Meet In the Middle Cross-Layer Adaptation for Audiovisual Content Delivery

Ismail Djama, *Member, IEEE*, Toufik Ahmed, *Member, IEEE*, Abdelhamid Nafaa, *Member, IEEE*, and Raouf Boutaba, *Senior Member, IEEE*

***Abstract*—This paper describes a new architecture and implementation of an adaptive streaming system (e.g., Television over IP, Video on Demand) based on cross-layer interactions. At the center of the proposed architecture is the Meet In the Middle concept involving both bottom-up and top-down cross layer interactions. Each streaming session is entirely controlled at the RTP layer where we maintain a rich context that centralizes the collection of i) instantaneous network conditions measured at the underlying layers (i.e.: link, network, and transport layers) and ii) user- and terminal-triggered events that impose new real-time QoS adaptation strategies. Thus, each active multimedia session is tied to a broad range of parameters, which enable it to coordinate the QoS adaptation throughout the protocol layers and thus eliminating the overhead and preventing counter-productiveness among separate mechanisms implemented at different layers. The MPEG-21 framework is used to provide a common support for implementing and managing the end-to-end QoS of audio/video streams. Performance evaluations using Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) objective video quality metrics show the benefits of using the proposed Meet In the Middle cross-layer design compared to traditional media delivery approaches.**

***Index Terms*—Cross-layer adaptation, forward error correction, link-layer quality, MPEG-21 multimedia framework, QoS metrics, real-time streaming.**

I. Introduction

THE IEEE 802.11 WLAN standard is being accepted and deployed in many different environments such as companies, universities, government institutions, and public places (airports, train stations, etc.). The achievement of WLANs is draining an unprecedented research interest that is translating into a tremendous commercial success. Its adoption is favored by the promises of the forthcoming 802.11n specifications topping 540 Mbit/s (raw throughput) with the help of MIMO technology. 802.11-based networks appear as a serious alternative for wired ethernet, paving the way for QoS-enabled and added-value service provisioning for ubiquitous users. Thanks to the rise of powerful video compression techniques, such as H.264 and MPEG-4, it is now possible to combine video, audio, and data within the same signal and transmit it over packet-based wireless networks. All these advances will help the emergence of new powerful multimedia applications with limitless possibilities and business promises. QoS-enabled multimedia services deployment is actually the main challenge to overcome in order to render WLANs an integral part of Network Operator's commercial offerings, and generally more attractive for the particular use.

Given the recent progress in both video encoding standards and wireless networks capacity, various in-door and out-door WLANs network operators are now more and more concerned by their ability to provide multimedia services with sustained QoS guarantees, while at the same time supporting a large number of heterogeneous wireless terminals. Providing QoS guarantees is an imperative in developing viable business models, while serving a maximum number of heterogeneous terminals is an obvious economical goal. Although the heterogeneity of terminal capabilities constitute an important burden on streaming systems, the QoS continuity of communication remains the main issue when streaming media over WLANs. Video adaptation techniques are clearly required in such environments with inherently varying link's capacity and quality. To achieve maximum efficiency, these techniques require a higher level of cross-layer interactions. This allows to i) translate application-layer QoS requirements into lower layers performance metrics such as link capacity and ii) reflect at the application-layer the short-term network fluctuations, captured at lower layers, such as interferences. QoS guarantees are particularly difficult to maintain in wireless environments where packet loss depends on the unpredictable occurrence and frequency of interferences. Error control techniques [e.g., FEC (Forward Error Correction) and ARQ (Automatic Retransmission reQuest)] are required in such environments. In particular, FEC is commonly used for real-time applications due to its proven scalability for multicast communications and the strict delay requirements of media streams.

With respect to video adaptation, the new MPEG-21 multimedia framework defines several parts for facilitating Digital Item (DI) consumption. MPEG-21 is intended to provide a general framework for multimedia access, delivery and consumption in order to enable seamless interoperability and content adaptation in distributed multimedia systems. More specifically, MPEG-21 defines two fundamental concepts which are Digital Item (DI) and User. DI represents an abstraction of a multimedia object (pictures, video or audio clip) and User is an abstraction of entities which interact with DIs. The MPEG-21 standard is divided into different parts. Each one handles a different aspect of the framework. Digital Item Adaptation (DIA), which is the MPEG-21 part 7 [1], specifies a set of tools to perform

Manuscript received February 22, 2007; revised July 26, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Anna Hac.

I. Djama and T. Ahmed are with the CNRS-LaBRI Lab, University of Bordeaux, F-33405, Talence, France (e-mail: djama@labri.fr; tad@labri.fr).

A. Nafaa is with the School of Computer Science and Informatics, University College of Dublin, Belfield, Dublin 4, Ireland (e-mail: nafaa@ieee.org).

R. Boutaba is with the School of Computer Science, University of Waterloo, Waterloo, ON, N2L 3G1, Canada (e-mail: rboutaba@bbcr.uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2007.911243

video adaptation, which includes Bitstreams Syntax Description (BSD), Usage Environment Description (UED), Adaptation QoS, and Universal Constraints Description (UCD). UED includes descriptive information related to user characteristics (e.g., user information and preferences, usage history, presentation preferences, audio language preferences, subtitles language preferences), terminal capabilities (e.g., codec capabilities and display capabilities, audio output capability, storage characteristics, battery capacities), network characteristics, and natural environment characteristics (e.g., location and time). The network characteristics provided by UED describes both static and dynamics network aspects. The static aspects are gathered in the network capabilities description (e.g., guaranteed bandwidth, delay transmission, error detection and correction), while the network conditions inform about the network dynamic status (e.g., packet delay, jitter, packet loss rate). An important feature of MPEG-21 resides in its ability to offer a rich, customizable, and interactive interface to users that enable them to change, at runtime, certain UED's attributes. The main challenge in future streaming systems is to enforce in a timely manner the UED requirements throughout the QoS adaptations implemented at different OSI layers. Note that terminal capabilities may also vary over the course of a multimedia session to reflect, for example, a codec reconfiguration/switching after a serious drop in battery level.

The way how network conditions are characterized by applications is an important factor that determines the effectiveness of adaptive mechanisms and remedial actions in face of transient network outages. In designing channel-aware video streaming systems, two interrelated challenging issues should be tackled: the accuracy of the effect of channel fluctuations and the effectiveness of application adaptation. The former consists in getting a thorough insight into channel fluctuations and theirs manifestations at application level by gathering a maximum number of QoS performance metrics at different levels of the protocol stack. Using QoS metrics of different protocol layers would deliver complementary information useful for building a consistent view of current delivery conditions. For instance, physical layer and transport layer measurements describe different manifestations of network fluctuations, and more importantly they give indications regarding different time scales. The latter challenging issue concerns the way the adaptation mechanisms interpret and react to those network fluctuations. For instance, 802.11 networks employ an inherent link-layer retransmission-based (ARQ) error control technique along with adaptive physical-layer coding/modulation to face frequent degradations in link quality. When used alone, 802.11 integrated adaptation mechanisms cannot avoid video degradations. In fact, these mechanisms are rather designed to increase the transmission reliability by either decreasing the transmission rate (coding/modulation) or increasing the average delay per packet transmission (ARQ). In certain circumstances, link quality degradations may be tackled at application-level (error control) with much more efficiency in terms of channel utilization. Particularly, application-level error control techniques (such as FEC) can improve application responsiveness by using physical layer signal strength measurements feedback. Indeed, metrics such as Received Signal Strength Indicator (RSSI) represent an efficient indicator to predict the increase of packet loss rate. RSSI can be used to trigger an adaptive FEC mechanism to face packet losses and prevent user-perceived quality degradations. However, when the link signal strength degrades below a certain level, it is more appropriate to use a stronger modulation and coding channel at the physical layer. It is worth noting that application-level techniques are not anymore efficient when a large number of corrupted link layer frames occur. Clearly, coordination is of utmost importance for adaptation mechanisms to be efficient.

Our ultimate goal is to design an efficient video streaming system that is media, user/terminal, and channel aware. Our approach is based on a cross-layer design involving interactions between the application and the underlying layers of the protocol stack and a better integration of video coding semantics/metadata with the network transport mechanisms. The standardized MPEG-21 digital item management framework offers an ideal context for cross-layer interactions between distinct streaming system layers. Hence it provides a better consideration to environmental and external factors such as end-users' perceived quality, overall UED descriptions, and channel characteristics reported by the underlying protocol layers. Our approach, called Meet-In-the-Middle (MIM) cross layer, attempts to conciliate both the bottom-up and the top-down approaches usually undertaken in cross-layer design. Using top-down awareness, application-level QoS requirements are interpreted and enforced at the lower layers using appropriate QoS mechanisms. Meanwhile, bottom-up awareness allows to delivering valuable information on the wireless channel conditions to readjust upper-layer QoS adaptation mechanisms.

Meet-In-the-Middle centralizes the treatment of information, resulting from multi-layer network measurements and instantaneous user- and terminal-triggered constraints, at the Real-Time Transport Protocol (RTP) layer. The latter protocol offers the ideal context to manage both high-level QoS requirements and lower layers network and link measurements. By maintaining a large context (parameters set) related to a given multimedia session, it is possible to handle several aspects of real-time communication such as media synchronization, adaptive application framing, rate adjustment, link layer adaptation, etc. The objective here is to integrate and control all audiovisual adaptation mechanisms in a coordinated manner to minimize possible redundancies and counter-productiveness between layers, and thus improving the overall system efficiency. In this paper, we introduce the MIM cross-layer approach as well as its underlying principles. We provide concrete deployment scenarios in terms of cross-layer interactions to validate the effectiveness of our proposal. We evaluate the performance of a simple MIM cross-layer architecture where user/terminal requirements are communicated at session initiation phase while application constraints and network conditions measurements are continuously collected from both transport layer and PHY/MAC layer. The validation is achieved through the implementation of a prototype that provides live IPTV and VoD services to WLAN's heterogeneous receivers.

The rest of this paper is organized as follows: Section II discusses related works on video streaming over IP networks with a special focus on cross-layer design approaches. Section III describes our proposed Meet-In-the-Middle (MIM) cross-layer adaptation and deals with the main implementation issues. Section IV presents performances evaluation results. Section VI concludes the paper.

II. Background and Related Works

A. Transport Services for Media Streaming

TCP and UDP are the most deployed transport services for real-time media streaming applications, although other protocols such as SCTP [2] and DCCP [3] have recently gained interest due to the support of new functionalities (e.g., flow control and multi-homing). Over the years, several enhancements of TCP and UDP have been proposed to better accommodate multimedia applications requirements. UDP-lite [4] is a typical example of these enhancements. By relaxing certain UDP and TCP constraints and implementing more efficient application-level error control schemes, audiovisual applications can reduce delays and excessive bandwidth consumption entailed by frequent packet retransmission. The IETF Audio Video Transport (AVT) working group has also specified the RTP (Real-time Transport Protocol) [5]. The origin of RTP stems from the growing need for more flexibility in managing various aspects of multimedia streams transmission. RTP may be used over any transport layer protocol (such as UDP, TCP, SCTP, etc.) though many redundant functionalities may exist with certain transport protocols. A typical example is the sequencing functionality provided by both TCP and RTP. Therefore, RTP is most commonly implemented over UDP. RTP follows the principles of application level framing and integrated layer processing proposed by *Clark and Tennenhouse* [6]. That is, RTP is intended to be malleable to provide the information required by a particular application and will often be integrated into the application level rather than being implemented as a separate layer. On the other hand, the RTP Control Protocol (RTCP), is used to monitor the quality of service and to convey information about the participants in on-going multimedia sessions. RTP/RTCP protocols are currently the "*de-facto*" Internet standards for real-time transport of various multimedia contents. Their specification is accompanied by documents (payload formats) that describe the specific encoding of different media types (RTP profile). RTP profile defines the meaning of each RTP payload and header fields depending on the transported media. RTP is usually implemented on top of UDP which offers no congestion control mechanisms and is thus unaware of network conditions and unfair towards other traffic flows. RTP/UDP may lead to an unfair behavior with respect to existing TCP flows. Since today's Internet traffic is dominated by TCP (Web and P2P traffic), it is crucial, at least from network operators perspective, that UDP traffic behaves in a TCP-friendly manner so as to ensure that TCP-based applications continue to receive acceptable quality of service. Otherwise such applications will keep entering in a congestion avoidance phase, causing performance collapse and serious drop in the network throughput.

The average throughput of TCP can be inferred from end-to-end performance measurements such as round-trip-time (RTT) and packet loss. This observation has led to the definition of TCP-Friendly Rate Control (TFRC- RFC 3448) [7] that can be performed by the end-hosts to maintain their rate within a certain level that matches a rate of an equivalent TCP session having the same conditions (i.e., the same session lifetime). The integration of lower level information such as packet loss and RTT into application-level processing was the first step towards the cross-layer design.

B. Cross-Layer Optimization

Cross-layer design has particularly arisen in wireless communications as an important paradigm to optimize the scarce wireless bandwidth utilization. It investigates situations where different OSI layers may cooperate to improve the ability of applications to achieve certain objectives such as QoS guarantees, power saving, or customization according to user preferences, etc.

Cross-layer design may be achieved by either integrating functionalities of different layers in a single protocol or simply establishing tight cooperation between adjacent (or non adjacent) layers.

In the former case, replication of information and redundancy of functionalities are avoided. This allows reducing the overhead and provides the means, thanks to the availability of a broad range of operational parameters in a single protocol, to implement advanced QoS mechanisms. The latter and most prevalent cross-layer design approach argues for richer inter-layer interactions to achieve better reactivity to network fluctuations and other external factors. The cross-layer parameters exchanged between layers depend on what functionality is being implemented as well as the objectives and constraints specific to the application being considered. A classification of useful parameters exchanged between layers in wireless communications can be found in [8]. In the following, we give three main categories of such parameters:

1) Channel state information (CSI) including location information, terminal capabilities, signal strength, interference level, etc.
2) QoS related parameters including delay, throughput, bit error rate (BER), and packet error rate (PER) measurements. Those parameters may be tracked at different layers.
3) Traffic patterns as perceived by each layer, including data traffic characteristics, knowledge of the data rate (constant or variable), data burstiness, data fragmentation, packet size (maximum transfer unit), and information about queue size.

Typically, cross-layer design approaches in wireless communications try to use the inherent variability of the wireless channel reported by the physical layer to adjust upper layers behaviors [9]. In fact, system components such as medium access control (MAC) protocols, routing algorithms, transport protocols, and application layer mechanisms may benefit from a certain degree of awareness about time-varying channel characteristics. The other way around, upper layer QoS constraints and application requirements may be translated into protocol behaviors enforced by appropriate mechanisms at the lower layers. For instance, by using transport layer statistics at RTP level, it is possible to adjust the application rate, error control techniques, buffering strategies, and so on.

C. Cross-Layer Video Streaming Architectures

Most of the works on cross-layer optimization have focused on MAC and PHY layers interactions in wireless environments. Very few works have considered higher level interactions such as the translation of user/terminal/application-level QoS requirements into well-proportioned QoS mechanisms.

In [10], the authors formalize the cross-layer design problem, discuss its challenges, and present a classification of possible solutions. The paper illustrates the relationships between layers and the challenge to find the best configuration that optimizes different metrics at different layers. The paper introduces the concept of "*coopetition*" between wireless stations, where a judicious mixture of competition and cooperation is often advantageous in competitive environments. When applied to wireless multimedia systems, "*coopetition*" changes the passive behavior of stations to adapt their transmission strategies to match available wireless and power resources, by enabling them to proactively influence the wireless systems dynamics through resource and information exchange.

In [11], the paper describes the recent advances in network modeling, QoS mapping, and QoS adaptation. The authors present a general block diagram of end-to-end QoS for video delivery over wireless networks. However, the end-to-end QoS is considered as a network-centric practice rather than application-level centric. Similarly works in [12]–[14], and [15] adopt a cross-layer design approach where both physical and MAC layer knowledge are shared with higher layers.

In [16], different error control and adaptation mechanisms available in several layers are evaluated for robust video transmission. Based on this evaluation, a new adaptive cross-layer protection strategy is proposed to enhance the robustness and efficiency of scalable video transmission. In [17], the authors propose a set of cross-layer techniques for adaptive video streaming over wireless networks. Data packets at application layer are decomposed on equal size radio link protocol (RLP) and FEC codes are applied based on RLP packets. The paper proposes also a priority-based Automatic Retransmission reQuest (ARQ) for corrupted RLP at application layer. The work in [18] present a new cross-layer content delivery architecture that is capable of receiving information from the network and adaptively tune transport parameters (e.g., bit rates) and other QoS mechanisms according to the underlying network conditions. The paper describes a service-aware IP transport architecture composed of a dynamic content-level audiovisual object classification model; a reliable application-level framing protocol with fine-grained TCP-friendly rate control and adaptive unequal error protection; and a service-level QoS mapping/packet-tagging algorithm for seamless IP differentiated service delivery. The obtained performance results demonstrate that by breaking the isolation of the OSI layers and by injecting content-level semantic and service-level requirements within the transport protocols, one can provide a more efficient support for multimedia services streaming.

More recently, a significant number of R&D activities are dedicated to QoS-sensitive multimedia delivery using cross-layer interactions. A number of such activities are conducted in the scope of European IST FP projects such as BRAIN, MIND, DRiVE, EVERES, PHOENIX, 4MORE, and ENTHRONE.

Both 4MORE (4G MC-CDMA Multiple-Antenna System on Chip for Radio Enhancements, IST-507039) [19] and PHOENIX (Jointly Optimizing Multimedia Transmission in IP- Based Wireless Networks, IST-001812) [20] projects address cross-layer integration issues for multimedia streaming architectures. PHOENIX aims to develop a scheme to let the application world (source coding, ciphering) and the transmission world (channel coding, link modulation and coding) to further interact using the IPv6 protocol (network world), in order to improve the performance of multimedia communications. NEWCOM (Network of Excellence on Wireless Communications, IST-507325) [21] aims at identifying gaps in European knowledge on cross-layer practices to ultimately prepare an action plan for filling those gaps. The ENTHRONE (IST-507637) [22] project proposes an integrated management solution which covers an entire audio-visual service distribution chain, including content generation and protection, distribution across networks, and reception at user terminals by integrating a cross-layer mapping of the application level QoS requirements to network level QoS concepts.

Similarly, our work falls in the area of audiovisual content delivery optimization using a cross-layer approach. Our approach however involves both top-down and bottom-up cross-layer interactions. Also, our work is not restricted to multimedia streaming over wireless networks. Different environments characteristics (terminals, users, multimedia applications, and networks) may be accommodated thanks to MPEG-21 tools. Finally, although the focus in this paper is on FEC and content adaptation techniques that are used as a proof of concept, other adaptation mechanisms may be as well envisioned.

III. Meet in the Middle Cross-Layer Adaptation

The ultimate goal of Meet In the Middle (MIM) cross-layer adaptation is to ensure a seamless translation of user/application level QoS requirements into network level QoS metrics. At the same time, it allows reflecting at the application-level important network measurements carried out by the underlying layers.

User and terminal capabilities along with application and content specifications are essential knowledge when translating upper-layer QoS requirements into effective streaming adaptation mechanisms. In this top-down translation, a continuous flow of information including user, terminal, and application specifications are collected and used as metadata encoded using the MPEG-21 standard. In the same time, RTP-level QoS monitoring data (e.g., loss rate, delay, and jitter) and QoS statistics collected from real-time link layer measurements (e.g., Signal strength), network layer (e.g., Diffserv and queue information), and transport layer (packet loss distribution)[1], follow a bottom-up approach in the protocol stack. Fig. 1 gives a horizontal view of the proposed cross-layer approach that integrates content, terminal, user, and network management into a seamless MIM cross-layer architecture. The network-level measurements are collected on a per multimedia-session basis and used as part of a single context tied to an individual RTP

[1]In this paper, we use the term "network-level" measurements to refer to different QoS measurements performed at the RTP layer and below (transport, network, and link layers). In contrast, higher-level (user, terminal, and coding) configurations are referred to as "service level" configurations.

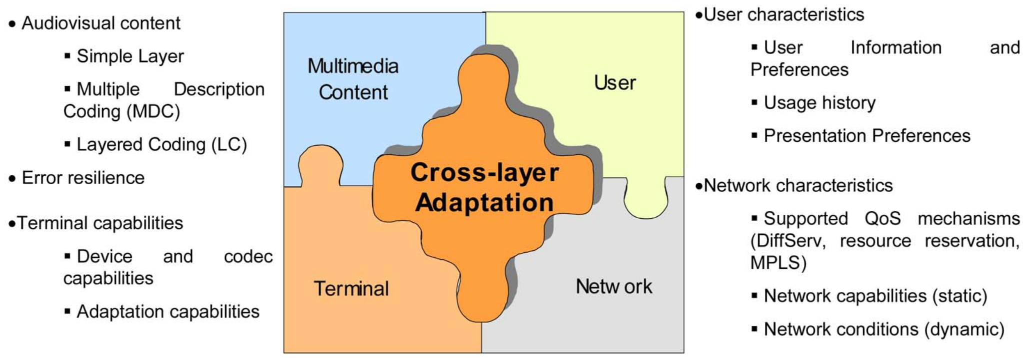

Fig. 1. Cross-layer adaptation based MPEG-21 metadata.

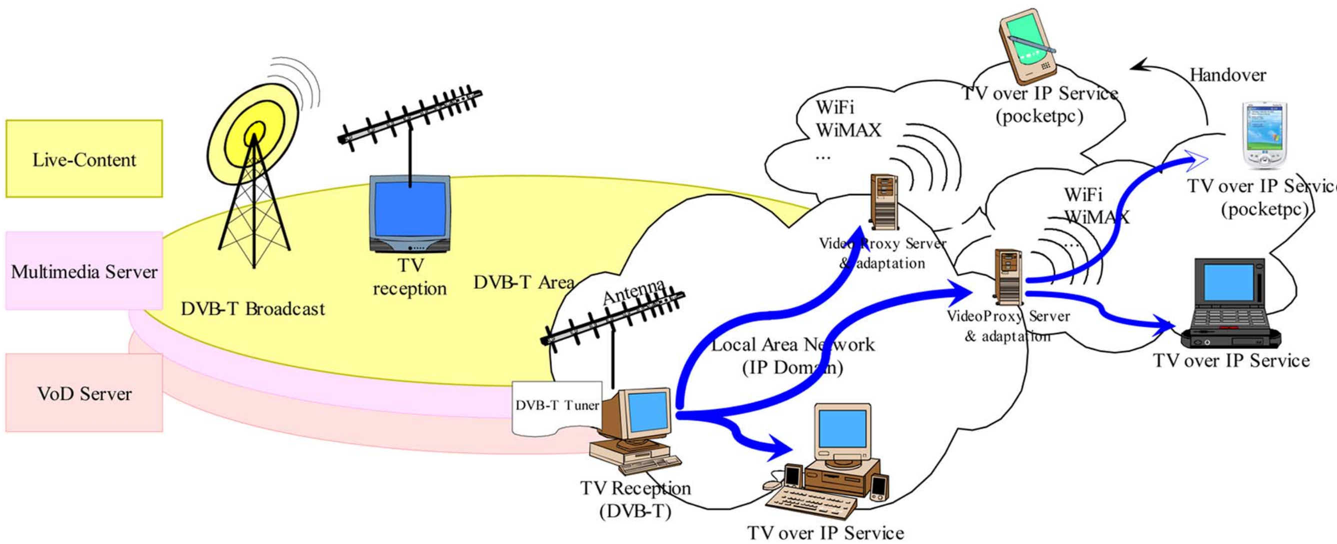

Fig. 2. Target architecture: Universal Media Access.

session (a context is maintained for each active multimedia stream). Thus, a different link-layer (respectively network, transport, application layer) adaptation strategy may take place for each handled multimedia stream regardless of possible QoS degradations experienced by other streams. For instance, link-layer transmission rate of a given stream may be deliberately decreased in response to harsh channel conditions due to receiver's distance from the access point (AP) or other environment-induced receiver's interferences. Based on the aforementioned QoS measurements, the server put into effect the appropriate combination of remedial actions available at different layers. Possible remedial actions include spatial/temporal resolution adaptation, per-pixel rate control, FEC redundancy adjustment, DiffServ re-marking, and MAC/PHY layer rate adaptation.

Live digital TV and VoD streaming services are used as a case study in our work, though MIM is not conceptually limited to this deployment scenario and may be, in practice, exploited with legacy Internet real-time and non real-time services. Our target architecture is described in Fig. 2. In this architecture, we particularly focus on WLAN's receivers. The WLAN is considered as a last-mile connection which is subject to quality degradations. We assume that multicasting multimedia streams between the various servers and TV receivers in the core and the wired portions of the network don't suffer from any degradation thanks to large capacities and resources over-provisioned. Quality degradations are essentially provoked by the wireless access network.

Signaling and "horizontal" interactions between receivers and servers may be supported by using several different signaling protocols (e.g., RTCP, RTSP, HTTP, and MMS). These signaling protocols allow receivers to report to the servers certain i) QoS metrics measurements achieved at different layers and ii) higher-level events triggered by terminal- and user- capabilities/preferences changes. We have extended the RTCP report to carry additional network-level QoS information such as loss distribution pattern, loss rate before, and after FEC recovery. Besides, we use RTSP together to transmit MPEG-21-compliant metadata regarding user, terminal, and encoding capabilities from the end-user's terminal to the streaming server. RTSP may be also used with an already active session to report on new events triggered by the user or the terminal. This protocol uses the IETF common format to express media and session descriptions, namely the session description protocol (SDP) [23]. Note that from the receivers' perspective, the content is exclusively located at the MIM-enabled Server (i.e., Video Proxy Server and Adaptation) as shown in Fig. 2. In the rest of this paper, we use the term video server to refer to the Video Proxy Server and Adaptation.

A. User/Application Level QoS Description Using MPEG-21 Metadata

SDP is basically used to describe i) session parameters for announcements or invitations and ii) the capabilities of a system, and possibly for providing a choice between a numbers of alternatives. In other words, SDP is used to convey information related to the multimedia stream before the session activation. This may include information related to both transport configuration (RTP and RTCP port number, RTP Payload Type definition, RTP payload header signalization, etc.) and multimedia encoding configuration (encoder parameters).

Current multimedia applications have higher demands in terms of advanced features and customizable configurations. However, SDP falls short in coping with these new demands. For instance, SDP does not provide information related to user profile (user information and preferences, usage history, presentation preferences, etc.), neither does it support the description of terminal capabilities such as adaptation capabilities. QoS parameters for different protocols such as traffic specification and flow specification or DSCP (Diffserv Code Point) for IP QoS differentiation are not supported by SDP. Yet, these QoS parameters need to be specified somehow (e.g., using out of band signaling). These deficiencies led the IETF MMUSIC working group to investigate a new generation of the SDP protocol namely SDPng (SDP next generation) to support various long-term extensions. On the other hand, MPEG-21 can fill the SDP gap by providing metadata related to the user, terminal, content and network characteristics. We believe that building SDP on top of MPEG-21 will provide the means to enable access to multimedia content under a wide range of delivery conditions and usage environments. In this perspective The IETF draft [24] presents a practical approach for harmonizing MPEG-21 with SDPng. In our work, SDPng is used to specify the context-layer of user preferences, content capabilities, and terminal requirements at session initiation stage. The specified parameters are then taken into account at the server level to appropriately configure the content and other adaptation mechanisms.

B. Cross-Layer Adaptation

Fig. 3 depicts the MIM cross-layer architecture and highlights the main top-down and bottom-up interactions between the involved layers. More precisely, the figure depicts the network architecture of a server which, unlike the receiver's, includes the MIM cross-layer adaptation engine. The latter is the core component centralizing i) server perception of current network-level conditions, ii) service-level configurations, and iii) current reported events and feedbacks from the receiver (i.e., client).

Detailed format of user characteristics, terminal capabilities, and content specifications are described in MPEG-21 metadata based on XML. The XML-compliant metadata coding constitutes the vehicle for vertical communications that take place between adjacent layers at each end-system. For example, and without loss of generality, a terminal that supports 8-bit color scheme could express its requirements down-to the RTP through an XML-based MPEG-21 element as follows (< m21-dia:ColorBitDepth blue="8" green="8" red="8"/>). Besides, the frame rate supported by a terminal could be expressed under the attribute "*refreshRate*" while the display capability of its screen could be

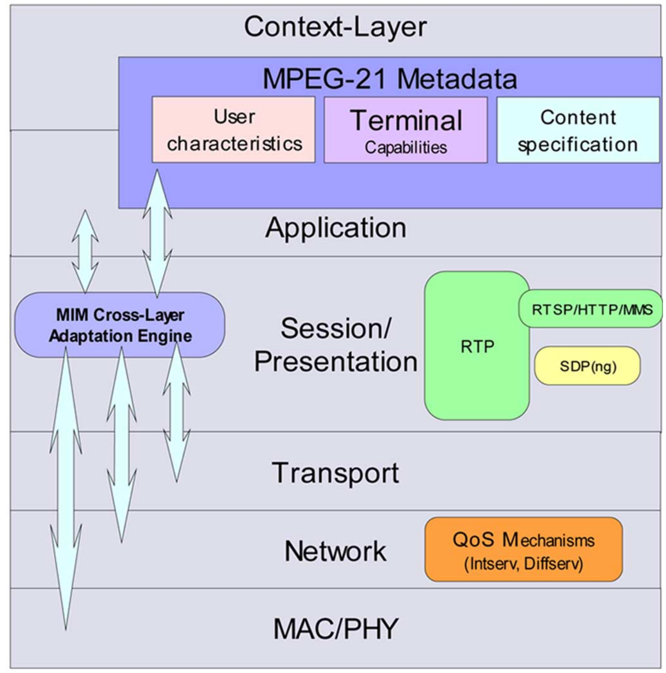

Fig. 3. MIM cross-layer interaction.

expressed in terms of pixels under the two attributes "*horizontal*" and "*vertical*." The transmission of terminal capabilities across the layers is an example of vertical communications that take place using MIM. Similarly, bottom-up interactions are used to communicate link and network performance measurements to upper layers as we describe it later in this paper.

XML-based requirements are enclosed into an SDPng message and then exchanged between end-systems (server and receiver) through the appropriate streaming session signaling (respectively maintenance) protocol such as RTSP. At the server, the SDPng message is parsed to extract the terminal (respectively user and content) capabilities that are again passed to RTP using the same MPEG-21 based XML encoding. Note that not only the receiver sends its capabilities to the server, but the server may as well send to the client its coding or transport capabilities at the streaming session initiation stage. This server to receiver communication may include video configuration specifics, streaming session IP address and port numbers, FEC stream decoding parameters, possible video rate adaptation granularity, and possible video temporal/spatial resolutions.

RTCP-based feedbacks reporting is another type of multimedia session signaling communication that is rather frequent and continuous during the session life-time. Typical information conveyed within RTCP reports are usual QoS metrics performance as well as certain advanced measurements such as loss pattern measurements. At the server side, this QoS feedback information is analyzed at the RTP level and then reflected at different layers by leveraging appropriate QoS adaptation mechanisms ranging from video encoding adaptation to PHY/MAC encoding and modulation re-setting. Note that not only network conditions are continuously reported, but also other user-centric parameters of UED (such as screen format, battery depletion level, user preferences, etc.) may be as well passed and continuously updated using RTSP.

At the network layer, an advanced interaction with upper and lower layers is necessary to ensure QoS continuity. A QoS matching and packet tagging algorithm is used to map application level QoS to network level QoS, and thus ensuring seamless IP differentiated service delivery. For instance, let's take the example of a video content structured as a single layer stream with I, P, and B frames. Each video frame could be carried over the IP Diffserv network with a particular DSCP code point. From the perspective of video distortion, it is known that an I-frame is more important than a P-frame which, in turn, is more important than a B-frame. Consequently, IP packets carrying I-frame fragments (or slices) will be marked with low drop precedence compared to IP packets carrying P-frame fragments, and so on.

Dynamic changing of context-related information should be also carefully considered (dynamic UED). Conventional advanced streaming systems collect network characteristics at the session initiation stage to statically configure basic video encoding parameters that usually remain unchanged throughout the streaming session lifetime. Using MIM, it is possible for the user to clearly express certain MPEG-21 based "*NetworkCharacteristic*" and thus put restrictions on maximum bandwidth ("*maxCapacity*"), average throughput, supported QoS classes, etc. These elements are also considered for dynamic adaptation. Some of them can be carried by end-to-end protocols such as RTCP. Several adaptation mechanisms can be performed at the server; some of which are described in the next section. It is worth noting that the main goal of this paper is not to focus on some adaptation scenarios or techniques but is rather to discuss and analyze the benefits of cross-layer adaptation in regards to collaboration between layers from MIM perspective.

C. Cross-layer Adaptation Strategies

This section describes three different cross-layer adaptation mechanisms: Link layer rate adaptation; adaptive FEC redundancy for video streams; and video content adaptation (temporal/spatial resolution and SNR level). All three mechanisms are based on MIM cross-layer interactions. We believe that the combination of these three adaptation mechanisms is sufficient to assess the ability of MIM cross-layer adaptation in facing changes in network conditions while meeting users' QoS expectations.

1) Link Layer Rate Adaptation: The 802.11 physical layer specifications introduce multirate capabilities which allow the physical layer to provide different channel coding and modulation. The link rate is strongly related to these two parameters. In fact, complex channel coding and modulation produce a higher link rates but are more sensitive to noise. On the other hand, lower link rates are based on stronger channel coding and modulation which are more resilient to noise. Therefore, it is more suitable to use a higher link rate when the wireless station are close to the Access Point and progressively decrease this link rate as the station moves away. The link layer rate adaptation automatically selects a transmission rate from a set of allowed rates based on the transmission conditions of the last frames. This automatic selection is performed by the Rate Control Algorithm (RCA) at Link layer. In [25], the authors identify three classes of RCAs: Statistics-based RCAs, SNR-based RCAs, and Hybrid RCAs. The Statistics-based RCAs maintain statistics information about the transmission conditions and the achievable throughput to adapt the rate consequently. Authors in [26], studied the performance of Auto Rate Fallback (ARF) algorithm, which belongs to statistics-based RCAs category since it is based on a continuous monitoring of the number of successfully acknowledged messages at link layer. The SNR-based RCA adapts the rate directly according to the SNR perceived by receivers since the appropriate channel coding and modulation are related to signal quality (SNR level). The link rate adaptation strategies proposed in [27] and [28] are based on these RCAs range. The third RCAs type, namely Hybrid RCAs, use a combination between the two above-mentioned techniques to take advantages of each of them and minimize their shortcomings.

The rate control algorithms (RCA) are used to allow maximizing the data throughput while minimizing the packet loss by switching between different rates (e.g., from 54 Mbit/s to 11 Mbit/s). In this context, work in [29] analyzes the streaming video quality and captures wireless LAN characteristics across network and wireless link layers. Authors investigate possible WLAN performance indicators that may be used to predict the streaming video quality. The results show that the wireless RSSI and average wireless link capacity are the most accurate indicators to predict the performance of streaming video over wireless LANs. In the similar optic, we have shown in [30] that when the signal strength decreases, the MAC frame error rate increases consequently. The "Sample" RCA, which can be classified in statics-based RCAs, uses a transmission time comparison achieved by different link rates. We demonstrated that adapting the video content according to link rate, decided by RCA, improves significantly the overall perceived video quality.

Based on this concept, it is clear that the correlation between the data rate and signal strength has to be taken into consideration in our MIM cross-layer adaptation where the signal quality measurements are used as input for triggering different adaptations.

2) Forward Error Correction (FEC) Adaptation: Packet loss is a problem that considerably affects the quality of received video quality at the client. It may lead to a very destructive effect on the reconstructed video sequence, because of video frames dependencies. This phenomenon is usually referred to as "*error propagation*." Packet loss may happen at different levels and due to different reasons. In wired packet-switched networks, congestion is the first cause of packet loss. Entire packets can be discarded by routers. Whereas in wireless networks, the transmission channel may cause frequent bit errors. Corrupted packets are discarded and hence considered lost.

Forward error correction can be applied at many levels from bit level up to packet level. In a bit level FEC, a bit is considered as a symbol while in packet level FEC, a symbol is a packet. Packet level FEC consists of producing " h" redundant packets from "k" original ones. FEC packet is generally based on erasure coding and its usefulness lies on that i) a single parity packet can be used to correct different single-packet losses in a group of packets, i.e., packets belonging to a given FEC block; ii) bit level FEC is unable to recover a completely lost or delayed packet; and iii) when using a bit level FEC, a corrupted packet is already detected and discarded at link layer (respectively transport layer) before being available at application level; hence, using a bit-level FEC at application level implies disabling error detection mechanisms (CRC and Checksum) of underlying layers.

Even though most of existing wireless access networks use adaptive coding and modulation schemes integrated to the link layer (see Section III-C.1), packet-level FEC protocols are usually required. As shown in [31], wireless communication experiences i) fast fading and white Gaussian noise, which are addressed by the integrated physical layer coding and ii) slow fading (e.g., when entering a tunnel), which is addressed by a packet level FEC encoding. These two levels of FEC encoding are complementary, each one addressing different problems. Clearly, there is a need for packet-level FEC protection, in addition to bit level FEC, to increase wireless multimedia communications reliability.

Typical packet-level FEC protocol that uses k media packets to produce n packets, among which $h = n - k$ parity packets, have the capacity to overcome up to h packet loss (when using MDS codes). This basically provides a resiliency against a maximum packet loss rate of $p = h/n$ when considering that even FEC packets may be affected by loss. Thus, based on the average packet loss rate measurements, such as those provided by the RTCP feedback, it is possible to adjust the level of redundancy each time as follows:

$$h = \frac{p \cdot k}{(1-p)}. \tag{1}$$

Based on the number (k) of media packets to protect and the measured mean loss rate (p), the number of FEC packets can be easily determined. Note that at the receiver, the loss rate (p) is measured at the transport level, which means before applying FEC recovery. This way, the server gets a consistent picture about the current network conditions in order to adjust the FEC redundancy. The overhead introduced by the FEC redundancy should be tackled by content adaptation mechanism such as transrating as proposed in this paper to maintain smooth bandwidth utilization.

Besides the transport-level monitoring and the network status measurements that may be carried out at different levels, delivering different indications are of great importance for FEC adaptation mechanisms. PHY/MAC layers signal strength measurements are among the most important QoS metrics in statistically shared environments like WLANs. Further, other advanced measurements such as loss pattern (packet loss distribution) may be useful for adjusting FEC transmission [32].

The frequency with which the network loss rate is reported to the sender may deteriorate the responsiveness of FEC schemes leading to suboptimal FEC efficiency. A high frequency would enhance the responsiveness at the sender while causing high variations between successive measurements (e.g., leading to instability of the system), not to mention the uncured excessive feedback overhead. In turn, a low frequency would trade-off good stability and low overhead for poor reactivity. In our case, we use RTCP reporting with a fixed frequency (up to 5% of RTP session bandwidth as recommended in the IETF standard). Furthermore, the RSSI is measured for each connected client at the access point which plays the role of proxy and video adaptation gateway. Clearly, using RSSI measurements, the adaptation gateway will have access to a coherent and up-to-date view of the network conditions perceived by each receiver. Indeed, the time-scale of physical layer measurement is very small and revealing of the short-term network conditions. This would considerably improve the responsiveness of FEC redundancy control as it takes into consideration the short-term degradation of signal quality. The MIM cross-layer FEC redundancy is based on both signal strength measurement at the client side and network packet loss ratio. It is adjusted dynamically to overcome as much as possible the video quality degradation. With the help of the content adaptation, the amount of bandwidth used can be maintained while FEC redundancy is added.

3) Content Adaptation: Transmitting packet video streams over WLAN encounters the problem of network capacity variation (i.e., bandwidth fluctuation) as the signal strength is unpredictable. The bandwidth of the path between the sender and the receiver is the most important characteristic that directly affects the quality of video services. It is however generally time varying and hardly predictable. If the sender transmits more than the available bandwidth, video packets may be lost or may experience excessive delays. A common technique to deal with bandwidth variation is to use adaptive content streaming. The server estimates the available bandwidth and then adapts its sending rate to match the available bandwidth. This technique is widely dependent on the video coding flexibility and features allowing for example video rate adaptation, multi-resolution streams adjustment, etc. The Scalable Video Coding (SVC) [33] defines three-dimensional scalability to allow for adaptation in heterogeneous environment by simply truncating appropriate bit streams parts. Temporal, spatial, and SNR (Signal to Noise Ratio) scalability are among the well known techniques used to tackle bandwidth variation and fluctuation. The temporal scalability is based on video frame rate, in which a higher quality layer corresponds to a higher video frame rate. In the spatial scalability, the quality layer has different video frame size and the quality increases by increasing the frame size of video. Finally, the SNR scalability is based on quantification factor which represents the visual quality of video pictures. These three dimensions are considered in our MIM cross-layer adaptation. Initially, content adaptation is performed at multimedia session initiation phase based on user preference (respectively terminal capabilities and encoding constraints) carried out using MPEG-21 UED; it is possible to select a streaming format out of various temporal/spatial resolution and SNR levels.

During multimedia transmission phase, we keep unchanged the temporal/special resolution for transmitted video. However, the SNR level is adapted by the server to control the video throughput. In fact, according to the cross-layer parameter (signal strength quality, change in link rate, and packet loss ratio) the server adjusts the FEC redundancy and adapts its sending rate by *transrating* the video content using quantification factor to overcome the overhead introduced by the FEC.

D. Implementation

The MIM cross-layer adaptation architecture is depicted in Fig. 4. It is composed of a media server and a client. The server streams audiovisual content to the client via an IP network using the RTP protocol.

Consider the use case of VoD streaming over IP with a multitude of users and 4 devices. At the connection phase, the client requests the content through the RTSP protocol. The

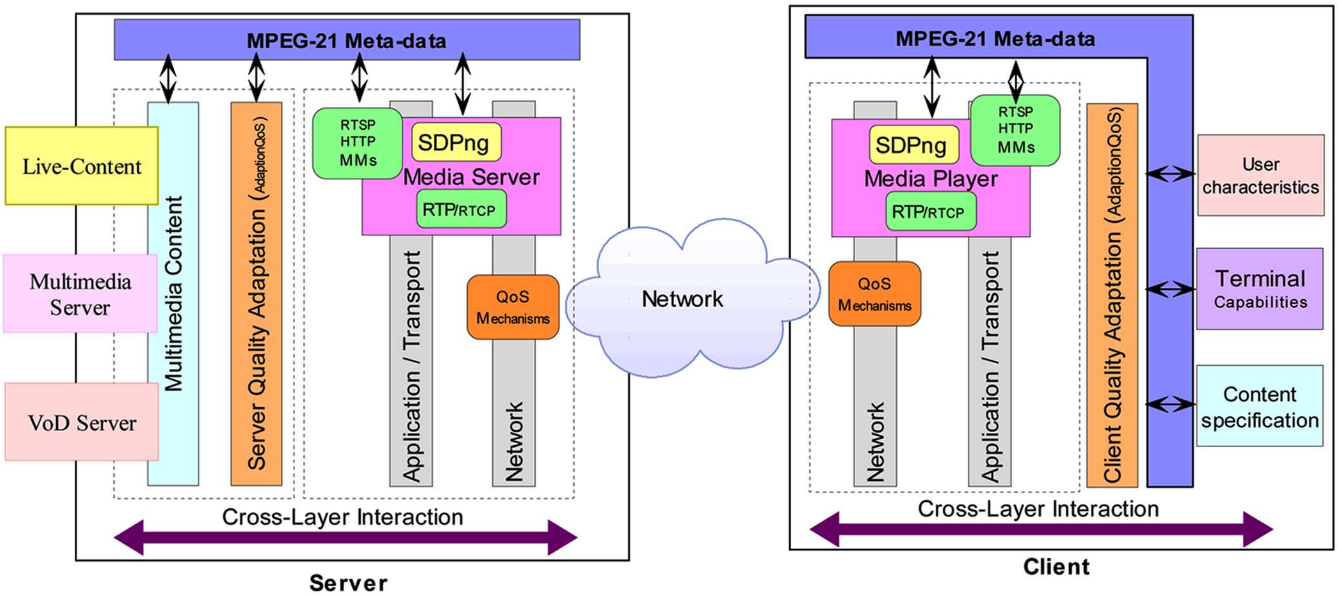

Fig. 4. MPEG-21-enabled cross-layer adaptation architecture.

MPEG-21 descriptions of UED are encoded in SDPng format and enclosed in HTTP or RTSP messages. The server's quality adaptation engine adapts the content based on UED descriptors (color scheme, refresh rate, screen size, codecs capabilities, and network characteristics, etc.). This represents the first MIM cross-layer interaction from the context-layer down-to the RTP layer. The server adapts dynamically to the changing conditions using the client signal strength, physical link rate, and packet loss-ratio that are carried through RTCP feedback. At this stage, the server adjusts the application-level QoS mechanisms such as FEC redundancy amount and video throughput.

During the streaming session lifetime, different types of QoS measurements can be collected and used to refresh UED descriptors made available at the server side for adaptation purposes. Such variety of measurements is not supported in our current implementation which is only limited to the aforementioned cross-layer parameters. In fact, standard RTCP "*fraction lost*" before and after FEC correction, "*cumulative number of packets lost*," and "*inter-arrival jitter*" allow the server to be continuously aware of long- and short-term changes in network conditions.

In our implementation, both the client and the server are based on open source project VideoLan (VLC) [34]. VLC is a highly portable multimedia server and player for various audio and video formats (MPEG-1, MPEG-2, MPEG-4, DivX, mp3, ogg ...) as well as DVDs, VCDs, DVB-S/T and various streaming protocols. It is used as a server to stream in unicast or multicast in IPv4 and IPv6 networks. VLC protocols have been extended to meet the requirements of our adaptation system. Our VoD server is based on VLC-0.8.5 running Linux 2.6.17 kernel.

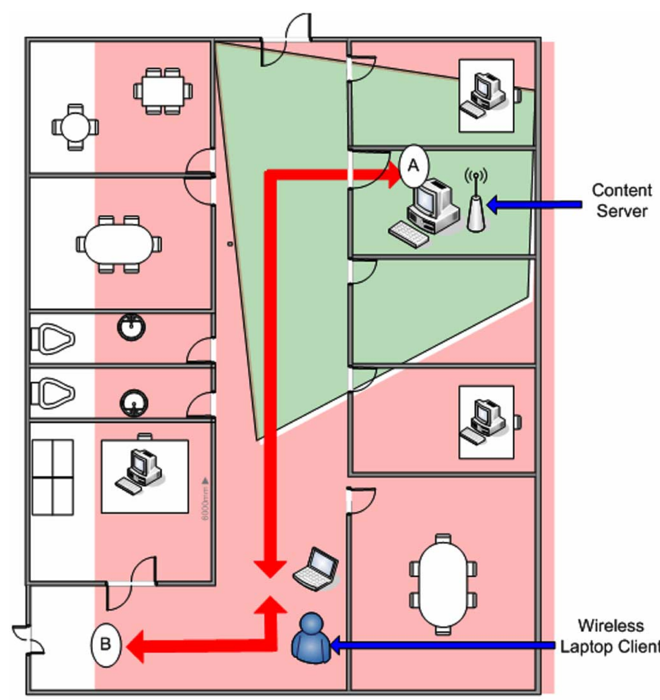

Fig. 5. Test-bed configuration and RSSI quality map of the environment.

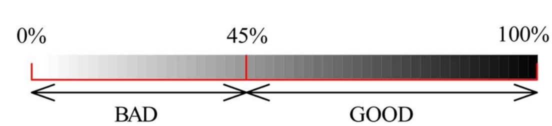

Fig. 6. RSSI percentage used in MIM.

IV. Performance Evaluation

We have deployed a test-bed to experiment with our proposed MIM cross-layer adaptation and evaluate its performance. The performance evaluation takes into account several aspects entailed by a practical deployment of the video streaming system, with a special focus on the appropriate QoS performance metrics to be measured by the network operator. The configuration of the test-bed is illustrated in Fig. 5.

In our experiments, we used an MPEG-4 coded Akiyo video sequence stream as a reference for testing. The Akiyo video sequence is 300-frames length with a video frame rate of 25

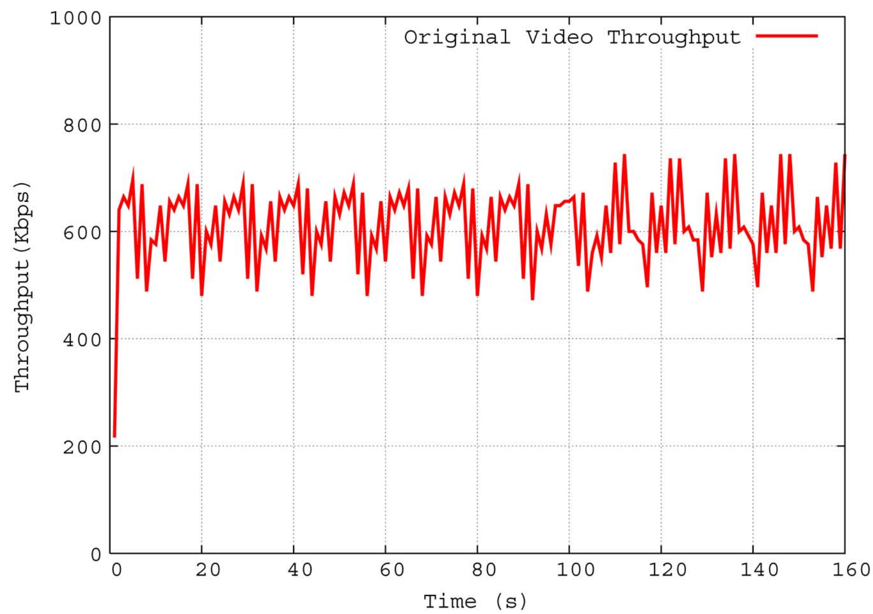

Fig. 7. Instantaneous throughput of the reference video.

frames per second. The server streams the video sequence continuously in a loop which we believe is fairly sufficient to highlight the main benefits of our architecture. The experiment duration is about 160 s during which the wireless mobile user moves from location A (content server) to location B, and then comes back to location A with a regular pace as illustrated by the red arrow in Fig. 5. The test-bed is deployed in a closed environment (laboratory), and the mobile user may be separated by up to three walls when it is at maximum distance from the Access Point. Both Access point and client wireless card are based on Atheros 802.11a/b/g chipsets. The Atheros chipset provides (through the MadWiFi open source driver [35]) two parameters to report on the channel state. The first is called "*signal strength*" which is measured in *dBm* with an interval range of $[-96, 0]$, the higher strength is 0 dBm and the lower is -96 dBm. The second parameter is called "*RSSI*" and provides a signal quality. The RSSI does not have unit and their range interval is [0, 94] where 0 represents the worst signal quality and 94 the best one. Obviously, there is a strong correlation between these two parameters variations as they both report different metrics of the link quality. However, for our MIM cross-layer adaptation, we choose to use the RSSI percentage by dividing the instantaneous RSSI value by the maximum RSSI which corresponds to index 94. The RSSI percentage provides a non-relative value which is more appropriate to measure the signal quality variation with acceptable accuracy and granularity. Fig. 6 shows the RSSI percentage used to steer the adaptation according to two ranges: 1) [0%–45%] is considered to represent bad link conditions and 2) [45%–100%] is considered to represent good link conditions. We also depict in Fig. 5 a rough estimation of the signal quality map (Good/Bad) measured in our lab environment. The locations where the signal quality is good are represented by green color, while red color represents locations where the signal quality is below the threshold (45%).

During the experiment, the same mobility pattern is each time assumed where the user moves from A to B. Due to time-

Packet loss rate	No loss	No FEC No overhead
	<5%	FEC(10,8) ➔ Overhead +25%
	>5%	FEC(10,7) ➔ Overhead +42%

Fig. 8. Example of simple adaptation strategies for scenario 2.

varying wireless channel conditions that depends on several unpredictable phenomenon and interferences, each run of the experiment produced slightly different results, but with always the same trend and within an acceptable disparity margin. In other words, the disparity in the measured signal quality is fairly minor. Initially, at time $\mathrm{t} = 0$ the mobile user moves from A to B with a speed of about 1 meter per second. At time $\mathrm{t} = 60$ seconds, the user arrives at location B. The user stays in location B until time $\mathrm{t} = 100$ seconds, where he experiences the worst channel condition. Afterward, the user returns back to location A and reaches its initial position at time $\mathrm{t} = 160$ seconds.

The Akiyo video sequence used in the experiment has a size of 368×242 pixels, an average throughput of 609 Kbps, and a peak rate of 760 Kbps as shown in Fig. 7. No audio traffic is sent during the experiment. In the rest of this section, we refer to the Akiyo video traffic as "*reference sequence*," which is available at [36]. In order to appropriately receive and display the video streams, we use a wireless laptop with 1.5 GHz processing capabilities and running a modified version of VLC-0.8.5. Video quality measurements are performed using objective metrics (PSNR and SSIM) that show the user-perceived quality gain entailed by the use of MIM cross-layer adaptation.

In the following, we do not consider the static adaptation carried by SDPng and performed at the connection phase. We focus only on different adaptation scenarios performed during the course of the streaming session. Although both adaptations have

		Packet loss rate		
		No loss	<5%	>5%
RSSI level	**Good** **[45% - 100%]**	- No FEC - No transrating	- FEC(10,8) ➔ **Overhead +25%** - Adapt the video ➔ **Transrating the video -25%**	- FEC(10,7) ➔ **Overhead +42%** - Adapt the video ➔ **Transrating the video -42%**
	Bad **[0%-45%[**	- FEC(10,8) ➔ **Overhead +25%** - Adapt the video ➔ **Transrating the video -25%**	- FEC(10,7) ➔ **Overhead +42%** - Adapt the video ➔ **Transrating the video -42%**	- FEC(10,7) ➔ **Overhead +42%** - Adapt the video ➔ **Transrating the video -42%**

Fig. 9. Example of simple adaptation strategies for scenario 3.

equal importance, we have focused on the dynamic adaptation for its ability to counter unforeseen events such as fluctuations in network conditions. With MIM cross-layer adaptation, the receiver continuously measures and reports the instantaneous packet loss ratio using RTCP. The server embedded at the access point measures the signal quality of all currently active users using the API of MadWiFi open source driver [35]. The above measurements (i.e., packet loss ratio, signal quality) are used by the server to re-adjust packet-level FEC redundancy transmission and to perform content adaptation. While the FEC increases the overhead, the video rate adaptation aims to reduce the video throughput. Both mechanisms applied together allow maintaining a smooth video bandwidth at the server. Note that the packet loss rates measures are collected at the receiver side before and after FEC recovery. This allows to accurately reporting on real network conditions experienced by the video traffic.

For the sake of comparison, we have tested and evaluated the performance of three streaming application scenarios, namely: a conventional streaming system, a streaming system with adaptive FEC only, and the MIM cross-layer adaptation. The evaluation focus on assessing the performance gain that may result from the combination of multi-layer QoS measurements such as transport-level loss rate and RSSI at PHY/MAC layer. The three tested scenarios are explained in the following.

- **Scenario 1:** reference sequence streaming with a conventional streaming system.
- **Scenario 2:** reference sequence streaming with adaptive streaming system. The server uses RTCP's packet loss rate measurements to adjust its FEC redundancy transmission. PHY/MAC-layer measurements are not collected. The adaptation strategies are shown in Fig. 8.
- **Scenario 3:** reference sequence streaming with the MIM cross-layer adaptation. We perform two different measurements to adapt the FEC redundancy: signal quality at link layer and packet loss ratio measurements at transport-level. The adaptation strategies are shown in Fig. 9.

The adaptation strategies introduced above are rather simple as they are intended to evaluate the benefits of the MIM cross-layer adaptation and to gain insight into the problem of coordinating different measurements at different layers. More specifically, we intend to show the advantage of using short-term and long-term performance measurements at different layers of the protocol stack such as signal quality and RTCP packet loss ratio. The RTCP loss threshold is fixed to 5% in the adaptation strategies in an arbitrary way so as to show the MIM cross-layer adaptation in our experimental test-bed. The objective is to capture the visual degradations (i.e., the subjective quality evaluation) entailed by packet loss and to emphasize the gain in performances achieved by MIM-based adaptation strategies. In real system, this threshold can be fixed by service provider according to the service level subscribed to by an end-user or according to service characteristics (loss ratio tolerated by content delivered to the client). More specifically, the loss threshold can be chosen directly by an end-user and integrated to the Usage Environment Description as part of the user preferences which are transmitted to the server during RTSP negotiation. The threshold (e.g., max loss rate) that determines the aggressiveness of MIM adaptation in response to network degradations is also dependent on the streamed media and their resiliency to packet loss. While it is commonly accepted that the video streams are, to certain extent, fairly resilient to packet loss, it is clear that different video encoding formats may present different resiliency levels. Further, it is also quite obvious that even different encoding efficiencies (rates, SNR level) with the same format results in different loss resiliencies. The more efficient the video encoding is, the more sensitive to error propagation effect, and the less the resiliency to loss the playback video is. The issue of establishing the optimal adaptation rules with the most appropriate adaptation thresholds is a multidimensional problem that should be addressed by taking into accounts both the user-to-service provider contracted SLA, the loss resiliency of the streamed media, and more generally the level of acceptance of users in terms of perceived video quality.

In order to deal with short term oscillations in signal quality measurement in scenario 3, we used a low-pass filter to smooth the consecutive measured value. The used low-pass filter is an Exponential Weighted Moving Average (EWMA) able to quickly detect unusual situations. It uses one exponential smoothing parameter to give more weight to recent observations and less weight to older observations or vice-versa, as shown in (2)

$$\mathrm{WRSSI} \leftarrow (1-\lambda) * \mathrm{RSSI} + \lambda * \mathrm{WRSSI}. \tag{2}$$

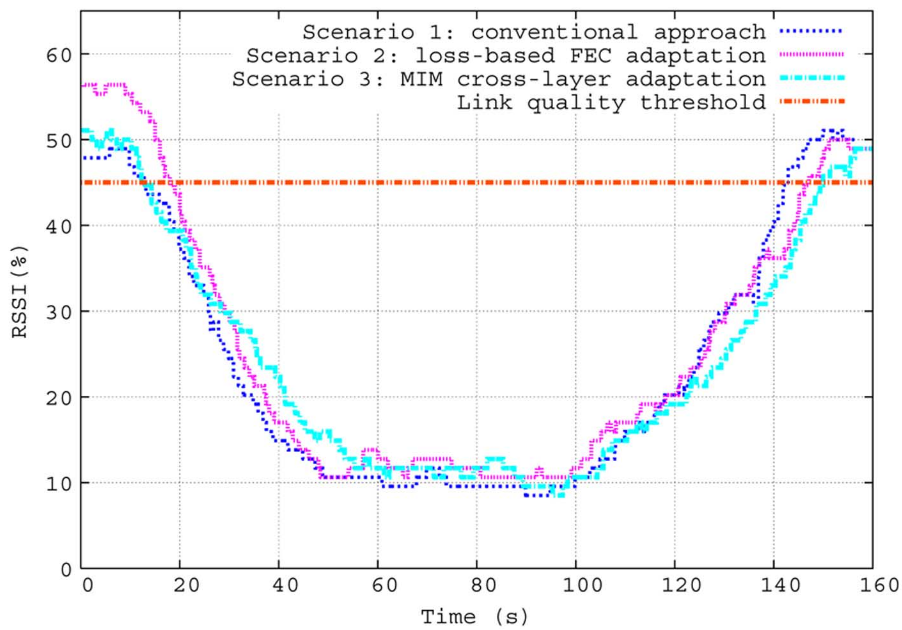

Fig. 10. Instantaneous smoothed measurement of Signal quality(%).

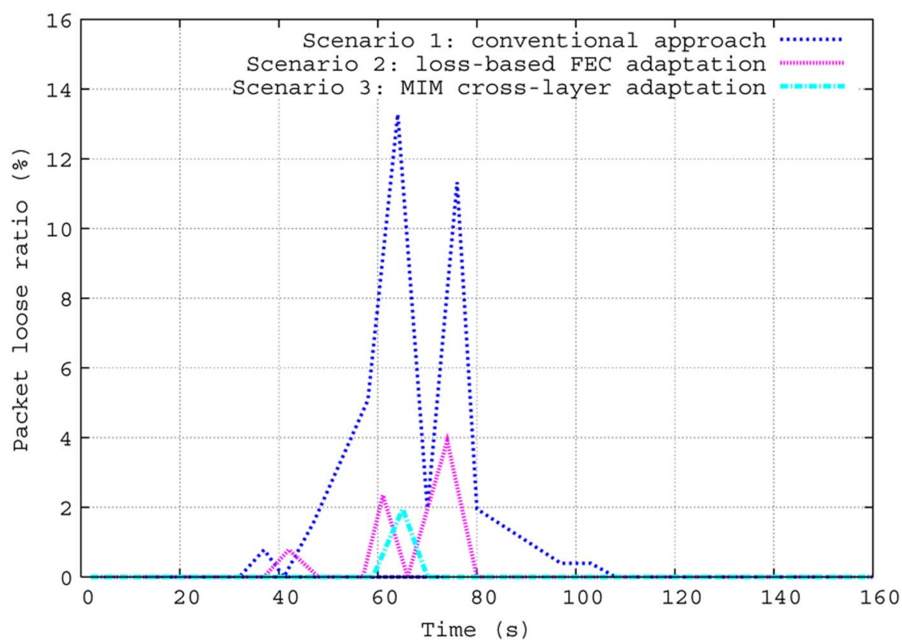

Fig. 11. Instantaneous packet loss ratio.

During the experiments, we choose $\lambda = 0.8$ to detect small shifts in link quality while limiting the effect of transient fluctuations. The adaptation decision is then based on the smoothed value of RSSI, namely WRSSI, rather than on instantaneous measurements.

Fig. 10 shows the measured signal quality experienced by the mobile user while roaming. The measurements are performed using Atheros driver which gives RSSI values ranging from 0 (0%) to 94 (100%) of quality percentage. As mentioned earlier, the signal quality threshold is set to 45%. In all three tested scenarios the experienced signal quality is more or less the same as the user moves, though there is some measurement variability between the different scenarios due to unpredictable noises and interferences which cannot be avoided when dealing with real experiments and test-beds. The signal quality drops significantly when the user is 30 meters away from location A with a RSSI of 10% as the average received signal quality. This leads to significant packet loss as shown in Fig. 11 for Scenarios 1 and 2. Here, packet loss ratios are measured after FEC recovery when FEC is used (Scenario 2 and 3). As expected, the measured packet loss ratio increases proportionally with the drop in the signal quality observed in Scenarios 1, 2, and 3 between $\mathrm{t} = 40$ s and $\mathrm{t} = 100$ s. The measured loss ratio for Scenario 1 can serve as a reference to asses the network-induced losses since no FEC recovering is

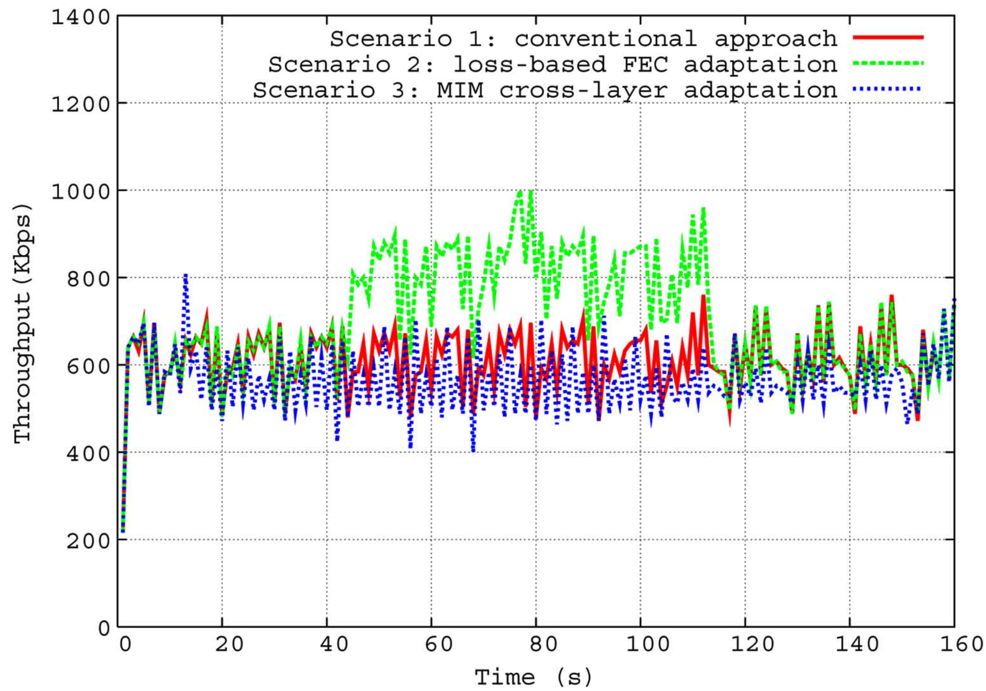

Fig. 12. Instantaneous video throughput achieved at the server side.

applied in this scenario. Scenario 3 used MIM experiences the lower loss as the signal drop, the FEC is systematically applied, and the content adaptation is used to reduce the overhead introduced by FEC packets.

Packet losses are translated into video frames dropping due to the cancellation of video-integrated error resiliency. Important packet losses with Scenario 1 are quite predictable since the server is not aware of network conditions and it does not perform any adaptation.

Scenario 2 shows better results than scenario 1 since the server uses the measured packet loss ratio reported by regular RTCP feedback to dynamically adapt the streaming process to network conditions. We can see that there is a significant oscillation in packet loss ratio as a consequence of the adaptive FEC transmission. In fact, as the loss ratio increases, the server responds by transmitting FEC packets to limit the experienced loss rates. However, as the server is unaware of current link layer measurements, it stops sending FEC packets and this leads again to significant packet loss. In contrast, the server in Scenario 3 keeps sending FEC packets since the measured signal quality is poor (bad). This allows the server to anticipate on packet losses that are most likely to occur frequently during poor link-quality periods. Furthermore, with the help of the video adaptation "*transrating*," the traffic sent by the server is not increased by the FEC redundant packet. This allows maintaining the same achieved throughput at the server side and avoiding the increase in the throughput which in turn may increase the packet losses. Fig. 12 shows the instantaneous throughput of the video traffic achieved at the server side before its transmission over the network. Scenarios 1 and 3 achieve a stable throughput during the experiment, while scenario 2 increases the throughput due to important transmission of FEC packets. The effect of the video adaptation in scenario 3 using MIM cross-layer adaptation is two folds: First, anticipating on eventual packet loss in the network by continuously monitoring the RSSI measurements. Second, the traffic transmitted by the server is maintained at the same level and aligned with the original video throughput. Thus, the FEC packets added by the adaptation mechanism do not contribute to increase the overall throughput of the video server. This reduces packet losses and enhances the overall video quality.

To assess the user-perceived QoS during link degradations, we use two relevant metrics to measure the objective video quality: Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [37]. The PSNR estimates the received image quality compared to the original image, while the SSIM measures the structural similarity between the original and the received image. The SSIM method has proved to be more accurate in respect to the Human Visual System (HVS). However, PSNR is widely used for measuring picture quality degradation based on mathematical analysis that proved to be fairly proportional to the human-perceived quality performances. It is derived from the root mean squared error. The PSNR for a degraded N1 ×N2 8-bit image f' compared to an original image f is computed according to (3) as follows:

$$\mathrm{PSNR} = 20 \times \log_{10} \frac{255}{\sqrt{\frac{1}{N_1 N_2}\sum_{x=0}^{N_1-1}\sum_{y=0}^{N_2-1}[f(x,y)-f'(x,y)]^2}}. \quad (3)$$

Fig. 13(a) and (b) show PSNR and SSIM results for scenario 1 and scenario 3 between frame #1000 and frame #2500. These frames are transmitted within the time interval elapsed between time $\mathrm{t} = 40$ seconds and time $\mathrm{t} = 100$ seconds which correspond to the poor link quality period where the signal quality is in its poor value (RSSI of 10%). Scenario 1 has a very

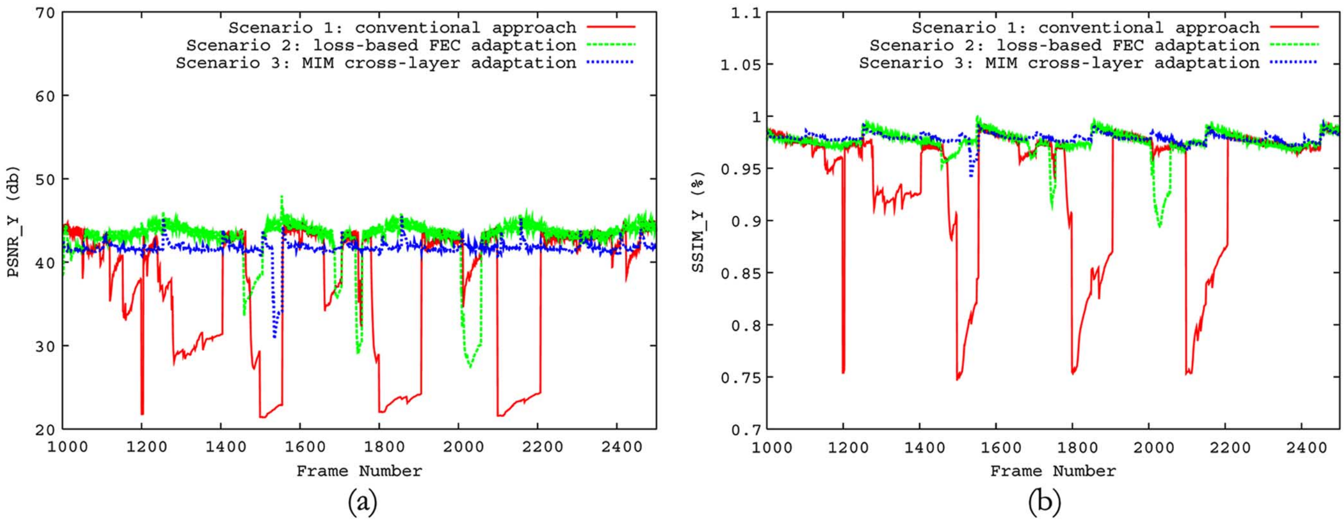

Fig. 13. Perceived video quality (PSNR) measurements for the different scenarios (PSNR and SSIM).

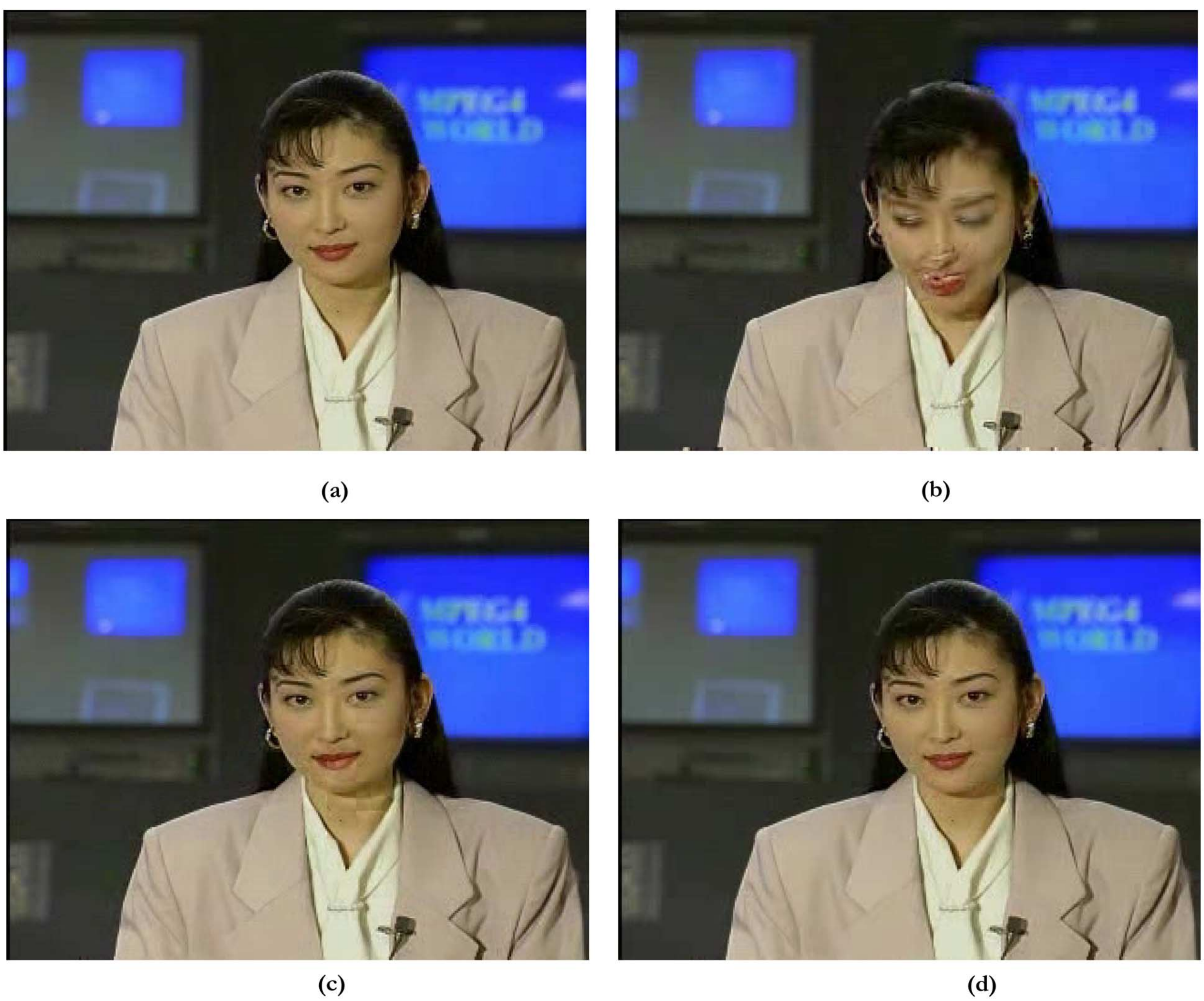

Fig. 14. Received video quality for Frame 1500. (a) original frame, (b) scenario 1, (c) scenario 2, and (d) scenario 3.

poor PSNR and SSIM quality. This is justified by the excessive packet loss ratio experienced during the aforementioned period. Scenario 2 produces an enhanced PSNR compared to scenario 3, though its value is oscillating during the period by reaching a minimum value of 28 dB. These oscillations affect considerably the subjective video quality.

The video adaptation performed by Scenario 3 affects slightly the PSNR values that keep stable during the critical period and maintain a good subjective video quality. These results are confirmed with the SSIM measurements. In fact, the SSIM of scenario 3 show better performances than scenario 2 and maintain a high level of structural similarity with an index near to 97%. These measurements demonstrate the benefits of MIM cross-layer adaptation in scenario 3 compared to 1 and 2. The received video of these different scenarios can be found in [36]. Coordination between different link quality measurements with different time-scales improves clearly the responsiveness and efficiency of streaming systems over wireless networks.

The snapshots of the received video for the original video [cf. Fig. 14(a)], scenario 1 [cf. Fig. 14(b)], scenario 2 [cf. Fig. 14(c)], and scenario 3 [cf. Fig. 14(d)] are compared for frame #1500. A significant enhancement in the perceived quality is noticeable when MIM cross-layer adaptation is applied.

Although the video is transrated in scenario 3, we cannot notice a difference between the original frame [cf. Fig. 14(a)] and the frame of scenario 3 [cf. Fig. 14(d)] using the original frame size [368×242 pixels]. However, the effect of packet loss on the objective video quality is clear in scenario 1 [cf. Fig. 14(b)] and scenario 2 [cf. Fig. 14(b)].

V. Conclusion

This paper introduced a new approach called Meet-in-the-Middle (MIM) for QoS-aware cross-layer design that conciliates both top-down and bottom-up interactions into an adaptive streaming system. MIM consists in gathering service-level characteristics (user, content, and terminal characteristics) and network-level status information into a single rich and dynamic context associated with each active video streaming session. Based on such context information, the server is able to perform a wide range of adaptive QoS mechanisms from content adaptation to PHY/MAC rate adaptation. In addition to the vertical cross-layer interactions that take place at each end-system, our approach involves horizontal communications between end-systems to enrich multimedia session context at the server side using additional measurements carried out at the receiver side. Performance evaluation of MIM cross-layer adaptation revealed a significant improvement of perceived quality compared to conventional approaches. In particular, combining different QoS metrics from different protocol layers proved to be useful in anticipating link quality degradations and thus increasing the responsiveness of the adaptation mechanisms. As future work we intend to develop analytical models for combining network-level metrics into meaningful network condition aggregates which can be used to improve the effectiveness of the adaptive streaming system. Another research direction is to further extend multimedia session context to include other information such as service level agreements. This would allow the service provider to offer a finer granularity of QoS guarantees at different prices.

References

[1] Information Technology—Multimedia Framework (MPEG-21)—Part 7: Digital Item Adaptation1st ed. Oct. 2004, ISO/IEC Std. 21000-7:2004.

[2] R. Stewart *et al.*, “RFC 2690: Stream control transmission protocol,” *Request For Comments, IETF*, Oct. 2000.

[3] E. Kohler, M. Handley, and S. Floyd, “RFC: 4340 datagram congestion control protocol (DCCP),” *Request For Comments, IETF*, Mar. 2006.

[4] L.-A. Larzon *et al.*, “RFC 3828 : The lightweight user datagram protocol (UDP-Lite),” *Request For Comments, IETF*, Jul. 2004.

[5] H. Schulzrinne *et al.*, “RTP: A transport protocol for real-time applications,” *IETF RFC 3550 (Obsolete 1889)*, Jul. 2003.

[6] D. Clark and D. Tennenhouse, “Architectural considerations for a new generation of protocols,” in *Proc. of ACM SIGCOMM’90*, Philadelphia, PA, Sept. 1990, pp. 200–208.

[7] M. Handley *et al.*, “TCP friendly rate control (TFRC): Protocol specification,” *IETF RFC 3448*, Jan. 2003.

[8] C. Verikoukis *et al.*, “Cross-layer optimization for wireless systems: A European research key challenge,” *IEEE Commun. Mag.*, vol. 43, no. 7, pp. 1–3, Jul. 2005.

[9] S. Shakkottai *et al.*, “Cross-layer design for wireless networks,” *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74–80, Oct. 2003.

[10] M. Van Der Schaar *et al.*, “Cross-layer wireless multimedia transmission: Challenges, principles, and pew paradigms,” *IEEE Wireless Commun. Mag.*, vol. 12, no. 4, pp. 50–58, Aug. 2005.

[11] Q. Zhang, W. Zhu, and Y.-Q. Zhang, “End-to-End QoS for video delivery over wireless internet,” *Proc. IEEE*, vol. 93, no. 1, pp. 123–134, Jan. 2005.

[12] S. Shakkotai, T. Rappaport, and P. Karlsson, “Cross-layer design for wireless networks,” *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74–80, Oct. 2003.

[13] J. Chen, T. Lv, and H. Zheng, “Cross-layer design for QoS wireless communications,” in *Proc. Int. Symp. Circuits and Systems, ISCAS*, 2004, pp. 176–220.

[14] L. T. Butala, “Cross-layer design for medium access control in CDMA Ad hoc networks,” in *EURASIP J. Appl. Signal Process.*, 2005, vol. 2005, no. 2, pp. 129–143.

[15] S. De, C. Qiao, D. A. Pados, M. Chatteriee, and S. J. Philip, “An integrated cross-layer study of wireless CDMA Sensor Networks,” *IEEE J. Select. Areas Commun.*, pp. 1271–1285, Sep. 2004.

[16] S. Krishnamachari, M. V. D. Schaar, S. Choi, and X. Xu, “Video streaming over wireless LANs: A cross-layer approach,” in *Proc. 13th Int. Packetvideo Workshop 2003*, Nantes, France, Apr. 2003.

[17] Y. Shan and A. Zakhor, “Cross layer techniques for adaptive video streaming over wireless networks,” in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2002, vol. 1, pp. 277–80.

[18] T. Ahmed, A. Mehaoua, R. Boutaba, and Y. Iraqi, “Adaptive packet video streaming over IP networks: A cross-layer approach,” *IEEE J. Select. Areas Commun.*, vol. 23, no. 2, pp. 385–401, 2005, in the.

[19] 4G MC-CDMA Multiple-Antenna System on Chip for Radio Enhancements [Online]. Available: http://www.ist-4more.org IST-507039

[20] Jointly Optimizing Multimedia Transmission in IP- Based Wireless Networks [Online]. Available: http://www.ist-phoenix.org IST-001812

[21] Network of Excellence on Wireless Communications [Online]. Available: http://newcom.ismb.it/public/index.jsp IST-507325

[22] [Online]. Available: http://www.enthrone.org/IST-507637

[23] M. Handley and V. Jacobsen, “SDP: Session description protocol,” *RFC 2327, IETF*, Apr. 1998.

[24] T. Guenkova-Luy *et al.*, “Harmonization of session and capability descriptions between SDPng and MPEG-21 digital item adaptation,” *IETF Draft, Work in Progress*, Feb. 2005.

[25] J. T. Haratcherev, K. Langendoen, R. Lagendijk, and H. Sips, “Automatic IEEE 802.11 rate control for streaming applications,” *Wireless Commun. Mobile Comput.*, vol. 5, pp. 412–437, 2005.

[26] T. Braskich, “Optimization of a link adaptation algorithm for voice over wireless LAN applications,” in *Proc. IEEE WCNC’05*, New Orleans, LA, Mar. 2005, vol. 3, pp. 1602–1607.

[27] D. Qiao *et al.*, “Goodput analysis and link adaptation for IEEE 802.11a wireless LANs,” *IEEE Trans. Mobile Comput.*, vol. 1, pp. 278–292, Dec. 2002.

[28] J.-L. C. Wu *et al.*, “An adaptive multirate IEEE 802.11 wireless LAN,” in *Proc. 15th Int. Conf. Information Networking*, Oita, Japan, Jan. 2001, pp. 411–418.

[29] M. Li *et al.*, “Weather forecasting—Predicting performance for streaming video over wireless LANs,” in *Proc. 15th ACM NOSSDAV’05*, Stevenson, WA, Jun. 2005.

[30] I. Djama and T. Ahmed, “A cross-layer interworking of DVB-T and WLAN for mobile IPTV service delivery,” *IEEE Trans. Broadcast.*, vol. 53, no. 1, pt. 2, pp. 382–390, Mar. 2007.

[31] H. Ernst *et al.*, “Transport layer coding for the land mobile satellite channel,” in *Proc. IEEE VTC’04*, Milan, Italy, May 2004, vol. 5, pp. 2916–2920.

[32] A. Nafaa *et al.*, "Joint loss pattern characterization and unequal interleaved FEC protection for robust H.264 video distribution over wireless LAN," *Comput. Networks J.*, vol. 49, no. 6, pp. 766–786, Dec. 19, 2005, et.
[33] Joint Scalable Video Model (JSVM). Hangzhou, Japan, Oct. 2006, 8.0, ITU-T VCEQ JVT-U202.
[34] VideoLan Open Source Project [Online]. Available: http://www.videolan.org/vlc/
[35] MadWiFi Open Source Driver [Online]. Available: http://madwifi.org/
[36] Download Page: Received Video Sequences [Online]. Available: http://www.labri.fr/perso/djama/ieee_tm.htm
[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

Ismail Djama (M'07) received the computer engineer degree from the Institut National d'Informatique, Algiers, Algeria, in 2003 and the master's degree in computer science from University of Besançon in 2005. He is pursuing the Ph.D. degree in computer science at CNRS-Lab LaBRI, University of Bordeaux, Talence, France.

His main research interests are the quality of service in next generation networks, cross-layer optimization, universal media access and delivery, multimedia adaptation, and multimedia protocols. Currently, he is working on European project ENTHRONE and different projects with CNES (Centre National d'études spatiales), Toulouse, France.

Toufik Ahmed (M'04) received the computer engineer degree (Honors) from the Institut National d'Informatique, Algiers, Algeria, in 1999 and the Master's and Ph.D. degrees in computer science from the University of Versailles, Versailles, France, in 2000 and 2003, respectively.

He was a Visiting Scientist at the School of Computer Science, University of Waterloo, Waterloo, ON, Canada, in 2002 and a Research Fellow at PRiSM Laboratory, University of Versailles, in 2003–2004. He is currently an Associate Professor at ENSEIRB School of Engineers/University of Bordeaux I, Bordeaux, France. His main research activities concern quality of service for multimedia wired and wireless networks, cross-layer optimization, and end-to-end signalling protocols. His work on quality of service and video delivering has led to many publications in major journals and conferences. He has worked on a number of projects including the IST IP ENTHRONE project funded by the European Commission.

Abdelhamid Nafaa (M'05) received the Masters and Ph.D. degrees in 2001 and 2005, respectively, from the University of Versailles, Versailles, France, where he was involved in several national and European projects including NMS, IST-ENTHRONE1, IST-ATHENA, and IST-IMOSAN.

He is a Marie Curie Research Fellow under the EU-FP6 EIF Marie Curie action that seeks broader synergy in the European research space. He has been granted the Marie Curie award to undertake research work at UCD in the area of multimedia services distribution over carrier-grade networks. Before joining UCD, he was an Assistant Professor at the University of Versailles and acted as Technology Consultant for U.S. and a European-based companies in the area of reliable multimedia communication over WiFi technology and IMS-based multicasting in DVB-S2 satellite networks, respectively. Recently, he has been involved in writing a successful EI-funded proposal VidAs that aim to develop P2P-based VOD services distribution for network operators. He is also involved in a successful FP7 proposal CARMEN that aim to develop a mixed WiFi/WiMax wireless mesh networks to support carrier-grade services. He is a co-author of over 25 technical journal or international conference papers on multimedia communications.

Raouf Boutaba (SM'01) received the M.Sc. and Ph.D. degrees in computer science from the University Pierre and Marie Curie, Paris, France, in 1990 and 1994, respectively.

He is currently a Professor of computer science at the University of Waterloo, Waterloo, ON, Canada. His research interests include network, resource, and service management in multimedia wired and wireless networks.

Dr. Boutaba is the founder and Editor-in-Chief of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and is on the editorial boards of several other journals. He is currently a distinguished lecturer of the IEEE Communications Society, the chairman of the IEEE Technical Committee on Information Infrastructure and the IFIP Working Group 6.6 on Network and Distributed Systems Management. He has received several best paper awards and other recognitions such as the Premier's Research Excellence Award.