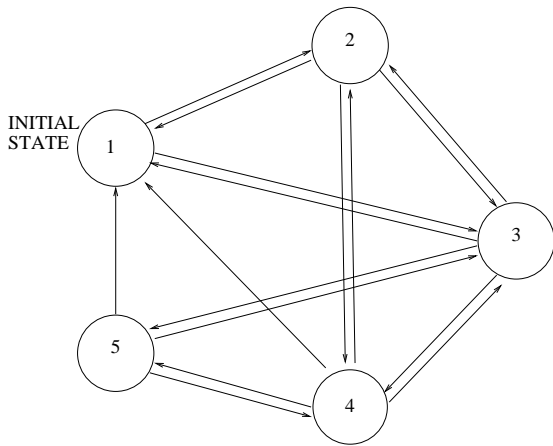
Fig. 2. A graphical representation of the APP-JSCC as a finite state machine with 5 states.

state	(qI, qP)	frame rate	GOP length
1	14,16	7.5	8
2	14,16	15	15
3a	14,16	30	30
3b	14,16	30	15
4	8,12	15	15
5a	8,12	30	30
5b	8,12	30	15

Typically, a low video quality value associated to a negative trend will cause a transition to a state characterized by a higher robustness. Given the bit-rate of the chosen state, the code rate available for signal protection is evaluated considering the total $R_{max} = R_s/R_c$ constraint, where R_s is the average source coding rate and R_c is the target average protection rate. That R_c target information is used either for embedded unequal error protection at the application level as in [9] or provided directly to the physical layer controller. If physical layer UEP is adopted, given the available total coded bit rate (R_{max}), the average channel coding rate (R_c) is derived by the application JSCC/D controller and proposed to the PHY controller. The knowledge of the bit-rate is of course approximated and based on rate/source parameters models developed by the authors or average values evaluated in previous controller steps.

As an example, for the first and second scenarios five different states have been chosen for the APP joint source and channel coding (JSCC) controller, characterized by different set of values for the above mentioned parameters. State 1 corresponds to the lowest source data rate (lowest video quality) and highest robustness, whereas state 5 corresponds to the highest source data rate (highest video quality) and lowest robustness. Thus, increasing the state number means increasing the robustness transmission at cost of a loss in the error free received video quality. Figure 2 depicts the finite state machine describing the APP–JSCC controller with the allowed transitions among states. More precisely, the number of possible parameter sets is seven, since state 3 and state 5 have two different options for the GOP length. The choice of the GOP length is made according to channel conditions: for average channel E_s/N_0 below a prefixed threshold the lower one is chosen, whereas for higher values of E_s/N_0 the higher one is preferred. The different sets of parameters are reported in Table II. The adaptive algorithm which has been tested takes into account the trend of the video quality feed-back from the source decoder. When there is a network congestion, indicated by an high value for the packet loss rate (PLR) feed-back in the NSI, the controller sets immediately the state to the first, characterized by the lowest source bit rate, in order to re-

duce as much as possible the amount of data which have to flow through the IPv6 network.

Additionally, the platform contains an alternative application controller for the H.264/SVC scalable video stream. To fully utilise the scalable video stream would require large amount of states to allow fine granular adjustments to the video stream, thus a different approach using fuzzy logic has been used. Fuzzy controller sets data rate and frame rate at the video sender by truncating the pre-encoded scalable video stream during the transmission.

B. Physical layer controller

The Physical layer controller's task is to decide, similarly as in [4], on the channel coding rate for each different sensitivity source layer, with the goal of minimizing the total distortion D_{S+C} with the constraint of the average channel coding rate R_c, provided by the application controller. Furthermore, the controller sets the parameters for bit-loading in multicarrier modulation, interleaver characteristics and performs a trade-off with receiver complexity.

C. Header Compression

In order to limit the potentially dramatic impact of errors on RTP/UDPLite/IPv6 headers over the wireless link, a header compression module based on IETF robust header compression (RoHC) recommendations has been introduced in the transmission chain. Located near the wireless part immediately after IPv6 wired network and mobility modelling modules, this module provides the compressed network header concatenated with the unchanged video payload to the Data link module. In practice, the uncompressed RTP/UDPLite/IPv6 header of size 60 bytes can be reduced in average to a value of 5 bytes only. This header size reduction implies a large increase in terms of robustness to errors: as an example, for a uniform bit error rate (BER) of 10^{-3} after radio link, the probability of error in the IP header drops from 44% to only 5%, which allows the authors to justify the hypothesis that network headers, when compressed, are small by comparison with the data payload.

IV. SIMULATION RESULTS

Example results obtained with the described demonstration platform are reported in Fig. 3. It shows a performance comparison between an example of video transmission adopting "classical" techniques and one exploiting the

proposed joint-adaptive strategies, both at application and at physical layer, as described in section III. The comparison is made in terms of subsequent video quality metric values, each obtained through the average over 1s. The value obtained have been normalised, in order to allow a comparison between the different metrics. Beside the PSNR the metrics considered are the structure similarity metric [7] and a no-reference metric originally developed for JPEG images [8]. The scenario considered is a WLAN supporting, at the radio link level, a coded bitrate of 12 Mbit/s. The video stream is coded according to the MPEG-4 standard and is supposed to be multiplexed with other real-time transmissions so that it occupies only an average portion of the available bandwidth, corresponding to a coded bitrate of 1 Mbit/s. A CIF resolution have been selected. In the non-adapted system, no RoHC techniques have been adopted, while in the adapted case the RoHC allows to reduce the overhead due to the network header transmission and a more effective protection can be exploited for video data. The channel codes are IRA-LDPCC with a "mother" code rate of (3500,10500), properly punctured and shortened in order to obtain different code rates. The resulting codewords are 4200 bits long. The code rate is fixed to 1/2 for the non adapted system, meaning that an EEP policy is adopted, while in the adapted case the code rate can change according to the SSI in order to perform UEP. Obviously, to allow a direct comparison between the two schemes, the average coded bitrate has been chosen to be the same with both the transmissions. In the first case, the modulation is a "classical" orthogonal frequency division multiplexing (OFDM) with 48 carriers for data transmission and a frame duration of 4 μs, whereas the PHY-JSCC implements also margin adaptive bit-loading techniques in the adapted system. The simulated channel is obtained according to the ETSI standard for channel A and it takes into account also a log-normal flat fading component with channel coherence time 5s, to consider the fading effects due to large obstacles. The figure is related to $E_b/N_0 = 9.2dB$, where E_b is the energy per coded bit. We may observe that the adapted system provides an evident improvement, with any considered metric. In particular, we observed an average gain of 3dB in terms of PSNR in the conditions under analysis.

V. Conclusions

A global approach for realistic network-aware joint source and channel system optimization has been outlined in the paper. The information to be exchanged among the system blocks are described, together with the techniques proposed to make them available by the relevant system blocks. After a short introduction on the role of the system "controllers" and of the scenarios envisaged for the performance evaluation of the said scheme, the demonstration platform realized in the framework of the PHOENIX project is described. Simulations for the considered scenarios show the gain obtained with the demonstrator when the adaptation options are activated, although more realistic assumptions than those commonly considered in literature, in terms of redundancy due to side information transmission, are made. The results could thus be seen as a proof of the real appli-
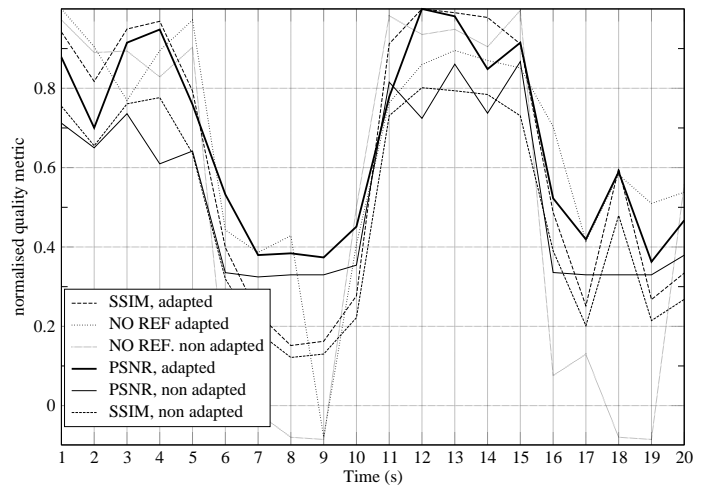

Fig. 3. Received video quality versus time with the JSCC adapted and non adapted system.

cability of joint source and channel coding techniques and of the gain achievable through cross-layer design.

VI. Acknowledgement

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, July-Oct. 1948.
[2] J. L. Massey, "Joint source and channel coding," in *Commun. Systems and Random Process Theory, NATO Advanced Studies Institutes Series E25*, J.K. Skwirzynski editor, pp. 279-293, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1978.
[3] J. Hagenauer and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia," *Proceedings of the IEEE*, vol. 87, no. 10, Oct. 1999.
[4] M.G. Martini, M. Chiani, "Rate-Distortion models for Unequal Error Protection for Wireless Video Transmission," *IEEE VTC 2004 Conference*, Milan, Italy, May 2004.
[5] L. Perros-Meilhac and C. Lamy, "Huffman tree based metric derivation for a low-complexity sequential soft VLC decoding," in *Proceedings of IEEE ICC'02*, New York, USA, vol. 2, pp. 783–787, April-May 2002.
[6] M. G. Martini, M. Chiani, "Proportional Unequal Error Protection for MPEG–4 video transmission", *Proc. IEEE ICC 2001*, Helsinki, June 2001.
[7] Z. Wang, L. Lu, A. C. Bovik, "Video quality assessment based on structural distortion measurement", *Signal processing: Image communication*, vol. 29, no. 1, Jan. 2004.
[8] Zhou Wang, Hamid R. Sheikh and Alan C. Bovik, "No-Reference Perceptual Quality Assessment of JPEG Compressed Images", *Proc. IEEE International Conference on Image Processing, pp. 477-480, Rochester, New York, September 22-25, 2002.*
[9] C. Lamy-Bergot, N. Chautru and C. Bergeron, "Unequal error protection for H.263+ bitstreams over a wireless IP network," *to appear in IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'06)*, Toulouse, France, May 2006.
[10] A. Zsiros, A. Flp, G. Jeney, "Easily Configurable Environment of UTRAN Physical Layer" *5th Eurasip Conference*, Szomolny, Slovakia, July 2005.
[11] L. Hanzo, S. X. Ng, W. Webb and T.Keller, "Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems", Second Edition. New York, USA : John Wiley and Sons, 2004.