6

Cross-Layer Design and Optimization in Wireless Networks

Vineet Srivastava
HelloSoft, Inc., Hyderabad, Andhra Pradesh, India

Mehul Motani
National University of Singapore, Singapore

6.1 Introduction

Communication networks have traditionally been engineered following the principle of protocol layering. This means designing specific network functionalities (such as flow control, routing, medium access) in isolation from each other, and putting together the complete system through limited interfaces between the *layers* performing these specific tasks. The layers, which are in fact distributed systems with collaborating entities distributed through the network [6, p. 20], are arranged in a vertical hierarchy. Each layer makes use of the services provided by the layers below itself and, in turn, makes its services available to the layers above itself. Inter-layer communication happens only between adjacent layers and is limited to procedure calls and responses. Familiar examples of layered communication architectures are the seven-layer Open Systems Interconnect (OSI) model [6, p. 20] and the four-layer TCP/IP model [39].

Even though layered communication architectures have been instrumental in the development of many communication networks, most notably the Internet, in recent times, the suitability of protocol layering is being questioned in the research community. This is largely due to wireless networks becoming an integral part of the communications

Cognitive Networks: Towards Self-Aware Networks Edited by Qusay H. Mahmoud
© 2007 John Wiley & Sons, Ltd. ISBN: 978-0-470-06196-1

infrastructure, and hence occupying the center stage of research and development activity. Researchers argue that while designing protocols at the different layers in isolation and running them without much interaction between the functionalities may have served well in the case of wired networks, doing so is not suitable for wireless networks due to the peculiarities of the wireless medium. To illustrate their point, researchers usually present a *cross-layer design* proposal. Broadly speaking, cross-layer design refers to protocol design done by actively exploiting the dependence between the protocol layers to obtain performance gains. This is unlike layering, where the protocols at the different layers are designed independently. There have been a large number of cross-layer design proposals in the literature; references [11], [48] and [54], among others, consolidate some of these proposals; reference [6] highlights the potential problems that cross-layer design, if done without appropriate care, can create.

While the presence of wireless links in a network makes strict layering of protocols look like an inefficient solution, the pursuit of cognitive networks makes it look grossly inadequate. This is because a cognitive network not only needs to provide network functionalities to the applications (and hence the users), it also needs to adapt to the application needs, and to the prevailing network conditions (for example to the traffic in the network, to the conditions of the underlying channel, or to the interference in the spectrum bands of operation) [7]. Such adaptability requires rich coordination between the traditional protocol layers, which is not feasible in the constraints of the layered architectures. Thus, when it comes to cognitive networks, cross-layer design is not just a matter of efficiency, but in fact it is central to the very concept of the network. The question thus is not 'whether', but rather 'how', cross-layer design and optimization should be done. For example, which layers should be coupled and how? What should be the roles of the different layers? On a broader note, should there be layers at all or should cognitive networks be built around fundamentally different architectures? As a matter of fact, such questions may not have a single answer given the multiplicity of the communication networks that we have around us, each with their own specific needs and purposes. (See also the concluding remarks in [19].)

The case for cross-layer design from a performance viewpoint, as articulated above, is easy to understand. Unfortunately, performance improvement brought about by cross-layer design is only one part of the picture. To be successful, any cross-layer design idea needs to be implemented and deployed in real systems. This means that the architectural considerations are equally, if not more, important. The success of layering is really due to its architectural merit: that different portions of the network can be developed and innovated upon by different vendors, and still be stitched together to yield a working system. With couplings introduced between the layers, as done with cross-layer design, the desirable architectural qualities of layering may be lost. Thus, a big challenge facing the proponents of cross-layer design is to evaluate the architectural consequences of their work, and to ascertain that their cross-layer design ideas are not eroding away the architectural merits of the layered architectures [34]. Architectural considerations involve questions like the coexistence of different cross-layer design ideas; determining which cross-layer design ideas, if deployed, may hamper further innovation; and so on. Admittedly, such questions are easier stated than answered or even addressed. Nonetheless, these questions have to be tackled head on for the promise of cross-layer design to be realized. Facilitating a healthy debate on such questions is our goal in this chapter. We do so by consolidating the different results, ideas and perspectives pertinent to the area of cross-layer design.

We start by looking at the different definitions of cross-layer design and capture them in a concise and general definition. Next, we look at the broad motivations for cross-layer design purely from a performance viewpoint and, in doing so, we present a quick survey of the research literature in this area. From a performance viewpoint, the motivations for cross-layer design are the unique problems created by the wireless networks, the avenue for opportunistic communications, and the new modalities of communications created by wireless links. Moving on, we categorize the different cross-layer design proposals from an architectural viewpoint, and look at the initial ideas for implementing cross-layer design proposals. Next, we take a brief look at some cross-layer design ideas that have found their way into commercial systems or ongoing standards. Having done the consolidation, we raise what we call the open challenges, which are some important problems researchers may want to address as they move forward. Next, we place the ongoing cross-layer design activity in a brief historic context after which we present our conclusions.

Even though the questions surrounding cross-layer design are multifaceted, we believe understanding cross-layer design at a broad level and not seeing it like a quick fix for specific problems created by wireless networks will help. Our aim in this chapter is to facilitate such holistic thinking incorporating the performance viewpoint, the architectural considerations as well as the future potential modalities of wireless communications about this extremely important and promising design paradigm.

6.2 Understanding Cross-Layer Design

6.2.1 The Different Interpretations

Cross-layer design is described in the research literature in different ways [52], [27]: as a protocol design methodology that exploits the synergy between the layers; as a joint design and optimization across more than one protocol layer; as a protocol design methodology that relies on, and creates, adaptability across the different layers by sharing information between them; and so on. An interpretation of cross-layer design, usually discussed in the context of multimedia transmissions, is that of a design methodology that leverages joint source and channel coding [67] which creates coupling between the application and the physical layers. Yet another interpretation of cross-layer design, usually discussed in the context of communications over fading channels, is that of a design methodology that takes into account the channel characteristics at the higher layers, and the stochastic arrival of traffic at the lower layers [5].

6.2.2 A Definition for Cross-Layer Design

In what follows, we present a definition for cross-layer design by viewing it as a violation of a reference layered architecture. Our definition is concise and yet it encompasses the aforementioned notions about cross-layer design. Our definition also draws a clearer contrast between layered protocol design and cross-layer design, regardless of the layered architecture in question.

As noted earlier a layered architecture, like the seven-layer Open Systems Interconnect (OSI) model [6, p. 20], divides the overall networking task into layers and defines a hierarchy of services to be provided by the individual layers. The services at the layers are realized by designing protocols for the different layers. The architecture forbids direct communication between non-adjacent layers; communication between adjacent layers is limited to procedure calls and responses.

In the framework of a reference layered architecture, the designer has two choices at the time of protocol design. Protocols can be designed by respecting the rules of the reference architecture. In the case of a layered architecture, this would mean designing protocols such that a higher layer protocol only makes use of the services at the lower layers and not be concerned about the details of how the service is being provided. Similarly, the services provided by a lower layer would not depend on the details of the higher layer (e.g., the application) that is requesting or using them. Following the architecture also implies that the protocols would not need any interfaces that are not present in the reference architecture.

Alternatively, protocols can be designed by violating the reference architecture, for example, by allowing direct communication between protocols at non-adjacent layers or by sharing variables between layers. Such violation of a layered architecture is cross-layer design with respect to the reference architecture.

- **Definition 6.1:** Protocol design by the violation of a reference layered architecture is cross-layer design with respect to the particular layered architecture.
- **Comment 6.1:** Examples of violation of a layered architecture include creating new interfaces between layers, redefining the layer boundaries, designing protocol at a layer based on the details of how another layer is designed, joint tuning of parameters across layers, and so on.
- **Comment 6.2:** Violation of a layered architecture involves giving up the luxury of designing protocols at the different layers independently. Protocols so designed may impose some conditions on the processing at the other layer(s).
- **Comment 6.3:** Cross-layer design is defined as a protocol design methodology. However, a protocol designed with this methodology is also termed as cross-layer design.

For exposition, consider a hypothetical three-layer model with the layers denoted by L_1, L_2 and $L_3.L_1$ is the lowest layer and L_3 the highest. Note that in such an architecture, there is no interface between L_3 and L_1. One could, however, design an L_3 protocol that needs L_1 to pass a parameter to L_3 at run-time. This calls for a new interface, and hence violates the architecture. Alternatively, one could view L_2 and L_1 as a single layer, and design a joint protocol for this 'super-layer.' Or, one could design the protocol at L_3, keeping in mind the processing being done at L_1, again giving up the luxury of designing the protocols at the different layers independently. All these are examples of cross-layer design with respect to the three-layer architecture in question.

Architecture violations, like those introduced by cross-layer design, clearly undermine the significance of the architecture since the architecture no longer represents the actual system. If many architecture violations accumulate over time, the original architecture can completely lose its meaning. Architecture violations can have detrimental impact on the system longevity, as has been argued for the case of cross-layer design in [34].

6.3 General Motivations for Cross-Layer Design

What motivates designers of wireless networks to violate the layered communication architectures? There are a three broad motivations. Several cross-layer design proposals aim to solve some unique problems created by wireless links. An example of such a problem is the classic case of a TCP sender mistaking a wireless error to be an indication of network congestion [52]. Another category of cross-layer design ideas aim to exploit the fundamental characteristics of the wireless medium opportunistically, for example by utilizing channel variations from fading at the higher layers. This is in line with the general goal of an adaptive protocol stack that responds dynamically to the changes in the network conditions. Yet another category of cross-layer design ideas make use of the new modalities of communications that wireless medium creates and that cannot be accomodated within the constraints of layering. An example here is node cooperation, as we discuss later. Cross-layer design with all three motivations are important in the context of cognitive networks of the future.

In this section, we elaborate further on the motivations for cross-layer design. We draw examples from the published research literature. Apart from clarifying the main motivations for cross-layer design, taking relevant examples from the literature also gives a good measure of the wide range of scenarios in which cross-layer design has been applied.

Since there is no overarching layered architecture that is followed in *all* communication systems, the reference layered architecture we assume is a five-layer hybrid reference model presented in [59, p, 44]. This model has the application layer, the transport layer, the network layer, the link layer which comprises the data link control (DLC) and medium access control (MAC) sublayers [6, p. 24], and the physical layer; we assume that all the layers perform their generally understood functionalities.

6.3.1 Problems Created by Wireless Links

6.3.1.1 TCP on Wireless Links

The first motivation for cross-layer design is that wireless links create some new problems for protocol design that cannot be handled well in the framework of the layered architectures. A transmission control protocol (TCP) sender erroneously mistaking a packet error on a wireless link to be an indicator of network congestion is an example [52]. This problem is often resolved by direct communication between the link layer and the transport layer, which is thus an example of cross-layer design motivated by a unique problem created by a wireless link.

6.3.1.2 Real-Time Multimedia on Wireless Links

Likewise, when an error-prone and shared wireless link is used for real-time multimedia communications, the error correction and resource management techniques employed at the lower layers of the stack may need to be adapted jointly with the compression and streaming algorithms applied at the higher layers. In fact, even on wired networks, rich coordination between the layers is needed to handle the real-time nature of multimedia transmission [69]. The time-varying nature of the wireless links, their relatively higher error probabilities and the delays caused in negotiating multiple access over the shared wireless medium further complicate matters and require a cross-layer design approach [51]. Here is thus another example where the problems created by wireless links necessitate violating the layered architectures. A good survey of cross-layer design ideas in this context is presented in [11]. We also refer the readers to [67] which explores the topic of joint source and channel coding as a special case of cross-layer design for the real-time multimedia communications problem, as we discussed earlier.

6.3.1.3 Power Control in Ad Hoc Networks

On a broader note, the wireless medium often creates problems in wireless networks that affect all layers of the traditional layered architectures at once. Some examples are power control in wireless ad hoc networks, energy management in wireless devices and security in wireless networks.

The problem of power control comes up because in a wireless ad hoc network, the transmissions of the different nodes interfere with each other unless they have been assigned orthogonal resources. Thus, even if all the nodes employ a large transmission power, they may not be able to communicate with one another because large transmission powers also lead to higher interference on the other nodes. Thus power control is needed.

Power control clearly influences the network topology, which is a concern of the network layer. It also impacts how much spatial reuse can be achieved, that is, how far apart can two ongoing communication sessions be without interfering with each other, which is a concern of the MAC layer. Power control is also linked to the processing at the physical layer, because the signal processing at the physical layer determines how stringent the requirements on the power control need to be. All these factors determine the end-to-end throughput. Furthermore, the transmitted power(s) determines the lifetime of the nodes (and the network) which one would want to maximize. Hence, the problem of power control cannot possibly be handled at any one layer in isolation, as is done while designing protocols in the framework of the layered architectures. It is thus a problem that, by its very nature, requires cross-layer design. It is no surprise then that a number of cross-layer design proposals in the literature have looked at power control in a cross-layer design framework. We shall discuss some of these ideas in Section 6.4.1.3.

6.3.1.4 Energy Efficiency and Security

Much like power control, energy efficiency of a communication device and network security (see [47] for example) are multifaceted issues that cannot be meaningfully addressed at any one layer of the protocol stack in isolation. The problem of energy-efficient design of wireless ad hoc networks is discussed thoroughly in [27] which also makes a case for cross-layer design. Reference [23] looks at the problem of energy management in wireless communications systems, and presents a thorough discussion on the cross-layer interfaces required for a rich coordination between the layers.

6.3.1.5 Vertical Handovers

Next, consider the problem of vertical handover in wireless networks. Vertical handover refers to a seamless transition of a multimode device from one network interface to another. For example, a person accessing the Internet on the road using the cellular mobile phone network might move into a building served by a wireless LAN. Ideally, from a user's perspective, such a change should be seamless and ongoing connections and data transfers should not be interrupted. To be fair, this problem of guaranteeing seamlessness as described above, though extremely important for mobile wireless networks, is not unique to wireless networks. As [31, p. 14] points out, a person who is accessing the Internet through a cable might also request for the same kind of seamlessness when changing the mode of access. However, we mention it here because the presence of wireless links creates avenues for taking into account cross-layer interactions when conducting vertical handovers, as discussed in detail in [57]. Thus, though not unique to wireless networks, this is another problem that touches all the layers of the protocol stack at once. Vertical handover is an integral component of the cognitive network because through effective vertical handovers the user can be kept connected to the network in the best possible mode at all times. In fact, vertical handovers are likely to have an immediate practical and commercial relevance as telecom service providers and equipment manufacturers worldwide are gearing towards converged networks and services.

6.3.2 The Optimistic Side

6.3.2.1 Fading Channels: Single User

Let us now look at a more optimistic side. It is well known that due to fading and multipath effects, wireless links can show tremendous time and/or frequency selectivity.

The variation in the channel quality creates new opportunities at the higher layers. As [4] puts it, fading allows the physical channel to be viewed as a 'packet pipe' whose delay, rate and/or error probability characteristics vary with time. Contrast this with a wired communication channel whose characteristics remain largely time-invariant. Reference [4] considers a buffered single user point-to-point communication system and proposes a rate and power adaptation policy based on the fluctuations of the channel and the buffer occupancy. Reference [63] considers a similar situation and also comes up with an optimal adaptive policy that minimizes a linear combination of the transmission power and the buffer overflow probability. Such adaptations are examples of cross-layer design motivated by the time-varying nature of the wireless channels.

6.3.2.2 Fading Channels: Multiple Users

In fact, the time variation in the quality of the wireless medium creates even more interesting opportunities in the case of multiuser networks. Consider the problem of downlink scheduling on a cellular network [52]. The situation is as follows: there are a number of users whose data is arriving in a stochastic fashion at the base station. The base station needs to schedule the transmissions of the different users. Since the network traffic is bursty, a fixed assignment of slots/frequency bands to the users is not efficient. Instead, the allocation of the channel to the different users should be done dynamically. In doing the scheduling, significant performance gains can be made if the base station takes into account the state of the downlink channel, and allows transmission to a particular user only if the channel between the base station and that particular user is in a 'good' state. This makes intuitive sense, since, when the channel for a particular user is bad, there is no point in scheduling the transmission for that user. For theoretical bases of such channel-dependent scheduling algorithms, we refer the readers to the results in [61], [62] and the references therein. The problem of resource allocation over single-user and multiuser (e.g., the multiple access and the broadcast) fading channels is also discussed in [5]. Also noteworthy in this context is the work presented in [68] that deals with the problem of transmitting bursty traffic over fading channels, and illustrates how channel variation can be 'exploited (rather than combatted)' to improve system performance. In an ad hoc network setting too, the time variation in the channels between the different users can be taken into account at the MAC layer to make throughput gains. (See [32] and the references therein.)

In short, fading, being unique to wireless links, does create new challenges and problems. But on the optimistic side, it also creates new opportunities. Generally speaking, layered architectures, with their fixed interfaces, appear too stiff to meaningfully address the time variations in the channel caused by fading. Making use of these opportunities for protocol design motivates cross-layer design, as evident from the several works that we have cited above.

6.3.3 The New Modalities

6.3.3.1 Multi-Packet Reception

Apart from the pessimistic and the optimistic views presented above, the wireless medium offers some new modalities of communication that the layered architectures do not accommodate. For instance, the physical layer can be made capable of receiving multiple packets [60] at the same time. This clearly changes the traditional balance of roles between the MAC layer and the physical layer, and the functionalities of the two layers can then be designed jointly, as illustrated in [60].

6.3.3.2 Node Cooperation

As another example, consider node cooperation. Nodes in a wireless network can cooperate with each other in involved ways. Node cooperation is actually a fairly broad concept. It covers a lot of ground, ranging from the nodes with single antenna each cooperating to create distributed coding systems [45] or to obtain the diversity gains provided by multiple-input multiple-output (MIMO) channels [37]; to cooperative beam forming where once again nodes equipped with single antennas cooperate to form antenna arrays; to relaying and mesh networking; and so on. Broadly, the motivation for node cooperation comes from the fundamental problems of wireless communications, such as the scarcity of spectrum, the energy limitations at the wireless devices and inability to accommodate multiple antennas at the same terminal. Cooperation may be imperative to increase the system capacity, the spectrum usage and energy efficiency for wireless communications devices [26]. Node cooperation is a very active field of research, with interest being shown from both the academia and the industry. We refer the readers to [19] and the several other articles in the same volume for a thorough and excellent introduction to the ongoing research in the area of node cooperation for wireless networks.

As far as this chapter is concerned, our interest in the actively researched area of node cooperation comes from the fact that it is a new modality of communications that wireless networks create. And because cooperation can fundamentally change the network model of a collection of point-to-point links that the traditional layered architectures have assumed (and that made sense for the wired networks), it is clear that incorporating node cooperation inevitably requires violating the layered architectures, and hence cross-layer design. See for example the discussion in [19] and also the discussion in [54], where the latter discusses in detail on what kind of architecture violations may be needed for accommodating coded cooperation in wireless ad hoc networks. Reference [70] presents a node cooperation scheme that generalizes the idea of hybrid automatic repeat request (ARQ) to multiple users; the exposition in [70] vividly illustrates the cross-layer interactions that cooperative communications necessitate.

6.3.3.3 Cognitive Radios

Another new modality that the wireless medium is ushering in is the idea of cognitive radios [29]. Cognitive radios can be understood as radios that gain awareness about their surroundings and adapt their behavior accordingly. For instance, a cognitive radio may determine an unused frequency band and use that for a transmission, before jumping to another unused band. Cognitive radios hold tremendous potential to increase the efficiency of wireless spectrum usage and create devices and systems that truly interact with the users. We refer the readers to [42] for a thorough discussion on the potential of the cognitive radio architecture.

Cognitive radios, at the very least, require enhanced sharing of information between the physical and the MAC layers of the protocol stack. Basically, unlike a conventional radio, the frequency band in which a cognitive radio may be operating at any given time depends on the channel occupancy measured at the physical layer, and conveyed to the MAC layer through an appropriate interface. This clearly requires cross-layer design.

6.3.4 Cognitive Networks and Cross-Layer Design

Cross-layer design done with all the three aforementioned motivations is integral to the cognitive networks of the future. Indeed, an efficient protocol stack that responds to the environment, network conditions and user demands is central to the very idea of cognitive networks. As we have seen above, the pursuit of achieving such efficiency and flexibility in wireless networks is not feasible by maintaining the strict boundaries between the different network functionalities, as done by layering. The way forward, thus, is inevitably cross-layer design.

6.4 A Taxonomy of Cross-Layer Design Proposals

In the previous section, we took a quick look at the broad motivations for the different cross-layer design ideas in the literature. Having seen *why* the presence of wireless links motivates designers to violate layered architectures, we now turn our attention to *how* the layered architectures are violated in the different cross-layer design proposals.

Based on the published research literature, we find that the layered architectures have been violated in the following basic ways:

1. Creation of new interfaces (Figure 6.1(a), (b), (c)).
2. Merging of adjacent layers (Figure 6.1(d)).
3. Design coupling without new interfaces (Figure 6.1(e)).
4. Vertical calibration across layers (Figure 6.1(f)).

We shall now discuss the aforementioned four categories in more detail and point out some relevant examples for the different categories. A few points are worth mentioning here. First, the examples that we point out are meant to be representative and not exhaustive. Secondly, the architectural violations that we identify can be combined to yield more

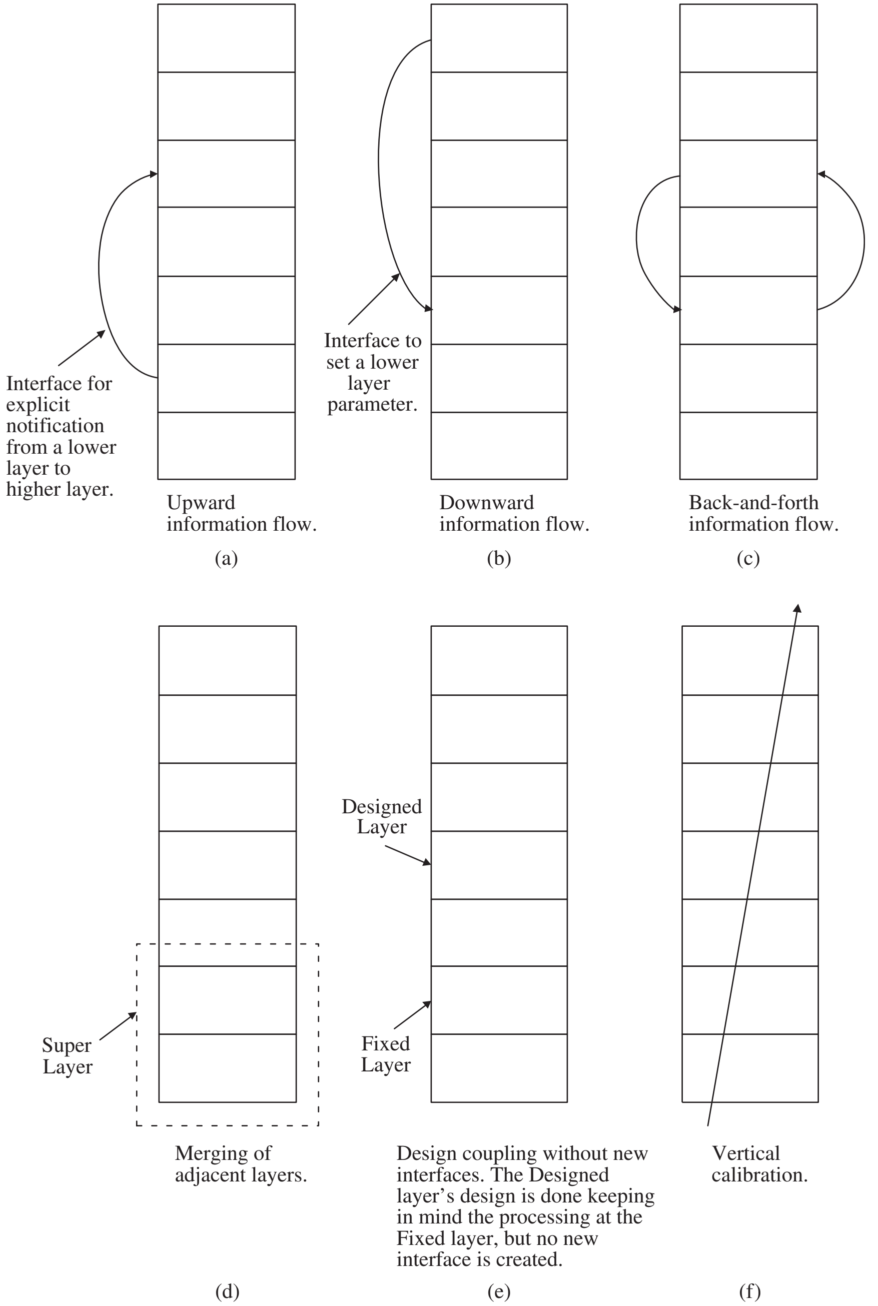

Figure 6.1 Illustrating the different kinds of cross-layer design proposals. The rectangular boxes represent the protocol layers

complex cross-layer designs. Finally, the reference layered architecture we assume is a five-layer hybrid reference model that we discussed in Section 6.3.

6.4.1 Creation of New Interfaces

Several cross-layer designs require creation of new interfaces between the layers. The new interfaces are used for information sharing between the layers at run-time. The architecture violation here is obviously the creation of a new interface not available in the layered architecture. We further divide this category into three categories depending on the direction of information flow along the new interfaces:

1. Upwards: from lower layer(s) to a higher layer.
2. Downwards: from higher layer(s) to a lower layer.
3. Back and forth: iterative flow between the higher and lower layer.

We now discuss the three subcategories in more detail and point out the relevant examples.

6.4.1.1 Upward Information Flow

A higher layer protocol that requires some information from the lower layer(s) at run-time results in the creation of a new interface from the lower layer(s) to the higher layer, as shown in Figure 6.1(a). For instance, if the end-to-end TCP path contains a wireless link, errors on the wireless link can trick the TCP sender into making erroneous inferences about the congestion in the network and as a result the performance deteriorates. Creating interfaces from the lower layers to the transport layer to enable explicit notifications alleviates such situations. For example, the explicit congestion notification (ECN) from the router to the transport layer at the TCP sender can explicitly tell the TCP sender if there is congestion in the network to enable it to differentiate between errors on the wireless link and network congestion [52]. Similarly [66], addressing the problems created for TCP by the disruption of links in the end-to-end route due to mobility in wireless ad hoc networks, also discusses the addition of more such explicit notifications between the link layer and the transport layer.

Examples of upward information flow are also seen in the literature at the MAC layer (link layer in general) in the form of channel-adaptive modulation or link-adaptation schemes [30], [32], [46], [50]. The idea is to adapt the power and data rates in response to the channel condition, which is made known to the MAC layer (link layer) by an interface from the physical layer. Reference [63] extends such a cross-layer adaptation loop by deciding the link layer parameters by considering both the channel condition as well as the instantaneous buffer occupancy at the transmitter, as we discussed in Section 6.3.2.

It is interesting to compare and contrast cross-layer design proposals that rely on upward flow of information to what can be called as self-adaptation loops at a layer. By a self-adaptation loop, we mean an adaptive higher layer protocol that respond to events that, within the constraints of layering, are directly observable at the layer itself. Hence, self-adaptation loops do not require new interfaces to be created from the lower layer(s) to the higher layer and cannot be classified as cross-layer designs. For example, consider the auto-rate fallback mechanism for rate selection [33] in wireless devices with multirate physical layers. The idea is that if some number of packets sent at a particular rate are successfully delivered, the data rate is increased, whereas, if a packet failure is experienced, data rate is dropped. In this case, the rate selection mechanism responds to the acknowledgements, which are directly observable at the MAC layer. Hence we do not treat auto-rate fallback as a cross-layer design. Similar examples of self-adaptations are provided in [7] and [28]. In fact, TCP provides another example of a self-adaptation loop since TCP changes its window size in response to the acknowledgements observed at the transport layer itself.

6.4.1.2 Downward Information Flow

Some cross-layer design proposals rely on setting parameters on the lower layer of the stack at run-time using a direct interface from some higher layer, as illustrated in Figure 6.1(b). As an example, the applications can inform the link layer about their delay requirement, and the link layer can then treat packets from the delay-sensitive applications with priority [65].

A good way to look at the upward and downward information flow is to treat them as notifications and hints, respectively, as proposed in [38]. Upward information flow serves the purpose of notifying the higher layers about the underlying network conditions; downward information flow is meant to provide hints to the lower layers about how the application data should be processed.

6.4.1.3 Back and Forth Information Flow

Two layers, performing different tasks, can collaborate with each other at run-time. Often, this manifests in an iterative loop between the two layers, with information flowing back and forth between the layers, as highlighted in Figure 6.1(c). Clearly, the architecture violation here are the two complimentary new interfaces.

As an example, we refer to the network-assisted diversity multiple access (NDMA) proposal [22], whereby the physical and the MAC layers collaborate in collision resolution in the uplink of a wireless local area network (LAN) system. Basically, with improvements in the signal processing at the physical layer (PHY), it becomes capable of recovering packets from collisions. Thus, upon detecting a collision, the base station first estimates the number of users that have collided, and then requests a suitable number of retransmissions from the set of colliding users. Then PHY signal processing lets the base station separate the signals from all the colliding users.

As another example, consider the problem of joint scheduling and power control in wireless ad hoc networks. Examples include [12], [24], [25], [36]. Basically, power control determines the effective topology of the network by determining which nodes can communicate with one another in a single hop. If the transmitted power is too large, then many nodes may be connected by a single hop, but the interference also would be large. On the other hand, keeping the power too small can make the network fragmented or create too many hops and hence added MAC contention. Protocols resulting from considering the joint problem of power control and scheduling often result in an iterative solution: Trying to keep the power level at an optimal level by responding to the changes in averaged throughput (e.g. see [25]). Reference [24] considers the joint scheduling and power control problem in time-division multiple access (TDMA) based wireless ad hoc networks. A scheduling algorithm chooses the users that will transmit, and then a power control algorithm determines if the transmissions of all the chosen users can simultaneously go on. If no, the scheduling algorithm is repeated. This iteration between scheduling and power control is repeated until a valid transmission schedule has been found. While there are differing views on which layer should power control belong to,[1] the collaborative nature of the cross-layer design mentioned above and the back and forth information flow that they require should be clear.

6.4.2 Merging of Adjacent Layers

Another way to do cross-layer design is to design two or more adjacent layers together such that the service provided by the new 'super-layer' is the union of the services provided by the constituent layers. This does not require any new interfaces to be created in the stack. Architecturally speaking, the super-layer can be interfaced with the rest of the stack using the interfaces that already exist in the original architecture.

Although we have not come across any cross-layer design proposal that explicitly creates a super-layer, it is interesting to note that the collaborative design between the PHY and the MAC layers that we discussed in Section 6.4.1.3 while discussing the NDMA idea tends to blur the boundary between these two adjacent layers.

6.4.3 Design Coupling Without New Interfaces

Another category of cross-layer design involves coupling between two or more layers at design time without creating any extra interfaces for information-sharing at run-time. We illustrate this in Figure 6.1(e). While no new interfaces are created, the architectural cost here is that it may not be possible to replace one layer without making corresponding changes to another layer.

For instance, [60] considers the design of MAC layer for the uplink of a wireless LAN when the PHY layer is capable of providing multipacket reception capability. Multipacket reception capability implies that the physical layer is capable of receiving more than one packet at the same time. Notice that this capability at the physical layer considerably changes the role of the MAC layer; thus, it needs to be redesigned. Similarly [55] considers the design of MAC layer in ad hoc networks with smart antennas at the physical layer.

6.4.4 Vertical Calibration Across Layers

The final category in which cross-layer design proposals in the literature fit into is what we call vertical calibration across layers. As the name suggests, this refers to adjusting parameters that span across layers, as illustrated in Figure 6.1(f). The motivation is easy to understand. Basically, the performance seen at the level of the application is a function of the parameters at all the layers below it. Hence, it is conceivable that joint tuning

[1] Power control is mentioned as a physical layer task in [12]. A different view is taken in [43] and [24] where power control is placed at the MAC layer and in [35] where power control is placed at the network layer.

can help to achieve better performance than individual settings of parameters – as would happen had the protocols been designed independently – can achieve.

As an example, [2] looks at optimizing the throughput performance of the TCP by jointly tuning power management, forward error correction (FEC) and ARQ settings. Similarly, [40] presents an example of vertical calibration where the delay requirement dictates the persistence of the link-layer ARQ, which in turn becomes an input for the deciding the rate selection through a channel-adaptive modulation scheme.

Vertical calibration can be done in a static manner, which means setting parameters across the layers at design time with the optimization of some metric in mind. It can also be done dynamically at run-time, which emulates a flexible protocol stack that responds to the variations in the channel, traffic and overall network conditions. Examples of dynamic vertical calibrations can be found in [49] and [1]. Both these references describe prototype systems that employ adaptations across several layers of the protocol stack.

Static vertical calibration does not create significant consideration for implementations since the parameters can be adjusted once at design time and left untouched thereafter. Dynamic vertical calibration, on the other hand, requires mechanisms to retrieve and update the values of the parameters being optimized from the different layers. This may incur significant cost in terms of overheads and also impose strict requirements on the parameter retrieval and update process to make sure that the knowledge of state of the stack is current and accurate. We again refer the readers to [49] and [1] for a thorough discussion of the challenges involved in implementing dynamic vertical calibrations.

6.5 Proposals for Implementing Cross-Layer Interactions

Alongside the cross-layer design proposals that we discussed in Section 6.4, initial proposals on how cross-layer interactions can be implemented are also being made in the literature. These can be put into three categories:

1. Direct communication between layers (Figure 6.2(a)).
2. A shared database across the layers (Figure 6.2(b)).
3. Completely new abstractions (Figure 6.2(c)).

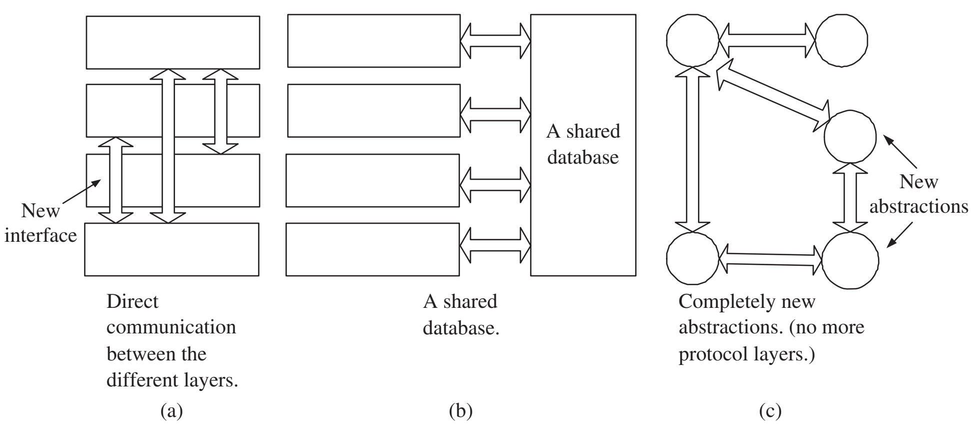

Figure 6.2 Proposals for architectural blueprints

6.5.1 Direct Communication Between Layers

A straightforward way to allow run-time information sharing between the layers is to allow them to communicate with each other, as depicted schematically in Figure 6.2(a). Note that this is applicable when there has to be run-time information sharing between layers (for example, in cross-layer designs that rely on new interfaces or in dynamic vertical calibrations). Practically speaking direct communication between the layers means making the variables at one layer visible to the other layers at run-time. By contrast, under a strictly layered architecture, every layer manages its own variables and its variables are of no concern to other layers.

There are many ways in which the layers can communicate with one another. For instance, protocol headers may be used to allow flow of information between the layers. Alternatively, the extra 'inter-layer' information could be treated as internal packets. The work in [64] presents a comparative study of several such proposals and goes on to present another such proposal, namely, the cross-layer signaling shortcuts (CLASS). CLASS allows any two layers to communicate directly with one another.

These proposals are appealing in the case where just a few cross-layer information exchanges are to be implemented in systems that were originally designed in conformance with the layered architectures. In that case, one can conceivably 'punch' a few holes in the stack while still keeping it tractable. However, in general, when variables and internal states from the different layers are to be shared as prescribed by such proposals, a number of implementation issues relating to managing shared memory spaces between the layers may need to be resolved.

6.5.2 A Shared Database Across the Layers

The other class of ideas propose using a common database that can be accessed by all the layers, as illustrated in Figure 6.2(b). See for instance [15], [41], [48] and [57]. In one sense, the common database is like a new layer, providing the service of storage/retrieval of information to all the layers.

The shared database approach is particularly well suited to vertical calibrations across layers. An optimization program can interface with the different layers at once through the shared database. Similarly, new interfaces between the layers can also be realized through the shared database. The main challenge here is the design of the interactions between the different layers and the shared database.

6.5.3 Completely New Abstractions

The third set of proposals present completely new abstractions, which we depict schematically in Figure 6.2(c). Consider, for example, the proposal in [10] which presents a new way to organize the protocols: in heaps and not in stacks as done by layering.

Such novel organizations of protocols are appealing as they allow rich interactions between the building blocks of the protocols. Hence, potentially they offer great flexibility, both during the design as well as at run-time. However, they change the very way protocols have been organized, and hence may require completely new system-level implementations.

6.6 Cross-Layer Design Activity in the Industry and Standards

So far we have been talking about the cross-layer design activity largely from the research literature. Interestingly, cross-layer design ideas have been making their way into several ongoing standardization activities and commercial products too. In this section we take a brief look at some of this industry activity.

6.6.1 3G Cellular Networks

The channel-dependent scheduling that we discussed in Section 6.3.2 has been incorporated into the high data rate (HDR) version of CDMA2000 (1xEV-DO) and enhanced general packet radio service (EGPRS) in the EDGE extension of GSM [52]. In these systems, the users periodically update the base station about the state of the downlink channel. This information is used by the base station for scheduling the transmissions to the different users.

6.6.2 Vertical Handovers

The problem of vertical handovers that we discussed in Section 6.3.3 is being addressed by the IEEE 802.21 standards group. The IEEE 802.21 group is defining standardized interfaces for sharing information from the lower layers (e.g. link layer and physical layer) with the higher layers (e.g. network layer). The IEEE 802.21 effort thus is an interesting example of convergence between the performance viewpoint, which dictates the need for cross-layer information sharing, and the architectural viewpoint, which requires standardized and well-defined interfaces between the modules.

6.6.3 Wireless Regional Area Networks

The IEEE 802.22 is working to standardize the operation of a wireless regional area network (WRAN). An excellent overview of this standardization effort can be found in [17]. The distinctive feature of the IEEE 802.22 is that it defines a wireless network that is to operate within the spectrum that has been assigned to the television broadcast systems. In order to protect the 'primary' users (the TV broadcasters) the 802.22 standard makes use of the cognitive radio techniques, thus representing another example of cross-layer design (between the MAC and the PHY layers).

6.6.4 Wireless Local Area Networks

The IEEE 802.11 standards cover the wireless local area networks that are primarily used within homes or small offices for wireless Internet access. The IEEE 802.11 a/b/g suite of technologies has found its way into many homes and offices worldwide. These technologies, however, do not perform well for multimedia traffic, primarily because the IEEE 802.11 MAC does not provide any kind of QoS guarantees. This problem has been addressed in an extension to the 802.11 MAC under the effort called IEEE 802.11e [44]. One of the features of IEEE 802.11e MAC that has a strongly cross-layer design flavor is in the realm of what is called the enhanced distributed channel access (EDCA). The idea is to categorize the incoming data from the higher layer into four priority classes at the MAC layer. This allows different kinds of data (for instance voice, video, time-insensitive data) to be treated differently at the MAC layer, such that their individual QoS requirements may be satisfied. In architectural terms, this is equivalent to having a direct communication interface between the higher layers and the MAC layer. The IEEE 802.11e has strong support from the industry. As an example, several of the IEEE 802.11e features have been incorporated into the 'Super-G' chipsets from the wireless LAN chipset vendor Atheros (see http://www.atheros.com).

6.6.5 Multimedia Over Wireless

In the context of multimedia over wireless networks, we point out Ruckus Wireless (http://www.ruckuswireless.com), a privately held company, which has also taken a cross-layer design approach remarkably similar to the 802.11e idea above in solving the problem of maintaining the QoS requirements of the different kinds of traffic types. The company calls its technology 'Smart-Cast'; the key idea is to apply a traffic-dependent scheduling. That is, at the MAC layer, the packets are classified according to their latency requirements, and a scheduling algorithm then determines the ordering in which the packets are transmitted over the wireless medium. The Ruckus Wireless system also features a proprietary technology called 'Beam Flex'. The idea in Beam Flex is to continuously monitor the interference in the wireless medium, and using the appropriate combination of the smart antenna system, route the traffic around the interference. This, thus, once again is an example of cross-layer design involving the physical layer and the MAC layer.

Digital Fountain (http://www.digitalfountain.com) provides error protection by employing a type of erasure correction code known as a fountain code (termed DF Raptor FEC by the company). This technology has been standardized for use in multimedia broadcast and multicast services (MBMS) in 3G. Basically, the use of error correction at the higher layers provides an enhanced line of defense against impairments caused by the wireless channel and network congestion. Fountain codes allow the receiver to recover lost packets without requiring the sender to keep track of and retransmit missing packets. This is done by cleverly introducing redundancy via the generation of a potentially large amount of encoded data for a given piece of content (say an mp3 file). The receiver then just has to collect a sufficient number of coded packets to decode the multimedia content, without keeping track of exactly which packets it has. When it has decoded the file, it can inform the sender to stop. Digital Fountain's DF Raptor FEC efficiently mitigates the need for end-to-end reliability at the packet level, and pushes it to the file level. Furthermore, since the codes in question are fountain codes, they are highly scalable and provide adaptive error protection in face of variations in the wireless transmission environment and the network quality. That the DF Raptor FEC technology takes a cross-layer design approach is evident in the fact that it can be employed at the transport layer or at the application layer and it uses forward error correction to provide end-to-end reliability.

6.6.6 Mesh Networks

In recent times there has been a considerable interest in the deployments of mesh networks. MIT Roofnet [8] is an example of an experimental mesh network deployed at Cambridge, Massachusetts, in the United States. The MIT Roofnet makes use of a routing protocol called ExOR which is in fact an integrated routing and link layer protocol [9]. ExOR relies on the broadcast nature of the wireless medium. (MIT Roofnet uses omnidirectional antennas.) Whenever the source has a packet to send, it broadcasts the packet which is received by a number of potential forwarders. One out of the set of potential forwarders, chosen based on the reliability of the link between the potential forwarders and the destination, forwards the packet further. This process continues until the packet is delivered at the destination. We refer the reader to [9] for more details. As far as we are concerned, we just notice that the cross-layer design involving the network and the link layers results in throughput improvements compared with the traditional approaches. Tropos Networks (http://www.tropos.com) deploys commercial wireless mesh networks. Though the details of Tropos' proprietary routing algorithm, the Predictive Wireless Routing Protocol (PWRP), are not known, we believe, based on the cursory descriptions that we have read, that PWRP also makes use of throughput information from the MAC layer in making the routing decisions at the network layer. This represents another example of cross-layer design between the network and the MAC layers.

This look at the standardization and the industry activity that we have taken is no doubt an extremely brief sampling of the activity in the industry. It, nonetheless, clearly illustrates that cross-layer design ideas are finding their way into commercial products and standards, which should be an encouragement for researchers in this field.

6.7 The Open Challenges

In Sections 6.3 through 6.6, we looked at the ongoing work in the area of cross-layer design. In doing so, we came face to face with the general motivations behind cross-layer design, the several different interpretations of cross-layer design, many examples of cross-layer design, some initial ideas on how cross-layer interactions may be implemented, and a quick snapshot on how cross-layer design ideas are getting into ongoing standardizations and commercial deployments. Having taken stock of the ongoing work, we now raise and discuss what we call the 'open challenges' for cross-layer design. Broadly speaking, we include two kinds of issues in the open challenges: questions about cross-layer design that, in our opinion, are important but are not getting sufficient attention in the research literature; and some questions regarding the fundamental nature of the wireless medium, questions whose answers will influence how communication architectures for the wireless networks should look like, and hence are pertinent to the cross-layer design effort. We note that some of these issues have been raised elsewhere in the literature too. As before, our purpose here is to consolidate the different issues and discuss their significance with respect to the cross-layer design activity.

The following are some open challenges for the designers proposing cross-layer design ideas:

1. How do the different cross-layer design proposals coexist with one another?
2. Will a given cross-layer design idea possibly stifle innovation in the future?
3. What are the cross-layer designs that will have the most significant impact on network performance, and hence should be most closely focused on?
4. Has a given design proposal been made with a thorough knowledge of the effect of the interactions between the parameters at different layers on network performance?
5. Under what network and environmental conditions be a particular cross-layer design proposal be invoked?
6. Can the mechanisms/interfaces used to share information between the layers be standardized?

7. What should be the role of the physical layer in wireless networks?
8. Is the conventional view of the network – that of a collection of point-to-point links – appropriate for wireless networks?

We now look at some of these issues in greater detail.

6.7.1 The Important Cross-Layer Couplings

While there are a number of cross-layer design proposals in the literature today, it is not clear which are the most important ones. A thorough cost-benefit analysis of the different cross-layer design proposals in terms of the implementation complexity versus the performance improvement is needed. To be fair, the important cross-layer design ideas depend on the specific network scenario considered and the metric of interest. For example, in the case of wireless ad-hoc networks, one can make the following inferences from the literature today: cross-layer design is needed between network layer and the MAC layer for wireless ad hoc networks, as we discussed in Section 6.6.6, since the functionalities of the two layers interact [3]; explicit notifications by new interfaces to the transport layer improve end-to-end performance [52]; making use of the channel knowledge at the MAC layer allows opportunistic usage of the channel and improves performance ([32] and the references therein); and energy, delay and security-related issues need to be handled across the layers in a holistic manner. It is time to move ahead from these general insights and harmonize the different cross-layer design ideas. This requires comparative quantitative study of the different cross-layer design proposals, and is an open challenge for the community. Also relevant in this context is the question of coexistence of cross-layer design proposals, which we discuss next.

6.7.2 Co-Existence of Cross-Layer Design Proposals

An important question to be answered is how different cross-layer design proposals can coexist with one another. To clarify by example, say the MAC layer in a stack responds to the variation in the channel by adjusting the data rate. The question is, will additionally adjusting the frame length at the link layer help further? How will an overriding control from, say, the transport layer, trying to control the link layer parameters, interact with these adaptation loops?

The question of coexistence of cross-layer design ideas is pertinent when it comes to determining whether some cross-layer design proposals can stifle further innovation. Let us say the physical layer and the link layer are optimized for a certain performance metric in a cross-layer design scheme. If this scheme is deployed first, can other schemes that also rely on some (other) cross-layer couplings, or those that assume no coupling between the link layer and the physical layer be deployed too at a later time?

Apart from presenting new cross-layer design proposals, designers need to start establishing which other cross-layer design interaction may or may not be employed together with their proposal. This has also been stressed in [34].

6.7.3 When to Invoke a Particular Cross-Layer Design?

The network conditions in a wireless network are usually time varying. In such a situation, one of the stated motivations behind cross-layer design is to achieve the network equivalent of impedance matching [52]. The idea is to make the protocol stack responsive to the variations in the underlying network conditions so that an optimal operating point is always maintained.

The pursuit of achieving such optimal operation throws up two complimentary challenges. First, designers need to establish the network conditions under which the proposed cross-layer designs would result in performance improvements. Reference [34] presents an example to illustrate how a cross-layer design involving an iterative optimization of throughput and power leads to a loss in performance under a certain pathological network condition. The example in [34] underscores the need for designers to establish the network conditions under which their design proposals should and should not be used. Secondly, efficient mechanisms to make a timely and accurate assessment of the state of the network need to be built into the stack, and the corresponding overheads need to be taken into account [1], [49]. This is also related to the question of interfaces between the modules, discussed next.

6.7.4 Standardization of Interfaces

The one thing that layering achieved was to present standardized boundaries and interfaces between the modules of the system, namely, the protocol layers. Now that the layered architecture is being violated in different ways, finding the new reference architecture becomes a challenge. What should be the boundaries between the modules? Should we stick to the traditional layer boundaries as in Figure 6.2(a) and (b) and determine the new interfaces from there, or should we look at completely new boundaries, as in Figure 6.2(c)? Or a combination? What should the interfaces between the modules look like?

Addressing this challenge requires greater synergy between the performance viewpoint and the implementation concerns than what is seen in the literature today. Basically, the organization of the modules (layers or otherwise) and the interfaces between them determine how efficiently can information be shared between them, at what kinds of overheads and delays. This, in turn, determines how effective cross-layer design proposals that rely on sharing dynamic information between the modules can be. Hence, proposers of cross-layer design relying on back and forth information flow between layers or dynamic vertical calibrations need to start considering the impact of delays in the retrieval/updating of information on the protocol performance. They also need to quantify the overheads associated with their cross-layer design proposals.

One point that should be mentioned here is that the interfaces get standardized during the process of drafting technical standards. A good example is the IEEE 802.21 standardization which is codifying the interfaces to be followed for the transfer of information between the different layers of the stack.

6.7.5 The Role of the Physical Layer

In wired networks, the role of the physical layer has been rather small: sending and receiving packets when required to do so from the higher layers. This is also the case in the current generation of wireless technologies. As we have seen in Section 6.4.3, advances in the signal processing at the physical layer can allow it to play a bigger role in wireless networks. This begs a question as to how much of a role should the physical layer play? This question is relevant to the cross-layer design effort because first, layered architectures like the OSI reference model do not allow much role for the physical layer besides providing a bit-pipe; and second, enhancements in the physical layer will have to be balanced by corresponding changes to the higher layers. Hence, figuring out the role to be played by the physical layer is an important question. Cross-layer designs relying on advanced signal processing at the physical layer can be an interesting research ground for the future.

6.7.6 The Right Communication Model

Wired networks, by their very nature, are essentially a collection of well-defined point-to-point communication links. The same cannot be said about wireless networks because the wireless medium is inherently broadcast and there is no clear-cut concept of a communication link in wireless networks. This gives rise to a fundamental question of whether it still makes sense to 'create' links in a wireless network, which is what has been done in all the networks of the past. As we discussed in Section 6.3.3, cooperative communications, when done at the link/physical layer, fundamentally changes the abstraction that the higher layers have of the network since the network no longer looks like a collection of point-to-point links [37]. By coupling physical, link and network layers in intricate ways, cooperative communications inevitably invite a cross-layer design approach. Hence, the open question facing the research community is to explore such modalities of communications in context of end-to-end network performance, and to design suitable communications architectures to support such modalities.

6.8 Discussion

We have been looking at the ongoing work in the area of cross-layer design for wireless communications and networks throughout this chapter. Implicit in this entire discussion, as it is in most of the published research, is the existence and acceptance of a reference layered architecture. A good number of communication systems broadly follow the five-layer model, which is a hybrid between the two most commonly taught models: the OSI seven-layer model and the four-layer TCP/IP model. The question is, how did these models come about in the first place?

The four-layer TCP/IP model came into being as part of an initiative undertaken by the United States (US) Department of Defense to connect different kinds of packet-switched data networks. The idea was to enable a communication device connected to a packet-switched network to communicate with any other communication device connected to any other packet-switched network. Reference [13] discusses the key considerations that motivated the development of some specific features of the TCP/IP protocols (for example a connectionless mode of communications with end-to-end flow and error control provided by the transport layer). In effect, the motivations elaborated in [13] highlight the design principles of the Internet since, in time, the TCP/IP model became the model that shaped the Internet.

On the other hand, the OSI seven-layer model came into being as a result of an initiative taken up by the International Standards Organization (ISO) to come up with a set of standards that would allow disparate systems anywhere in the world to communicate with each other. The idea was to allow interoperability between communication equipments developed by the different vendors. The basic seven-layer reference model was published in early 1980s and it became an international standard (ISO 7498) in 1983 [20]. The OSI model had split the lowest layer of the TCP/IP model into three separate layers. It had also done the same for the highest layer.

Interestingly, in the case of the TCP/IP model, the protocols at the different layers were developed before formal layer definitions had been made [56], [59, p. 39]. Basically, as the TCP/IP was getting bundled with the UNIX operating system, the popularity of these protocols in the research community shot up [56]. In due course, layer definitions were attached to the model [59, p. 39] and thus came the four-layer TCP/IP model. By contrast, in the case of the OSI seven-layer model, the layers were defined *before* the development of protocols [59]. As mentioned earlier, the idea behind the OSI model was to standardize the protocol development effort, and to do so, a flexible architecture was provided in the seven-layer model. Subsequent to the publication of the seven-layer model, protocols at the different layers were developed and published as separate international standards.

Reference [59] provides an insightful comparison and critique of both the models. On a broader note, it discusses the relative merits of the contrasting approaches: architecture before protocols, or protocols before architecture. Interestingly, engineers actually implementing the protocols, or considering implementation issues, have long raised questions about the suitability of strict protocol layering, whether defined (as in OSI) or adopted (as in TCP/IP). For instance, [16] advocated the concept of ‘soft-layers’ that allow more flexible information sharing between adjacent layers, [58] discussed in detail the inefficiencies and problems that resulted in implementing layered OSI systems, and [14] and [18] talked of integrated layer processing – none because of wireless links in the network, or because of a desire to create cognitive networks! In fact based on the feedback from the designers, the original OSI model was revised and a new reference was published. An interesting summary of the major changes to the original reference model and their motivation, peppered with palpable political overtones, can be found in [21].

Seen in this historical light, there is really no surprise that the increasing proliferation of wireless networking and the pursuit of cognitive networks is ushering in a large number of cross-layer design proposals. After all, the layered architectures themselves have been created in response to specific technical and industrial needs, and are not in any way all encompassing. What we are witnessing today are, by and large, minor tweaking to these well-understood layered architectures. It is entirely conceivable that in future the architectures of today will evolve and may or may not look similar to the ones today. Every cross-layer design proposal serves to highlight a specific shortcoming of traditional protocol layering for the scenario in question. Over time these ideas, if significant, will be incorporated into the standards or proprietary systems, and hence the architectures will evolve.

That said, we hasten to add that while the protocol layers as we have been taught in networking courses may not survive, the core principle of modularity inevitably will, as that is the very principle that underpins the modern economy. Possibly a new kind of layering will evolve for wireless networks. As researchers in the field, we will do well to see our cross-layer design ideas in the bigger context of specific architectures that we are looking at, and hence evaluate their coexistence with other ideas, their impact on the deployability of other ideas, and so on. In other words, even though cross-layer design does inevitably look like the way forward, researchers should look not just at the performance viewpoint, but also at the larger architectural consequences of their work. In stating this, we are unanimous with [34].

6.9 Conclusions

In this chapter, we consolidated the several different ideas and results falling under the broad purview of cross-layer design. We saw that cross-layer design can be defined as a violation of a reference layered architecture. Cross-layer design is done to either solve a unique problem or exploit a new opportunity or new modality created by wireless links. We taxonomized the different cross-layer design proposals according to the kind of architecture violations they represent, looked at the initial ideas for implementing cross-layer interactions and briefly described some cross-layer design ideas that have made their way into commercial products and industry standards. We then raised some open questions related to cross-layer design that researchers may want to address as they move forward. Finally we placed the cross-layer design of today in a brief historic perspective.

Through the broad-based discussion that we undertook in this chapter, we have seen both the wide-applicability of cross-layer design ideas, as well as the inevitability of cross-layer design in the pursuit of cognitive wireless networks of the future. There is little doubt that cross-layer design holds tremendous potential to unleash the true potential of wireless communications. However, for that potential to be fulfilled, it is imperative for the designers to view cross-layer design holistically by considering both performance and architectural viewpoints. We hope that by consolidating the varied ideas, categorizing the initial thoughts and by raising some open questions, this chapter will facilitate such holistic thinking about cross-layer design.

References

[1] Akyildiz, I., Altunbasak, Y., Fekri, F. and Sivakumar, R. (2004) AdaptNet: an adaptive protocol suite for the next-generation wireless Internet. *IEEE Communications Magazine*, **42**(3), 128–36.

[2] Barman, D., Matta, I., Altman, E. and Azuozi, R.E. (2004) TCP optimization through FEC, ARQ and transmission power tradeoffs. *Proceedings of the International Conference on Wired/Wireless Internet Communications*, Frankfurt.

[3] Barrett, C., Drozda, M., Marathe, A. and Marathe, M.V. (2002) Characterizing the interaction between routing and MAC protocols in ad-hoc networks. Proceedings of the ACM Annual International Symposium on Mobile Ad-Hoc Networking and Computing (MobiHoc'02), pp. 92–103, June, Lausanne.

[4] Berry, R.A. and Gallager, R.G. (2002) Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory*, May, 1135–49.

[5] Berry, R.A. and Yeh, E. (2004) Cross-layer wireless resource allocation. *IEEE Signal Processing Magazine*, September, 59–68.

[6] Bertsekas, D. and Gallager, R. (1992) *Data Networks*, 2nd edn, Prentice Hall, New Jersey.

[7] Bhagwat, P., Bhattacharya, P., Krishna, A. and Tripathi, S.K. (1996) Enhancing throughput over wireless LANs using channel state dependent packet scheduling. *Proceedings of the IEEE Infocom'96*, pp. 1133–40.

[8] Bicket, J., Aguayo, D., Biswas, S. and Morris, R. (2005) Architecture and evaluation of an unplanned 802.11b mesh network. *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking (MobiCom'05)*, pp. 31–42, August.

[9] Biswas, S. and Morris, R. (2005) Opportunistic routing in multi-hop wireless networks. Proceedings of the ACM (SIGCOMM'05), pp. 69–74, August.

[10] Braden,, R., Faber, T. and Handley, M. (2002) From protocol stack to protocol heap – role-based architecture. *Proceedings of the Hot Topics in Networking (HOTNETS I)*, March, Princeton, NJ.

[11] Burbank, J.L. and Kasch, W.T. (2005) Cross-layer design for military networks. *Proceedings of the IEEE Military Communications Conference (MILCOM'05)*, pp. 1912–18, October.

[12] Chiang, M. (2004) To layer or not to layer: balancing transport and physical layers in wireless multihop networks. *Proceedings of the IEEE Infocom'04*, pp. 2525–36, March, Hong Kong.
[13] Clark, D.D. (1988) The design philosophy of the DARPA Internet protocols. *Proceedings of theACM Symposium on Communications Architectures and Protocols (SIGCOMM'88)*, pp. 106–14.
[14] Clark, D.D. and Tennenhouse, D.L. (1990) Architectural considerations for a new generation of protocols. *Proceedings of the ACM Symposium on Communications Architectures and Protocols (SIGCOMM'90)*, Philadelphia.
[15] Conti, M., Maselli, G., Turi, G. and Giordano, S. (2004) Cross-layering in mobile ad hoc network design. *IEEE Computer Magazine*, February, 48–51.
[16] Cooper, G.H. (1983) *An argument for soft layering of protocols*, Master's thesis, Masachussets Institute of Technology.
[17] Cordeiro, C., Challapali, K., Birru, D. and Shankar, N.S. (2005) IEEE 802.22: the first worldwide wireless standard based on cognitive radios. *Proceedings of the IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN'05)*, pp. 328–37, November.
[18] Crowcroft, J., Wakeman, I., Wang, Z. and Sirovica, D. (1992) Is layering harmful? *IEEE Network*, **6**(1), 20–4.
[19] Cui, S. and Goldsmith, A.J. (2006) Cooperation techniques in cross-layer design. In F.H.P. Fitzek and M.D. Katz (eds.), Cooperation in *Wireless Networks: Principles and Applications*, Springer.
[20] Day, J. (1989) The reference model for open systems interconnection. In C.A. Sunshine (ed.), *Computer Network Architectures and Protocols*, 2nd edn, Plenum Press, New York.
[21] Day, J.D. (1995) The (un)revised OSI reference model. *ACM Computer Communication Review*, **25**(10), 39–55.
[22] Dimić, G., Sidiropoulos, N. and Zhang, R. (2004) Medium access control – physical cross-layer design. *IEEE Signal Processing Magazine*, **21**(9), 40–50.
[23] Eberle, W., Bougard, B., Pollin, S. and Catthoor, F. (2005) From myth to methodology: cross-layer design for energy-efficient wireless communication. *Proceedings of the ACM Design Automation Conference (DAC'05)*, pp. 303–8, June, Anaheim, CA.
[24] Elbatt, T. and Ephremides, A. (2004) Joint scheduling and power control for wireless ad hoc networks. *IEEE Transactions on Wireless Communications*, **3**(1), 74–85.
[25] Elbatt, T., Krishnamurthy, S., Connors, D. and Dao, S. (2000) Power management for throughput enhancement in wireless ad-hoc networks. *Proceedings of the IEEE International Conference on Communications (ICC'2000)*, pp. 1506–13, New Orleans.
[26] Fitzek, F.H.P. and Katz, M.D. (2006) Cooperation in nature and wireless communications. In F.H.P. Fitzek and M.D. Katz (eds.), Cooperation in *Wireless Networks: Principles and Applications*, Springer.
[27] Goldsmith, A. and Varaiya, P. (1995) Capacity of fading channel with channel side information. *IEEE Transactions on Information Theory*, **43**, 1986–92.
[28] Hara, S., Ogino, A., Araki, M., Okada, M. and Morinaga, M. (1996) Throughput performance of SAW-ARQ protocol with adaptive packet length in mobile packet data transmission. *IEEE Transactions on Vehicular Technology*, **45**, 561–9.
[29] Haykin, S. (2005) Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, **23**, 201–20.
[30] Holland, G., Vaidya, N. and Bahl, P. (2001) A rate-adaptive MAC protocol for multi-hop wireless networks. *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking (MobiCom'01)*, pp. 236–51, July, Rome.
[31] Jamalipour, A. (2003) *The Wireless Mobile Internet*, John Wiley & Sons, Co., Chichester, UK.
[32] Ji, Z., Yang, Y., Zhou, J., Takai, M. and Bagrodia, R. (2004) Exploiting medium access diversity in rate adaptive wireless LANs. *Proceedings of the ACM Annual International Symposium on Mobile Computing and Networking (MobiCom'04)*, pp. 345–59, October, Philadelphia.
[33] Kamerman, A. and Monteban, L. (1997) WaveLAN II: a high-performance wireless LAN for the unlicensed band. *Bell Labs Technical Journal*, 118–33.
[34] Kawadia, V. and Kumar, P. (2005) A cautionary perspective on cross layer design. *IEEE Wireless Communications Magazine*, **12**, 3–11.
[35] Kawadia, V. and Kumar, P.R. (2005) Principles and protocols for power control in ad hoc networks. *IEEE Journal on Selected Areas in Communications*, **5**, 76–88.

[36] Kozat, U., Koutsopoulos, I. and Tassiulas, L. (2004) A framework for cross-layer design of energy-efficient communication with QoS provisioning in multi-hop wireless networks. *Proceedings of the IEEE Infocom'04*, pp. 1446–56, March, Hong Kong.

[37] Laneman, J.N. (2006) Cooperative diversity. In F.H.P. Fitzek and M.D. Katz (eds.), Cooperation in *Wireless Networks: Principles and Applications*, Springer.

[38] Larzon, L.-A., Bodin, U. and Schelén, O. (2002) Hints and notifications. *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC'02)*, pp. 635–41, March, Orlando, FL.

[39] Leiner, B.M., Cole, R., Postel, J. and Mills, D. (1985) The DARPA Internet protocol suite. *IEEE Communications Magazine*, **23**, 29–34.

[40] Liu, Q., Zhou, S. and Giannakis, G.B. (2004) Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links. *IEEE Transactions on Wireless Communications*, **3**(9), 1746–55.

[41] Mahonen, P., Petrova, M., Riihijarvi, J. and Wellens, M. (2006) Cognitive wireless networks: your network just became a teenager. *Proceedings of the IEEE Infocom'06*.

[42] Mitola, J. (2006) Cognitive radio architecture. In F.H.P. Fitzek and M.D. Katz (eds.), Cooperation in *Wireless Networks: Principles and Applications*, Springer.

[43] Muqattash, A. and Krunz, M. (2004) A single-channel solution for transmission power control in wireless ad hoc networks. *Proceedings of the ACM Annual International Symposium on Mobile Ad-Hoc Networking and Computing (MobiHoc'04)*, pp. 210–21, Tokyo.

[44] Ni, Q. (2005) Performance analysis and enhancements for IEEE 802.11e wireless networks. *IEEE Network*, **19**(7), 21–7.

[45] Nosratinia, A., Hunter, T.E. and Hedayat, A. (2004) Cooperative communications in wireless networks. *IEEE Communications Magazine*, **42**(10), 74–80.

[46] Qiao, D. and Choi, S. (2001) Goodput enhancement of 802.11a wireless LAN via link adaptation. *Proceedings of the IEEE International Conference on Communications (ICC'01)*, pp. 1995–2000, Helsinki.

[47] Radosavac, S., Moustakides, G.V. and Baras, J.S. (2006) Impact of the optimal MAC layer attacks on the network layer. *Proceedings of the 4th ACM Workshop on Security of Ad-hoc and Sensor Networks (SASN'06)*, pp. 135–46, October.

[48] Raisinghani, V.T. and Iyer, S. (2004) Cross-layer design optimizations in wireless protocol stacks. *Computer Communications*, **27**, 720–4.

[49] Sachs, D.G., Yuan, W., Hughes, C.J. *et al*. (2004) *GRACE: a hierarchical adaptation framework for saving energy*, Technical report, University of Illinois, Urbana Champagne.

[50] Sadeghi, B., Kanodia, V., Sabharwal, A. and Knightly, E. (2002) Opportunistic media access for multirate ad-hoc networks. *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking (MobiCom'02)*, September, Atlanta.

[51] Schaar, M.V.D. and Shankar, N.S. (2005) Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms. *IEEE Wireless Communications Magazine*, August, 50–8.

[52] Shakkottai, S., Rappaport, T.S. and Karlsson, P.C. (2003) Cross-layer design for wireless networks. *IEEE Communications Magazine*, **41**, 74–80.

[53] Srivastava, V. and Mehul, M. (2005) The road ahead for cross-layer design. *Proceedings of the IEEE Broadnets'05*, pp. 551–60, October.

[54] Srivastava, V. and Motani, M. (2005) Cross-layer design: a survey and the road ahead. *IEEE Communications Magazine*, **43**(12), 112–19.

[55] Sundaresan, K. and Sivakumar, R. (2004) A unified MAC layer framework for ad-hoc networks with smart antennas. *Proceedings of the ACM Annual International Symposium on Mobile Ad-Hoc Networking and Computing (MobiHoc'04)*, pp. 244–55, Tokyo.

[56] Sunshine, C.A. (1989) A brief history of computer networking. In C.A. Sunshine (ed.), *Computer Network Architectures and Protocols*, 2nd edn, Plenum Press, New York.

[57] Sur, A. and Sicker, D.C. (2005) Multi layer rules based framework for vertical handoff. *Proceedings of the IEEE Broadnets'05*, pp. 571–80, October.

[58] Svobodova, L. (1989) Implementing OSI systems. *IEEE Journal on Selected Areas in Communications*, **7**, 1115–30.

[59] Tanenbaum, A.S. (1996) *Computer Networks*, 3rd edn, Prentice-Hall, New Jersey.

[60] Tong, L., Naware, V. and Venkitasubramaniam, P. (2004) Signal processing in random access, *IEEE Signal Processing Magazine*, **21**, 29–39.

[61] Tse, D. and Hanly, S. (1998) Multi-access fading channels: part I: polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Transactions on Information Theory*, **44**, 2796–815.

[62] Tse, D. and Hanly, S. (1998) Multi-access fading channels: part II: Delay-limited capacities. *IEEE Transactions on Information Theory*, **44**, 2816–31.

[63] Tuan, H.A. and Mehul, M. (2003) Buffer and channel adaptive modulation for transmission over fading channels. *Proceedings of the IEEE International Conference on Communications ((ICC'03)*, pp. 2748–52, July, Anchorage.

[64] Wang, Q. and Abu-Rgheff, M.A. (2003) Cross-layer signalling for next-generation wireless systems. *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC'03)*, pp. 1084–9, March, New Orleans.

[65] Xylomenos, G. and Polyzos, G.C. (2001) Quality of service support over multi-service wireless Internet links. *Computer Networks*, **37**, 601–5.

[66] Yu, X. (2004) Improving TCP performance over mobile ad hoc networks by exploiting cross-layer information awareness. *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking (MobiCom'04)*, pp. 231–244, October, Philadelphia.

[67] Zhai, F., Eisenberg, Y. and Katsaggelos, A.K. (2005) Joint source-channel coding for video communications. In A. Bovik (ed.), *Hand of Image and Video Processing*, 2nd edn, Elsevier Academic Press.

[68] Zhang, J. (2002) Bursty traffic meets fading: a cross-layer design perspective. *Proceedings of the Allerton Conference on Communications, Control and Computing'02*.

[69] Zhang, Q., Zhu, W. and Zhang, Y.-Q. (2001) Resource allocation for multimedia streaming over the Internet. *IEEE Transactions on Multimedia*, **3**, 339–55.

[70] Zhao, B. and Valenti, M.C. (2005) Practical relay networks: a generalization of hybrid-ARQ. *IEEE Journal on Selected Areas in Communications*, **23**, 7–18.