MULTIRATE DELIVERY OF SCALABLE VIDEO WITH PROGRESSIVE NETWORK CODES

Michele Sanna, and Ebroul Izquierdo

School of Electronic Engineering and Computer Science, Queen Mary, University of London
Mile End Road, E1 4NS, London, UK
email: { michele.sanna, ebroul.izquierdo}@eecs.qmul.ac.uk

ABSTRACT

The future scenario in video content distribution will rely upon large interconnected systems. Multiple platforms need differentiated services. Scalable Video Coding is the upcoming standard solution for decoding multiple versions of the video from the same bitstream. We propose Network Coding to delivery seamlessly different versions of the video to users with different requirements on the same network. Network coding increases the network rate and provides error control at network level, sensibly improving the overall quality of transmission.

1. INTRODUCTION

Universal systems for video streaming are meant to provide a centralized source of multimedia content to be delivered to different classes of users. On the one hand, systems like Digital Video Broadcasting (DVB) differentiate the physical interface between terrestrial (DVB-T), satellite (DVB-S) and cable (DVB-C) systems. On the other hand, the backbone interfacing the final physical medium with the central source carries a video streaming signal for these platforms with deeply different display requirements. IP-TV systems for instance, rely on a Content Delivery Network (CDN) in which the users meet different rates and reception conditions. Different services might be delivered depending on the class of service paid by the user.

Classic video codecs are designed to work under definite display characteristics and channel rates. To differentiate the service, the source needs to encode different versions of the same video, with a great employment of resources. Scalable Video Coding (SVC) is a novel video coding paradigm that exploits the variety of the transmission channels and user requirements [1]. Decomposition of the signal provides a layered structure of the data. Partial decoding is performed by dropping unwanted parts of the bistream, allowing adaptation of the video to the channel and to the user's requirements.

Network Coding (NC) was introduced by Ahlswede *et al.* in 2000 as a novel technology enabling network transmission with increased rate, up to the theoretical network capacity [2]. It allows intermediate nodes to retransmit to the other nodes a function of the received information. NC overpasses the traditional forwarding at intermediate nodes, which can be regarded as a special case of network coding. Consider the butterfly network in Fig. 1. With conventional networking (Fig. 1,a) the central node forwards one of the incoming packets at a time, thus serving both messages *a* and *b* in turn to one of the two receivers. With network coding (Fig. 1,b) the central node can transmit a combination (e.g., bit-wise XOR) of the input messages, thus serving both receivers with both messages in the same transmission slot.

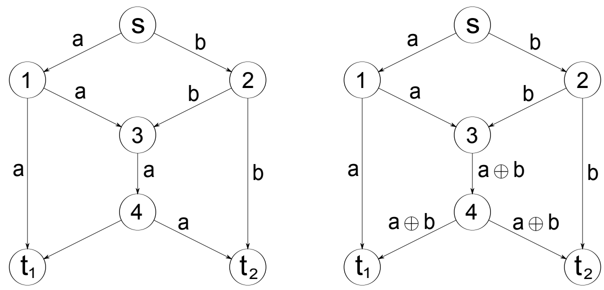

Figure 1: The two-source two-sink butterfly network without (a) and with (b) network coding.

Network Coding boosts the usage of network resources and increases the potential quality of an SVC streaming system. Most of the literature in SVC transmission via NC deals with peer-to-peer dissemination [3]. Multicast streaming of SVC in a client-server fashion would benefit from a proper design of the network code that exploits the hierarchy of the video layers. We propose the use of the algebraic model of network coding transmission [4] to design network codes for transmission of differentiated information flows.

Progressive codes were studied for priority transmission [5]. A subspace of the source coding space is delivered by partial decoding of independent codeblocks partitions. We use progressive codes to allow differentiated services to reach the various receivers with a single transmission from the source. We want to spread the coded video throughout the network and ensure that the base layer is received by all terminals and the enhancement layers are received depending on the availability of network rate.

Errors and losses are amplified by network coding transmission due to the coded nature of the information. Under severe channel conditions, the video quality rapidly falls down. By assuming the role of a coding operator, network coding allows error control against link failures at network level. Network Error Correction (NEC) allows coding redundancy in the spatial dimension, boosting the resistance against channel impairments with respect to traditional Forward Error Correction (FEC) only [6].

We propose a combination of multirate network coding and NEC to deliver differentiated video to heterogeneous receivers with Unequal Error Protection (UEP) of the video layers. We analyze a case study of multicast network where these factors can be easily combined to provide superior video quality to all receivers with respect to traditional transmission with packet forwarding. We make use of the wavelet-based aceMedia Scalable Video Codec (aceSVC) which has demonstrated performance comparable to the H.264/SVC standard [7]. Our method needs knowledge of the network code at the source, but it does not need a particular design of the network code, thus any generic network code can be used.

The paper is organized as follows: Section 2 recaps the Scalable Video Coding paradigm and the wavelet-based aceSVC codec. Section 3 introduces the algebraic approach to network coding and the multirate transmission. Section 4 presents simulation results on a case study and Section 5 concludes the paper.

2. WAVELET-BASED SCALABLE VIDEO CODING

Scalable Video Coding (SVC) copes with the need of delivering video with differentiated rates and display requirements [1]. A scalable approach to video coding performs a single coding operation and supports multiple decoding configurations from the same embedded bitstream. Unlike conventional coders, that would need several transmissions for a large variety of receivers, SVC avoids overloading the network and decreases the coding computational load. In a CDN dedicated to distributed video, edge routers serve the respective groups of users by receiving and forwarding the video stream adapted to the respective platforms. The communication in the backbone is to be optimized to deliver the different versions of the video.

We make use of the aceSVC wavelet-based codec which has scalability properties in terms of frame rate, spatial resolution and quality (SNR) [7]. The aceSVC codec uses a combination of Motion Compensated Temporal Filtering (MCTF) for temporal scalability, discrete wavelet transform (DWT) for temporal and spatial scalability and Embedded ZeroBlock Coding (EZBC) of textures for quality scalability. The decomposition is organized in layers, starting from the basic display configuration, where each enhamcement layer increases the frame-rate, resolution or quality (T/S/Q). For instance, a three layer spatial decomposition of a 4CIF (704×576) sequence can be: A base layer, which is QCIF (176×144), a first enhancement layer with CIF resolution (352×288), and a second final layer with full resolution 4CIF.

The coded data is organized in elementary units, called atoms. Each atom contains at least one enhancement level of quality for a determined T/S, or vice versa, for a single GOP, as shown in Fig. 2. The atoms are organized in an embedded bitstream. The extractor operates by reading a subset of atoms from the base layer (0/0/0) up to the desired T/S/Q to reduce instantly the video stream from the full quality to that configuration.

2.1 Channel Coding for Scalable Data

The distribution of the errors in the video layers can be critical. SVC streaming performance are boosted by a thoughtful allocation of resources between source and channel coding to exploit the limited channel resources. In Layered Coding (LC) the hierarchy of the data implies that base and lower layers are more critical for successful decoding, both in terms of visual degradation and because the video is not decodable without the base layer. Higher layers have less impact on the visual quality than the base layers and the motion vectors, and can also be dropped if needed [8].

Error protection performance of a scalable data stream is boosted by means of Unequal Error Protection (UEP).

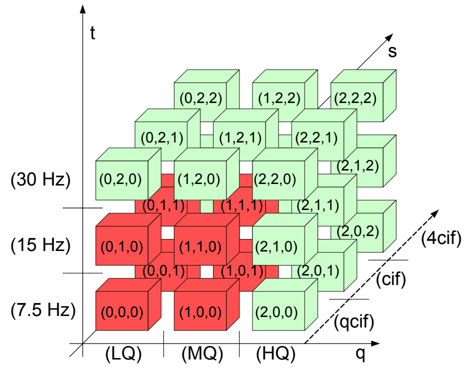

Figure 2: Representation of atoms of an SVC stream with 3 levels of temporal, spatial and quality scalability [7].

UEP increases the video quality under adverse channel conditions by differentiating the redundancy bits for the protection of different portions of data. Typical techniques are the independent coding of parallel streams with different FEC codes [9].

In multirate streaming, the video is splitted to multiple streams, which are sent independently and to different receivers. In the next section we explain how to separate the coding spaces of network coding to perform UEP at network level.

3. NETWORK CODING FOR MULTIRATE VIDEO DISTRIBUTION

The major impact of network coding in networking is that the transmission rate in multicast scenarios achieves the *max–flow* rate of the network which is unreachable with traditional packet forwarding [2]. The algebraic model of network coding is used here to design the routing at the source so that the correct subspaces of the source coding space are spanned at the receivers. This achieves differentiation and scalability of the data received by the sink nodes.

3.1 Algebraic transmission model

A network can be modeled as a directed graph $\mathscr{G} = (V, E)$, where V is a set of vertices and E is a set of edges. A vertex is designated as the source node s and the subset $T = [t_1, t_2, \ldots, t_{|T|}]$ of V includes the sink nodes. Every edge has unitary capacity for coding purposes. The actual capacity of the links is modeled with multiple parallel edges. In linear network coding the message sent at the output edges of each node are linear combinations of the messages at the input edges [10]. Given a non-source node i with input edges $d \in In(e)$ and output nodes $e \in Out(e)$ we refer to i as $head(e)$ and $tail(d)$. Being U_d the input messages in $d \in In(e)$, the message U_e transmitted on e is:

$$U_e = \sum_{d \in In(e)} \beta_{d,e} U_d \tag{1}$$

The symbols and coefficients belong to a Galois Field of size $q = 2^m$, where m is the number of bits of the bitstream corresponding to one symbol.

An algebraic approach to model the manipulation of data in NC has been proposed in [4]. By assuming a synchronized transmission at link level, on a higher level the transfer characteristic to each receiver can be resumed in a system matrix M_t, to be used as follows. The network codewords at the source are vectors $\mathbf{x} = [x_1, x_2, \ldots, x_h] \in \mathscr{C}$, with $x_i \in GF(q)$. The codeword received by sink t is a vector $\mathbf{y}_t = [y_1, y_2, \ldots, y_{n_t}]$ obtained by the network transformation as:

$$\mathbf{y}_t = \mathbf{x}A(1-F)^T B_t = \mathbf{x}M_t, \tag{2}$$

where the involved matrices are defined as follows. The adjacency matrix F contains the local coding kernels and is defined as an $|E| \times |E|$ matrix:

$$F_{e,d} = \begin{cases} \beta_{d,e} & \text{if } tail(e) = head(d) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

An $h \times |E|$ matrix A is defined for the routing of the h RV processes at the source node s:

$$A_{i,d} = \begin{cases} \alpha_{i,e} & \text{if } tail(e) = s \\ 0 & \text{otherwise} \end{cases}, i = 1, 2, \ldots, h, \tag{4}$$

and a set of $|E| \times n_t$ matrices $B_t, t \in T$ for the routing of the received messages to the n_t output symbols at receivers $t \in T$:

$$B_{t;e,j} = \begin{cases} \varepsilon^{(t)}_{e,j} & \text{if } head(e) = t \\ 0 & \text{otherwise} \end{cases}, j = 1, 2, \ldots, n_t, \quad t \in T. \tag{5}$$

The network can be seen as a coding operator that projects a source coding space $\Omega_{h,q}$ with length h into the receiver spaces, by the system matrices M_t. Such projected space has length n_t (number of input edges to the sink) and is a subspace of $\Omega_{h,q}$, with $r_t \le n_t$ dimensions (where r_t is the number of paths from source to sink t.

The transmission can be considered error free on a higher level by assuming that local codes are applied on node-to-node connections to provide link-level error control, at the expenses of a reduced rate for source coding. Additionally, with NEC error control can be exploited at the network coding level. Channel impairments are considered as a random addition of the symbol traveling on an edge:

$$\mathbf{y} = (\mathbf{x}A + \mathbf{z})(1-F)^T B_t = \mathbf{x}M_t + \mathbf{z}F_t, \tag{6}$$

where $\mathbf{z} \in GF(q)^{|E|}$ is the error vector.

If the source codebook has rate $\omega < h$, the source codewords have minimum distance $d_{min} = h - \omega + 1$ from one another. Error control functionalities are available at the receivers with rate $\omega < r_t < n_t$. A linear network code can be regarded as a $(\omega, r_t, d_{t,min})$, $t \in T$, where r_t is the rate and $d_{t,min}$ is the minimum coding distance at the receiver t. The receivers are able to correct up to $\lfloor \frac{d_{min,t}-1}{2} \rfloor$ link errors and detect up to $d_{min,t} - 1$ errors.

Randomized routines for code construction have high probability of keeping the minimum coding distance among the codewords at the receiver. This probability grows as the field size increases. Since the design of multirate codes assumes a network with infrequent changes, deterministic construction algorithms could be used as well to calculate a Minimum Distance Separable (MDS) network code.

3.2 Progressive codes for multirate information flow

At the source, the video layers are reduced to parallel streams, eventually encoded with error-control codes and mapped onto the symbols in the corresponding positions of the source codeword. Since the video layers are hierarchically organized, sets of lower video layers can can be mapped onto subspaces of $\Omega_{h,q}$. Consider subspaces $V_{\omega_i,q} \subset \Omega_{h,q}$ with dimension ω_i. Various sink nodes receive different flows from the source (r_t). Such rate ω_i is ensured to each sink if r_t paths, with $\omega_i \le r_t \le n_t$, correctly span $V_{\omega_i,q}$.

We assume that all receivers should receive the symbols spanning the space $V_{\omega_0,q}$ in order to decode the base layer. Any other enhancement layers can be sent by extending the coding space with the additional paths to the receivers with higher rate. Enhancement layers are transmitted in parallel along additional dimensions thus providing some receivers with an enhanced data stream and without jeopardizing the reception of the base space by the others.

The idea is to have a transfer characteristic by stacked coding as a result of a proper design of the matrix A, which routes the h codeword symbols (and thus the source streams) into the paths to the receivers. In order to successfully decode a proper subset of source symbols, progressive coding makes sure that the messages received by the sink nodes are linear combinations of only the messages intended for them. The receivers decode a subspace from a set $C^{(t)}_i$ of $\omega_i + d_{t,min} - 1$ network symbols, independent from the symbols of the rest of the coding space. This means that if

$$\begin{aligned} \mathbf{y}_t(C^{(t)}_i) &= [x_1, \ldots, x_{h_i}, x_{h_{i+1}} \ldots, x_h] * M_t = \\ &= [x_1, \ldots, x_{h_i}, 0, \ldots, 0] * M_t, \end{aligned} \tag{7}$$

then decoding of $V_{\omega_i,q}$ at the receiver $t \in T$ is independent from the decoding of $V_{\omega_j,q}, j \ge i$.

When the topology allows, the routing matrix at the source can be designed to allow progressive coding. Elementary row operations, such as Gaussian elimination, are used to reduce the matrix

$$M = [M_1, M_2, \ldots, M_{|T|}] \tag{8}$$

to a row-echelon form. Such a matrix, when reduced to row-echelon form, has the most possible zero elements in the lower left area. For example, a case with three receivers, with rates $(2,3,4)$, if the network allows, can be reduced to the following row-echelon form transfer matrix.

$$M = \left[\begin{array}{cc||ccc||cccc} \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet \end{array}\right], \tag{9}$$

where $\bullet$ stands for non-zero element. The first receiver receives a full rank system of linear equations whose variables are a subset of two out of the four source symbols, so it is able to decode the base layers transmitted on the first two paths. The second receiver decodes two source symbols, which carries the data of another video layer, and the third receiver can detect the whole codeword.

A subspace $V_{\omega_i,q}$ with $d^{(i)}_{t,min} \ge d_{t,min}$, where $d_{t,min}$ is the minimum distance of $\Omega_{h,q}$, has additional error protection capabilities. A non-regular distance coding space yields to unequal error protection by differentiating network error correction among the coding space partitions.

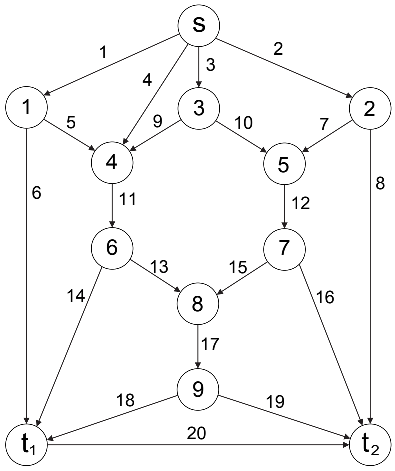

Figure 3: Network with multirate flows to the receivers.

3.3 Video layers organization

Layers are grouped into non overlapping sets and organized in independent streams, such as:

$$\begin{aligned}&Stream_0 = \{MV, Q_0\},\\&Stream_1 = \{Q_1, Q_2, \ldots, Q_{n_1}\},\\&Stream_2 = \{Q_{n_1+1}, \ldots, Q_{n_2}\},\\&\ldots\\&Stream_i = \{Q_{n_{i-1}+1}, \ldots, Q\},\end{aligned}$$

and with source rates (without error-control coding):

$$R_{stream_i} = \sum_{i=n_{i-1}+1}^{n_i} R_i. \tag{10}$$

Due to network-constrained rate, for channel coding each stream is assigned a coding rate proportional to the source coding rate and the dimension of the respective coding subspace:

$$R_{stream_1}/r_{FEC}^{(1)} * \omega_1 = R_{stream_2}/r_{FEC}^{(2)} * \omega_2 = \ldots \tag{11}$$

4. SIMULATION RESULTS

We show now a case study of transmission of the scalable video on a network with a single source node, multiple receivers with different rates and intermediate nodes with network coding capability. Our exemplary network and testbed is shown in Fig. 3.

We encode a video sequence with the wavelet-based aceSVC codec with quality scalability into 4 layers, with the objective of transmitting the base layer to the first receiver and all 4 layers to the second one. We apply a deterministic network code construction algorithm, which ensures the minimum coding distance at all receivers. The network transfer characteristics, after proper design of the routing matrix A as explained in section 3.2, has a shape like:

$$M_1 = \begin{pmatrix} \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet \\ 0 & 0 & \bullet \\ 0 & 0 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & 0 & 0 & 0 \end{pmatrix}, \tag{12}$$

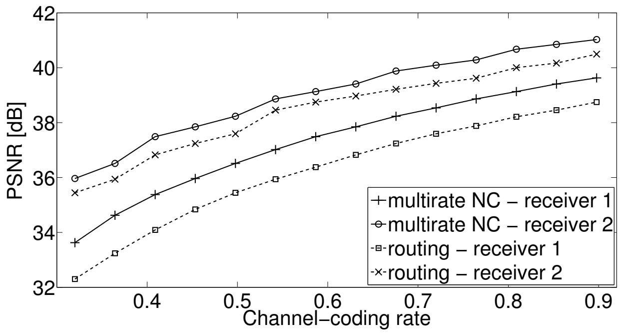

Figure 4: Peak Signal-to-Noise ratio (PSNR) of city sequence at CIF resolution at the two receivers against the information rate, by assuming a channel rate of 300 kbps on each link.

where we can see that the linear system at the first receiver depends only on the first three variables. For comparison purposes we assume that unresolved bottlenecks reduce the rates from $r_1 = 3$ and $r_2 = 4$ to $r_1 = 2.5$ and $r_2 = 3.5$ with routed transmission.

At first we assume that error control is conducted at link level. Error-protection codes are applied locally at each link, so that an error-free transmission is seen by the network coding layer, up to a sustainable channel error rate. The full rate of the network is exploited and no NEC is applied, leaving the error control at transport level. We allocate the source video bitrate accordingly with the available information rate. The visual quality at the two receivers is shown in Fig. 4 by means of the Peak Signal-to-Noise Ratio (PSNR) between the transmitted and the received video. The curves represent maximum quality achievable in error-free transmission and its decrease as the channel coding rate increases. The increased rate allows coding at the source with higher quality.

We consider the impact of errors on the visual quality. Errors have a strong impact in network coding transmission, due to the propagation between the received streams. We test a combination of multirate transmission and Unequal Error Protection (UEP) at network level (with NEC). The throughput rates are $r_1 = 3$ and $r_2 = 4$. We use two partitions of the coding space. The base stream is mapped in a $(\omega_0, r_1^{(0)}, d_{1,min}^{(0)}) = (1,3,3)$ code which delivers three symbols to both receivers from a single source symbol. This allows correcting, for each codeword, error dominated by a pattern with one link error (for the definition of dominant error pattern refer to [11]). Layers from 1 to 3 are transmitted to sink t_2 with a $(1,1,1)$ code ($d_{2,min}^{(1)} = 1$) and delivered only to the second sink node thanks to the extra coding path. We use irregular LDPC codes for FEC at transport level.

We compare network-coding transmission with traditional routing, which, due to reduced throughput allows LDPC coding at transport level with reduced protection. The video layer $q0$ is sent in a multicast session (with coding rate $3/5$ in this example as opposed to the coding rate of $1/3$ of NEC), whereas $q1$ to $q3$ are sent separately only to the second receiver. The simulation results are shown in Fig. 5 for the first and second receiver. Video quality is compared by means of PSNR. We can see the higher quality received by the second sink, thanks to the increased bitrate and the extra

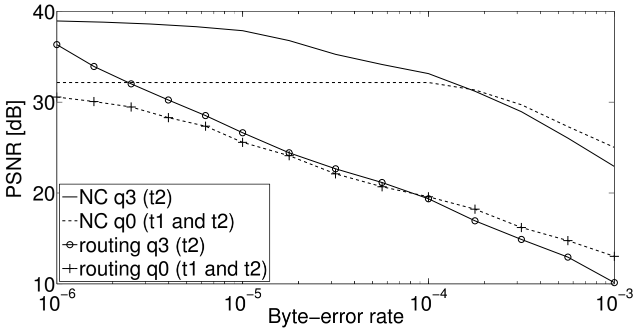

Figure 5: PSNR of video received at nodes t_1 and t_2. Network code is $(\omega_0, r_t^{(0)}, d_{t,min}^{(0)}) = (1,3,3), t \in T$, $(\omega_1, r_2^{(1)}, d_{2,min}^{(1)}) = (1,1,1)$. Routing rate is 2.5 for t_1 and 3.5 for t_2. Sequence: City, CIF resolution, 30 fps, rates $(q0, q1, q2, q3) = (384, 480, 576, 672)$ kbps. Base field size $q = 16$

video layers. It can be noticed the effect of quality degradation typical of scalable data with unequal error protection. After a certain break-point the higher layers, which have less error protection, are deeply affected by visual errors. Without a mechanism of error concealment (i.e., dropping this stream and decoding only the lower layers) the visual quality drops. Lower layers also decrease the quality for higher error rates due to the stronger error protection rate.

Protecting the base stream with a strong network error correction code keeps a constant quality of the layer $q0$ thanks to the possibility of correcting link errors for each codeword. The quality contribution of layer $q3$ to the reception at t_2 is comparable in the two cases of routing and network coding. They both decrease as the error rate grows. The difference resides in the quality of the base layer, which is kept high by the network error correction code. Application of NEC admits low rates $(1/3$ in our case) but outperforms any other form of error protection. NEC decouples the incisiveness of errors in adjacent time slots, so that, even with shorter block length, the probability that an LDPC block is affected by an undecodable number of errors, decreased. This is shown in Fig. 5, where below a certain error rate the reception of the base layer is not affected at all.

Network error correction increases robustness at the expense of reduced information rates. As shown by the experiments, strong protection of lower layers brings more benefit than a higher rate and video quality at the source under severe channel conditions. The efficiency is sensibly boosted as the network and the number of receivers increase, but this might lead to a difficult designing of the source routing matrix to deliver the respective subspaces. On the other hand, the design of the routing matrix can be done entirely at the source, which needs to know only the chosen network code. There is no need to design a particular network code, because any generic code, calculated randomly or deterministically (to keep the coding distance), can be used.

5. CONCLUSION

We proposed a method for designing network codes distribution of multirate scalable video on a network capable of performing coding at intermediate nodes. We showed how to use the algebraic model of network coding to deliver seamlessly multirate video to receivers with different demands. We also showed a case study in which unequal network error correction achieves superior robustness against channel errors and leads to better visual quality than the case with routing and LDPC codes.

Acknowledgments

This research was partially supported by the European Commission under contract FP7-247688 3DLife and FP7-248474 SARACEN

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the h.264/avc standard,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, no. 9, pp. 1103–1120, 2007.

[2] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, “Network information flow,” *Information Theory, IEEE Transactions on*, vol. 46, pp. 1204 –1216, jul 2000.

[3] E. Kurdoglu, N. Thomos, and P. Frossard, “Scalable Video Dissemination with Prioritized Network Coding,” in *International Conference on Multimedia & Expo, StreamComm Workshop, ICME’11*, 2011.

[4] R. Koetter and M. Médard, “An algebraic approach to network coding,” *Networking, IEEE/ACM Transactions on*, vol. 11, pp. 782 – 795, oct. 2003.

[5] Y. Lin, B. Li, and B. Liang, “Differentiated data persistence with priority random linear codes,” in *Distributed Computing Systems, 2007. ICDCS ’07. 27th International Conference on*, p. 47, june 2007.

[6] R. W. Yeung and N. Cai, “Network error correction. I: Basic concepts and upper bounds, and II: Lower bounds.,” *Commun. Inf. Syst.*, vol. 6, no. 1, pp. 19–54, 2006.

[7] N. Sprljan, M. Mrak, T. Zgaljic, and E. Izquierdo, “Software proposal for wavelet video coding exploration group,” tech. rep., M12941, 75th MPEG Meeting, ISO/IEC JTC1/SC29 /WG11/MPEG2005, Bangkok, Thailand, January 2006.

[8] N. Ramzan, S. Wan, and E. Izquierdo, “Joint source-channel coding for wavelet-based scalable video transmission using an adaptive turbo code,” *J. Image Video Process.*, vol. 2007, January 2007.

[9] M. van der Schaar and H. Radha, “Unequal packet loss resilience for fine-granular-scalability video,” *Multimedia, IEEE Transactions on*, vol. 3, pp. 381 –394, dec 2001.

[10] S.-Y. Li, R. Yeung, and N. Cai, “Linear network coding,” *Information Theory, IEEE Transactions on*, vol. 49, pp. 371 –381, feb. 2003.

[11] Z. Zhang, “Theory and applications of network error correction coding,” *Proceedings of the IEEE*, vol. 99, pp. 406 –420, march 2011.