Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Computer Communications 27 (2004) 720–724
computer communications
www.elsevier.com/locate/comcom

Cross-layer design optimizations in wireless protocol stacks

Vijay T. Raisinghani[a,b,*,1], Sridhar Iyer[b]

[a] *Tata Infotech (ATG), Bombay, India*
[b] *KR School of IT (IIT), Bombay, India*

Received 23 October 2003; accepted 24 October 2003

Abstract

The performance of applications on mobile devices is affected by the device constraints of memory, processing power, battery life and the variations in the wireless network. The variations in the wireless network will be compounded in the next generation networks—3G and beyond—when the devices move across heterogeneous networks. To allow interoperability with the Internet, existing standard protocol stacks would be deployed in the new networks and mobile devices. However, these protocol stacks which are *architected* and *implemented* in a layered manner do not function efficiently in mobile wireless environments. Cross-layer feedback in the protocol stack would be useful to improve the efficiency of these protocol stacks.

In this paper, we discuss the benefits of cross-layer feedback on the mobile device and present a representative survey.

Keywords: Cross-layer feedback; Protocol stack; Mobile device

1. Introduction

The new wireless networks—3G and beyond [1,2]—are expected to be all-IP and using standard protocol stacks, e.g. TCP/IP [3] (Transmission Control Protocol/Internet Protocol), to ensure interoperability.

The standard protocol stacks are *architected* [4] and *implemented* in a layered manner and function inefficiently in mobile wireless environments [5]. This is due to the highly variable nature of wireless links and the resource-poor nature of mobile devices. For wired networks some of the early work [6–8] shows that information exchange between protocol layers or *cross-layer feedback* is useful. Some examples of cross-layer feedback are: (1) TCP packet loss information communicated to the application layer to enable application adaptation, (2) link/MAC layer tuning the transmit power of the physical layer based on the bit-error rate information from the physical layer.

Cross-layer feedback optimizations may be implemented at the intermediate nodes [9,10] or mobile hosts (MH). We focus on cross-layer feedback in the MH since we believe that it would be easier to implement changes on the end-devices than in the network.

We believe that improving user satisfaction is the ultimate goal of improving application performance on wireless devices. Cross-layer feedback on a mobile device would help in improving application performance and thus user satisfaction. However, to further enhance user satisfaction it is essential to incorporate dynamic user requirements also into the protocol stack. For example, it would be useful to allow a user to indicate and change application priorities dynamically. Further, since a large fraction of the battery is consumed by the network interface [11], it is imperative that the various protocol layers adapt and collaborate to optimize power consumption.

In Section 2 we discuss the various cross-layer feedback possibilities and the potential benefits, in Section 3 we present related surveys and in Section 4 we present the summary and conclusion.

2. Cross-layer feedback

Cross-layer feedback means interaction among the layers in the protocol stack. For the sake of convenience, cross-layer feedback can be categorized as follows:

* Corresponding author. Address: KR School of IT (IIT), Bombay, India.
E-mail addresses: rvijay@it.iitb.ac.in (V.T. Raisinghani), sri@it.iitb.ac.in (S. Iyer).
[1] He is a PhD student sponsored by Tata Infotech Ltd.

0140-3664/$ - see front matter
doi:10.1016/j.comcom.2003.10.011

Upper to lower layers. For example, delay or loss constraints of the application communicated to the link layer to enable the link layer to adapt its error correction mechanisms; user defined application priority communicated to TCP to increase the receiver window of the application with a higher priority.

Lower to upper layers. For example, TCP packet loss information given to the application layer so that the application can adapt its sending rate; physical layer transmit power and bit-error rate information communicated to the link/MAC layer to enable adaptation of error correction mechanisms.

In Section 2.1 we present examples of cross-layer feedback for each layer. We also discuss the benefits and indicate some of the disadvantages.

2.1. Physical layer

The function of physical layer is to transmit raw bits over a certain distance with minimum bit errors, using a suitable power level.

The information available at the physical layer is transmit power, bit-error rate and coding/modulation in use. However, for *explicit* measurement of bit-error rate the receiver will have to provide feedback to the sender.

2.1.1. Interaction with upper layers

Application or user: The application layer or the device user may tune the physical layer parameters to improve throughput or download software for another physical layer [12] (see Section 2.5 also). However, one disadvantage is that software downloading itself could consume high amounts of power.

Network: The bit-error rate on an interface could be used as a guide by the network layer to select the appropriate interface.

Link/MAC: Ebert and Wolisz [13] discuss *protocol harmonization* for MAC and physical layer for IEEE 802.11 [14]. They investigate the effects of packet length, transmit power and bit-error rate. Their results show that minimum energy is consumed for transmission if an *optimal* transmit power is used for a packet. Further, this optimal transmit power is proportional to the packet length. Also, they show that varying the packet length according to the BER also helps reduce energy consumption. They report that fragmentation into packets of size 500 bytes for a BER $> 10^{-5}$ leads to the largest reduction in energy consumption (also see Section 2.2).

Battery aware physical layer. The physical layer may also adapt its coding/modulation depending on the battery status.

2.2. Link/MAC layer

The functions of link/MAC layer are improving link reliability through forward error correction (FEC) and Automatic Repeat reQuest (ARQ); avoiding/reducing collisions; fragmenting data into frames so as to ensure reliable transmission with minimal overhead.

The information available at the link/MAC layer is current FEC scheme, number of frames retransmitted, frame length, point in time when the wireless medium is available for transmission and hand-off related events.

2.2.1. Interaction with upper layers

User: Link throughput information can indicate to the user the kind of application performance that should be expected. The user may then decide which applications can be run.

Application: At the link/MAC layer the frames from different applications may be treated differently. For example, frames of applications with a low delay requirement may be transmitted on priority. Similarly, FEC/ARQ may be improved for applications with a high reliability requirement. The above is based on the idea of a *multi-service link layer* [15], for QoS (Quality of Service, i.e. delay, loss, jitter requirements of applications) in the Internet, which adapts the link layer services based on the traffic class. However, such schemes may increase the processing overhead and hence power consumption.

Transport: When channel conditions are poor, retransmissions at the link layer result in delays which could lead to TCP retransmissions and thus reduced throughput [16]. To avoid this, TCP and link layer could exchange retransmission information. In a IEEE 802.11 [14] environment [17] show that increasing MAC level retransmissions to avoid TCP retransmissions, decreases the power consumption.

Network: Mobile-IP [18] is used for IP hand-off whenever the mobile device changes sub-nets. Hand-off in Mobile-IP depends on the detection of a network change at the IP layer. This information may not be available as quickly as the signal strength changes monitored continually at the link layer. Thus, link layer hand-off information can be used to reduce the hand-off latency for Mobile-IP [19,20]. On similar lines, IP micro-mobility protocol, Cellular-IP [21] uses signal strength of the base station beacons for hand-offs.

2.2.2. Interaction with lower layers

Physical: Based on current channel conditions the error control mechanisms at the link layer may be adapted to reduce the transmission errors [22,23]. Lettieri and Srivastava [22] show around 50% improvement in goodput and 20% improvement in transmission range by using the optimal Maximum Transmission Unit (MTU) for a particular BER. In a GSM case study, Ludwig et al. [23] show that by increasing the frame length the throughput can be increased by 18–25%, depending on the radio conditions.

For tuning physical layer power see Section 2.1.

Batter aware link/MAC layer. See Section 2.1 for optimization in collaboration with physical layer.

From the examples in the previous sections, it can be seen that adaptation of error control mechanisms at the link/MAC layer along with transmit power control at the physical layer can help in substantial reduction in power consumption and improvement in throughput.

2.3. Network layer

Network layer functions are routing, addressing, selecting the network interface, and IP hand-off [18] to maintain IP connectivity in *foreign* networks.

The information available at the network layer is Mobile-IP hand-off initiation/completion events and the network interface currently in use.

2.3.1. Interaction with upper layers

Application or user: An application could control its sending rate based on Mobile-IP hand-off indications.

A device may have multiple wireless network interfaces that can provide different levels of service. For example, a wireless LAN interface may provide lesser delays and higher throughput as compared to a GPRS interface on the same device. Depending on the application or user needs, the network layer could select an appropriate network interface. However, continuing a session uninterrupted onto another interface is an open research area.

Transport: Mobile-IP hand-off delay may lead to reduced throughput due to the TCP retransmission time-out (RTO) and back-off mechanism. TCP can be informed about the event of Mobile-IP hand-off to reduce the retransmission latency. A *fast retransmit* [24] can be initiated on the MH using this information. Depending on the hand-off conditions, this helps in reducing TCP retransmission latency by up to 75% and improving throughput by up to 25% [24].

2.3.2. Interaction with lower layers

Link/MAC: see Section 2.2; Physical: see Section 2.1.

From the aforementioned, it seems that the Mobile-IP hand-off indications to the application and TCP would be quite useful in conserving the battery and increasing throughput.

2.4. Transport

The transport layer is concerned with establishing end-to-end connections over the network. Mobile networks are characterized by large delays, packet losses and high bit-error rates. Transport protocols like TCP interpret this as a congestion loss which reduces its throughput [24]. Cross-layer feedback may also be beneficial in case of protocols like UDP or RTP, however, due to space limitations we restrict our discussion to TCP.

The information available with TCP is round-trip time (RTT), RTO, MTU, receiver window, congestion window, number of packets lost and actual throughput (or goodput).

2.4.1. Interaction with upper layers

User: A user may assign priorities to the running applications. In case the applications are downloading some data, this higher priority would indicate the need for higher download bandwidth. To enhance user satisfaction, TCP may map the higher priority of an application to a larger receive window [25,26]. Further, a user could provide information about an impending disconnection. This information can be used by TCP to increase its RTO values (also see Section 2.6). Also, TCP may provide packet loss and goodput information to the user. The user may shutdown some non-critical applications based on this input. This information may probably be crucial for enhancing user satisfaction.

Application: Applications may indicate their QoS requirements to TCP. Based on this information TCP may manipulate the receiver windows. On the other hand, TCP may provide packet loss and goodput information to the application. The application can use this input to adapt its sending rate.

2.4.2. Interaction with lower layers

Network: see Section 2.3; Link/MAC: see Section 2.2.

2.5. Application

The application layer is the interface to the user for running user tasks. For example, web browsing, downloading a file using FTP, sending e-mail, watching a video clip, etc.

The existing applications were designed for wired networks and do not perform well in wireless networks. Application adaptation based on information from lower layers would be useful in improving application performance over wireless networks.

An application layer can communicate to other layers the application's QoS needs, i.e. the delay tolerance, acceptable delay variation, required throughput and acceptable packet loss rate.

2.5.1. Interaction with upper layers

User: A user's requirement can be captured by an application and communicated to the lower layers. The mobile device could then be reconfigured to satisfy user needs [12].

2.5.2. Interaction with lower layers

Transport: see Section 2.4; Network: see Section 2.3; Link/MAC: see Section 2.2.

Physical: Multi-media applications like video, use various standard coding techniques for video transmission. Information about channel conditions can be used to adapt the coding. For example, if the bandwidth is low, a lower quality video coding may be used which requires lesser bandwidth. Similarly, an e-mail application could defer downloading the file attachments in an e-mail when the channel conditions are poor. Some of the proposals for application adaptation are Refs. [27–29].

Power saving based on application information. If the application can tolerate some delays, it may be possible to switch off the network interface card intermittently [30]. Information about the type of coding used by a video-application could be used to discard some frames at the network interface to save power [31]. However, this will reduce the video quality.

From the previous discussion, probably information about channel conditions from the physical and link/MAC layers would be crucial in improving application performance. Also, tuning of the link layer error control mechanisms based on the application QoS requirements, seems to be essential in improving application throughput.

2.6. User

We consider the user to be the uppermost layer of the protocol stack. We believe that user requirements should be taken into account to enhance *user perceived QoS*. The motivation for this is that the user decision could be contrary to the system decision but it could lead to improved user satisfaction. For example, (1) for a user a FTP download may be more important than a streaming video, (2) a user may know that a disconnection is imminent in an approaching tunnel while the system will *know* it only after the signal is affected, (3) the system may decide to conserve battery by not downloading some information while the user may, at that instant, feel that the information (e.g. a *stock quote*) is more important than saving battery.

The user will need information from the lower layers to use the mobile device effectively. It seems that the most crucial one will be link throughput information from the link layer. This will help the user decide about the applications that can be run and also indicate to the user the kind of performance that should be expected.

2.6.1. Interaction with lower layers

Application: see Section 2.5; Transport: see Section 2.4; Network: see Section 2.3; Link/MAC: see Section 2.2; Physical: see Section 2.1.

Battery status. Depending on the battery status, the user may instruct the system to optimize power consumption sacrificing performance and vice versa.

In this section we presented the different possibilities of cross-layer feedback with a discussion about the benefits. In Section 3 we present the related surveys.

3. Related surveys

Jones et al. [32] present a survey of work addressing energy efficient and low-power design within all layers of the wireless network protocol stack. Zorzi and Rao [33] discuss the impact of higher order error statistics on the various layers of the protocol stack. Power aware protocols in ad hoc networks are discussed in Ref. [34]. References therein provide insight into the various power aware protocol proposals and design issues. Badrinath et al. [35] present a conceptual framework for network and client adaptation. They survey the various proposals for application adaptation and map it to the conceptual framework. Barakat et al. [36] present a survey on the various suggestions to improve TCP behavior over heterogeneous networks.

4. Summary and conclusion

Many surveys predict that in the next few years the number of wireless devices accessing the Internet will far exceed the number of wired devices. The need for a paradigm shift from strictly layered protocol stacks to cross-layer feedback is clear from the benefits of cross-layer feedback.

Although, a number of cross-layer ideas hold promise, we believe that *user feedback* and *battery constraints* may drive the algorithms for cross-layer interactions. Further research efforts are needed to combine subjective user inputs with the quantitative inputs from the various layers to enhance user satisfaction in wireless networks.

References

[1] A. Jamalipour, S. Tekinay (Eds.), Fourth Generation Wireless Networks and Interconnecting Standards, Vol. 8 of IEEE Personal Communications, 2001.

[2] UMTS Forum, Glossary, http://www.umts-forum.org/glossary.asp (2003).

[3] W. Richard Stevens, TCP/IP Illustrated, vol. I, The Protocols, AWL, 1994.

[4] ITU, Information Technology—OSI—Basic Reference Model, X.200, 1994, July.

[5] G. Xylomenos, G.C. Polyzos, Internet protocol performance over networks with wireless links, IEEE Network 13 (4) (1999) 55–63.

[6] D.D. Clark, The Structuring of Systems using Upcalls, in: ACM Symposium on Operating Systems, 1985, pp. 171–180.

[7] G.H. Cooper, The Argument for Soft Layer of Protocol, Technical Report TR-300, Massachussets Institute of Technology, Cambridge, MA, 1983, May.

[8] D.D. Clark, D.L. Tennenhouse, Architectural Considerations for New Generations of Protocols, in: ACM SIGCOMM, Philadelphia, PA, 1990, pp. 200–208.

[9] H. Balakrishnan, V.N. Padmanabhan, S. Seshan, R.H. Katz, A comparison of mechanisms for improving TCP performance over wireless links, IEEE/ACM Transactions on Networking 5 (6) (1997) 756–769.

[10] P. Bhagwat, P. Bhattacharya, A. Krishna, S.K. Tripathi, Using channel state dependent packet scheduling to improve TCP throughput over wireless LANs, Wireless Networks 3 (1) (1997) 91–102.

[11] M. Stemm, R.H. Katz, Measuring and reducing energy consumption of network interfaces in hand-held devices, IEICE Transactions on Communications E80-B (8) (1997) 1125–1131.

[12] A. Aghvami, T. Le, N. Olaziregi, Mode switching and QoS issues in software radio, IEEE Personal Communications 8 (5) (2001) 38–44.

[13] J.-P. Ebert, A. Wolisz, Combined tuning of RF power and medium access control for WLANs, Mobile Networks and Applications 6 (5) (2001) 417–426. special issue on Mobile Multimedia Communications (MoMuC'99).

[14] IEEE Std 802.11-1997, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (18 November 1997).

[15] G. Xylomenos, G.C. Polyzos, Quality of service support over multiservice wireless internet links, Computer Networks 37 (5) (2001) 601–615.

[16] A. DeSimone, M. Chuah, O. Yue, Throughput Performance of Tranport-layer Protocols Over Wireless Lands, in: IEEE GLOBECOM, 1993.

[17] M. Methfessel, K.F. Dombrowski, P. Langendörfer, H. Frankenfeldt, I. Babanskaja, I. Matthaei, R. Kraemer, Vertical optimization of data transmission for mobile wireless terminals, IEEE Wireless Communications 9 (6) (2002) 36–43.

[18] C. Perkins, IP Mobility Support for IPv4. RFC3344, 2002, August.

[19] A. Sanmateu, F. Paint, L. Morand, S. Tessier, P. Fouquart, A. Sollund, E. Bustos, Seamless mobility across IP networks using mobile IP, Computer Networks 40 (1) (2002) 181–190.

[20] J.C.-S. Wu, C.-W. Cheng, N.-F. Huang, G.-K. Ma, Intelligent handoff for mobile wireless Internet, in: R. Cáceres, L.F. Chang, R. Jain (Eds.), Special Issue on Wireless Internet and Intranet Access, vol. 6, Kluwer, Dordrecht, 2001.

[21] A.G. Valk, Cellular IP: a new approach to Internet host mobility, ACM SIGCOMM Computer Communication Review 29 (1) (1999) 50–65.

[22] P. Lettieri, M.B. Srivastava, Adaptive Frame Length Control for Improving Wireless Link Throughout, Range and Energy Efficiency, in: INFOCOM, vol. 2, 1998, pp. 564–571.

[23] R. Ludwig, A. Konrad, A.D. Joseph, R.H. Katz, Optimizing the end-to-end performance of reliable flows over wireless links, Wireless Networks 8 (2/3) (2002) 289–299.

[24] R. Cáceres, L. Iftode, Improving the performance of reliable transport protocols in mobile computing environments, IEEE Journal on Selected Areas in Communications 13 (5) (1995) 850–857.

[25] V.T. Raisinghani, A.K. Singh, S. Iyer, Improving TCP Performance over Mobile Wireless Environments using Cross-Layer Feedback, in: IEEE ICPWC, New Delhi, India, 2002.

[26] V.T. Raisinghani, S. Iyer, User Managed Wireless Protocol Stacks, 23rd ICDCS, Poster, 2003.

[27] B.D. Noble, M. Satyanarayanan, D. Narayanan, J.E. Tilton, J. Flinn, K.R. Walker, Agile Application-Aware Adaptation for Mobility, in: 16th ACM Symposium on Operating System Principles, ACM, St Malo, France, 1997.

[28] A. Alwan, R. Bagrodia, N. Bambos, M. Gerla, L. Kleinrock, J. Short, J. Villasenor, Adaptive mobile multimedia networks, IEEE Personal Communications 3 (2) (1996) 34–51.

[29] H. Liu, M.E. Zarki, Adaptive Source rate control for real-time wireless video transmission, Mobile Networks and Applications 3 (1) (1998) 49–60.

[30] R. Kravets, P. Krishnan, Application-driven power management for mobile communication, Wireless Networks 6 (4) (2000) 263–277.

[31] P. Agrawal, S. Chen, P. Ramanathan, K. Sivalingam, Battery Power Sensitive Video Processing in Wireless Networks, in: IEEE PIMRC, Boston, 1998.

[32] C.E. Jones, K.M. Sivalingam, P. Agrawal, J.C. Chen, A survey of energy efficient network protocols for wireless networks, Wireless Networks 7 (4) (2001) 343–358.

[33] M. Zorzi, R. Rao, Perspectives on the impact of error statistics on protocols for wireless networks, IEEE Personal Communications 6 (5) (1999) 32–40.

[34] A. Goldsmith, S. Wicker, Design challenges for energy-constrained ad hoc wireless networks, IEEE Wireless Communications 9 (4) (2002) 8–27.

[35] B. Badrinath, A. Fox, L. Kleinrock, G. Popek, P. Reiher, M. Satyanarayanan, A conceptual framework for network and client adaptation, Mobile Networks and Applications 5 (4) (2000) 221–231.

[36] C. Barakat, E. Altman, W. Dabbous, On TCP performance in a heterogeneous network: a survey, IEEE Communications Magazine 38 (1) (2000) 40–46.