Video Communication on Mobile Wireless Ad-hoc Networks

Gagan Gupta
gagan.gupta@arc.com

David Lissmyr
lissmyr@stanford.edu

EE384B: Multimedia Communication over the Internet
Project Report

March 2007

Abstract

Wireless mobile ad-hoc networks are made up of cooperating mobile nodes that do not need any supporting infrastructure or centralized access point. In addition to challenges posed by wireless networks to multimedia communication, the inherent characteristics of mobile ad-hoc networks make multimedia communication more difficult as paths are frequently broken, link-capacities fluctuate and the nodes usually have tight power constraints. Much work has been done to enhance individual layers in the network protocol stack to address these challenges. Recently various research groups have proposed improvements based on cross-layer designs, where optimizations are not made locally but jointly over multiple layers. Our work explores the reasons for using cross-layer designs and analyzes both the advantages and trade-offs of such designs. We take a look at how different parts of the network layer stack may be affected due to such techniques. We present a survey of different cross-layer techniques that have been proposed. Many challenging problems in cross-layer designs are yet to be resolved. In the survey we propose a general framework that facilitates interaction and optimization across the entire networking protocol stack to explore cross-layer designs. We also identify a few other areas for future work.

Acknowledgments

We would like to thank our project supervisor, Prof. *Fouad Tobagi,* and teaching assistant, *Sangwook Ha,* for their help, comments and insights. This work would not have been possible without their guidance.

1 Introduction

Multimedia communication and mobile communication have seen rapid growth in the last decade and will continue to see an even more rapid growth in the coming years. It was only inevitable for multimedia content to be delivered over mobile networks. Given the commercial success it has seen, real-time audio-visual communication will become an integral part of next generation wireless communication.

Mobile ad-hoc networks, often referred to as MANETs, consist of autonomous nodes that are free to move and form arbitrary networks. Nodes may drop out or get added as they move around. They communicate using the wireless medium, so that links can be created between nodes within a certain distance of each other. Nodes situated farther away can be reached using the multi-hop routes in the network. Moreover, links are constantly created and broken as nodes move around or drop in and out. Such networks are useful where permanent infrastructure is unavailable or not cost-effective.

Ad-hoc networks will find applications where permanent communication infrastructure is absent or too expensive such as in remote areas, non-permanent business settings, personal area networks, low cost community internetworking, etc. They are highly flexible and are easy to set up. They therefore find a wide range of applications ranging from personal area networks to 4G mobile networks. We also sense that ad-hoc networks will fuel commercial success of peer-to-peer multimedia information exchange in mobile personal networks.

Mobile wireless ad-hoc networks have been an important topic of research during the last few years. Since these networks are characterized by limited bandwidth, variable link capacities, changing topology and power-constraints, they pose a big challenge when it comes to media communication, with problems ranging from QoS to transmission error management. As nodes appear and disappear and as they move around, paths are frequently broken. Robust and efficient protocols are therefore needed to establish paths to accommodate media communication.

Of all types of content, motion video, even with state-of-the-art compression, is most the most demanding in terms of bit-rate and user experience. The challenges are further compounded due to high error rates, fading signals and intermittent loss of connection in mobile networks.

Much work has been done in networking protocols to improve end-user experience of viewing video content on such networks. While individual layers in the network protocol stack (OSI model) have been enhanced to handle such traffic, recently work has been done to improve optimization using cross-layer techniques in which multiple layers cooperate with each other. Such techniques are showing promising results since they adapt based on a more global view of the system.

In this survey paper we enumerate the challenges posed by ad-hoc networks for multimedia communication and describe the different cross-layer techniques that have been proposed to address them.

Rest of the document is organized as follows. In section 2 we provide an overview of challenges posed by ad-hoc wireless networks for video communication. In section 3 we provide an overview of cross-layer designs and why they improve video streaming. In section 4 we provide details on an example of video coding standard, H.264/AVC, which forms a vital cog in cross-layer designs. In sections 5, 6 and 7 we provide details on cross-layer designs proposed in the academia. To the best of our knowledge we are unaware of actual commercial implementation of such techniques.

The area of cross-layer design is relatively new and much work remains to be done. In section 8 we also present our thoughts on how future research could be made easy and some areas of future work in the field.

2 Video Transmission over Ad-hoc Networks

It is difficult to provide guaranteed quality of service (QoS) in mobile networks due to the following reasons[10]:

1. Wireless links are highly error prone. Electromagnetic waves used in mobile radio communication travel between the sender and receiver. During propagation the waves get affected by absorption, reflection, diffraction and scattering. When the sender and/or the receiver are mobile, these physical phenomena create time-varying channel conditions and hence affect channel performance.

   Objects such as hills, trees, walls, etc. between sender and receiver cause shadowing thereby attenuating the radio frequency (RF) signal (path loss). Since the waves travel in all directions once transmitted they can arrive at the receiver from multiple directions with different delays causing phase and frequency shifts. Superposition of the waves causes interference resulting in amplitude attenuation. Both phenomena (called fading) can cause lengthy burst of errors, loss of synchronization and intermittent loss of connection.

   Transmission errors can range from single bit errors to burst errors.
2. Ad-hoc networks suffer from loss of connectivity as nodes drop off.
3. Since nodes in ad-hoc networks are typically user devices they use low power transmission signals that readily suffer fading.
4. Due to the dynamic nature of the network it can suffer from all of the above issues simultaneously.

Low bit-rate video coding relies heavily on interframe coding, i.e. previously encoded frames are used to predict next frame. Errors in one frame can hence have a cascading impact on subsequent frames. And one can always expect some amount of residual error to occur in such error-prone channels. Since video transmission is delay sensitive, transmission and queuing delays also become important factors in quality of service (QoS). A delayed packet is as good as lost since the receiver cannot make us of it in a timely fashion. Video streams also generate a large amount of data. It is therefore important to avoid congestion and ensure that the network operates within its capacity.

Due to above reasons ad-hoc networks pose the most serious challenge to video transmission.

Transmission errors can be handled by using forward error correction (FEC) to some extent. The error conditions vary widely and worst case FEC protection makes the bit-rate impractical. Closed-loop automatic repeat request (ARQ) retransmission scheme fares better than FEC. However user experience of video content requires real-time, fixed delay playback synchronized with audio making only retransmission based error handling impractical. A combination of the FEC and ARQ schemes can be used. A balance needs to be maintained between channel coding and source coding. Given a fixed bandwidth available between sender and receiver a strong channel coding scheme can result in weaker source coding scheme thereby delivering a poor quality video signal at the receiver.

Transmission errors can also be reduced by judiciously controlling the signal power to reduce interference[3]. Strong signal can improve reception quality at the receiver but also increases the chances of interfering with other users and degrading their performance. Hence an appropriate balance needs to be maintained for signal power also.

Given the range of issues ad-hoc networks can suffer from, individual layers of the network stack cannot address them adequately. Meeting end-to-end performance requirements of demanding applications requires interaction between layers. Error handling needs to be tackled by multiple layers simultaneously. For e.g. by controlling signal power, channel coding, transport protocol as well as using various source coding and destination decoding algorithms. Furthermore these algorithms need to be adaptive because network conditions can change dynamically during transmission of a video stream. This requires feedback of channel conditions to the upper layer so that adaptive decisions can be made.

3 Cross-layer design

Networks are traditionally organized as a stack of layers, each one interacting with the layer directly below and above itself. This structure makes the design of a networking system easier since once the outputs and inputs of each layer have been defined, they can be implemented independently from one another. A change of technology in one layer can be performed without having to alter the whole stack.

In the case of low-latency data communication over wireless ad-hoc networks, cross-layer implementations have been proposed. In this case, the layers do not operate independently but share information and perform joint optimizations. Recently, many cross-layer designs have been proposed for different levels of the network protocol stack and applications to maintain a desired QoS.

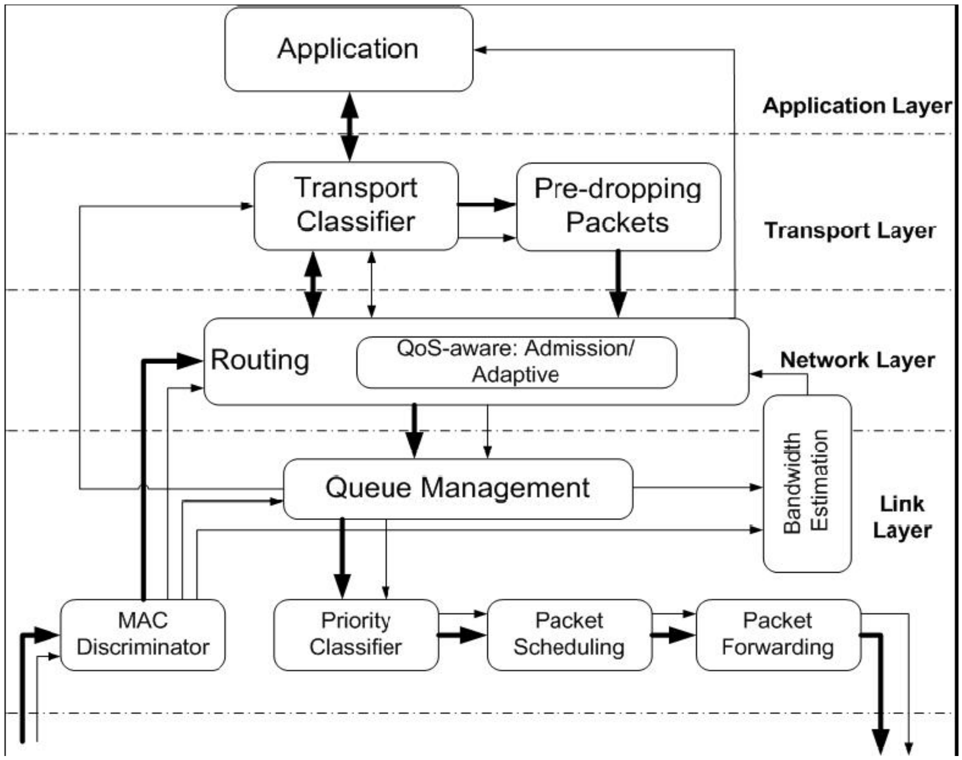


Figure 1 A QoS architecture proposed in [20]

Figure 1 shows one such QoS model based on cross-layer design technique proposed in [20]. It shows possible interactions between various layers of the protocol stack. Bold lines are flow of data packets and narrow lines indicate flow of control information exchanged between different layers. Information is exchanged between layers which can then take adaptive measures to maintain QoS. For e.g. routing protocols can avoid weak links or application layer can adapt its transmission rate based on network throughput and latency. In general, QoS can be defined to include overall network utilization, throughput, latency, distortion at receiver-end, overall power consumption and other metrics. Cross-layer designs strive to simultaneously optimize for one more of these parameters.

Cross-layer design is a very broad term that can be applied to many various kinds of designs. Some tend to use this term for designs where there is just a piece of information going from a layer to an other one. But the traditional layering already includes information transmissions between different layers. Adding new pieces of information in the interface between two layers can therefore not really be called cross-layer design since strict layering is still preserved. Cross-layer design would rather imply a joint optimization between different layers.

This joint optimization might either be a single algorithm that would make (multi-dimensional) optimization on information inherent to each of the layers and output decisions for each of them. It might also preserve a stronger form of layering by consisting of “dynamic parallel optimizations”. One layer would compute an optimum and give the result to the other layer that would compute its own optimum. Once this is done, this last optimum would be sent to the first layer that would update its optimum and pass the new information to the second layer and so on until the process converges.

Need for cross-layer design

The layered construction of the protocol stack was designed to be adapted to traditional networks. Such networks have wired links and fixed nodes. They offer the possibility to establish a relatively stable path and keep a low packet loss that can be treated using error control and error concealments methods. Wireless mobile ad-hoc networks do not come up with routes the same way traditional networks do. The topology constantly changes and the link qualities fluctuate, leading to packet loss rates that go far beyond what can be treated using classical error concealment strategies. Recovering from lost packets is crucial to the reconstruction of the video. Upper layers therefore need to have some knowledge of the channel quality variations in order to reach an optimal operating point. Added to this, media communication over such networks has strict QoS requirements that make joint-optimizations even more crucial.

Advantages

Cross-layer designs make use of the interactions between layers and develop these to reach better results. They also enhance adaptability at all layers using this information exchange.

Drawbacks

As previously mentioned, cross-layer design can significantly increase the design complexity. Since layers can no longer be implemented independently, any change in technology or within one layer might directly affect the other ones too. Layers allow designers to perform optimizations within each layer without having to take into account the constraints inherent to the other layers. Cross-layer designs specifically use joint-layer optimizations and therefore lose this convenience of implementation.

We therefore lose some advantages of the layered design such as the ability to standardize and to deploy new protocols. Moreover, joint optimizations between various layers may involve complex equations with multiple variables and therefore complicate the algorithms.

Employing cross-layer does not require one to abandon the OSI reference model. Abandoning layering would clearly be impractical and would lead to major problems in terms of implementation, standardization, debugging, upgrading and more. Cross-layer designs will therefore keep some level of separation while authorizing information exchange and joint optimizations. Such optimizations can involve any of the different layers. Some examples of optimization of existing protocols for multimedia communication involve (i) the Physical and MAC layers: adaptive modulation and MAC optimizations and (ii) the Physical and Network layers: adaptive power control and routing optimizations.

There are many others that include the transport layer and some form of encoder control (application layer) such as joint optimization of the bit rate of the video encoder and the hop counts of the paths established by the network layer.

In section 4 we provide details on the latest video coding standard H.264/AVC which includes features that help in cross-layer design. Many cross-layer designs eventually rely on the ability of the encoder to control the bit rate and data it produces and on the decoder to perform error concealment when errors occur. Cross-layer designs control the encoder based on prevailing network conditions and knowledge of error concealments the decoder can perform to maintain high quality of video reception at the receiver. Hence the source coding system is an important component in cross-layer design. We have used H.264/AVC only as an example in our survey. Similar features also exist in other coding standards.

In section 5 we provide details on cross-layer techniques that work on the upper layers of the protocol stack, from the application layer to the network layer. In section 6 we present a survey of techniques that predominantly involve the lower layers of the protocol stack, from the network layer to the physical layer. In section 7 we present a few techniques that take a more comprehensive view of the protocol stack when applying optimizations.

All techniques presented use peak signal-to-noise ratio (PSNR) to measure quality of video received at the receiver. They measure effectiveness of their schemes based PSNR value of test streams with and without application of their schemes.

4 H.264/AVC

ITU-T's Video coding Experts Group (VCEG) and Moving Picture Expert Group (MPEG) collaborated to define a new video coding standard, H.264/AVC [31] in 2003. It also forms a new part, called AVC, of the MPEG-4 standard. H.264/AVC offers improved coding efficiency and network adaptation as compared to all previous video coding standards such as MPEG-4 [17], H.263 [19] and MPEG-2 [21] and enables significant bit-rate reduction at the same quality level. Many new tools have been added in the standard to improve coding efficiency, error resilience and adaptability to differing conditions and usage.

Wireless networks, including cellular networks, contain different types of transmission modes to allow for transport of different types of traffic. Video transmission services could include transmission of multimedia messages, conversational video and pre-coded stored video. In cellular networks conversational video may be transmitted using packet-switched or circuit switched services [12]. These different types of transmissions impose different requirements and it is desirable that the video coding standard be capable of addressing them.

Due to emerging business models in which end-user's costs are proportional to transmitted data volume and limited resource of bandwidth and transmission power, compression efficiency is vital for multimedia wireless applications. Furthermore, it is desirable that the coding standard syntax not affect the bit-rate and be flexible to allow transmission in different environments and networks using different current and future protocols [26]. Wireless transmissions are error prone and hence the third important feature for a coding standard is to be error resilient and adaptable to changing transmission conditions. During the process of standardizing H.264/AVC error resilience techniques were included keeping in mind IP-based and wireless transmission. For these reasons H.264/AVC is becoming a popular video coding standard and has been recently adopted as the standard for mobile TV. Therefore for our survey we have used H.264/AVC as an example to illustrate error resilience and other features that are included in coding standards that cross-layer techniques can exploit to deliver optimum quality video to the end user.

The H.264/AVC standard defines two conceptual "layers", the video coding layer (VCL) and the network adaptation layer (NAL) as shown in Figure 2 [27]. VCL specifies video signal coding. NAL specifies interface between the video codec and the external world. The NAL layer allows the assembly of the data to be transmitted to adapt to the underlying medium (wired or wireless). The syntax used to assemble data takes into account characteristics of the transmission medium and enables cross-layer techniques to take advantage of the syntax to adapt to network conditions dynamically.

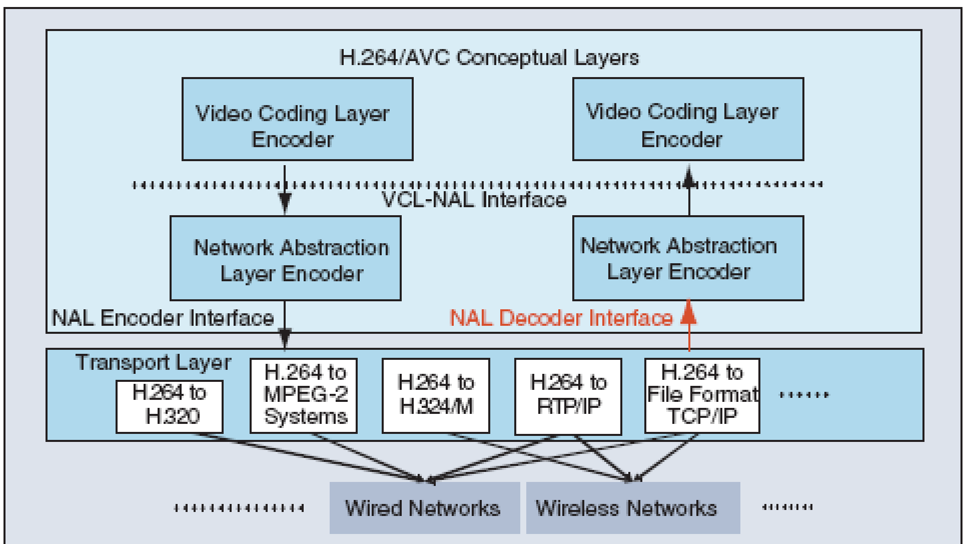


Figure 2. H.264/AVC in a transport environment: The network abstraction layer interface enables a seamless integration with stream and packet-oriented transport layers (from [7]).

Figure 2 H.264/AVC conceptual layers [27]

In general the H.264/AVC standard defines many tools and techniques in which both the encoder and the decoder can operate to minimize degradation of picture quality due to transmission errors.

The standard defines ways for a decoder to conceal errors when data is lost or arrives corrupted. These are helpful when it is impractical to ask for retransmissions. For intra coded macroblocks the decoder can use weighted interpolation to estimate values of missing pixels using pixels from neighboring macroblocks. For predicted frames motion vectors of neighboring blocks may be used to predict motion vectors of the missing macroblock or collocated macroblock from reference frame may be copied in its place.

The standard also defines multiple tools available to the encoder to improve error resiliency of the coded video stream [27]. They are listed below.

1. Slice structured coding: Allows grouping of arbitrary number of similar macroblocks in one slice. No intra-frame prediction happens across slice boundaries. This reduces image degradation due to lost packets.
2. Flexible Macroblock Ordering (FMO) allows arrangement of the coded macroblocks in a non-raster scan order which randomizes the data. So loss of a segment of data results in randomizes errors in the video frame making it more likely that neighboring data is available to conceal errors in lost macroblocks.
3. Arbitrary slide ordering (ASO) allows slices of a picture to appear in any order. This avoids increase in delay even when data arrives out of order in packet switched networks.
4. Data partitioning allows up to three partitions for transmission of coded information. First partition contains header and motion information. Intra and inter coded transform coefficients are included in the second and third partitions respectively. Higher priority can be assigned to partition containing more important data, in general intra information. Thus degradation of picture quality due to packet loss can be reduced if the network provides unequal error protection.
5. H.264/AVC allows intra coding of individual macroblocks which can be selected randomly or in a predetermined fashion or adaptively based on channel distortion characteristics.
6. Redundant coded slices - same slice is predicted using different reference frames and transmitted.
7. Multiple reference frames for coding. This increases the chances that a reference frame is available at the receiver for decoding.

Entire frame may not be intra-coded. Macroblocks (MBs) can be intra-coded occasionally or when inter-prediction yields poor result. Different modes of coding MBs have been defined in the standard to enable low entropy compression. Intra-coded MBs can be selected randomly or in a pattern and transmitted periodically. This limits quality degradation when losses occur.

Of course the strongest lever one can apply on an encoder is the quantization parameter, QP that directly controls the bit rate it generates.

Different applications can make effective use of a subset of these techniques. Used individually or in combination these techniques have shown to retain a high quality of decoded video stream measured using either PSNR or SNR to measure quality as shown in [22, 26,13, 29 and 30] for error rates up to 5-7%. In general they find that using small packet sizes of about 100 bytes, FMO and introducing intra coded MBs either periodically or randomly give good overall results. In this survey we focus on only those techniques that can be used for cross-layer techniques in which the encoder or the transmitter dynamically adapts to the changing channel conditions. For e.g. various error concealment techniques that can be applied by the decoder when errors and packet loss have already occurred are not further described here.

H.264 allows transmission of video to adapt to changing channel conditions. For stored content that will be streamed, several versions of the pre-encoded content based on expected channel conditions can be maintained. The actual version streamed can be switched based on the changing channel conditions. For small fluctuations non-reference frames can be dropped. For larger fluctuations, versions can be switched at I-frames that are also indicated as instantaneous decoder refresh (IDR) (typically used for random access and fast forward/rewind). Yet another way to switch H.264/AVC encoded streams is at synchronization-predictive (SP) pictures. SP frames allow drift-free switching between versions. The standard also defines SI frames which have properties of I frames but are transmitted only if transmission errors occur.

Forward error correction (FEC) in combination with interleaving can also be applied (channel coding). However it increases the overhead in the payload. Increased coding efficiency allows use of additional bit rate for forward error correction.

Rate-Distortion Encoder Control

The parameters provided to the encoder can be selected based on the desired quality of the video (distortion) at a desired bit rate (rate control). In general, lower the bit rate, higher the distortion and higher the bit rate lower the distortion.

The QP value can be adapted on a frame by frame basis to ensure constant bit rate. If the highest QP value cannot achieve the required bit rate, frames may be dropped.

A rate-distortion (R-D) optimization algorithm can be used to select encoder parameters to produce a coded output that has minimum distortion at a given bit rate. The desired bit rate can be derived on the basis of existing network conditions.

Lagrangian optimization technique is used to achieve this optimization in [26]. A set, O, of coding options are defined for each MB. Assume that a certain mode o introduces distortion D(o) and the corresponding bit rate R(o), the rate constrained mode decision selects the coding option $o*$ such that the Lagrangian cost function is minimized as

$$o^* = \arg\min(D(o) + LR(o)) \quad o \in O \tag{1}$$

L is the Lagrange parameter for appropriate weighting of rate and distortion. The Lagrange mode selection can be used for QP value, MB mode, reference frame selection and motion vector search. Distortion can be measured using squared sample differences (SSD) or sum of absolute differences (SAD). Rate is measured using number of bits used for encoding. Lagrange parameter depends on the QP value selected.

The selection of coding mode in (1) can be modified based on influence of the random lossy channel. In case of error-prone transmission the expected decoder distortion can be used in (1) for a given encoding mode *o*. The expected distortion at the decoder can be computed based on channel conditions, coding mode used and error concealment used by the decoder.

In general, using periodic intra coded MBs with inter predicted MBs gives good results. Keeping packet sizes relatively small, especially for wireless transmission, helps reduce distortion due to packet losses. Packets formed on the basis of slice boundaries give good results.

A feedback channel between receiver and sender can be used to communicate status of the process so that the encoder can react accordingly. Information communicated by the receiver can vary from receipt/loss of packets or per frame status as proposed in [28]. Depending on the type of fame that is lost at the receiver the sender sends the following P or I frame along with redundant downsampled version of the same frame to counter burst errors and allow the decoder to prevent error drift in subsequent frames. The proposed scheme dynamically adjusts the picture quality based on changing network conditions.

The R-D optimization method has been used to generate channel-adaptive results in which the encoder is aware of channel error probability. Simulation based experiments using R-D optimization in [22] show that displayed video continues to show high PSNR even at high error rates. An adaptive control scheme based on R-D optimization is also presented in [32] in which the QP to a MPEG-4 encoder is controlled to adjust bit rate based on packet loss estimation, trip delay and observed data rate information received by the sender from the decoder.

Authors in [11] have tested the error resilience tools for a wireless network in which they have modeled error and packet losses as random and burst packet losses. They tuned the encoder based according to expected packet loss in the network and have shown how careful selection of parameters results in maintaining high PSNR value of the decoded video stream. The tools they used individually were (i) insertion of random intra coded MBs (ii) use of multiple reference frames (iii) different grouping of MBs in a frame and (iv) FMO.

5 Cross-layer Design across Application, Transport and Network Layers

Current video encoders do not encode all the frames of a video the same way. To keep it simple, some frames will be intracoded; other ones will be coded predictively, using one previous frame as basis. Classically, the way the different frames are encoded depends on the video itself. Be it with H.264 or with MPEG as shown in [4], if a frame is very similar to the previous one – which is often the case in videos – the second frame will often be predictive coded. In [4], the authors study how to adapt this choice to the video, with intra-coded frames whenever a scene change occurs. The encoder basically tries to maximize the encoded video quality for a certain bit rate.

In the case of video transmission over MANETs, it might be interesting to influence the encoder's choice based on the state of the network. In fact, the video quality on the sender side is often not the same as the quality on the receiver side. Some packets maybe lost in which case they often can not be resent in time. A frame arriving to the decoder after the time at which it should have been decoded is in most cases useless (we will later see cases where these can become useful). This may result in frames not being decoded on the receiver side. In this case, if a following frame is predictively coded with the lost frame as a basis, the decoder will not be able to properly decode this frame. Fewer predictively coded frames will reduce such problems but it will also lead to a smaller compression ratio. To maximize the quality on the receiver side, coordinating the encoder and the network layer would therefore be of interest.

1. By knowing the current packet loss rate as well as the bandwidth, the encoder can adapt the way it encodes the frames as well as its compression ratio
2. Several techniques propose use of multiple paths between the sender and the receiver. In such cases, the primary path may either be used to send all the data and the other paths would be used as back-up paths, or the different paths maybe used to send different descriptions of the video [2,24].

Application layer – Transport layer interaction

The first case is developed in [7]. The authors propose a cross layer feedback control to let the encoder adapt itself to the current network topology. Based on evidence showing that paths with higher hop counts lead to smaller throughputs, the application would adapt the quantization parameter of the encoder, and therefore the encoding rate, according to the hop-count of the paths chosen by the network layer or based on the occurrence of link breakages. But such a strategy basically consists of passing information from the network to the application layer, without any joint optimization. Cross-layer design is not really invoked. Instead of having the network layer making its choices and letting the application layer then adapt, a real cross-layer optimization would also lead to paths chosen according to the media characteristics. For high-entropy parts of the media, paths with high throughput may be chosen where as low-entropy parts of the image sequence may be accommodated by other paths. The network layer would therefore make its choices not independently but jointly "with" the application layer.

Multiple Description Coding

Video transmission over networks often uses path diversity and a traditional approach to this is layered coding. The problem with layer coding is that if one of the descriptions is missing, it might not be possible to use the other descriptions to decode.
A more popular approach that solves this problem is the multiple description coding MDC. With this approach, different compressed descriptions of the video are generated so that any one of the descriptions is enough to be able to decode with a reasonable quality, i.e. there is no hierarchy between the descriptions. The more descriptions a receiver receives, the better the received video quality. The results in a higher bit rate than layered coding but it is more resistant to path breakages.

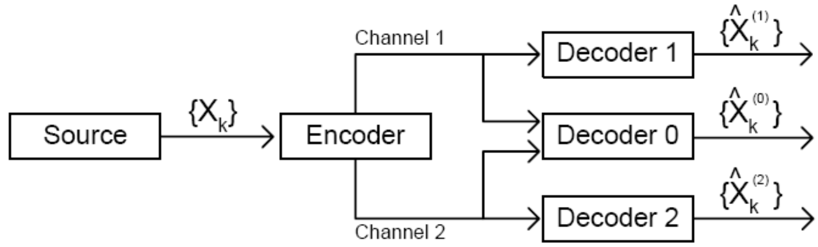


Figure 3 Multiple description codec with two descriptions ([33]). If the two descriptions are received, the central decoder is used to produce the best quality. Else one of the other two side-coders is used.

Multiple descriptions can be generated using various techniques. The most straightforward ones are time-domain and spatial-domain separations. In the first one, the frames having an odd number may be grouped to form one description with the even-numbered ones forming the other description. The second one consists in spatially dividing each frame into two subframes. One way is to include in each subframe every other pixel of the original frame. A frame of size n*m would lead to two subframes of size n*(m/2). More advanced techniques include the Multiple Description Scalar Quantization (MDSQ), in which different quantizers are used, and the Multiple Description Transform Coding (MDTC) using multiple correlating factors. All these are described in more detail in [34]. Another example is shown in Figure 3.

Most of these descriptions are *video-centric*, meaning that the algorithms only take into consideration information from the video stream to produce the descriptions. They do not make use of information relevant to the network. We believe that a cross-layer design allowing to adapt the descriptions to the network would yield better results. As we will see later, many protocols use both multiple paths and multiple descriptions but in most cases, the paths are chosen based on the number and size of the description flows. There is no feedback from the network to the algorithm producing the descriptions themselves. Allowing feedback would truly lead to a cross-layer design where paths and descriptions could be jointly optimized. For simplicity, we may just consider two parameters of the network: mobility and density of nodes. Clearly, higher the mobility, higher is the difficulty to maintain paths. Likewise, the lower the density, the fewer paths we can find. Using this two-dimensional classification of network configurations, we believe that descriptions can be computed in different ways.

A high density – low mobility network makes multi-path routing easier and allows us to fully use multiple descriptions, whereas low density networks may offer just a single path, in which cases a single or double-description scheme would be sufficient.

In [15], the authors propose a scheme over a single path with two descriptions, one with high and the other with low resolution in order to have a higher protection against packet losses. The authors, stating that maintaining other links may be difficult, concentrate on a single path. Their simulations do not compare their protocol to protocols that use multiple paths. Moreover, they consider bursty packet losses but with a path that keeps the same properties outside of these bursts. In reality, path changes may have to occur due to link breakages, which most of the time results in paths having completely different hop-counts and bandwidths.

Another way [7] of creating multiple streams is to create for each encoded frame a "redundant" packet containing a highly compressed version of the frame. This packet will come handy if the other streams do not reach the receiver. This strategy is probably best suited to address path losses. When a path is broken, several consecutive packets will be lost until a new path is established. The redundant packets, taking another path, will therefore help keep the video running, though with a lower quality. The authors here consider three different strategies to recover from link breakages:

1. Once the new path is established, the frame to be sent is intra-coded. This will help the receiver re-synchronize rapidly.
2. Once the new path is established, the frame to be sent is predictively coded using as basis a frame that is believed to have been received before the path breakage occurred. This strategy needs a lower bit-rate but needs the ability to estimate what the latest received frames were. Moreover, the delay during which the link was broken can be quite long and thus could lead to very small similarities between the two frames. In this case the predictive coding loses its advantages.
3. If the receiver has enough buffer or the delay permits it, the sender might resend all the frames that could not be sent during the route change.

Using the first strategy, the authors have been able to maintain an acceptable quality on a video stream sent over a primary path suffering from a link breakage, whereas these same strategies without another supporting stream such as the redundant packet suffer from much bigger quality drops as shown in Figure 4.

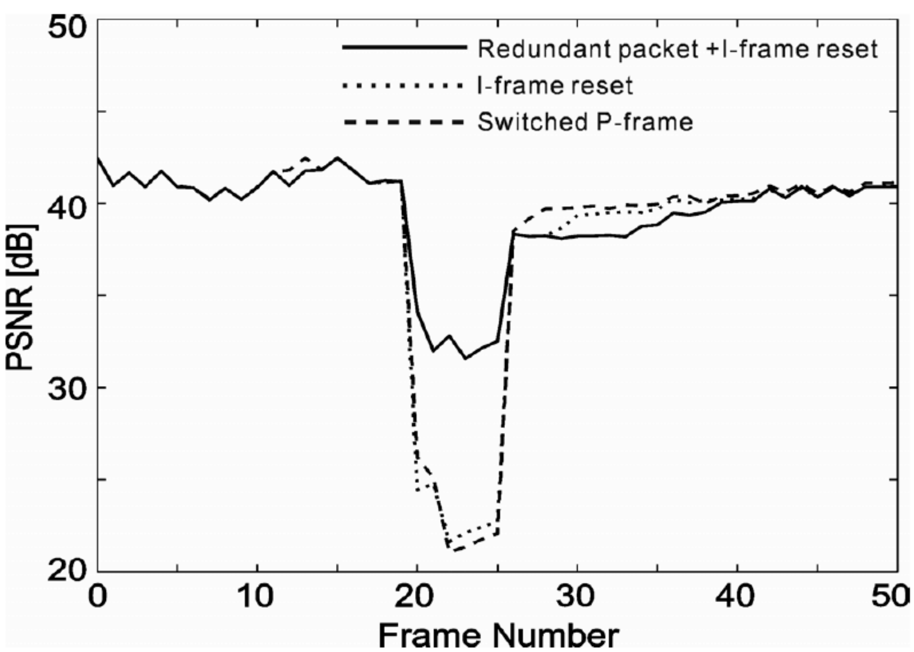


Figure 4 Recovering from link breakage using redundant packets and I frames.

This shows the advantage of using different descriptions. Moreover, the problem of link breakage is addressed in an interesting way, but the possibilities offered by cross-layer design and multiple-path, multiple-stream strategies are not fully used in this work. The descriptions are determined beforehand, there is no room – beside the quantization parameter for instance – for the encoder to fully adapt to the topology of the network. A network where for example two equivalent disjoint paths can be established would probably be better exploited using two equivalent streams rather than a big stream and a small "redundant" one.

[24] goes one step further using a truly multiple-path, multiple-stream approach. Clearly, this is the most complicated approach in terms of optimization.

To sum up, multiple description videos pose the following questions:

- How many descriptions do we want to set up?
- Should these be routed along a single or multiple paths?
- How are these paths chosen?

To answer these questions the authors of [16] have used a model and mathematical framework of ad-hoc networks. Their first assumption is that using a double-description is largely enough, as any higher number of descriptions only yields a marginal quality improvement. The main source of improvement then relies on the paths chosen. We describe this further in the following sub-section.

Multiple paths and path choice

In traditional networks, multiple paths are used to balance loads over different links instead of concentrating all the traffic on one path. As for wireless ad-hoc networks, finding several paths becomes important to ensure robustness in the communication. Using a classical simulation pattern on ns-2, with nodes randomly moving and using a multipath routing protocol based on Dynamic Source Routing their results confirm that using 3 paths yields lower packet losses than for 1 or 2 paths, but that any number of paths higher than 3 does not really improve the results as shown in Figure 5.

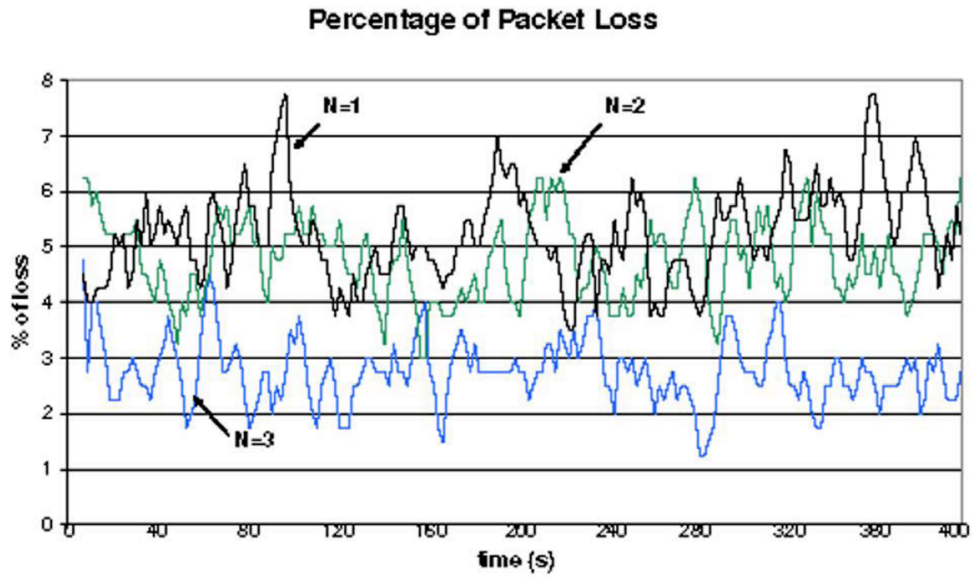


Figure 5 packet losses when 1, 2or 3 paths are used.

Multiple description coding using multi-path routing has been proven [24] to be a perfect match to wireless ad-hoc networks since these are mesh networks, i.e. multiple paths exist between the sender and the receiver. Even though the paths are not very stable, they can often be considered to have uncorrelated failure events, leading to low probabilities of having all the paths and different descriptions being down at the same time.

Knowing this, creating multiple paths – whether it is for multiple- or single-description video communication – is a difficult problem. Different works have shown (such as [5], [24]) that multiple description and multiple-path routing is indeed successful in reaching a better decoded video stream than layered coding and single-path strategies.
Most of these, though, use a set of paths that are given beforehand. They therefore do not address a real cross-layer implementation that would select both descriptions and paths as part of a join optimization.

Establishing a path in a wireless ad-hoc network is big research area that has led to many propositions. In terms of multipath routing too, there are many existing algorithms such as:

- Computing the k-shortest paths
- Node- or link-disjoint paths
- Braided multiple paths: a primary path is established between the sender and the destination, and for each node of this primary path, the algorithm looks for the best path that does not contain this node. The algorithm therefore offers detours around each of the nodes of the primary path.

All these algorithms create multiple paths using information from the network only. They are traditional in the sense that they do not need any cross-layer design, the network layer only needs information relevant to its own level.

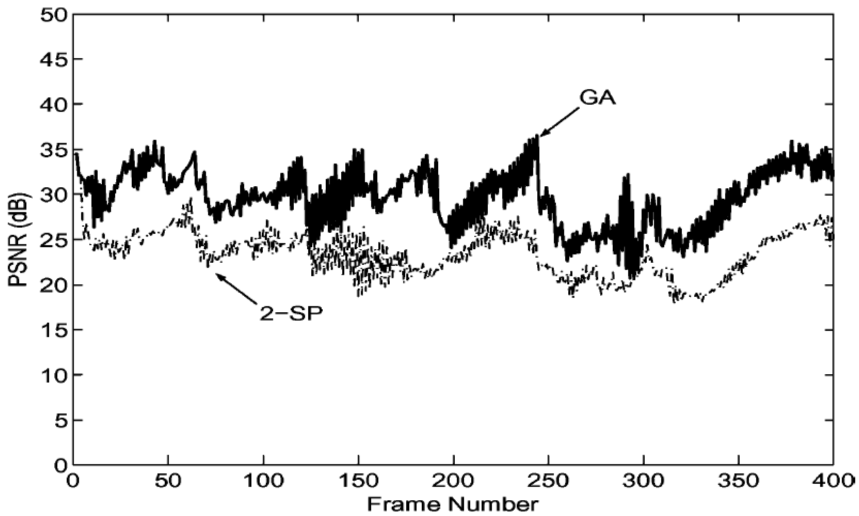
As shown in [16], the paths chosen need to take into consideration the application layer's requirements and performance. A cross-layer approach is necessary, especially when the paths have to be adapted to the multiple-description coding offered by the encoder.

One existing approach is to see the path-finding process as an optimization over the video quality on the receiver side. The authors of [5] propose such a strategy with their QoS-provisioning algorithm. They make use of different metrics (a reliability metric computed using the SINR and a mobility metric) to create a new metric that would correspond to the "quality" of a link and be QoS-provisioning. Using this metric, the possible paths given by the underlying routing protocol are compared. The authors suggest using AODV (Ad-hoc On Demand Distance Vector) or DSR (Dynamic Source Routing) as these have the ability to discover several paths. The weak point of this algorithm is that the path discovery process is done independently by the underlying protocol. It returns a mere subset of the set of possible paths and excludes any cross-layer intervention when creating this subset.

Similarly, the authors of [33] develop an approach using a given multiple description encoder. It is developed in the context of overlay networks. A function taking as input known link parameters such as packet loss rate, bandwidth, delay and jitter, and outputting the estimated video quality on the receiver site is once again used to compare paths. The optimization is on a parameter that directly concerns the application layer and is weakly cross-layer, but the optimization problem is then solved using an exhaustive search. This is clearly time and resource-consuming due to the exponential size of the solution space.

As shown in [16], such a function would include a complex ratio of high-order exponentials that can not be decomposed. We have here one of the drawbacks of cross-layer design that we mentioned earlier. The joint-optimizations lead to complicated equations on big exponential-size solution spaces. Approaches such as the trajectory-based metaheuristics (including simulated annealing and tabu search) reportedly lead to unsatisfactory solutions as they tend to be get caught in local optimums. The authors of [16] therefore proposed a solution based on Genetic algorithms to address this optimization problem.

Comparing the results obtained on a video sent on paths determined through these Genetic Algorithms (GA) and paths determined by previously mentioned network-centric algorithms, 2-Shortest Paths (2-SP) algorithm and Disjoint Path Selection Protocol (DPSP) yield results as shown in Figure 6.



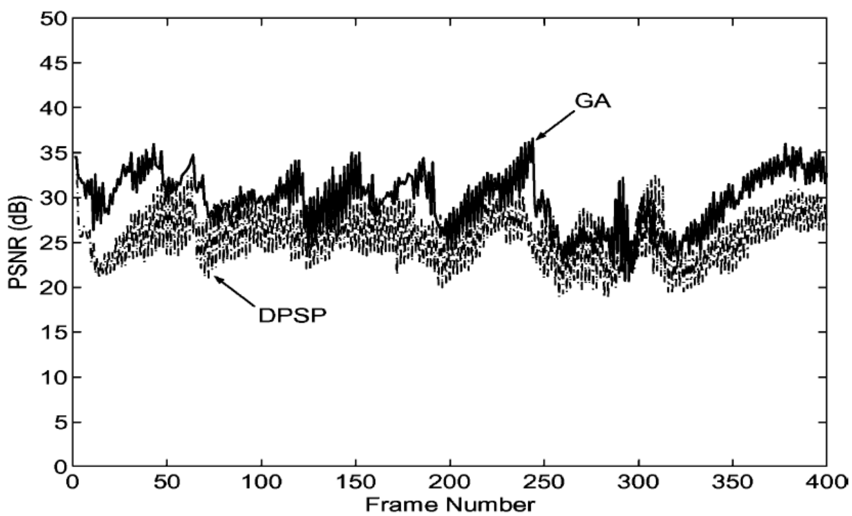


Figure 6 Results of genetic algorithms [16].

Still, such algorithms have a high computational complexity and require powerful computing in the nodes. Implementations of the algorithms show that one can reach acceptable computing times when working with laptop or other powerful devices. But it is very questionable as to how less powerful nodes such as mobile phones and devices may be able to handle such algorithms. We believe that using the "generational" property of Genetic algorithms, one may adapt these to less-powerful devices. In fact, while computing the best paths, the algorithm creates different "generations" of sets of paths with an improving quality. Video communication may therefore be started before the algorithm has finished selecting the paths, using the intermediary sets of paths that already have a good quality, and later update these paths using the better ones computed by the Genetic algorithm.

As we have seen, path determination in a jointly optimized environment is a difficult problem that has most of the time been addressed in an unsatisfying way either because the cross-layer characteristic has been put aside or because of the exponential size of the solution space could not be explored in a quick manner.

Cross-layer scheme to adapt to congestion

Yet another opportunity for cooperation is between source coding and congestion control implemented in the transport layer. In wireless networks it is difficult to determine whether a route is congested because packets can also get lost due to channel errors. In such a case statistical model can be used to determine losses due to errors. In [14] authors have modeled a system that consists of a source encoder and a buffer which is filled at a source rate. The buffer is transmitted at a certain send rate. The receiver also has a buffer which receives at a receive rate. The receive rate may be different from send rate due to packet losses. The authors create a model which establishes source rate and send rate such that source/receiver buffers do not overflow/underflow. The model uses end-to-end delay, round-trip delay, link packet loss ratio, overall packet loss ratio, amount of data receiver receives and the receiver buffer size to estimate the rates. A periodic feedback mechanism is established to send this information from the receiver to the source. The source rate is determined by factoring in the send rate, buffer status and video quality (by way of a minimum PSNR). As the feedback arrives at regular intervals the rate is adjusted such that a smooth transition in rate and quality are effected, as necessary. A window of intervals is used for the purpose. Send rate is increased/decreased if the receiver buffer will underflow/overflow. Rates are not adjusted for the case when packets are lost due to link errors. The link packet loss ratio is provided by the MAC layer. Authors have shown that their scheme effectively adapts source rate and send rate based on the packet losses in the network and maintains a minimum level of video quality. They have demonstrated throttling of send rate as packets are lost due to congestion using simulation. Source rate is adjusted by controlling QP value for base layer and by terminating the coding process for the enhancement layer when rate needs to be reduced.

Interestingly, in [14] the authors take into account the buffers at the source and receiver making their scheme more pragmatic. Although they deliberately ignore link errors when adjusting the rates we will see later that others adjust rates for all types of errors. So inherently their scheme is sub-optimal. Their scheme relies heavily on periodic feedback. They do not comment on effect of delay or loss of feedback on their scheme. As we will see in the next section use of buffer makes the encoder response sluggish. It would have been interesting to see what was the best size to use for the feedback window in order to balance the smoothness and reaction time of the adaptive response.

6 Cross-layer Design across Network, MAC, Link and Physical Layers

Wireless networks pose different challenges to the lower layers of the networking protocol stack as compared to wired networks. Wireless links are not independent entities. Each link can cause interference in other neighboring links. More complex medium access protocols are required to minimize the interference perceived by the receiver. At the same time the protocols also need to maximize the throughput by exploiting spatial reuse and concurrent transmission. Minimum use of power is also often a requirement in wireless networks since many devices in a mobile network are battery powered. Often, all of these are conflicting requirements and an ideal protocol strives to strike an optimal balance [25].

Transmission rate at the physical layer can be increased/decreased depending on the low/high interference seen at the receiver. The interference is directly controlled by the MAC protocol used. Increasing transmission power yields a higher bit rate but also increases interference. Therefore power control is tightly coupled to MAC and physical layer protocols. Yet another factor that has an impact on interference caused and transmission rate achieved is the routing protocol employed by the network layer. A routing protocol that tries to use a single hop between two distant nodes at low transmission rate may not perform as well as one that uses multiple nodes close to each other to transmit data using higher transmission rate.

To improve achieved bit rate, strong forward error correcting (FEC) coding can be used. However that degrades the source coding at the application layer. In wireless networks the physical layer has an additional attribute of signal strength (power). Typically bit rate and power can be considered to be parameters of the physical layer or the link layer. Bit rate at the physical/link layer is controlled by using appropriate coding such as repetition coding or convolution codes. For medium access, MAC protocols may allow only one node to transmit at a time using time division multiple access (TSMA) or CSMA or token passing. MAC protocols also create an exclusion region around the destination such that all nodes in the region are silent during the transmission while nodes outside the region can still be active. The MAC protocol may also create total exclusion in which only one node in the entire networks sends at a time whereas the other extreme would be to allow all nodes to be active simultaneously.

Therefore, protocols at one layer affect performance of other layers making it imperative to use cross-layer techniques in order to design optimal wireless networks.

The author in [25] has proposed cross-layer schemes to derive optimal wireless networks by adapting network layer, link layer and physical layer protocols. The author optimizes the maximum achievable data rate and fairness for two types of networks: (i) ultra wide band (UWB) in which link rates are linear functions of signal-to-interference-and-noise-ratio (SINR) and (ii) narrow band networks (like 802.11 or CDMA) in which the link rates are sub-linear, typically log of SINR.

In UWB networks the author recommends adapting the bit rate based on interference detected by the receiver and creating an exclusion region based on signal strength. Nodes outside the established exclusion region are allowed to transmit. The exclusion region is established based on the transmission power of the source by the MAC protocol. For low-rate networks with small power constraints the author shows that the exclusion region is small enough to include just the destination node.

Interestingly the author's study shows that some protocol choices do not require cross-layer consideration. For e.g. MAC protocol need not depend on the routing protocol. At the same time the ideal routing protocol is one that uses minimum energy and suffers minimum loss – this translates into using a route made up of smallest hop lengths from source to destination instead of a direct path. Any standard routing protocol such as AODV or DSR can be used for the purpose. The study also shows that nodes should transmit using maximum allowed power.

In narrow-band networks the author shows that the exclusion region size depends on power used by the source as well as link lengths and hence the MAC protocol now has to depend on the routing protocol. Furthermore the routing protocol used – multi-hop versus direct link – also depends on the power constraints. For small power constraints multi-hop routes are better and for large power constraints direct paths are better. Hence the routing protocol now depends on link/physical layer.

While the work done by the author does demonstrate the interplay between protocol layers for wireless networks to achieve highest throughput, he/she has not considered energy consumption and latency as factors. The author has also taken a general view of transmission and does not take into account characteristics and constraints posed by specific type of data such as audio/video. However we believe that as a fundamental study it provides insights into the interdependence of the lower layers.

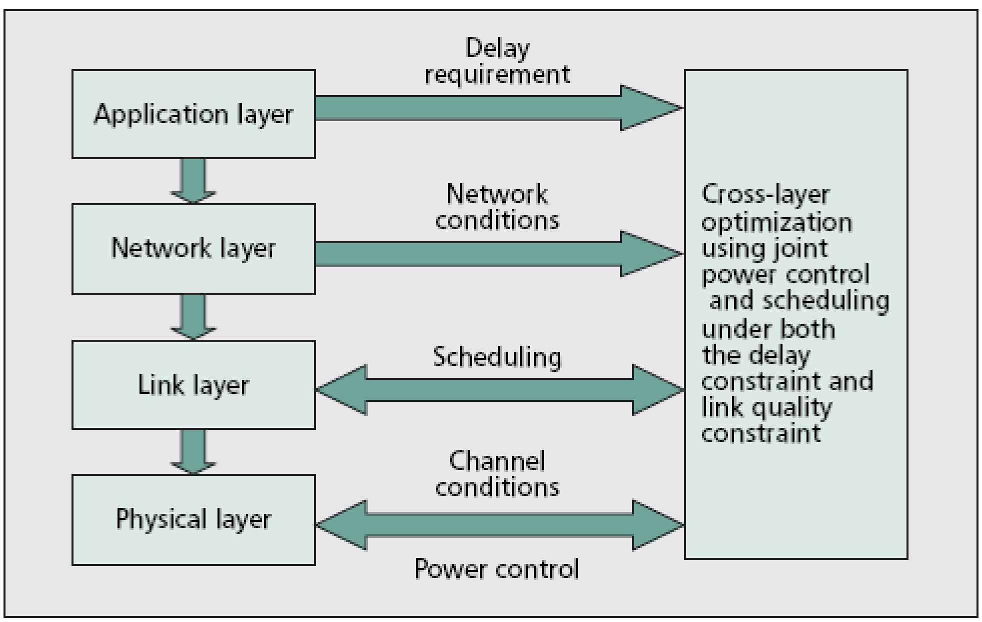


Figure 7 Cross-layer scheme suggested in [6]

Authors in [6] have presented a cross-layer scheme in which physical layer and link layer protocols cooperate to achieve delay constrained, high quality video transmission. The authors have assumed a CDMA network with two non-overlapping frequency channels – one for data and other for control. SINR values of each node are broadcast and available to all other nodes. Fading values are measured at the receiver and communicated to the source. The scheme also takes inputs from the application layer and network layer when making decisions at link and physical layers as shown in Figure 7. The scheme controls delay seen by a packet by controlling power. It estimates delay from source to destination based on the number of hops in the route computed by the network layer, a G/G/1 model of queuing system of the network, estimated traffic generated at a node and observed delay at the node. If the computed delay is more than the delay target specified by the application layer, the hop counts are reduced iteratively until the target can be met by assuming that transmitting power can be increased. The transmission schedule for each packet (either generated locally or received) at a node is determined based on remaining lifetime and remaining hops of a packet. Shortest life packet that needs to travel the farthest receives higher priority. Based on past fading value, effect of Doppler spread is incorporated into estimating the power required to transmit by the source for a given value of SINR of the receiver. Authors have proposed a distributed algorithm in which each node iteratively computes the power it will use such that overall power consumption in the network is minimized. If in a given slot nodes do not converge the node with lowest SINR defers transmission to next slot. Figure 8 shows the block diagram of the system. Authors have demonstrated increase in achieved throughput and reduction in the total power consumption, using simulations.

Authors in [6] optimize the scheme for meeting delay constraints and minimizing power consumption in the network but ignore the aspect of maximum achievable throughput. It is unclear whether their scheduling priority scheme impacts fairness of access to bandwidth. It would appear that nodes that are further apart may get higher share of the bandwidth than the nodes close to each other.

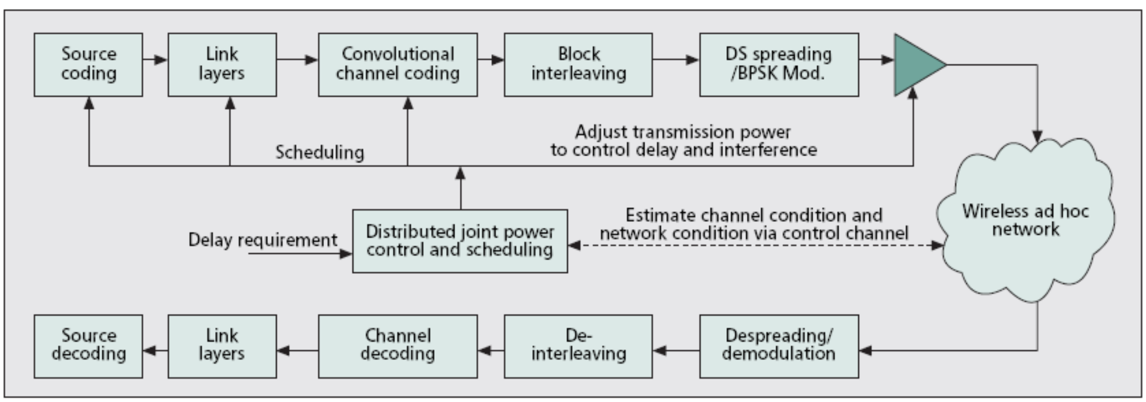


Figure 8 Block diagram of system that employs cross-layer scheme in [6]

There is an interesting contrast between the above two schemes in [25] and [6]. In [6] the authors propose reducing hop counts to reduce delay, but as seen in [25] this can reduce the maximum achievable throughput. Author in [25] proposes using maximum power per node after establishing the exclusion region at destination for max throughput, but if power is to be conserved as is done in [6] then one has to give up throughput. In [25] the MAC protocol was independent of the route selected but in [6] the scheduling used for transmission was directly dependent on number of hops in the route. The main difference between the two is that the type of traffic is considered in [6]. This comparison demonstrates how the type of traffic, such as video with delivery time constraints, in a network can make protocols and their interaction more complex and non-intuitive.

Authors in [9] present a cross layer design in which the physical layer cooperates with the application layer for real time video transmission in a MIMO wireless network. The design maximizes channel capacity for a given target constant bit error rate (BER) while maintaining a minimum guaranteed data rate. The video encoder in the application layer produces a layered coded bit stream consisting of a base layer and an enhancement layer. Authors contend that the encoder by itself cannot respond to changing channel conditions in time due to factors such output buffer size which introduces non deterministic delay depending on occupancy, type of frame being encoded (I or P or B) and complexity of the image. (This is in direct contrast to the scheme in [14]). Hence the authors propose that the physical layer make the decision to adapt the bit rate. Physical layer uses sufficient power to transmit the base layer at a constant BER therefore guaranteeing that the base layer is transmitted. Authors have derived what such power level would be based on the channel conditions. For the enhancement layer the physical layer chooses maximal rate possible within a coherence period (when transmission is not interrupted) using nominal power and truncates the remaining bits outside of it.

While the authors in [9] claim their scheme to be cross-layer, it is only the physical layer that makes the appropriate decision once the encoder has produced the bit stream. A more adaptive scheme would have been to let the encoder adapt if it could within the required time period.

7 Cross-layer Design Across all Layers

When video is transmitted over wireless networks, typically channel coding is used in order to recover from errors. Errors can occur because of loss of packets due to queuing delay, channel fading, interference, multipath effects and link failures due to mobility. Therefore the effective transmission rate is reduced and has to be shared between source coding and channel coding. If a source produces more data than the effective transmission rate then further data is lost.

Forward error correction (FEC) and automatic repeat-request (ARQ) are two techniques commonly used to mitigate transmission errors. ARQ requires retransmission of lost or packets with errors. Since this introduces additional delay in transmission it is impractical to use ARQ for real-time multimedia transmission. FEC techniques take up additional transmission bandwidth but require no feedback. The receiver attempts to recover missing or incorrect data using the redundancy built into the coding scheme. FEC coding is also effective in dealing with bursty errors which are common in wireless ad-hoc networks. The amount of bandwidth required for FEC can be controlled using appropriate coding scheme.

One way to maintain as high a quality of video as possible would be to control the rate in accordance with the effective transmission rate between a source and destination in a network. Authors in [35] take a closer look at characteristics of data transmission and incorporate feedback mechanism in the network layer. They adapt the routing protocols such as ad-hoc on-demand distance vector (AODV) and optimized link state routing (OLSR) to maintain additional data such as packet-loss rate, bandwidth and interference conditions for each hop in the routing table at each node. They compute an effective transmission rate based on this info for a given route which is used to select a channel coding scheme to meet given delay and prevailing channel conditions. This improves end-to-end quality of service (QoS) by taking into account both effective transmission rate and channel error effects.

Authors use Reed-Solomon (RS) channel coding in which successive *k* packets are aligned vertically. Vertically aligned q-bit symbols from the aligned packets are used to generate *n-k* parity packets using an RS(*n, k*) scheme. *n* packets are then transmitted using RTP/UDP/IP protocol over the wireless network. By controlling the RS(*n, k*) scheme the authors control the amount of FEC data transmitted. For e.g. they propose a QCIF image be transmitted using 9 packets in which case they can vary the number of parity packets from 0 to 6. Higher the number of parity packets, lower the effective transmission rate but better the ability of the receiver to correct errors.

The effective transmission rate reduces as the number of hops along the path increase. Authors in [35] assume an ad-hoc network of homogeneous nodes generating same amount of traffic in a similar pattern. They first formulate a mechanism to compute effective bandwidth for multihop transmission (effective bandwidth decreases as more and more nodes get affected at each successive hop). In [36] authors have created a model in which a node X_i successfully transmits to another node X_j as long as all other nodes X_k that transmit simultaneously are not within a certain radius of X_i:

$|X_k - X_j| \geq (1+\delta)\,|X_i - X_j|$ Where $|X_k - X_j|$ is distance between X_k and X_j.

The parameter δ defines the "guard band" between two neighboring transmitters. Authors in [36] show that based on the above interference model, as number of nodes in the network increase the capacity goes to zero. In [35] the authors assume a more realistic scenario in which each node is aware of all neighboring nodes that can cause interference by way of modified network layer protocol. Assuming a simplistic model in which all traffic travels across *L* hops, a given transmission rate W_i and number of interference neighbors c_i for each node in the route, the authors derive a formula for an effective transmission rate, $R_{effective}$ between a source and a destination:

$$R_{effective} \propto \frac{1}{L}\min\left(\frac{W_i}{c_i}\right)$$

Authors then propose that the routing algorithm be modified to provide the above information L, W_i, and c_i to the source node when the route is established. Hop count L is usually available to the source in systems using AODV protocol. The transmission rate W_i can be made available for each node if this information is included in the routing update messages sent back to the source from the destination. RTS/CTS mechanism commonly used in wireless networks can be used to establish number of neighboring nodes c_i when nodes reply back to RTS. Or probing packets can be sent periodically to establish this count (RTS mechanism may establish incorrect value when nodes do not have any data to send out). Based on connectivity, routing protocols will have a choice of routes whose effective rate can be computed as per above equation. The route that maximizes the bound on effective transmission can be selected. If the route is already established, all information is available to compute $R_{effective}$. If a new route or a route change is performed then the information is gathered as the route is established and $R_{effective}$ is computed.

Authors then propose a cross-layer rate-control scheme in which application layer, network layer and MAC layers cooperate to produce a video bit stream for a given delay constraint and channel condition as shown in Figure 9.

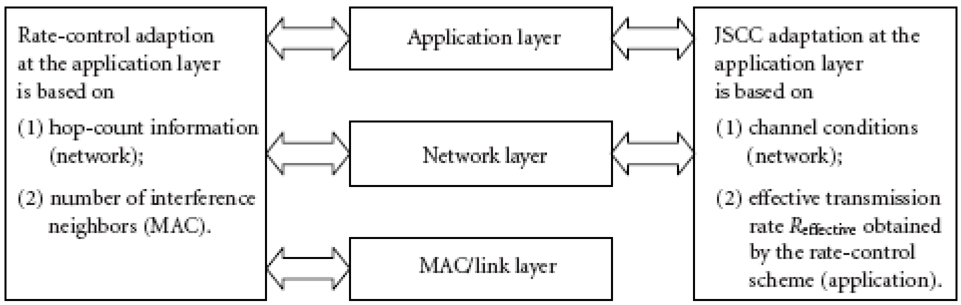


Figure 9 Cross-layer scheme proposed in [35]

As described earlier, first $R_{effective}$ is computed. The routing protocol also collects information on channel conditions such as delays and packet-loss. This information is then used to select the optimal source/channel coding rates. Authors' scheme supports a given combination of source coding and channel coding rates. Channel coding is performed using RS coding as described earlier. The algorithm selects a combination that produces the highest PSNR for the given frame such that transmission delay for such a bit stream (including queuing delay) does not exceed a given maximum delay. Transmission delays at each node are communicated to the source node using the routing protocol. Authors also propose a scheme to estimate packet loss and burst length over the entire route using a Gilbert channel model and a distortion estimation scheme when source coding bit rate is reduced in order to accommodate channel coding. Distortion also depends on the video sequence, the intracoding scheme and channel errors. The following algorithm is then used for coding:

1. For a given delay constraint, find a feasible set of RS codes and source coding rates.
2. Using the distortion model, estimate distortion for each of the feasible pairs of RS/source codes.
3. Select the feasible pair with the lowest distortion for frames within the current routing update interval.

Authors show that their scheme produces better results than when no FEC coding is used at all and when a fixed FEC coding scheme is used.

This is one of the rare studies in which authors take in to account such a large number of factors at different layers to control the video transmission system. While the authors in [35] have compared their scheme against fixed FEC and no-FEC schemes, they have not compared their scheme to any other cross-layer scheme which would have been very interesting. They also assume an ideal and homogeneous system in which the source generates packets from encoding a frame at fixed intervals. We believe this is a limitation in their study since it is not taking into account the variable queuing delay at the node. While they do propose models for estimating interference, effective transmission rates, loss patterns and distortion, it is unclear whether any experiments were run to validate their models. Their distortion estimation can be enhanced by taking into account error concealment employed at the receiver. Furthermore, their scheme involves interaction between layers ranging from application down to MAC, it is unclear what is the impact of the delay caused as their algorithm computes the effective coding rates for real-time video. A better understanding of the complexity of the algorithm is essential to assess practical implementation of the scheme.

Another thorough and comprehensive cross-layer design has been proposed in [23]. The scheme jointly optimizes source coding, packet scheduling, routing, link capacity assignment and link layer techniques to deliver high quality real-time video. It is applied dynamically at each node in the network to adapt to changing channel conditions. Figure 10 shows the interactions between various layers that the scheme proposes.

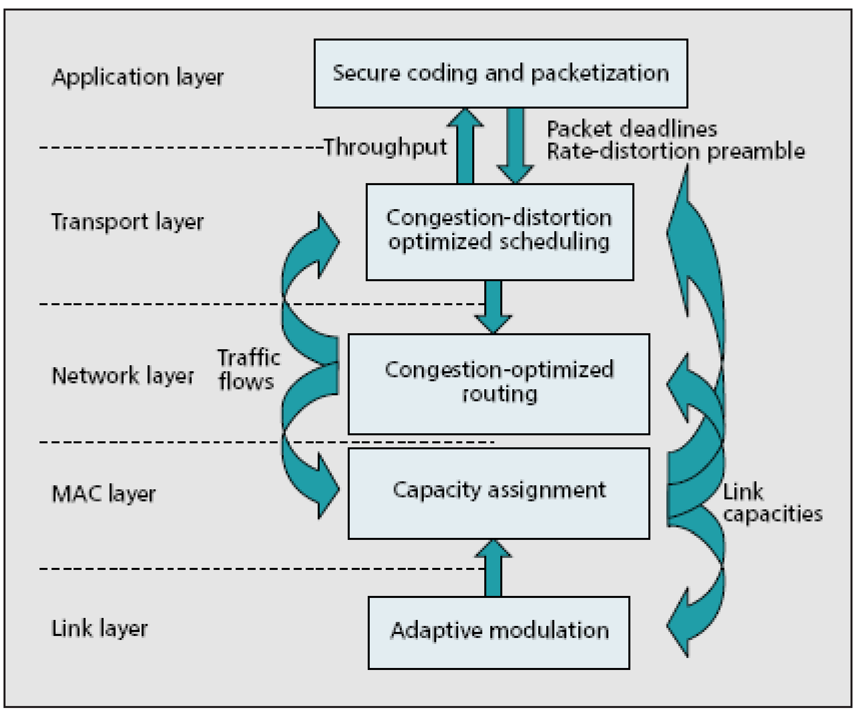


Figure 10 Cross-layer scheme proposed in [23]

At the link layer link rates are maximized as per channel conditions thereby increasing capacity of the network. Based on this information, the MAC layer assigns time slots, codes or frequency bands to each of the links. The network layer cooperates with the MAC layer to assign flows that minimize congestion. Optimal capacity and network flow assignments are established by iteratively exchanging sub-optimal solution between the two layers. At the transport layer transmission and re-transmission of video packets are scheduled based on a congestion-distortion optimization. Rate-distortion trade-off and delay requirements are used to select appropriate coding rate by the application layer.

Link layer parameters that can be adapted to improve the data rate are modulation, coding, transmitter power, target BER and symbol rate. Authors in [23] propose two adaptive link layer techniques: (i) given current SINR and link layer parameters, optimize throughput and (ii) for a fixed packet length, optimize symbol rate and constellation size for maximal throughput. Authors provide a mechanism to compute optimal packet length for given symbol system, probability of error which depends on modulation type and link SINR. In general, for high SINR regions where link quality is high, maximum symbol rate for the highest constellation size results in larger packet lengths and for low SINR regions error rate is decreased by adding redundancy. Authors compute the throughput achieved on a link based on the choice of optimal values for the above parameters.

For the case when multiple nodes within a region may transmit, a transmission strategy (such as multi-hop, single-hop, spatial reuse, etc.) is used to coordinate medium access. Once the achievable rate for each link is computed the time share for each node is established for the given transmission strategy. Thus an achievable capacity region that characterizes data rates simultaneously achievable between pairs of source-destination nodes can be established by considering all different transmission strategies as shown in figure 3 [37].

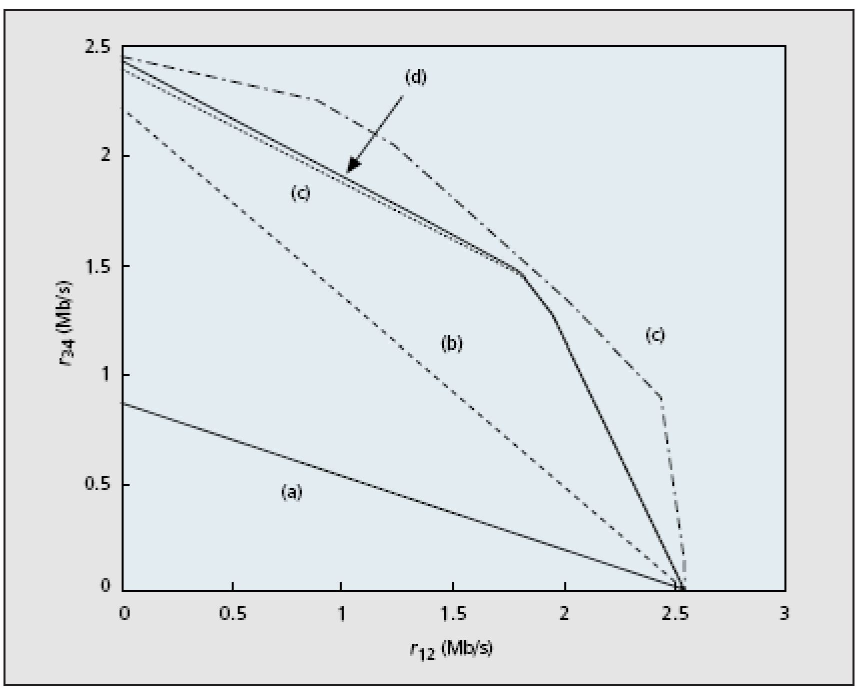


Figure 11 Capacity regions of an example ad-hoc network. (a) Single-hop routing, no spatial reuse. (b)Multi-hop routing, no spatial reuse. (c) Multihop-routing with spatial reuse. (d) Two –level power control added to (c). (e) Successive interference cancellation added to (c).

An operating point within this region can then be selected based on optimal flow such that congestion is minimized. Authors provide a mechanism to derive the optimum capacity and flow such that link utilization is maximized over the links of the network. This operating point gives the bit rate that can be used. Authors recommend that multipath routes be used to split the flow. Path diversity provides higher aggregate data rate through spatial reuse and have uncorrelated loss patterns although they increase contention at MAC layer because more links are involved. Also the complexity of maintaining multiple routes is higher. The cross-layer scheme allows for efficient utilization of available link bandwidths. Authors show that this cross-layer coordination allows a network to sustain much higher data rates while maintaining high video quality measured using PSNR as compared to oblivious layer routing even if individual layers within the protocol stack are optimized as shown in Figure 12.

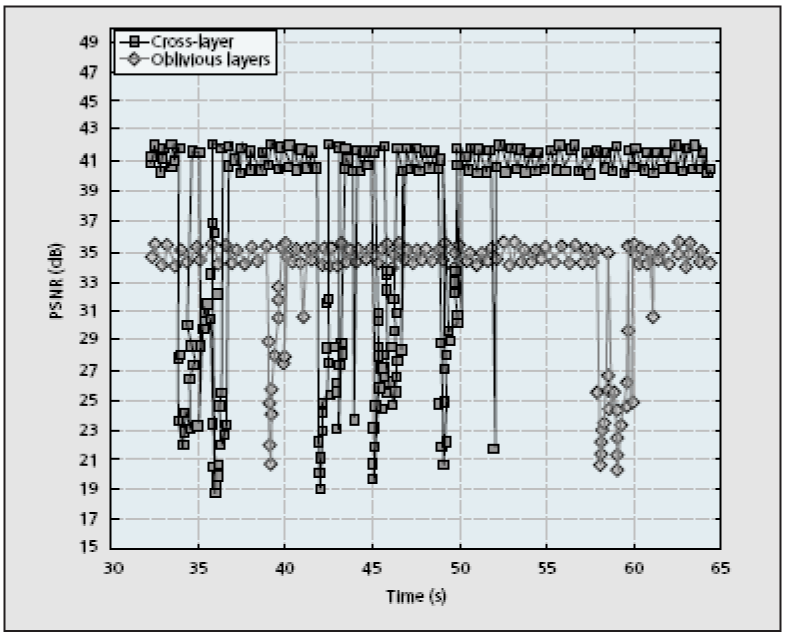


Figure 12 Video quality results of cross-layer and oblivious layer optimizations in wireless network.

At the transport layer, protocols such as TCP are unaware of delay requirements and importance of packets. Authors propose a congestion-distortion (CoDiO) optimized scheduler that limits end-to-end delay and maintains video quality. The scheduler transmits most important packets such that network congestion is minimized. Important packets are prioritized over less important ones which may even get dropped in flight. The scheduler also avoids transmitting packets in bursts so that queuing delay is minimized.

The application layer determines the optimal rate for the video stream based on rate-distortion characteristics, delay constraints and prevailing network conditions. Authors have derived a model in [38] that measures video quality using encoder quantization and packet loss due to congestion. This model is used to determine the highest sustainable data rate in conjunction with the capacity-flow optimization described earlier. Experiments show that delay constraints reduce the effective network capacity and as the traffic approaches this capacity the delay gets exponentially worse. Hence it is essential to operate the network within its capacity limit.

Cross-layer scheme in [23] is a very comprehensive scheme that optimizes across all the network layers. It would have been interesting to measure gains of different portions of the scheme and perform a cost-gain analysis to see which optimization gives the best results and what effort is required for the same. Also the authors have not compared their scheme with schemes proposed by others even if they are not as comprehensive. This scheme is also complex and requires assessment for practical implementation.

8 Future Areas of Research

Although the field is relatively new there are ample studies that demonstrate the utility of cross-layer designs in enabling efficient utilization of a wireless network and high quality user experience. Much work has been done on optimization across successive layers of the network protocol stack. There are also some studies that have looked at optimizing more layers of the stack. Typically, proposals include exchange of data between one or two layers which are applied on theoretical models to assess how each layer should adapt locally. In our discussion we call this "local optimization". In general, proposals consist of one or more such local optimizations.

In this section we share potential areas for future research. The areas in which future work can be done are in enhancing media-aware routing protocols, cooperation between more layers of the protocol stack, better characterization of the network for more accurate analysis and in general a mechanism to compare and measure gains of different techniques. We have also proposed a unified framework to facilitate such research.

8.1 Enhanced Routing Protocols

Most cross-layer studies do not go deeper into the characteristics of the underlying routing protocols. These protocols have been used in some of the work to gather information on other nodes or to establish specific links. In most cases, the simulations are done using either a reactive or a pro-active protocol, but without specifically studying how these two types of routing protocols are relevant for multimedia communications.

Nodes using a pro-active protocol will constantly gather information about their neighbours and nodes situated farther away, and maintain routing tables so that a path to a given node can be computed at any time without much delay. Even if there is no multimedia traffic, a pro-active protocol will keep the network busy through this information exchange. This becomes critical for low-bandwidth networks or networks with a very high number of nodes.

On the other hand, reactive protocols will only seek to establish a path when it is demanded. This leads to a higher delay in establishing a path but also lower bandwidth usage when path changes are not needed.

Based on this, the first distinction that can be between the two pertains to the delay-tolerance of the communication in progress. It turns out that real-time video communication is very sensitive to delay and most of the time can not accommodate the delay required for path changes with a reactive protocol. On the other hand, stored videos are often dispatched using a bigger buffer at the receiver, thus being able to accommodate higher delays and jitters. Moreover, since the video is already stored and ready before the communication even begins, the sender knows in advance how much bandwidth it will need and for how long. It also knows which parts of the video have higher entropy and which parts have lower entropy. The system should therefore be able to adapt itself in advance to the changing properties of the video feed, thus making it possible for a reactive protocol to be used. Real-time video of course do not have this property since the sender would have to react *live* to the video characteristics, without being able to do path changes and adaptations in advance.

In [1], the authors have quickly compared the reactive protocol AODV (Ad-Hoc On demand Distance Vector) and a new link-state protocol (pro-active protocol), while using multi-path routing. They consider a real-time video communication and confirm the fact that a very important delay (up to half a second) can arise whenever there is a link break whereas the proactive protocol can maintain this delay below 20ms as shown in Figure 13 and Figure 14.

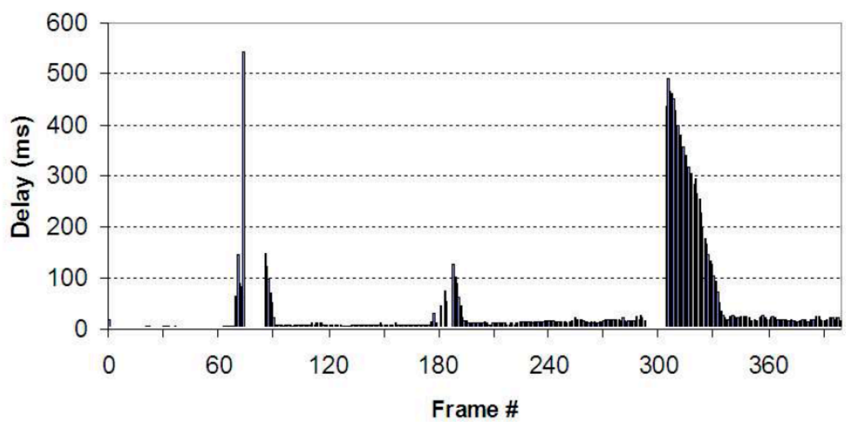


Figure 13 Delay when AODV is used

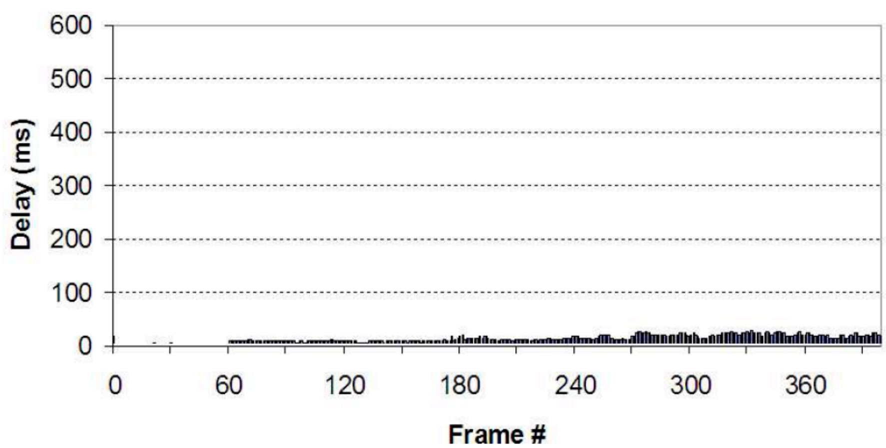


Figure 14 Delay when a pro-active protocol is used

Furthermore, information sharing is an important aspect of cross-layer design since most joint-optimizations rely on sets of parameters and metrics found in different layers. Since pro-active protocols constantly gather information and update their tables, they are a good source of information for such designs. Reactive protocols also give information even though the update rate is not the same. For example [5, 7, 18] are based on reactive protocols, with [18] making use of the hop-count computed by AODV or DSR (Dynamic Source Routing). As mentioned earlier, the hop-count is a possible metric to evaluate the throughput of a path and is used as such to adapt the quantization parameter of the encoder.

[16] makes use of the pro-active protocol OLSR (Optimized Link State Routing) to gather as much information as possible. Along with the neighbour sensing used in OLSR, link characteristics such as bandwidth, loss rate and delay are stored at each node and frequently updated.

The choice between pro-active and reactive protocols is often considered to be a choice between low delay and low bandwidth usage. However as we have highlighted above this choice goes beyond such a consideration as information exchange is a crucial step in any cross-layer design. Moreover, many designs we have seen need paths to be determined in an *application-centric* way rather than a *network-centric* one. Both classes of protocols can be adapted to be more application-centric. Proactive protocols have an advantage when it comes to dynamically adapting the paths to the network topology and the encoder needs. This aspect of routing protocols needs to be explored further.

8.2 Analyzing Proposals and Schemes

At the moment it is not easy to assess relative quality of the proposed schemes. Different schemes use different models to model the network conditions under different assumptions. Simulation studies also use different benchmarks to measure their effectiveness.

While all studies propose new schemes, they do not comment on implementation cost and potential impact on the overall system. We have seen how certain cross-layer design propositions lead to algorithms that need large computational power. Some of the studies we have seen need at some point an exhaustive search in a space of exponential size. Even if [16] proposes genetic algorithms to reduce the complexity, we still believe that most of the devices that will be used in MANETs (such as handheld devices) do not have the efficiency and power required to perform these computations in a short time.

Moreover, most of the studies simulate a single video stream over a given network of nodes. It would be interesting to see how multiple streams going through multiple pairs of senders-receivers would be handled. With multiple paths generated for multiple communications going on, the load on the network can rapidly grow and lead to major congestion. Nodes participating in various communications - either as sender, receiver or just as intermediate node - would need to make choices not only for each stream individually, but also between the different streams. If for example a node is an intermediate node for two different streams, it would prioritize the "weakest" stream, i.e.

the one that would be hardly hit if the node stopped relaying the packets. Likewise, video streams sent from a sender to multiple receivers at the same time are an area of research that should have many applications in the future but has not lead to much work yet.

A study that compares gains from different local optimizations as well as their cost will be valuable to guide future work as also suggested in [23]. It is likely that interaction between just a few layers may yield the maximum benefit. At the same time it would be interesting to know whether some layers should operate independently for better overall performance, as was seen in [25] in the case of UWB type networks.

While all studies use simulations to assess their schemes, there seems to be little work done in terms of measuring gains in an actual setting.

Unlike for internet, no study seems to have been performed that has measured parameters in an actual wireless network to create a model based on real-life measurements that the researchers can use.

More work is needed in the above areas.

8.3 A Framework for Cross-layer Design

Many proposals include feedback to control the coding rate of the application layer. But not many (except [23]) use the data available from application layer to feed-forward to downstream layers for optimization. For stored video (and also possibly for real-time video) information such as frame complexity can be used so that downstream layers can allocate for resources ahead of time. For e.g. network layer can work on creating appropriate routes (multi-path or single-path), MAC layer and link layer can budget for the appropriate bit rates. The information can be used to reserve resources dynamically as complexity changes during transmission and network conditions change due to link breakages and error conditions.

Another area of future research would be to devise a single set of cooperative algorithms that takes inputs from different layers and perform a global optimization across all layers from application to physical. For e.g. an algorithm that takes into account, image complexity, and requests resources from the lower layers which take into account the number of routes to be established as per current congestion conditions in the network and select optimum routes. For the selected routes, the MAC layer and link layer then decide on scheduling, FEC and power control scheme based on channel conditions to maximize throughput of the system and maintain high quality of video delivery to the receiver.

We propose a framework that can be used for such research as shown in Figure 15.

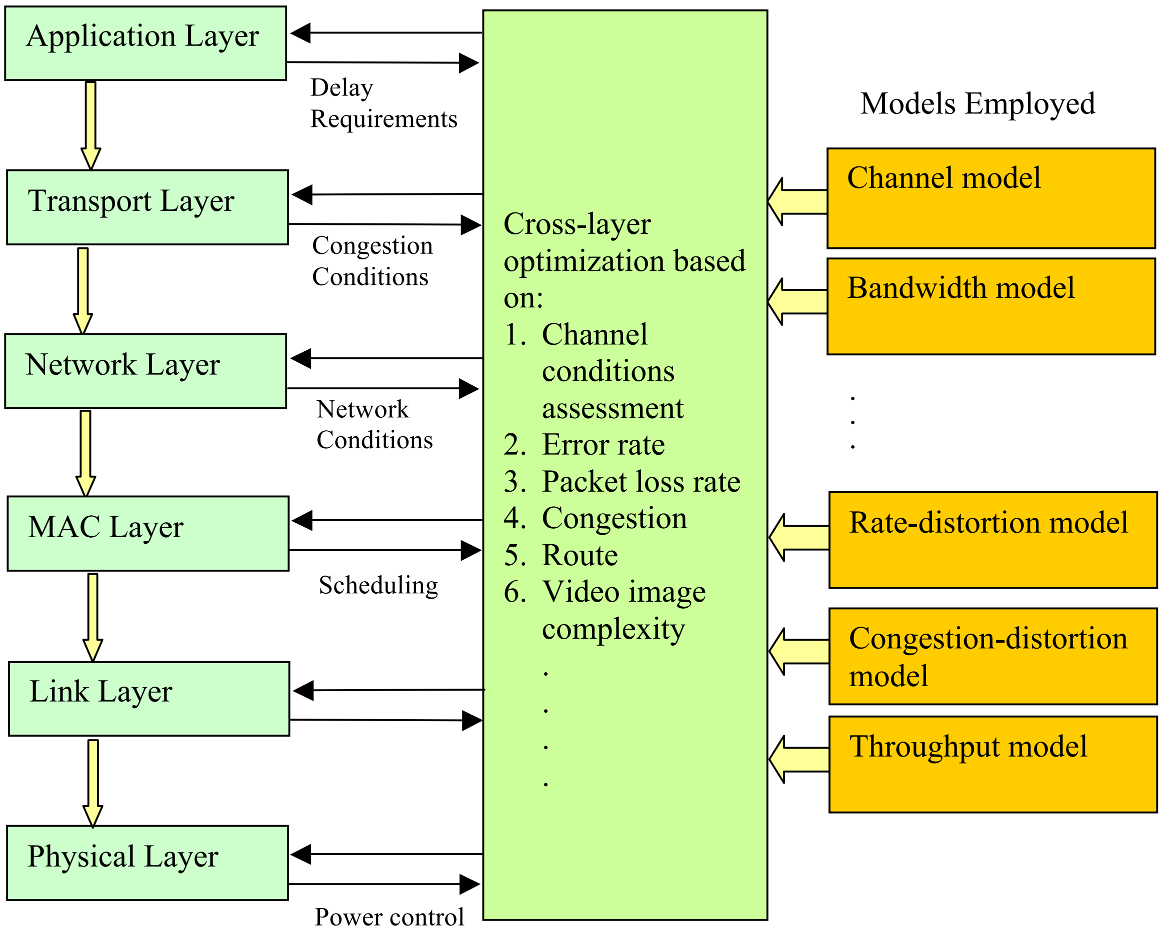


Figure 15 Block diagram of the proposed unified framework for cross-layer design research.

The framework consists of the following components:

1. OSI model protocol stack. Implementations of the layers may be modified as necessary. Note that for wireless networks we have shown MAC and link layers separately.
2. A set of inputs that characterize prevailing network conditions. They are provided by the different layers as they are gathered using existing or modified protocols.
3. A set of models that characterize the different aspects of the network. These models are derived apriori.
4. A set of cross-layer optimization algorithms that use the inputs and models to and generate outputs. The outputs are used by the protocol layers to adapt as desired by the optimization algorithms.

The framework proposed defines a set of inputs that may be provided by different layers and a set of models that are used by the optimization algorithms. The models may be changed as can the algorithms. This allows researchers to use different models for different types of wireless networks as well as refined models for the same network. This framework allows for free flow of information from all layers to each other. The actual implementation of each layer can of course be modified to accommodate for desired changes.

Inputs provided to the cross-layer algorithms come from the layers implemented within the source node as well as from feedback received from the nodes along the route including the destination. The different types of inputs that may be generated are listed below. Information gathered from feedback is marked as "fed back" as opposed to information gathered locally.

1. Link layer
    a. FEC coding rate
    b. SINR
    c. Sustainable bandwidth (local and fed back)
2. MAC layer
    a. Channel loss ratio of links (local and fed back)
    b. Transmission delay (local and fed back)
    c. Number of interference neighbors (local and fed back)
3. Network layer
    a. Hop count of the route
    b. Transmission delay (fed back)
4. Transport layer
    a. Packet loss ratio computed by the destination (fed back)
    b. Queuing delay (fed back)
5. Application layer
    a. Receiver buffer size and status (fed back)
    b. Amount of data received by the receiver (fed back)
    c. End-to-end delay (fed back)
    d. Distortion estimate (fed back)
    e. Delay constraint
    f. Image complexity

Various models that can be employed to model network behavior are:

1. Optimal packet length estimation to maximize throughput [23]
2. Point-to-point link throughput [23]
3. Capacity region estimation [23]
4. Congestion-distortion model
5. Rate-distortion model
6. Interference model [35]
7. Throughput capacity model [35]
8. Channel loss model
9. Distortion estimation model
10. Channel coding policy model
11. Error concealment model
12. Bit error rate estimation model
13. Power optimization model

Cross-layer design implementations tend to be complex. A framework on the above lines will help provide insights into the system. It will also provide enough freedom to the designer to try various combinations of optimizations, both tried and untried so far. Coupled with the ability to alter models and algorithms it will enable design of better algorithms which are more adaptive, distributed, robust and scalable [23].

8.4 Other Areas

1. Another area that has not been explored is to analyze traffic and mobility patterns in ad-hoc networks and how they affect error rates and delay in the network. It would be interesting to study the impact when the protocols adjust to changing topologies. Their behavior models can be created and applied in cross-layer techniques
2. Node density and mobility are other interesting aspects of a mobile network. It may be helpful to determine average density and mobility of the nodes in the network. A highly mobile node may be more prone to error or link breakage. A routing protocol may want to avoid such a node or establish another route or request the application layer to introduce more redundancy in the bit stream in anticipation of poor transmission by the node.

   The density and mobility are general metrics that can be rapidly computed through neighbor sensing and the information can then be propagated to other nodes in the network. These metrics could be used to determine the robustness and duration of paths when they are being established or when their status change. Moreover, the specific mobility of nodes in a route can help make a choice of intermediate nodes and help in relaying information where one might prioritize packets coming from very mobile nodes.

9 Conclusion

Mobile ad-hoc networks pose significant challenges for multimedia communication. While work as been done in optimizing individual network layers, recently cross-layer techniques have been proposed. Cross-layer designs involve cooperation of multiple layers in the OSI model networking stack to optimize for a given set of parameters. As we have seen, mathematical models and simulations have been used to confirm that cross-layer designs perform better than traditional designs. Studies have shown that cross-layer designs are able to deliver higher quality video in face of data errors, packet losses and queuing delays in mobile wireless networks. At the same time, cross-layer designs can also be used to optimize for other system parameters of the network such as overall throughput, bit rates, fair network access, and power consumption among others.

So far we have seen that cross-layer designs have been mostly applied to the upper-layers of the stack, such as multiple-path, multiple-stream strategies. More work is required to be done to apply cross-layer designs to the lower layers and more importantly across the entire protocol stack.

Although some studies classify themselves as cross-layer designs, they only use information exchanged between adjacent layers. We believe that substantial gains will come about when joint-optimization is applied across more and more layers as is seen in some studies that employ true cross-layer designs.

The designs we have seen employ a wide range of models and parameters in the optimization process. Therefore, we believe, a unified framework for future research would be very helpful. We have proposed one such framework and have identified few areas for further work in cross-layer designs. There is also a lack of real-life tests to help assess various proposed schemes in real-life settings. Future experiments involving large geographical areas and real-life moving nodes would be helpful in validating these advances.

Finally, the main drawback of cross-layer designs is that they increase the complexity of the algorithms, thus requiring powerful devices for implementation. Balancing complexity and gains is yet another area for future research.

10 References

[1] J. Karlsson, H. Li, J. Eriksson, "Real-time video over wireless ad-hoc networks," *Proc. of Intnl. Conf. on Computer Comm. and Networks,* pp. 596- , 2005.

[2] W. Wei, A. Zakhor, "Multiple Tree Video Multicast Over Wireless Ad-hoc Networks", *IEEE Transactions on circuits and systems for video technology,* vol. 17, no. 1, January 2007.

[3] Q. Qu, R. Appuswamy and Y. Chan, "QoS guarantee and provisioning for real-time digital over mobile ad-hoc CDMA networks with cross-layer design", *IEEE Wireless Communications*, Oct. 2006, pp. 82-88.

[4] Z. Liu, G. He, Z. Liu, "An adaptive control scheme for real-time MPEG-4 video over ad-hoc networks", *Proc. of Intnl. Conf. on Wireless Comm., Networking and Mobile Computing,* vol. 2, pp. 23-26, Sept. 2005.

[5] V.C. Frias, G.D. Delgado, M.A. Igartua, J.A. Delgado, J.M. Diaz, "QoS Provision for video-streaming applications over ad-hoc Networks", *Eurocon 2005*, November 2005.

[6] Q.Qu, R. Appuswamy and Y. S. Chan, "QoS guarantee and provisioning for real-time digital Video over Mobile Ad-hoc CDMA Networks with Cross-layer Design," *IEEE Wireless Communications*, pp. 82-88, Oct. 2006.

[7] H. Gharavi, K. Ban, "Dynamic adjustment packet control for video communications over ad-hoc netwoks", *IEEE Intnl. Conf. on Communications*, vol. 5, pp. 3086-3090, June 2004.

[8] A. Abd El Al, T. Saadawi, M. Lee. "Improving Interactive Video in Ad-hoc Networks Using path Diversity", *IEEE Intnl. Conf. on Mobile Ad-hoc and Sensor Systems,* pp. 369-378, Oct. 2004.

[9] A. Jeyaraj, L. Cheng, M. E. Zarki, "Cross layer Design for Real Time Wireless Video," *2005 Intnl. Conf. on Wireless Networks, Comm. and Mobile Computing*, pp. 1503-1508, 2005.

[10] B Girod and N. Farber, "Chapter 12: Wireless Video, Compressed video Over Networks", November 1999.

[11] C.M. Calafate and M.P Malimbres, "Testing the H.264 error-resilience on wireless ad-hoc networks," *EC-VIP-MC 2003, 4th EURASIP Conference*, pp 789-796.
[12] L. Liu, X. Ye, S. Zhang and Y. Zhang, "H.264/AVC error resilience tools suitable for 3G mobile video services," *Journal of Zhejiang University SCIENCE*, pp. 41-46, 2005.
[13] H. O. Inge, J. Ola and P. Andrew, "RTP-based broadcast streaming of high definition H.264/AVC video: an error robustness evaluation," *Journal of Zhejiang University SCIENCE A*, 2006 7(Suppl. I):19-26.
[14] P. Zhu, W. Zeng and C. Li, "Cross-layer design of source rate control and QoS-aware congestion control for wireless video streaming," *IEEE Intnl. Conf. on Multimedia and Expo,* pp. 1133-1136, July 2006.
[15] L. Bin, H. Feng, S. Lifeng, Y. Shiqiang, "Unbalanced Multiple Descsription Coding Scheme for Video Transmission over Wireless Ad-hoc Networks", *ICME 2006*, Jul. 2006.
[16] S. Mao, T. Hou, X. Cheng, H.D. Sherali, F. Midkiff, Y-Q. Zhang, "On Routing for Multiple Description Video over Wireless Ad-hoc Networks", *IEEE Transactions on Multimedia*¸ October 2006.
[17] ISO/IEC 14496-2: "Information technology – coding of audiovisual objects-part2: visual," Geneva, 2000.
[18] H. Gharavi, K. Ban, "Rate adaptive video transmission over ad-hoc networks", *Electronics Letters,* June 2004.
[19] ITU-T Recommendation H.263, "Video Coding for Low bit rate Communication," version 1, Nov. 1995; version 2, Jan 1998; version 3, Nov. 2000.
[20] L. Chen and W. Heinzelman, "Network architecture to support QoS in mobile ad-hoc networks," *IEEE Intnl. Conf. on Multimedia and Expo,* vol. 3, issue 27-30, pp. 1715-1718, Jun 2004.
[21] ISO/IEC 13818-2: "Generic coding of moving pictures and associated audio information-Part2: Video," 1994, also ITU-T Recommendation H.262.
[22] S. Wegner, "H.264/AVC Over IP," *IEEE Trans. Circuits Sys. Video Technol.*, vol. 13, no. 7, pp. 645-673, July 2003.
[23] E. Setton, T. Yoo, X. Zhu, A. Goldsmith and B. Girod, "Cross-layer design of ad-hoc networks for real-time video streaming," *IEEE Wireless Communications Magazine,* August 2005, vol. 12, no. 4, pp. 59-65.
[24] S. Mao, S. Lin, S. Panwar, Y. Wang, E. Celebi, "Video Transport Over Ad-hoc Networks: Multistream Coding with Multipath Transport", *IEEE Journal on selected areas in communications,* vol. 21, no. 10, December 2003.
[25] B. Radunovic, "A Cross-Layer Design of Wireless Ad-Hoc Networks," Jul. 2005.
[26] T. Stockhammer, M. M. Hannuksela and T. Wiegand, "H.264/AVC in Wireless Environments," *IEEE Trans. Circuits Sys. Video Technol.*, vol. 13, no. 7, pp. 657-673, July 2003.
[27] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer and T. Wedi, "Video coding with H.264/AVC: tools, performance, and complexity," *IEEE Circuits and Systems Magazine*, First Quarter, pp. 7-28, 2004.

[28] H. B. Yu, C. Wang and S. Yu, “A novel error recovery scheme for H.264 Video and its application in conversational services,” *IEEE Trans on Consumer Electronics,*” vol.50, no. 1, pp. 329-334, Feb. 2004.
[29] L. Liu, S. Zhang, X. Ye, Y. Zhang, “Error resilience schemes of H.264/AVC for 3G conversational video services,” *Proceedings of the 2005 The Fifth Intnl. Conf. on Computer and Info. Technol.*, 2005.
[30] E. Masala, C.F. Chiasserini, M. Meo, J.C. De Martin, “Real-time transmission of H.264 video over 802.11-based wireless ad-hoc networks,”
[31] ITU-T Recommendation H.264, “Advanced video coding for generic audiovisual services,” Mar 2005.
[32] Z. Liu, G. He and Z. Liu, “An adaptive control scheme for real-time MPEG4 video over ad-hoc networks,” *IEEE*, pp. 1153-1155, 2005.
[33] A. Begen, Y. Altunbasak and O. Ergun. Multi-path Selection for Multiple Description Encoded Video Streaming
[34] V.K. Goyal. Multiple description coding: Compression meets the network.
[35] Q. Qu, Y. Pei, J. Modestino, X. Tian and B. Wang, “Cross-layer QoS control for video communications over wireless ad-hoc networks,” EURASIP journal on Wireless Communications and Networking, 2005:5, pp. 743-756.
[36] P. Gupta and P.R. Kumar, “The capacity of wireless networks,” *IEEE Trans. Inform. Theory,* vol. 46, no. 2, pp. 388-404, 2000.
[37] S. Toumpis and A. Goldsmith, “Capacity regions for wireless ad-hoc techniques,” *IEEE Trans. On Wireless Comm.,* vol. 2, no. 4, pp. 736-748, July 2003.
[38] E. Setton, X. Zhu and B. Girod, “Minimizing distortion for multipath video straming over ad-hoc networks,” *Intnl. Conf. on Image Processing, Singapore,* pp. 1751-1754, Oct. 2004.