Cross-Layer Design for Resource Allocation in 3G Wireless Networks and Beyond

Hai Jiang, Weihua Zhuang, and Xuemin (Sherman) Shen, University of Waterloo

ABSTRACT

Cross-layer design approaches are critical for efficient utilization of the scarce radio resources with QoS provisioning in the third-generation wireless networks and beyond. Better system performance can be obtained from information exchanges across protocol layers, which may not be available in the traditional layering architecture. This article provides an overview of cross-layer design approaches for resource allocation in 3G CDMA networks, summarizes state-of-the-art research results, and suggests further research issues. In addition, a cross-layer design approach for real-time video over time-varying CDMA channels is proposed, where link layer resource allocation benefits from information in both the application and physical layers. Simulations results are given to demonstrate the effectiveness of the proposed approach.

INTRODUCTION

Third-generation (3G) wireless networks, also known as International Mobile Telecommunications 2000 (IMT-2000), aim at providing multimedia mobile services and achieving a maximum bit rate of 2 Mb/s. The deployment of 3G networks has already begun in different regions, and researchers have been proposing how 3G networks will evolve to beyond 3G or fourth-generation (4G) networks. To achieve a successful and profitable commercial market for 3G and beyond, network service designers and providers need to pay much attention to efficient utilization of radio resources. Although the available bandwidth is much larger in 3G and beyond networks (compared to 2G networks), it is still critical to efficiently utilize radio resources due to fast growth of the wireless subscriber population, increasing demand for new mobile multimedia services over wireless networks, and more stringent quality of service (QoS) requirements in terms of transmission accuracy, delay, jitter, throughput, and so on.

In order to meet the "anywhere and anytime" concept, the future wireless network architecture is expected to converge into a heterogeneous, all-IP architecture that includes different wireless access networks such as cellular networks, wireless local area networks (WLANs), and personal area networks (PANs, e.g., Bluetooth and ultra-wideband networks). It is well known that the success of today's Internet has been based on independent and transparent protocol design in different layers, a traditional network design approach that defines a stack of protocol layers (e.g., the Open System Interconnection [OSI] protocol stack). Using the services provided by the lower layer, each protocol layer deals with a specific task and provides transparent service to the layer above it. Such an architecture allows the flexibility to modify or change the techniques in a protocol layer without significant impact on overall system design. However, this strict layering architecture may not be efficient for wireless networks when heterogeneous traffic is served over a wireless channel with limited and time-varying capacity and high bit error rate (BER). Efficiently utilizing the scarce radio resources with QoS provisioning requires a cross-layer joint design and optimization approach. Better performance can be obtained from information exchanges across protocol layers. This article provides an overview of cross-layer design approaches over 3G and beyond wireless networks, summarizes state-of-the-art research results, and identifies further research issues. The article focuses on code-division multiple access (CDMA) networks, because:

- CDMA is a major candidate for 3G and beyond systems.
- CDMA offers more flexibility than time-division multiple access (TDMA) to cross-layer design for resource allocation due to its capability to support simultaneous transmissions.

In the following we first summarize recent research on cross-layer design for CDMA resource allocation, then present our research work on video service over CDMA time-varying channels.

0163-6804/05/$20.00 © 2005 IEEE

CROSS-LAYER DESIGN FOR CDMA RESOURCE ALLOCATION

One major challenge in multimedia services over CDMA cellular networks is QoS provisioning with efficient resource utilization. Compared to circuit-switched voice service in the 2G CDMA systems (i.e., IS-95), heterogeneous multimedia applications in future IP-based CDMA networks require a more complex QoS model and more sophisticated management of scarce radio resources. QoS can be classified according to its implementation in the networks, based on a hierarchy of four different levels: bit, packet, call, and application. Transmission accuracy, transmission rate (i.e., throughput), timeliness (i.e., delay and jitter), fairness, and user perceived quality are the main considerations in this classification. This classification also reflects the principle of QoS categories from the customer point of view:

- Bit-level QoS — To ensure some degree of transmission accuracy, a maximum BER for each user is required.
- Packet-level QoS — As real-time applications, such as voice over IP (VoIP) and videoconferencing, are delay-sensitive, each packet should be transmitted within a delay bound. On the other hand, data applications can tolerate delay to a certain degree, and throughput is a better QoS criterion. Each traffic type can also have a packet loss rate (PLR) requirement.
- Call-level QoS — In a cellular system, a new (or handoff) call will be blocked (or dropped) if there is insufficient capacity. From the user's point of view, handoff call dropping is more disturbing than new call blocking. Effective call admission control (CAC) is necessary to guarantee a blocking probability bound and a smaller dropping probability bound.
- Application-level QoS — Bit- and packet-level QoS may not directly reflect service quality perceived by the end user. On the other hand, application layer perceived QoS parameters are more suitable to represent the service seen by the end user, for example, the peak signal to noise ratio (PSNR) for video application, and the end-to-end throughput for data application provided by the responsive Transmission Control Protocol (TCP).

To guarantee the bit- and packet-level QoS requirements of mobile stations (MSs), an effective link layer packet scheduling scheme with appropriate power allocation is necessary. Specifically, the power levels of all the MSs should be managed in such a way that each MS achieves the required bit energy to interference-plus-noise density ratio (denoted E_b/I_0), and the transmissions from/to all the MSs should be controlled to meet the delay, jitter, throughput, and PLR requirements. A centralized scheduler at the base station (BS) benefits from more processing power and more available information than a distributed one. For the downlink, the BS has information on the traffic status of each MS. For the uplink, each MS needs to send a transmission request upon new packet arrivals and update its link status to the BS, as shown in Fig. 1. The request and update information can be transmitted in a request channel or piggybacked in the transmitted uplink packets to avoid possible contention in the request channel, and can be stored in the MS's profile at theBS. The BS responds by broadcasting transmission decisions to MSs.

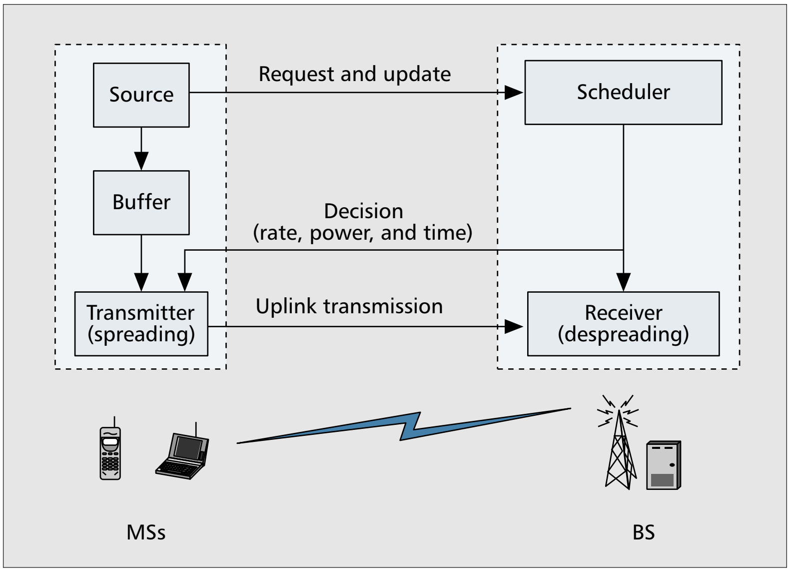


Figure 1. *The centralized scheduler for the uplink transmission.*

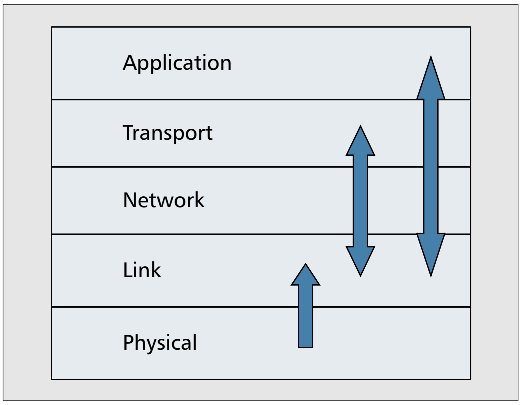


Figure 2. *The cross-layer information for IP-based CDMA resource allocation.*

To efficiently utilize scarce radio resources and achieve overall QoS satisfaction, cross-layer information is necessary. In traditional layering architecture, the link layer has statistical knowledge of the lower physical layer, such as the average channel capacity. However, to exploit the CDMA time-varying channel, it is better for the link layer to have knowledge of instantaneous channel status. Also, to guarantee the application-level QoS such as an acceptable visual quality of video services or a guaranteed TCP throughput of data services, the application or transport layer should be jointly designed with the link layer. In a five-layer reference model, Fig. 2 shows three possible cross-layer information directions, from physical to link layer, from link to transport layer and vice versa, and from

With the capability to support simultaneous transmissions in a CDMA system, multiuser diversity can be employed more effectively and flexibly than the traditional channel-aware scheduling schemes for a TDMA system.

link to application layer and vice versa. They lead to three cross-layer design approaches: channel-aware scheduling, TCP over CDMA wireless links, and joint video source/channel coding and power allocation, as discussed in the following.

CHANNEL-AWARE SCHEDULING

In a multiple access wireless network, the radio channel is normally characterized by time-varying fading. To exploit the time-varying characteristic, a kind of diversity (*multiuser diversity*) can be explored to improve system performance. The principle of multiuser diversity is that for a cellular system with multiple MSs having independent time-varying fading channels, it is very likely that there exists an MS with instantaneous received signal power close to its peak value. Overall resource utilization can be maximized by providing service at any time only to the MS with the highest instantaneous channel quality.

Multiuser diversity can be applied to CDMA networks successfully. For the downlink (or uplink), assume at each instant t only one MS i in a cell is receiving (or transmitting) with the target E_b/I_0 value Γ_i (at the receiver side) while other MSs are idle. Then the achieved data rate of MS i is given by

$$R_i(t) = \frac{W}{\Gamma_i} \cdot \frac{P^{\max} \cdot h_i(t)}{I_i^o(t) + v_i(t)} \qquad (1)$$

where W is the total downlink (or uplink) bandwidth used by all the MSs, $P^{\max}$ is the maximum power limit of the transmitter, h_i is the channel gain of MS i, and I_i^o and v_i are intercell interference and background noise power at the receiver side of MS i's connection, respectively. On the right side of the equation, the second term is the received signal to interference-plus-noise ratio (SINR), represented by $\mathrm{SINR}_i(t)$. In the downlink with the same required E_b/I_0 value for all MSs, obviously the maximum system throughput can be achieved if at any time the BS only transmits to the MS with the highest instantaneous channel quality (i.e., with the highest $\mathrm{SINR}_i(t)$). For the uplink, if there is sufficient power at MS transmitters and there is a limit on received power at the BS, when only the MS with the best channel quality transmits, the minimum transmit power is needed. This leads to minimum interference to neighbor cells, thus increasing overall system capacity in a multicell environment.

With the capability to support simultaneous transmissions in a CDMA system, multiuser diversity can be employed more effectively and flexibly than traditional channel-aware scheduling schemes for a TDMA system. An MS does not need to wait until it has the best channel quality among all MSs, but rather can transmit as long as its channel is good enough. All the MSs are divided into two sets: bad channel state MSs if the channel gain is F dB less than the average value, and good channel state MSs otherwise. The value F is called the *good/bad threshold*. The scheduler keeps track of the obtained services and channel states of all the MSs. MSs in a bad channel state postpone their transmissions until they have a good channel state, use a relatively small weight in resource allocation [1], or use fewer code channels in multicode CDMA. In a good channel state a previously sacrificed MS will be compensated (i.e., get a larger weight or more code channels).

It should be mentioned that for real-time traffic (e.g., voice or video) with a delay constraint, if an MS is in a bad channel state for a relatively long period, its packets will be discarded when multiuser diversity is employed, as it has to wait until a good channel state. Hence, it is challenging to apply multiuser diversity to real-time traffic. An effective way is to incorporate the packet delay in the scheduling decision, as in our proposed scheme discussed later.

TCP OVER CDMA WIRELESS LINKS

For data services, TCP guarantees error-free delivery. TCP was originally designed for wireline networks with a reliable physical layer, where packet loss mainly results from network congestion. In such networks TCP adjusts its sending rate based on the estimated network congestion status so as to achieve congestion control or avoidance. In a wireless environment TCP performance can be degraded severely as it interprets losses due to unreliable wireless transmissions as signs of network congestion and invokes unnecessary congestion control. To improve TCP performance over the wireless links, several solutions have been proposed to alleviate the effects of non-congestion-related packet losses [2], among which snoop TCP and explicit loss notification (ELN) are based on cross-layer design. In snoop TCP, TCP layer knowledge is used by link layer schemes, while in ELN, the network layer takes advantage of cross-layer information from the physical layer.

When a TCP connection is transmitted over CDMA cellular networks, further considerations are needed. First, CDMA capacity is interference limited. TCP transmission from an MS generates interference to other MSs. It is desired to achieve acceptable TCP performance (e.g., a target throughput) and at the same time introduce minimum interference to other MSs (i.e., to require minimum low-layer resources). Second, power allocation and control in CDMA can lead to a controllable BER, which affects TCP performance.

For a TCP flow i over CDMA cellular networks, if there are packets to be transmitted, up to a fixed number (denoted M_i) of link layer packets can be scheduled to be transmitted in each link layer frame with E_b/I_0 value Γ_i. M_i is called the *target number of scheduled packets* for flow i, and (M_i, Γ_i) is called the *link layer design parameter vector*. It can be seen that TCP interacts with link layer resource allocation. Specifically, TCP dynamically adjusts the sending rate of TCP segments (which will be fed into the link layer transmission queue) according to network congestion status (e.g., packet loss and round-trip delay); on the other hand, the link layer design parameter vector (M_i, Γ_i) ultimately determines the packet loss rate and transmission delay over the wireless link, and therefore affects the TCP performance. From cross-layer parameter design based on this interaction, a feasible set of link layer design parameter vectors can be obtained, which achieves the target TCP throughput. If the design parameter vector is chosen (among the feasible parameter vectors) to minimize the amount of required resources, the optimal resource allocation can be achieved for the TCP flow [3].

Joint Source/Channel Coding and Power Allocation for Video Services

Video transmission is an important component of multimedia services. Typical video applications include mobile videoconferencing, video streaming, and distance learning. Due to realtime nature, video services typically require QoS guarantees such as a relatively large bandwidth and a stringent delay bound.

For video services over a CDMA channel with limited capacity, an effective way is to pass source significance information (SSI) from the source coder in the application layer to the channel coder in the physical layer. More powerful forward error correction (FEC) code (and therefore more overhead) can be used to protect more important information, while no or weaker FEC may be applied to less important information. Such joint source/channel coding is a crosslayer approach, called *unequal error protection* (UEP). UEP can easily be performed with Bose-Chaudhuri-Hocquenghem (BCH) codes, Reed-Solomon (RS) codes, and rate-compatible punctured convolutional (RCPC) codes with different coding rates for packets with different priorities. UEP can also be implemented by means of power allocation in CDMA systems; for example, transmission power can be managed so that a more important packet experiences a smaller error probability [4]. In case of capacity shortage, UEP schemes can result in more graceful quality degradation (and thus smaller distortion, or higher PSNR) than equal error protection (EEP). Based on channel capacity, the optimal transmission rate and power allocation for packets of each priority can be found to minimize the average distortion of the received video by means of an optimization formulation over CDMA channels [5]. It outperforms uniform power allocation, as it exploits the degree of freedom added by CDMA power allocation.

A video codec has the ability to adjust its source coding rates. This flexibility can also be exploited to improve system performance. Consequently, it is desirable to employ a joint source/channel coding scheme that allocates bits for source and channel coders so as to minimize the end-to-end distortion under a given bandwidth constraint [6].

With interference-limited capacity, it is important to take into account the power management in CDMA systems when designing the source and channel coding. More flexibility can be obtained when power allocation is considered jointly with source and/or channel coding. A large source coding rate can lead to low quantization distortion, and a high transmission accuracy level (i.e., high E_b/I_0 from power allocation) can achieve low channel-error-induced distortion. However, for CDMA a large source coding rate (and therefore a large transmission rate) and a high E_b/I_0 value are conflicting objectives. Hence, the source coding rate and E_b/I_0 can be jointly designed such that the resulting PSNR is maximized, subject to a joint constraint on them [7]. Transmission power consumption minimization can also be achieved in joint source/channel coding andpower allocation schemes subject to acceptable video distortion [8]. Apparently, when transmission power consumption is reduced, CDMA system capacity can be enlarged. However, the above optimization is complicated to achieve. The case is worse when time scheduling for multiplexed video traffic is implemented. Further investigation is necessary.

Applying multiuser diversity to real-time traffic is very challenging due to the delay requirement of real-time traffic. To address this problem, we propose a cross-layer approach termed dynamic-weight generalized processor sharing (GPS) for wireless video resource allocation with service differentiation over CDMA networks.

Proposed Cross-Layer Approach for Video over Time-Varying CDMA Channels

Applying multiuser diversity to real-time traffic is very challenging due to the delay requirement of such traffic. To address this problem, we propose a cross-layer approach termed *dynamic weight generalized processor sharing* (DWGPS) for wireless video resource allocation with service differentiation over CDMA networks. To implement UEP and multiuser diversity, the proposed cross-layer approach can benefit from information in both the application and physical layers. Our focus is on video transmission in the uplink, as resource allocation in the multiple access uplink is much more complex than that in the broadcasting downlink. However, the proposed solution should also be applicable to downlink video transmission.

Dynamic-Weight Generalized Processor Sharing

For resource allocation, the well-known GPS [9, 10] is an ideal fair scheduling discipline originally proposed for wireline networks. The basic principle of GPS is to assign a fixed weight to each session, and allocate bandwidth to all sessions according to their weights and traffic loads. GPS can provide each session with a minimum service rate. Also, a tight delay bound can be guaranteed by the GPS server for each session if its traffic is shaped by a leaky bucket regulator. The minimum service rate and tight delay bound guaranteed in GPS may seem attractive to realtime video transmission. However, as compressed video traffic is usually bursty, its peak rate is likely to be much greater than its average rate. Therefore, a large weight should be assigned to a video session in order to guarantee the peak rate. This means a video session will get a large portion of the total capacity whenever it has traffic to transmit, thus leading to service degradation of other sessions. On the other hand, if the peak rate cannot be guaranteed, the delay bound of video traffic cannot be guaranteed either because of the latency in the leaky bucket regulator. In order to apply GPS discipline to video transmission and extend it to wireless networks, dynamic weights in GPS are introduced in our research, and the approach is DWGPS.

In DWGPS, each raw video frame is compressed to several batches of link layer (LL)

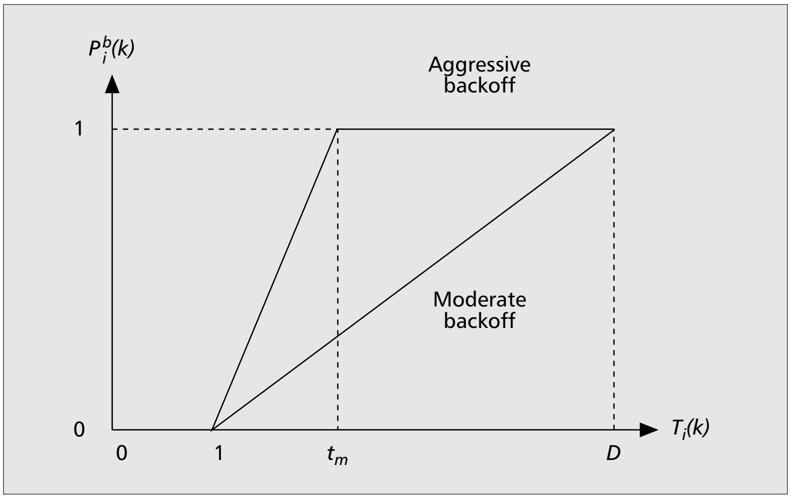


Figure 3. *The backoff probability of moderate and aggressive backoff schemes.*

packets according to the priority of the coded information. When generated, each batch is classified into one of total L classes (numbered from 1 to L) according to its priority. Upon the arrival of each video frame, the MS creates a transmission queue for each batch of the video frame, assigns a timer with a timeout value to each batch, and reports to the BS the batch class and batch arrival size in LL packets. The batch class and batch arrival size are determined based on information passed from the application layer. The timeout value in a unit of LL frame[1] reflects the maximum tolerable delay over the wireless link. The timer will decrease by one after every LL frame. If the timer expires, any LL packets remaining in the associated batch transmission queue are considered useless and discarded, and the batch transmission queue is deleted.

A *session* is defined as an active batch in the transmission queue. Therefore, a video sequence may have multiple sessions simultaneously. At LL frame k with a total number $N(k)$ of active batches at all MSs, an active batch i is assigned a DWGPS weight

$$\phi_i(k) = g_i(k)\frac{S_i(k)}{T_i(k)}, \quad 1 \le i \le N(k) \tag{2}$$

where $g_i(k)$ is the importance weight of batch i's class at LL frame k, and $S_i(k)$ and $T_i(k)$ are the remaining size and remaining timer value of batch i at the beginning of LL frame k, respectively. Equation 2 is reasonable because the larger a batch's remaining size, the more capacity it requires; and the smaller the batch's timer value, the more urgent the batch's delivery. The selection of the importance weight is based on the criterion that a batch from a higher-priority class will be assigned a relatively larger weight in DWGPS, corresponding to a higher transmission rate of this batch than those of lower-priority classes to better protect higher-priority classes during capacity shortage. In DWGPS the importance weight selection is quite flexible. An optimization approach can be used for importance weight selection to protect higher-priority classes as much as possible. The importance weight can also be configured to achieve a target ratio of LL packet loss rates of the L traffic classes in order to avoid starvation of low-priority classes.

In Eq. 2 $\{S_i(k)/T_i(k)\}$ is the average service capacity amount required by batch i in each subsequent LL frame before the batch times out. With the weight proportional to its average required capacity portion, each batch is expected to be served smoothly rather than in burst within the delay bound. If a batch is expected to transmit all backlogged LL packets before it times out, all other batches in the same class are expected to do the same. Similarly, if a batch is expected to lose a portion of LL packets, all other batches in the same class are expected to have the same share of packet loss. Therefore, fairness can be achieved among different traffic flows. In this work, fairness means that all batches in the same class deliver successfully a similar portion of their arrival traffic (thus leading to a similar packet loss rate).

To implement DWGPS, upon each new batch arrival, the MS reports to the BS the batch class and batch size. For each LL frame, the BS determines how many LL packets can be transmitted from each active batch of each MS, and broadcasts the decision to the MSs. Therefore, the BS only needs to store the class number, remaining size, and timer value of each batch. And the information exchange overhead is not significant.

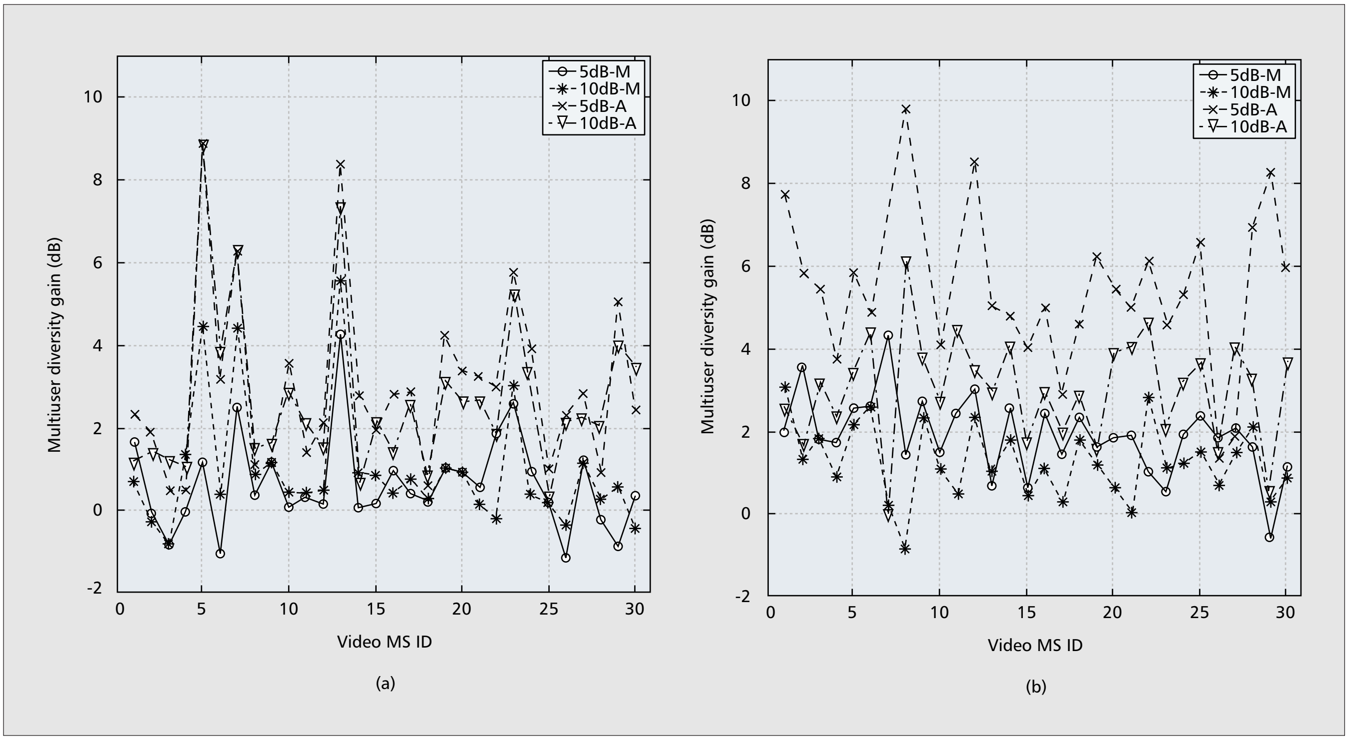
MULTIUSER DIVERSITY ADAPTATION

To incorporate multiuser diversity and consider the delay bound of real-time traffic, in DWGPS, each batch i with a bad channel at LL frame k is kept idle with a probability $P_i^b(k)$ (called *backoff probability*) at this LL frame. If an MS is in a bad channel state, we say all of its batches are in a bad channel state. Intuitively, the smaller a batch's timer value, the more urgent the batch's transmission, and the smaller the backoff probability for this batch should be. Figure 3 shows moderate and aggressive backoff schemes when batch i is in a bad channel state at LL frame k, where D is the wireless delay bound and t_m is the point of timer value $T_i(k)$ from which the backoff probability of the aggressive backoff scheme is linearly decreased. It is worth noting that the relation of backoff probability vs. timer should depend on the channel fading rate: if the channel fades fast, relatively large backoff probabilities should be used in a bad channel state with the expectation that the channel quality will get better and the affected batches will be compensated soon. Furthermore, the good/bad threshold F should be determined carefully. This threshold affects the probability of a batch being considered in a bad channel state, and indirectly affects the performance of the backoff scheme. In our research, F could be determined based on experimental results.

Although originally designed for video transmissions, DWGPS can also be applied to real-time voice applications such as VoIP over CDMA with a required delay bound. Furthermore, data traffic is usually deemed delay-insensitive, but this may not always be true in practice. For example, for Web browsing, packets will be discarded if they cannot be delivered successfully within a deadline. It is desired that data traffic should have a delay bound, which can be much larger than that of

---

[1] *In this article LL frame means link layer time-frame, while video frame, I-frame, and P-frame all mean a frame of picture in a video sequence.*



■ **Figure 4.** *Multiuser diversity gains for MSs in the diversity DWGPS schemes in the slow and fast fading environments with* F = *5 or 10 dB, and moderate (represented by M) or aggressive (represented by A) backoff: a) slow fading environment; b) fast fading environment.*

voice or video. In this context DWGPS is also effective in supporting data traffic.

PERFORMANCE EVALUATION

Computer simulations are carried out to evaluate the performance of DWGPS. Consider 30 video test sequences (with IDs ranging from 1 to 30) are transmitted from 30 MSs to their correspondence nodes. Each raw video test sequence is in Quarter Common Intermediate Format (QCIF) with a duration of 3000 LL frames at a rate of 10 video frames/s, and is compressed by an MPEG-4 coder with a *base layer* and an *enhancement layer*. In the base layer, only I-frame and P-frame are used. B-frame is not used due to the additional delay involved in its video compression and decompression process. Hence, there are three classes of batches from each video sequence: I-frame batch in the base layer (called IB batch), P-frame batch in the base layer (called PB batch), and batch in the enhancement layer (called E batch). IB batch has the highest priority, E batch the lowest.

It is observed that in DWGPS, IB class traffic receives the best service (in terms of packet loss rate) and E class traffic receives the least. Hence, DWGPS can implement UEP effectively.

For multiuser diversity adaptation, we define a multiuser diversity gain for an MS in a diversity DWGPS scheme as the ratio of average transmission power for an LL packet from the MS in the nondiversity DWGPS scheme to that in the diversity DWGPS scheme. Figure 4 shows the multiuser diversity gain of the 30 MSs for $F = 5$ dB (or 10 dB) and moderate (or aggressive) backoff scheme in the slow fading (with normalized fading rate 0.01) and fast fading (with normalized fading rate 0.28) environments. In comparison with $F = 10$ dB, $F = 5$ dB brings about more service degradation in terms of LL packet loss rates to high-priority batch classes (not shown here due to space limitation). This is because $F = 5$ dB leads to a larger probability of an MS being considered in a bad channel state, resulting in a larger probability of a batch being kept idle, which imposes more capacity requirements on later LL frames. Note that the probability of a channel 10 dB less than average quality is 10 percent, while the probability of a channel 5 dB less than average quality is 27 percent. For $F = 10$ dB, the aggressive backoff scheme outperforms the moderate backoff scheme in terms of multiuser diversity gain as shown in Fig. 4, at the cost of negligible (or non-negligible) service degradation in terms of LL packet loss rates of high-priority classes in the fast (or slow) fading environment. Hence, $F = 10$ dB, and the aggressive (or moderate) backoff scheme is a good choice for fast (or slow) fading.

CONCLUSION

In this article the fundamentals of recent research efforts in cross-layer design for resource allocation in future CDMA-based 3G and beyond networks have been presented, and a novel cross-layer approach for video service over time-varying CDMA channels is proposed. In cross-layer design the overall system performance can be improved by taking advantage of the available information across different layers. To achieve this, an appropriate signaling method is necessary. Cross-layer signaling can be implemented by the following means [11]:

- Cross-layer information is stored in packet headers.

For a CDMA network supporting heterogeneous voice/video/data traffic with different QoS requirements, it is critical to consider the tradeoff among the cross-layer approaches for different traffic types, and to achieve desired overall system performance with efficient resource utilization.

- A third-party network service takes care of the management of the cross-layer information.
- System profiles are used to collect and store cross-layer information, and access to the system profiles is provided to the related layers.

Recent research has provided preliminary results for cross-layer design over all-IP CDMA networks. Further research efforts should include:

- Joint source/channel coding and power allocation for multiplexed video services with time scheduling: when several video flows are multiplexed, it is desired to use time scheduling for efficient bandwidth utilization. It is challenging to jointly allocate source coding rate, channel coding rate, and power level to video flows in a multiplexed environment.
- Cross-layer design for differentiated services (DiffServ)-based QoS: DiffServ has emerged as a scalable solution to ensure QoS in IP networks. When a data application transported by TCP is provided with DiffServ QoS, there is an interaction between TCP congestion management and the IP-layer traffic conditioning/forwarding mechanism. Poor performance can result from mismatch between TCP and IP layer mechanisms [12]. When DiffServ is applied to CDMA wireless networks, this interaction should be extended to the lower link/physical layers. To address this, effective cross-layer design from the TCP to the IP and finally to the link/physical layers is necessary to achieve the desired DiffServ QoS performance.
- Cross-layer design for heterogeneous voice/video/data traffic: a cross-layer design approach usually focuses on a specific traffic type. For a CDMA network supporting heterogeneous voice/video/data traffic with different QoS requirements, it is critical to consider the tradeoff among cross-layer approaches for different traffic types, and to achieve desired overall system performance with efficient resource utilization.

ACKNOWLEDGMENTS

This work was supported by the Premier's Research Excellence Award (PREA) from the Ontario Government and by the Strategic Project (STPGP 257682 - 02) of the Natural Science and Engineering Research Council (NSERC) of Canada.

REFERENCES

[1] L. Xu, X. Shen, and J. W. Mark, "Fair Resource Allocation with Guaranteed Statistical QoS for Multimedia Traffic in Wideband CDMA Cellular Network," *IEEE Trans. Mobile Comp.*, vol. 4, no. 2, Mar.-Apr. 2005, pp. 166–77.

[2] G. Xylomenos *et al.*, "TCP Performance Issues over Wireless Links," *IEEE Commun. Mag.*, vol. 39, no. 4, Apr. 2001, pp. 52–58.

[3] H. Jiang and W. Zhuang, "Cross-Layer Resource Allocation for Integrated Voice/Data Traffic in Wireless Cellular Networks," to appear, *IEEE Trans. Wireless Commun.*

[4] I.-M. Kim and H.-M. Kim, "Efficient Power Management Schemes for Video Service in CDMA Systems," *Elect. Lett.*, vol. 36, no. 13, June 2000, pp. 1149–50.

[5] S. Zhao, Z. Xiong, and X. Wang, "Joint Error Control and Power Allocation for Video Transmission over CDMA Networks with Multiuser Detection," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 12, no. 6, June 2002, pp. 425–37.

[6] L.P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint Source-Channel Coding for Motion-Compensated DCT-Based SNR Scalable Video," *IEEE Trans. Image Proc.*, vol. 11, no. 9, Sept. 2002, pp. 1043–52.

[7] Y. S. Chan and J. W. Modestino, "A Joint Source Coding-power Control Approach for Video Transmission over CDMA Networks," *IEEE JSAC*, vol. 21, no. 10, Dec. 2003, pp. 1516–25.

[8] Q. Zhang et al., "Power-Minimized Bit Allocation for Video Communication over Wireless Channels," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 12, no. 6, June 2002, pp. 398–410.

[9] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case," *IEEE/ACM Trans. Net.*, vol. 1, no. 3, June 1993, pp. 344–57.

[10] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple Node Case," *IEEE/ACM Trans. Net.*, vol. 2, no. 2, Apr. 1994, pp. 137–50.

[11] Q. Wang and M. A. Abu-Rgheff, "Cross-layer Signalling for Next-generation Wireless Systems," *Proc. IEEE WCNC 2003*, pp. 1084–89.

[12] P. Giacomazzi, L. Musumeci, and G. Verticale, "Transport of TCP/IP Traffic over Assured Forwarding IP-Differentiated Services," *IEEE Network*, vol. 17, no. 5, Sept.–Oct. 2003, pp. 18–28.

BIOGRAPHIES

HAI JIANG [S'04] (hjiang@bbcr.uwaterloo.ca) received a B.S. degree in 1995 and an M.S. degree in 1998, both in electronics engineering, from Peking University, Beijing, China, in 1995 and 1998, respectively. He is currently working toward a Ph.D. degree at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His current research interests include QoS provisioning and resource management for multimedia communications in all-IP wireless networks.

WEIHUA ZHUANG [M93, SM'01] (wzhuang@bbcr.uwaterloo.ca) received B.Sc. and M.Sc. degrees from Dalian Maritime University, China, and a Ph.D. degree from the University of New Brunswick, Canada, all in electrical engineering. Since October 1993 she has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where she is a full professor. She is a co-author of the textbook *Wireless Communications and Networking* (Prentice Hall, 2003). Her current research interests include multimedia wireless communications, wireless networks, and radio positioning. She is an Associate Editor of *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Vehicular Technology*, and *EURASIP Journal on Wireless Communications and Networking*.

XUEMIN (SHERMAN) SHEN [M'97, SM'02] (xshen@bbcr.uwaterloo.ca) received a B.Sc. (1982) degree from Dalian Maritime University, China, and M.Sc. (1987) and Ph.D. (1990) degrees from Rutgers University, New Jersey, all in electrical engineering. Currently, he is with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where he is a professor and associate chair for graduate studies. His research focuses on mobility and resource management in interconnected wireless/wireline networks, UWB wireless communications systems, wireless security, and ad hoc and sensor networks. He is a co-author of two books, and has published more than 200 papers and book chapters in wireless communications and networks, control, and filtering. He was Technical Co-Chair for IEEE GLOBECOM '03, ISPAN '04, QShine '05, IEEE Broadnets '05, and WirelessCom '05, and is Special Track Chair of the 2005 IFIP Networking Conference. He serves as Associate Editor for *IEEE Transactions on Wireless Communications*; *IEEE Transactions on Vehicular Technology*; *Computer Networks*; *ACM/Wireless Networks*; *Wireless Communications and Mobile Computing* (Wiley); and *International Journal Computer and Applications*. He has also served as Guest Editor for *IEEE JSAC*, *IEEE Wireless Communications*, and *IEEE Communications Magazine*. He received the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, for demonstrated excellence of scientific and academic contributions, and the Distinguished Performance Award in 2002 from the Faculty of Engineering, University of Waterloo, for outstanding contributions in teaching, scholarship, and service.