

Cross-layer Designs: a Survey

Md. Mohiuddin Khan
Department of Computer Science and Engineering
Islamic University of Technology (IUT)
Dhaka, Bangladesh

ABSTRACT

Computer communication has been going through major changes throughout the last decades. While wireless technologies have been widely adopted, various domains and implementations like wireless sensor Networks (WSN), mobile ad-hoc networks (MANETs), wireless mesh networks have emerged. Since TCP/IP was a protocol designed for wired networks, wireless transmission poses unique challenges to the well-defined and rigid protocol stack. The well-known layers of the OSI model or its practical counterpart TCP/IP model were too strict in some cases to provide with all the services necessitated by these new domains. These issues make way to cross-layer design where the traditional boundaries among layers are violated in different ways to achieve performance gain. In this paper, we gather the motivation behind the cross-layer design, illustrate some representative examples and draw conclusions for the future challenges.

General Terms

Design, Performance, Standardization, Wireless communication, Protocol architecture

Keywords

Cross-layer design, wireless networks.

1. INTRODUCTION

The TCP/IP protocol architecture has been the foundation behind the success and proliferation of Internet. As described in [1], sound architectures play an important role to build long-lasting standards that can be adopted in various levels. TCP/IP, or its theoretical counterpart OSI model is a prime example of a well defined standard. The whole communication flow from one entity to another is broken down according to tasks and divided among layers. Table 1 [2] lists the primary functions of the five layers. Each of the layers contributes to the whole information flow by completing its assigned task and passing the processed information to the upper/lower adjacent layer. The adjacent layers can communicate via predefined interfaces. Therefore, one layer does not need to know about the inner workings of the other layers as long as the services provided by the layers remains the same. As for a trivial example, a web browser residing in the application layer can work both with wired and wireless network connections as the internal intricacies of changing of the physical media are dealt with at the lower layers, which keep providing the expected services to the upper layers. As a result, a programmer designing a web browser can focus only on the application, barring aside tiny details of the network connection and the physical media. Thus, the layered design leads to an effective abstraction among layers which helps designers to create protocols for one layer independently of the others.

Table 1. TCP/IP layers and their primary tasks

Layer	Tasks
Application	User service: remote log-in, _file transfer, web access etc.
Transport	Port addressing, segmentation and reassembly, flow control, error control, connection control
Network	Logical addressing, routing
Data link	Framing, physical addressing, flow control, error control, access control
Physical	Physical characteristics of interface and media, representation and synchronization of bits, data rate

The modularity and independence of the layers deliver well for traditional wired networks, which were the only target at the time of designing the architecture. As we enter in the era of wireless networks, problems with the rigid architecture begin to emerge. Wireless networks, such as cellular networks and wireless local area networks (WLAN) bring new characteristics like high error rates, mobility, multipath propagation etc., which were not present in wired networks. Changes in the architecture were demanded by the fundamentally different characteristics of wireless medium than its wired counterpart. As for a classic trivial example in this regard, we can refer to the performance of TCP in wireless networks. Traditional TCP regards that packet loss happens due to congestion and applies congestion avoidance procedures whenever possible. In wireless networks, packet loss can occur even if there is no congestion, due to the characteristics mentioned above. As a result, performance of TCP significantly degrades in wireless networks, as summarized in [3].

Apart from the basic paradigm shift from wired to wireless networks, related qualitative changes were also imminent. On one side, a plethora of networks like ad hoc networks, personal area networks and wireless sensor networks emerged. On the other hand, users become mobile and share multimedia content in real time. Quality of service (QoS) has become a buzzword. All these factors demand more services from the protocol architecture. Hence, novel ideas come into being which not only depend on the defined interaction of adjacent layers but also use additional interfaces between adjacent and non-adjacent layers. These are the so called cross-layer designs, rightfully named after their fundamental property.

In this paper, a brief survey of the existing cross-layer designs is presented. The rest of the paper is organized as follows: section 2 gives a brief introduction to cross-layer design, section 3 presents an intuitive categorization of the ways

layers can interact, section 4 illustrates some case studies according to the common areas where cross-layer designs are applied. Section 5 presents some recommendations on how the quality of cross-layer designs can be measured. Section 6 concludes the paper.

2. BASICS OF CROSS-LAYER DESIGN

2.1 Definition

It is only fair to give a simple definition of the topic that is discussed throughout this paper. We can define cross-layer design as follows:

A design where the boundary among the protocol layers is violated by sharing internal information, helping layers to become aware of the changes in the others and hence provide higher quality of service to the user.

2.2 Cross-layer signaling methods

Obviously, cross-layer designs involve cross-layer signaling which is not defined in the protocol architecture. These signaling methods should consume as scarce resources as possible, reducing the overhead. In [6], the authors present a survey of the prevalent cross-layer signaling methods. They discussed about the following four methods of signaling:

Packet headers: Information can be encoded in layer headers which can later be used by some other layer to glean the desired information. This can be compared to have pipe like flow of signals among the layers.

ICMP messages: in IP based networks, Internet Control Message Protocol (ICMP) messages can be used for signaling. However, as ICMP messages are always encapsulated by IP packets, the messages have to traverse through the network layer, even if the interacting layer pairs are data link and physical or transport and application.

Network service: An additional storage in the network can be used to store the layer parameters. This storage can be managed by some distributed servers which gather channel and link states and serve the layers as needed.

Local profiles: This approach is similar to the previous one, but instead of using some other locations, layer parameters are stored and shared locally, inside the host.

2.3 Concerns about cross-layer designs

The fundamental principle of cross-layer design involves crossing the defined boundaries of a well-defined architecture. Sharing cross-layer information is a sensitive issue. The possibilities become endless when the layer boundary is crossed. In some cases it can lead to severe consequences which cannot be predicted beforehand. As cautioned in [1], *“unbridled cross-layer design can lead to spaghetti design which can stifle further innovation”*. The authors give an example of such situations where a rate-adaptive MAC protocol suffers from lower data rate when combined with minimum hop routing, as the minimum hop routing chooses longer hops having weaker signal strength, making the adaptive MAC protocol to opt for lower data rates. Another important aspect of any cross-layer design is its compatibility with the legacy designs. Achieving performance gain is of mere importance if the design cannot practically exist and

interact with other technologies.

A careful review of the related literature reveals that the cross-layer designers are usually reluctant to provide quantitative estimation of the overhead created by the design. They provide the performance gain resulting from the cross-layer design but many a times do not discuss on the additional parameters like delay, processing power and memory space, which are demanded by the new design. As the dependency among the layers is significantly altered, researches should exhaustively investigate the general scenarios and provide the performance results. Each of these factors should be kept in mind while designing and presenting a cross-layer design.

3. CATEGORIZATION OF CROSS-LAYER DESIGNS

3.1 Interaction between physical and link layer

The lower two layers are mainly responsible for setting up transmitting power, error correction, fragmentation of frames etc. Their functions are so closely related that each of the layers can make good use of the information available in the other.

Chien et al. [8] describe such a proposal where high level of adaptability is gained through using the cross-layer information. They provide adaptation of both link and physical layer. For link layer, they propose to vary the frame length at medium access control (MAC) layer according to the channel condition and also to vary the FEC/ARQ conditions according to bit-error rate (BER). For physical layer, they propose the adaptability of processing gain and equalization. Applied together, these adaptations bring significant improvement to battle the degraded channel conditions. In [9], a similar mechanism is proposed where the authors exploit the relation among frame length, transmit power and energy consumption. They cite the poor performance of TCP in wireless as a motivating factor and describe a combined tuning of physical and MAC layer. In [10], modifications in the data link layer are proposed so that the MAC frame lengths can be adapted according to the error rates in physical layer.

A good look at all the proposals mentioned above reconfirms the relation of physical and data link layer and shows how one can make one layer adaptive using information from another layer, namely exploiting the relation among BER, frame length and transmit power. This, as we will see, leads to more sophisticated levels of adaptation involving higher layers of the architecture.

3.2 Interaction of physical and link layer with the upper layers

These types of cross-layer design involve regulating various parameters in the upper three layers based on the lower two layer information. For example, [11] proposes cross-layer handoff management protocol (CHMP) which enhances the Mobile IP scheme for wireless networks. Here, link layer and network layer share information to decide the optimum timing for handoff. A handoff trigger unit is proposed which collects information from the link layer and the network layer. The link layer provides mobile terminal's speed information. The network layer at first discovers the neighboring base stations which may be involved in the forthcoming handoff. Then an

estimation of the signaling delays resulting the possible handoff to each of the stations is calculated beforehand. Taking these delays and the mobile terminal's speed in consideration, the handoff is initiated at a time which reduces the possibility of the packet loss due to handoff latency.

An example of layer-wide design is presented in [12] where cross-layering is used to build a framework for real-time video streaming in ad-hoc networks. Here, apart from the link layer and MAC layer interaction, which falls in the category described in the previous subsection, MAC layer and network layer worked jointly to determine the optimal network flow. These two layers share information to and out the network flow with minimal congestion. Moreover, the transport layer uses application layer delay constraints and link layer capacities to achieve a *congestion-distortion optimized scheduling* which avoids redundant transmission of packets.

Another common example in this category is the tuning of application layer parameters by getting information from the lower two layers and vice-versa. These two approaches are responsible for channel-adaptive application layer and application-priority aware link/physical layer, respectively. Interestingly, information flow in both these directions are exploited in proposals like [13] which incorporates both channel state and content priority information to offer better streaming of video in wireless networks.

3.3 Merging of layers

So far we have talked about cross-layer designs that use information from other layers. A more *revolutionary approach*, as named in [4], includes merging of more than one layer to achieve performance improvement. This type of approach is suitable to application areas which have a specific usage. For example, [14] proposes such a design for target tracking in wireless sensor network, which indeed serves a very specific purpose. The authors argue that target tracking, unlike traditional data communication, is not address-centric, rather event and location centric. Hence, they omit network and transport layer and create a three layer design with application, MAC and physical layer. Another example of this category is [15] where the authors compare performances of various possible cross-layer approaches in Bluetooth scatternets and show that the best performance is achieved by merging all the layers into a single one.

4. EXAMPLES OF CROSS-LAYER DESIGN AND CASE STUDIES

The discussion in the previous section provides a view of the cross-layer design from the layer or architecture perspective. It is more interesting to observe the problems and situations that demand those designs. In this section, we discuss such scenarios to which cross-layer designs are widely deployed. Cross-layers designs tend to exploit communication patterns with specific behavior that varies from traditional communication. Example of this type of design includes [14] where the location centric communication of the nodes is exploited. Another common area of cross-layer design is real-time multimedia transmissions which demand an efficient adaptation of the rate of transmission according to the link capability and vice versa. More about these examples follows in the next subsections.

4.1 Multimedia transmission over wireless networks

4.1.1 Scheduling of video streaming over UMTS

In [16], the authors propose a cross-layer design for Universal Mobile Telecommunication Systems (UMTS), which is one of the third generation mobile communication systems. Just as the other wireless networks, UMTS also suffers from high error rates in dealing with streaming video. The authors measure UMTS traffic in a real life network and analyzed the acquired data to and out interesting trends. As usual for wireless links, they found the errors in transmission to be bursty, grouped together in certain amounts of time, leaving wider time periods without any error. Moreover, they discover that the bursts of errors are periodic in nature. These patterns in the errors provide a certain amount of probable error-free time slots. Keeping this in mind, they categorize the video in different levels of priority. They work with H.264 video codec. In video compression terminologies, there are three types of frames: I-frames, P-frames and B-frames. While I-frames are independent, P and B frames have to be decompressed using their reference I-frames. Therefore, losing an I-frame during the transmission is much more costly than losing a B or P frame. Thus, the authors consider I-frames to be of higher priority than the other frames. Having knowledge about the error periodicity and priority of data, they apply cross-layer design. They send the error information from link layer to layer 3 (Radio network controller layer) so that the higher priority data (i.e. the I-frames) is scheduled to be sent at the error-free time. Their proposal suffers from some delay in transmission as it waits for the error-free slots to send the I-frames. However, they show this delay is tolerable as long as there is a sufficient amount of buffer present in the receiving terminal.

4.1.2 Congestion-optimized video scheduling

Congestion occurs when sender or receiver try to communicate more data than they are capable of handling. TCP deploys congestion control through its congestion window mechanism, which, as stated earlier, is not very effective in wireless networks. Setton et al. [17] used a mobile ad hoc network using 802.11 to implement a cross-layer scheduling for video transmission. They used a distortion model based on their earlier works. The distortion model helps the application layer to decide a balance between video encoding distortions and packet loss distortions. The application layer receives information about the fluctuations in network capacity from the transport layer and adjusts the video quality likewise. On the other hand, the transport layer depends on information from the application layer and takes decisions such as when to transmit and which packet to transmit. The transport layer also decides whether to drop some lower priority packets and transmit the more important ones.

4.1.3 Application-driven cross-layer optimization for video streaming

In [13], the authors propose a cross-layer architecture for video streaming. They call the approach an application-driven optimization as the main focus is to improve the user experience with using application layer information. A joint optimization of application, physical and data link layer is proposed.

For their experiment, the authors optimize physical, data link and application layer. One uncommon yet greatly desired characteristics of cross-layer design is covered by the authors: they provide an analysis of the overhead caused by the cross-layer modifications. The authors show that the design suffers from negligible amount of delay when lower number of video frames is sent but the delay increases linearly with the increase of video frames. In addition to that, they propose to handle only an effective subset of the abstracted parameters to reduce the processing overhead.

4.1.4 Cross-layer architecture for scalable video transmission

Huusko et al. [18] define a cross-layer framework for video transmission. The architecture contains a wired network from the sender to router and a wireless network as the last hop, which represents the possible mobile user. Additionally, the architecture regards only IPv6 because it heavily depends on IPv6's header structure, QoS support and routing abilities.

In the application layer, the streaming server adapts the bit rate and frame rate of the video at regular intervals based on the feedbacks from the lower layers. A fuzzy logic based controller is used for this purpose. Their resulting graphs show that the controller increases the bit rate and frame rate while the channel conditions remains satisfactory and decrease them accordingly. Besides this controller at the application layer, another level of adaptation is also proposed in data link layer. A NS-2 simulation shows that they are achieving fewer amounts of packet drops using their scheduler. However, their NS-2 simulation consists of a very simple case of two mobile nodes and they assume a presence of a large enough buffer at the transmitter to avoid congestion loss. As a matter of fact, real-life scenarios may not follow these assumptions. A more detailed level of simulation exploring diverse conditions would have provided a better idea on the effectiveness of their solution.

4.2 Ad hoc networks

4.2.1 MobileMan architecture for ad hoc wireless networks

Conti et al. proposed the MobileMan architecture [20], which provides a novel stack wide proposal of cross-layer design for mobile ad hoc networks.

Their architecture is compatible with the existing standards, allows a stack-wide context awareness and reduces duplicate data sharing. Though the MobileMan architecture proposed some novel ideas, they have not yet detailed the design of the Network Status component upon which all the cross-layer interaction depends. In a way, it is comparable to a single point of failure: if the Network Status is not designed in an efficient way, the performance gain would collapse drastically. In addition, they do not also describe the impact of overhead caused by all the cross-layer information shared through this component.

4.2.2 Congestion control

In [21], the authors propose a cross-layer design to control congestion in multi-hop ad hoc wireless networks. They start by referring to the traits of TCP in multi-hop networks: instability and unfairness. Instability occurs due to TCP's timeout mechanism which leads to opt for slow start, thus reducing both congestion window and achieved performance. Unfairness is created due to MAC layers exponential backoff

algorithm which provides preferential link access to the station that starts later.

Depicting flaws in the current solutions, the authors propose their cross-layer solution named "Cross-layer congestion control for TCP" (C³TCP). The performance of the proposed system is first analyzed with respect to a string topology. The results show that the proposed solution works on par with the traditional TCP when no cross-traffic is present. In presence of cross-traffic, the solution worked better and rightly keeps the throughput close to the available bandwidth. A comparison with TCP Vegas and TCP Westwood also shows improved performances in cross-traffic scenarios. The comparisons in a more complex grid topology showed similar results. However, as (C³TCP) also deploys a new module for cross-layer design, it is not quite clear how their solution is different from TCP Vegas and Westwood with respect to modifying traditional TCP.

5. DISCUSSION AND RECOMMENDATION

Cross-layer designs open possibilities where the variation in information available to each of the layers can be exploited to provide a level of adaptability which results in significant performance improvement. As for example, we have discussed how cross-layer designs are successfully applied to video transmission. This is particularly suitable because video streams have different kinds of frames with varying priority. In general, all the cross-layer designs make use of these priorities by adapting to the priority of the frame to be sent. The same is true for the designs which take feedback from lower layers and adapts the upper layers. They make use of the varying levels of network conditions and adapt the upper layer parameters accordingly. Hence, cross-layer designs provide us a way to utilize the variations in parameters. Cross-layer designs also provide opportunities to exploit specific characteristics of novel networks like wireless sensor networks or Bluetooth scatternets, where it may occur that some of the layers of the traditional protocol stack are creating additional overhead as for the particular nature of the communication.

However, cross-layer design is not without the drawbacks. As discussed throughout the previous sections, additional interdependence among layers can lead to potential problems. One way to defy these problems is to extensively analyze the designs, creating all the probable scenarios. Alarmingly, this is one very aspect where the researchers show reluctance. They focus to show the performance gain of the parameters deemed important by themselves, sometimes neglecting the wider scenario. As for example, we have seen an adaptive video transmission design where PSNR is used as the only measurement of the user perceived quality of the video. However, the relation between perceived quality and PSNR is highly reduced when the video content and codec type do not remain the same [24]. Examples like this are not scarce in the realm of cross-layer design. Hence, authors should strictly define the premise of the performance parameters that are to be improved by a cross-layer design. Additionally, most of the cross-layer design, pertaining to their novel nature, uses simulation studies to draw results. This may be a potential bottleneck as we know simulations do not always represent real-life network conditions.

Resulting from the aforementioned examples and discussions, we recommend the following four criteria against which a

cross-layer design should be measured. These four characteristics ensure that a cross-layer design is effective as well as verified and compatible.

Performance: Achieved performance gains due to the cross-layer design. This is the reason the cross-layer design is introduced in the first place.

Overhead: Computational overhead incurred by the design. The overhead should be quantified and demonstrated in approved standards. A comparison of performance against overhead is desired.

Resilience: Compatibility of the new design in varying environments. For example, a cross-layer design for wireless video streaming should provide improved performance not only in suitable conditions but also in congested scenarios.

Conflict susceptibility: Interaction behavior with other designs. For example, a cross-layer MAC may not work with some specific routing protocol. Probable scenarios involving common interactions should be investigated in depth.

6. CONCLUSIONS

In this paper, we have presented a brief overview of the cross-layer design. We presented the motivation behind it, categorized it and presented some case studies. Finally, we have outlined the measuring criteria that can be used to verify a cross-layer design from a holistic point of view. Cross-layer design was a demand of the time, to which researches have replied with great enthusiasm. It has been applied to various domains, achieving a significant performance gain and better quality of service. Nevertheless, one should keep in mind that adopting to cross-layer design does not mean to abolish the layered architecture, rather complement them. The recent trend shows an inclination towards cross-layer design. We can assume that in near future proposals which redefine the layered architectures completely will emerge. The focus should be on the compatibility of the new designs with the existing ones and keeping the additional overhead in control, achieving a balanced position to support the present ideas and be prepared for the upcoming ones as well.

7. ACKNOWLEDGEMENT

This work is derived from a seminar course at RWTH Aachen University distributed systems group, under the supervision of Mr. Florian Schmidt.

8. REFERENCES

- [1] V. Kawadia and P. Kumar, "A cautionary perspective on cross-layer design," *IEEE Wireless Communications*, pp. 3-11, 2005.
- [2] B. Forouzan and S. Fegan, *Data communications and networking*. McGraw-Hill Science Engineering, 2003.
- [3] G. Xylomenos, G. Polyzos, P. Mahonen, and M. Saaranen, "TCP performance issues over wireless links," *IEEE Communications Magazine*, vol. 39, no. 4, pp. 52-58, 2001.
- [4] F. Aune, "Cross-layer design tutorial," Technical report, Norwegian University of Science and Technology, Dept. of Electronics and Telecommunications, 2004.
- [5] S. Floyd, "TCP and explicit congestion notification," *ACM SIGCOMM Computer Communication Review*, vol. 24, no. 5, p. 23, 1994.
- [6] Q. Wang and M. Abu-Rghe, "Cross-layer signaling for next-generation wireless systems," in *IEEE WCNC*, vol. 2, pp. 1084-89, 2003.
- [7] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine*, vol. 43, pp. 112-119, December 2005.
- [8] C. Chien, M. Srivastava, R. Jain, P. Lettieri, V. Aggarwal, and R. Sternowski, "Adaptive radio for multimedia wireless links," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 5, pp. 793-813, 1999.
- [9] J. Ebert and A. Wolisz, "Combined tuning of RF power and medium access control for WLANs," *Mobile Networks and Applications*, vol. 6, no. 5, pp. 417-426, 2001.
- [10] P. Lettieri and M. Srivastava, "Adaptive frame length control for improving wireless link throughput, range, and energy efficiency," in *IEEE INFOCOM*, vol. 2, pp. 564-571, 1998.
- [11] S. Mohanty and I. Akyildiz, "A cross-layer (layer 2+ 3) handoff management protocol for next-generation wireless systems," *IEEE Transactions on Mobile Computing*, vol. 5, no. 10, pp. 1347-1360, 2006.
- [12] E. Setton, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," *IEEE Wireless Communications*, 2005.
- [13] S. Khan, Y. Peng, E. Steinbach, M. Sgroi, and W. Kellerer, "Application-driven cross-layer optimization for video streaming over wireless networks," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 122-13-, 2006.
- [14] L. Song and D. Hatzinakos, "A cross-layer architecture of wireless sensor networks for target tracking," *IEEE/ACM Transactions of Networking (TON)*, vol. 15, no. 1, pp. 145-158, 2007.
- [15] B. Raman, P. Bhagwat, and S. Seshan, "Arguments for cross-layer optimization in Bluetooth scatternets," in *Proceedings of Symposium on Applications and the Internet*, pp. 176-184, 2001.
- [16] W. Karner, O. Nemethova, P. Svoboda, and M. Rupp, "Link Error Prediction Based Cross-layer Scheduling for Video Streaming over UMTS," *15th IST Mobile & Wireless Communications Summit, Mykonos, Greece*, 2006.
- [17] E. Setton, X. Zhu, and B. Girod, "Congestion-optimized scheduling of video over wireless ad hoc networks," in *IEEE International Symposium on Circuits and Systems*, vol. 4 p. 3531, 2005.
- [18] J. Huusko et al., "Cross-layer architecture for scalable video transmission in wireless network," *Signal Processing: Image Communication*, vol. 22, pp. 317-330, 2007.
- [19] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications." <http://tool.ietf.org/html/rfc3550>. Accessed January 2010.

- [20] M. Conti, G. Maselli, G. Turi, and S. Giordano, "Cross-layering in mobile ad hoc network design," *Computer*, vol. 37, no. 2, pp 48-51, 2004.
- [21] D. Kliazovich and F. Granelli, "Cross-layer congestion control in ad hoc wireless networks," *Ad Hoc Networks*, vol. 4, pp. 687-708, 2006.
- [22] D. Kliazovich and F Granelli, "TCP Vegas: End to end congestion avoidance on a global Internet," *IEEE Journal on selected areas in communications*, vol. 13, no. 8, pp. 1465-1480, 1995.
- [23] S. Mascolo, C. Casetti, M. Gerla, M. Sanadidi, and R. Wang, "TCT Westwood: Bandwidth estimation for enhanced transport over wireless links," in *Proceedings of the 7th annual international conference on Mobile computing and networking*, p. 297, ACM, 2001.
- [24] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp 800-801, 2008.