On the Combination of Multi-Layer Source Coding and Network Coding for Wireless Networks

Krigslund, Jeppe; Fitzek, Frank; Pedersen, Morten Videbæk

Published in:
Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2013 IEEE 18th International Workshop on

DOI (link to publication from Publisher):
10.1109/CAMAD.2013.6708078

Publication date:
2013

Document Version
Early version, also known as pre-print

Link to publication from Aalborg University

Citation for published version (APA):
Krigslund, J., Fitzek, F., & Pedersen, M. V. (2013). On the Combination of Multi-Layer Source Coding and Network Coding for Wireless Networks. In *Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2013 IEEE 18th International Workshop on* (pp. 1 - 6). IEEE (Institute of Electrical and Electronics Engineers). https://doi.org/10.1109/CAMAD.2013.6708078

On the Combination of Multi–Layer Source Coding and Network Coding for Wireless Networks

Jeppe Krigslund
Department of Electronic Systems
Aalborg University, Aalborg, Denmark
acticom GmbH, Berlin, Germany
jepkri@es.aau.dk

Frank Fitzek
Department of Electronic Systems
Aalborg University, Aalborg, Denmark
acticom GmbH, Berlin, Germany
ff@es.aau.dk

Morten V. Pedersen
Department of Electronic Systems
Aalborg University, Aalborg, Denmark
Steinwurf ApS, Aalborg, Denmark
mvp@es.aau.dk

Abstract—**This paper introduces a mutually beneficial interplay between network coding and scalable video source coding in order to propose an energy-efficient video streaming approach accommodating multiple heterogeneous receivers, for which current solutions are either inefficient or insufficient. State of the art in media streaming typically acknowledges network and source coding being complementary. Despite this, these research topics are treated separately. By cultivating advantageous behaviour and features in the network and source coding structures a correlation between coding complexity and channel quality is developed. A linear coding structure designed to gracefully encapsulate layered source coding provides both low complexity of the utilised linear coding while enabling robust erasure correction in the form of fountain coding capabilities. The proposed linear coding structure advocates efficient support of multi-resolution video streaming.**

I. Introduction

Advancements of the communication technology within consumer electronics have enabled a vast diversity of connected devices with various capabilities. These types of devices span from wearable electronics such as wristwatches and glasses over vehicular equipment to more familiar devices like phones, tablets, laptops and televisions. All providing means for wireless connectivity. Industrial equipment has experienced similar transformation. To accommodate for this variety of devices, content and service providers must utilise adaptive and scalable solutions fitting a wide range of client devices. Such solutions have been subject to intense research recently. Especially for distribution of media content, as video data allocates over half of the Internet traffic today.

A popular approach for video provisioning in the Internet is transcoding (such as the standardised MPEG DASH), where a video stream is provided in several different qualities. All streams are individually encoded, and thus also individually decodable. Such approach is beneficial for each individual receiver, as the desired video quality is efficiently compressed and thus requires a minimum of network activity to receive. However, as a unique data stream is transmitted for each individual receiver, the required bandwidth is directly proportional to the amount of simultaneous receivers. Furthermore, the service provider requires storage or live-encoding of multiple video streams of increasing quality. Alternatively, multicast transmissions can be utilised in order to increase the network friendliness of video streaming to multiple receivers. This however introduces two problems with respect to both reliability and accommodation of video quality for all receivers: (1) multicast channels are inherently unreliable due to lack of feedback, thus the video stream will be erroneous. (2) Each provided video quality is to be transmitted individually, thus requiring excessive bandwidth.

Prior research have proposed multiple description coding [1] as a solution, where multiple low quality streams (descriptions) can be combined into one higher quality stream. While these capabilities are promising in a scalable video scenario, the current support from the video coding community is at a strictly experimental level. A layered coding approach provides similar scalability within a single video stream, where each layer corresponds to an increasing video quality. Furthermore, each added layer only introduces a small overhead [2]. Thus, a multicast video stream using layered coding may support heterogeneous devices while preserving network friendliness. However, due to multicast communication being inherently unreliable, forward error correction is needed to accommodate for erasures. Utilising a block code (such as Reed-Solomon or Random Linear Coding) is common practice for such erasure correction. Such codes induces an all-or-nothing behaviour, negating the benefits of the layered structure and the error concealment features that may be embedded in video coding. Protecting each layer individually may enable partial decoding of the layers, but does not ensure reception of the lower layers. Furthermore, each layer then requires an individual amount of overhead, increasing the bandwidth allocated by the service.

Fuelled by the increase in connectivity prior research efforts have investigated wireless multi path and mesh network topologies. In general two research directions co-exist within the field of wireless mesh network and multi path scenarios: One direction focusing on improving performance in these networks by intelligent planning and coordination along with simple Network Coding (NC) [3, 4], while the other approach relies on advanced coding and opportunistic behaviour [5–7].

This work was partially financed by the Green Mobile Cloud project granted by the Danish Council for Independent Research (Grant No. 10-081621), the Colorcast project granted by Danish Ministry of Science (Grant No. 12-126424/FTP), and by the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement No. 285969 [CODELANCE]

978-1-4673-4919-2/13/$31.00 ©2013 IEEE

Even though the size and shape of both wireless networks and the connected devices are in rapid evolution, the approach used for distributing content such as video remains the same. This is eventually bound to change.

We propose a novel approach to distribution of scalable video content to multiple heterogeneous receivers by creating a synergy between linear NC and scalable video coding. The designed network coding structure provides fountain code capabilities for increased robustness while preserving possible low complexity decoding in order to accommodate for devices with limited computational power. The basic principle is to support a wide range of devices with varying link qualities by creating a correlation between decoding complexity and channel quality and allowing devices to only receive and extract a sub-stream of the video.

In the following, Section II includes coding structures of H.264/SVC video. This knowledge is then utilised in a proposed network coding structure in Section III.

II. Video Coding Structure

In the field of source coding for video, three approaches are generally considered: Single Description Coding (SDC), Scalable Video Coding (SVC) and Multiple Description Coding (MDC). Each approach pose both advantages and drawbacks, all providing different means for creating scalable content.

In order to provide a video streaming service for a heterogeneous set of receivers, different video qualities must be available in order to accommodate receivers either high or low quality video. The SVC extension to the H.264 standard (denoted H.264/SVC) provides such capabilities, enabling low quality video to smaller and less capable devices, while providing the full quality for capable receivers.

A wide range of coding algorithms and mechanics are utilised by the H.264 video encoding in order to provide efficient video compression. The coded data are encapsulated in structures and substructures that should be kept pristine in order to minimise the impact of lost data. The work in [2, 8, 9] provides a full overview of H.264. Following is a summary of the structures encapsulating the compressed video data used further in this paper.

Slices: H.264 operates on parts of frames, slices, which then again consist of macro blocks, which essentially are small rectangular areas of pixels in YUV 4:2:0 format [8]. Each slice categorised as either an I, P or B-slice, depending on the type of prediction used for the macro blocks within the slice. That is, for I-slices all contained macro blocks use intra-prediction, making the slice independent on other slices. Thus it can be decoded without prior information of the video. P and B-slices contain at least one macro block that uses respectively one or two motion compensated prediction signals, and thus making the slice dependent on prior or subsequent information in the video. The inter-frame dependencies in a coded video stream originates from these slices. Every slice can be decoded independently of other data in the respective picture.

NAL Units: The Network Abstraction Layer (NAL) provides network-friendliness for the underlying compressed video data bit stream by separating the coded data such as slices into NAL units, structured integer-length byte sequences of limited size. These especially serves the purpose of accommodating packet based networks such as e.g. IEEE 802.11 and 3G, where packets are either received correctly or lost. By intelligently packing the coded data into such structures, the impact of lost data can be restrained. The NAL decoder interface specified in the H.264/AVC standard assumes that NAL units are received in correct decoding order, if not lost [9]. As the NAL units are expected pristine by the decoder interface, these should be treated indivisible in a streaming scenario.

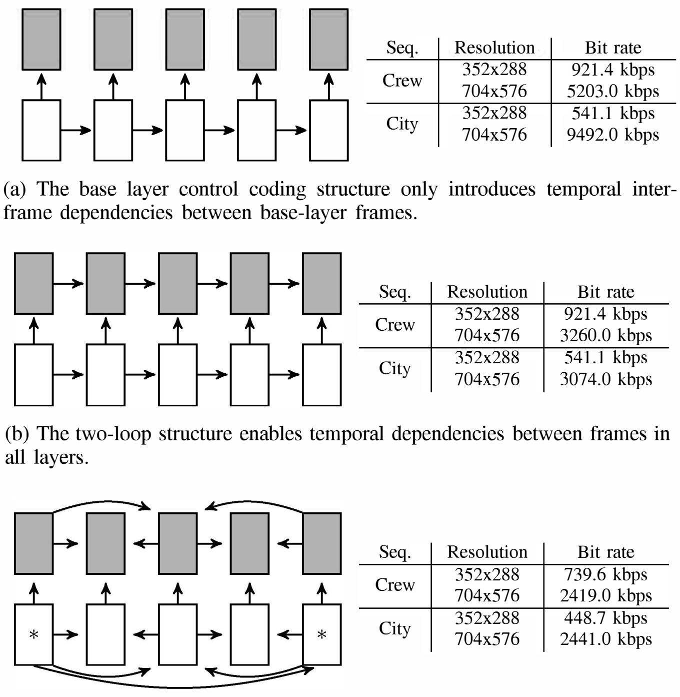
Access Units: Sets of NAL units containing information that results in one decoded picture is denoted an access unit. Furthermore, the data contained within an access unit may also contain redundant picture information about the picture described by the entire access unit. This is in order to accommodate decoders in the phase of recovery from earlier lost NAL units. Such data may be discarded by a decoder if unneeded, or simply entirely omitted by the respective encoder [9]. The compressed video data encapsulated by an access unit is further denoted as a frame.

Instantaneous Decoding Refresh: At the very start of every independently decodable video sequence is an Instantaneous Decoding Refresh (IDR) access unit. Such access units contain an intra-frame, a coded picture that always can be decoded independently of previous pictures in the coded NAL stream [8].

The above mentioned structures are embedded in the coded video stream in order to divert from the bit stream nature of source coding and provide increased robustness and usability for packet based networks [8]. Thus these structures should be respected by further error correction coding in order to keep the amount of errors in the video stream at a minimum.

A scalable video stream poses a structure in the inter-frame dependency that can be configured to provide certain beneficial features, depending on the application. Spatial and temporal scalability can be configured along with coding delay and the period between each IDR. The resulting bit rate of the video is then dependent on the configured features, which may be a limiting factor in a streaming scenario.

The structure of the inter-frame dependency may influence the behaviour of a streaming scenario significantly, depending on the error pattern and additional error correction. That is, if a sequence of coded pictures all introduce a co-dependency, lost data may introduce errors in a large part of the remaining pictures. Depending on the encoded structure. For H.264/SVC the inter-frame dependencies can be configured temporally only, as the spatial dependency structure is inherently specified by the layered coding approach. In order to confirm this difference in behaviour when subject to loss three different scalable video structures are included in the further investigation, illustrated in Figure 1. Each investigated video coding structure is used to encode two video sequences. The “Crew” and “City” sequences. The quality of these coded video sequences is plotted in Figure 2, though only presenting the enhancement layer quality. The base layer provides similar Y-PSNR values, but for a lower resolution. The video structures each pose



Seq.	Resolution	Bit rate
Crew	352x288	921.4 kbps
	704x576	5203.0 kbps
City	352x288	541.1 kbps
	704x576	9492.0 kbps

(a) The base layer control coding structure only introduces temporal inter-frame dependencies between base-layer frames.

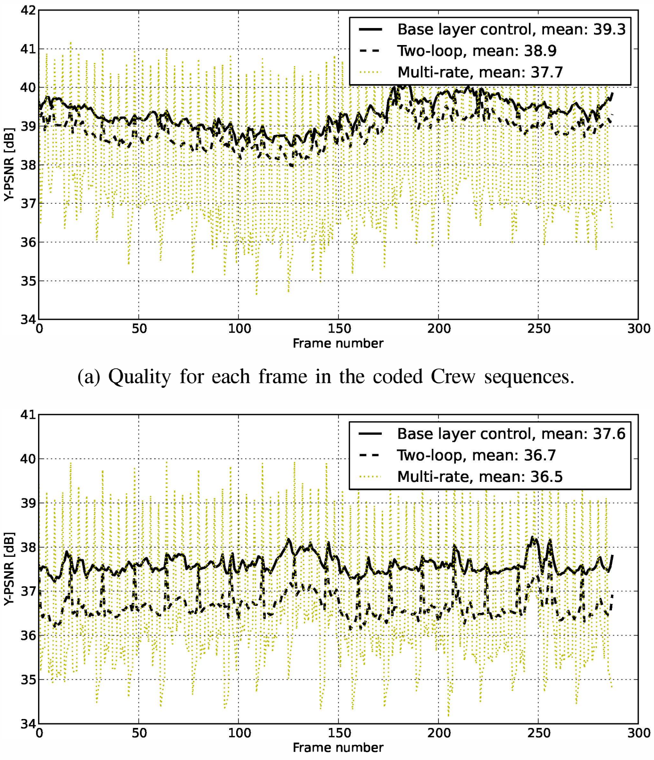
Seq.	Resolution	Bit rate
Crew	352x288	921.4 kbps
	704x576	3260.0 kbps
City	352x288	541.1 kbps
	704x576	3074.0 kbps

(b) The two-loop structure enables temporal dependencies between frames in all layers.

Seq.	Resolution	Bit rate
Crew	352x288	739.6 kbps
	704x576	2419.0 kbps
City	352x288	448.7 kbps
	704x576	2441.0 kbps

(c) The multi-rate structure defines a temporal inter-frame dependency structure the enables temporal scaling as well as spatial. Frame rate of 6 and 12 frames per second are available from the stream as well, though not listed here.

Fig. 1: The investigated scalable video coding structures and specifications of the respective coded sequences. All sequences are coded at a top layer frame rate of 24 frames per second. Light squares indicate base layer frames, dark is enhancement layer. Tables indicate bit rate of each layer.



(a) Quality for each frame in the coded Crew sequences.

(b) Quality for each frame in the coded City sequences.

Fig. 2: Video quality of the coded video sequences in terms of Y-PSNR for each frame. Only quality for the top layer video stream is presented.

advantages and drawbacks in a video streaming scenario:

Base layer control: By only specifying inter-frame dependency between base layer frames, the higher quality video in the enhancement layer can be enabled/discarded after convenience on a per frame basis. This can be helpful in scenarios with fading channel, e.g. streaming on mobile devices. However, the trade-off is the increase in video bit rate, which is also an important factor for such scenarios. From Figure 2 it is shown that this structure provides a consistent quality for all frames in the coded video sequences.

Two-loop: Increasing the inter-frame dependency to include all layers in the structure decreases the responsiveness of the enhancement layer. This results in a video quality that can only be adjusted at each IDR. Compared to the base layer control structure, the bit rate is lower due to the increased inter-frame dependency, while the video quality is preserved (cf. Figure 2). Furthermore, the IDR access units (every 16th frame) provide a slightly better quality.

Multi-rate: Temporal scalability is added in this structure by intelligently using bi-directional dependence between coded pictures in both spatial layers. This provides efficient video compression, resulting in a low bit rate. The spatial enhancement layer can only be enabled at each base frame belonging to the lowest available frame rate (marked with $*$ in Figure 1c). Furthermore, these base layer frames also constitute the backbone of the coding structure, as the remaining frames are heavily dependent on these. Thus, should these frames be lost, the impact of the video stream would be significant. This is also expressed in the resulting Y-PSNR in Figures 2a and 2b, where every 4th picture is of significantly higher quality. As the majority of frames is of lower quality, this lower quality may provide a better indicator of the perceived quality than the average Y-PSNR.

Allowing a lower quality in the two-loop encoding structure may enable a bitrate comparable to that of the multi-rate structure for a perceived video quality remains comparable to the multi-rate quality. As the multi-rate encoding structure dictates a strict inter-frame dependency structure, most of the information in such encoded video is embedded in the low-rate video (every 4th frame). This implies that transmission erasures will most likely occur in these frames, which may cause corruption in a significant part of the video stream if not corrected. Oppositely, the inter-frame dependency for the two-loop structure is loosely defined, which may reduce the impact of unrecovered erasures.

Additionally, all of the investigated video coding structures introduce the problem of ensuring successful reception of the base-layer. The bit stream structure of a video stream, along with the inter-frame dependency should be taken into account in a designed NC and Forward Error Correction (FEC) scheme intended for multiple heterogeneous receivers. Likewise, the inter-frame dependency structure should be chosen according to the deployed scenario.

Video source coding poses properties useful for an error correction scheme deployed in broadcast scenario where feedback is undesired. Firstly, for video streaming data loss is not fatal: The H.264 video coding standard was designed for video streaming [8], meaning that a decoder should be able to cope with missing data and thus continue functional operation after data loss. Data loss may however introduce errors in the video stream. Should the inevitable happen, and lost data is causing erroneous pictures in the video stream, this effect will be corrected at the next IDR access unit, if not before. Adjusting the IDR period will thus affect the maximum time a video stream can suffer from corruption caused by an erasure. This though poses a trade-off in increased bit rate.

III. Network Coding for Video

We design a structured network coding approach that allows low complexity decoding for receivers with fair channel conditions, while high complexity decoding is necessary for extended erasure correction. The proposed coding structure takes into account the underlying structure of a scalable video stream, and provides support for multiple hierarchically dependent layers in such stream.

The hierarchical dependency structure of SVC video favours the lower layer video streams, rendering any enhancement layer unusable before the subjacent layers are successfully received. A linear intra-session NC approach may strictly ensure such receiving order.

By combining the fountain code capabilities of Random Linear Network Coding (RLNC) [10] with a structured approach where packets are transmitted uncoded, it is possible to accommodate receivers with limited capabilities. Additionally, extended error correction opportunities of higher complexity can be co-enabled for capable nodes. The features utilised in the proposed intra-session network coding approach is presented in the following sections:

A. Systematic Coding

Transmitting the packets contained in a generation uncoded, and only coding the redundant packets effectively lowers both the encoding and decoding process. An uncoded source packet x_i from a generation $G = \{x_1, \ldots, x_n\}$ can be expressed as a linearly coded packet $U_i = \sum_{j=1}^{n} \delta_{ji}\ x_j$, where δ is the Kronecker delta notation. Thus, such uncoded source packets are perfectly viable combinations and can be included directly in the decoding process, while remaining usable independently. Systematic coding also enables the encoder to start the coding process before a whole generation of data has been generated, reducing the encoding delay.

B. Parity Packets

As the systematic coding phase completes for the generation G, it is possible to construct a packet that is always innovative (for incomplete decoders) without introducing the complexity of linear coding. Such packet is constructed by $P = \sum_{j=1}^{n} x_j$ and is denoted a parity packet. Should only one of the systematic packets be lost, the parity packet can be used to recover this dropped packet by XORing it with the received uncoded packets. Otherwise, it can be included in the decoding process, where it with certainty will provide one degree of freedom regardless of the utilised finite field size.

C. Random Linear Combinations

By randomly combining original source packets x_i into coded packets of the form $C_i = \sum_{j=1}^{n} c_j x_j$. A potentially unlimited amount of coded packets can be produced. The variable c_j specifies a random coefficient from a finite field. This random structure introduces a probability of linearly dependent coded packets. This can though be neglected for higher field sizes (e.g. 2^8 or higher [11]). Furthermore the impact of linear dependency is reduced with larger generation sizes [10, 11].

D. Expanding Window Coding Structure

The set of source packets within a generation that is included in the combined coded packets is further denoted as the "coding window". It is possible to address the hierarchical structure of SVC by varying the size of the coding window. Limiting the coding window to only include source packets containing data from the base layer video stream allows decoding of the base layer alone. If enough of these coded packets are received. However, expanding the coding window to further include enhancement layer source packets produces coded packets that, when decoded, results in both base layer data and enhancement layer data. In this decoding process the base layer coded packets can beneficially be included. This approach ensures that the hierarchical dependency in SVC is fulfilled. Prior research has also combined SVC and expanding window linear coding for the purpose of unequal error protection, where the redundancy of the lower layer video is increased to ensure decoding of the base layer [12].

E. Respecting the Video Structure

To reduce impact of losses the internal structures embedded in video coding (cf. Section II) should be respected for easy recovering. By making sure that access units are always located in the start of an encoded symbol it is ensured that an unrecovered symbol will not affect multiple access units, thus minimising the impact of erasures. Figure 3 illustrates a simplified example of this data packing. In practice, the generations contains a larger amount of symbols than illustrated.

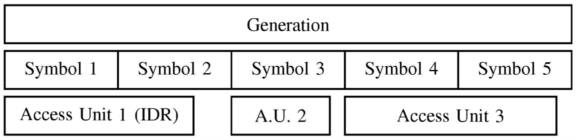


Fig. 3: Alignment of access units to symbols in a generation.

Structuring the coding generations such that the first access unit in a generation is an IDR access unit may emphasise the renewing qualities of such access unit. Should a preceding

$$\begin{bmatrix} \mathbf{U_1}^{(1)} \\ \mathbf{U_2}^{(1)} \\ \mathbf{U_3}^{(1)} \\ \mathbf{P}^{(1)} \\ \mathbf{C_1} \\ \vdots \\ \mathbf{C_{m-1}} \\ \mathbf{U_4}^{(2)} \\ \mathbf{U_5}^{(2)} \\ \mathbf{U_6}^{(2)} \\ \mathbf{P}^{(2)} \\ \mathbf{C_m} \\ \vdots \\ \mathbf{C_n} \end{bmatrix} = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ 1 & 1 & 1 & & & \\ c_{11} & c_{12} & c_{13} & & & \\ \vdots & \vdots & \vdots & & & \\ c_{(m-1)1} & c_{(m-1)2} & c_{(m-1)3} & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ c_{m1} & c_{m2} & c_{m3} & c_{m4} & c_{m5} & c_{m6} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & c_{n3} & c_{n4} & c_{n5} & c_{n6} \end{bmatrix} \begin{bmatrix} \mathbf{x_1}^{(1)} \\ \mathbf{x_2}^{(1)} \\ \mathbf{x_3}^{(1)} \\ \mathbf{x_4}^{(2)} \\ \mathbf{x_5}^{(2)} \\ \mathbf{x_6}^{(2)} \end{bmatrix}$$

Fig. 4: An example encoding matrix of the proposed NC coding structure.

generation be unrecoverable e.g. due to a burst in packet loss, instantly decodable video is then ready in the next generation.

By combining the expanding window coding structure with systematic coding, parity packets and randomly combined packets it is possible to define a linear coding approach that consists of two coding phases. (1) an expanding window systematic phase and (2) a random linear codign phase. The systematic phase provides low complexity with few error correction capabilities, while the random linear coding phase provides extensive error correction with a trade-off in increased coding complexity. The two phases are mutually beneficial in the sense that packets received in the systematic phase will effectively lower the complexity for decoding processes of the random linear coding phase, due to the instantly available uncoded symbols. Likewise will the random linear coding phase provide additional error correction, should the systematic phase not suffice. However, should this impose undesired complexity for the receiving node, this node can opt to just use the systematic packets received and then deal with the errors that may occur in the video. Examples of the coding follows.

F. Coding Structure Examples

A video stream containing two layers of video quality, both contributing with roughly the same bit rate, is to be distributed to multiple heterogeneous receivers. A network coding generation encapsulating the source code is set to cover a total of six packets. Three from each video layer. The generation is then encoded with the proposed coding structure, illustrated in Figure (4).

The encoding follows the coding vectors in Figure (4) chronologically from the top and down: symbol ${U_1}^{(1)}$ is transmitted first, then ${U_2}^{(1)}$ and so forth. The amount of redundancy applied to the lower layer (packets $P^{(1)}, C_1, ..., C_{m-1}$) can be adjusted to fit the desired error correction. Section III-G proposes how such redundancy can be chosen optimally. A heterogeneous set of receivers is considered in the example:

Small display, few capabilities: With a small display, only low quality video is needed to provide a proper video experience, thus only the base layer packets are needed. This node may acquire generated packets ${U_1}^{(1)}, {U_3}^{(1)}, P(1)$ from the base layer. The missing packet can the be extracted with XOR operations: ${U_2}^{(1)} = {U_1}^{(1)} \oplus {U_3}^{(1)} \oplus P^{(1)}$. The node can then present the video in the needed quality, and stop receiving further data from this generation. Battery consumption is reduced as computational power and network operation is kept at a minimum.

Small display, further capabilities: Should a node not be able to recover the needed base layer video by only utilising the additional parity packet, the node must choose between erroneous video, or do additional work to attempt to correct the error. Consider a node only receiving packets ${U_1}^{(1)}, {U_3}^{(1)}$. The full base layer can not be extracted from these, but the packets alone will with high probability provide at least some video data, but most likely also errors. The scale of these errors are unknown. The node can choose these errors and limit the needed computational and network efforts, or further attempt to receive the enhancement layer and the coded linear packets that follows in the random linear coding phase of the transmission.

Large display devices: For nodes requiring a higher video quality, the choice between further error correction and erroneous video poses fewer complications as the amount of redundant data needed to be received is limited. Only an amount of data corresponding to the total size of base layer and enhancement layer is to be received. The more received from the systematic phase, the lower complexity. However, if the complexity introduced by coding is undesired, the node can opt to only use what is received in the systematic phase.

G. Choosing Optimal Sub–Layer Redundancy

The amount of redundancy that should be added to each of the lower layers (the sub-layers) in the video structure introduces a trade-off between network friendliness and energy consumption at the receivers. From a network efficiency perspective the optimal approach is not to add redundancy to any layer but the top enhancement layer (omitting packets $P^{(1)}, C_1, ..., C_{m-1}$ in Figure (4)). For receivers requiring higher layers of the video, the redundant packets for the lower layers poses a probability not being innovative, as these receivers can not stop receiving packets: they are waiting for the higher layers to be transmitted. Due to the fact that the redundant packets for the top enhancement layer is going to be transmitted under all circumstances, all receivers might as well use these packets as repair packets. However, this approach is inefficient in terms of energy consumption at the receivers, as the devices only needing the low quality video will most likely need to both receive and decode the full enhancement layer, requiring additional network activity and network coding effort. The following proposes an approach to optimising the sub-layer redundancy wrt. minimising energy consumption.

A sub-layer redundancy can be specified such that a minimum of total packets is processed by the receivers, assuming the distribution of receivers requiring the low and high quality video (further denoted D_L and D_H respectively) is known or estimated, This metric can be used under the assumption that the amount of total packets received by the nodes is directly proportional to the energy consumption in receiving and processing these packets. Furthermore an arbitrary estimate of channel packet loss, denoted r, is utilised.

A two-layered scenario is used as an explanatory example. The size of the base layer is defined as L_1, while the total size of base layer and enhancement layer, the generation size, is denoted g. The amount of redundancy packets including parity packet is denoted R_1. In order to successfully extract only the base layer, a node must receive L_1 out of the first $L_1 + R_1$ transmitted packets. Due to the random packet loss, only a subset of D_L manages to do this. The average size of this subset, ${D_L}'$, is given by the binomial distribution

$$ {D_L}' = D_L \cdot \sum_{i=L_1}^{L_1+R_1} \binom{L_1+R_1}{i} \cdot (1-r)^i \cdot r^{(L_1+R_1-i)}. \quad (1) $$

The remaining subset, ${D_L}^{\dagger} = D_L - {D_L}'$, must then receive and decode the enhancement layer in order to extract the needed base-layer. Thus the total amount of packets needed to be received by the D_L receivers is given by $\Sigma D_L = {D_L}' \cdot L_1 + {D_L}^{\dagger} \cdot g$.

Naturally, all D_H nodes need to receive g packets. Furthermore, because these nodes must endure the full base layer transmission phase they might receive non-innovative base layer packets, which inherently introduces an overhead. As the channel conditions are assumed similar for all receivers, average overhead, g^{+} is again the binomial distribution:

$$ g^{+} = \sum_{i=L_1}^{L_1+R_1} \binom{L_1+R_1}{i} \cdot (1-r)^i \cdot r^{(L_1+R_1-i)} \cdot (L_1+R_1-i), \quad (2) $$

This overhead only applies to the D_H receivers, as the D_L nodes surely does not encounter the problem of receiving too much base layer data. They should simply stop receiving when they have enough. Thus, in total the amount of packets the D_H nodes will receive is then $\Sigma D_H = D_H \cdot (g + g^{+})$. The optimal base layer redundancy R_1 is then found by solving the optimisation problem

$$ \underset{R_1 \in \mathbb{N}}{\arg\min} (\Sigma D_H + \Sigma D_L). \quad (3) $$

This optimisation problem can be expanded to include several layer, though with increased complexity. Furthermore, this only suggests the optimal redundancy at the sub-layers. The redundancy applied to the top layer can be adjusted to accommodate for arbitrarily high packet losses if needed, as it is assumed that this extra redundancy does not cause increased power consumption at the receivers that has already decoded the generation. Figure 5 gives an example of this optimisation process in a scenario with 20 total receivers $(D_L + D_H)$, a sub-layer size, L_1, of 20 packets and a total size of the generation g, of 80 packets, with packet loss, r, of 10 %. The amount of D_L receivers is then increased and the optimal sub-layer redundancy expressed as a function of this.

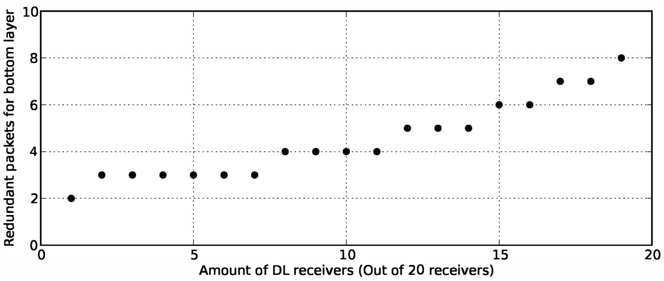


Fig. 5: Example of the optimal amount of redundant packets.

IV. Conclusion

The potential of utilising beneficial features in the structure of network and video source coding has been investigated with respect to wireless broadcast streaming scenarios. The findings advocate that intelligent structuring of the utilised codes can introduce a coding complexity proportional to the error rate on the channel, enabling advanced FEC coding for capable devices while low-end devices may utilise no or low-complexity FEC coding given proper channel conditions. Interplay between the coding structures enables energy-efficient scalable video streaming to heterogeneous devices. A scenario where state-of-the-art methods fall short, in terms of either network efficiency, reliability or content scalability.

References

[1] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, 2001.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the h.264/avc standard," *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[3] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the Air: Practical Wireless Network Coding," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 497–510, June 2008.

[4] S. El Rouayheb, A. Sprintson, and C. Georghiades, "On the index coding problem and its relation to network coding and matroid theory," *Information Theory, IEEE Transactions on*, vol. 56, no. 7, pp. 3187–3195, 2010.

[5] S. Biswas and R. Morris, "Opportunistic routing in multi-hop wireless networks," *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 1, pp. 69–74, Jan. 2004.

[6] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," in *Conf. on App., tech., archi., and prot. for comp. comm. (SIGCOMM)*. New York, NY, USA: ACM, 2007, pp. 169–180.

[7] T. Ho, B. Leong, M. Médard, R. Koetter, Y. Chang, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proc. IEEE ISIT'03*, jun 2003.

[8] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[9] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/avc in wireless environments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 657–673, 2003.

[10] D. J. C. MacKay, "Fountain codes," *IEE Proceedings-Communications*, vol. 152, no. 6, pp. 1062–1068, 2005.

[11] J. Heide, M. V. Pedersen, F. Fitzek, and M. Médard, "On code parameters and coding vector representation for practical rlnc," in *IEEE International Conference on Communications (ICC) - Communication Theory Symposium*, Kyoto, Japan, jun 2011.

[12] D. Vukobratovic and V. Stankovic, "Unequal error protection random linear coding for multimedia communications," in *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, oct. 2010, pp. 280 –285.