Cross-layer Architecture for Scalable Video Transmission in Wireless Network

J. Huusko J. Vehkaperä, P. Amon, C. Lamy–Bergot, G. Panza
J. Peltola and M. G. Martini

Abstract

Multimedia applications such as video conference, digital video broadcasting (DVB), and streaming video and audio have been gaining popularity during last years and the trend has been to allocate these services more and more also on mobile users. The demand of Quality of Service (QoS) for multimedia raises huge challenges on the network design, not only concerning the physical bandwidth but also the protocol design and services. One of the goals for system design is to provide efficient solutions for adaptive multimedia transmission over different access networks in all-IP environment. The joint source and channel coding (JSCC/D) has already given promising results in optimizing the multimedia transmission. However in practice, arranging the required control mechanism and delivering the required side information through network and protocol stack have caused problems and quite often the impact of network has been neglected in studies. In this paper we propose efficient cross-layer communication methods and protocol architecture in order to transmit the control information and to optimize the multimedia transmission over wireless and wired IP network. We also apply this architecture to a more specific case of streaming of scalable video stream. The scalable video coding has been an active research topic recently and it offers simple and flexible solutions for video transmission over heterogeneous networks to heterogeneous terminals. In addition it provides easy adaptation to varying transmission conditions. In this paper we illustrate how the scalable video transmission can be improved with efficient use of proposed cross-layer design, adaptation mechanisms and control information.

Key words: Scalable Video Coding (SVC), Cross-layer design, WLAN, joint source and channel coding, Multimedia adaptation layer, header compression, UDP-Lite, RTP packetization

1 Introduction

The evolution of wireless telecommunication systems can be divided into short term and long term evolution towards a global and integrated system, which will meet the requirements of both users and industrial world, and which could make efficiently use of emerging technologies. The expectations of evolution, whether called as short term i.e. increasing the bandwidth with new radio and access technologies or as longer term i.e. co-operative, converging networks are in the end very similar in many ways. For end-user side the expectations of the future system include for example the good service quality and improved quality of experience (QoE), easy access to applications and services, improved usability of services, enhanced security and reasonable cost. Similarly on the service and network provider side minimizing the operational and capital expenditures by easy quality of service provisioning and network/security management, flexibility of configuration and reconfigurability of the system and maximization of the network capacity are the expected values. Fulfilling these expectations will be a challenging task for system designers, who are aiming at producing flexible next generation wireless systems that interconnect, in a transparent way, a multitude of heterogeneous networks and systems. Optimal allocation of system and application resources can be achieved with the co-operative optimization of communication system components in different layers, and this is in particular the case for multimedia processing and transmission. The increased amount of the wireless network components in the whole transmission system and the demand of better QoS and QoE are guiding the work to more optimized co-operation of different elements in the whole multimedia system.

Traditionally, the two encoding operations of compression and error correction are separated from each other, following Shannon's well-known separation theorem [1], which states that source coding and channel coding can, asymptotically with the length of the source data, be designed separately without any loss of performance for the overall system. However, it has been shown that separation does not necessarily lead to the less complex solution [2] and separation is not either always applicable [3], especially for multimedia transmission. Thus the JSCC/D techniques that include a co-ordination between source and channel encoders have been recently investigated, and techniques have been developed [4–6] to improve both encoding and decoding processes while keeping the overall complexity at an acceptable level [7].

In order to benefit from joint source channel coding in real systems, control information needs to be transferred through the network and system layers. Unfortunately, the impact of the network and networking protocols are quite often discarded while presenting the joint source and channel coding systems and only minimal effort is put into finding solutions for providing efficient

inter-layer and network signaling mechanisms. Some work has however carried out in order to provide cross-layer protection strategies for video streaming over wireless network, such as combining the adaptive selection of application forward error correction (FEC) and medium access control (MAC) layer automatic repeat request (ARQ) as presented in [8,9].

There are already some mechanisms in use for generic information exchange between the different system layers, as the quality of service (QoS) features, namely Differentiated Services (DiffServ) [17] and Integrated Services (IntServ), which provide means for an application to reserve resources and specific service level from the interconnecting IP network by mapping the application requirement at network protocol level. Another example of the inter-layer signaling can be found from IEEE 802.11e standard where the QoS provisioning is performed between the application and the medium access layers.

The QoS information consisting of the IP packet priorities, to drop them selectively, is not alone sufficient as an optimization method for multimedia transmission.More detailed information needs to be delivered in order to fully optimize the end-to-end transmission in a cross-layer manner. Some of the possible methods to arrange the control information delivery between physical and application layer for are discussed for example in [10], which describes the use of two additional adaptation layers in protocol stack, in order to transmit cross-layer information within protocol stack and through the network. Another possible solution for transferring the required controlling information is to extend the current protocols such as Internet Protocol version 6 (IPv6) or Internet Control Message Protocol version 6 (ICMPv6) through the definition of new options and message types.

The presented solutions are potential candidates for transferring the control information through both wired and wireless network but they do not solve fully the problem of transferring control information through protocol layers from application layer to physical layer and vice versa. Furthermore they do not propose solutions to use this information for end-to-end optimization, which requires to take into account all protocol layers and particularly applications.

Within multimedia applications, scalable video coding has gained a lot of attention recently from both researchers and standardization committees [11]. The scalability has been a part of several recent video coding standards (e.g. MPEG–2, MPEG–4, H.263+) but it has not reach a wide popularity in the industry mainly because of poor compression perfomance. Currently, a new scalable video coding standard is under standardization process by Joint Video Team (JVT) and it provides comparable compression efficiency and flexible 3-D scalability [12]. The layered structure of the scalable video stream and different priorities among the layers support the usage of unequal error protection (UEP) [13,14] techniques and prioritization of layers on network level

e.g. using Differentiated Services (DiffServ) [15]. Combination of scalable video and different cross-layer solutions have been studied in several papers for example in [16], where end-to-end transmission control protocols and congestion control for scalable video has been studied. In [8], combining of the adaptive selection of application FEC and MAC layer ARQ together with scalable video has been studied.

In this paper we introduce the innovative IST-PHOENIX cross-layer architecture, which considers not only the joint optimization of source and channel coding, but also includes the interconnecting network in more detail than in previous works. One of the originality of this paper is the definition of the needed cross-layer information and the solutions how this information can be delivered and utilized on different system layers within the system architecture defined in Chapter 2. In Chapter 3 we describe the methods to optimize multimedia transmission in network level, and the mechanisms for transferring transparently the control information through the network and within the protocol stack. In Chapter 4 we concentrate in more detail on optimization of scalable video transmission in PHOENIX architecture and we present some examples how the proposed cross-layer information can be utilized to adapt scalable video according the varying transmission conditions.

2 PHOENIX System Architecture

Fig. 1 represents the overall system architecture, including the informative signals and control signals for transmitting cross-layer control information through the system in order to implement efficient adaptation techniques such as unequal error protection (UEP) and soft input source decoding schemes. The system utilizes several new control lines for control signal delivery and application and physical layer controller units in a way that those are also exploitable in cross-layer communication.

The PHOENIX system consist of routers and nodes using both wireless and wired connection on network side. In our application scenarios we have assumed the last hop in the network as an error-prone wireless channel in order to demonstrate the possible mobile user case and to benefit from particular control signals (e.g. source a-posteriori and decision reliability information).

In addition to traditional system blocks, the architecture includes both physical and application layer controller units. Controllers are used for supervising the other components to adapt to changing conditions by providing them information about network and channel conditions and user requirements. For the controlling purpose, we have defined signaling mechanisms for both transmitter and receiver side.

The following cross-layer control information have been considered to be the most relevant in our system:

- Source significant information (SSI);
- Source a-priori information (SRI);
- Channel state information (CSI);
- Network state information (NSI).
- Decision reliability information (DRI);
- Source a-posteriori information (SAI);

2.1 Required Cross-Layer Signals

The source significant information (SSI) is generated by the source coder and it represents the information on the sensitivity of source bits to channel errors. The SSI can be exploited by UEP techniques and it needs to be synchronized with payload. Due to strict relation to the media stream, the transmission of SSI signals provides a significant challenge to the interlaying network and protocol communications.

The source a-priori information (SRI) is produced by the source encoder and used by the destination source decoder. The SRI information is any information that the decoder can utilize for improving the QoS and possibly helping the soft decoding e.g. type of video or variable length code (VLC) size. The SRI is synchronized with the associated video stream, more and more in relation of its accuracy, and it is generated by and targeted to the same JSCC/D devices as SSI. However, the amount of delivered SRI information is lower than SSI information.

The channel state information (CSI) delivers the actual conditions of each wireless channel through which the media stream is directed. The CSI signals go in the reverse path with respect to the video data packets, hence it is not strictly synchronized with them. Furthermore, the CSI frequency should be much lower than the packet rate to be considered almost negligible in terms of additional overhead.

The network state information (NSI) reports about the availability of network resources across the data path. Such information can be represented by the QoS performance parameters such as delay, delay-variation (jitter) and packet loss. NSI can be effectively exploited at the source encoder to better tune the amount of the generated rate and coding parameters in general, as well as at each radio transmitter node. The NSI information goes towards the source terminal and it is not synchronized with the media stream. However, the NSI reports must be frequently updated and thus an automatic scaling mechanism with respect to the number of destination terminals is required in order to

5

well accomplish even large multicast sessions without significantly loading the network, especially in the uplink direction.

The decision reliability information (DRI) provides further elements related to the channel decoding process. It is based on the concept of "soft" decision, where the final result about the value of a bit is not simply either 0 or 1, but also the level of certainty or soft value of it. This implies that the size of DRI can be bigger than the data itself. The DRI is generated by radio receiver and used by the destination terminal in order to better tune the source decoding process and hence improve the resulting QoS. In practice the DRI will be limited to the last wireless hop in order to avoid the complexity of network and reduce the signalling load.

The source a-posteriori information (SAI) results from the analysis of the decoding process of the video stream. It is generated by the client terminal side source decoder and exploited at the radio receiver to set the working parameters of the channel decoder and demodulator module. The SAI travels in the direction from the destination terminal(s) to the source terminal and it is not strictly synchronized with the video data packets.

In the wired part of system the cross-layer signals related to source and channel decoding process are not needed. The NSI information and also the CSI information can be used in the wired part of the network to enhance the congestion control and congestion notification in the network. It is also possible to utilize the SSI information for QoS provisioning at media-aware network elements.

3 PHOENIX Network Architecture

A central concept of the cross-layer system is the Network Transparency and cross-layer information exchange, but also modifications at protocol level are required in order to increase the overall performance of the multimedia transmission system. The Network Transparency expresses the abstract idea of making the underlying network infrastructure almost invisible to all the entities involved in the system. The primary goal of Network Transparency is to transfer cross-layer control information through the IP network, in a transparent manner, in spite of strict rules of ISO OSI reference model.

On the network level, only the IPv6 is matter of interest in PHOENIX cross-layer architecture due to its improved routing capabilities, scalability, header structure and support of QoS compared to IPv4. The IPv6 does not implement natively protocol checksum, thus the upper layer checksum mechanism becomes mandatory. Since the PHOENIX system supports bit error resilient audio/video encoders, the transport protocol checksum would not be needed for whole data frame, and thus for multimedia data it is rational to prefer partial checksum over full checksum mechanism. With partial checksum only the important parts such as packet headers are protected, while the payload can be left unprotected.

Several different protocols supporting the partial checksum, such as UDP-Lite [19] and Datagram Congestion Control Protocol (DCCP) [20] have been developed in order to enhance the video transmission by maximizing the end-to-end throughput. The standard practice for implementing the partial checksum is to set fixed value for the checksum length. However, the audio/video streams can consist of different amount of important data, which should be protected. and the cross-layer system should be able to control and change the length of protecting checksum area when necessary. Thus, the PHOENIX architecture support Real-time Transport Protocol (RTP) over UDP-Lite multimedia transmission with dynamic partial checksum coverage.

The main results concerning Network Transparency for a single PHOENIX cross-layer-compliant video flow are summarized in the Table 1. For each control and signal information the most suitable way to carry it, together with the related overhead are presented. The transmission mechanisms for PHOENIX system were analyzed based on the overhead caused by the transmission and the required synchronization with the video stream. For the SSI control signals two different solutions can be used. Although the IPv6-hop-by-hop header provides tolerable overhead (only a few kilobytes per second) for SSI, it is more feasible to use RTP with scalable video coding since the RTP payload for scalable extension of H.264/AVC includes already the SSI information as described in Chapter 4. The SRI can also utilize the IPv6-hop-by-hop header. The ICMPv6 has been selected for the NSI and CSI transmitting over the network and for NSI also the RTP control protocol (RTCP) is usable. The DRI and SAI, which are utilized and generated at receiver can cause high bandwidth consumption, thus these signals are feasible only for receiver, which is connected directly to the wireless network.

The aforementioned transmission mechanisms comply with the standard protocol solutions and thus some of the signals, such as NSI, SSI and CSI could be utilized at any media-aware network element. Also, the standard solution

does not incure any additional overhead in transmission and the transmission overhead caused by the headers can be minimized with standard header compression mechanisms. However, since the SSI and SRI information need to be synchronized strictly with the video payload, we have also considered to generate an extra header for the video payload. The proposed extra header for ciphering information, SRI and SSI signal information as well as for quantization is illustrated in Fig. 2.

The header consist of ciphering initialization vector (Cipher key) and information whether frame is ciphered of not (Cipher flag). The extra header carries also the video header size, video format, number of reference frames and frame type in the SRI fields. The rest of the header fields are reserved for the SSI information (the number of SSI classes and the rate and length for each class) and for the number of quantization bits. In order to minimize the transmission overhead caused by the extra header, the maximum total size of the extra header is reduced to 12 bytes, which implies the use of maximum of 3 SSI classes per frame. In order to avoid transmission overhead in wireless channel and to improve the robustness to errors, we have implemented the robust header compression (RoHC) [21] with the corresponding protocol profiles.

3.2 Medium Access Control Partial Checksum Mechanism

Traditionally in TCP/IP networking the robustness of the system is guaranteed with effective protocol checksum mechanisms. Especially the protocol headers should be always protected efficiently in order to avoid severe system-wide failures. However, with multimedia data it could be more beneficial to leave the actual payload unprotected or to provide different protection levels in different parts of the data. This kind of method is usually referred as a unequal error protection (UEP) and it is performed at the physical and application level. The UEP schemes defined for video and audio de/coding, do not however consider or be directly related with data transmission. In order to transfer the multimedia data efficiently, also the link layer mechanisms should provide competent error detection and protection in unequal manner by protecting the header information but leaving the video or audio payload itself unprotected. We have considered new flexible and efficient solution for data link layer error detection, which is closely related to upper layer (transport and network layer) error detection mechanisms for multimedia data. The proposed partial checksum solution is also used with the Multimedia Adaptation Layer, which is defined with several simulation results in this paper.

We have introduced several improvements for the IEEE 802.11 carrier sense multiple access (CSMA) medium access control (MAC) mechanism, in order to support efficiently the multimedia transmission. These proposed features

include the arranging of partial checksum mechanism for multimedia data, a new data frame format and identification mechanism for the variety of cross-layer signals. Although the proposal is based on CSMA with collision avoidance, similar mechanisms can be provided for any access mechanism with some modifications.

In the IEEE 802.11 WLAN systems, the identification of the transmission frame and setting the type of the frame is performed using the two byte frame control (FC) field of the corresponding MAC-frame. We have enhanced the IEEE 802.11 data frame structure with additional multimedia data frame subtype, which is identified with FC field notation at the transmission frame header. Using the FC field to identify the multimedia frame we can assure that the MAC frame will be correctly identified at the receiver side and the protocol data unit is passed correctly also to the upper layer protocols. Introducing only the subtype modifications, it is also possible to provide full backward compatibility for the system with the original IEEE 802.11a/b/g standards. Similar frame identification method is applied also for several cross-layer control signals and new MAC control frame types have been defined for those.

In order to support multimedia transmission and scalable video, modifications has been made at MAC level transmission frame compared to the generic IEEE 802.11 frame structure. A new header field for checksum coverage area (CRC length) has been included and the frame check sequence (FCS) field has been moved to the MAC header part. The frame format for the the new multimedia subtype is illustrated in Fig. 3.

In some cases with cross-layer system supporting the unequal error protection (UEP) schemes, the length of the partial checksum needs to be defined dynamically. Thus, we have introduced both length and Frame Check Sequence (FCS) fields at the frame header. With this kind of arrangement it is possible to change the checksum coverage area and calculate the cyclic redundancy check (CRC) for specified coverage area more easily than for example in the case of the generic data frame structure. Since the coverage area is already located on header part of transmission frame, the identification of header and payload part of the transmission frame is easy at the receiver side. Since the FCS field is included in header part it needs to be set always to zero before the CRC calculation in order to avoid confusion in CRC check. At the transmitter side the value of coverage area can be inherited from the upper-layer protocols or it can be set static for each stream. Depending on the used upper-layer protocol suite, we have defined several modes for coverage area, which can be set per stream.

Since the channel is error-prone and the IPv6 does not include header error detection mechanism, the header should be verified for errors as early as possible. Thus, the minimum coverage for the partial checksum at the link layer

level should include the transmission frame header and the IP header part of the transmission frame payload. However, we suggest that the transport layer header is included in the CRC check in multimedia transmission. Thus, the length of the coverage depends on the used transport protocol. When using the RTP/UDP-Lite/IP header compression the minimum coverage area depends on the size of the compressed header, and the checksum should always cover the both MAC header and the compressed upper layer headers. For cross-layer information delivery the extra header, which is illustrated in Fig.refextraheader should be also included in CRC check.

4 Cross-layer Mechanisms for Scalable Video Transmission

In previous sections, we have defined a general architecture for cross-layer wireless video transmission and different system blocks of the architecture together with cross-layer information transmission techniques. While the proposed architecture is general to be used together with several radios and video codecs [22,23], we would like to present some special aspects related to scalable video transmission utilizing the proposed architecture. The proposed architecture can also be used with scalable video but due to different bitstream structure, some cross-layer mechanisms are realized slightly differently for scalable video than for traditional non-scalable video. In this section, we will present some novel cross-layer techniques for scalable video from the network aspect.

Our goal is to optimize video streaming performance in a network topology that consists of wireless hops in addition to wired links. Optimization is enabled by performing source rate adaptation on a regular time interval for the video stream at streaming server in the case of changing transmission capacity of the system. More faster rate adaptation is implemented at the data link layer allowing rapid reaction to channel state changes of the corresponding wireless channel. Faster adaptation is based on prioritization of the most important portions of the scalable video bitstream in the Multimedia Adaptation Layer (MAL) so that the most important portion is transmitted first, and the rest of the portions are then transmitted in decreasing order of importance. This adaptation solution is useful especially in the case of scalable video stream.

4.1 Scalable extension to H.264/AVC

New scalable video coding standard is currently under standardization process at the joint video team (JVT) of ISO/IEC MPEG and ITU-T VCEG standardization groups. Developed scalable video coding (SVC) standard will be integrated as an extension to H.264/AVC [25] in the form of Annex G. The

SVC extension to H.264/AVC adds signal-to-noise ratio (SNR) and spatial scalabilities in addition to already existing temporal scalability, and therefore expanding the use of H.264/AVC to several different application settings [12]. In SVC, scalability is achieved by taking advantage of the layered approach already known from previous video coding standards e.g. MPEG-2, MPEG-4 and H.263. The difference between the scalability used in previous standards and the one proposed by SVC is a more advanced inter and intra-layer prediction of both texture and motion information.

Scalable video bitstream is usually divided into layers which are classified as base layer and enhancement layer(s). This is also the case for the SVC. Information from lower layers is used to remove the redundancy between different layers. Consequently, the dependency between layers makes the prioritization of different layers during transmission suitable. The importance and dependency between the layers is important when adapting and protecting the bitstream on lower layers and it should be provided to the other system layers using e.g. SSI information exchange. In the literature, the protection on lower layers is usually achieved using forward error correction (FEC) on application or physical layer [14,13]. Together with application layer FEC, H.264/SVC which is an extension of the H.264/AVC, can be used in the proposed cross-layer architecture eventhough the bit error resilience tools are not specified in H.264/AVC and H.264/SVC [24]. Also, a sophisticated decoder can be used to hide bit errors. Sophisticated decoder can utilize the correct data before detected bit error and it can use error concealment techniques to conceal the corrupted data. In this paper, we will concentrate more deeply on the cross-layer adaptation of scalable video on different system layers and decoding issues of erroneous scalable bitstream are not deeply investigated.

4.2 RTP Payload for SVC

A first criterion for the design of an RTP payload format for SVC is the backwards compatibility to the payload format for H.264 [27]. A second design aspect relates to the reuse of the Network Abstraction Layer (NAL) unit header in the RTP payload header. The current specification of SVC defines a NAL unit header of three bytes for SVC video data. This header includes in addition to the first byte already incorporated in H.264 a second byte with priority information and a flag for discardable streams and a third byte including information about the index of the three scalability dimensions. The SSI information concerning the importance of every NAL can be thus easily integrated into the RTP header of SVC access unit and transmitted through the protocol stack. The definition of NAL units containing H.264 video data remains unchanged, which is especially a one byte NAL unit header mainly containing the NAL unit type.

The streaming servers can base their adaptation decisions on the SSI information provided in the RTP payload header. Three approaches can be differentiated. Simple algorithms can base their adaptation decision on the priority (SSI) information provided in the second byte of the NAL unit header. The higher the value, the less important is the related video packet and the more likely will it be discarded from the bit-stream in case of bit-rate constraints. The second option is the usage of the three scalability parameters included in the third byte. More complex adaptation decisions can be realized, e.g. the size of the video picture can be separately controlled. Most sophisticated algorithms combine the first and the second approach realizing e.g. optimal rate-distortion stream thinning by the implementation of so-called quality layers [28].

Using the RTP payload format for SVC, adaptation of the video stream can also be done in the network using media-aware network elements, not only at the server. The information for the adaptation decision is provided by NSI (mainly RTCP feedback). NSI allows estimating the available data-rate in the network and consequently controlling the adaptation decision (mainly video bit-rate). The advantage of bit-stream adaptation in the network compared to adaptation at the application level is a faster controlling of the data-rate. Still the adaptation at application level is important in order to avoid sending unnecessary video information.

4.3 Application Layer Controller for Scalable Video

To fully utilize the advances of the scalable video stream, adjustments to the video stream are required. Preliminary adjustments for resolution, frame rate and bit rate are made during the negotiation phase (e.g. with Session Initiation protocol (SIP) or Real Time Streaming Protocol (RTSP)) but in order to utilize changing transmission capacity more efficiently, adjustments are needed also during the transmission. At the application layer, adjustments to scalable video parameters should be made on regular time interval (e.g. one second interval) but not too often in order to keep system more simple. Faster adaptation should be made at lower layers of the wireless link for example on link or physical layer. In this subsection we will concentrate on longer time period controlling utilized on the application layer and we will discuss the short time adaptation in the subsection 4.4.

We have selected to use a novel fuzzy logic-based controller at application layer of the streaming server to adjust the bit rate and frame rate of the compressed video according the cross-layer information available from the different OSI system layers. Fuzzy controller sets bit rate and frame rate of the video at the video sender by utilizing the SSI information provided by the RTP payload

header of the packets. As input parameters, in addition to SSI information, the proposed controller uses NSI information (RTCP statistics), previous control state information (current bit rate and frame rate) and CSI information (estimate of the channel state in the form of signal-to-noise ratio (SNR)) as additional cross-layer information from the physical layer. The control logic is based on several fuzzy logic rules (54 rules) that model the different states of the transmission system. Fuzzy combining is used to collect results from different rules and to provide output values for the bit rate and frame rate. The controlling logic of the proposed controller can be summed using the following rules:

- Try to increase the bit rate if the transmission conditions are good;
- If the transmission conditions changes suddenly, the bit rate has to be changed more dramatically;
- Try to find a good compromise for bit rate and frame rate by utilizing available feedback information;
- Try to find the maximum transmission capacity of the system and utilize it efficiently.

In order to evaluate the proposed fuzzy logic-based controlling algorithm, we have simulated a transmission of H.264/SVC video stream over wireless IP network using a simulation platform which is based on the introduced PHOENIX architecture. The simulation architecture includes simulation models for all OSI system layers from the physical layer up to the application layer present in IP based video communication networks. The scenario considered is video transmission using RTP/UDP-Lite/IPv6 and WLAN radio with a maximum bit rate of 12 Mb/s. The capacity of the channel is shared among several users and only 2 Mb/s is considered for the user. Controlling of different system layers is done at one second interval which was chosen as a compromise between the availability of the feedback information and the cost of regular updates. Consistently, H.264/SVC video is encoded using group of pictures (GOP) with length of 1 s.

A non-selective block fading channel has been used to evaluate the performance of the proposed controller. Foreman test sequence (CIF resolution, duration 10 seconds) is transmitted through IPv6 network where the last hop is considered wireless. Fig. 4 illustrates the adaptation capability of the proposed controller and Fig. 5 represents the transmission conditions in terms of wireless channel's SNR and PLR of the system. As it can be seen from Fig. 4, the controller adapts the transmission rate based on changing transmission conditions presented in Fig. 5. When the state of the wireless channel and packet loss rate (PLR) of the system changes quickly, the fuzzy logic algorithm reacts on this by changing the rate of the sent video faster than in a situation where the channel state changes slowly. If the channel state is good and steady, the algorithm increases the bit rate moderately. The adaptation

of frame rate during the same simulation was also studied. Selected frame rate is illustrated in Fig. 5. As it can be seen from this figure, the proposed algorithm changes the frame rate of the transmitted video according the changing transmission conditions. The simulation scenario represents transmission of video in highly varying wireless channel in order to validate the efficiency of the proposed controlling algorithm. In steady state transmission conditions, the algorithm tries to find the maximum transmission capacity of the system available for certain user and utilize it efficiently which is the case during the first 5 seconds.

4.4 Multimedia Adaptation Layer (MAL) for Scalable Video

The packet classification based on SSI-information utilization, can be performed at the network (IP) level using Differentiated Services (DiffServ) DS field. For wireless networks, including also layer 2 access points and switches, which do not support layer 3 routing capabilities a lower level packet classification should be used in addition to the IP level classification.

Based on the scalable video bitstream layer prioritization suitability, we propose a mechanism for adapting the video transmission to rapidly changing wireless channel and network conditions above the link layer but after the network layer. One of the main requirements for the architecture is to be general enough to work with different access networks from IEEE 802.11 (WiFi) and IEEE 802.16 (WiMAX) to 3GPP and UMTS.

The data link layer MAL scheduling mechanism for scalable video is illustrated in Fig. 6. On the transmitter side, MAL consists of a packet classifier, for classifying the incoming packets from IP-layer to queues based on frame type or priority information provided by SSI information included in RTP header; it also includes transmit and receive buffers, and the scheduler engine. The scheduler engine takes advantage of CSI and NSI information in transmission frame formation. Depending on the current channel and network state the scheduling engine forms the data link frame from the available base layer and enhancement layers video frames, which can be transmitted with a certain level of assurance over the wireless network. The scheduling engine also inserts a frame identifier to the data link frame, which is further utilized at the receiver side to identify the incoming data frame format. The proposed MAL architecture is required for fast adaptation to rapid channel and network changes. The MAL solution is able to provide a medium access and physical layer independent solution for efficient multimedia transmission for the systems which are able to provide channel state information such as signal-to-noise ratio and/or signal strength. The architecture can be easily deployed e.g. with WiFi, WiMAX, and UMTS. For the WLAN, the MAL schedul-

14

ing engine is enhanced to support also the MAC frame control identification. The different SVC layers are marked in the FC field in order to improve the scheduling and identification of the video frames. Thus, the MAL completes the aforementioned MAC partial checksum mechanism and MAC cross-layer information exchange functionalities for IEEE 802.11 standard family.

We simulated simple application scenario with NS-2 network simulator. The scenario consisted two mobile nodes connected through WLAN access network. The nodes were located on a planar field and used Dynamic Source Routing (DSR) to discover each other and maintain connectivity. The first node transmitted a 10-second H.264/SVC pre-encoded video stream of Foreman video clip (CIF resolution and encoded bitrate of 1000 kb/s) with the base layer and two enhancement layers. The transmitter had a large interface buffer, which was well above the bandwidth-delay product in order to avoid congestion losses. The buffers were configured as a flat drop-tail queues without packet differentiation and as a priority queues (PriQ).

The following three transport methods were investigated.

- "SVC" – The H.264/SVC stream was transmitted over User Datagram Protocol (UDP) without any differentiation based on layered video coding.
- "SVC/PriQ" – Three different DiffServ codepoints (DSCP) were used in marking the packets, with DSCP "0" representing the base layer. Support for the traffic prioritization (PriQ), no notification of packet drops and connection loss.
- "SVC/MAL" – The MAL was added to the previous method in which the packets went through a special treatment in queue management, channel state information notifications.

The Fig. 7, represents the timeline of packet arrivals and packet drops in the defined scenario for each of the listed transmission method. The upper dots represents the arrived video packets for each video layer and the highlighted dots below illustrates the dropped packets. As seen from Fig. 7 the SVC and SVC/PriQ methods perform quite similarly in terms of packet drops, whereas SVC/MAL provides significantly less packet drops in overall due to the bit rate adaptation. The improved channel utilization with SVC/MAL can be seen also from the Fig. 8, which illustrates the throughput of SVC base layer. In the Fig. 8, the throughput is illustrated with discrete peak values and the 4 per moving average. As seen from the figure, the MAL clearly recovers better from connection loss in sense of delivering most important i.e. layer 0 packets instantly when connection is restored. In [29] we have also compared the MAL prioritization with constant bitrate traffic and simple priority queues. Although the plain priority queues show only a little advantage compared to the non-prioritizing transmission in the delivery of highest priority packets, it can be assumed that the use of even simple priority queues will be more useful

15

than the non-prioritizing transmission if there is more than just one video transmission. The MAL shows some improvement already with single video transmission and combining the short-term MAL prioritization with long-term fuzzy logic based application controlling, the efficiency of video transmission can improve increasingly with high traffic load.

5   Conclusions

In this paper we presented a novel PHOENIX architecture, which enhances the cross-layer approach for multimedia transmission in all-IP environment. Our cross-layer architecture proposal includes the complete transmission chain from application layer source coding to wired and wireless channel models, required cross-layer signaling mechanisms and full network functionality. For cross-layer control information delivery, we have proposed several different protocol solutions which are optimal in terms of network overhead caused by them. At the protocol level we proposed several new enhancements for standard solutions, such as link layer partial checksum, dynamic modifications for transport level checksum and enhanced header compression mechanism for optimizing the adaptive video transmission and for efficient transmitting of the cross-layer information.

In order to deepen the PHOENIX cross-layer approach, we specified cross-layer communication mechanisms in the case of scalable video coding. The proposed mechanisms utilize the cross-layer information introduced by the PHOENIX architecture. For the SVC cross-layer architecture we defined the fuzzy logic based application controller mechanism, which is able to maintain efficiently the data-rate and the perceived video quality for the video stream as high as possible. We introduced also a new RTP payload format for SVC, which supports the delivery of source significance and scalability information, and makes the adaptation of video stream possible also at the network level. Finally, we proposed the Multimedia Adaptation Layer bitstream prioritization mechanism for SVC in order to provide fast reaction for bitstream to channel and network state during transmission. Simulation results for long-term fuzzy controlling and short-term bitstream prioritization are presented and results indicate efficient and improved adaptivity of the scalable video in varying transmission conditions. The results indicate that gain can be achieved in the usage of system resources when proposed cross-layer mechanisms together with proposed signalling are used during the system design. Also, a usage of the proposed mechanisms and signals do not require from service providers to change their whole network infrastructure. This enables the usage of these mechanisms also together with existing network solutions.

16

Acknowledgment

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal* 27, pp. 379–423, 623–656, Jul.-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] S. B. Zahir Azami, P. Duhamel and O. Rioul, "Joint source-channel coding: panorama of methods," in *Proceedings of CNES workshop on data compression*, Toulouse, France, Nov. 1996.

[4] J. Hagenauer and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia," *Proceedings of the IEEE* 87 (10), Oct. 1999.

[5] K. Sayood and J.C. Borkenhagen, "Use of residual redundancy in the design or joint source/channel coders," *IEEE Transactions on Communications* 39, pp. 838–846, Jun. 1991.

[6] M. Park and D.J. Miller, "Joint source-channel decoding for variable-length encoded data by exact and approximate MAP sequence estimation," *IEEE Transactions on Communications* 48 (1), pp. 1–6, Jan. 2000.

[7] 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* 2, pp. 783–787, , New York, USA, Apr.-May 2002.

[8] M. van der Schaar, S. Krishnamachari, S. Choi and X. Xu, "Adaptive Cross-Layer Protection Strategies for Robust Scalable Video Transmission Over 802.11 WLANs," *IEEE J. Select. Areas Commun.* 21, pp.1752–1763, Dec. 2003.

[9] M. van der Schaar and S. Shankar N "Cross-Layer Wireless Multimedia Transmission: Challenges, Principles, and New Paradigms," *IEEE Wireless Communications* 12 (4), pp. 50–58, Aug. 2005.

[10] S. Mérigeault and C.Lamy, "Concepts for exchanging extra information between protocol layers transparently for the standard protocol stack," in *Proceedings of IEEE ICT'03*, Tahiti, French Polynesia, Feb. 23–Mar. 1st. 2003.

[11] J.-R. Ohm, "Advances in Scalable Video Coding," *Proceedings of IEEE* 93 (1), pp. 332–344, Jan. 2005.

[12] T. Wiegand, G. Sullivan, J. Reichel, H. Schwartz, and M. Wien, "Joint Draft 7 of SVC Amendment," *Joint Video Team (JVT)*, JVT-T201, Klagenfurt, Austria, July 2006.

[13] U. Horn, K. Stuhlmüller, M. Link, and B. Girod, "Robust Internet video transmission based on scalable coding and unequal error protection," *Signal Processing: Image Communication* 15, pp. 77–94, Sept. 1999.

[14] M. van der Schaar and H. Radha, "Unequal Packet Loss Resilience for Fine-Granular-Scalability Video," *IEEE Transactions on Multimedia* 3 (4) , pp. 381–394, Dec. 2001.

[15] J. Shin, J.W. Kim, and C.-C. Jay Kuo, "Quality-of-Service Mapping Mechanism for Packet Video in Differentiated Services Network," *IEEE Trans. on Multimedia* 3 (2), pp. 219–231, June 2001.

[16] F. Yang, Q. Zhang, W. Zhu, and Y.-Q. Zhang, "End-to-End TCP-Friendly Streaming Protocol and Bit Allocation for Scalable Video Over Wireless Internet," *IEEE Journal on Sel. Areas in Comm.* 22 (4), pp. 777–790, May 2004.

[17] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "End-to-End QoS for Video Delivery Over Wireless Internet," *Proceedings of IEEE* 93 (1), pp. 123–134, Jan. 2005.

[18] M. Conti, G. Maselli, G. Turi and S. Giordano, "Cross-Layering in Mobile Ad Hoc Network Design," *IEEE Computer* 37 (2), pp. 48–51, Feb. 2004.

[19] L-A. Larzon, M. Degermark, S. Pink, L-E. Jonsson and G. Fairhust, "The Lightweight User Datagram Protocol (UDP-Lite)," IETF RFC 3828, Jul. 2004.

[20] E. Kohler, M. Handley and S. Floyd, "Datagram Congestion Control Protocol," IETF RFC 4340, Mar. 2006.

[21] C. Bormann(editor), "RObust Header Compression (ROHC): framework and four profiles: RTP, UDP, EPS, and uncompressed," IETF RFC 3095, July 2001.

[22] M.G. Martini, M. Mazzotti, C. Lamy-Bergot, J. Huusko, and P. Amon, "Content Adaptive Network Aware Joint Optimization of Wireless Video Transmission," to be published in *IEEE Communication Magazine*, Jan. 2007.

[23] C. Lamy-Bergot et al., "Joint Optimization of Multimedia Transmission over an IP wired/wireless link," in *Proceedings of EuMob Symposium 2006*, Alghero, Italy, Sept. 2006.

[24] C. Bergeron and C. Lamy-Bergot, "Soft-input decoding of variable-length codes applied to the H.264 standard," in *Proceedings of IEEE MMSP'04*, Sienna, Italy, Sept. 2004.

[25] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on CSVT* 13 (7), pp. 560–576, July 2003.

[26] T. Rathgen, P. Amon, A. Hutter, "Storage and RTP transport of scalable video," in *Proceedings of International Workshop on Systems, Signals and Image Processing (IWSSIP 2005)*, Chalkida, Greece, Sept. 2005.

[27] S. Wenger, M. M. Hannuksela, T. Stockhammer, M. Westerlund, D. Singer, "RTP Payload Format for H.264 Video," IETF RFC 3984, 2005.

[28] I. Amonou, N. Cammas, S. Kervadec, S. Pateux "Layered quality optimization for SVM," *ISO/IEC JTC 1/SC 29/WG 11*, M11704, Hong Kong, China, Jan. 2005.

[29] J. Vehkaperä, K. Pentikousis, M. Majanen, J. Huusko, and J. Peltola, "Improving Scalable Video Delivery with Cross-Layer Design," in *Proceedings of 1st MediaWiN workshop*, Athens, Greece, May 2006.

Table 1
Solutions for cross-layer control signal transmission mechanisms

Control signal	Suitable mechanism	Estimated overhead and comments
SSI	IPv6 Hop-by-Hop and RTP for SVC	IPv6 overhead of few KBps and high synchronization with the stream
SRI	IPv6 Hop-by-Hop or Destination	Overhead of few KBps and high synchronization with the stream
CSI	ICMPv6	Overhead of less than 1 Kbyte/sec for CSI updating period up to 50 ms; slight synchronization with the video data.
NSI	RTCP	Low overhead with suitable frequency of 200 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). These ontrol signals should in practice be sent only when the wireless receiver is also the data traffic destination.
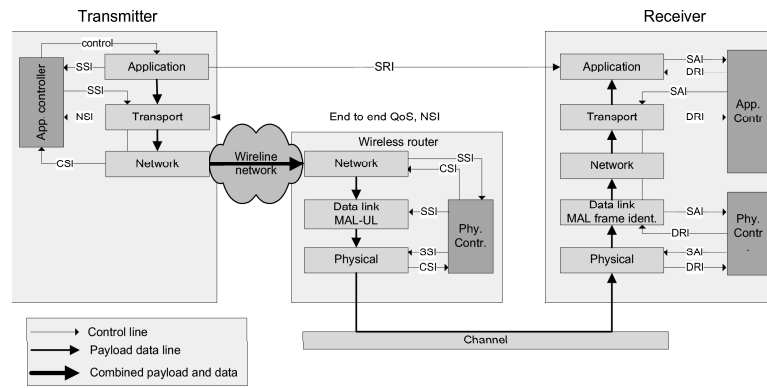
Fig. 1. Phoenix Cross-layer Architecture and Control Signals
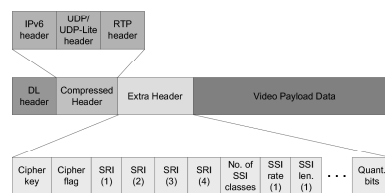
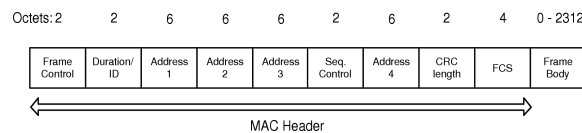Fig. 2. Extra header structure for cross-layer information

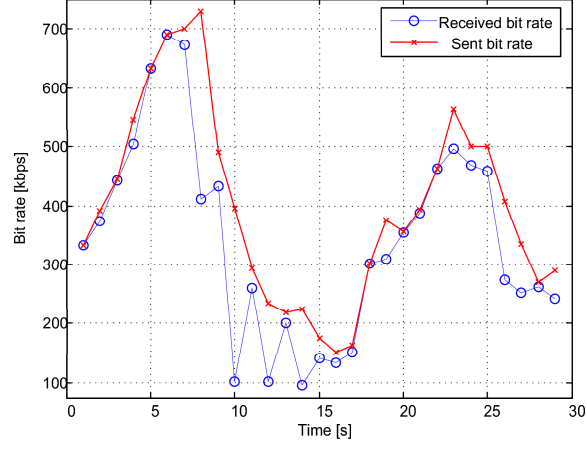Fig. 3. Data Frame Format for Multimedia

20

Fig. 4. Received and sent (selected by the controller) bitrate versus time achieved using the PHOENIX architecture.
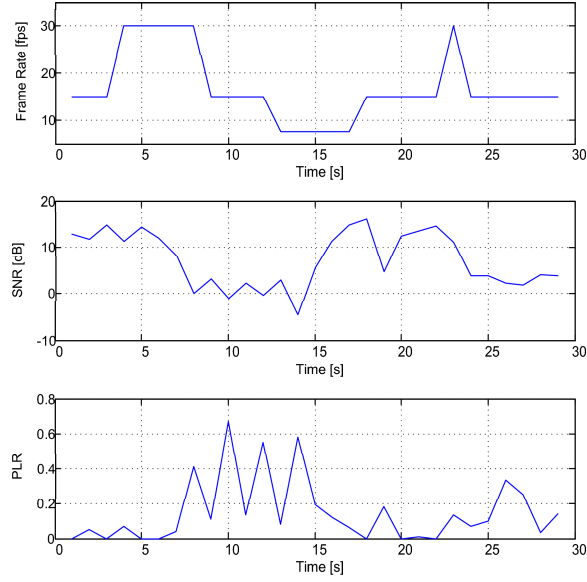

Fig. 5. Frame rate (selected by the controller), SNR, and PLR versus time achieved using the PHOENIX architecture.
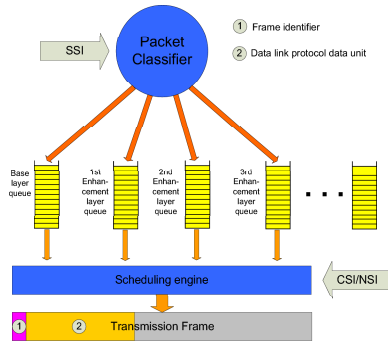

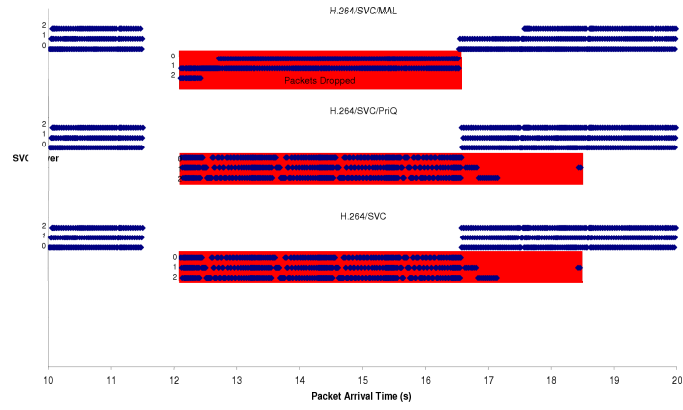Fig. 6. Multimedia Adaptation Layer uplink scheduler

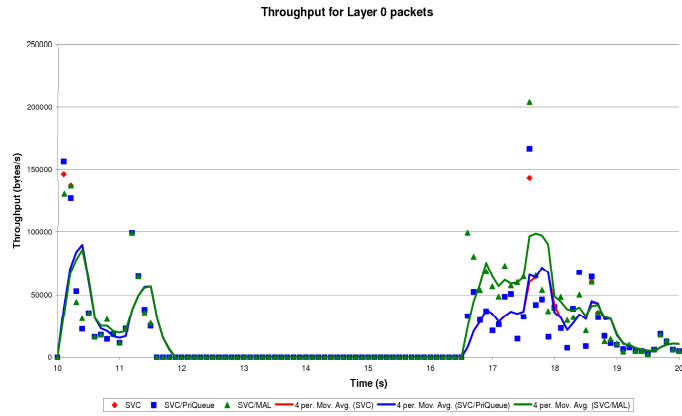Fig. 7. Timeline of packet arrivals and drops with and without MAL

Fig. 8. Throughput for SVC base layer with and without MAL