1ST QUARTER 2008, VOLUME 10, NO. 1

IEEE
COMMUNICATIONS
SURVEYS
The Electronic Magazine of
Original Peer-Reviewed Survey Articles

www.comsoc.org/pubs/surveys

CROSS-LAYER DESIGN PROPOSALS FOR WIRELESS MOBILE NETWORKS: A SURVEY AND TAXONOMY

FOTIS FOUKALAS, VANGELIS GAZIS, AND NANCY ALONISTIOTI, UNIVERSITY OF ATHENS

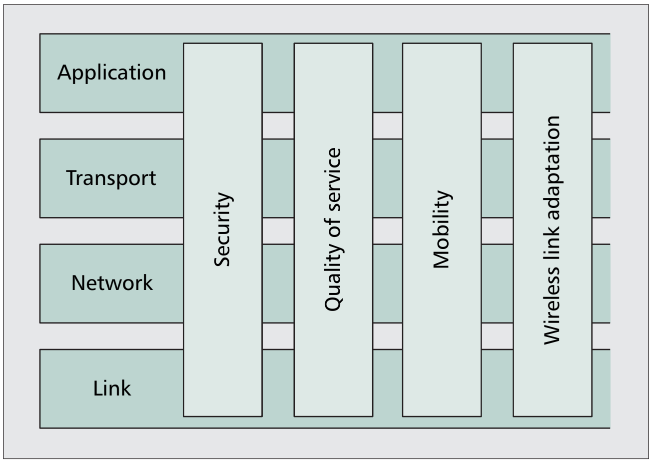
ABSTRACT

Third-generation (3G) and beyond 3G mobile communication systems must provide interoperability with the Internet, increase throughput for mobile devices, and optimize their operation for multimedia applications. The limited ability of traditional layered architectures to exploit the unique nature of wireless communication has fostered the introduction of cross-layer design solutions that allow optimized operation for mobile devices in the modern heterogeneous wireless environment. In this article we present the major cross-layer design solutions that handle such problems, and discuss cross-layer implementations with a focus on functional entities that support cross-layer processes and the respective signaling. In addition, we consider the associated architectural complexity and communication overhead they introduce. Furthermore, we point out the major open technical challenges in the cross-layer design research area. Finally, we conclude our article with a summary of cross-layer approaches developed thus far and provide directions for future work.

The continuing evolution of mobile communications has spawned several radio technologies (e.g., orthogonal frequency-division multiplexing [OFDM], code-division multiple access [CDMA]) and mobile network architectures (e.g., Third Generation Partnership Project [3GPP], 3GPP2) over the last few years. This technological proliferation has, in turn, brought on an unprecedented increase in the number of wireless access standards and their associated protocol stacks. The need for ever faster standardization cycles and the urgent demand to support access to the Internet by these mobile network architectures called for a "mix-and-match" approach to the definition of the associated protocol stacks. As individual protocols are typically specified with different assumptions in mind, the end-to-end performance of these protocol stacks in deployed mobile networks has not always been satisfactory.

Stratification, the composition mechanism for protocol frameworks, renders each protocol layer impervious to the functionality embedded within other protocol layers. Inside a protocol stack, exchange of control and data information may take place only between adjacent protocol layers and is supported by the concept of a service access point (SAP). A SAP provides access to a selected subset of protocol functionality via a precisely defined set of primitive operations. A particular protocol layer may offer and/or use more than one SAP, depending on its function and the information it needs to exchange with its adjacent protocol layers. This black box paradigm lies at the heart of the standard open systems interconnection (OSI) reference model and has been the prevalent design approach since the dawn of all modern networking architectures [1]. In this respect, any attempt to violate the OSI reference model is considered a cross-layer design. In the same context, in wireless networks any abstract model that rationalizes in a nontrivial manner the cross-layer interactions from the physical to the transport layer in order to allow information transfer across the layers' stratification boundaries is also considered a cross-layer design. In most cases cross-layer design jointly attunes a number of lower layers' parameters (e.g., channel state information) to upper-layer functions like transport and routing [2].

Regarding the underlying motivation, cross-layer design addresses problems of wireless network performance whose cause can be traced back to the original design assumptions underpinning the architecture of the employed network architectures and their protocol stacks (i.e., the black box paradigm). The well-known case of TCP's performance over wireless links is one of the most commonly cited applications of cross-layer design, but it is not the only one. Recently, numerous research efforts from around the globe have used



■ **Figure 1.** *A cross-layer coordination model [5].*

cross-layer solutions to improve the performance of wireless communication systems and protocol stacks in selected application areas. These range from the case of TCP's operation over wireless networks to more advanced topics such as maximizing the amount of users per service area (i.e., per radio cell) and adapting the encoding of multimedia content according to the state of the wireless channel.

Nowadays, the potential of cross-layer design for improving critical performance aspects of modern wireless networks is widely recognized, and the ability to support cross-layer interaction patterns throughout the protocol stack is considered an important property of beyond 3G mobile communication systems. The present article surveys existing cross-layer design applications and summarizes their key properties in a comprehensive taxonomy of cross-layer design approaches. The rest of the article is organized as follows. We present the architectures and models proposed so far within the concept of cross-layer design in wireless networks and identify key functional entities in cross-layer management procedures. We present a taxonomy into which existing cross-layer design efforts are classified and identify their salient functional properties. We present and categorize cross-layer signaling approaches. We identify open technical challenges associated with cross-layer design proposals and provide a qualitative investigation. We provide a comparative discussion of cross-layer design solutions and a summary of the lessons learned from the survey. Finally, we conclude the article with directions for future work.

CROSS-LAYER ARCHITECTURES, MODELS AND ENTITIES

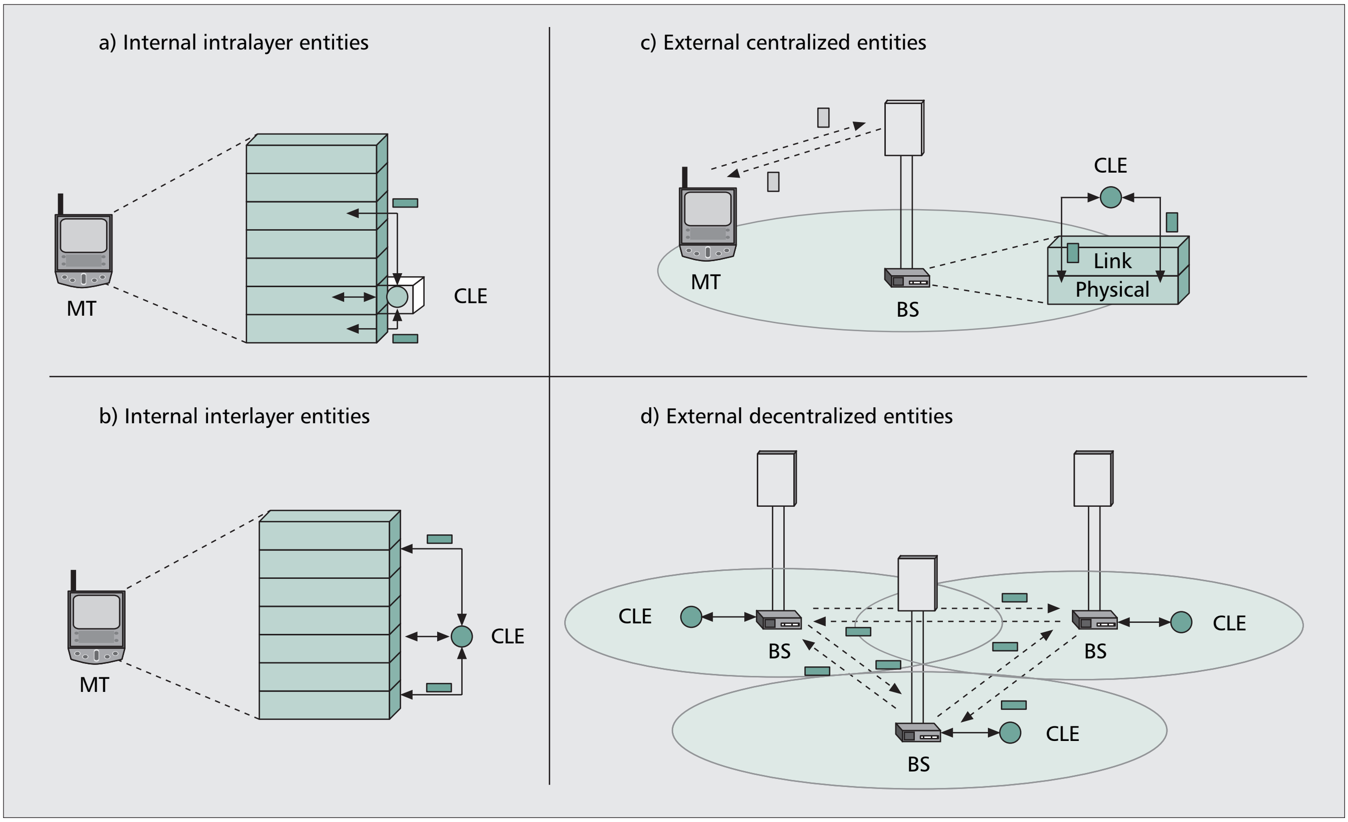
As so eloquently stated in [3], it is architecture that facilitates the decomposition of a system's functions into modular components that operate in unison to realize its purpose. By employing components optimized for specific purposes, a modular architecture allows for gains in performance and transparent system upgrades. At a design level, modularity is achieved by abstracting the functionality offered by each module via appropriate interfaces [4]. For instance, the OSI reference model encapsulates protocol functionalities into entities (i.e., protocol layers) that form the modules of the proposed architecture, while abstracting the internal details of each module through the interfaces it exposes to its adjacent layers [1].

Cross-layer design (CLD) allows communication to take place even between nonadjacent layers through additional entities introduced into the system's architecture. However, there is no reference model that specifies the functionality each new entity (i.e., module) must realize in a cross-layer design solution. To address this, [5] has proposed a model for determining the functionality that each new CLD entity might support. This model introduces four different planes that extend across the protocol layers of the OSI reference model in a visually vertical manner. Each of these so-called coordination planes encapsulates the behavior of a CLD algorithm or protocol targeted at solving a specific problem. In wireless mobile devices these problems include security, mobility, quality of service (QoS), and adaptation of the wireless link, thus leading to four coordination planes.

The security plane (Fig. 1) — The security plane coordinates encryption protocols and security technologies across different layers. Thus far, several encryption methods are available at various protocol layers. SSH and SSL provide end-to-end encryption at the transport and the application layer, and IPSec provides an end-to-end encryption at the network layer; in IEEE 802.11a/b/g wireless networks, Wired Equivalent Privacy (WEP) has been superseded by Wi-Fi Protected Access (WPA) for encryption. If each layer, independent of other layers, carries out encryption, unnecessary duplication of encryption functionality occurs, thus consuming more power, wasting valuable processing resources, and degrading network performance. Hence, there is the problem on which a CLD technique focuses is to determine which particular protocol layer should perform encryption. Thus, if the use of encryption schemes offered by different layers is coordinated by the security plane pertaining to CLD, the selection of a single encryption scheme suitable for whatever security requirements apply is possible — and, of course, desirable.

The QoS plane (Fig. 1) — Several QoS solutions have been proposed so far involving various protocol layers, such as RTP and TCP receiving QoS information from the application layer, and the integrated services (IntServ) and differentiated services (DiffServ) architectures developed by the Internet Engineering Task Force (IETF) support IP QoS. Developed according to the OSI reference model, these solutions do not support cross-layer communications, and QoS requirements are not conveyed to layers further along the protocol stack. In the time-varying wireless environment, however, the need to communicate protocol state information from the physical and link layers to the application layer, and to exploit it for improved QoS (e.g., in real-time data flows) is compelling. The provision of QoS information between nonadjacent protocol layers requires a cross-layer design. Hence, the QoS coordination plane must facilitate the communication of QoS information and coordinate the provision of QoS across multiple layers.

The mobility plane (Fig. 1) — Mobility supports the movement of wireless terminals from one service area to another through handovers to appropriate radio access points (i.e., cellular base stations). There are two handover categories: horizontal handover, where the mobile device moves between access points of the same technology, and vertical handover, dealing with mobile device movements between access points of different technologies. In both cases, upper layers must be able to mitigate the effects of handover, so mobility-related functionality must support the generation of notifications about handovers [18]. That will facilitate a smooth — and, ideally, seamless — transition of the mobile device's applications to the new wireless technology. To this end, the mobility coordination plane would take care of adapting the upper-layer services to the underlying wireless technologies.



■ **Figure 2.** *The different entities which (a, b) coordinate cross-layer management procedures in a protocol stack; and (c, d) which process cross-layer information in a network deployment.*

The wireless link adaptation plane (Fig. 1) — This plane addresses effects specific to the wireless link, i.e., channel fading, bit error rate (BER) variations, and transmission delays. These properties can affect the performance of upper layers, particularly that of TCP, which erroneously considers packet losses attributed to the instant state of the wireless channel as being caused by congestion in the end-to-end path. Several cross-layer solutions to indicate the actual cause of packet losses occurring at lower layers have been proposed thus far (e.g., the TCP-sleep protocol identifies losses related to channel fading effects). In this case the automatic repeat request (ARQ) protocol at the data link layer tries retransmissions. Obviously, retransmission of lost packets from both TCP and ARQ could halve the congestion window and, thus, the utilization of the wireless link. To avoid such rate degradations, the coordination between TCP and ARQ protocols is necessary [17].

Another important CLD aspect is the management of cross-layer interactions in a way that can guarantee the system's smooth operation. To this end, the aforementioned management model specifies an interlayer coordination manager responsible for the central coordination of CLD processes. In general, CLD introduces management entities that operate as either an optimizer of performance or a scheduler of some kind, depending on the problem at hand. Such an entity may reside within the protocol stack of the affected system, in which case it is considered an internal entity, or in an external network node. In the former case, the internal entity may be either an interlayer entity that coordinates the operation of all protocol stack layers or a set of intralayer entities, each of which is collocated with a protocol layer (Fig. 2). In the case of external entities, these may be centralized and hosted by a specific network node or distributed over several network nodes.

TYPES OF CROSS-LAYER MANAGEMENT ENTITIES

Internal Interlayer Entities

Interlayer Cross-Layer Manager (Fig. 2) — Reference [5] proposed a central interlayer coordination manager that applies cross-layer algorithms in any protocol stack layer. The coordination manager receives notifications for events occurring at protocol layers and is thus aware of the specific state any protocol layer is in. For instance, in the TCP case the congestion window and BER threshold state variables are used to trigger the connection initiated and link lost events, respectively.

Interlayer Cross-Layer Optimizer (Fig. 2) — Reference [6] proposed a CLD architecture that jointly optimizes the operational parameters of multiple layers via a cross-layer optimizer (CLO) entity responsible for optimizing N layers based on abstracted layer parameters. The abstraction of the layer parameters reduces the number of parameters the CLO needs in order to optimize the layer functionality. The benefit of this approach is that it provides a technology-independent way of interacting with each protocol layer. Consequently, the CLO can be deployed in heterogeneous networks comprising different wireless technologies and access systems. Layer abstraction identifies the parameters that expose the capabilities of the corresponding layer, thus enabling calculation of the proper values that optimize a specific objective function by the CLO. For instance, in the case of audiovisual transmission, the objective function to optimize may be the average peak signal-to-noise ratio (PSNR) that translates to the video quality perceived by a user. The performance criterion of the cross-layer optimization is the average PSNR between the encoded and displayed video stream, calculated through the rate distortion (RD) factor. The CLO employs a reconfiguration procedure to distribute the values of the abstracted parameters to the corresponding protocol layers. Each protocol layer is then responsible for matching the abstracted parameters and values into its own (i.e., internal) parameters that adapt its mode of operation. This approach incurs communication overhead due to the cross-layer information (i.e., the RD information) being conveyed from the video server to the CLO as well as some processing overhead during the reconfiguration process.

Internal Intralayer Entities

Intralayer Cross-Layer Optimizer (Fig. 2) — Reference [7] introduced a model for designing and implementing cross-layer feedback to allow direct communication between any pairs of layers in the protocol stack. This CLD model is called ÉCLAIR and consists of the following modules:

- The tuning layer (TL) provides an interface for invoking control information at a particular protocol layer. As control information specifies the behavior of protocols, the actual protocol behavior can be changed by properly manipulating its control information.
- The optimizing subsystem (OSS) activates optimization algorithms. The OSS collects control information from the TL through the protocol optimizers and adapts the protocol's behavior during runtime. To this end, the OSS contains a set of protocol optimizer (PO) entities. A PO implements an algorithm that addresses a specific cross-layer optimization. Hence, several specialized PO entities can be implemented and deployed according to the optimization purposes.

The ÉCLAIR architecture has been used in a feedback loop to control the bandwidth of running applications by tuning the receiver window of each TCP connection. In the ÉCLAIR architecture, applications set up the desired TCP window size to advertise their restrictions on network throughput. ÉCLAIR assigns a priority to each application and calculates the appropriate receiver window based on that priority. More specifically, the TL for the TCP layer (TCPTL) uses the priority to calculate the receiver window for each application. In this particular approach, changes in the protocol stack will only affect TL entities and the functionality of PO entities; hence, a PO does not depend on protocol layer code. All the same, the additional function calls between the OSS and the TL through the several POs incur internal overhead. However, ÉCLAIR achieves the following design goals:

- Minimal or zero processing overhead within the protocol stack since the OSS is executing concurrently with the protocol stack.
- Several PO entities can be dynamically deployed and ported to multiple technologies.

Intralayer Cross-Layer Scheduler (Fig. 2) — Reference [8] proposed a cross-layer adaptation framework for 802.16e orthogonal frequency-division multiple access (OFDMA) systems. It strives to achieve the highest system performance by exploiting cross-layer information between the medium access control (MAC) and physical layers. The MAC layer consists of the scheduling and resource allocation components that comprise a MAC scheduler and resource controller, respectively. The scheduler's algorithm determines the number of packets that could be transmitted to each user. The resource controller allocates the frequency bands for each user by using a channel-aware subcarrier allocation algorithm. In addition, a user grouper organizes individual users into groups according to the subchannel type of the 802.16e OFDMA standard.

The scheduler works in conjunction with the resource controller to increase the achievable throughput of users based on the channel quality information (CQI). Moreover, a hybrid ARQ protocol is deployed for purposes of link adaptation. The use of HARQ aids in the selection of the modulation and the coding scheme. Although the involvement of HARQ increases overall throughput, it also introduces a considerable amount of overhead when a large number of retransmissions occur. In 802.16e systems, retransmissions are associated to certain control messages that allow a mobile terminal to identify the correct packets in a data burst. However, these control messages can accrue up to 60 percent of the resources that should be allocated to mobile users. Consequently, allocation of channel resources should take into account in the amount of control messages and retransmissions.

External Entities

External Radio Scheduler (Fig. 2) — Reference [9] proposed a scheduling strategy for wideband CDMA (WCDMA) systems such as the Universal Mobile Telecommunication System (UMTS). This strategy exploits cross-layer information to improve system performance in terms of capacity and delay. It assigns users to priorities based on short-term channel variations instead of using only long-term ones. In a WCDMA system the radio base station (BS) provides each user with a transmission power $P_i(t)$. The BS sets the available transmission power P_T and, using a downlink fast control mechanism, notifies each wireless device of the minimum transmission power. Due to the slowly varying radio channel conditions, the power fluctuates around an average value.

By merging the rapid channel fluctuations of WCDMA, [9] proposed a scheduling scheme that prioritizes radio transmissions using a function that exploits these rapid channel fluctuations. This function takes into account not only the channel state (e.g., as typical multiuser diversity does), but also the channel variation experienced by each user. These priorities are evaluated for the downlink channel. To handle downlink radio conditions, a radio scheduler located in the BS need not use signaling to invoke power-related information since it uses fast power control information from the power control mechanism. Hence, in principle, the proposed scheme does not affect system performance, and its deployment in UMTS mobile networks could increase the number of served users. Simulation results quantify the realistically achievable gain of this strategy as up to 30 percent in capacity and 35 percent reduction in average channel access times.

External Centralized Cross-Layer Optimizer (Fig. 2) — Reference [10] proposed a CLD solution to address QoS provisioning over IP-based CDMA networks. The authors proposed a centralized cross-layer scheduler located at the BS that interacts with mobile terminals to exchange information regarding its traffic, power level, etc.eteras. This cross-layer scheduler supports a Dynamic Weight Generalized Processor Sharing (DWGPS) scheduling scheme according to which a video frame from the application layer is compressed to several batches of link layer (LL) packets according to its priority. To this end, the mobile terminal sends the batch class and batch size to the BS. The BS is also aware of the maximum tolerable delay over the wireless link as denoted by the timeout value.

This proposal takes into account the multiuser diversity gain that denotes the ratio of average transmission power for an LL packet. A piece of information also considered is the good/bad threshold F, since the bad channel state of a batch affects the backoff probability. Consequently, this threshold must be carefully set to avoid degrading the effectiveness of the backoff functionality. Whether a channel is in a bad or good state is a critical issue that must be estimated carefully,

Management entity	Objectives	Incurred overhead
Internal interlayer entities		
Interlayer cross-layer manager	Cross-layer design implementation based on an internal interlayer cross-layer manager, that manages the entire protocol stack.	N/A
Interlayer cross-layer optimizer	A cross-layer optimizer that is responsible for optimizing *N* layers with respect to an application-oriented function.	It causes external overhead when the cross-layer information is passed from the network to the terminal.
Internal intralayer entities		
Intralayer cross-layer optimizer	Each layer includes protocol optimizers that employ the appropriate algorithm for optimization purposes.	It incurs internal overhead due to the additional function calls.
Intralayer cross-layer scheduler	The scheduler and the controller can jointly improve users' throughput based on the channel quality information (CQI).	The control messages and retransmissions that the MS needs in locating its packet within a burst incur external overhead.
External entities		
External radio scheduler	The radio scheduler prioritizes radio transmissions based on the channel state and the channel variation of each user.	By employing fast power control information in base stations, external overhead due to signaling is avoided.
External centralized cross-layer optimizer	The MS sends to the BS the batch class, the batch size and the maximum tolerable delay indicating the timeout value.	The information exchange overhead is not significant. However, the good/bad state for a MS is a critical issue and must be estimated carefully.
External distributed cross-layer optimizer	The decentralized scheduler is employed to schedule L links satisfying simultaneously the interference constraints imposed by the distributed network model.	A decentralized scheduling approach introduces signaling overhead depending on the network size and topology.

■ Table 1. *Types of cross-layer functional entities and their impact.*

since in CDMA systems a mobile terminal does not transmit only when it has the best channel quality but also when its channel gain is no more than F dB less than the average value. Hence, the backoff probability must be small when the timer value for transmission is low, in which case the batch must be transmitted urgently. Moreover, the channel fading rate affects the backoff probability and consequently the timer's value.

Distributed Cross-Layer Optimizer — The majority of cross-layer designs focus on single-hop wireless networks (i.e., cellular networks). The aforementioned approaches concern the conventional case where a single access point serves mobile devices using radio cells. On the other hand, cross-layer design for resource allocation has already been applied in multihop wireless networks. Reference [11] discusses cross-layer design algorithms that operate in a distributed fashion. A decentralized scheduler is employed to schedule L links simultaneously satisfying the interference constraints imposed by the distributed network model and the associated general interference model. Such links are dominated by interference; consequently, they may suffer significant capacity losses in packet transmission and reception. Naturally, the introduction of a distributed and decentralized scheduling scheme introduces additional signaling overhead that,ultimately depends on the actual network size and topology [2].

Table 1 depicts the types of cross-layer management entities dealing with cross-layer optimization, scheduling, and information management in general. It also presents the associated overhead according to the reviewed works.

CROSS-LAYER SOLUTIONS FOR MOBILE COMMUNICATIONS

CLD solutions have been proposed to address different problems that arise due to the evolution of wireless mobile communications. The provision of Internet services over mobile communication networks has been the driving force in this evolution. On the other hand, the need to serve the largest possible population of users in next-generation mobile communication networks can be satisfied through cross-layer design that exploits valuable properties of the wireless channel. Cross-layer design optimization solutions can provide improved QoS to the mobile terminal for its multimedia applications. In the following we present a (nonexhaustive) list of cross-layer solutions developed for wireless mobile communications.

IMPROVING THE PERFORMANCE OF TCP OVER WIRELESS NETWORKS

Many CLD solutions have been introduced in order to improve TCP's performance over wireless links. Drawing the problem in outline, TCP invokes error and congestion control mechanisms such as the retransmission of TCP segments and the reduction of the congestion window whenever losses are detected, even though these may not be a result of congestion (i.e., losses caused by data corruption in the wireless medium). The inability of TCP to correctly identify the cause of packet losses is tackled by indicating explicitly either network congestion or transmission errors. In addition, the properties of underlying technologies mainly at the physical and MAC layers are exploited in an effort to improve TCP's performance [2]. The following subsections present the relevant cross-layer solutions.

Indicating Network Congestion — Wireless link reliability is questionable due to transmission errors. Link-layer mechanisms that tackle this deficiency are forward error correction (FEC), ARQ, and HARQ. The FEC mechanism enables the receiver to detect and correct errors [12]. As opposed to FEC, ARQ does not provide any error detection or correction, but solely grants frame retransmission from transmitter to receiver [13]. HARQ, in general, is a combination of FEC and ARQ. Particularly, it corrects transmission errors; however, if the channel quality is not at a good level, the receiver performs error detection before requesting retransmission. HARQ is recently deployed in 3G wireless systems and, in conjunction with adaptive modulation and coding (AMC), improves the performance of TCP over 3G wireless systems [14]. Nonetheless, even though these mechanisms improve the wireless channel's reliability, TCP will still treat all losses as congestion-related.

In wired networks, when congestion occurs, it is dealt with by the Adaptive Queue Management (AQM) mechanism offered by network routers. Consequently, it prevents the potential delays due to duplicate acknowldegments (ACKs) and packet retransmission. More specifically, router networks support Random Early Detection (RED) algorithms based on the average queue length exceeding a threshold. Explicit Congestion Notification (ECN) notifies the receiver of congestion in the end-to-end communication path. The congestion is indicated using a 2-bit-long ECN field in the IP header. On the other hand, ECN-capable TCP contains two additional fields in the TCP header for TCP-endpoint to TCP-endpoint signaling [15]. If the sending TCP entity is informed of congestion-related losses, it will avoid redundant retransmissions and thus facilitate the proper operation of congestion control at the TCP layer. However, as mentioned, the ECN mechanism was designed for a wired network and does not indicate the transmission error on the wireless link [16].

Indicating Transmission Errors — The Explicit Loss Notification (ELN) scheme notifies the TCP sender of packet losses caused by reasons unrelated to network congestion [19]. A snoop agent located at the BS treats a packet loss as a corruption in the wireless medium; however, this agent does not provide local recovery through packet retransmissions as does the agent in the snoop protocol [16]. More specifically, it retains a sequence block of ACKs indicating the successful transfer of packets from the sender to the receiver. It compares the previous with the newly arrived ACK value. If there is a gap in the sequence of received ACKs due to a packet loss, it sets the ELN bit. The ELN bit is contained in the TCP checksum since there is currently no specific bit in the TCP header for ELN. Whenever an ACK is received successfully, the agent cleans up the old block and retains a new one. The notification is passed to the sender by using either TCP header options in the packet header or Internet Control Message Protocol (ICMP) messages.

ECN and ELN are technology-independent and do not require any particular wireless network architecture or radio technology. There are, however, cases where the susceptible performance of TCP can be improved by exploiting the characteristic features of the underlying radio technology.

Exploiting Properties of Underlying Technologies — Due to the nonorthogonal nature of signals in a CDMA system, the interference among a user's substreams degrades the TCP performance of each user. When TCP users demand substantial throughput, interference among the associated radio signals downgrades the TCP transmission capacity [20]. Hence, in CDMA networks (e.g., multicarrier CDMA) where the available capacity is interference-limited, the objective is to improve TCP's performance with minimum impact on interference. However, such a solution requires cooperation between link layer resource allocation and TCP. Therefore, TCP should exploit the wireless link layer properties in order to achieve the target TCP throughput with the minimum possible amount of resources. In [21] the required resource amount depends on a resource vector denoted (M, Γ). The M value denotes the packets that can be scheduled for transmission in a slot, and the Γ value denotes the required bit-energy-to-interference-plus-noise density ratio (E_b/I_0) for all M packets. In the transport layer in particular, M_i is the target number of scheduled packets for TCP flow i, with a Γ_i value for the signal-to-interference-plus-noise ratio (SINR) level of the link layer unit (i.e., frame). The resource vector (M_i, Γ_i) determines the packet loss rate and transmission delay over a wireless link.

AMC is a widely known technique that is pertinent to matching the transmission rate to time-varying channel conditions [22]. It has been deployed in both WCDMA and WiMAX wireless broadband networks, and can realize several benefits for TCP's performance over wireless links [23, 24]. Reference [26] advocates a CLD approach that effectively conflates AMC with TCP in order to maximize TCP throughput. In particular, while sustaining a prescribed packet error rate (PER) P_0, better TCP performance can be achieved in terms of throughput by maximizing the data rate the AMC is able to render. By selecting the channel-dependent parameters such as the average of the received signal-to-noise ratio (SNR) g, the mobility-induced Doppler spread f_d, the fading parameter m, and the number of packets K the data link layer's queue can serve as well, the TCP throughput is improved for a prescribed P_0.

INCREASING THE RATIO AND CAPACITY OF SERVED USERS

Increasing the Ratio of Served Users — In cellular networks, multiple access methods have been used for the transport of voice data between BSs and mobile devices. These include time-/frequency-division multiple access (TDMA/FDMA) methods. Data transmission and telephony, however, create bursty traffic. In such a case, resources allocation like the allocation of a fixed number of time slots by transmitters in the TDMA method leads to underutilization due to the nature of traffic. Thus, even if a user is not transmitting any data (i.e., voice), the transmitters have already assigned the time slot to him/her; hence, the dedicated voice channel aimlessly consumes bandwidth [13]. The same problem is posed in FDMA-based communication systems. Consequently, the way to handle bandwidth efficiently is to allocate it in a flexible manner by using cross-layer techniques.

Asynchronous CDMA uses spectrum more efficiently than static allocation methods (TDMA/FDMA) when traffic is bursty in nature. More specifically, flexibility in spectrum allocation is provided by scheduling algorithms that decide which users are permitted to transmit during a specific time slot.

Objectives	Interactions	Summary
Improving the performance of TCP over wireless networks	Between data-link layer, network and transport layers.	Using notification mechanisms, the indication of network congestion and wireless channel errors is accomplished. The exploitation of parameters (i.e., channel conditions, frame size) from the lower layers helps improve TCP performance.
Increasing the ratio and the capacity of served users	Physical and data-link layer co-design and consolidation.	The ratio of the served users is increased by using the multi-user diversity and the power control mechanisms respectively in a cross-layer fashion. The co-design of AMC at the physical layer and HARQ at the data-link layer provides more capacity to users.
Application and optimization at the mobile application layer	Application and physical layer synchronization. Data -link layer involvement is possible.	Application-driven cross-layer optimization approach is evaluated using an application objective function (e.g., PSNR). The lower layers provide multi-user diversity, power control and knowledge of the frame's type. Statistics-based estimations at the data link layer enhance the optimization process.

■ Table 2. *The application categories on which most cross-layer design solutions focus.*

The decision depends on channel state information (CSI) fed back from users to BSs and is known as channel-state-dependent scheduling. In this case mobile users send CSI feedback to the access points indicating the effective received rate at which data can be transmitted. Such information is generally the received SNR from a certain user [20]. Thus, there is no need to continually reallocate time slots; on the contrary, only users with a good level of BER consume time slots. The gain achieved by this mechanism is known as the multiuser diversity gain [13].

Furthermore, 3G wireless systems have deployed cross-layer methods that exploit CSI in order to increase the served users and reduce undesired power radiation. These methods consist of policy scheduling algorithms that take into account information related to the user's channel state. As mentioned above, [9] has proposed such a scheduling algorithm using cross-layer information to provide more capacity and smaller delays for each mobile user. This algorithm gives priorities to each user based on short-term channel variations instead of using only long-term variations as was typically done in CDMA wireless networks. The short-term information expresses the knowledge of the transmission power variation at the frame level, while the long-term information implies the expected future conditions, which include the expected future value of the required transmission power for the next frame. The proposed algorithm has been simulated for UMTS downlink channels and could be applied in the current UMTS radio resource management (RRM) framework. The selection of users in terms of transmission power allocation has been made with respect to the following priority rules:

- Users with better channel conditions (i.e., requiring lower transmission power in the downlink) should have higher priority.
- Users with improving channels (i.e., the current needs in terms of transmission power are lower than those needed in the last N frames) can be afforded higher priority.
- Users that experience bad channel conditions for a rather small number of consecutive frames should be provided with a priority that compensates for their transmission delays.

Increasing the Capacity of Served Users — Reference [10] has also applied the multiuser diversity concept to real-time traffic. Due to the time-varying feature of CDMA channels, multiuser diversity provides the opportunity to use channel-aware scheduling methods employing the good/bad threshold value. More specifically, the data rate that can be achieved in CDMA channels is related to the SINR of one mobile station (MS). In CDMA the interference caused is intercell interference. In the downlink, if the BS only transmits to the MS with the highest SINR at time instant t, the maximum system throughput will be achieved. On the contrary, in the uplink, if only one MS transmits to the BS, the minimum intercell interference will be achieved. In a multicell environment, channel-aware scheduling leads to an increase in total system capacity. However, in a CDMA system an MS does not transmit only when it has the best channel quality, but also when its channel gain is not F db less than the average value. As mentioned above, the F value implies the good/bad threshold.

Moreover, the capacity of served users is realized through the spectral efficiency. Reference [27] describes a CLD between physical and data link layers in order to achieve higher spectral efficiency. This combination contains an AMC scheme and an ARQ mechanism at the data link layer. Given that the maximum number of retransmissions must be within delay constraints, an optimal design of the AMC scheme at the physical layer should be targeted to maximize the spectral efficiency. If the maximum number of retransmissions allowed per packet is N_r^{max}, the probability of packet loss after N_r^{max} retransmissions must be no larger than an upper bound imposed by the application requirements (e.g., video transmission). After that, the joint design consists of an AMC that satisfies a PER upper bound at the physical layer and an ARQ that grants the N_r^{max} upper bound at the data link layer. The experimental results of this approach show that a small number of retransmissions in conjunction with the chosen modulation-coding pair (mode), the latter determining the SNR level of the communication channel, can improve spectral efficiency in terms of bits per transmitted symbol. Of course, an arbitrary increase of retransmissions severely downgrades spectral efficiency.

In much the same concept, [28] combines AMC with an HARQ scheme against the truncated ARQ used in [27]. More specifically, [28] uses a type-I HARQ mechanism for packet retransmission and aims for maximum optimization under a prescribed delay constraint. The target PER determines the target BER for a transmission block and, in turn, the carrier-to-noise ratio (CNR) region of the AMC scheme can be determined. However, contrary to [27], which used a fixed packet size, [28] takes into account a variable packet size L. It is demonstrated that by adjusting the packet size in conjunction with the CNR region's level, a high spectral efficiency can be achieved. Furthermore, [29] combines type-II HARQ with AMC against the pure ARQ mechanism. Moreover, it presents a comparison between type-II and type-I HARQ mechanisms by numerical results. Particularly, the authors observed that type-II HARQ improves spectrum efficiency more than type-I HARQ does in the high CNR region specified by the

AMC scheme. On the other hand, the smallest packet size L amplifies spectral efficiency.

Reference [30] also proposed a CLD methodology based on an AMC scheme at the physical layer in conjunction with a scheduling mechanism at the MAC function of the data link layer. It proposes a scheduler that considers the channel state of the physical layer and the queue state of the data link layer. The queue is a finite-length buffer that is implemented at the gateway of the wireless access network. The adopted access mechanism is the time-division multiplexing (TDM)/TDMA. For each user i the numbers of time slots that can actually be scheduled are assigned depending on both the channel state and queue state. The reserved time slots are conditioned by the prescribed QoS levels for each user.

ADAPTATION AND OPTIMIZATION AT THE MOBILE APPLICATION LAYER

Cross-layer optimization for wireless video streaming can offset the end-to-end distortion caused in the received video at the application layer [6]. Such a cross-layer solution is accomplished by exchanging information between tje source and channel coder in the application and physical layers, respectively. This cross-layer information is known as source significance information (SSI) [31]. Unequal Error Protection (UEP) is such a cross-layer approach that protects important information from impairments caused by channel errors. This approach is also known as Joint Source/Channel Coding (JSCC) [32]. In CDMA networks the use of power control improves the error probability by combating the negative effects of interference. More specifically, the power level is determined by the energy-per-bit to multiple access interference (MAI) density ratio $\gamma_b = E_b/N_0$. Reference [32] presents an evolutionary approach tp JSCC that proposes joint control between the source coding and power control in terms of the source rate R_s of the video codec and the average γ_b of the physical layer, respectively. This approach is known as Joint Source Coding-Power Control (JSCPC) and attempts to achieve an end user's QoS level by adjusting the combination of R_s and γ_b.

However, adaptation at the application layer such as video streaming could benefit from link adaptation as well. Reference [33] presents such a combined solution that improves real-time streaming video quality by adapting the video encoding rate at the application layer according to MAC layer statistics-based information (e.g., throughput) in conjunction with the physical layer. Especially in the case of multiuser access, throughput predictions at the MAC layer give feedback to the systems' combination.

On the other hand, application layer optimization should be based on a user-perceived video quality factor like the quantitative parameter PSNR that indicates video distortion. To this end, in JSCPC methodology the optimal operational distortion rate for a single user with effective bandwidth requirement W_{eq} in hertz is given as

$$PSNR(W_{eq}) = \max PSNR(R_s, \gamma_b). \tag{1}$$

Reference [6] seeks to maximize the expected user-perceived video quality. This goal can be achieved by selecting the optimal parameter values for each group of pictures (GOP). A GOP is a sequence of groups of consecutive frames of the video stream. The expected user-perceived video quality implies that the video reconstruction quality at the user's side is

$$D = D_s + D_L. \tag{2}$$

The D_s term is a source rate function and defines the reconstruction quality in the error-free case, called *source distortion*. The D_L term is a packet loss rate function and represents the distortion derived from the transmission. The D_L is called *loss distortion*. By combining the so-called distortion profile information (i.e., the rate vector and distortion matrix) plus the transmission probabilities from the two-state Markov packet burst loss model, the radio link parameters can be chosen. Using this approach, the expected quality can be achieved for a particular application, and radio link layer parameters can be set with respect to the desired objective function (i.e., the PSNR) depending on the value of D.

Moreover, JSCC in conjunction with the multiuser diversity methodology can improve the QoS in time-varying CDMA channels. Reference [10] proposes such a cross-layer approach called multiuser adaptation, which exploits information in the application and physical layers. In previous sections we discussed the LL units named batches as well as the class and arrival batch size that represent the crucial cross-layer information. In the transmission queue within the BS, each batch LL unit is managed according to its priority for the corresponding session. In order to exploit multiuser diversity in cross-layer design, each batch i is determined as in bad channel state when the channel fades fast, as mentioned above. Whether a batch is in a bad channel state depends on the aforementioned good/bad threshold F. Comparing the threshold values $F = 10$ dB and $F = 5$ dB, it is inferred that in the case of $F = 5$ dB the service degradation in terms of LL unit losses is greater than in the case of $F = 10$ dB. This is because the threshold $F = 5$ dB sustains the bad channel state probability of one MS, and the corresponding batch is considered idle for a longer period of time.

Thus far we have classified and discussed a representative list of CLD proposals in terms of their objectives. Table 2 presents the interactions between layers that are accomplished by each of these solutions.

CROSS-LAYER SIGNALING

Given that there are no particular restrictions in the way cross-layer signaling takes place, various approaches have been adopted by the implementations thus far [34]:

- An additional packet header carries forward or indicates cross-layer information (CLI).
- Header options report changes in lower layers.
- Profiles and labels contain and indicate the meaningful CLI, respectively.
- A network service collects and distributes related CLI.

Inevitably, cross-layer signaling will incur some overhead; [34] points out the evaluation criteria of cross-layer signaling methods and presents a representative comparison. All the same, research in cross-layer signaling and notification mechanisms must strive to find efficient answers to a set of intriguing questions: What is the best way of indicating CLI, a header option or an additional header? Should cross-layer signaling be an in-band or an out-of-band one? In what format is CLI stored and conveyed? In which part of the network should an entity (i.e., server, router, manager etc.) be located for collecting and managing CLI? What is the overhead introduced by signaling during the setup phase or throughout the session?

To this end, next we extend the presentation of [34], pointing out the different types of cross-layer signaling mechanisms such as in-band and out-of-band signaling, the employed protocols, the format of transport messages/files, and the distribution of introduced overhead. We focus on the signaling protocol, the distribution pattern on network nodes (servers, routers, gateways, etc.) involved in it, and the potential overheads it incurs. As the detailed investigation of CLD signaling options is still in progress [35], a detailed exhaustive presentation is beyond the scope of this article.

SIGNALING MECHANISMS AND PROTOCOLS

Reference [41] presents the design concept of interlayer signaling and notification termed the *hints and notifications* (HANs) concept. Hints are messages traveling in the downward direction (i.e., from higher toward lower layers of the protocol stack) and are merely additional information on packet internals. Hints can be either an additional header or an in-band protocol (e.g., IPv6) [34]. On the other hand, notifications are messages rising from the lower protocol layers to notify the upper protocol layers regarding the current operational state (i.e., channel state, network congestion). To this end, notifications could be header options of in-band protocols (e.g., IP and TCP header options for ECNs) [35]. However, separate out-of-band protocols such as ICMP and RTCP are also candidates for notifying upper layers [31, 37].

In case of local profiles, two mechanisms are required, as discussed in the following. One mechanism is a transport mechanism that, in respect to the local adaptation (LA) concept presented in [38], could be performed by an XML-based mechanism or, alternatively, using an additional packet header. In this way a collection of parameters called layer-independent descriptors (LIDs) are transferred in order to be organized and stored in profiles locally. Each packet carrying audiovisual data is associated with an LID profile that indicates the adaptation capabilities of the content to lower protocol layers. This indication is a label in the IP packet header that associates the current packet with a specific profile, resulting in a so-called LID-label binding established in network routers. The Resource Reservation Protocol (RSVP), an out-of-band signaling protocol, is used to distribute the binding between LIDs and labels across routers [35].

On the other hand, [39] facilitates the transmission of CLI by means of a network service. The network service is realized through a so-called wireless channel interface (WCI). The WCI plays the role of mediator between network operators and mobile clients. The WCI employs an *abstraction engine* that first treats the parameters gathered and then distills them in an abstraction format that is meaningful to the mobile clients. A WCI server gathers parameters from the radio BSs and other network elements, and makes them available or forwards them to its clients through a proxy server that also implements the abstraction engine. Mobile clients receive or retrieve these parameters from the network server in the form of XML descriptors. However, a mechanism is required to select parameters from the network. SNMP is used for monitoring and selecting parameters that can be deployed as either an in-band or out-of-band mechanism [40].

SIGNALING ASPECTS AND EVALUATION

Signaling messages derived from lower layers are directed in most cases to the transport and application layers. The propagation of messages to the TCP layer, whether by an in-band (e.g., ECN) or out-of-band (e.g., ICMP, RTCP) approach, ends up in a specific socket. On the other hand, the propagation of notifications to the application layer ends up in a particular application. Frequent notification propagations pose internal overhead problems when those signals are delivered to application and transport layers [35]. Therefore, the higher the number of sockets and applications, the greater the incurred internal overhead. Moreover, the message type format (i.e., header options, additional header) affects the associated host's performance [36].

However, considering signals transferred from one side (the sender) to the other side (the receiver), the focus for localizing the possible incurred overhead should be additionally on the signaling mechanism (i.e., in-band or out-of-band). It is generally argued that in-band messages create less internal overhead than a separate message (i.e., out-of-band) does [36]. On the other hand, out-of-band messages preserve the normal protocol flow from additional external overhead [35]. All the same, in such a case the processing (i.e., internal at end systems and routers) overhead due to additional packet flow can be considerable [34]. Furthermore, the additional control path maintained for the out-of-band messages also incurs external overhead.

Overheads are incurred during session setup or raised through a session. In the case of local profiles, there is no need for transferring CLI throughout the session; hence, the external overhead of cross-layer signaling is limited to the session setup phase [38]. In network service signaling, either a mobile client or a WCI server initiates an application session. In this phase negotiations regarding cross-layer parameters and formats take place between mobile clients and the WCI server through the proxy [39]. Thus, the majority of external overhead is limited in the negotiation phase in this CLD signaling method. On the other hand, internal overhead will depend on the complexity of the abstraction engine.

Figure 3 illustrates the different ways of conveying CLI between and across protocol layers. Signaling aspects such as transport and signaling mechanisms in cross-layer communication are summarized in Table 3.

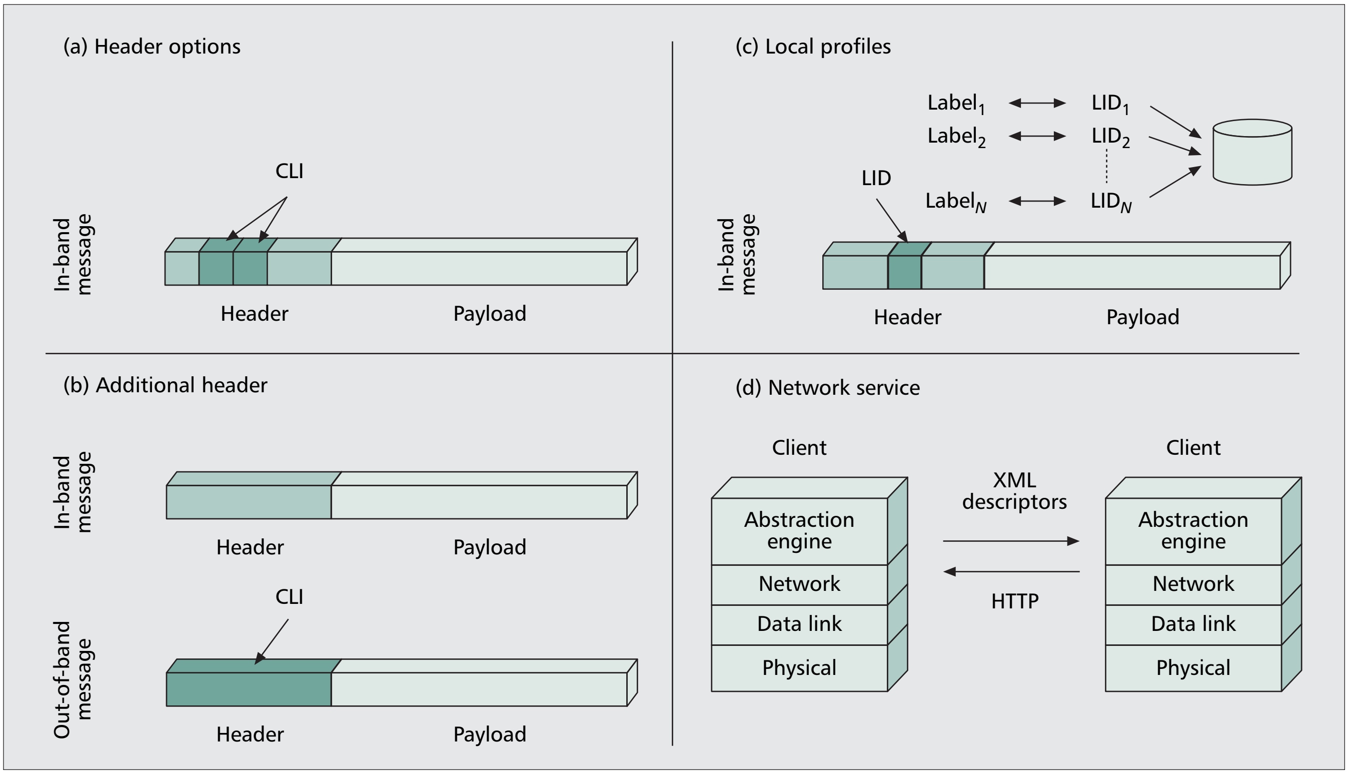
ELABORATING ON OPEN TECHNICAL CHALLENGES

As the area of cross-layer design is further developed, concerns are being voiced regarding its architectural repercussions, calling for a more cautionary approach in its use as a design artifact [3]. Some of the open challenges that lie on the path towards an efficient resolution of these concerns are [4]:

Standardization of Interfaces/Mechanisms for CLD — As previously discussed, the CLD architecture should provide the functionality for its own modules. To this end, an important question concerns the potential interfaces between these modules. The need for information exchange and sharing between nonadjacent protocol layers will determine these interfaces. Moreover, the layer parameters will indicate the flow of information between layers and therefore the direction of information exchange inside the protocol stack [3]. To this end, layer abstraction can expose the mechanisms and parameters of each layer [6].

Standardization can provide a unique vehicle for smoothly deploying various cross-layer design solutions in next-generation mobile communication networks. However, the investigation, specification, development, and, ultimately, standardization of cross-layer entities, interfaces, and algorithms to meet the need for cross-layer optimizations and dynamic interaction patterns between the protocol layers remain an open technical challenge.

The Coexistence of Different CLD Solutions — The main consideration for this issue is whether CLD solutions that intend to solve the same problem could be independently deployed in a transparent manner. For example, how can a set



■ **Figure 3.** *The ways of exchanging and indicating cross-layer information.*

of different cross-layer scheduling algorithms based on AMC be deployed at different times without changing the physical layer with regard to the set of exploited parameters? Are there common mechanisms different CLD approaches may use? If so, the dynamic deployment of different algorithms could be achieved with minimal impact on each individual layer's implementation. However, that does not preclude the case of adopting a custom mechanism for each particular circumstance.

The Role of the Physical Layer in CLD — In wireless networks the physical layer plays an important role. Advanced signal processing at the physical layer provides valuable functions such as rate adaptation, channel-aware scheduling, and subcarrier allocation. Nonetheless, the inherent variability of the wireless medium may impact the function of network layer protocols, thus affecting end-to-end performance. CLD mainly relies on the unique features of the physical layer to achieve better QoS over multicell wireless networks such as CDMA and OFDM.

IDENTIYING CLD MECHANISMS AND INTERFACES

Notification-Based CLD — With regard to the class of notification-based CLD approaches, we have identified the need for the following categories of mechanisms.

Congestion and Loss Recognition at the Link Layer — In wired networks congestion is addressed through an adaptive queue management (AQM) mechanism in network routers that support RED algorithms based on the average queue length exceeding a specific threshold [15]. However, in CDMA wireless networks mechanisms such as RED cannot be provided because the uplink does not provide any shared buffer [42]. In this case the estimation of load is used as a proxy measure of congestion. As the uplink in CDMA networks is interference-limited, this depends on the intracell interference experienced by one mobile user due to the transmission of other mobile users in the same cell. In the downlink case the load estimation is estimated based on the power transmitted from the BS since the downlink is power-limited.

On the other hand, ELN uses an agent running at the BS that inspects the loss due to corruption in the wireless medium [16]. More specifically, the agent retains a sequence block of ACKs indicating the successful transfer of packets from the sender to the receiver. By comparing the previous with the newly arrived ACK value, a packet loss can be detected and the ELN bit set accordingly.

Congestion and Loss Indication to the Network and Transport Layer — As mentioned previously, notifications are conveyed by IP header options or ICMP messages at the network layer. For example, ECN uses a 2-bit-long ECN field in the IP header. These fields are the ECN-capable transport (ECT), which indicates the ECN capability of the end node, and the congestion experienced (CE) field used by routers to indicate the congestion on the end-to-end path [15]. ELN uses in-band ICMP signaling at the IP layer [16]. Another explicit notification, Explicit Bad Station Notification (EBSN), is implemented as a type of ICMP message [43]. In this approach the BS sends ICMP messages to the TCP sender to request either the postponement of timeouts or retransmission of packets [44].

On the other hand, to convey cross-layer information between a pair of TCP endpoints, a TCP receiver adds the ECN-Echo (ECE) flag in the TCP header to inform the TCP sender that a CE packet has been received. The TCP sender sets the congestion window reduced (CWR) flag in the TCP header to inform the data receiver that the cwnd parameter has been reduced [15]. Another approach to explicit notifications, Explicit Wireless Loss Notification (EWLN), uses the sequence number of the ACK at the TCP header in order to indicate the specific packet loss sent from the TCP receiver to the sender over the wireless link in the EWLN [45]. Alternatively, the TCP checksum is used to determine the ELN bit [16].

Features	Header options	Packet header	Local profiles	Network service
Signaling mechanism	In-band	In-band or Out-of-band	In-band (header options indicate labels)	In-band or out-of-band (depends on SNMP)
CLI transport mechanism	No transport (notification)	Through packet header	XML descriptors or packet headers	XML descriptors (access network) SNMP (core network)
Overhead type and localization	Internal (low) external (low)	Internal (low) external (medium)	Throughout the session set-up (deterministic)	Throughout the negotiation phase (occasional)

■ Table 3. *Signaling aspects in cross-layer communication.*

TCP Sender Reactions upon Receiving Notifications — In the ECN-capable TCP, *cwnd* reduction occurs upon detection of the CE field. However, the TCP sender should not reduce its *cwnd* value more than once per window of data and must reset the retransmit timer when the *cwnd* is 1 MSS (maximum segment size). If the sending TCP continues sending ECE messages while the *cwnd* is 1 MSS, the retransmission throughput drops to one packet per round-trip time (RTT) [15].

Contrary to ECN, an ELN-capable TCP sender is fully aware of which segment is lost and retransmits it without reducing *cwnd* or any kind of congestion control action. The TCP sender identifies the lost segment by the corrupted counter field received in the TCP option field in conjunction with the checksum field (ELN bit) in the received ACK [16]. In the same way as [44, 45], an ICMP-RETRANSMIT message and an EWLN bit request retransmission of the segment by the TCP sender when the BS has failed to retransmit that packet and when the receiver detects errors in a TCP segment, respectively. In EWLN, the ACK contains the sequence number of the corrupted packet that has been lost.

It is known that, in addition to the retransmissions at the transport layer, the BS can initiate retransmissions at the link layer. In such a case the BS must notify the TCP sender, which must undertake the following actions: hold the retransmit timeout (RTO) of the corresponding packet the link layer tries to retransmit, and update the source timer based on the estimation of the RTT in order to prevent packet retransmissions [43][44].

User-Driven Notification — In certain cases, users may want to control the download rate on their devices. To this end, a priority parameter indicates the user's requirements in terms of bandwidth for each running application. Based on this parameter, the TCP receiver can adjust its window for each application [7]. The calculated receiver window is passed to the TCP sender through an ACK as the conventional TCP does. Thus, for each application, the TCP sender will not exceed a particular amount of sent data before it waits for a new ACK.

Cross-Layer Optimization and Scheduling — We discussed notification-based mechanisms that improve TCP's performance by mitigating the undesired effects of TCP congestion control. There are also CLD solutions that improve TCP's performance or offer better QoS to upper protocol layers by means of cross-layer optimization and scheduling mechanisms. In cross-layer optimization several alternative combinations of protocol layer parameters can be pursued [25]. To this end, such an optimization problem can be formulated by combining several parameters from different layers across the protocol stack. The aim of the following paragraphs is to identify and/or clarify some common parameters to be combined in cross-layer optimization procedures.

Error Correction at the Data Link Layer — References [12, 46] propose a mechanism at the link layer dedicated to correcting packet errors instead of retransmitting them. They rely on adaptive FEC by which the sender is able to select the appropriate FEC taking into account an estimated PER calculated at the link layer [12]. A link layer agent is placed at the mobile node and the mobile host. At the mobile node the link layer agent estimates the PER and RTT of the TCP session [46].

Adaptive Modulation and Coding Scheme at the Physical Layer — This approach deploys AMC at the physical layer to enhance TCP's throughput performance [26]. The AMC scheme serves a finite-length queue of K packets at the data link layer. The PER estimation at the link layer depends on mode n of the AMC scheme and the received γ. The channel fading is adjusted by the so-called Nakagami parameter m, and the channel state transition probabilities are modeled by a finite state Markov chain model. The latter is affected by the mobility-induced Doppler spread. By changing the K, γ, and m parameter values, TCP's throughput is adjusted to match the target PER.

Queuing Control at the Data Link Layer — In [21] TCP's throughput is enhanced by controlling the queue performance and the packet loss event of TCP in the wireless link. In the proposed cross-layer optimization concept, link layer efficiency is determined by the number m of LL units to be transmitted by multicarrier CDMA (MC-CDMA) in a slot and the SINR Γ value for all m LL units. Taking into account the received power for an LL unit of flow i in a slot, the amount of resources in a slot is defined as a function of the resource vector (m, Γ). In the link layer a RED-like buffer is used for retransmission of LL units. By calculating the mean queuing delay for TCP packets in the wireless link using the RED queuing mechanism, the target TCP throughput can be realized, keeping the resource vector (m, Γ) in optimal values.

Retransmissions at the Wireless Link — In [27] a CLD approach that jointly combines AMC at the physical layer with truncated ARQ at the data link layer enhances the average spectral efficiency in terms of transmitted bits per symbol. The average RTT and packet length N_p at the data link layer manifest delay constraints. On the other hand, the probability of packet loss after the maximum number of retransmissions specify the PER upper bound. Given the PER upper bound, the γ regions (i.e., the received SNR bounds) of the AMC scheme in conjunction with the $N_r^{\max}$ optimize system performance in terms of transmitted bits per symbol. The channel fading is adjusted by the Nakagami parameter m. On the other hand, the number of $N_r^{\max}$ can enhance or downgrade the PER in these γ regions.

In the same context, the CLD solution in [28] adopts type-I HARQ in order to minimize buffer size and augment spectral efficiency. The spectral efficiency optimization results rely on both $N_r^{\max}$ and packet length L changes. Additionally, the work presented in [29] optimizes the average spectral efficiency using type-II HARQ with rate-compatible punctured convolutional (RCPC) codes. Changing the rates of RCPC codes, the spectral efficiency is improved in the range of γ regions. Moreover, changes in the packet length L improve the spectral efficiency for a specific γ region.

Estimation and Control of Power at the Physical Layer — Scheduling strategies and priorities in CDMA networks can be relied on to provide the estimated required power for a user in the next transmission frame [9]. Moreover, given the received power strength inferred by fast power control information in BSs and the imposed packet delay threshold, prediction of the SINR parameter value for the next frame and, consequently, the BER and PER, can be calculated [47]. Therefor, based on measured power from the power control mechanism, predictions of the radio channel's condition determine the priorities of packet scheduling.

Video Reconstruction and Adaptation at the Application Layer — An application-driven cross-layer optimization approach could be relied on to measure the video quality perceived by users. The PSNR is a quantitative parameter that implies the reconstruction quality of an image pertinent to the original image at the receiver.

For video transmission over wireless links, the video reconstruction quality at the receiver is the sum of the source distortion D_s and the expected loss distortion D_L. The D_s expresses the error-free image and should be sent along with the video bitstream. On the other hand, the D_L is related to the packet loss rate caused during transmission [6]. The packet loss rate estimated at the data link layer results from information about the physical layer such as the modulation scheme (binary phase shift keying [BPSK], quaternary PSK [QPSK], etc.), channel coding, channel estimates (i.e., SNR), and transmit power.

In the same context a PSNR value is expressed as a function of the video's source and channel condition characteristics depicting the capability of video transmission over wireless links when a JSCPC is deployed [32]. The video source and channel condition characteristics are exposed by the video source coding rate R_s and the power level in a CDMA network expressed as the energy-per-bit-to-interference-density ratio, respectively. The data rate at the application layer is adapted by changing the quantization step size of the encoder. Additionally, statistical parameters from the MAC layer (e.g., throughput, spectral efficiency) allow smoother video adaptation at the application layer [33].

Multiuser Diversity at the Physical Layer — In [10] the application's video frames are encapsulated into LL timeframes called batches. On each batch a weight is assigned that indicates the video stream's priority. The MS creates a transmission queue for each batch of the video frame and assigns a timer with a timeout value to each batch. The BS knows the class number, remaining size, and timer value of each batch. At the link layer, batches are adapted to channel variations caused by channel fading. The multi-user diversity (the good/bad threshold) determines the batch's timer value and, thus, the probability that an LL frame is kept *idle* (i.e., the backoff probability). The faster the channel fades, the larger the backoff probability and (intuitively) the timer value.

Co-Channel Interference Controller in Multicell Systems — In OFDM systems a mobile user experiences interference from BSs using the same subcarrier. Hence, an optimization problem that should be solved in OFDM systems is the subcarrier allocation for each user. The work in [8] employs an optimization algorithm that considers the maximum achievable rate of the kth user on the jth subcarrier. In addition, a resource controller at the MAC layer determines the proper rate for each user based on channel quality feedback information such as the average SINR or MCS level. The SINR parameter is affected by the power received from the serving BS as well as that from its neighbor counterpart.

Furthermore, in multicell systems users served by the same subcarrier in adjacent cells (i.e., co-channel users) form a set. Reference [48] deems a set of users feasible if the BER at all receivers does not exceed a certain threshold. This feasibility depends on several factors, like the BS-user link gains, the modulation scheme that defines the SINR and BER capabilities, and the transmitting power from each BS, which are crucial for controlling the interference at the receivers. By imposing transmit power constraints at each BS and controlling the co-channel interference using a centralized controller, [48] achieves a large rate in each subcarier. One algorithm determines the allocation of users to subcarriers in an OFDM multicell environment considering both the interference caused by BSs to a new user that enters the network and the interference caused by the BS serving the new user to previous users that have already joined the network. Another algorithm assumes a user as preferable when the increase in the SINR of the other users in the same subcarrier is minimal. In the last algorithm a subcarrier is allocated to the user with the larger gain in the subcarrier; subsequently, the total power budget is allocated to subcarriers according to water filling, imposing a power constraint to approach the SINR for a set of subcarriers and remove subcarriers when the constraint is violated.

Multiuser Scheduler at the MAC Sublayer — In [30] a CLD solution proposes a scheduling policy for a TDMA system served by AMC and a finite-length queue of K packets per user at the physical and data link layers, respectively. For any particular user, the maximum number of packets that can be transmitted at time t depends on the AMC mode n and the number of time slots b reserved for one user. The key parameters of this CLD approach consist of the channel condition parameters such as Doppler spread f_d, SNR γ, and Nakagami parameter m as well as the resource management parameters such as time slots b, PER P_0, and queue size K. The objective of this CLD solution is the minimization of radio resources, b and K, while guaranteeing the prescribed QoS.

For OFDM systems, [8] provides scheduling for each user for different modes by exploiting information from the physical layer (i.e., channel matrix, SINR, MCS level, velocity, and location) as well as from the MAC layer (i.e., fairness and QoS in terms of packet delay and packet loss rate). The scheduler specifies the scheduling of users and the number of packets that can be scheduled in the current frame. Variable channel-aware scheduling is applied by using the SINR reported by the mobile terminals on the uplink control channels.

THE COEXISTENCE OF CLD SOLUTIONS

For different CLD solutions to harmoniously coexist, the commonalities between them must be identified, including the common sets of layer parameters. Although the complete identification of CLD commonalities is beyond the scope of this article, in the next paragraphs we discuss the most prominent ones.

For instance, notification-based CLD solutions employ the aforementioned mechanisms listed in Table 4. To identify

	Mechanism	Parameters
Application layer	User-driven notification	Application priority
Transport layer	Congestion and loss indication	TCP header checksum
		TCP header options
	Cwnd reduction	Congestion window
	Retransmissions	SN of corrupted packet
	Timeout reset or pause	RTT estimation
	Receiver window control	Receiver window
Network layer	Congestion and loss indication	ICMP message
		IP header options
Data link layer	Congestion recognition	RED (average queue length)
	Loss recognition	Agent (ACKs list block)
Physical layer	Congestion recognition	Load estimation (intra-cell interference, BS transmit power)

■ Table 4. *Mechanisms and parameters involved in notification-based CLD.*

congestion, a RED-like mechanism that relies on the average queue length at the link layer should be supported. On the other hand, to support notifications about packet losses, a link layer agent that retains an ACK list is an appropriate solution. At the network layer, ICMP messages and IP header options represent the out-of band and in-band signaling, respectively. At the transport layer, the TCP header options and other fields (e.g., checksum) may be used to indicate network congestion and packet loss. It is realized that upon receiving a notification about congestion or loss of a segment, the TCP sender will either reduce the *cwnd* or retransmit the segment, respectively. To avoid unnecessary retransmissions performed by the BS (local recovery) and the TCP sender as well, the latter should postpone or reset the RTO for the particular lost packet. Besides, in application-driven notification, the receiver window is calculated based on user preferences associated with each user's open sockets.

To summarize the discussion, a clearly open issue lies in properly choosing how to identify and signal congestion from the link layer in wireless networks. Physical layer characteristics (interference level, transmitting power budget, etc.) in infrastructure networks (i.e., CDMA) as well as in ad hoc networks (i.e., wireless LANs [WLANS]) can support this task, as mentioned previously. However, further investigation is required for an integrated end-to-end ECN solution in heterogeneous networks [42].

Apart from identifying some generic mechanisms used by notification-based CLD approaches, research should also investigate the coexistence of cross-layer optimization and scheduling approaches. One coexistence aspect is the deployment of several algorithms/techniques that rely on the same coupling between different layers [4]. For instance, solutions that combine the AMC at the physical layer and an ARQ-like protocol (e.g., type-I, type-II HARQ) at the data link layer consider the same parameters, such as RTT, N_p at the data link layer, and the mode n and SNR γ at the physical layer (Table 5). A cross-layer optimizer that uses these parameters as input can activate different algorithms at different time instants. This objective is undertaken by the co-channel interference controller operating in a multicell environment where the modulation scheme, interference level (SINR), transmit power, and BS-user link gain serve as input to different subcarier allocation algorithms (Table 5). With regard to the power control estimation at the physical layer, the received power strength measurement (SINR) is a crucial parameter for scheduling strategies at the MAC layer along with cardinal parameters such as the time slots b and queue size K at the data link layer (Table 5).

THE ROLE OF THE PHYSICAL LAYER

It is obvious that unique physical layer features, such as AMC, power control, subcarrier allocation, and multiuser diversity, feature prominently in CLD applications. These features depend on parameters like the transmitted power from the BS, the modulation and coding scheme, the BER level induced by additive Gaussian noise and co-channel, intracell, and intercarrier interferences according to the deployed multicell network, the BS-user link gain, as well as the velocity and location of the MS.

Similarly, while the estimation of SNR concerns the evaluation of modulated and coded wireless channels, the SINR is a good metric for evaluating channel strength in multiple access wireless systems characterized by fading channels where intercell interference must be added to channel noise. It has been argued [27, 28] that in cross-layer optimization and scheduling the SINR should be related to the PER in evaluating the effectiveness of a CLD proposal. Such mappings are important in simulating and evaluating the performance of a CLD proposal at the physical layer [8, 46].

However, SINR-PER mapping is not an easy task [46] as the calculation of additive interference is affected by the actual network topology [2], and bit errors are correlated [27]. Consequently, when topologies are different, the only parameter required to evaluate a CLD is the SINR-PER mapping given that the SINR is already well defined. Further investigation of this mapping is a key aspect of cross-layer optimization and scheduling approaches for PHY-MAC co-design and coordination.

On the other hand, spatial processing via multiple-input multiple-output (MIMO) techniques is an advanced signal processing technique that alleviates cross-layer scheduling at the data link layer [49]. Reference [50] proposes a solution based on MIMO and AMC at the physical layer to support QoS guarantees at the data link layer in terms of effective bandwidth and effective capacity. Such cross-layer modeling that capitalizes on spatial diversity and spatial multiplexing of MIMO systems is also an area for further investigation.

DISCUSSION AND LESSONS LEARNED

EVALUATION CRITERIA FOR CLD MODELS

Besides the need for cross-layer architecture, the relation between such architecture and the resulting performance is also important. Cross-layer design architectures can lead to considerable improvements in throughput and delay performance. To this end, the following evaluation criteria should be considered [3]:

- **Unintended interactions**: A CLD architecture must figure

	Mechanism	Parameters
Application layer	Video reconstruction and adaptation	Source distortion, expected loss distortion (packet loss rate P) Video source-coding R_s (quantization step-size)
Data link layer	Error correction	Link layer agent (PER, RTT)
	Queuing control with RED	Length of K packets, queuing delay
	Retransmissions	Number of Retransmissions N_r^{max}, average RTT, packet length N_p
	MAC layer scheduler	Time slots b, queue of K packets per user, MCS level (SINR)
Physical layer	Adaptive modulation and coding	Mode n, SNR γ, Rate code
	Power control	Received power strength measurement (SINR)
	Multiuser diversity	Good/Bad threshold (channel fading)
	Co-channel interference controller	Subcarier rate (BS-user gain, transmit power, SINR)

Table 5. *Mechanisms and parameters involved CLD optimization and scheduling.*

out the effect of interactions caused on a separate part or layer of the protocol stack. Cross-layer architectures usually do not take into account the impact of the interactions they introduce between nonadjacent layers.

- **Dependency**: Many different but possibly interrelated parameters are involved in CLD architectures. The relation between these parameters should be considered using a dependency graph.
- **Stability**: When a parameter in the dependency graph is controlled by two different loops implemented on the same protocol, the stability of the entire scheme must be carefully evaluated.

Furthermore, some kind of cost-benefit analysis for CLD architectures that takes into account the following costs will be necessary:

- The evaluation of an objective function with a large set of variables may introduce excessive delays.
- The abstraction process for choosing the states and capabilities of different layers also introduces some communication overhead.
- A cross-layer architecture that does not provide a good level of modularity is more difficult to manage and optimize.

LESSON LEARNED

Cross-Layer Architectures, Models, and Entities — Evidently, the modularity of the overall architecture is a critical issue in system design. An unwisely designed architecture will suffer from excessive delay as a result of suboptimal interactions between its modules. For example, [7] established a high level of modularity by specifying control interfaces for each module in the architecture; however, the additional functional or procedure calls incur internal overhead in the protocol stack. Cross-layer designers should take these unintended interactions into account. To this end, designers should validate their architecture to identify code deadlocks (i.e., parallel control loops) that may seriously affect system stability. The abstraction process may also introduce processing and communication overheads [6]. An abstraction of a particular layer can be represented by a small set of selected parameters that clearly expose the associated layer's capabilities. Obviously, the overhead associated with the exchange and processing of these parameters will be proportional to their cardinality [10]. In addition, the translation of abstracted parameters into layer-specific parameters and actual modes of operation (and vice versa) could introduce significant signaling overhead if the set of selected parameters is excessively large. At protocol layers residing below the network layer, the unnecessary transmission of control messages due to the retransmission of packets at the data link layer may also waste precious wireless channel resources [8]. In such a case the data link layer's efficiency is degraded, and the achievable throughput drops substantially. Apparently, when a scheduler refrains from collecting channel state information (i.e., no SNR feedback), no external overhead results [9].

Cross-Layer Signaling — Thus far, we have elaborated on cross-layer signaling mechanisms between the mobile clients and the radio access network, and considered the way these mechanisms are realized in an end-to-end networking context. For such mechanisms, customarily one can adopt one of the four available options (Fig. 3). It is evident that an important aspect of CLD is a way to indicate cross-layer signaling between peers, for which two primary alternatives are available: reusing a simple option (or options) in the current header format of the employed transport protocol or using an additional header. In the first case, the options in the current header are considered mostly as an in-band signaling method (e.g., ECN, ELN, EBSN). In the latter case, one must also consider that an additional header will often require a separate out-of-band signaling protocol (i.e., ICMP, RTCP, RSVP). On the other hand, local profiles and a network service convey valuable CLI for local or application layer adaptation, respectively. Both these cases will undoubtedly introduce external overhead when the session starts since all the LIDs and XML descriptors must be transmitted for the first time. However, once a session starts, the offered QoS may be affected by the need to appropriately manipulate packet headers and the size of header itself as well as the transformation of the XML files. To summarize, the CLI can be conveyed by either XML files or an additional packet header, and the way of indicating CLI could be assigned to a header option (i.e., label, ECN).

Therefore, it is evident that the choice of cross-layer indications or the way of passing CLI may differ according to application requirements. To this end, there is work in progress within IETF that elaborates on the different types of cross-layer signaling and notification [35]. In addition, [51] argues for an in-band mechanism when signaling QoS requirements. In most cases, in-band signaling does not introduce a new header but uses selected options in the current header such as ECN, which is a well-known in-band solution [35]. Finally, [36] outlines several aspects on cross-layer communication to help CLD practitioners improve protocol performance.

CONCLUDING REMARKS

As the evolution of wireless communication technologies continues, cross-layer approaches are being increasingly studied as a design approach for confronting unintended performance degradations as well as supporting new processes for beyond 3G mobile communication networks. Not surprisingly, over the last decade cross-layer design has developed into a hot research topic that is still growing. Currently, several research efforts are seeking ways to integrate cross-layer design solutions into wireless communication standards for purposes of allocating resources to mobile users, scheduling access to shared resources with higher throughput, and achieving better quality of service for multimedia applications. Many of these solutions employ functional entities that support cross-layer processes for mobile devices. We envisage an evolution toward 4G networks based on these kinds of functional entities, protocols, and processes that provide optimization and reconfiguration at both the terminal and the network.

REFERENCES

[1] H. Zimmermann, "OSI Reference Model — The ISO Model of Architecture for Open System Interconnection," *IEEE J. Trans. Commun.*, vol. 28, no. 4, Apr. 1980.

[2] L. Georgiadis, M. J. Neely and L. Tassiulas, *Resource Allocation and Cross-Layer Control in Wireless Networks*, Now Publishers Inc., Apr. 24, 2006.

[3] V. Kawadia and P.R . Kumar, "A Cautionary Perspective on Cross-Layer Design" *IEEE Wireless Commun.*, vol. 12, no. 1, Feb. 2005, pp. 3–11.

[4] V. Srivastava and M. Motani, "Cross-Layer Design: A Survey and The Road Ahead," *IEEE Commun. Mag.*, vol. 43, no. 12, Dec. 2005, pp. 112–19.

[5] G. Carneiro, J. Ruela, and M. Ricardo, "Cross-Layer Design in 4G Wireless Terminals," *IEEE Wireless Commun.*, vol. 11, no. 2, Apr. 2004, pp. 7–13.

[6] S. Khan *et al.*, "Application-Driven Cross-Layer Optimization for Video Streaming Over Wireless Networks," *IEEE Commun. Mag.*, vol. 44, no. 1, 2006, pp. 122–30.

[7] V.T. Raisinghani and S. Iyer, "Cross-Layer Feedback Architecture for Mobile Device Protocol Stacks," *IEEE Commun. Mag.*, special issue on Cross-Layer Protocol Engineering, Jan 2006.

[8] T. Kwon *et al.*, "Design and Implementation of a Simulator Based on a Cross-layer Protocol between MAC and PHY Layers in a WiBro Compatible IEEE 802.16e OFDMA System," *IEEE Commun. Mag.*, Dec. 2005, pp. 136–46.

[9] R. Ferrús *et al.*, "Cross-Layer Scheduling Strategy for UMTS Downlink Enhancement," *IEEE Commun. Mag.*, vol. 43, no. 6, June 2005, pp. s24–28.

[10] H. Jiang, W. Zhuang, and X. Shen, "Cross-Layer Design for Resource Allocation in 3G Wireless Networks and Beyond," *IEEE Commun. Mag.*, Dec. 2005.

[11] X. Lin, N. B. Shroff, and R. Srikant, "A Tutorial on Cross-Layer Optimization in Wireless Networks," *IEEE JSAC*, Special Issue on Non-Linear Optimization of Communication Systems, 2006.

[12] M. Miyoshi, M. Sugano, and M. Murata, "Improving TCP Performance for Wireless Cellular Networks by Adaptive FEC Combined with Explicit Loss Notification," *IEICE Trans. Commun.*, vol. E85-B, no. 10, Oct. 2002.

[13] S. Shakkottai, T.S. Rappaport, and P.C. Karlsson, "Cross-Layer Design for Wireless Networks," *IEEE Commun. Mag.*, vol. 41, no. 10, Oct. 2003, pp. 74–80.

[14] H. Zheng and H. Viswanathan, "Optimizing TCP Performance with Hybrid ARQ and Scheduler," *Proc. 14th Int'l. Symp. Pers., Indoor, Mobile Radio Commun.*, vol. 2, Sep., 2003, pp. 1785–89.

[15] K. Ramakrishnan, S. Floyd, and D. Black "The Addition of Explicit Congestion Notification (ECN) to IP," RFC 3168, Sept. 2001.

[16] H. Balakrishnan *et al.*, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links," *IEEE ACM Trans. Networking*, Dec. 1997.

[17] D. Kliazovich and F. Graneill, "A Cross-Layer Scheme for TCP Performance Improvement in Wireless LANs," *IEEE GLOBECOM*, vol. 2, no. 2, Dec. 2004, pp. 840–44.

[18] T. Goff *et al.*, "Freeze-TCP: A True End-To-End TCP Enhancement Mechanism For Mobile Environments," *Proc. 19th INFOCOM*, Sept. 2000, pp. 1537–45.

[19] H. Balakrishnan and R. H. Katz, "Explicit Loss Notification and Wireless Web Performance," *Proc. GLOBECOM Internet Mini-Conf.*, Nov. 1998.

[20] P. Bender *et al.*, QUALCOMM, Inc., "CDMA/HDR: A Bandwidth Efficient High-Speed Wireless Data Service for Nomadic Users," *IEEE Commun. Mag.*, July 2000.

[21] H. Jiang and W. Zhuang, "Cross-Layer Resource Allocation for Integrated Voice/Data Traffic in Wireless Cellular Networks" *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, Feb. 2006, pp. 457–68.

[22] Goldsmith and S. G. Chua, "Adaptive Coded Modulation for Fading Channels," *IEEE Trans. Commun.*, vol. 46, no. 5, May 1998, pp. 595–602.

[23] IEEE 802.16 Working Group, "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," 2004.

[24] 3GPP TS 25.308 V6.4.0, "High Speed Downlink Packet Access (HSDPA); Overall Description; Stage 2," rel. 6, 2007.

[25] M. van der Schaar and S. Shankar, "Cross-Layer Wireless Multimedia Transmission: Challenges, Principles, and New Paradigms," *IEEE Wireless Commun.*, vol. 12, no. 4, Aug. 2005, pp. 50–58.

[26] Q. Liu, S. Zhou, and G. B. Giannakis, "TCP Performance in Wireless Access With Adaptive Modulation and Coding," *Proc. IEEE ICC*, June 20–24, 2004.

[27] Q. Liu, S. Zhou, and G.B. Giannakis, "Cross-Layer Combining of Adaptive Modulation and Coding With Truncated ARQ Over Wireless Links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, Sept. 2004.

[28] D. Wu and S. Ci, "Cross-Layer Design for Combining Adaptive Modulation and Coding with Hybrid ARQ," *Proc. Int'l. Conf. Commun. and Mobile Comp.*, July 3–6, 2006, pp. 147–52.

[29] D. Wu and S. Ci , "Cross-Layer Combination of Hybrid ARQ and Adaptive Modulation and Coding for QoS Provisioning in Wireless Data Networks," *Proc. 3rd Int'l Conf. Quality of Service in Heterogeneous Wired/Wireless Networks*, Aug. 7–9, 2006, no. 47.

[30] Q. Liu, S. Zhou, and G.B. Giannakis, "Cross-Layer Scheduling with Prescribed QoS Guarantees in Adaptive Wireless Networks," *IEEE JSAC*, vol. 23, no. 5, May 2005.

[31] S. Mérigeault and C.Lamy, " Concepts for Exchanging Extra Information Between Protocol Layers Transparently for the Standard Protocol Stack," *Proc. IEEE ICT '03*, Feb. 23-Mar. 1, 2003.

[32] Y. S. Chan and J. W. Modestino, "A Joint Source Coding-Power Control Approach for Video Transmission over CDMA Networks," *IEEE JSAC*, vol. 21, no. 10, Dec. 2003.

[33] I. Haratcherev *et al.*, "Optimized Video-Streaming Over 802.11 by Cross-Layer Signaling," *IEEE Commun. Mag.* Special Issue on Cross-layer Protocol Engineering for Wireless Mobile Networks, vol. 44, no. 1, Jan. 2006.

[34] Q. Wang and M.A. Abu-Rgheff, "Cross-Layer Signalling for Next-Generation Wireless Systems," *IEEE Wireless Commun. and Networking Conf.*, vol. 2, no. 16–20 Mar. 2003, pp. 1084–89.

[35] P. Sarolahti and S. Floyd, "Cross-Layer Indications for Transport Protocols," Internet draft, IETF, Mar. 2007. work in progress.

[36] P. Sarolahti, S. Floyd, and M. Kojo, "Transport-Layer Considerations for Explicit Cross-Layer Indications," Internet draft, IETF, Oct. 2006, informational.

[37] P. Sudame and B. R. Badrinath, "On Providing Support for

Protocol Adaptation in Mobile Wireless Networks," *Mobile Networks and Apps.*, vol. 6, no. 1, Jan.–Feb. 2001, pp. 43–55.
[38] A. Takács *et al.*, "The Layer-Independent Descriptor Concept," *J. Computers*, vol. 1, no. 2, May 2006, pp. 23–32.
[39] B. Kim, "A Network Service Providing Wireless Channel Information for Adaptive Mobile Applications: Part I: Proposal," *Proc. IEEE ICC '01*, June 2001.
[40] J. Case *et al.*, "A Simple Network Management Protocol," RFC 1157, May 1, 1990.
[41] L. Larzon, U. Bodin, and O. Schelen, "Hints and Notifications," *Proc. Wireless Commun. and Networking Conf.*, Mar. 2002.
[42] V. A. Siris, "Resource Control for Elastic Traffic in CDMA Networks," *Proc. ACM MOBICOM*, Sept. 2002.
[43] B. S. Bakshi *et al.*, "Improving Performance of TCP over Wireless Networks," *Proc. IEEE ICDCS '97*, 365–73.
[44] S. Goel and D. Sanghi, "Improving TCP Performance over Wireless Links," *Proc. TENCON '98*, Dec. 1998, pp. 332–35.
[45] B. Zhang and M.N. Shirazi, "Implementation of Explicit Wireless Loss Notification Using MAC-Layer Information," *Proc. IEEE Wireless Commun. and Networking Conf.*, vol. 2, Mar. 2003, pp.1339–43.
[46] B. Liu, D. L. Goeckel, and D. Towsley. TCP-Cognizant Adaptive Forward Error Correction in Wireless Networks," *Proc. IEEE GLOBECOM*, Nov. 2002.
[47] M. Saied and A. El-Atty, "Efficient Packet Scheduling with Pre-Defined QoS Using Cross-Layer Technique in Wireless Networks," *Proc. IEEE ISCCC '06*.
[48] I. Koutsopoulos and L. Tassiulas, "Cross-Layer Adaptive Techniques for Throughput Enhancement in Wireless OFDM-Based Networks," *IEEE/ACM Trans. Net.*, vol. 14, no. 5, Oct. 2006.
[49] C. Antón-Haro *et al.*, "Cross-Layer Scheduling for Multi-user MIMO Systems," *IEEE Commun. Mag.*, vol. 44, no. 9, Sept. 2006, pp. 39–45.
[50] J. Tang and X. Zhang, "Cross-Layer-Based Modeling for Quality of Service Guarantees in Mobile Wireless Networks," *IEEE Commun. Mag.*, vol. 44, no. 1, Jan. 2006, pp. 100–06.
[51] J. Harper and P. McGeer, "Requirements for In-Band QoS Signalling," Internet draft, Network Working Group, Jan. 2007, work in progress.

BIOGRAPHIES

FOTIS FOUKALAS (foukalas@di.uoa.gr) received a Dipl.-Ing. degree in electrical and computer engineering from Aristotle University of Thessalonica, Greece, in 2001. In 2003 he received an M.Sc. degree from the National Technical University of Athens. In May 2003 he joined the Communication Networks Laboratory at the Department of Informatics and Telecommunications of the University of Athens. Since December 2003 he has been working on his Ph.D. in the same department. His research interests are in the area of physical and data link layer techniques enabling cross-layer optimization and reconfigurability.

VANGELIS GAZIS (gazis@di.uoa.gr) holds B.Sc., M.Sc., and Ph.D. degrees from the Department of Informatics and Telecommunications at the University of Athens, and an M.B.A. degree from the Athens University of Economics and Business). He is a senior researcher at the Department of Informatics and Telecommunications ,and has participated in national (OTE-DECT, GUNet, GEANT-2) and European projects (MOBIVAS, ANWIRE). He specializes in reconfigurable mobile systems and protocol stacks for beyond 3G mobile, ontology languages for autonomic systems, reflective and component middleware, adaptable services, and open API frameworks for telecommunications.

NANCY ALONISTIOTI (nancy@di.uoa.gr) has B.Sc. and Ph.D. degrees in informatics and telecommunications from the University of Athens. She specializes in reconfigurable systems and networks for beyond 3G, adaptable services, pervasive computing, and context awareness. She has participated in national and European projects, (CTS, SS#7, ACTS RAINBOW, EURESCOM, MOBIVAS, ANWIRE, E2R, LIAISON) and is co-editor of *Software Defined Radio, Architectures, Systems and Functions* (Wiley Series on Software Radio).