

Improving last-hop multicast streaming video over 802.11

Igor Kozintsev, Jeff McVeigh
Intel Corporation

Abstract

In this paper we address the problem of robust real-time video streaming using multicast over wireless LANs. This scenario represents a significant challenge for existing 802.11 networks because it requires transmission over multiple unreliable channels to heterogeneous receivers with potentially different connection bit rates and very limited feedback information to the sender. The wireless channel is modeled as an equivalent multiple access packet erasure channel at the network level. We formulate the problem as an optimization of a regret criterion across the space of all users with the objective of maximizing the aggregate received quality. We formulate a proposed system solution that efficiently combines scalable source coding based on client rate adaptation coupled with channel packet erasure coding. Further work is required to validate the solution based on real-world wireless conditions in home and enterprise environments.

1 Introduction

The rapid adoption of broadband wireless local area networks (WLAN), specifically IEEE 802.11a/b/g, has generated significant demand for the wireless delivery of high quality video within the home and enterprise environments. Most research and commercial solutions to date have focused on the unicast wireless scenario, which attempts to deliver robust content to a single client via an access point within the WLAN. While the unicast case is still not a fully solved problem, our interest is on a "last hop" multicast, or broadcast, scenario where a large number of clients (potentially hundreds) attempt to wirelessly stream the same content from a single access point. Example applications of this scenario include the online gaming, redistribution of television content to various rooms within a home and the broadcast of a camera feed to audience members at a sporting event, conference, or lecture, where each client may contain local storage to enable limited control on the content. While we seek a solution to this problem for the 802.11 family of

WLANs, the concepts presented here are applicable to other packet wireless networks that support rate adaptation.

The key problems related to video streaming over WLANs are dynamic fluctuations in channel quality and higher bit error rate, and correspondingly higher packet error rate, compared to wired networks. These characteristics are due to fading, interference and multi-path effects between the access point and client. Dynamic rate adaptation is employed in 802.11 to adjust the modulation scheme to achieve higher throughput based on channel characteristics. This implies that multiple clients might have significantly different link rates (e.g., 1, 2, 5.5 or 11 Mbps for 802.11b), which further complicates the multicast scenario compared to a single-hop wire-line scenario. While rate adaptation is effective at decreasing packet error rate in the presence of degraded channel quality, the throughput of other clients is negatively affected when some clients are connected at lower link rates due to the shared channel characteristic of the 802.11. Figure 1 illustrates a hypothetical last-hop multicast environment, where compressed media data is streamed to devices with different link rates.

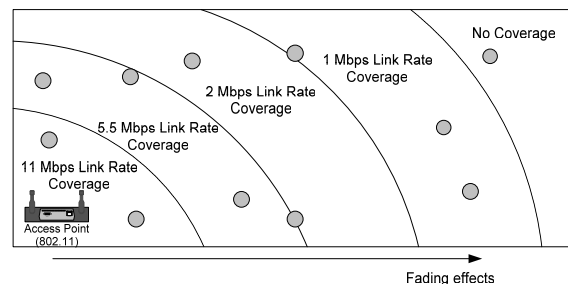


Figure 1. Last-hop scenario with multicast streaming to multiple clients with different channel characteristics. Clients are represented by grey circles and the concentric arcs provide the bounding region for hypothetical link rates based on rate adaptation.

The use of scalable media coding techniques has been

previously proposed to address client heterogeneity and ensure graceful quality degradation for the multicast scenario [8, 17, 19]. However, these hierarchical representations are sensitive to the position of packet loss, i.e., the received quality is a function of which packets are erased. Priority Encoding Transmission was proposed in [7] to eliminate this dependency, which allows for different protection of scalable layers based on their importance. Several proposals have been offered for the solution of the optimal amount of parity bits to protect each layer [12, 16, 14]. A near-optimal solution was described in [14], termed the Multiple Description Coding using Forward Error Correction (MDFEC), that can be solved using a Lagrangian optimization based $O(N)$ complexity algorithm.

All known previous multicast streaming work has assumed that all packets are transmitted within a wireless network using the same modulation scheme or link data rate. In these formulations, the individual clients may receive varying throughput based on their channel characteristics, but the probability of receiving each packet transmitted from the access point is assumed to be constant for each specific client. Our approach differs in that we explicitly include the effect of dynamic rate adaptation into our problem formulation.

The major contribution of this paper is the incorporation of rate adaptation into the previously developed MDFEC algorithm. This is coupled with a systems approach that utilizes channel packet erasure coding and quality of service provisioning to achieve robust multicast streaming for a very large number of heterogeneous clients.

The rest of the paper is organized as follows. Section 2 provides an overview of scalable video coding and develops a model between compression ratio and perceived video quality, which is used to solve for the optimal aggregate quality across all multicast clients. Section 3 develops a packet erasure model of the 802.11 channel and provides details on appropriate channel coding techniques. Section 4 formulates our problem in terms of a minimum regret criterion. Section 5 discusses the system-level interaction for improving multicast streaming, and Section 6 provides concluding remarks and areas for future investigation.

2 Scalable video coding

The majority of current digital video applications operate within fixed transmission channel or storage capacity constraints. An illustrative example is high-definition digital television as defined by ATSC [1], where compressed source content is modulated using 8-VSB and broadcast terrestrially within a fixed 6 MHz bandwidth. Since the channel bandwidth is dedicated and known a priori, the source content is optimally encoded by compressing to a single target bit rate that fits within the channel bandwidth

at a specific combination of spatial resolution and temporal sampling frequency (e.g., up to 19 Mbps, 1280x720 resolution, 60 frames per second). This process is referred to as *single-layer* or non-scalable coding since a single bitstream representation is obtained that satisfies the specific transmission channel constraints and client capabilities. Single-layered coding is also utilized effectively in digital cable, direct-broadcast satellite, and DVD applications, which are all based on the MPEG-2 video (Main Profile) standard [2].

Internet-based streaming video, conversely, operates in an environment where the effective transmission channel rate can vary dynamically and dramatically depending on network load and last-mile connectivity. These constraints preclude the use of a single, fixed bitstream representation for all clients. Bitstream switching is typically employed for point-to-point and multicast Internet streaming applications [11], where multiple, independent bitstreams are compressed and stored at the server with each bitstream tuned to particular connection data rate, channel loss characteristics, and client processing and display capabilities. In the unicast scenario, the client monitors the packet transmission rate and loss statistics and communicates with the server to send the representation that will yield the highest quality over the current channel. This technique has been extended to multicast delivery, where the individual streams are assigned to different multicast addresses. Clients then subscribe to the multicast address that provides the minimum distortion for the client's channel characteristics. Due to the temporal dependency within each stream, switching can only occur at specific synchronization points, which limits the potential rate of adaptation.

While the bitstream switching technique has achieved commercial success for unicast and multicast Internet streaming, this solution is not applicable for our target last-hop multicast wireless scenario. Specifically, the wireless channel is shared between clients that may have dramatically different connection rates, the throughput for each client may vary instantaneously due to multi-path fading, interference and transmitter-receiver distance, and the server (which is assumed to be a home PC or home media server in our scenario) has relatively limited storage and processing capabilities compared to a dedicated streaming server and is likely incapable of encoding multiple independent streams in real-time for each potential client channel. *Scalable video coding* is well suited to our scenario as it makes effective use of the shared channel and can quickly adapt to channel characteristics without the delay associated with synchronization points required for bitstream switching techniques.

Scalable coding techniques typically consist of a base layer representation and a collection of one or more enhancement layers. The base layer is independently decodable, while the enhancement layers hierarchically depend

on the base layer and/or lower enhancement layers. The enhancement layers can provide scalability across one or more orthogonal vectors of perceived visual quality, namely: 1) *spatial scalability* for improved image resolution or detail, 2) *temporal scalability* for improved smoothness of motion, and 3) *SNR scalability* for improved image fidelity.

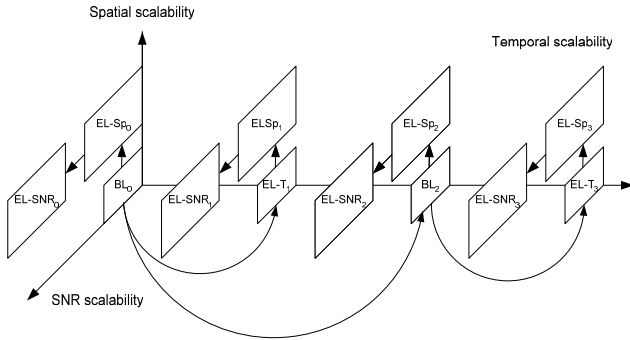


Figure 2. Example scalable bitstream representation.

Figure 2 illustrates an example scalable bitstream utilizing all three scalability vectors. The lowest bit rate, and hence lowest quality, representation is provided by the base layer (BL), shown as frames BL_0 and BL_2 in this simple example¹. Clients with higher throughput can obtain improved smoothness of motion through the addition of the temporal enhancement layer (frames $EL-T_1$ and $EL-T_3$). Similarly, improved image resolution is provided by the spatial enhancement layer ($EL-S_p$), and finally improved image fidelity by the SNR enhancement layer ($EL-SNR$). This example hierarchical representation can service clients with up to four separate throughputs with an aggregate bit rate equal to the sum of the base layer and three enhancement layers.

The use of scalable coding techniques is not without its liabilities. These include degradation in rate-distortion performance compared to single-layered techniques for a specific rate, increased encoder / decoder complexity, and the interdependence between layers, which requires appropriate source-channel coding techniques in error prone environments [15, 10]. The rate-distortion "penalty" is the most significant factor for the lack of commercial deployment of scalable coding techniques even for applications that do not operate within fixed channel environments.

It is not surprising that a scalable representation would yield inferior rate-distortion performance compared to a single-layer representation due to the inherent overhead and

¹This example is not intended to be definitive, but instead is used to illustrate the interaction between scalability vectors and inter-layer dependencies.

redundancies that must be present in the hierarchical representation. Figure 3 depicts the rate-distortion performance for a theoretical single-layer and scalable coding scheme. We assume that the scalable scheme uses a base layer technique equivalent to the single-layer representation, which yields equivalent distortion at the base layer rate ($R'_0 = R_0$). Beyond this rate, the scalable representation requires an increased bit rate to achieve equivalent distortion (e.g., $R'_1 > R_1$). The superiority for single-layered coding is obvious for applications that operate within a fixed channel. However, for this example where we assume clients with four different transmission channels, it is clear that the scalable coding scheme requires lower aggregate bit rate provided that $R'_3 < R_0 + R_1 + R_2 + R_3$.

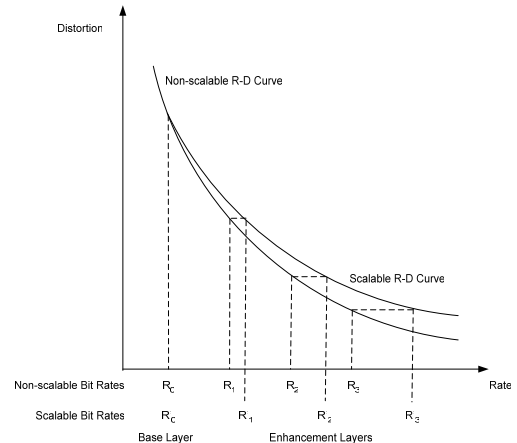


Figure 3. Rate-distortion comparison between theoretical single-layer and scalable coding schemes.

The reduction, or elimination, of the rate-distortion penalty is an area of active and challenging research. Recent advances have been obtained through the utilization of motion-compensated temporal filtering (MCTF), which is based on the lifting representation of a temporal subband filter bank [18]. MCTF provides improved coding efficiency in the wavelet domain in the presence of object motion. This work has gained the attention of the standards community, and forms the current basis of the MPEG-21 work item on Scalable Video Coding (SVC), which recently reached first Working Draft (WD 1.0) status. The current specification² calls for a base layer utilizing the recently defined H.264 / AVC standard [5], and a MCTF-based enhancement layer [6]. The H.264-compliant base layer provides compatibility with current and future single-layer systems that utilize this efficient codec, while the MCTF-based enhancement layers provide medium to fine-grain scalability across temporal, spatial and SNR vectors.

²Since SVC is still in the early definition stage, significant algorithmic changes are likely to occur prior to final standardization.

Since MPEG-21 SVC is anticipated to reach final standardization by ?????, we assume a scalable coding scheme similar to the current SVC definition. The objective then is to calculate the optimal bit allocation across the base and enhancement layers to achieve high-quality, robust multi-cast wireless streaming to a large number of clients within a specified service area. As we are interested in maximizing the aggregate perceived quality for all clients, we first construct a model of quality as a function of bit rate or, equivalently, compression ratio.

As a proxy for a true SVC implementation, we utilize an internally developed real-time H.264 implementation [13]. Although this is a non-scalable solution, we assume that an SVC encoder will be able to relatively closely track the operational rate-distortion curve of the H.264 implementation across a specified bit rate range (i.e., we assume that continued advances in the specification will yield a negligible rate-distortion penalty for the scalable coding capability).

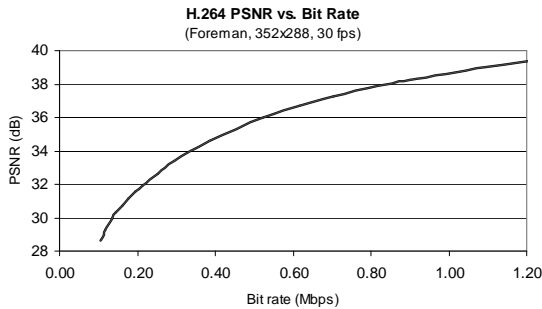


Figure 4. Peak signal-to-noise ratio versus bit rate for H.264 encoding of Foreman sequence (352x288 resolution, 30 frames per second).

Figure 4 shows a familiar rate-distortion plot of the H.264 encoder for the Foreman sequence, where distortion here is represented by the peak signal-to-noise ratio (PSNR), which is a logarithmic function of the mean-squared error between the reconstructed and original video frames. While PSNR is not directly correlated to perceived quality, it provides a simple to calculate quantitative measurement and is used extensively for comparisons between encoding algorithms; hence, we use it here for the basis of our quality model.

Since we are interested in the quality relationship across sequences with differing resolutions and frame rates, we similarly generated PSNR results for a variety of low, standard, and high definition sequences. These results were normalized with respect to compression ratio, where the compression ratio is defined as the ratio between the uncompressed source, assuming a 4:2:0 chrominance sampling,

and the compressed bit rate. The results are shown in Figure 5, which demonstrates the inverse relationship between compression ratio and bit rate.

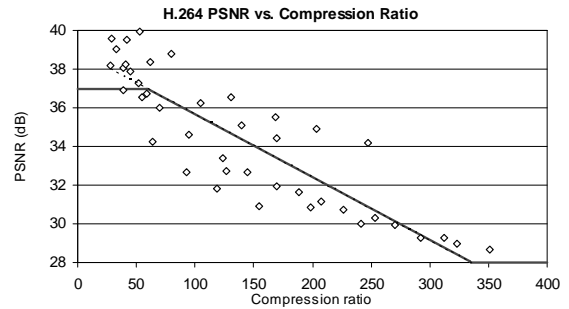


Figure 5. PSNR versus compression ratio for a variety of sequences encoding using an H.264 implementation.

Although PSNR vs. bit rate plots typically show a logarithmic relationship, this clustering over sequences with varying characteristics indicates that we can approximate with a linear relationship. We calculated a least-squares fit over this data, which is represented by the dotted diagonal line. The linear model can then be used to estimate the PSNR for a given bit rate and source resolution and frame rate.

Since we are ultimately interested in quality as perceived by an end-user, we propose to take some liberties and establish a mapping between PSNR and perceived quality. It is noted that perceived quality quickly saturates above a certain PSNR threshold. We postulate that 37 dB is a reasonable value for this threshold, which implies that a bit rate that yields a PSNR greater than 37 dB is equivalent perceptually to a lower bit rate (higher compression ratio) that yields a PSNR of 37 dB. Hence, there is no value, in terms of perceived quality, in using a higher bit rate than that which would achieve this threshold.

Similarly, perceived quality also reaches a minimum level of acceptance at a second PSNR threshold. We conjecture that 28 dB is a reasonable value for this threshold. This implies that for low bit rates (high compression ratios) that yield a PSNR of less than 28 dB, we assume that the received quality is so poor that it is equivalent to not even receiving the content. Hence, we must stay above this threshold by increasing bit rate or scaling back on the source resolution or frame rate. These thresholds are depicted on Figure 5 as the solid, discontinuous line.

Our final step in the modeling process is to map the thresholds and linear region to a measure of perceptual quality. The ITU has established a subjective video assessment methodology, as described in [4]. In this process, users sub-

jectively rate content based on comparison to original content using a 5-point rating systems: 5 = imperceptible, 4 = perceptible, but not annoying, 3 = slightly annoying, 4 = annoying, and 5 = very annoying. For simplicity, we use a direct mapping between our PSNR least-squares model and thresholds to this scale. Figure 6 depicts our model for mapping compression ratio to perceptual quality as defined by the ITU subjective scores.

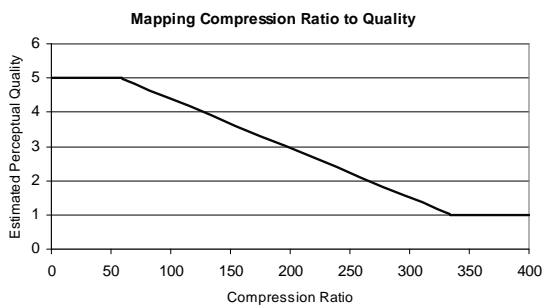


Figure 6. Mapping relationship between compression ratio and estimated perceptual quality.

This model, as we discuss in subsequent sections, can provide a practical alternative to the actual operating Distortion-Rate characteristics in real-time streaming applications. An approximate D-R function is used to establish the optimal bit rate partitioning of the scalable coding representation to achieve the highest aggregate perceived quality for all clients in our multicast scenario. Additionally, use of this model will eliminate the case where certain clients receive an excessively high bit rate stream at the expense of clients that can only receive a lower data rate (this is achieved through the upper threshold). It will also ensure the minimum level of quality received by clients that are only able to receive at low link rates.

3 IEEE 802.11 WLAN channel

The original IEEE 802.11 b standard [3] a 2.4GHz spread-spectrum wireless LAN capable of operation at bit rates of 1, 2, 5.5 and 11 Mbit/s (using a spread-spectrum BPSK modulation). IEEE 802.11 'a' and 'g' boosted the rate up to 54 MBit/s by using more advanced physical (PHY) layers and it also widened the range of available rates to provide more flexibility for channel adaptation. Available proprietary solutions advertise higher rates and the IEEE 802.11 groups are in the process of developing even faster wireless networking standards. As a result, effective throughput of WLANs can now achieve about 20-25 MBit/s and this finally enables applications like wire-

less video streaming. Figure 7 illustrates a typical effective throughput achieved by an 802.11b network under normal conditions. As a rule of thumb the effective throughput is 50 – 60% of the wireless rate due to media access and control overhead and for an 11 Mbit/s TCP connection is close to 5 – 5.5 Mbit/s.

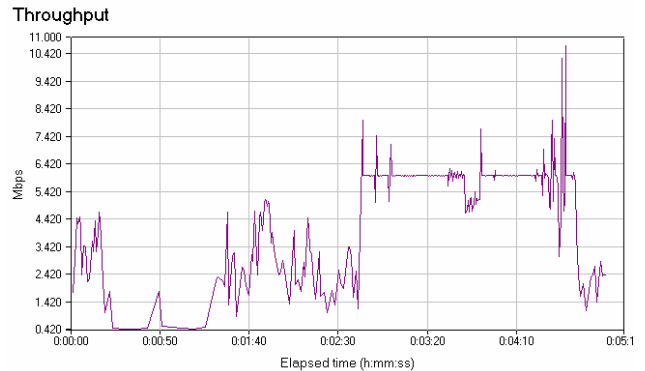


Figure 7. Example of throughput variation over time in 802.11 b network.

Effective throughput differs significantly for different PHY rates. Figure 8 illustrates the Packet Error Rate dependence on Signal to Noise Ratio in the channel for different rates/modulation schemes for the packet size of 1024 Bytes. While the actual PER values may differ depending on implementations the relative behavior remains more or less the same: in most channel SNR regimes only a single modulation scheme performs relatively well (i.e., above around 10% PER in the Figure 8). This PER dependence on the channel SNR justifies the approximation in Figure 1 where the SNR is pictured to be proportional to the distance of a receiver from the sender. Another parameter of the 802.11 PHY that determines the the effective packet throughput is the size of the PHY frames. Longer packets in a high SNR regime offer higher throughput compared to the shorter packets due less relative overhead of headers, acknowledgements and so on. Conversely, when the channel gets noisier it may make sense to switch to shorter packets that have higher chances to get through. This trade-off provides an additional degree of freedom in the overall optimization problem of media multicast.

Throughout the process of improving PHY layer of 802.11 networks, the Media Access Control (MAC) essentially remained the same. The IEEE 802.11 standard uses the same logical link layer as other 802-series networks (including the 802.3 wired Ethernet standard), and uses compatible 48-bit hardware Ethernet addresses to simplify routing between wired and wireless networks. Using radio transceivers for the network physical layer is complicated by the inability of radio transceivers to detect colli-

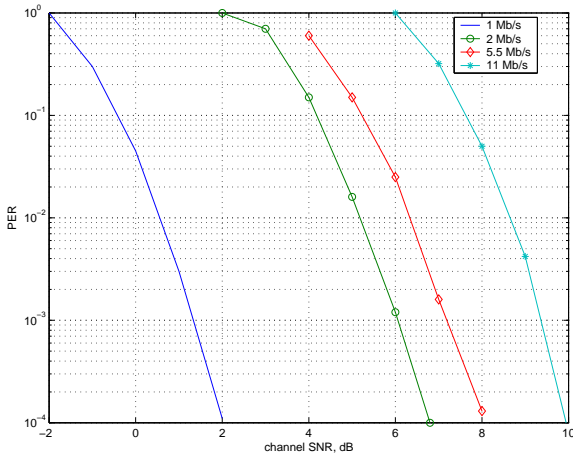


Figure 8. Typical Packet Error Rate for different wireless transmission rates as a function of Signal to Noise Ratio.

sions as they transmit, and the potential for devices outside the network to interfere with network transmissions. Communication is also hampered by the hidden node problem; two widely spaced nodes on the network may be unable to communicate with each other directly, and yet still interfere with transmissions to an intermediate point. To address these limitations, a complex Media Access Control that includes retransmissions of corrupt packets and collision avoidance is used. Link layer acknowledgements are used to detect corrupt packets and initiate fast retransmissions, perhaps with different rate. Because in this paper we are targeting a multicast transmission that does not use ACKs at the link layer, extra arrangements may be required to control the transmit rate and other parameters that are configured with the help of ACK mechanism.

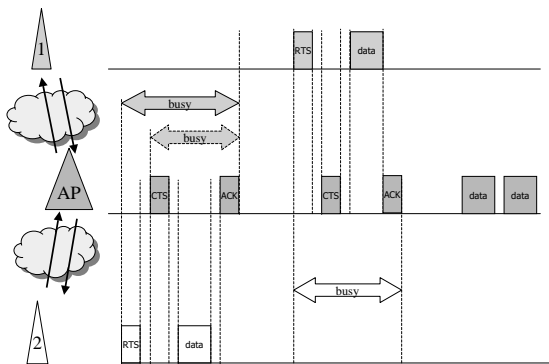


Figure 9. Example of IEEE 802.11b LAN operation.

Fig. 9 illustrates a Distributed Coordination Function (DCF) MAC using a typical example of two mobile stations communicating with an Access Point (AP). The Request

To Send (RTS) / Clear To Send (CTS) mechanism is often used to avoid a hidden node problem. Station 2 accesses the channel (after sensing the channel as free for some interval of time) by sending an RTS frame. At this time all other stations that received the RTS frame set the state of the channel as “busy” for the duration of time advertised in the RTS frame. The AP broadcasts the CTS frame so that all other stations that are out of range of station 2 (“hidden” nodes) can also set the channel state as “busy”. After this procedure Station 2 sends a data frame that is acknowledged by the AP. A similar process is then repeated by station 1 when it needs to transmit (the time between the acknowledgement frame and the RTS frame from station 1 is available for all stations to initiate transmission). Finally, the AP multicasts several unacknowledged data packets (in this example not using the RTS/CTS mechanism).

It was clear that the original 802.11 MAC protocol was inadequate for providing QoS for media streaming applications. IEEE 802.11 'e' group has been working on improving this situation and the results of its work are expected to become a standard in the near future. In the meantime, a subset of IEEE 802.11 'e' draft was adopted by several companies and is known as Wireless Multimedia Extension (WME). Among others, WME provides several important tools to improve QoS for multimedia including: maintaining 4 separate queues for wireless streams with different priorities, ability to use unacknowledged transmission mode and block acknowledgements, enhanced DCF media access. In practice, the QoS features give a basic set of tools to optimize the LAN-wide traffic - efficient way of using these tools is an ongoing research and development effort.

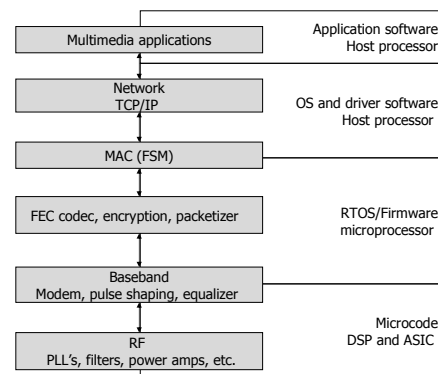


Figure 10. System view of a typical software and hardware stacks for wireless media streaming.

Modeling of an 802.11 channel can be done at multiple levels. Figure 10 illustrates a typical software and hard-

ware stack for a PC-like platform running a wireless media streaming application. The IEEE 802.11 MAC/Logical Link Control (LLC) and physical (PHY) layers are represented by the lower for blocks in the diagram and in practice are implemented as a combination of DSP/ASIC responsible for the RF and latency-critical processing, and the microprocessor implementing the MAC functionality. In contrast to this, TCP/IP stack along with all OS software and the actual media streaming applications are executed on a host processor. It is usually a common approach to design such multi-layer systems by allowing only adjacent layers to communicate with each other and thus separating the functionality and design efforts for those layers into independent problems. Under this practical constraint, user applications see the wireless channel as an IP packet channel with erasures—much like the wired Ethernet. Therefore in this paper we model the wireless network channel as a packet erasure channel at the network layer level. On the other hand, the specifics of media streaming in general and especially media multicast over WLAN require some cross-layer interaction to efficiently adapt source and channel coding. This avenue has been explored in academic research of Joint Source/Channel coding that was shown to offer potential gains over a traditional source-channel coding separation. This approach is yet to find its way into real-life systems.

To illustrate our channel model in Figure 11 we show a sample trace³ which we obtained using an 802.11 AP and mobile station. The AP was sending UDP multicast packets at a constant rate and the receiving station was recording the sequence numbers of the correctly received packets. Each point on the trace corresponds to an average of about one hundred packets received. Note that the channel behavior in this experiment is the result of the interaction between the actual RF channel, the hardware responsible for PHY and the software/firmware responsible for MAC/LLC and IP implementation.

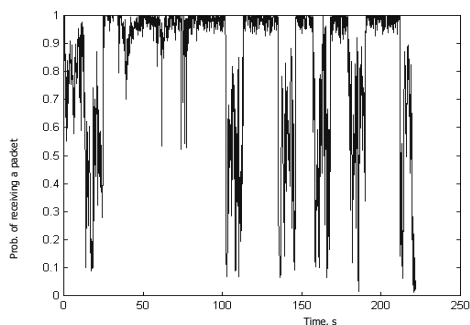
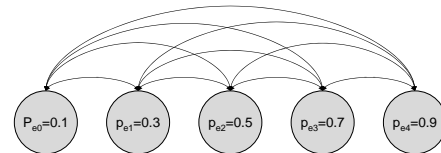


Figure 11. Packet erasure probability as a function of time in IEEE 802.11b trace.

³The trace shows a channel with a rather poor quality because of significant distance used in this measurement.

In the simplest case the packet channel illustrated in Figure 11 can be approximated as a memoryless packet erasure channel when packets are dropped independently of each other with probability p_e . However, by examining the trace in Figure 11 (and the one in Figure 7) we can conclude that there is significant amount of correlation between adjacent samples which suggests a better model with memory for approximating the channel. One possibility is to use a multi-state Markov channel model. Figure 12 illustrates a 5-state Markov model approximation and a sample transition probability matrix obtained for a trace shown in Figure 11 uniformly quantized to probabilities of packet erasure of 0.1, 0.3, 0.5, 0.7 and 0.8. The estimated transition probability matrix can be used to predict the next state of the channel with high fidelity. For the sample trace shown in Figure 11 prediction error is within 10% and the maximum error of one state (i.e., 0.2 in our case). This in practice means that the channel state can be very efficiently estimated using only sparse measurements of the packet error rates at the clients, which, in its turn, allows adapting coding scheme for c



0.7368	0.2174	0.0055	0	0
0.2211	0.4609	0.1803	0.0278	0.0014
0.0421	0.2957	0.5191	0.2083	0.0036
0	0.0261	0.2678	0.5278	0.0356
0	0	0.0273	0.2361	0.9595

Figure 12. Markov channel model with five states and a sample transition probability matrix.

3.2 Coding for Packet Erasure Channels

Two general classes of methods can be used to combat packet losses in a packet erasure channel. The Forward Error Control (FEC) is a method of transforming the data block into a set of packets of bigger size to ensure that if a number of a lost packets is below some designed threshold the original data can be extracted intact. This method (FEC) therefore provides error resilience by increasing the amount of data to be sent. The FEC does not require the return channel (aside from statistics and control data required for adaptation). There is no guarantee that the data will arrive to the receiver without errors. For multimedia streaming, however, the delay requirements often dominate the error-free transmission requirements and hence, error-free transmission is not required. Automatic Repeat Request (ARQ) is an alternative approach to robust data communications. It operates by requesting the transmitter to resend the erased

packets using the return channel. ARQ is implemented both in the WLAN MAC (for unicast traffic) and also at the TCP/IP level if TCP is used. Thus ARQ requires two-way communication channel to be present, which in the case of WLAN is the same physical medium as the forward channel so effectively ARQ also expands the data size because of acknowledgements and retransmissions. The difference from the FEC, however, that ARQ is inherently channel adaptive since only lost packets are resend (while FEC adds overhead to all packets). On the other hand, ARQ may introduce significant delays due to roundtrip propagation time and processing time. The last condition significantly limits the application of ARQ to multimedia communications.

Previous work on media streaming also suggested combining FEC and ARQ methods together (sometimes called Hybrid ARQ) which might be also applicable to the multicast scenario when the number of acknowledgements is not significant. In our approach we chose to avoid Acknowledgements and retransmissions at all and adopt an Unequal Error Protection (UEP) mechanism for protecting scalable media streams in the multicast applications.

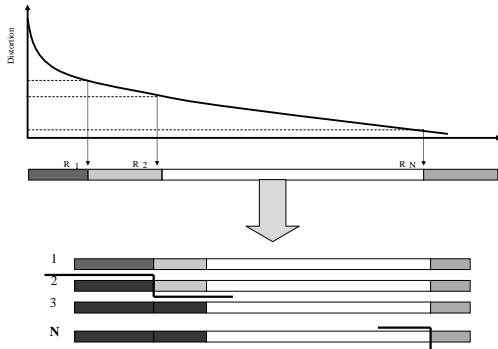


Figure 13. Scalable bit stream partitioned into m layers and encoded using packet erasure correcting scheme.

A part of source data (for example, a few macroblocks, a video frame or several frames) are first encoded using a scalable/progressive coding scheme as outlined in Section 2. Next, the encoded stream is divided into N segments that are mapped to N data packets as shown in Figure 13. The i -th segment (quality layer) is split into i equal parts and a systematic $(N, i, N - i + 1)$ Reed-Solomon code is applied to it to obtain the contribution of the i -th layer to each of the N packets. The Reed-Solomon code ensures that the i -th resolution layer can be decoded if the number of erasures does not exceed $N - i$. In other words, if in Figure 13 only (any) two packets are received out of total N the decoding can be performed up to the rate of at least R_2 . This encoding scheme also guarantees that the first part of the bitstream is decoded successfully before any following part can be decoded.

For a unicast (single user) scenario the problem of designing the optimal channel coding scheme (as outlined above) for a given progressive source is equivalent to finding the locations of $N - 1$ partitions (some partitions may occur in the same places meaning that certain channel rates are not used) subject to the total rate constraint. A fast, near-optimal solution, of complexity $O(N)$, based on Lagrangian optimization (MDFEC algorithm), to optimally partition the bit stream is described in [14]. As the input the MDFEC algorithm expects the Distortion-Rate function, channel statistics and the total rate constraint. The expected distortion is given by the equation ([14]):

$$E[d] = q_0 E + \sum_{j=1}^N q_j d(R_j)$$

where $q_i(N)$ is the probability of receiving i out of N packets, $d(r)$ is the distortion at rate r , E is the source variance, and $\mathbf{R} = (R_1, R_2, \dots, R_n)$ is the rate partition (see Figure 13). The MDFEC algorithm attempts to find the minimum of $E[d]$ with respect to partitions R_j .

4 Problem formulation for media multicast

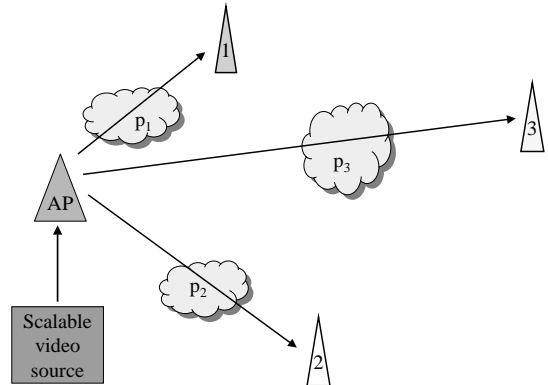


Figure 14. Wireless Multicast scenario with three users having different probabilities of packet erasure (p_1 , p_2 and p_3).

In this section we formulate the problem of media multicast over a wireless LAN. We consider a specific case of a scalable data stream communication using UEP codes over several memoryless packet erasure channels with known probabilities of erasure as illustrated in Figure 14. We start with an observation that when there is only a single client, it is clear that an optimal coding scheme should seek to maximize the received user quality (minimize distortion) given a set of constraints on rate, bandwidth, power, etc. However,

we cannot directly apply this idea to the multicast scenario, since the scheme that maximizes the received quality for one client may not be the optimal one for other clients, due to the heterogeneity among the receivers. Hence, for the multicast scenario, we should attempt to maximize some overall quality criterion subject to a set of given constraints.

To state the problem formally, let $D(\mathbf{x}, \hat{\mathbf{x}})$ be a distortion measure, where \mathbf{x} and $\hat{\mathbf{x}}$ are original and decoded data vectors (for example video frames). For real-valued sources we use Mean Squared Error (MSE) defined as $D(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n} \|\mathbf{x} - \hat{\mathbf{x}}\|^2$. Let $K = (K_e, K_d)$ be a set of block encoders/decoders. That is if $\mathbf{x}_0^{n-1} = x_0, \dots, x_{n-1}$ are samples from the source $s \in S$ then the encoder is a mapping from the source alphabet A_S^n into the channel alphabet A_C^m , i.e., $\mathbf{y}_0^{m-1} = K_e(\mathbf{x}_0^{n-1})$. The signal is transmitted through a channel $c \in C$ to produce the output \mathbf{z}_0^{m-1} . The decoded signal is then given by $\hat{\mathbf{x}}_0^{n-1} = K_d(\mathbf{z}_0^{m-1})$. For each coding scheme $k \in K$ we define its measure of robustness over the multiple users as:

$$\rho_k = L(D_k(i) - D_{opt}(i))$$

where L denotes some function (i.e., L_2 norm, weighted L_2 norm, maximum, etc) $D_k(i)$ is the distortion of the coding scheme k for the i -th user and $D_{opt}(i)$ is the optimal performance achievable by the coding scheme designed solely for user i . Additional constraints (blocklength, etc.) are taken care of by specifying the classes of encoders and decoders. Weighting is also easily incorporated in to this problem formulation. The performance of a coding scheme can then be optimized for a desired quality metric ρ_k . In a specific case of scalable source coding and UEP channel coding for packet erasure channel, K_e maps source data into a set of packets that are transmitted over a packet erasure channel to the users who finally map the received packets back to the source data using decoder K_d , perhaps with some quality degradation.

Assuming that rate-distortion curve is convex (not always true in practice for operational rate-distortion curves), the expected distortion must also be convex in the rate partition, as the weighted sum of convex functions with positive weights is another convex function. Hence, $D_i = (E[d_i] - E[d_i]_{min})$ is also convex in the rate partition.

Interesting special case can be further investigated if we assume a minimax optimization of D_i 's. That is assuming that $L(\cdot) = \max(\cdot)$ implies D has to be convex in the rate partition, as the maximum/supremum of a set of convex functions is convex [9]. Let $\mathbf{Q}_i = (q_0 q_1 \dots q_n)^T$ represent the transmission profile of the i^{th} client and let $\mathbf{P} = (d(R_0) d(R_1) \dots d(R_n))^T$, where $d(R_0)$ is the source variance. Then we can write the expected distortion of the i^{th} client as

$$E[d_i] = \mathbf{Q}_i^T \mathbf{P}$$

It was demonstrated that for the case of 2 users, the optimal minimax solution is $D_1 = D_2$. Then at $D_1 = D_2$,

$$(\mathbf{Q}_1^T - \mathbf{Q}_2^T) \mathbf{P} = E[d_1]_{min} - E[d_2]_{min}.$$

$E[d_1]_{min}$, $E[d_2]_{min}$, \mathbf{Q}_1 , and \mathbf{Q}_2 are known quantities and hence solving this equation for \mathbf{P} , we obtain the optimal rate partition. For more than 2 users we can simply repeat this procedure for all possible pairs.

Up to now in our derivations of optimal partitioning we implicitly assumed that the transmission happens at the same wireless rate. This is not true in practice where, for example in 802.11 'b' there are 4 possible rates that allow to improve the effective throughput depending on the actual channel conditions⁴. We can account for possibility to have multiple rates by introducing another index m to correspond to a particular modulation scheme. In this case probability of packet erasure for client i becomes a vector with elements $p_{e_i}^m$ and the rate profile becomes \mathbf{R}^m . The problem for a single user can be formulated as the one of finding R_i and the corresponding modulation rates m_{R_i} that minimizes the expected distortion subject to the total rate constraint⁵. In practice to solve this optimization problem we can use MD-FEC algorithm followed by an optimization over modulation index in an iterative manner. We are currently in the process of investigating this approach and extending it for a multicast problem formulation.

5 System approach to media multicast

In the previous section our goal was to formulate and solve the media multicast problem from a more theoretical point of view. The goal of this section is to outline a system view of a typical media multicast scenario and outline the practical aspects of our approach.

Regardless of the actual criteria used to optimize source and channel coding the algorithm for Unequal Error Protection of a scalable video source is expecting as an input the set of source and channel parameters. As we demonstrate in Figure 15 the optimization algorithm (in practice implemented as a part of user application) receives the source Distortion Rate function, that can be approximated by a few parameters as described in Section 2.

In order to find the optimal coding strategy the algorithm requires the information about the existing users and their corresponding channel conditions. Initial discovery can be performed using a standard mechanisms like Universal Plug and Play (UPnP) where devices manifest their presence and capabilities using a standard IP mechanism.

⁴Another tradeoff not discussed in here is based on ability to adjust the packet length for particular channel conditions.

⁵Note that this total rate also takes into account modulation rate which was not the case in the original formulation.

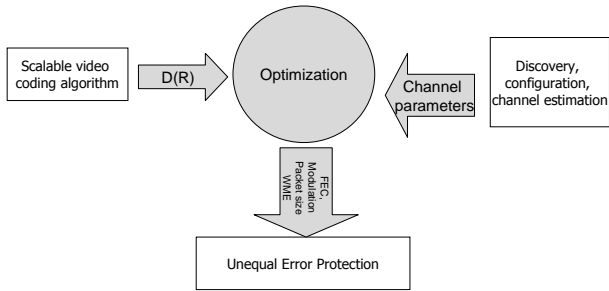


Figure 15. Implementation of video multicast over 802.11.

The devices themselves are also responsible for collecting the channel statistics (i.e., current probabilities of packet erasures) and communicating these information back to the multicast source performing optimization. As we demonstrated in Section 3 the channel state can be efficiently estimated using a simple Markov model with only several data points (one per 0.1 – 0.3 seconds) needed to be collected and communicated back to the data source. The optimization algorithm then decides on the set of source and channel coding parameters (rate partitioning and code protection) as well as the modulation, packet size and QoS parameters passed directly to the 802.11 MAC level. In practice, the video multicast traffic will be assigned the top WME priority to ensure that the background best effort connections do not interfere with video (but at the same time have opportunity to transmit when channel conditions permit).

6 Concluding remarks

In this work we address the problem of reliable video multicast over 802.11 networks. The combination of scalable source coding and Unequal Error Protection channel coding is an attractive theoretical solution for this problem, which can also be efficiently implemented in real-life systems. This is a new area of research and development and the answers on a few problems outlined in this work are yet to be found.

References

- [1] ATSC, "ATSC digital television standard", ATSC doc. A/53, September 1995.
- [2] ISO/IEC 13818-2:1996(e), "Information technology - Generic coding of moving pictures and associated audio information: Video", 1996.
- [3] Information technology - telecommunications and information exchange between systems - local and metropolitan area networks-specific requirements - part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications. 1997.
- [4] ITU-R BT 500-10, "Methodology for the subjective assessment of the quality of television pictures", March 2000.
- [5] ITU-T and ISO/IEC JTC1, "Advanced video coding for generic audiovisual services," ITU-T Recommendation H.264, ISO/IEC 14496-10 AVC, September 2003.
- [6] ISO/IEC JTC1/SC29/WG11, "Technologies under consideration for Working Draft 1.0 of ISO/IEC 21000-13 Scalable Video Coding," Doc. N6519, July 2004.
- [7] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan. Priority encoding transmission. *IEEE Trans. Inf. Theory*, (42):1737–1744, November 1996.
- [8] A. Balachandran, A. Campbell, and M. Kounavis. Active filters: delivering scalable media to mobile devices. In *Proc. of NOSSDAV 97*, 1997.
- [9] D. Bertsekas. *Nonlinear programming*. Athena Scientific, Belmont, 1995.
- [10] L. Cheng, W. Zhang, and L. Chen. Rate-distortion optimized unequal loss protection for FGS compressed video. *IEEE Trans. Broadcasting*, 50(2):126–131, June 2004.
- [11] G. Conklin, G. Greenbaum, K. Lillevoldand, Y. Lippman, and A. Reznik. Video coding for streaming media delivery on the internet. *IEEE Trans. on Circuits and Systems for Video Technology*, 11:269–281, March 2001.
- [12] G. Davis and J. Danskin. Joint source and channel coding for image transmission over lossy packet networks. In *Conf. Wavelet Applications to Digital Image Proceedings, SPIE*, Denver, CO, Aug. 1996.
- [13] V. Iversen, J. McVeigh, and B. Reese. "Real-time H.264/AVC codec on Intel architectures", to appear. In *Proc. IEEE Int. Conf. Image Proc.*, October 2004.
- [14] K.-W. Lee, R. Puri, T.-E. Kim, K. Ramchandran, and V. Bhargavan. An integrated source coding and congestion control framework for video streaming in the internet. In *Proc. of INFOCOM 2000*, March 2000.
- [15] A. Majumdar, D. Sachs, I. Kozintsev, K. Ramchandran, and M. Yeung. Multicast and unicast real-time video streaming over wireless LANs. 12(6):524–534, June 2002.
- [16] A. E. Mohr, E. A. Riskin, and R. E. Ladner. Graceful degradation over packet erasure channels through forward error correction. In *Proc. Data Compression Conference*, Snowbird, UT, Mar. 1999.
- [17] S. McCanne, V. Jacobson, and M. Vetterli. Receiver-driven layered multicast. In *ACM SIGCOMM 96*, 1996.
- [18] W. Sweldens. A custom-design construction of biorthogonal wavelets.
- [19] D. Wu, T. Hou, and Y.-Q. Zhang. Scalable video transport over wireless IP networks. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communication*, pages 1185–1191, Sept. 2000.