Cross-layer techniques for real-time video streaming over wireless networks

I. INTRODUCTION

The transmission of multimedia data over wireless networks is one of the most challenging tasks posed to the communication systems in the last years. The massive amount of data that characterizes multimedia signals, together with the strict Quality of Service (QoS) requirements on bandwidth, delay, and delay jitters, makes the traditional transmission protocols inadequate. The current Internet infrastructures can only support best-effort services, which impose varying network conditions and can not grant any maximum delivery delay. These inconveniences are further exacerbated by the introduction of wireless channels and the interconnection of heterogeneous networks, which increase the variability of transmitting conditions. Previous works have shown that the QoS level can be significantly increased adapting, at each layer, the parameter values to the transmitted information and the network conditions. However, these results can be significantly improved allowing the separate protocol layers to interact and tuning their parameter setting via a joint optimization procedure. This report provides a short insight about some of the cross-layer schemes for wireless video communications that are proposed in literature.

II. FROM INDEPENDENT LAYERS TO INTERACTING LAYERS

The layered OSI (Open System Interconnection) architecture [1], which provides a loose abstraction for many devices and applications for data transmission, divides all the networking tasks into an hierarchy of seven layers (protocol stacks). At each layer, services are put into action primarily according to protocol headers, which each layer prepends to data and sends to the lowest levels, and secondarily according to procedure calls. Each layer exchanges information with the immediately adjacent ones: communication with non adjacent layers is strictly forbidden. The OSI architecture relies on the principle of modularity, which is well known and widely accepted in system design [2]. Each layer is regarded as a black box whose behavior is specified only by the interface (information hiding). Modularity provides the abstraction necessary for designers to understand the overall system and assures the independent development of each module. In relation to this approach, Kawadia and Kumar [3] state: “Modularity accelerates development of both design and implementation by enabling parallelization of effort. Designers can focus their effort on a particular subsystem with the assurance that the entire system will interoperate. A good architectural design can thus lead to quick proliferation. When subsystems are standardized and used across many applications the unit per cost is reduced, which in turn increases the usage . . . Architecture also leads to the longevity of the system.” The same authors ascribe the success of Internet more to the architecture than to its algorithmic structure.

Standard OSI architecture distinguishes seven layers:

- **Physical**: it deals with modulation, power level and transmission of raw bits on the physical channel;
- **Data link/MAC**: it deals with link reliability through FEC (Forward Error Correction) and ARQ (Automatic Repeat reQuest), medium access protocol to avoid/reduce collisions, framing of data to ensure reliable transmission with minimum overhead;
- **Network**: it introduces the concept of address and path for packets routing;
- **Transport/Session/Presentation**: the distinction between these three layers is something blurred. They deal with establishing end-to-end connection through the network. Typical information available at this level are
round trip time, retransmission timeout, maximum transmission unit, number of packets lost and actual throughput;

- **Application**: it is the interface to the user for running user tasks over the net (*web-browsing, email, video clip . . .*).

Nonetheless, those implementations based on protocol stack OSI architecture are not efficient for wireless networks [4] because of their structural difference. In wireless networks there is no possibility of *switch*, i.e. the capability to receive data from a node shutting off all others. Nodes simply radiate energy everywhere and transmissions mutually interfere. Therefore, we must take into account the SINR (*Source Interference Noise Ratio*) value, together with the SNR value. While this fact could be regarded as a disadvantage, it offers many potential advantages since nodes have new possibilities to cooperate [5]. For example:

- node A could cancel interference created by node C on node B;
- node A could relay packets from node D to node E:
  - amplifying the received signal;
  - decoding the packet, regenerating it and then retransmitting it.

Moreover wireless channels have a high variability over space and time due to the motion of the devices, and to changes of surrounding physical environment. The states of channels can asynchronously switch from “good” to “bad” within few milliseconds and vice versa.

To face this problem, several cross-layer schemes have been proposed in literature. Resource management, data protection strategies and adaptation to channel conditions are *jointly* optimized across layers. The key result is that, actively exploiting the dependence between protocol layers, a significant performance improvement can be obtained. Excellent reviews of various cross layer proposals are provided in [6] and [7]. Two examples of cross layer feedback are given in section IV.

Moreover, a significant research effort has been spent in investigating efficient communication protocols between different OSI levels. In [8] an attempt to create a general taxonomy and to classify *how* layers interact is performed. Layers interact essentially in four different ways:

- creation of new interfaces without changing layers design;
- merging adjacent layers;
- design coupling between layers without creating new interfaces;
- calibration or joint adjustment of parameters across layers.

All the proposed approaches are intended to find the optimal parameter setting that permits to maximize the Quality of Service (QoS) experienced by the end-users. The following section will provide a formal description of this problem.

### III. Joint Optimization of Parameter Settings

A formal definition of cross layer optimization problem is given in [9]. Let $S_{L_i}$ denote the set of adaptation and protection strategies available at layer $L_i$. The set of *joint* adaptation and protection strategies is thus defined by the Cartesian product $S = S_{L_1} \times S_{L_2} \times \cdots \times S_{L_7}$.

Let $S$ an element of $S$, the solution of the cross layer optimization problem is the solution of the following equation

$$S^{opt}(x) = \arg\max_S Q(S(x))$$

where $Q(.)$ is a perceived/objective multimedia quality measure (Mean Opinion Score, Source Noise Ratio, PSNR . . . ) and $x$ denotes the current channel conditions vector. Equation (1) is subjected to different constraints, e.g.

$$\text{delay}(S(x)) \leq D_{max}$$
$$\text{power}(S(x)) \leq P_{max}$$
where $D_{\text{max}}$ and $P_{\text{max}}$ denote respectively the maximum accepted delay and the maximum transmission power available at each station. Beyond device and service constraints, system constraints are commonly considered as bandwidth and fairness (service distribution equity).

It is interesting to point out the greater complexity of cross layer optimization with respect to the layered one. Denoting with $N_L$ the number of strategies available at layer $L$, the search space has cardinality $N_{\text{crosslayer}} = N_{L_1} N_{L_2} \cdots N_{L_7}$, instead of $N_{\text{layered}} = N_{L_1} + N_{L_2} + \cdots + N_{L_7}$, because, in this second case, optimizations are pursued in a disjoint way. As [10] reports, the problem is further complicated by:

- difficulty to derive tractable analytical expressions for $Q$, delay, power, fairness etc.;
- computational complexity of the proposed solving methods;
- continuous changing of channel conditions $x$, requiring constant update of system parameters.

Beyond a plethora of analytic solution techniques (convex programming, Lagrange duality, stochastic stability and many others [10]), alternatives ways, which are mainly based on heuristic approaches [11], are actually explored. In this report we could mention learning and classification methods [12], bargaining theory [13]. It is worth observing that the usage of heuristic techniques suggests cross fertilization possibilities with the field of Artificial Intelligence.

The Difficulties, that occur in various cross layer proposals, fall between the horns of a dilemma:

- the actual performance of analytic solutions is dubious since, despite their optimality demonstrations, they are based on simplifying hypotheses and they require significant computational complexity. We remind that the rate of convergence to the solution is extremely important, since the dynamic nature of wireless channels requires fast converging solutions;
- ad hoc optimizations usually lack an analytic model that represents the global system behavior and avoids unintended negative consequences. The necessity of an holistic approach is evidenced in [8]: an example of the law of unintended consequence is given in section IV.

Other problems are related with software engineering issues in designing cross layer schemes. The interactions between different layers can be implemented in three different ways [8]:

- direct communication between layers;
- a shared database across the layers;
- a radically new architectural design.

In the first case (direct communication), the internal variables of one layer are made visible to the others against the principle of information hiding. We remind that undisciplined interactions could lead to well-known bad programming practices (spaghetti or goto programming) affecting negatively the stability and longevity of an architecture.

The need for a holistic approach, which avoids a poor performance because of a bad interaction between conflicting partial optimizations, is the driving force behind the development of the second and third approach [6], [8]. The implementation of a shared database defines a new meta-layer providing the services of storing and retrieving information to the other layers and, at the same time, an interface and a decoupling module to the optimization programs. This approach is particularly well-suited to vertical calibration.

The third approach in the list [14] paves the way for enhanced interactions between different levels but requires a new system level implementation. This makes the third solutions less probable than the previous ones since the definition and the spreading of new architectural designs have to deal with the resistance of older software systems.

IV. SOME EXAMPLES OF CROSS-LAYER FEEDBACK

In the following, we present two examples of cross layer feedback whose performance has been proved by experimental results. A third example reports a possible unintended consequence due to conflicting transmission configurations between different layers and so it points out the necessity of an holistic approach.

A. TCP on wireless networks

The current deployment of the TCP protocol assumes all losses as being congestion related. Whenever a loss occur over a wireless channel the TCP source reacts to this as though it was do to congestion and thus decreases its
sending rate and consequently the network throughput. On the contrary, differentiating between congestion related losses and wireless channel related losses (this can be done by notification from physical layer to TCP) significant improvement of network throughput is obtained [15].

B. Multiuser Diversity Gain

The so-called Multiuser Diversity Gain is another example of cross layer feedback where the significant gains, due to channel dependent scheduling algorithms, are demonstrable with very easy analytical arguments [16], [7]. For the sake of concreteness, we now present an example that illustrates the possibility of unintended interaction between non-coordinate optimizations.

C. Rate-Adaptive Mac and Minimum-Hop Routing

The Rate-Adaptive Mac protocol, a modification of the IEEE 802.11 MAC protocol introduced in the 802.11e specification, allows the transmission of data at higher rates by changing the modulation scheme according to whether the channel quality is good or bad. Each node chooses the transmission rate among a set of available rate values before transmitting every packet. At first, the sender transmits the Request to send (RTS) packet at the lowest data rate (the base rate). The receiver measures the signal strength and figures out the maximum rate at which data can be received. This rate is then communicated to the sender in the clear to send (CTS) packet. The following data and acknowledgement (ACK) packet are transmitted at this data rate.

Minimum-Hop Routing chooses the longest possible hop in the path. An example is the Destination Sequenced Distance Vector (DSDV) protocol [17]. DSDV builds routing tables by sending hello packets to the neighbors. Hello packets are broadcast packets (they are sent at the base rate and thus have a large range) that contain cumulative routing information.

Combining Rate-Adaptive Mac with DSDV leads to a worse performance than the original system. In fact choosing the longer hop allows low data rates only since signals have lower strengths. The end-to-end throughput improves turning off the Rate-Adaptive Mac and transmitting at the highest data rate. In this way, longer hops are not active (they do not send any data when the channel is not good enough to transmit at the highest rate), and only short hops are transmitting at high data rates [3].

V. CROSS-LAYER SOLUTIONS FOR WIRELESS VIDEO COMMUNICATIONS

The previous section has presented some cross-layer architectures that improve the performance of the transmission adapting the parameter setting of each layer through an inter-layer message exchange. In the following paragraphs we will focus on video transmission, and will present some of the cross-layer schemes that have been designed so far to allow a reliable transmission over wireless networks. Unlike wired packet switched networks that suffer channel losses and delay, the wireless networks have to deal with a time-varying, error-prone physical channel that can be also severely bandwidth constrained. The channel conditions can change rapidly over time due to noise, interference, multipath and the movement of the mobile host. In such a context, transmission schemes have to dynamically adapt both to the application requirements and to the channel conditions. The research efforts in the area of robust video transmission over wireless channels have mainly focused on studying adaptive error-control strategies at the application layer. Error-control techniques such as forward-error correction (FEC) and automatic repeat request (ARQ) are, in fact, necessary to maintain high quality media delivery. FEC is a channel coding technique protecting the source data at the expense of adding redundant data during transmission. FEC has been commonly suggested for applications with strict delay requirements [18]. In the case of media transmission, where delay requirements are not that strict or the round-trip delay is small (e.g., video/audio delivery over a single wireless channel), ARQ is applicable and usually plays a role as a complement to FEC. Unequal error control can also be adopted if it is important to take into account the different importance of different parts of media [19].

However, each network layer has its own optimized adaptation and protection mechanisms, and hence a joint cross-layer optimization is desirable in order to provide an high overall performance for a video transmission.
A. The joint source-channel coding (JSCC) approach

The joint source-channel coding (JSCC) is a possible approach to effectively reduce the errors occurred during transmission by allocating the resources between source codes and channel codes (bit allocation problem) [20]. It is important to derive an analytic model describing the relation between media quality and source/channel parameters. The most common metric to evaluate media quality is the expected end-to-end distortion $D_T$, where $D_T$ consists of source distortion $D_S$ and channel distortion $D_C$. Source distortion is caused during the media source encoding (i.e., quantization, motion estimation, rate control). Channel distortion occurs when parts of media stream are lost due to network congestion, or incorrectly received due to wireless channel noise. Therefore, the bit allocation problem can be formulated as the optimization problem: \[
\min_{D_S, D_C} D_T(s.t. R_S + R_C \leq R_T)
\]
where $R_T$ is the total available bandwidth, and $R_S$ and $R_C$ are the rates for source coding and channel coding, respectively. JSCC schemes are thus proposed to achieve the optimal end-to-end quality by adjusting the source and channel coding parameters, simultaneously (see also [21]). From the source-coding point of view, there are a lot of possibilities that can be jointly considered with channel coding (see [22], [23], [24]).

B. Hybrid FEC/ARQ cross-layer techniques

Most of the adopted solutions are based on FEC techniques, i.e. they do not need to interact with the receiver to deal with errors and losses. This fact is highly desirable for low-latency communications since using a feedback channel to send transmission reports implies extra delays in the communication. However, the adoption of feedback information has proved to be quite effective in the quality of the transmitted information, as it was shown by Girod et al. in [25], [26]. Many papers in literature tries to find an efficient trade-off between these two tendencies by adopting hybrid ARQ/FEC schemes that adaptively protect the transmitted information using channel codes and retransmissions at network layers.

In [27] Zhai et al. design an effective algorithm that optimize this combined strategy choosing whether a packet must be retransmitted or not according to some feedback information and delay constraints.

In [28], an efficient and robust transmission of video over 802.11 Wireless Local Area Networks (WLANs) is considered; the authors propose a novel error-protection method that can provide adaptive QoS to layered coded video by utilizing priority queuing at the network layer and retry-limit adaptation at the link layer. The proposed cross-layer protection system can provide not only priority delivery services, but also UEP to the different video streams, by adapting different retry-limit settings in the media access control (MAC) for the multiple queues containing the different video streams priorities.

Krishnamachari et al. in [29] evaluate different error-control and adaptation mechanisms available in the different layers for robust transmission of video, namely, MAC retransmission strategy, application layer FEC, bandwidth-adaptive compression using scalable coding and adaptive packetization strategies. The authors propose a “cross-layer protection” system by performing tradeoffs between throughput, reliability, and delay depending on the channel conditions and application requirements. They model the end-to-end distortion of MPEG-4 FGS video for various channel conditions with different UEP strategies, and they show that the model matches the results that can be obtained by simulations. Based on this model, a strategy for the adaptive selection of application layer FEC, maximum MAC retransmission limit, and packet sizes, depending on the channel condition, to maximize the video quality is developed.

C. Cross-layer approach: FEC and ARQ at RLP packet

In [30], the authors propose a set of building blocks for channel and application adaptive wireless streaming applications. They exploit two well-known principles in wireless video communication: the first has to do with the fact that different parts of a video bit stream are of different importance, and hence need to be protected via FEC and ARQ to different degrees; the second has to do with adaptivity to channel conditions by dropping unimportant packets. They propose a novel packetization scheme so that forward error correction (FEC) codes can be applied within an application packet at radio link protocol (RLP) packet level, but not a bit level, rather than across different application packets and thus reduce delay at the receiver. Traditional cross-packet FEC, instead, encodes $k$ application layer packets and generates $n-k$ parity packets.

Furthermore, a priority-based automatic repeat request (ARQ), together with a scheduling algorithm, is applied at
the application layer to retransmit only the corrupted RLP packets to improve the wireless bandwidth efficiency. In doing so, the application layer packet is constructed in such a way that an exact integer number of RLP packets is generated after decomposition of any application layer into RLP packets. Based on the analysis of the delay, bandwidth utility and video quality variations, the scheme proposed in [30] outperforms the traditional scheme if the same conditions are given or even a tighter condition is given to the analyzed scheme. The reason is that the adaptation scheme is done at the application layer, but at the granularity of the RLP layer. The approach can therefore combine the flexibility and programmability of the application layer adaptations, with low delay and bandwidth efficiency of link layer techniques.

D. Source-Adaptation-Based Wireless Video Transport

In wireless networks, channel state information is hard to obtain in a reliable manner due to the rapid change of wireless environments. It is known, however, that the source motion information is always available and can be obtained easily and accurately from video sequences. In [31], it is proposed a novel cross-layer framework that exploits only the motion information inherent in video sequences and efficiently combines a packetization scheme, a cross-layer forward error correction (FEC)-based unequal error protection (UEP) scheme, an intracoding rate selection scheme and a novel intraframe interleaving scheme. The results shown in [31] demonstrate that the proposed approach is very effective in dealing with the bursty packet losses occurring on wireless networks without incurring any additional implementation complexity or delay.

E. Cross-Layer based on Coopetition

The authors in [9], propose a new paradigm for wireless communications based on coopetition, which allows wireless stations to optimally and dynamically adapt their cross-layer transmission strategies to improve multimedia quality and utilization of wireless resources. When applied to wireless multimedia systems, coopetition fundamentally changes the passive way stations currently adapt their transmission strategies to match available wireless and power resources, by enabling them to actively influence the wireless systems dynamics through resource and information exchange. To allow coopetition, the authors propose a new way to architect the wireless multimedia communication system by jointly optimizing the protocol stack at each station and the resource exchanges among stations. In this paradigm, information about resources and constraints (e.g., QoS requirements, multimedia traffic characteristics, experienced channel conditions) of the various stations can be disseminated to all network members (stations), and used as available optimization criteria for their own communication subsystem.

F. Cross-Layer optimization of network and application layers

During multimedia packet transmissions, network congestions and the limited amount of data that can be stored in queues lead to the need of discarding some of the transmitted packets. Network congestions may happen whenever the overall traffic load overcomes the available network resources, and some of the packets to be transmitted are either buffered or are discarded by traffic policer that monitor the network status [32], [33]. On the other hand, as long as a node can not access a shared transmission medium, outcoming packets are buffered in a queue that could get soaked with packets quite easily in case the size of the cell is underdimensioned with respect to the number of users. Moreover, an excessive waiting time in the queues could makes a video packet obsolete since it can not be decoded in time, and therefore, the video source decoder considers its late arrival like a loss.

In order to provide a control over losses or excessive delays, network elements and edge nodes need to adopt an appropriate scheduling strategy that shapes the incoming and outcoming traffic according to a specific policy (IntServ, DiffServ [34], [35]). The main purposes of these schemes are to prevent users/services that violates their traffic limitations from jeopardizing the QoS of other connections and to protect well-behaving users/services against such traffic violations. At the edge of the network each packet is classified, and a service class is assigned (trTCM [36], MPLS). This class determines how the network elements handle packets along the data path according to different scheduling and queue management schemes (DropTail, RED, WRED, RIO). As for video applications, these strategies turn out to be crucial in the characterization of the video quality perceived by end users since the significant amount of information that needs to be transmitted must be accurately controlled [37], [38].
Generally classifiers label packets according to their size and the buffer levels, and their video content is not taken into consideration. Recently, new techniques have been proposed showing that a classifying strategy aware of the significance of each packet in the decoding process improves the QoS experienced by the end user.

One of these solutions is proposed by Kim et al. in [39] characterizing the distortion produced by packet losses in a layered video coder. According to the distortion model derived in the paper, it is possible to design an optimization technique that assign to each packet a priority level according to its impact on the overall quality and the quality layer which it belongs to. The authors propose an efficient solution for minimizing the average distortion keeping the total cost under a certain value (in this case the cost consists in the price that the user has to pay in order to transmit a packet with a certain service level).

A similar approach was proposed by Luna et al. in [40], considering the possibility of either minimizing the overall distortion or minimizing the overall cost required to keep the average distortion under a certain threshold. At the same time, the total bit rate is constrained too in order to fit the coded stream into the available bandwidth. Like the technique mentioned before, the optimal solution is found through a Lagrangian relaxation problem which is solved here through a Dynamic Programming procedure. The distortion model proposed by the author allows a simple pruning algorithm that significantly reduces the computational load.

This approach was lately extended by Zhai et al. in [41], where a joint optimization algorithm that controls both the source coder and the packet classifier is presented. This optimization scheme determines both the characteristics of the video signals coded by the packet and its service class constraining the maximum end-to-end delay. The packet is then transmitted over a DiffServ-supported network that either blocks or forwards packets according to their priority class. The reported results show that the quality of the reconstructed sequence can be improved up to 1.3 dB.

A different approach proposed in [42] performs a sort of cross-layer Unequal Error Protection (UEP) assigning each packet to different service classes according to the syntax elements they carry. The Data Partitioning (DP) mode specified within the H.264/AVC standard partitions the bit stream related to each coded slice into three different types of packets (A,B, and C). As a consequence, the relevance of each packet in the decoding process changes according to the contained information, and the adoption of an algorithm that adapts the protection level to the packet type can significantly improve the quality of the reconstructed sequence. The proposed algorithm takes advantage of the Enhanced Distributed Channel Access (EDCA) mechanism defined within the standard IEEE 802.11e. This protocol standardizes different Access Categories (ACs), which are associated with a different queue and a different set of channel access parameters. Assigning each packet type to a different AC, it is possible to modulate the throughput for each packet type varying the loss probability. The same approach has been adopted by Bernardini et al. in [43] to transmit a Multiple Description video stream specifying different throughputs to each description. Experimental results prove that at high BER values adopting the MDC approach in place of the traditional single description H.264/AVC with DP enables proves to be more effective.

Other approaches take advantage of network layer information in order to drive properly the coding choices at higher levels. In this case the cross-layer strategy is only intended to provide upper levels with a report of the network condition as accurate as possible. In [44] the authors propose a cross-layer optimization algorithm that takes advantage of the routing information in ad-hoc networks. In case the number of hops provided by the AODV algorithm that interlies between the source node and the destination node increases, the amount of redundant information is increased in order to tune appropriately the channel coder.

G. Cross-Layer optimization of link and application layers

Previous sections have shown how it is possible to adapt the protection level of transmitted information at application and network layers according to the channel conditions and the characteristics of the coded video signal. However, at link layer the reliability of the transmission can also be affected by the transmission power since the SNR level could be increased leading to a higher throughput. On the other hand, this may lead to interference for other users or an inefficient use of the available battery energy. All these issues make transmission power control a fundamental element in the design of wireless networks.

Transmitter power can be adapted in order to extend battery life in mobile devices, grant a certain QoS level, and reduce the interference to other users. Despite this technique could be critical in case of non-orthogonal
communications (like in case of CDMA transmissions) because of interference problems, it can be efficiently applied to other modulation techniques that involve temporal and frequency separation.

A first cross-layer optimization strategy is presented in [45], [46], where source coding parameters and data rate are changed in order to control the transmission delay and the overall coding/channel distortion while managing efficiently the energy supply at the mobile device. The resulting algorithm employs a Dynamic Programming strategy that tries to minimize the expected distortion constraining the transmission delay and the energy involved in the transmission of the sequence. Because of the number of different parameters that have to be optimized, the complexity of the whole algorithm is high and it must be reduced through some simplification in the distortion model.

In a more recent approach [47], Zhai et al. introduce some additional protection at link layer. The overall scheme involves two different FEC codes: a first RFC2763-based RS code at application layer and a second RCPC code at link layer. The authors propose an efficient algorithm that tunes the channel code rate of both approaches according to the video content and the estimated channel state.

In a more recent paper [21], the same authors integrate the power management based protection strategy with the solution discussed in the previous works increasing the number of protocol layers involved in the optimization process. Once again a DP approach is designed and there are still complexity issues to be solved.

H. Cross-layer optimization between application, data link and physical layers

In [48], [49] and [50] a cross-layer optimization between application layer, data link layer, and physical layer to optimize the end-to-end quality of the wireless streaming video application and to use efficiently the wireless resources, is proposed. The authors include the application layer in the joint optimization because the end-to-end quality observed by the users directly depends on the application and the application layer has information about the impact of each successfully decoded piece of multimedia data on the perceived quality. The architecture analyzed consists of the process of parameter abstraction, a cross-layer optimizer, and the process of decision distribution. They consider a video streaming server located at the base station and multiple streaming clients located in mobile devices. The users are sharing the same air interface and network resources but they request different video contents.

At the base station, an architecture as shown in Figure 1 is proposed to provide end-to-end quality-of-service optimization. Three main blocks can be seen:

![Cross-Layer architecture proposed in [48].](image)

1) **Parameter abstraction**: state information parameters are collected from the application layer and the radio link layer and they are transformed into parameters that are comprehensible for the cross-layer optimizer, so-called cross-layer parameters. In a single-user scenario, for example, the abstract key parameters from the data link layer and physical layer could be: transmission data rate $d$, transmission packet error ratio $e$, data packet size $s$, and the channel decorrelation time $t$. At application level, instead, the abstracted parameters could include the source data rate, the number of frames per second, the size (in bytes), and the maximum delay of each frame. Other important information for the optimizer is the distortion-rate function and the
so-called loss distortion profile, which shows the distortion $D_i$ that is introduced in case the $i$th frame of the GOP is lost.

2) **Cross-layer optimization:** the cross-layer optimizer performs the optimization with respect to a particular objective function. From a given set of possible cross-layer parameter tuples, the tuple optimizing the objective function is selected. The choice of a particular objective function $\Gamma$ depends on the goal of the system design, and the output (or decision) of the optimizer might be different for different objective functions. In the example of streaming video, one possible objective function in a single user scenario is the $MSE$ between the displayed and the original video sequence, that is, the sum of loss distortion $MSE_L$ and source distortion $MSE_S$:

$$MSE = MSE_L + MSE_S$$

where $MSE_L$ can be computed from the distortion profile by

$$MSE_L = \sum_{i=1}^{N} D_i P_i$$

where $N$ is the number of frames in one GOP, $P_i$ is the probability that the $i$th frame is the first frame lost during transmission of this GOP and $D_i$ is the mean square error that is introduced by this loss. The parameter $D_i$ is taken from the measured distortion profile and is usually different for each GOP. The parameter $P_i$ can be derived using the data link layer and physical layer parameters introduced before. For a multiuser situation, different extensions of the $MSE$ are possible. For example, the objective function can be the sum of $MSE$ of all the users. That is,

$$\Gamma = \sum_{k=1}^{K} MSE_k$$

where $MSE_k$ is the $MSE$ of the user $k$. Other definitions of $\Gamma$ can be found in [48].

3) **Decision:** after the decision on a particular cross-layer parameter tuple is made, the optimizer distributes the decision information back to the corresponding layers.

In [48] and [49], the authors provide simulation results to evaluate the performance of the joint optimization, with three users, each of which requests a different video sequence. Their study reveals that the proposed architecture can provide a potential way to improve the performance and, even when a small number of degrees of freedom of the application layer and the radio link layer is considered, they obtain significant improvements in user-perceived quality of the streaming video by joint optimization.

In the end, in [51], the cost due to applying the cross-layer design (Fig. 1) and the trade-off between cost and performance are analyzed.

The approach proposed in [52] presents the same holistic approach applied to a layered video coder. Different QoS classes are defined and adaptively assigned to each frame and each quality level. In this process the transmission power is optimized for each frame varying the protection level in order to optimize the overall resulting quality. The paper also specifies the message interaction that takes place during the transmission of the information in order to find the optimal parameter setting. Despite the previous techniques resort to DP algorithms to find the optimal parameter setting, in this case the optimization problem is solved through a closed form equation.

While all these techniques do not imply a sender/receiver interaction, other works introduce in the optimization process the possibility of retransmitting part of the information according to the arrived packet acknowledgments. In [53], Zhang et al. propose a cross-layer optimization for scalable video over 3G wireless networks that combines FEC protection, delay-constrained ARQ, and power control to minimize the expected distortion at the decoder.

Some of these approaches try to deal also with the message exchange problem since each layer must be appropriately tuned according to the other layers and the channel conditions. Papers [52], [54] deal with this issue in defining the optimization strategy, while in [50] the authors face the problem of cross-signaling analyzing the costs implied by the cross-layer optimization itself.
VI. CONCLUSION

The report presents a quick overview of cross-layer strategies for wireless video communications that have been proposed in literature. Most of the solutions still present complexity issues since the more layers are jointly optimized the more variables are to be adaptively set. Therefore, a lot of efforts are invested in trying to simplify this problem from a computational point of view. Moreover, most of the quoted works deal with QoS support for unicast media streaming only. QoS provisioning for multicast systems implies a whole set of new solutions. Mobility also poses new challenges to the cross-layer architectures since handoffs have proved to be crucial in maintaining an acceptable media quality. In the end, wireless ad-hoc networks require new coding solutions that tries to cope with their dynamic changing topology and the presence of interfering nodes.

REFERENCES
