OPTIMIZING WIRELESS MULTIMEDIA TRANSMISSIONS THROUGH CROSS LAYER DESIGN

Haitao Zheng

Wireless Technology Research, Bell-Labs, Lucent Technologies, U.S.A

ABSTRACT

In this paper we have addressed some of the key design issues associated with providing multimedia service over wireless networks. Adaptation to continuously changing wireless environments was achieved through a cross layer design framework, which promotes communication and interaction across multiple protocol layers. We have also presented two innovative processes to improve the performance of video conferencing and streaming video over wireless network in terms of adaptation and cross layer design.

1. INTRODUCTION

Given that the Internet traffic increases dramatically and users desire ubiquitous Internet access, the next generation of networking systems will be data-centric with the addressed mobility consideration. As a result, the circuit-switched 1G and 2G wireless networks have gradually evolved into packet-switched 2.5G technologies such as GPRS, EDGE and 1X-EVDV to provide packet data services and further improve voice capacity, which will eventually be phased out by the 3G/4G wireless technologies. In recent years, enormous amount of efforts has been to support IP in wireless networks. Protocols and programming languages including WAP, WML and J2ME have been developed to adapt Web content to the limitations of handhold devices. Mobile IP networks have been designed to maintain consistent transport-layer quality as the remote terminal is constantly in motion. However, in developing IP-based wireless data networks, significant difficulties remain to be addressed, as they are summarized next.

Dynamic link characteristics - The process of a mobile device transmitting and receiving radio signal through the air makes wireless transmission vulnerable to noise and interference. The shadowing effect, multi-path fading and interference from the other devices make channel conditions varying unpredictably over time. Changing the transmission rate as the channel varies does improve efficiency but results in data rate oscillation. Furthermore, mobility introduces difficulty in channel estimation and prediction, thus raises error rate. Two approaches have commonly been used to address this problem. The first approach employs sophisticated channel coding and interleaving technologies. This approach, however, heavily relies on the quality of channel estimation. The second approach, the link layer ARQ mechanism performs error control by retransmitting lost frames[1]. Although insensitive to the quality of channel estimation, this approach introduces latency and delay jitters to IP packet flow. The tradeoff between latency and reliability depends on the ARQ persistence, which defines the willingness of the protocol to retransmit lost frames to ensure reliable transmission[2]. The persistence can be expressed in terms of time or the maximum number of retransmissions.

Asymmetric data rate - Mobile terminal has limited power so that the uplink data rate is usually less than the downlink data rate. This limitation is less stringent since most data applications are asymmetric.

Resource Contention - As in wireline networks, users share channel resources in wireless networks. When multiple users run a variety of applications, the most salient issue is the significant variability in terms of Quality of Service (QoS) requirement such as error rate, latency and bandwidth. The resource contention problem is already quite challenging in wireline networks. As the result of mobility and unpredictable link variation, dynamic network topology makes wireless networks even harder to coordinate. The Medium Access Control (MAC) layer uses a scheduler to determine the next user to be served based on individual user's channel condition and QoS requirement[3]. Currently, this scheduler is developed only for downlink transmission because only the base station gathers all the user information. The uplink transmission is typically made through contention, yielding high delay jitters.

Overall, high transmission errors and variable latency are the major causes of data loss in wireless networks. For wireline networks, link and sub-networks normally have relatively stable transmission rate at low error rate. Data loss is primarily due to network congestion and buffer exhaustion. As such, many techniques have been developed to support efficient packet transmission over wireline networks. Unfortunately they are not applicable to wireless networks. For example, in wireline networks, adding bandwidth can solve latency problem since bandwidth is not a paramount concern. However, in wireless environment, this is quite difficult due to adverse channel condition and limited battery life of the mobile device.

2. ADAPTATION THROUGH CROSS LAYER DESIGN

An important aspect of wireless networks is dynamic behavior. The conventional protocol structure is inflexible as various protocol layers can only communicate in a strict manner. In such a case, the layers are designed to operate under the worst conditions, rather than adapting to changing conditions. This leads to inefficient use of spectrum and energy. Adaptation represents the ability of network protocols and applications to observe and respond to the channel variation. Central to adaptation is the concept of cross layer design [4][5]. Cross layer design for the three key layers in the overall protocol stack (i.e., application-layer, transport-layer, and network- and link-layer) are reviewed in this section.

Application Transmission Adaptation - This approach refers to the application's capability to adjust its behavior to changing network and channel characteristics. Wireless networks often have to deal with adverse conditions where handoffs, deep fading, and bad carrier signals result in high rate of packet losses. Only adaptive applications can cope with these challenging circumstances. For multimedia delivery, a media server can track packet losses and adjust media source rate and protection level accordingly [6][7][8][9][10]. Whereas this level of adaptation is system independent and application specific, an application is able to reconfigure itself accurately only if it could identifies the underlying network and channel variations.

Transport-layer Transmission Adaptation - Instead of application-layer adaptation, the adaptation can be shifted to the underlying transport layer, making it transparent to the application layer, so that applications originally developed for wireline networks remain intact. One drawback of this level of adaptation is that it is impossible to implement a complete adaptation if part of it is application specific. The protocol should differentiate various packet loss patterns (i.e., packet losses due to network congestion or from channel errors[11][12][13]) and invoke congestion control and rate adaptation accordingly. TCP and its variants provide reliable connections by retransmitting the lost packets. However, the resulting latency is in general too large for real-time and streaming media applications. For this reason, most streaming applications use UDP protocol with an unreliable packet delivery. However, by discarding corrupted packets, UDP does not distinguish between packet losses due to congestion and corruption. Alternatively, UDP-Lite applies partial checksum to some parts of a packet (i.e., packet payload) and reduces packet loss rate [14]. CUDP conducts a precise error detection and recovery through error location information from link-layer [15].

Network- and Link-layer Transmission Adaptation - The application characteristics, such as QoS requirement and packet priority, could be used in coordinating the link- and network-layers. In particular, the persistence level of the link layer ARQ mechanism should adapt to each application's latency and reliability requirements, while the link layer scheduler allocates radio resources to various packet flows based on their QoS priorities. The adaptation, however, requires the link- and network-layers to be able to distinguish different packet flows, which in general can be achieved by an explicit indication of the QoS requirement associated with each packet flow[2]. Note that in some systems, the transport- and link-layer both conduct error recover by using FEC coding and retransmissions. The balance between both schemes is important for the optimal usage of the overall communication resources[16]. Meanwhile, the network could operate efficiently by using the link- and physical-layers information, such as rate fluctuation and error condition, to distribute channel resources.

In addition, the adaptation relies on each layer's ability to estimate current and even predict future network and channel conditions. Within a protocol stack, the link-layer must detect its present status including link availability, congestion and error conditions, and signal it to upper layers for appropriate adaptation[15]. A proper form of the information exchange across multiple layers is crucial to the effectiveness of the adaptation.

3. INTEGRATING THE ADAPTATION FOR WIRELESS MULTIMEDIA TRANSMISSION

In general, in building an efficient wireless network, we strive to create a series of protocol layers that communicate, interact and thus yield continuously improved applications and services. Next, we will highlight some of the innovative processes to improve the performance of video conferencing and streaming video over wireless network in terms of adaptation and cross layer design.

Case I: Optimizing Link Layer Scheduler, Hybrid ARQ and Link Adaptation for Wireless Video Conferencing

Channel characteristics play an important role in determining the amount and the type of data traffic that can be carried with sufficient QoS over the channel. This is especially true for multi-user wireless systems where multiple users share the bandwidth, and each user's channel varies independently due to interference and fading. Scheduling algorithms[17][18] can exploit multi-user diversity due to independent channel variations, and thus are crucial to system performance. Certain scheduling algorithms designed for shared packet channel, e.g. maximum rate scheduling, only consider the overall system throughput, so that the transmission would take place to those mobile devices with

good channel condition. However, real-time applications like video conferencing require timely packet delivery. Thus, allowing transmissions to take place only when the channel condition is good may decrease the percentage of packets being delivered in time. Here lies the tradeoff n which the scheduler must decide whether it is worthwhile to transmit to a particular mobile device even through that decision may decrease the overall channel capacity.

Link adaptation[19], which continuously adjusts the modulation and coding schemes (MCS), provides an efficient way of maximizing the instantaneous usage of wireless channel. It enables the use of spectral-efficient high-order MCSs when channel conditions are favorable, and reverts to the MCSs that are more robust but with lower transmission rates when channel conditions degrade. Traditionally, the MCS selection is adjusted based on the channel quality feedback to maximize the throughput while maintaining certain target FER. For voice transmission, the mapping is selected to maintain a 1% FER, while for data transmission the mapping is usually generated to maximize the (averaged) throughput taken into account hybrid ARQ operation. For real-time data applications, the mapping design should consider both packet delay and throughput.

We conduct Opnet based network simulations to evaluate the optimization of scheduler, link adaptation and HARQ scheme for wireless video conferencing over UMTS High-Speed Downlink Packet Access (HSDPA)[20]. A single base station supports three user equipments(UE). The average signal-to-interference-and-noise ratio(SINRs) for UE 0, 1 and 2 are 3dB, 2dB and −1dB, so that in average UE 0 is the best user and UE 2 is the worst. The downlink packet delay is measured from the time a packet is generated at the wire-line videoconference party to the time the packet is received by the wireless videoconference party. We only consider downlink direction at this time and ignore uplink performance. The cumulative distribution function (CDF) of packet delay corresponding to proportional fair(PF) and maximum rate(MAX) schedulers are shown in Fig. 1. The mapping designs are 1% FER based and HARQ fitted mapping design. The latter transmits more aggressively and requires retransmissions. We observe that MAX scheduler results in performance imbalance between good user (UE 0) and bad user(UE 2), while PF scheduler improves user fairness. Fig. 2 shows that HARQ fitted mapping helps bad user (UE 2) while 1% FER based mapping design favors good users. Hence, appropriate aggressiveness is essential to system performance and should be selected carefully based on application QoS requirements. Fig. 3 compares the UE 2 delay performance with round robin(RR), PF and MAX schedulers, where PF and MAX are significantly robust compared to RR. This also indicates that the choice of scheduler is quite important for delay sensitive applications.

Case II: Modified UDP with Link Layer Error Indicator for Streaming Video over Wireless

UDP is generally used for video streaming. However, UDP is unable to distinguish between packet losses caused by network congestion and by channel errors. For this reason, it is more appropriate to use UDP over wireline networks than over wireless networks. UDP-Lite, on the other hand, ignores channel errors unless they corrupt packet header. By doing so, it shift the error handling responsibility to the application. When packet level FER coding is deployed to provide error control, CUDP is superior to UDP-Lite because CUDP utilizes error indication from the link layer. In general, the transmission unit at link layer is smaller than that at network- and transport-layers, so that link layer error indications provide a precise estimation of the error location. A packet FEC decoder can use the error locations to group erroneous data blocks to erasures and double the error recovery capability. Without packet FEC coding, the error indication is still beneficial because it can assist video decoder to locate errors by formatting the corrupted link layer unit as all "1"s.

The performance of UDP, UDP-lite and CUDP was compared in [15] in terms of streaming MPEG video through a UMTS-similar system, where CUDP achieves 2-6 dB of PSNR improvement over UDP, and 5-10 dB over UDP-Lite. As congestion packet loss increases, the advantages diminish because network congestion becomes the dominant impairment.

4. CONCLUSIONS

This paper provided an overview of the intelligent video over wireless networks. Adaptation to continuously changing wireless environments was achieved through a cross layer design framework, which promotes communication and interaction across multiple protocol layers.

REFERENCES

[1] Y. Wang and S. Lin, "A modified selective-repeat type-II hybrid ARQ system and its performance analysis," *IEEE Transactions on Communications*, vol. COM31, pp. 593-- 608, May 1983.

[2] Advice to link designers on link Automatic Repeat reQuest (ARQ), Internet Draft, Mar. 2002, draft-ietf-pilc-link-arq-issues-04.txt.

[3] Jalali, R. Padovani, R. Pankaj, "Data throughput of CDMA-HDR: A high efficiency high data rate personal communication wireless system," Proceedings of the IEEE Vehicular Technology Conference, Tokyo, Japan, May 2000.

[4] "A Multilayered Approach to Mobile Networking", Stanford University Project Report.

[5] L.Tong, Q. Zhao and G. Mergen, "Multipacket reception in random access wireless networks: From signal processing to

optimal medium access control", IEEE Communications Magazine V39, 11, NOV, 2001, p108-112

[6] B.Girod and N.Färber, "Wireless Video", in Compressed video over network, edited by A.Reibman and M.T.Sun, Marcel Dekker, 2000.

[7] M. van der Schaar and H. Radha, "Unequal Packet Loss Resilience for Fine-Granular-Scalability Video", IEEE Trans. on Multimedia, December 2001.

[8] H. Liu and M.E. Zarki, "Adaptive source rate control for real-time wireless video transmission", Mobile Networks and Applications, 3:49-60, 1998.

[9] S. Aramvith, I.-M. Pao, and M.-T. Sun, "A rate-control scheme for video transport over wireless channels," IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, pp. 569--580, May. 2001.

[10] C.Y. Hsu, A. Ortega, M. Khansari, "Rate control for robust video transmission over burst Error wireless channels", IEEE Journal on Selected Areas in Communications V17, 5, MAY, 1999, p756-773.

[11] S.Cen, P.C.Cosman and G.M. Voelker, "End-to-end differentiation of congestion and wireless losses", SPIE MMCN2002, San Jose, CA, Jan. 18-25, 2002.

[12] N.K.G. Samaraweera, "Non-congestion packet loss detection for TCP error recovery using wireless links", IEEE Proceeding on Communications, vol. 146, No. 4, Aug. 1999.

[13] S.Biaz and N.Vaidya, "Discriminating congestion Losses from Wireless Losses using Inter-Arrival Timers at the Receiver", Technical Report 98-014, Computer Science Department, Texas A&M University, June 1998.

[14] L.Larzon, M.Degermark and S.Pink, "UDP Lite for Real Time Multimedia Applications", Proceedings of the QoS mini-conference of IEEE International Conference of Comm. (ICC'99), Vancouver, Canada, June 1999.

[15] H. Zheng and J. Boyce, "An Improved UDP Protocol for Video Transmission over Internet-to-Wireless Networks", IEEE Trans.on Mult., Vol.3, No.3, Sept. 2001, pp356-364.

[16] A. Chockalingam and G. Bao, "Performance of TCP/RLP Protocol Stack on Correlated Fading DS-CDMA Wireless Links",IEEE Trans. On Veh. Tech.,Vol. 49, pp.28-33, Jan. 2000.

[17] T. Enderes, S. C. Khoo, C.A. Somerville and K. Samaras, "Impact of statistical multiplexing on voice quality in cellular networks", Mobile Networks and Applications, pp. 153-161, vol. 7, iss. 2, April, 2002.

[18] A. Jalali, R. Padovani, R. Pankaj, "Data throughput of CDMA-HDR: A high efficiency high data rate personal communication wireless system," Proceedings of the IEEE Vehicular Technology Conference, Tokyo, Japan, May 2000.

[19] K. Blachandran, S.R. Kadaba and S. Nanda, "Channel quality estimation and rate adaptation for cellular mobile radio", IEEE Journal on Selected Areas on Communications, vol. 17, no. 7, pp. 1244-1256, July 1999.

[20] 3GPP TSGRAN TR 25.848, Physical Layer Aspects of UTRA High Speed Downlink Packet Access, Release 2000.
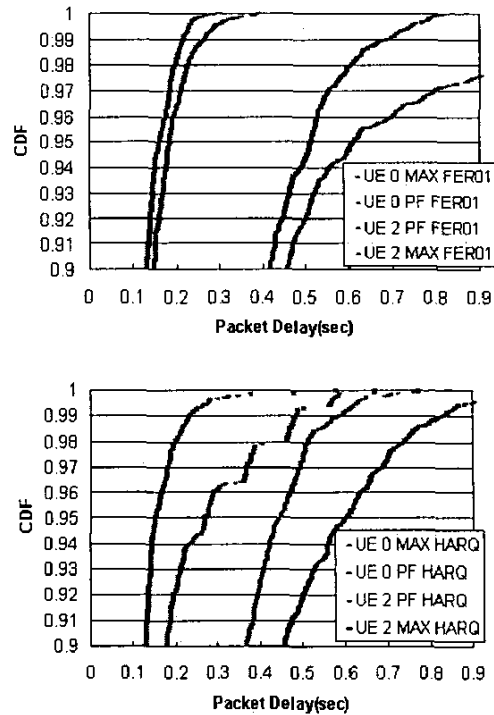
Figure 1. Delay Performance of PF and MAX Schedulers with (up) 1% FER based mapping (down) HARQ fitted mapping. (labels are for curves from left to right)
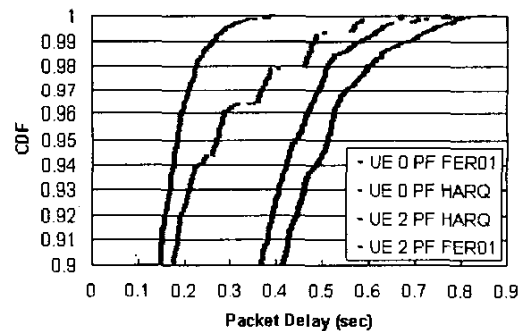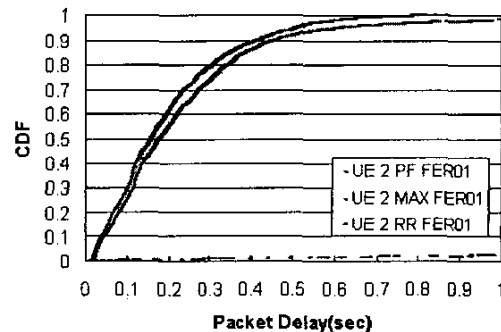

Figure 2. Delay Performance of PF and two mapping designs.

Figure 3. UE 2 Delay Performance of RR, PF and MAX schedulers