Supporting Digital Video in a Managed Wireless Network

Paramvir Bahl, Microsoft Corporation

ABSTRACT Several problems have to be overcome before personalized interactive video communication services over radio frequency networks can become a reality. The challenge of being able to support wireless video services is significant since video is generally recognized as being bandwidth-hungry, error-sensitive, and sometimes delay-intolerant; radio channels, on the other hand, are characterized as having limited capacity and high bit error rates, and being time-varying. Under such hostile conditions the author explores issues in supporting real-time digital video communications over infrastructure-based wireless networks. The guiding philosophy behind the work described is that robust wireless video communications is possible if the different components within the network, operating system, and application layers copperate with one another and with the overall system.

With the emergence of tetherless networks in homes, schools, hotels, airports, conferences, and other such places of congregation, and with society's increasing reliance on the easy availability of multimedia information on the Internet, we are surely moving toward widespread use of handheld wireless video communicators. Even today we can get off-the-shelf miniature video cameras, powerful low-power processors, high-density memory modules, miniature high-resolution displays, inexpensive batteries, and reasonably fast RF transceivers which when put together allow us to construct useful handheld wireless multimedia appliances. The technology trends are clear; practitioners are finding innovative ways of increasing efficiency while reducing size and system cost, which in turn is propelling the buildup of the necessary communications infrastructures. Researchers continue to work on improving both software and hardware efficiency by, for example, developing faster RF transceivers that employ novel antenna designs with improved noise reduction circuitry. Together, researchers and practitioners are cooperating and creating hardware and software standards that build on the shoulders of well-thought-out mechanisms and powerful algorithms in the hope that these standards will fuel the industry into developing inexpensive robust high-quality interactive wireless video communication products for the masses [1–3].

Supporting robust video communications over RF networks is a hard problem primarily because of three factors:
- Scarcity of bandwidth
- Time-varying error characteristics of the transmission channel
- Power limitations of the wireless devices

The problem becomes even harder, albeit interesting, if we include the repercussions of user mobility on network service guarantees. In our exposition of this subject we ignore the impact of user mobility and focus only on the characteristics of the wireless channel and the wireless device. We claim without elaboration that there are a significant number of multimedia applications that can benefit from explicit support for wireless video even when user mobility is limited to within the range of the wireless-to-wired access point for the lifetime of the connection.

Unlike wireline networks where the increase in the number of users and their demands can be met by adding more fiber or other similar wired media, the transmission capacity in wireless networks cannot be increased arbitrarily. Thus, in video communications, due to the large quantity of data involved, compression is almost always used in the management and movement of digital video over wireless networks. Unfortunately, the currently pervasive video coding standards are unsuitable for transmission over RF channels. Currently video communications are carried out using source coders and channel coders designed independent of each other. The tradition of separating source and channel coding is based on the solid theoretical foundation of Shannon's celebrated *separation principle*, which basically states that this separation is optimal [4]. For example, in a point-to-point transmission using a known time-invariant channel, one can design the best possible channel coding method to approach channel capacity; that is, achieve a rate R b/s such that $R \leq C$ where C is the channel capacity in bits per second. Then the task of the source coder is simply to do the best job it can in compressing the input signal so that the compressed bit rate will match the channel rate. However, when considering wireless video communications, there are compelling reasons not to adhere blindly to the separation principle. For example, Shannon's work makes no assumptions about the error characteristics of the channel on which the data would traverse; nor does it take into account the optimization possible in channel resource utilization through statistical multiplexing.

The problems arising from scarcity of bandwidth are aggravated by the fact that radio channels are highly unpredictable. This unreliability comes from errors that arise due to distinct propagation phenomena such as multipath fading, shadowing, path loss, noise, and interference from other users, all of which have a multiplicative effect on the transmitted signal, causing it to deteriorate. Multipath propagation, caused by superposition of radio waves reflected from surrounding objects, gives rise to frequency-selective fading, resulting in rapid fluctuations of the phase and amplitude of the signal. This usually happens when the receiver moves over a distance on the order of a wavelength or more. Shadowing, caused by the presence of large physical objects (buildings, walls, etc.) which preclude a direct line of sight between the radio transmitter and receiver, is a medium-scale effect: field strength variations occur when the antenna is displaced over distances larger than a few tens or hundreds of meters. The result is strong signal power attenuation. Path loss causes the received power to vary gradually due to signal attenuation determined by the geometry of the path profile in its entirety. All this is in addition to the local propagation mechanisms, which are determined by terrain features in the immediate vicinity of the antennas. The combined effect of these phenomena is that the receiver has to deal with a bitstream that is corrupted by both random bit errors and burst errors. Video coding techniques developed without regard to such time-varying channel impairments behave poorly because optimal codes within them are destroyed in such a manner that even with a few bits in error, the entire image signal is rendered useless to the end user. This is true for all currently popular first-generation video compression standards, including the International Organization for Standardization's (ISO) Motion Picture Experts Group (MPEG)-1 and MPEG-2 standards, and the International Telecommunication Union's (ITU) H.261, and H.263 v. 1 standards[1] — coders designed with little regard to the error characteristics of the channel [5–8].

Figure 1 illustrates the effect of random bit errors on an MPEG-2 coded video sequence. The first set of images show the phenomena of temporal propagation, where an error induced in the 16th frame continues to have an effect through the 31st frame of the video sequence. The second set of images illustrates the phenomena of spatial propagation, where due to corruption in the motion vectors an additional person appears in the original picture.

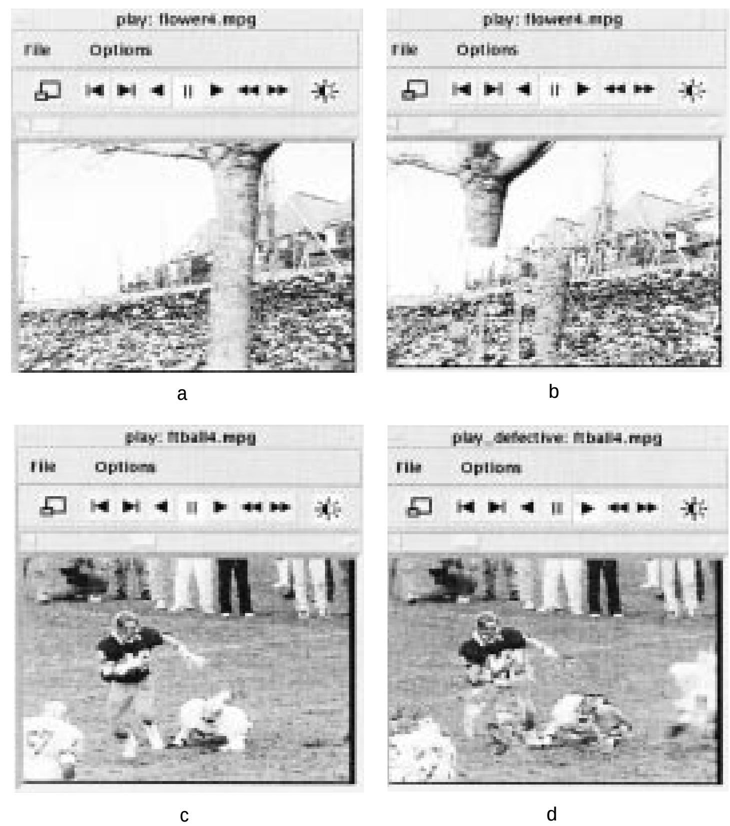


Figure 1. *The effect of temporal and spatial error propagation in current ISO and ITU video codecs: a) frame #16; b) frame #31; c) original; d) with spatial error propagation.*

Equalization techniques that reduce intersymbol interference can be used to alleviate some of the problems; however, these come at the expense of reduced system efficiency, increased system cost, and increased processing delay. Efficiency and processing delays are affected since most of these techniques depend on obtaining training sequences to learn the channel; and for time-varying RF channels the system is forced to adjust equalization often, potentially making the overhead unacceptable for time-critical applications. At a higher layer, error protection schemes that use sophisticated convolution or block codes may be employed to alleviate the error induction problem, but they aggravate bandwidth problems since several more bits have to be added to the video bitstream [9]. Similarly, automatic repeat request (ARQ)-type retransmission procedures, while improving error recovery against burst errors, aggravate latency (jitter) problems. While it is true that we cannot completely eliminate these overheads, we can reduce the dependency on such techniques while improving the final quality of the displayed video by letting the application and source encoder do their part. Thus, error resiliency in video applications via error detection, recovery, and concealment become critical requirements for a video encoder.

Current second-generation cellular and cordless communications standards are also inadequate for transporting compressed video over RF channels. The resource allocation algorithms, traffic scheduling mechanisms, and channel access protocols for these systems are biased toward integrated packet voice and data communications. Digital video communications was considered desirable but not essential for the success of these systems; consequently, its characteristics were not accommodated by the designers of these systems. Careful evaluation of the dominant time-division-multiplexing-based channel access schemes proposed in the literature reveals that there are four main reasons why the proposed medium access control protocols (MACs) perform poorly when digital video is incorporated into the wireless environment.

First, most of the popular channel access protocols are unable to guarantee sustained bandwidth, bounded delay, and hence quality of service (QoS) guarantees for video communications. QoS is an essential ingredient for the success of interactive visual communications; without it, under heavy loads video tends to exhibit poor and sometimes intolerable (choppy) quality.

Second, several prominent channel access schemes apply time-assigned speech interpolation (TASI) as the multiple

[1] *The most recent low-bit-rate video coding standard, H.263 v. 1, was originally developed for visual communications over plain old telephone service (POTS) and for public switched telephone network (PSTN) multimedia terminals. PSTN is characterized by low delay, an error rate typically better than 10^{-6}, and channel conditions that remain constant with time. For such an environment H.263 works reasonably well. However, wireless radio networks incur higher bit error rates (typically 10^{-2}) with time-varying channel characteristics. In such an environment we studied the performance of H.263 video and found that it performs poorly.*

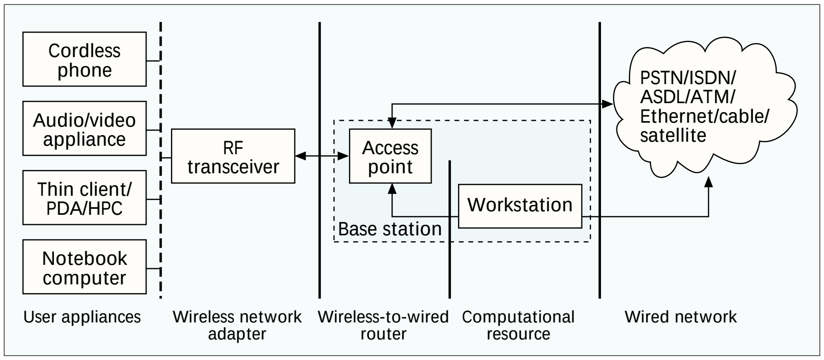


Figure 2. *Reference architecture for a managed wireless network.*

access method. The frame length is equal to the voice codec packet generation period. The motivation for this is that as soon as a voice connection succeeds in making a reservation, voice packets can be transmitted without any additional delay. The voice connection is guaranteed one slot in each subsequent frame for the duration of the talk spurt, and since one voice packet is generated in one frame time, both delay and buffering are bounded. Unfortunately, for video this choice of frame length has no meaning. A one-slot-per-frame guarantee is not useful for video because it does not prevent excessive delays, buffer overflows, and consequent degradation in visual quality.

Third, in almost all protocols dynamic reservations for voice packets, considered higher priority, is allowed while data packets, considered lower priority, have to contend for every slot. Real-time packet video falls somewhere in the middle — in general, it has higher priority than data and lower priority than voice. It wouldn't be appropriate to treat video as voice; the high priority coupled with the high demand for bandwidth would surely overwhelm network resources and degrade all ongoing connections. Similarly, it isn't appropriate to treat video as data; contention for every slot would lead to frequent collisions, causing excessive delays, resulting in lowering both the quality of ongoing connections and the overall bandwidth utilization of the system. Thus, video has to be treated as a separate entity with its own set of requirements.

Fourth, schemes that completely rely on contention to obtain slot positions perform poorly under heavy load. When video is introduced in a resource-strapped wireless network, the amount of data in the system is increased to the point where collisions are bound to escalate. Even a mixture of different permission probabilities for different traffic sources is not enough to alleviate this problem. Excessive collisions, as noted above, can drastically reduce system performance.

This lack of support in terms of absence of error-robust video compression algorithms, video-supportive transportation protocols, and video supportive resource management and control algorithms have rendered visual communications over wireless radio networks impractical in present-day RF communication systems.

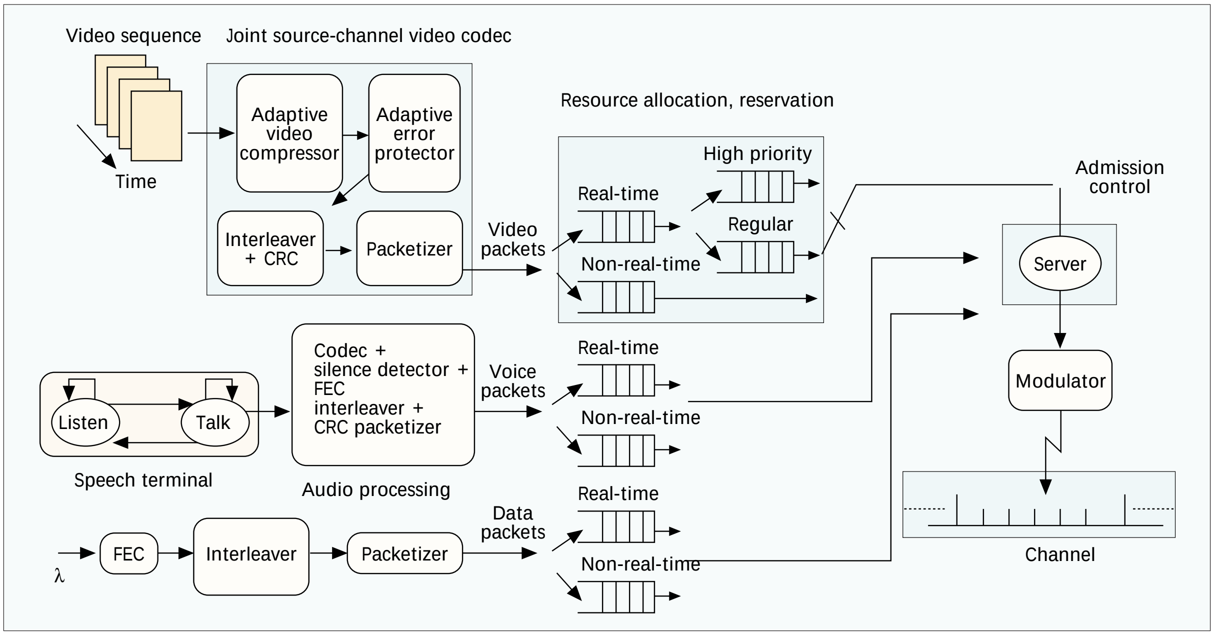
In this article we focus on the problem of providing high-quality interactive video communications over a managed wireless network. A managed wireless network, as opposed to an ad hoc wireless network, is one in which the control and management of the network is centralized. All wireless end nodes communicate with a central *access point* or base station which is connected to the wired network. Figure 2 illustrates the reference architecture for such a network. Implicit in this figure is our assumption that the wireless device is always within range of the access point and there is at least a "reasonable" amount of bandwidth available to accommodate voice, data, and compressed video traffic simultaneously. We assume that the mobility profile of users is characterized as low-speed sporadic movement and the usage environment is the user's home, or lounges in places such as airports, hotels, and hospitals, or rooms in places such as universities, conference centers, and theaters. We examine the specific challenges of video compression, bandwidth partitioning, bandwidth reservation, bandwidth allocation, bandwidth utilization, video traffic scheduling, and channel access protocols all of which are critical components in a wireless network that provides explicit support for video communications.

THE DEVICE MODEL

From a high level, we conceptualize the transmission path for voice, video, and data traffic in a wireless multimedia device as in Fig. 3. To combat the dual problems of low bandwidth and high random bit error rates with a high possibility of burst errors, each traffic class is suitably rendered for transmission before being handed over to the network stack. For example, in the case of digital video the incoming video sequence is compressed and split into orthogonal substreams with differing priorities. The various substreams are fragmented, packetized, and classified as appropriate. Header information, including connection identifier, virtual path identifier, priority identifier, and packet sequence number, is added to each packet. These packets are then handed over to the transport layer, which along with the network layer negotiates the bandwidth with the access point (base station), reserves resources for the prioritized substreams, sets up end-to-end connections, and manages the scheduling of connections according to time constraints. At the link layer, depending on their priority level, packets are made error-resilient through the use of powerful forward error correcting (FEC) codes (e.g., concatenated Reed-Solomon and rate-punctured convolution codes). To combat burst errors, the FEC-coded video is interleaved over multiple time slots. Voice and data connections are also processed in a similar manner to compensate for problems with the transmission channel. The voice subsystem is augmented with silence detection circuitry so that packets are transmitted only when the speaker is in the talking state. Data connections are multiplexed with voice and video connections. Real-time voice packets are given the highest transmission priority, non-real-time video connections the lowest. Real-time video packets are guaranteed a prenegotiated QoS. The channel access protocol ensures that network access is provided to all traffic classes without letting any single traffic class shut out the others. System bandwidth is managed by the bandwidth manager at the access point (or any other computing resource connected to the network) according to an algorithm that minimizes the maximum blocking probability for the different traffic classes.

CHANNEL-ADAPTIVE VIDEO CODECS

In acknowledgment of the anticipated importance of wireless video applications, standardization committees have been moving forward in creating audio-video standards that take into account the hostile nature of the wireless environment. For example, the ITU's current focus is on enhancing its low-bit-rate video compression standard [6] by adding negotiable operational modes such as spatial, temporal, and signal-to-noise ratio (SNR) scalability, error tracking, independent segment coding, and reference picture selection, all of which add



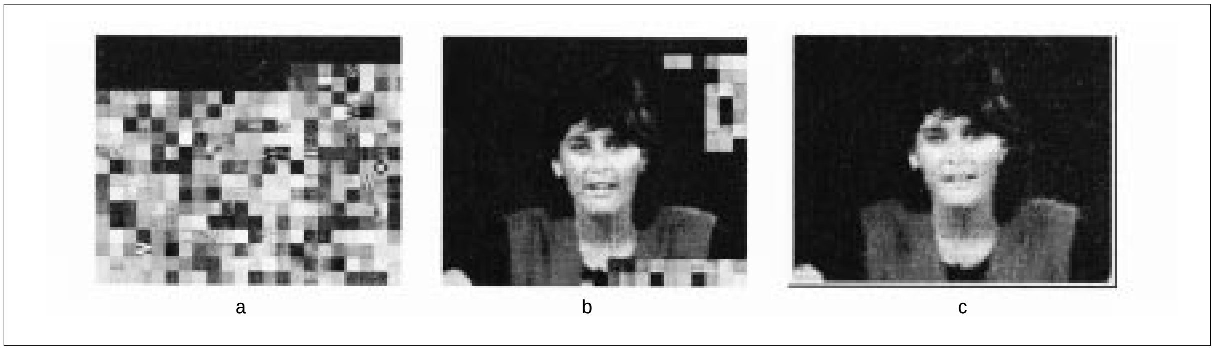
■ **Figure 3.** *Conceptual model for a video-capable wireless device.*

robustness to the video bitstream while minimizing delay latency [10]. Similarly, ISO's MPEG-4 committee is investigating options such as data partitioning, reversible variable-length codes, object segmentation, and resynchronization, which have been shown to improve transmission reliability over wireless channels [11, 12]. While these techniques are an important step toward alleviating some of the problems, more can be done.

Multiresolution decomposition of images, such as wavelet or subband decomposition, has drawn considerable attention in the past for the transmission of video over networks with dissimilar speeds [13]. Generally speaking, in this method a pair of operators splits the signal into two orthogonal parts with no overlap between them. This has an effect of splitting the time-frequency plane into two halves. The operators themselves are decimating operators, so the samples in the two parts add up to the total number of signal samples. One practical method of decomposing an image signal is to filter it into low- and high-frequency components using a conjugate pair of decimating quadrature filters. This procedure can be applied iteratively to produce as many parts (frequency bands) as desired. After such decomposition, frequency bands can be selectively dropped to match the bandwidth of the underlying channel. The decoder is able to reconstruct a lower-resolution image even if it does not receive most of the bands.

The general ideas of these methods can be applied for communications in wireless networks as well. With multiresolution decomposition, the coarse version of the image signal (containing the lower frequencies) can be better protected against transmission errors than the detail information. This form of unequal error protection is much more desirable than having to protect the entire coded image. However, this alone does not eliminate error propagation problems since each frequency band still has to be compressed, and the techniques available for doing that are the intra- and intermode compression strategies which come with known ramifications. Coding the primary (or most important) frequency band of sequential frames independently and coding the rest with respect to each other is one possible solution to limiting error propagation. Spatially dividing the image frame into regions and coding these separately according to their importance is another way to facilitate unequal error protection and limiting the effect of error propagation.

We have investigated this latter approach and found it quite promising [14]. Specifically, we combine content-sensitive spatial decomposition with multiresolution coding. We extract spatial information from video frames, creating regions that are then decomposed into subbands of different perceptual importance before being compressed and transmitted independently. The segmentation algorithm is an adaptation of a simple *split-and-merge* algorithm presented in [15]. After the initial application of the algorithm, subsequent segmentation is done only when there is significant motion activity in the frame; otherwise, the segmentation mask is reused for sequential frames. A low-complexity two-tap Harr filter provides a four-level decomposition of each region followed by the perennial motion compensation, quantization, entropy coding, and run-length coding techniques. The system applies unequal error protection, prioritized transmission, and reconstruction from incomplete data to guarantee a minimum spatial and temporal resolution at the receiver. When transmission errors cause corruption in regions, rendering them undecodable, or when dynamic reduction in system bandwidth causes some of the regions not to reach the decoder in a timely manner, the receiver is still able to reconstruct the transmitted frame from partial information. The complete frame is reconstructed at the receiver using a combination of the current and previous regions that were received correctly. The temporal difference between the original and substitute regions dictates how good or bad the final picture looks. When this difference is large, visual quality is impaired by a tearing effect; however, when the previous good region is from an immediately preceding frame, the quality is generally acceptable. With this observation, the tearing effect is reduced considerably by letting the receiver use the reverse channel to demand from the transmitter immediate transmission of the video regions it had to substitute in order to reconstruct the current frame. The transmitter in compliance adapts the transmission priorities of the regions in its queue to ensure that the requested video regions reach the receiver with the

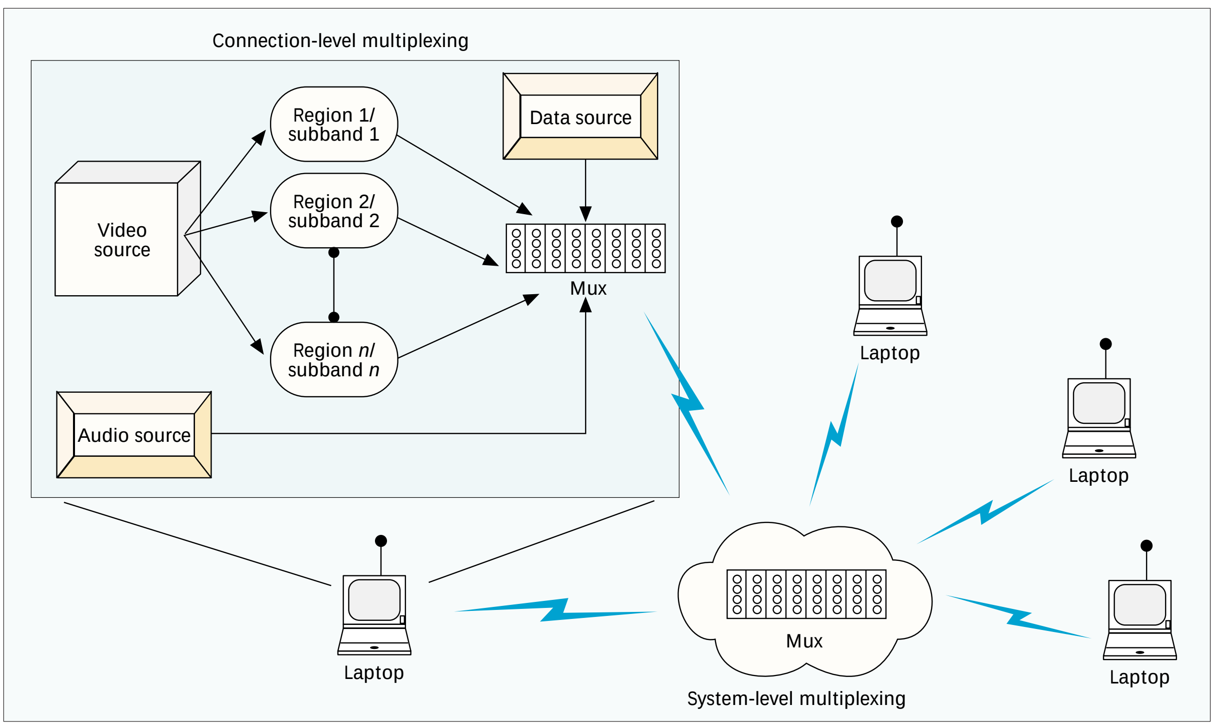


■ **Figure 4.** *Alleviating error propagation with spatial and frequency decomposition: a) bitstream corruption of the* Miss America *sequence in ITU's H.263; b) after region segmentation; c) iafter region segmentation, frequency segmentation, and prioritized transmission (PSNR: 3.2.34 dB).*

next transmission. With such a scheme in place, the difference between the current video region and those stored in the receiver's region store is never too great, and the tearing effect is mitigated.

With content-sensitive segmentation the system adaptively controls the bit rate allocation, transmission priorities, and error protection within each video frame. Region segmentation bounds both spatial and temporal error propagation within frames and, when combined with an inter-region statistical multiplexing scheme (next section), ensures optimal utilization of the reserved transmission bandwidth. This technique exhibits better error concealment, better temporal resolution, and better bandwidth utilization properties than achievable by current standards [14]. In Fig. 4 we illustrate the power of this technique.

Before concluding this subsection, we point out that the emerging ISO MPEG-4 v. 1 standard defines the notion of video objects (VOs) and their temporal instances, video object planes (VOPs), which when put together form video sequences. Although the initial motivation behind creating VOs was to provide increased flexibility for editing purposes, they can as easily be used for improving the transmission of video over wireless networks. Specifically, VOs constituting the video sequence may be ranked according to their importance. This ranking can happen automatically (e.g., by determining the amount of motion activity associated with the object) or statically (e.g., by ranking the head and shoulders of a talking person over the background). The priority given to the VO affects the subsequent processing of the bitstream, including coding, protection, bandwidth reservation, packet scheduling, transmission, and reconstruction. The difference between MPEG's VOs and our video regions is that VOs have a much stronger semantic meaning than video regions. For this reason extracting VOs from a sequence is much more complicated, compute-intensive, and consequently power-draining than the



■ **Figure 5.** *Comibining connection-level and system-level statistical multiplexing.*

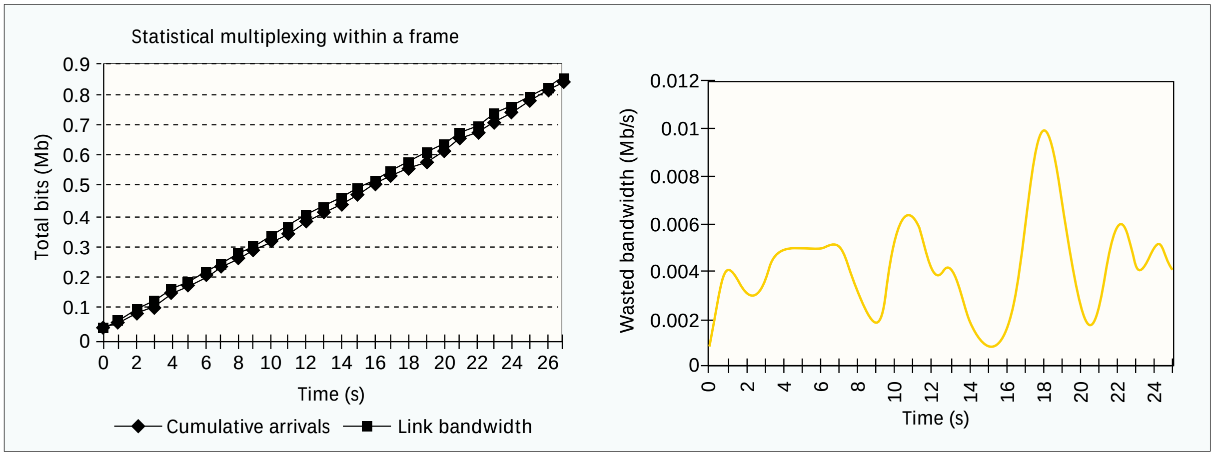


Figure 6. *Intraframe statistical multiplexing for a region-segmented video codec.*

simple *split-and-merge* algorithm we used. MPEG-4 v. 1 standardizes neither the segmentation method nor the error concealment techniques. These are left to the different vendors and are a "value-added" feature for individual offerings.

Resource Reservation and Utilization

It is a well-accepted fact that variable bit rate (VBR) video codecs are network-unfriendly. On one hand, since VBR traffic is delay-sensitive, a resource reservation scheme seems to be the right choice; on the other hand, because VBR video is unpredictably bursty, if resources are reserved according to peak rates the network is underutilized when the peak to average-rate ratios are high. These two opposing characteristics have resulted in a common belief that it is unlikely that performance guarantees can be provided to such bursty sources with very high network utilization.

Region segmentation and MPEG-4's VO approach can make VBR video codecs network-friendly. Specifically, the peak bit rate for the most important region or VO in an image is reserved at connection establishment time. It is likely that most of the time the compressor will produce bits far below this peak number. To avoid underutilizing and wasting allocated bandwidth, we have introduced the notion of inter-frame (or inter-region) statistical multiplexing [16]. The bandwidth left over after the primary region has been transmitted is used to transmit the remaining regions. Also, packets whose retransmission has been requested by the receiver (as in ARQ schemes) are sent using the leftover bandwidth that was reserved for the primary region. In essence, the idea is to combine statistical multiplexing at the system level with statistical multiplexing at the connection level to achieve optimum bandwidth utilization. Figure 5 illustrates the conceptual model for this mode of operation.

To corroborate our claim we segmented each image of the *Miss America* sequence into five regions and then compressed each region using the multiresolution motion-compensated algorithm outlined in the previous section. Bandwidth was reserved for the primary (head and shoulder) region and was equal to its peak rate of 24 kb/s. The average bit rate for this region was 8 kb/s. The second, third, and fourth regions had an average bit rate of 3, 2.5, and 2.8 kb/s, respectively. Transmission priority was given in the order of the lowest frequency subband of the main region, followed by the lowest frequency subband of the remaining regions, followed by the subsequent higher-frequency subbands. Any bits left over after the reserved bandwidth was used up were transmitted using any available unreserved bandwidth. It should be noted, as pointed out previously, that in steady state the video decoder can reconstruct the full frame whenever the primary subband of the main region is received. Figure 6 illustrates the difference in the amount of bandwidth reserved and used. In all our experiments the reserved bandwidth is completely used, and all transmitted image frames were displayable since the primary region always reached the receiver.

Resource Partitioning and Commitment

The dichotomy in the strategies for assigning resources is fixed versus dynamic assignment. In fixed assignment a connection for a particular traffic class can only be accepted if the bandwidth assigned for that class is available. In dynamic bandwidth assignment, other traffic classes are allowed to use (borrow) with various strategies. Many such strategies have been proposed, with a trade-off between complexity and performance [17]. In either case, the traffic-carrying capacity of the wireless network is strongly dependent on the efficiency of the bandwidth allocation scheme. For a fixed allocation scheme the efficiency is directly linked to system capacity, and for dynamic use of bandwidth a good nominal allocation reduces the need to borrow, and thus, the penalty associated with borrowing (locking out bandwidth for other traffic classes).

At the two extremes of resource partitioning strategies are the Complete Sharing (CS) and Complete Partitioning (CP) (also called Mutually Restricted Access or MRA) strategies, and in between are the rest, generally referred to as *hybrid* strategies. Figure 7 illustrates the three general bandwidth-partitioning strategies, and Table 1 lists the differences between them. The reader is referred to [18] for a more comprehensive survey of such schemes.

A natural question to ask is, which of the three is the best scheme for guaranteeing QoS for visual communications? Clearly CS is not suitable. CP, on the other hand, can deliver, but as noted in Table 1 is wasteful of bandwidth. Priority Borrowing is thus the most viable candidate. We have extended Priority Borrowing to include static bandwidth reservation and called it *Priority Sharing with Restrictions* (PSR) [17]. Figure 8 illustrates this scheme.

At connection establishment time bandwidth for the main subband or region of the video frame is allocated from the *Reserved* portion the available spectrum. The amount of bandwidth reserved by the system for such static reservations is determined as in [17]. The remaining spectrum is divided among voice, data, and video (for secondary and tertiary subbands or regions) users and is used for dynamic burst-level reservation. In terms of priority, voice users are given the highest priority, followed by interactive-video and data users, in that order. Table 2 shows an example of who can borrow from whom.

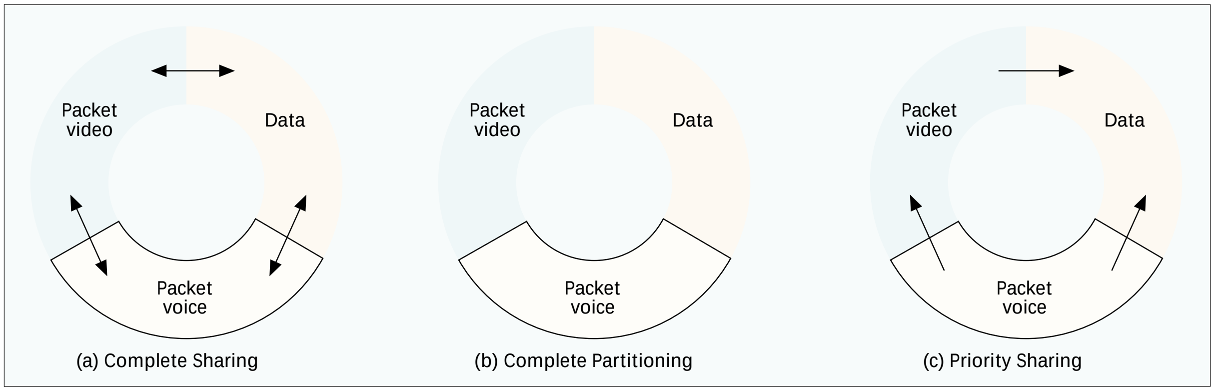


Figure 7. *Different flavors of bandwidth partitioning schemes.*

Complete Sharing	Complete Partitioning	Priority Sharing
No protection from overloading of other classes	Complete protection from overloading of other classes	Protection from overloading of other classes can be built in
Blocking probability for each traffic class is not adjustable	Blocking probability for each traffic class is easily adjustable	Blocking probabilities are tunable, depending on the flavor of the PS scheme
Bandwidth-efficient	Bandwidth-inefficient	Bandwidth-efficient

Table 1. *Comparing bandwidth-sharing policies.*

After deciding on the allocation strategy, finding the optimal partitioning point is a very difficult task since it can be directly modeled by an NP-complete graph-coloring problem. Because of the algorithmic intractability of finding the exact optimum, various suboptimal solutions have been proposed in the literature [18]. In fact, the intractability of finding the optimum is present already in the simplest situation, when the traffic consists of voice calls only and the statistics of the offered traffic class are completely known. The problem becomes even more difficult when the wireless network carries integrated nonhomogenous traffic, a situation occurring naturally in wireless multimedia networks. In this case estimating the blocking probability of connections and its application in resource allocation strategies is complicated since there are no closed formulas that can easily be applied for optimal resource partitioning, and it is unreasonable to rely on advance knowledge regarding the detailed statistical properties of traffic classes since their deviation from well established models is not uncommon. Consequently, any solution to the problem of system-level resource partitioning among different traffic classes under incompletely known conditions should have at least the following properties:

- It should be robust and insensitive to statistical assumptions, depending only on measurable quantities (e.g., average rates of the aggregated traffic flow for different classes) and not on any detailed statistics of the traffic mix and/or arrival process.
- The allocation should be based on minimizing a bound on the connection blocking probabilities that is proven to be asymptotically optimal.

From a practical viewpoint, insensitivity to traffic characteristics is highly advantageous since detailed statistical information is typically unavailable or uncertain, and optimality is important because it signifies that for large systems it is sufficient to know aggregate flow rates. In [17] we propose a bandwidth partitioning mechanism that has both these properties. In terms of implementation, the algorithm is a simple iterative process that converges to the optimal solution at a geometric rate of convergence, making it well suited even to the case when the aggregate load values change and the bandwidth allocation has to be recomputed from time to time.

Traffic Scheduling and Channel Access

Having proposed an efficient channel adaptive video codec along with complimentary bandwidth reservation, utilization, and partitioning strategies, we focus our attention next on channel access and an algorithm for traffic scheduling which takes into account the specific needs of time-bounded video traffic in conjunction with the traditional voice and data traffic. While there have been a plethora of research articles written on the subject of medium access control (MAC) protocols and many different schemes have been proposed, relatively few have given importance to the characteristics of the video traffic generation process. In this subsection we describe a variation of the popular time-division multiple access (TDMA) protocol that, together with an adaptive slot-scheduling algorithm, provides multimedia connections with timely access to the channel.

Our protocol, called *Adaptive Reservation Multiple Access Protocol* (ARMAP), supports the transmission of multiple traffic classes including voice, data and video, providing QoS guarantees to video and a high priority to voice connections. Specifically, ARMAP allows both static and dynamic bandwidth reservations. Static reservations are lifetime reserva-

	Data	Voice	Video-dynamic	Video-static
Data	—	BP	BP	BP
Voice	B	—	B	BP
Video-dynamic	B	BP	—	BP
Video-static	X	X	X	—

B Borrowing allowed
BP Borrowing allowed, preemption possible
X Borrowing not allowed
— Don't care

Table 2. *Rules for priority sharing with restrictions.*

tions made at connection establishment time and are for video connections only. The amount of bandwidth to reserve, for example, may be equal to the estimated peak bit rate of the primary subband of the main region or main VO (see previous section). Bandwidth for static reservations comes from the video-static portion of PSR (Fig. 8 and Table 2) and is necessary to guarantee spatial and temporal resolution of the video signal at the receiver. Dynamic burst-level reservations are for improving the visual quality of the images. Real-time voice connections also reserve bandwidth dynamically at the beginning of each talk spurt and are guaranteed zero transmission delay if their reservation request is accepted. A novel slot scheduling algorithm provides timely and contention-free channel access for ongoing video connections. Briefly, the algorithm is based on monitoring and subsequently exploiting the regularity in the packet generation process for each individual video connection, providing them with a special minislot for requesting transmission slots. The strength of the algorithm is that these minislots are provided in synch with each node's video compression cycle (defined as the process of capturing, compressing, and packetizing a video frame). When sufficient network resources are available, the scheduler provides reservation minislots to the transmitter at a rate equal to the video compression cycle rate; when resources are insufficient, the transmitting node scales down its video compression cycle rate to match the reservation minislot generation rate. This dual adaptation ensures optimum bandwidth usage and energy conservation, since valuable battery energy isn't wasted in capturing and compressing video-frames that can't be transmitted.

Simulation reveals that ARMAP achieves a promising combination of bandwidth efficiency and QoS for time bounded video traffic [19].

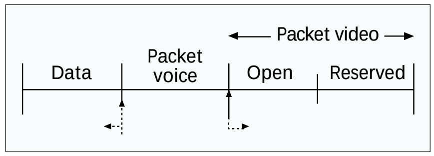


Figure 8. *Priority sharing with restrictions.*

Software Architecture

The software architecture corresponding to the conceptual model of Fig. 2 and to the mechanisms and algorithms discussed so far is illustrated in Fig. 9. The network is assumed to be a heterogeneous mixture of technologies with the wireless segment being the primary bottleneck toward providing guaranteed QoS services to video applications.

Looking at the architecture of the wireless end node in Fig. 9, a video application uses a modified socket interface to send control messages such as connection setup and teardown, bandwidth negotiation and reservation, and delay specifications to the receiver. The format and semantics for these requests is similar to Resource Reservation Protocol (RSVP) messages [20]. The connection establishment request carrying QoS descriptors is forwarded to the Admission Controller at the access point. The Admission Controller decides to accept or negotiate with the requester in consultation with the Bandwidth Manager, which determines whether enough resources are available to accommodate the request, and in consultation with the Security Manager, which authenticates the requestor. The Admission Controller at the access point may reject the request if not enough bandwidth is available on the wireless segment. In the case of wired segments alternate paths may be explored to determine if the request can be accommodated elsewhere. If accepted, the request is forwarded onto the Internet passing through several Admission Controllers across the different networks. End-to-end network connections are established after the acceptance and commitment of resources by the appropriate network entities between the sender and receiver. Once the connection has been established time-bounded audio-video packets are sent using a version of the Real Time Protocol (RTP) [21] modified to accommodate the particular video codec being employed. A Packet Classifier module maps the application

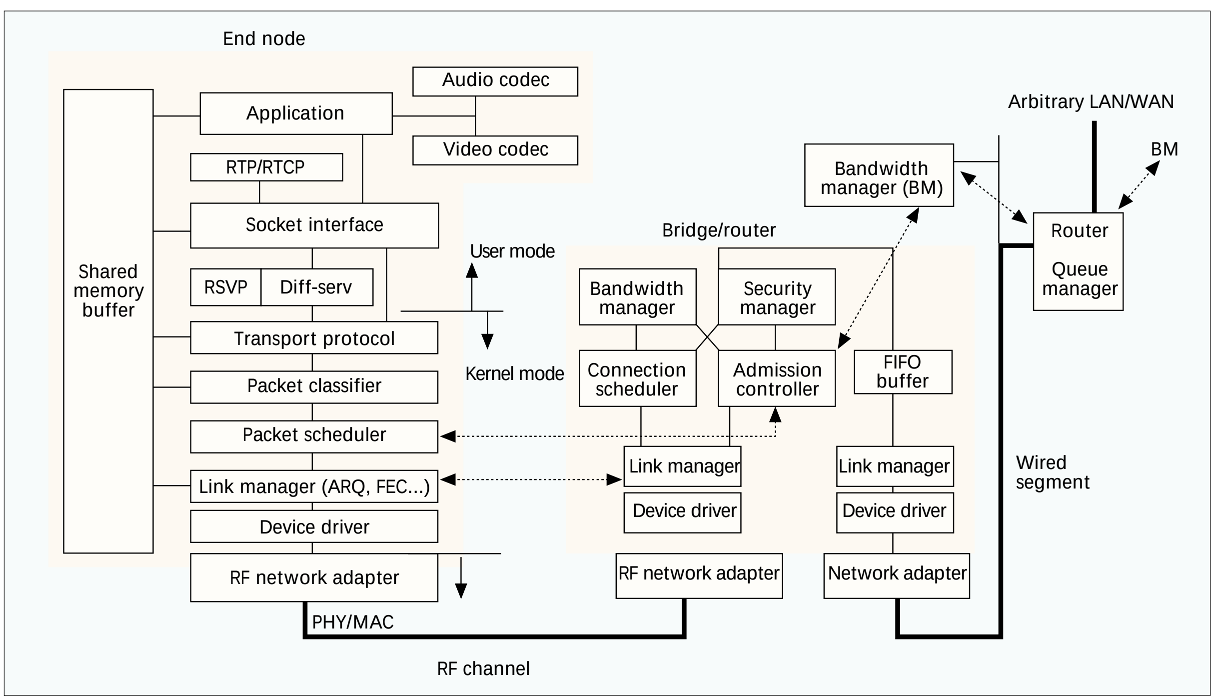


Figure 9. *Software architecture for packet flow from a wireless node to a wired segment.*

(video codec) specified priorities to the network-supported priority classes. The classifier may use precedence bits provided in the IP header and/or the MAC header to achieve this in a manner similar to supporting differentiated service classes [22]. Once each packet has been classified, the Packet Scheduler schedules the packets for transmission on the RF interface according to their priorities (Fig. 2). The Connection Scheduler module at the access point runs the slot-scheduling algorithm described in the previous subsection.

In terms of implementation, the Bandwidth Manager software may reside on any workstation, router, or switch on the same subnet. The Bandwidth Manager for the wireless segment manages the allocation, partitioning, and commitment of bandwidth according to the algorithms described earlier. For performance reasons the memory buffer containing the audio-video data is shared between the kernel and the user space and is accessible to the application and the different layers of the network stack.

CONCLUSIONS

The thesis of this work is that personalized interactive wireless video communication services are possible if the application, operating system, and network layers in the end node cooperate with one another and with the system as a whole. In this article, we touch on a variety of issues that are critical to the successful management and timely delivery of video over RF networks. In particular, we presented a content-sensitive region-based multiscale motion-compensated video coding algorithm that achieves joint source/channel coding in the delay as well as loss and corruption dimensions. By allowing region-by-region reconstruction, the perceptual delay of the video becomes less than the worst-case network delay, and the traffic capacity is increased because of the worst-case delay relaxation. Hand in hand with this codec, we presented a bandwidth reservation and statistical multiplexing scheme that allows the transmitter to send critical sections of the video data needed to achieve a guaranteed temporal and spatial resolution at the receiver without underutilizing the reserved bandwidth.

At the system level, we address the problem of balancing the needs of the different traffic classes with the need of the system to maximize the number of ongoing connections by partitioning network bandwidth appropriately. Next we present a slot-scheduling algorithm that exploits the inherent periodicity in the video packet generation process to provide energy-conserving, timely, and contention-free access to the channel. Finally, we present a brief overview of the software architecture corresponding to the ideas presented in this article. As we conclude, it is important for us to mention that in our exposé of this subject we do not discuss several additional areas currently under rigorous investigation by researchers that are relevant to wireless video transmission. Notable are the recent advances made in channel-adaptive error control and rate control strategies, energy-conserving algorithms, and software and hardware implementation details of video communicators. Furthermore, mobility management in wireless networks can affect system design across the board. Comprehensive coverage of all such issues would require a much longer article.

REFERENCES

[1] P. W. Baier, P. Jung, and A. Klein, "Taking the Challenge of Multiple Access for Third-Generation Cellular Mobile Radio Systems — A European View," *IEEE Commun. Mag.*, vol. 34, no. 2, Feb. 1996.

[2] J. S. Dasilva *et al.*, "European Third-Generation Mobile Systems," *IEEE Commun. Mag.*, vol. 34, no. 10, Oct. 1996, pp. 68–83.

[3] ETSI RES-10-12, "High Performance Radio Local Area Networks (HIPERLAN) Type 1, Functional Specification," May 1997.

[4] C. E. Shannon, "Communications in the Presence of Noise," *Proc. IRE*, vol. 37, Jan. 1949.

[5] ITU Rec. H.261, "Video Codec for Audiovisual Services at P x 64 kbits/s," CDM XV-R 37-E, CCITT, 1990.

[6] ITU Rec. H.263, "Video Coding for Low Bit-Rate Communications," ITU-T SG 15, 1996.

[7] ISO/IEC 11172-2, "Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to about 1.5 Mbits/sec: Part 2: Video," Aug. 1993.

[8] ISO/IEC 13818-2, "Information Technology — Generic Coding of Moving Pictures and Associated Audio," 1995.

[9] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*, Prentice Hall, 1983.

[10] N. Färber, B. Girod, and J. Villasenor, "Extensions of ITU-T Recommendation H.324 for Error-Resilient Video Transmission," *IEEE Commun. Mag.*, this issue.

[11] A. Puri and A. Eleftheriadis, "MPEG-4: An Object-based Multimedia Coding Standard for Supporting Mobile Applications," *ACM Mobile Networks and Apps. J.*, 1998.

[12] R. Talluri, "Error Resilient Video Coding in ISO MPEG-4 Standard," *IEEE Commun. Mag.*, this issue.

[13] *Proc. IEEE*, Special Issue on Wavelets, Jelena Kovacevic and Ingrid Daubechies, Eds., vol. 84, no. 4, Apr. 1996.

[14] P. Bahl and W. Hsu, "Adaptive Region-Based Multi-Scaled Motion-Compensated Video Coding for Error Prone, Bandwidth-Limited Communication Channels," *Proc. Conf. Broadband Networking Technologies, SPIE's Int'l. Symp. Voice, Video and Data Communications*, Dallas, TX, Nov. 1997.

[15] C. Diab, R. Prost, and R. Goutte, "Block-Adaptive Subband Coding of Images," *ICASSP*, 1990, pp. 2093–96.

[16] P. Bahl and I. Chlamtac, "H.263 Based Video Codec for Real-Time Visual Communications over Wireless Radio Networks," *Proc. IEEE ICUPC*, San Diego, CA, Oct. 1997.

[17] P. Bahl, I. Chlamtac, and A. Farago, "Resource Assignment for Integrated Services in Wireless ATM Networks," *Int'l. J. Commun. Sys.,* Special Issue on Personal Communication Systems, John Wiley, May 1998.

[18] *IEEE Commun. Mag.*, Special Issue on Bandwidth Allocation in ATM Networks, vol. 35, no. 5, May 1997.

[19] P. Bahl, "ARMAP — An Energy Conserving Protocol for Wireless Multimedia Communications," PIMRC '98, Boston, MA, Sept. 1998.

[21] H. Schulzrinne *et al.*, "RTP: A Transport Protocol for Real-Time Applications," Internet draft, Jan. 1996.

[20] R. Branden *et al.*, "Resource ReSerVation Protocol (RSVP) — Version 1 Functional Specification," RFC 2205, Sept. 1997.

[22] Baker *et al.*, "IP Precedence in Differentiated Services," Internet draft, Apr. 1998.

ADDITIONAL READING

[1] *IEEE Commun. Mag.*, Special Issue in Bandwidth Allocation in ATM Networks, vol. 35, no. 5, May 1997.

[2] *IEEE Pers. Commun.*, Special Issue on Wireless ATM vol. 3, no. 4, Aug. 1996 and *IEEE Commun. Mag*., Special Issue on Mobile and Wireless ATM, vol. 35, no. 11, Nov. 1997.

BIOGRAPHY

PARAMVIR BAHL [SM] (bahl@microsoft.com) holds a Ph.D. in computer systems engineering from the University of Massachusetts Amherst. He is currently with Microsoft Corporation, where he is carrying out research in peripatetic computing, multihop multiservice ad hoc networks, mobility-aware self-configuring operating systems, and real-time audio-visual communications. Prior to Microsoft, he spent nine years at Digital Equipment Corporation where he initiated, led, and contributed to several seminal multimedia projects, including the industry's first desktop hardware adapters for audio/video compression, the world's fastest software implementations of ISO and ITU audio/video codecs, and fast scalable image and video rendering algorithms. In 1994 he received Digital's multiyear Engineering Fellowship Award for carrying out research in real-time visual communications over wireless networks. He is vice chair of ACM SIGMOBILE and founding editor-in-chief of *ACM Mobile Computing and Communications Review*. He serves on the editorial boards of the *IEEE Journal on Selected Areas in Communications* (*JSAC*), *Wireless Communication Series*, and has served as guest editor for *ACM Mobile Computing and Applications* and *JSAC* on special topics related to wireless multimedia communications. He serves on a number of organizing and technical program committees of IEEE and ACM conferences including MobiCom, INFOCOM, PIMRC, MoMuC, and MMT.