Joint Scalable Video Coding and Packet Prioritization for Video Streaming over IP/802.11e Heterogeneous Networks

Thomas Pliakas  
Dept. of Information & Communication Systems  
University of Aegean  
Karlovasi, Samos, Greece  
tpliakas@aegean.gr

George Kormentzas  
Dept. of Information & Communication Systems  
University of Aegean  
Karlovasi, Samos, Greece  
gkorm@aegean.gr

Sofia Tsekeridou  
Athens Information Technology - AIT  
Building B7, 19.5 km Markopoulo Ave  
Peania, Athens, Greece  
sots@ait.edu.gr

ABSTRACT
The paper discusses scalable video streaming traffic delivery over heterogeneous DiffServ/WLAN networks. A prototype architecture is proposed and further validated that explores the joint use of packet prioritization and scalable video coding (SVC) together with the appropriate mapping of 802.11e access categories to the DiffServ traffic classes. A complete set of simulation scenarios, involving four different video sequences using the scalable extension of H.264/MPEG-4 AVC, demonstrates the quality gains of both scalable video coding and prioritized packetization.

Keywords
802.11e, DiffServ, End-to-End QoS, H.264/MPEG-4 AVC, Packet Prioritization, SVC, WLAN

1. INTRODUCTION
IP technology seems to be able to resolve the inter-working amongst the diverse fixed core and wireless access technologies. At the network level, the end-to-end QoS provision could be established through the appropriate mapping amongst the QoS traffic classes/services supported by the contributing underlying networking technologies [11] [12]. A QoS cross layer architecture based on error resilience features of H.264/MPEG-4 AVC can be applied for further improvements on end-to-end QoS. Building on this background, this work involves a DiffServ-aware IP core network and a 802.11e access network and examines end-to-end QoS issues regarding scalable video streaming and prioritized packetization based on data partitioning (DP) for delivering multimedia traffic across fixed and wireless network domains.

The Differentiated Services (DiffServ) [4] approach proposed by IETF supports (based on the DiffServ Code Point (DSCP) [8] field of the IP header) two different services, the Expedited Forwarding (EF) that offers low packet loss and low delay/jitter and the Assured Forwarding (AF), which provides better QoS guarantees than the best-effort service. Differences amongst AF services imply that a higher QoS AF class will give a better performance (faster delivery, lower loss probability) than a lower AF class.

The 802.11e [3] standard addresses the issue of QoS support in wireless LANs. The MAC protocol of 802.11e supports multiple access categories (ACs). A higher priority access category has a smaller minimum contention window thus has a higher probability to access the channel. Different access categories can have a different maximum contention window and inter-frame spacing interval (IFS). The 802.11e defines four access categories; AC3 corresponds to the highest access priority, and AC0 to the lowest.

The basic coding scheme for achieving a wide range of spatio-temporal and quality scalability is scalable video. For Signal-to-Noise Ratio (SNR) scalability the most appropriate technique for video delivery over heterogeneous networks, is the scalable extension of H.264/MPEG-4 AVC [13]. In order to support fine-granular SNR scalability, progressive refinement (PR) slices have been introduced in the scalable extension of H.264 [14]. A base representation of the input frames of each layer is obtained by transform coding similar to H.264 [18]. The corresponding Network Abstraction Layer (NAL) units (containing motion information and texture data) of the base layer are compatible with the single layer H.264/MPEG-4 AVC. Furthermore, by employing data partitioning, the H.264 encoder partitions the compressed data in separate units of different importance. The packets, with assigned priority, are sent to a QoS-aware network to receive different forwarding treatments. Mapping these prioritized packets to different QoS levels causes them to experience different packet loss rates with this differential forwarding mechanism. The quality of the base representation can be improved by an additional coding of the so-called PR slices. The corresponding NAL units can be arbitrarily truncated in order to support fine granular quality scalability or flexible bit-rate adaptation.

To address end-to-end QoS problem scalable video streaming traffic delivery over a heterogeneous IP/802.11e network, the paper proposes and validates through a number of NS2-based simulation scenarios an architecture that explores the joint use of packet prioritization and scalable video coding together with the appropriate mapping of 802.11e access categories to the DiffServ traffic classes. This work extends previous authors’ papers [11] [12] dealing with joint scalable-technologies.
video coding and packet prioritization over IP/UMTS and IP/DVB heterogeneous networks.

The rest of the paper is organized as follows. In Section 2, the proposed scalable video coding techniques and prioritization framework for providing QoS guarantees for scalable video streaming traffic delivery over a heterogeneous DiffServ/WLAN network is presented. In Section 3, we demonstrate how video-streaming applications can benefit from the use of the proposed architecture. Finally, Section 4 draws the conclusions and discusses directions for further work and improvements.

2. PROPOSED ARCHITECTURE

The proposed architecture integrates the concepts of scalable video streaming, prioritized packetization based on the H.264 data partitioning features and mapping DiffServ classes to MAC differentiation of 802.11e. The proposed architecture is depicted in Figure 1. It consists of three key components: (1) Scalable video encoding (Scalable extension of H.264/MPEG-4 AVC), (2) prioritized packetization according based on data partitioning, and (3) DiffServ/802.11e class mapping mechanism in order to assure the optimal differentiation and to achieve QoS continuity of scalable video streaming traffic delivery over DiffServ and 802.11e network domains. Each one of these components is discussed in detail in the following subsections.

2.1 Scalable Video Coding

Scalable Video Coding should meet a number of requirements in order to be suitable for multimedia streaming applications. For efficient utilization of available bandwidth, the compression performance must be high. Also, the computational complexity of the codec must be kept low to allow cost efficient and real time implementations. When compared against other scalable video coding schemes, the fine granular scalability method is outstanding due to its ability to adapt to changing network conditions more accurately.

2.1.1 Scalable Extension of H.264/MPEG-4 AVC

In order to provide FGS scalability, a picture must be represented by an H.264/AVC compatible base representation layer and one or more FGS enhancement representations, which demonstrate the residual between the original predic-

![Figure 1: Overall Architecture](image)

For the encoding of the enhancement representation layers a new slice called Progressive Refinement (PR) has been introduced. In order to provide quality enhancement layer NAL units that can be truncated at any arbitrary point, the coding order of transform coefficient levels has been modified for the progressive refinement slices. The transform coefficient blocks are scanned in several paths, and in each path only a few coding symbols for a transform coefficient block are coded [16].

2.2 Prioritized Packetization

We define two groups of priority policies, one for BL and one for EL. These policies are used from the Edge Router of the DiffServ-aware underlying network to map the packets to the appropriate traffic classes. The packetization process can affect the efficiency as well as the error resiliency of video streaming.

For the BL, at the Video Coding Layer (VCL), an additional type of slice, besides the three partitions (A, B, and C) obtained when DP is enabled, that represents Instantaneous Decoding Refresh (IDR) pictures. The IDR access units contain information that cannot be included into the three partitions, like the intra-picture (coded picture that can be decoded without needing information from previous pictures) where no data partitioning can be applied.

The order in which the slice units are sent is constant. The first transmitted slice units transmitted contain the Packet Set Concept (PSC) information, such as picture size, display window, optional coding modes employed, macroblock allocation map, etc. This higher-layer meta information should be sent reliably, asynchronously, and before transmitting video slices.

The next transmitted slice units contain the IDR picture. Since IDR frames may contain only I slices without data partitioning, they are usually sent at the start of video sequences (just after the PSC). The slice units following the IDR frames contain one of the three partitions (A, B, or C).

The NAL is responsible for the encapsulation of the coded slices into transport entities of the network. Each NAL unit (NALU) could be considered as a packet that contains an integer number of bytes, including a header and a payload. The header specifies the NALU type, and the payload contains the related data. The most important field of the NAL header is the NalRefIdc (NRI) field [7].

The NRI contains two bits that indicate the priority of the NALU payload, where 11 is the highest transport priority, followed by 10, then by 01, and finally, 00 is the lowest. Accordingly, the incoming VCL layer slices are differentiated and encapsulated into NALUs by enabling the NRI field in the NAL header. Table 1 depicts the relation between the type of the BL content and the corresponding DiffServ classes.

The first digit of the AF class indicates forwarding priority and the second indicates the packet drop precedence.

<table>
<thead>
<tr>
<th>NRI</th>
<th>AF Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>7</td>
</tr>
<tr>
<td>01</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 1: DiffServ Classes allocation for NRI

<table>
<thead>
<tr>
<th>DiffServ Classes</th>
<th>Slice Type</th>
<th>NRI Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF</td>
<td>PSC</td>
<td>11</td>
</tr>
<tr>
<td>AF11</td>
<td>IDR A</td>
<td>10</td>
</tr>
<tr>
<td>AF12</td>
<td>B C</td>
<td>01, 00</td>
</tr>
</tbody>
</table>

Table 2: DiffServ/802.11e classes coupling

<table>
<thead>
<tr>
<th>Traffic Class</th>
<th>DiffServ Classes</th>
<th>AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>EF</td>
<td>3</td>
</tr>
<tr>
<td>Class 2</td>
<td>AF11</td>
<td>2</td>
</tr>
<tr>
<td>Class 3</td>
<td>AF12</td>
<td>1</td>
</tr>
<tr>
<td>Class 4</td>
<td>Best Effort</td>
<td>0</td>
</tr>
</tbody>
</table>

The PSC packets obtain the highest priority. Furthermore, as information carried in both partition A and IDR are essential for decoding an entire video frame, it is important to give these slices more priority than partition B and C. Based on these rules, the NAL layer marks the different NALUs.

2.3 DiffServ/802.11e QoS Classes Coupling

In order to integrate the 802.11 network domain with the core network domain, and to achieve QoS consistency across the DiffServ IP and 802.11e network, we have to map 802.11e access categories to predefined DiffServ classes. A direct mapping approach as proposed by [10] is adopted. Table 2 shows the mapping of the predefined DiffServ classes according to the DiffServ specification, where the first digit of the AF class indicates forwarding priority and the second indicates the packet drop precedence, and the 802.11e access categories for the proposed mapping approach.

The packets, with assigned priority, are sent to the DiffServ network to receive different forwarding treatments. Mapping these prioritized packets to different QoS DS levels causes them to experience different packet loss rates with this differential forwarding mechanism. In addition to the prioritized dropping performed by DiffServ routers, traffic policing can be carried out at intermediate video gateways (between different network domains), using packet filtering. When the IP packets are encapsulated in MAC frames, each frame should be allocated to a priority queue, or an access category.

3. FRAMEWORK EVALUATION

This section evaluates the performance of the proposed framework through a set of simulations. A NS-2 based simulation environment with the appropriate extensions [1] for simulating 802.11e WLANs is adopted.

Four YUV QCIF 4:2:0 color video sequences consisting of 300 to 2000 frames and coded at 30 frames per second are used as video sources. Each group of pictures (GOP) is structured as IBPBBPBB and contains 36 frames, and the maximum UDP packet size is at 1024 bytes (payload only). The scalable extension of H.264/MPEG-4 AVC encoder/decoder provided by [2] is used for encoding YUV sequences. The video frames are then encapsulated into RTP packets using a simple packetization scheme [15] (by one-frame-one-packet policy). The size of each RTP packet is maximally bounded to 1024 bytes. The generated video packets are delivered through the DiffServ at the form of UDP/IP protocol stack. The 802.11b is employed for the physical layer, which provides four different physical rates. In our simulation, the physical rates are fixed to 11 Mbps for data and 2Mbps for control packets. Table 3 depicts the MAC Parameters for the simulations.

Additionally, the streaming node station generates background traffic (500 kbps) using constant bit rate (CBR) traffic over User Datagram Protocol (UDP). This allows us to increase the virtual collisions at the server’s MAC layer. Furthermore, we include five wireless stations where each station generates 300 kbps of data using CBR traffic in order to overload the wireless network.

A unique sequence number, the departure and arrival timestamps, and the type of payload that identify each packet. When a packet does not reach the destination, it is counted as a lost packet. Furthermore, not only the actual loss is important for the perceived video quality, but also the delay of packets/frames and the variation of the delay, usually referred to as packet/frame jitter. The packet/frame jitter can be addressed by the so called play-out buffers. These buffers have the purpose of absorbing the jitter introduced by the network delivery delays. It is obvious that a big enough play-out buffer can compensate any amount of jitter. There are many proposed techniques in order to develop efficient and optimized play-out buffer, dealing with this particular trade-off. These techniques are not within the scope of the described testbed. For our experiments the play-out buffer is set to 1000 milliseconds.

In order to measure the improvements in video quality by employing H.264/MPEG-4 AVC, we use the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity (SSIM) [17] metrics. PSNR is one of the most widespread objective metric for quality assessment and is derived from the Mean Square Error (MSE) metric, which is one of the most commonly used objective metrics to assess the application level QoS of video transmissions [5].

Let’s consider that the video sequence is represented by \( v(n, x, y) \) and \( v_n(x, y) \), where \( n \) is the frame index and \( x \) and \( y \) are the spatial coordinates. The average PSNR of the decoded video sequence among frames at indices between \( n_1 \) and \( n_2 \) is given by the following equation:

\[
PSNR = 10 \log_{10} \frac{V^2}{MSE}
\]  

where \( V \) denotes the maximum greyscale value of the luminance. The average \( MSE \) of the decoded video sequence among frames at indices between \( n_1 \) and \( n_2 \) is given by:

![Figure 2: Simulation Setup](Image)
AIFS
Queue length
50
CW
70
15
1023
31
50
1023
50
4
Max Retry limit
31
50
7

Table 3: 802.11 MAC Parameters

<table>
<thead>
<tr>
<th>Access Category</th>
<th>AIFS</th>
<th>CW_{min}</th>
<th>CW_{max}</th>
<th>Queue length</th>
<th>Max Retry limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC3</td>
<td>50</td>
<td>7</td>
<td>15</td>
<td>50</td>
<td>8</td>
</tr>
<tr>
<td>AC2</td>
<td>50</td>
<td>15</td>
<td>31</td>
<td>50</td>
<td>8</td>
</tr>
<tr>
<td>AC1</td>
<td>50</td>
<td>31</td>
<td>1023</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>AC0</td>
<td>70</td>
<td>31</td>
<td>1023</td>
<td>50</td>
<td>4</td>
</tr>
</tbody>
</table>

The best-effort network, we use a network that implements the best-effort model. The mapping of packets is based on Table 1. The DiffServ routers implement WRED queue management. In this scenario, according to Figure 4, the overall PSNR is better than without using prioritization. The $P_{avg}$ value is 31.054dB. Figure 5 depicts the SSIM metric of both scenarios (BE and QoS-enabled networks) for foreman video sequence.

Figure 3: Scalable video transmission over best-effort networks

Figure 4: Scalable video transmission over DiffServ/802.11e Heterogeneous Network

We repeat the same measurement for four different YUV video sequences consisting of 300 to 2000 frames. For all the scenarios, we consider the simple but efficient error concealment scheme described in the previous section. The average PSNR and SSIM for the above scenarios are shown in Table 4, where:

- in Scenario 1 we transmit scalable H.264/MPEG-4

$MSE = \frac{1}{XY(n_2 - n_1 + 1)} \sum_{n=n_1}^{n_2} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} M^2$ \hspace{1cm} (2)

where $M$ is defined as:

$M = [v(x, y, n) - v_{av}(x, y, n)]$ \hspace{1cm} (3)

Note that, the PSNR and MSE are well-defined only for luminance values. As it mentioned in [5], the Human Visual System (HVS) is much more sensitive to the sharpness of the luminance component than that of the chrominance component, therefore, we consider only the luminance PSNR.

SSIM is a Full Reference Objective Metric [6] for measuring the structural similarity between two image sequences exploiting the general principle that the main function of the human visual system is the extraction of structural information from the viewing field. If $v_1$ and $v_2$ are two video signals, then the SSIM is defined as:

$SSIM(v_1, v_2) = \frac{(2\mu_{v_1}\mu_{v_2} + C_1)(2\sigma_{v_1v_2} + C_2)}{(\mu_{v_1}^2 + \mu_{v_2}^2 + C_1)(\sigma_{v_1}^2 + \sigma_{v_2}^2 + C_2)}$ \hspace{1cm} (4)

where $\mu_{v_1}$, $\mu_{v_2}$, $\sigma_{v_1}$, $\sigma_{v_2}$, $\sigma_{v_1v_2}$ are the mean of $v_1$ the mean of $v_2$, the variance of $v_1$, the variance of $v_2$ and the covariance of $v_1$ and $v_2$. The constants $C_1$ and $C_2$ are defined as:

$C_1 = (K_1L)^2$ \hspace{1cm} (5)

$C_2 = (K_2L)^2$ \hspace{1cm} (6)

where $L$ is the dynamic range of pixel values and $K_1 = 0.01$ and $K_2 = 0.03$, respectively. [22] defines the values of $K_1$ and $K_2$.

At the first scenario, we examine the transmission of H.264 scalable video streams consisting of two layers. The BL is encoder at 256 Kbps, while the EL is encoded at 512 Kbps. As video source is used the Foreman YUV QCIF video sequence (176x144) consisting of 400 frames. The underlying network for the first measurement is a simple Best Effort network, like Internet, without implementing any QoS model for guarantee end-to-end video quality. The video frame is sent every 33 ms for 30 fps video. Figure 3 shows the PSNR graph for the experimental scenario described above. The Y axis represents the PSNR value in dB while the X axis represents the frame number of video sequence.

As one may observe from Figure 3, during severe network congestion caused by interference by background traffic, the PSNR values are between 10dB and 12dB. The average value of PSNR, $P_{avg}$ is 29.038dB. Note that, a frame is counted as lost also, when it arrives later than its defined playback time.

We repeat the same measurement, but instead of using a best-effort network, we use a network that implements the
mapping of 802.11e access categories to the DiffServ traffic classes. The proposed prioritization scheme in conjunction with the proposed DiffSer/802.11e classes coupling have improvements in the overall quality of the received video, by isolating the losses and the delays to packets carrying less important partitions.

5. REFERENCES

### Table 4: Average PSNR/SSIM for scalable H.264/MPEG-4 AVC video streams

<table>
<thead>
<tr>
<th>Video</th>
<th>Frame</th>
<th>Scen.1</th>
<th>Scen.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highway</td>
<td>2000</td>
<td>28.339</td>
<td>9762</td>
</tr>
<tr>
<td>Mother</td>
<td>961</td>
<td>28.892</td>
<td>31.021</td>
</tr>
<tr>
<td>Salesman</td>
<td>444</td>
<td>28.523</td>
<td>31.210</td>
</tr>
<tr>
<td>Foreman</td>
<td>400</td>
<td>29.038</td>
<td>31.054</td>
</tr>
</tbody>
</table>

AVC video stream in a best effort network.

- in Scenario 2 we transmit scalable H.264/MPEG-4 AVC video stream over a DiffServ/802.11e heterogeneous network.

As it seems in Table 4, the proposed prioritization scheme improves the overall quality of the received video. By isolating the losses and the delays to packets that contain B and C partitions we can achieve significant gains to video quality. By distributing the traffic to all traffic classes, we can achieve equal or even better video quality, in the lowest price, by sending lowest traffic to the cost effective EF/AC3 traffic. From the network provider perspective, the utilization of the network is more efficient, by serving more users, at the level of quality they pay.

4. CONCLUSIONS
Nowadays, continuous media applications over heterogeneous all-IP networks, such as video streaming and video-conferencing, become very popular. Several approaches have been proposed in order to address the end-to-end QoS both from the network perspective, like DiffServ and 802.11e access categories, and from the application perspective, like scalable video coding and packetized prioritization mechanisms. The paper addresses the end-to-end QoS problem of scalable video streaming traffic delivery over a heterogeneous DiffServ/802.11enetwork. It proposes and validates a number of NS2-based simulation scenarios a framework that explores the joint use of packet prioritization and scalable video coding, by evaluating scalable extension of H.264/MPEG-4 AVC, together with the appropriate

![Figure 5: SSIM measurements of scalable video transmission over DiffServ/802.11e Heterogeneous Networks](image-url)