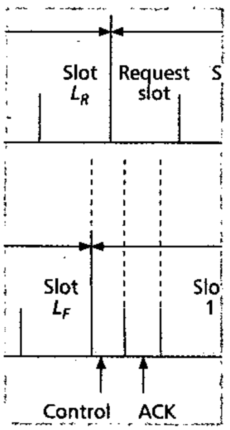
4G Mobile Communications: Toward Open Wireless Architecture

Quality-of-Service Provisioning in Future 4G CDMA Cellular Networks

Hai Jiang and Weihua Zhuang, University of Waterloo



It is well accepted that the 4G wireless network architecture will include different wireless access networks, glued together by Mobile IP to provide seamless Internet access to mobile users. A major challenge in establishing such a heterogeneous architecture is QoS provisioning in different wireless networks.

Abstract

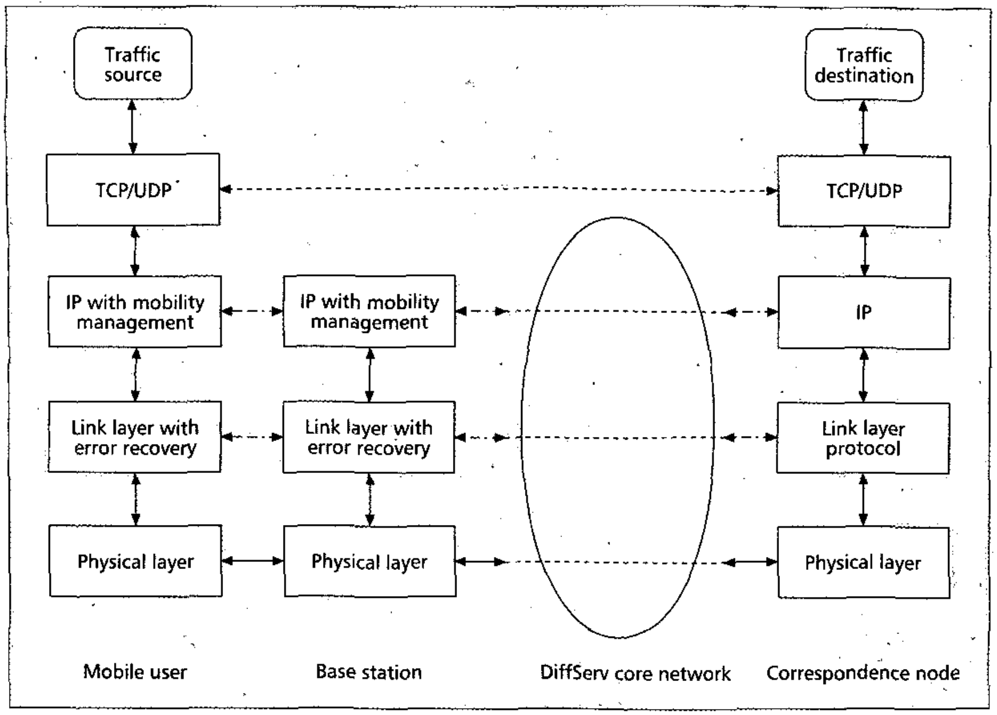
It is well accepted that the 4G wireless network architecture will include different wireless access networks, glued together by Mobile IP to provide seamless Internet access to mobile users. A major challenge in establishing such a heterogeneous architecture is QoS provisioning in different wireless networks. QoS provisioning is more complex in a CDMA cellular system due to the interference-limited capacity. In this article we propose a vertically coupled protocol architecture to provide QoS in 4G CDMA cellular networks. This architecture combines the transport layer protocols and link layer resource allocation to both guarantee the high-layer QoS requirements and achieve efficient resource utilization in the low layer. A packet-switching MAC scheme is also provided to achieve efficient multiplexing. The MAC scheduler uses only per-flow information in packet scheduling, thus significantly reducing the computation complexity and system overhead.

Introduction

The past decade has witnessed the astounding development speed of mobile communication networks. The second-generation (2G) cellular systems, such as GSM and IS-95, brought a revolution from analog to digital technology, and are today's mainstream systems. They can provide voice and low-rate circuit-switched data services. The third-generation (3G) systems, also known as International Mobile Telecommunications 2000 (IMT-2000), are aimed 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 are thinking how 3G networks will evolve to fourth-generation (4G) systems. With the explosive demand for high-speed data services, in 4G systems mobile users will be serviced with a data rate as high as 20 Mb/s. Also, in order to meet the "anywhere, anytime" concept, the future 4G network architecture is expected to converge into a heterogeneous, all-IP architecture, which includes different wireless access networks such as 4G cellular networks, wireless LAN (WLAN), Bluetooth, and ultra-wideband systems. IP is employed as the common network layer protocol. Different wireless networks can be glued together by Mobile IP to provide seamless Internet access to mobile users.

Integrating multiple subsystems into 4G brings about many challenges such as subsystem interworking and end-to-end quality of service (QoS) provisioning to different Internet applications. QoS provisioning is not easy even in wireline networks due to varying QoS requirements and the heterogeneous nature of multimedia traffic (e.g., time-varying transmission rates). In a wireless environment, the situation is worse due to the limited radio resources and the wireless channel characteristics. Furthermore, when an IP-compliant mobile user is handed over between different wireless access networks, the equivalent resource amount should be determined and allocated in the new network to guarantee QoS during the traffic lifetime. This task is more complex in a 4G code-division multiple access (CDMA) cellular system as the system capacity is interference-limited, and power should also be considered as a kind of resource along with bandwidth. When different applications are supported, an efficient packet-switching medium access control (MAC) protocol is required in 4G CDMA cellular systems in order to achieve high multiplexing gain and efficient resource utilization.

For QoS provisioning in IP networks, the differentiated services (DiffServ) [1] approach has emerged as an efficient and scalable solution to ensure Internet QoS based on handling of limited traffic classes. However, current research on DiffServ mainly focuses on the wireline network. Only limited work is done on DiffServ over all-IP wireless networks, and most of the work focuses only on the transport and network layers, without consideration for the utilization of valuable resources in the lower (i.e., link and physical) layers. To address the above issues, in this article we propose a vertically coupled protocol architecture for 4G cellular networks to provision QoS to mobile users who have subscribed to DiffServ. The proposed architecture combines the transport layer protocols and link layer resource allocation to both guarantee



■ Figure 1. *The protocol stack architecture in 4G CDMA cellular networks.*

UDP is suitable for real-time applications (VoIP, voice chat, etc.) as it does not use retransmissions to guarantee reliable delivery. It is used in the system as the transport layer protocol for voice premium service.

transport layer QoS requirements and achieve efficient resource utilization in the link layer. The transport layer QoS is guaranteed with minimal equivalent resources required at the link layer and therefore at the physical layer. Besides, the MAC scheduler in the link layer is based on per-flow information, thus reducing the computation complexity and system overhead compared to other MAC schemes [2, 3], which use per-packet information to determine scheduling priority.

System Architecture

Consider a 4G cellular system connected to a DiffServ core network. Here we focus on transmissions in the reverse link, as resource allocation in the multiple access reverse link is much more complex than in the broadcasting forward link. For a scenario where a mobile user wants to initiate a connection to a fixed correspondence node, the base station is responsible for resource allocation in the wireless link. It is also the edge router of the DiffServ core network. In the base station, packets from mobile users are classified into limited service classes via a packet marking mechanism according to service level agreements (SLAs). In the DiffServ core network, no per-flow information is needed. A core router provides, based on the marked field, differentiated aggregate treatments to different classes of packets. Thus, the core network is kept simple, and the base station is responsible for complicated functionality such as per-flow traffic conditioning and marking. Figure 1 shows the protocol stack architecture under consideration. We do not consider the effect of the IP layer on system performance as the IP layer only generates a relatively fixed amount of overhead to overall performance. In other words, only interactions between the transport layer and the link layer are considered to determine system performance.

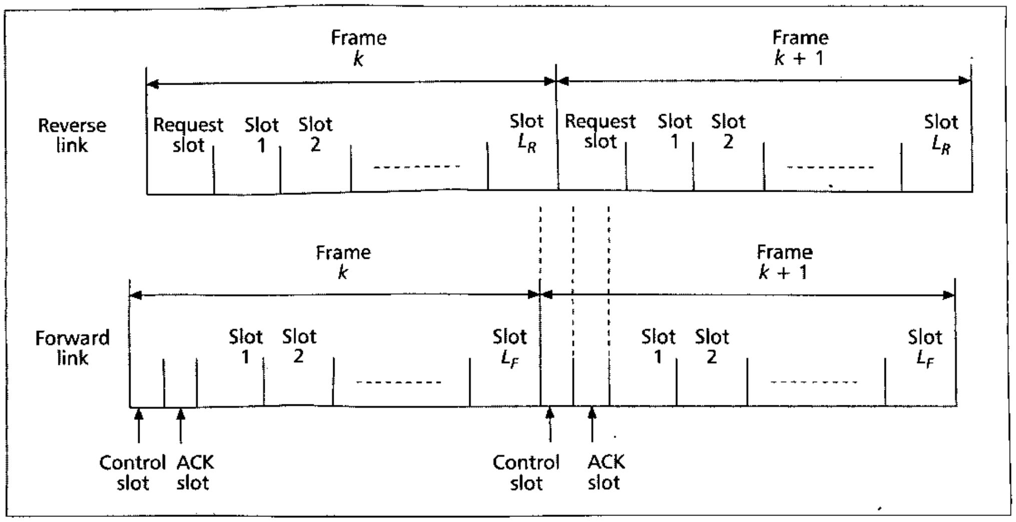
Traffic Classes

We use the traffic classes defined in DiffServ: premium service for voice traffic and assured service for data traffic. For simplicity, we do not include video traffic in this article. Premium service is aimed at providing a low-loss, low-delay, and low-jitter connection, and is intended for real-time applications such as voice over IP (VoIP) and videoconferencing. Such a service appears to endpoints as a virtual leased line. Assured service is to provide a reliable connection with a target transmission rate, referred to as committed information rate (CIR), even during network congestion.

Transport Layer Protocol

User Datagram Protocol (UDP) [4] is suitable for real-time applications (VoIP, voice chat, etc.) as it does not use retransmissions to guarantee reliable delivery. It is used in the system as the transport layer protocol for voice premium service. When a voice user is on talk spurt, the UDP packets will be generated periodically at a rate depending on the encoding rate, UDP packet size, and packet header size. In our system the QoS requirement for voice premium service is a guaranteed peak information rate (normally the transmission rate on talk spurt) with a small delay bound. If a UDP packet cannot be delivered to the end user within this bound, it will be dropped; the dropping probability is also bounded.

If a flow is scheduled to transmit several packets in a frame, the scheduler tries to put them in the same slot by orthogonal multi-code (MC) CDMA. The packets will arrive at the base station without interference with each other because of the same transmission environment, thus increasing the system capacity.



■ Figure 2. *Time frame architecture of the FDD forward and reverse links.*

On the other hand, Transmission Control Protocol (TCP) can provide reliable end-to-end transmission over unreliable IP service, which is suitable for assured service. The NewReno version of TCP [5] is considered here, as it is a popular implementation in today's Internet. The QoS requirement for data assured service is the guaranteed CIR with reliable end-to-end transmission (i.e., the achieved TCP throughput should be at least the required CIR).

Under the assumption of a transparent IP layer, each transport layer (TCP or UDP) packet is segmented to a number of link layer packets for transmission over the error-prone wireless link.

Link Layer Time Frame Architecture

In the link layer, we consider a hybrid time-division/code-division multiple access (TD/CDMA) packetized architecture, as it has flexibility in time scheduling and can allow simultaneous transmissions from different users. Figure 2 shows the time frames of the forward and reverse links operating in frequency-division duplexing (FDD) mode. The frame architecture for the forward link is illustrated because acknowledgments (ACKs) for the reverse link transmissions are sent from the base station over the forward link. This diagram is similar to the one used in [2]. The main difference is that our model ensures that the ACK for any packet transmitted in a reverse link frame is expected to be received before any packet slot of the next reverse link frame.

In the TD/CDMA architecture, time is partitioned into fixed-length frames. In the reverse link, each frame is divided into a request slot and L_R packet slots; in the forward link, each frame is divided into a control slot, an ACK slot, and L_F packet slots. In each type of slot, CDMA multiplexing is used with a fixed spreading gain. A link layer information packet can be accommodated in a packet slot. ACKs for forward link transmissions are piggybacked in the transmitted information packets of the reverse link. If a mobile user wants to initiate a call, it issues a request in the request slot based on slotted ALOHA. If the request is received and admitted by the base station, the mobile user will receive an acceptance notification in the control slot of the next forward link frame. The acceptance notification also tells the mobile user which code is assigned and which slot can be used by this user in the first available frame. For information packets transmitted in the reverse link, the ACK slot in the forward link is used by the base station to send the following to the mobile users: ACKs for packets successfully received in the last frame and information about how many packets can be scheduled in the next reverse link frame. It can be seen that if a packet is scheduled in a reverse link frame, the ACK can be expected to be received before any packet slot of the next reverse link frame. This is important for timely packet retransmissions.

If a flow is scheduled to transmit several packets in a frame, the scheduler tries to put them in the same slot by orthogonal multicode (MC) CDMA. The packets will arrive at the base station without interference with each other because of the same transmission environment, thus increasing the system capacity.

Link Layer Error Recovery

TCP was originally developed for wireline networks with reliable physical links. When a TCP connection includes a wireless link in its end-to-end path, its performance suffers from severe throughput degradation, as TCP considers the packet losses due to unreliable wireless transmission as an indication of network congestion and cuts back its transmission rate and subsequently increases it in a sluggish mode, thus underutilizing the available wireless bandwidth. On the other hand, UDP itself does not provide mechanisms to ensure timely delivery or other QoS guarantees, but relies on lower-layer services to do so. In order to improve the performance of TCP/UDP over a wireless channel, it is necessary to resort to link layer error control.

Two classes of error control mechanisms are commonly used: forward error correction (FEC) and automatic repeat request (ARQ). FEC corrects errors at the expense of redundancy and computation complexity. In ARQ, by the use of error detection code included in the transmitted packet, errors can be detected at the receiver end and retransmission can be requested via a feedback channel. In our system, hybrid FEC/ARQ complemented by cyclic redundancy check (CRC) is used, as it is recommended by recent studies as a more appropriate candidate for wireless transmissions. The hybrid FEC/ARQ is designed based on transmission accuracy and delay requirements for different traffic classes.

For an assured service flow transmitted with TCP, as it is delay-tolerant and requires high reliability, unlimited retransmissions are allowed. If a link layer packet is transmitted successfully in a reverse link frame, an ACK is fed back before the next reverse link frame's packet slots, as shown in Fig. 2. Otherwise, this packet is retransmitted repeatedly in the following frames until it is received successfully. The main concern about unlimited retransmissions is the possibility of introducing competing retransmissions between the TCP and the link layer ARQ because of premature TCP timeouts and out-of-order delivery of TCP packets. This adverse effect is negligible in our system, because:

- With power control, the link layer packet loss events are approximately independent.
- From numerical analysis, the successful link layer packet transmission probability in optimal resource allocation is high (e.g., 90 percent).
- The ACKs for the reverse link transmissions can be sent back in time.
- TCP implementation usually uses a coarse granularity of timeout value (say 500 ms) and adjusts this value according to the measured round-trip time (RTT).

All these factors determine that the probability of premature timeouts and out-of-order delivery in TCP is negligible. The mobile user employs a random early detection (RED) [6] queue as the TCP packet buffer, because RED has the ability to control the queuing delay and prevent consecutive packet losses. The packet losses in the RED queue will be recovered by TCP error control mechanisms. As a result, the assured service mobile user will see an end-to-end transmission with no error. The throughput requirement of assured service is guaranteed based on the coupling of the TCP layer and link layer.

For a delay-sensitive voice flow transmitted with UDP, limited retransmissions are considered. The retransmission limit is determined based on the voice traffic delay bound. If a link layer packet cannot be received by the base station successfully within the retransmission limit, this packet and the subsequent ones of the same UDP packet will be dropped. Retransmission limit is jointly determined with the received signal bit energy to interference-plus-noise density ratio (SINR) at the physical layer to guarantee both the delay bound and packet loss rate bound of the UDP packets.

Optimal Resource Allocation for a Single User

In a CDMA system, due to the nonorthogonal nature of the signals simultaneously transmitted from different users, the system capacity is interference-limited. Let (m_i, Γ_i) denote the resource vector for traffic flow i in a link layer frame, where m_i is the instantaneous packet number to be transmitted by orthogonal MC-CDMA (i.e., in a slot) and Γ_i is the SINR for all the m_i packets. We first study the capacity of a slot, under which constraint flows with different resource vectors may be allocated to transmit in the same slot. For simplicity, consider a single-cell environment here. Let N denote the number of active flows. This allocation is feasible in a time slot only when [7]

$$\sum_{i=1}^{N} \frac{m_i \Gamma_i}{G + m_i \Gamma_i} < 1;$$

where G is the processing gain. From the above inequality, if the total amount of resources in each slot is normalized to 1, we can define the amount of resources in a slot required by flow i as

$$C^s(m_i, \Gamma_i) = \frac{m_i \Gamma_i}{G + m_i \Gamma_i},$$

which is also the instantaneous resource amount of flow i in the frame.

The value of m_i can change from frame to frame; so can the resource amount $C^s(m_i, \Gamma_i)$. After weighting $C^s(m_i, \Gamma_i)$ by the steady-state probability distribution of m_i, we can obtain the average amount of long-term resources required by flow i, denoted C_i^e and referred to as the equivalent resource amount for flow i. One objective in this research is to guarantee the transport layer QoS with minimal C_i^e (translating to a minimal amount of resources required at the physical layer).

For a premium service voice flow i, a UDP packet in the transport layer is segmented into M_U link layer packets. When a UDP packet is generated, it requires that all the M_U link layer packets be delivered successfully to the base station within D_U link layer frames. The link layer scheduling policy for a UDP packet is as follows. In the first frame after the UDP packet arrival, these M_U packets are transmitted in a slot with SINR value Γ_U, which is a link layer design parameter and needs to be optimized. In each of the subsequent $D_U - 1$ frames, any packet will be retransmitted if not received successfully in the previous frame. If any of the M_U packets cannot be received successfully by the base station within the D_U frames, the corresponding UDP packet will be dropped.

The scheduling policy for an assured service flow i is: if it has packets to transmit, a fixed number (denoted M_i) of link layer packets are scheduled to be transmitted in each frame (with SINR value Γ_i). M_i is called the target number of scheduled packets for flow i, and is an upper bound of m_i. The link layer design parameter vector (M_i, Γ_i) for assured service flow i needs to be determined.

The retransmission limit is determined based on the voice traffic delay bound. If a link layer packet cannot be received by the base station successfully within the retransmission limit, this packet and the subsequent ones of the same UDP packet will be dropped.

Traffic type	CIR	Number of flows	τ	Optimal point		
				M_i	Γ_i	C_i^e
A	60 kb/s	10	200 ms	4	5.03	0.1269
		(IS: 1–10)	40 ms	4	4.98	0.1273
B	300 kb/s	10	200 ms	20	5.11	0.4114

Table 1. *Parameters for assured service data flows in the simulations.*

It can be seen that, a user's transport layer QoS (e.g., delay and packet loss rate for voice, throughput for data) is determined by the link layer design parameter (vector). After an exhaustive search, we can derive the feasible set of Γ_U's for a voice flow and (M_i, Γ_i)'s for a data flow that satisfy the transport layer QoS requirement. Among the feasible set, if we choose the design parameter (vector) that minimizes the equivalent resource amount, the optimal resource allocation is achieved. A detailed description of this procedure is not given here because of the mathematical complexity. A discussion of the data assured service case is given in [8].

MAC Packet Scheduling Algorithm

When different flows are multiplexed, the scheduling policies for all the flows may not be satisfied in each frame, so a MAC scheduler is essential to schedule packets from these flows. Many factors affect the design of a packet-switching MAC scheduler, such as traffic characteristics, QoS requirements, and resource availability. All these factors should be taken into account to achieve fairness and efficiency.

Similar to the case in wireline DiffServ networks, premium service voice flows are given priority in the proposed scheduler as they are delay-sensitive. In each frame, the scheduler guarantees to transmit all the link layer packets remaining in each voice flow's transmission queue (i.e., new packets and packets for retransmissions). This can be supported by an appropriate call admission control strategy. Data flows share the residual resources in each frame; thus, if C_v stands for the sum of the instantaneous resource amounts of all the premium service voice flows in this frame, all assured service data flows will share the remaining available instantaneous resource amount $L_R - C_v$ (L_R is the total instantaneous resource amount in each frame). Furthermore, if not all the expected instantaneous packet numbers (m_i for a voice or data flow i, obtained from the scheduling policies) from all the flows can be scheduled in a frame, we should determine a new feasible instantaneous number m_i^* of packets to actually be scheduled from each assured service flow, and in future frames try to compensate for the assured service flows that experience performance degradation.

In order to achieve both long-term and short-term fairness among assured service flows, we introduce a long-term fairness weight w_i^l and a short-term fairness weight w_i^s for assured service flow i, calculated based on information from the coupled higher layers.

Long-Term Fairness Weight — As the edge router of the DiffServ network, the base station has a marker to classify the incoming traffic from assured service flow i. It measures the transmission rate of this flow during a relatively long period and marks the arrival packets *in-profile* or *out-of-profile* based on comparison of the measured rate r_i and the rate requirement CIR_i. Here we choose the long-term fairness weight,

$$w_i^l = \frac{\mathrm{CIR}_i}{r_i}.$$

Obviously, as r_i decreases from the CIR value, this weight increases.

Short-term Fairness Weight — TCP, the transport layer protocol of assured service, avoids network congestion through controlling the sender's congestion window. The short-term transmission rate of TCP flow i can be represented by the ratio of the product of the instantaneous TCP congestion window (W_i) and TCP packet size (S_T) to the measured smoothed TCP RTT (λ_i) at the sender. Here we choose the short-term fairness weight to be

$$w_i^s = \frac{\mathrm{CIR}_i}{W_i \cdot S_T / \lambda_i}.$$

In each frame, for each premium or assured service flow i, the base station has the information of the expected instantaneous packet number m_i and optimal SINR value Γ_i. For a premium service flow, the m_i packets are guaranteed to be scheduled. For assured service data flow i, we define a weight $w_i = C^s(m_i, \Gamma_i) \cdot w_i^l \cdot w_i^s$ to share the available instantaneous resource amount $(L_R - C_v)$ for data flows in the frame, and can further estimate a new instantaneous number m_i^* of packets to be actually scheduled from this flow. If not all the m_i packets from all the premium service flows and all the m_i^* packets from all the assured service flows can be scheduled in the frame, we reduce the available instantaneous resource amount for data flows from $(L_R - C_v)$ to $(L_R - C_v)(1 - \delta)$, where δ is a small value. If all the packets are scheduled successfully, the remaining instantaneous resource amount in this frame is used to schedule extra assured service packets, based on the long-term fairness weights of the assured service flows (in decreasing order).

For the above scheduling, some information needs to be exchanged between the mobile user and the base station. For voice flow, the mobile user only needs to inform the base station of the beginning of a talk spurt or a silence duration. For data flow i, the short-term fairness weight w_i^s should be updated to the base station upon a TCP packet arrival at the transmission queue in the mobile user, as W_i and λ_i change at this rate. All the update information is transmitted in the reverse link request slot, using more powerful channel coding to avoid transmission collision, or piggybacked at the end of the transmitted information link layer packets (if any) to reduce contention in the request slot.

It can be seen that the packet scheduling procedure requires per-flow rather than per-packet information. Thus, the computation complexity and system overhead are expected to be reduced significantly compared to other scheduling schemes [2, 3], which use per-packet information to determine scheduling priority.

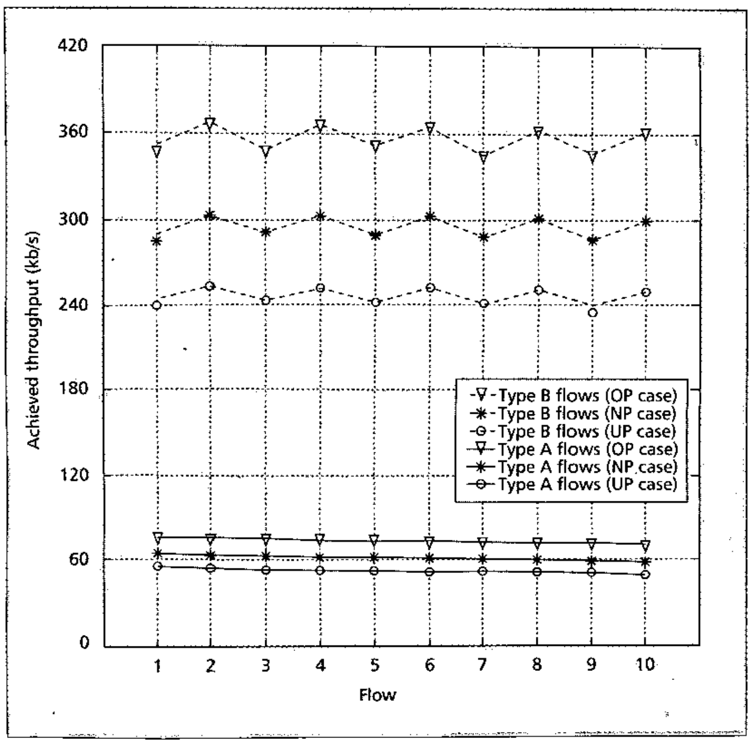
Performance Evaluation

The performance of the proposed vertically coupled protocol architecture is evaluated by computer simulations. The UDP transmissions of premium service voice flows and TCP transmissions of assured service data flows in a cell are simulated..In the link layer, the frame length is 10 ms, which includes eight packet slots.

Parameters of Assured Service Data Flows — We consider two types of data traffic with throughput requirements CIR = 60 kb/s (type A) and CIR = 300 kb/s (type B), respectively. As shown in Table 1, 10 long-lived TCP flows (with ID numbers ranging from 1 to 10) for each type are simulated; five (with odd ID numbers) have fixed wireline domain delay τ = 200 ms, and the other five (with even ID numbers) have τ = 40 ms. Each TCP flow runs on a link layer with unlimited retransmissions. Via numerical analyses for single users, the optimal link layer design parameter vector (M_i, Γ_i) is (4, 5.03), (4, 4.98), (20, 5.11), and (20, 4.97) for the four subtypes, respectively. The corresponding C_i^e values are also given in Table 1. It can be seen that the sum of equivalent resource amounts for all 20 data flows is 5.423. In the simulations, the achieved TCP throughput is traced through the accumulative ACKs from the correspondence node.

Parameters of Premium Service Voice Flows — Voice traffic is simulated based on the on-off model, with an average on duration 0.352 s and an average off duration 0.650 s. In the talk spurt (on state), a UDP packet is generated every five frames, which is further segmented into six link layer packets. The wireless delay bound for a UDP packet is D_U = 3 frames, and the UDP packet loss rate bound is 1 percent. The optimal Γ_U value 4.02 and the equivalent resource amount 0.0126 (derived from numerical analysis) are used for each voice flow in the simulations. Three cases of traffic load scenarios are considered:

- Underprovisioned (UP): 20 data flows and 250 voice flows are supported. The sum of the required equivalent resource amounts of all the flows is 8.6, larger than the available resource amount (i.e., 8).
- Near-provisioned (NP): 20 data flows and 204 voice flows are supported. The sum of the required equivalent resource amounts is 8, equal to the available resource amount.
- Overprovisioned (OP): 20 data flows and 150 voice flows are supported. The sum of the required equivalent resource amounts is 7.3, less than the available resource amount.



■ Figure 3. *Achieved TCP throughput for data flows in UP/NP/OP cases.*

The simulation for each case runs for 30,000 frames, and the statistics are collected in the last 24,000 frames. As the premium service voice flows are given priority in each frame, it is observed during all the simulations that their delay bound and packet loss rate requirement are guaranteed. Figure 3 shows the achieved TCP throughput for data flows (type A 1–10 and type B 1–10) in the UP, NP, and OP cases. It can be seen that each data flow achieves a throughput around its CIR in NP, and this throughput increases approximately 20 percent (decreases approximately 16 percent) in OP (UP). All small-CIR flows (type A) achieve almost the same performance, while the achieved throughput for large-CIR flows (type B) fluctuates in each case where flows with τ = 40 ms obtain higher throughput than flows with τ = 200 ms. This is consistent with the observation in previous research: the networks have a bias against flows with large RTTs and large CIRs [9]. When both RTT and CIR are large (e.g., type B traffic with τ = 200 ms), it takes more time for the TCP window to reach the steady state after reacting to packet losses, thus achieving a relatively small throughput. The above fluctuation (below 10 percent) should be acceptable, taking into account the fact that assured service is intended to provide a coarse QoS assurance level. Thus, the analytical equivalent resource estimation for voice and data flows is sufficiently accurate, and the proposed MAC packet scheduling scheme can achieve effectiveness and reasonable fairness.

Conclusion

In this article we propose a vertically coupled protocol architecture for provisioning QoS in 4G CDMA cellular networks. In the proposed verti-

Combined with a call admission control strategy, our scheme provides a solution to QoS-guaranteed DiffServ in future 4G cellular networks. This work also should provide helpful insights for call admission control in 4G networks.

cal coupling, optimal resource allocation can be achieved for voice and data traffic with transport layer QoS guarantee. A MAC packet scheduling scheme is proposed to achieve efficient statistical multiplexing, which is based only on per-flow information. Combined with a call admission control strategy, our scheme provides a solution to QoS-guaranteed DiffServ in future 4G cellular networks. This work should also provide helpful insights for call admission control in 4G networks.

Acknowledgments

This work was supported by the Premier's Research Excellence Award (PREA) from the Ontario Government.

References

[1] S. Blake *et al.*, "An Architecture for Differentiated Services," IETF RFC 2475, Dec. 1998.

[2] I. F. Akyildiz *et al.*, "A Slotted CDMA Protocol with BER Scheduling for Wireless Multimedia Networks," *IEEE/ACM Trans. Net.*, vol. 7, Apr. 1999, pp. 146–58.

[3] V. Huang and W. Zhuang, "QoS-Oriented Access Control for 4G Mobile Multimedia CDMA Communications," *IEEE Commun. Mag.*, Mar. 2002, pp. 118–25.

[4] J. Postel, "User Datagram Protocol," IETF RFC 768, Aug. 1980.

[5] S. Floyd and T. Henderson, "The NewReno Modification to TCP's Fast Recovery Algorithm," IETF RFC 2582, Apr. 1999.

[6] S. Floyd and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance," *IEEE/ACM Trans. Net.*, vol. 1, Aug. 1993, pp. 397–413.

[7] A. Sampath *et al.*, "Power Control and Resource Management for a Multimedia CDMA Wireless System," *Proc. IEEE PIMRC '95*, vol. 1, Sept. 1995, pp. 21–25.

[8] H. Jiang and W. Zhuang, "Quality-of-Service Provisioning to Assured Service in the Wireless Internet," *Proc. IEEE GLOBECOM'03*, Dec. 2003.

[9] M. Baines *et al.*, "Using TCP Models to Understand Bandwidth Assurance in a Differentiated Services Network," *Proc. IEEE GLOBECOM '01*, vol. 3, Nov. 2001, pp. 1800–05.

Biographies

HAI JIANG [S] received a B.S. degree in 1995 and an M.S. degree in 1998, both in electrical engineering, from Peking University, Beijing, China. He is currently working toward a Ph.D. degree at the University of Waterloo, Canada. His current research interests include QoS provisioning and resource management for multimedia communications in all-IP wireless networks.

WEIHUA ZHUANG [M'93, SM'01] (wzhuang@bbcr.uwaterloo.ca) received a Ph.D. degree in electrical engineering from the University of New Brunswick, Canada. Since October 1993 she has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where she is a 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 received the Premier's Research Excellence Award (PREA) in 2001 from the Ontario Government. She is an Associate Editor of *IEEE Transactions on Vehicular Technology*.